

Transfer Learning for Leisure Centre Energy Consumption Prediction

P Banda,¹ M.A Bhuiyan,¹K Zhang¹ and A Song¹

RMIT University, Melbourne 3000, Australia

Abstract. Demand for energy is ever growing. Accurate prediction of energy demand of large buildings becomes essential for property managers to operate these facilities more efficient and greener. Various temporal modelling provides a reliable yet straightforward paradigm for short term building energy prediction. However, newly constructed buildings and recently renovated buildings, or buildings that have energy monitoring systems newly installed, do not have sufficient data to develop accurate energy demand prediction models. In contrast, established buildings often have vast amounts of data collected which may be lying idle. The model learned from these buildings with huge data can be useful if transferred to buildings with little or no data. An ensemble tree-based machine learning algorithm and datasets from two leisure centres and an office building in Melbourne were used in this transfer learning investigation. The results show that transfer learning is a promising technique in predicting accurately under a new scenario as it can achieve similar or even better performance compared to learning on a full dataset. The results also demonstrated the importance of time series adaptation as a method of improving transfer learning.

Keywords: Transfer Learning · Energy Use Prediction · Leisure Centre

1 Introduction

Efficient use of energy is undoubtedly a subject of great importance in sustainability as the increase of world population and economic growth keep adding pressure on energy supply. One key area of efficient energy management is in the building sector. According to [23], almost 39% of total energy consumption in the US is from buildings. China's energy usage by buildings is expected to reach as high as 35% by 2030 [5]. In Europe, buildings account for 40% energy usage which is equivalent to 36% CO₂ emissions [18]. Recreational facilities are attracting attention because of the increased awareness of health and fitness in modern lifestyle [6]. Sports facilities account for 8% of the total building energy usage in Europe [8]. Leisure centres, for example in Australia, often offer an array of different activities under one roof, such as swimming pools (indoor and outdoor), physical fitness centres, spas, and children's play park. Such arrangement exhibit complex and high energy use profiles that present energy management challenges for building managers. Accurate prediction of energy use at these

leisure centres is of paramount importance for the building managers or owners as it enables them to make informed decisions to manage better and optimise the operational performances of their buildings.

Widely used techniques in the literature for building energy consumption prediction, are broadly classified under engineering methods (white box methods such as EnergyPlus by [7], and DOE-2 by [3], statistical methods, and artificial intelligence (black box) methods [1]. Comprehensive discussions on the above techniques and their advantages and disadvantages can be found in recent reviews [9], [10] and [22]. In this study, we introduce transfer learning to facilitate the prediction process. The main thrust of transfer learning is the notion that it ignores the condition that training and testing data must obey the same distribution. Just like human beings can acquire knowledge while learning tasks and leverage that knowledge to solve related tasks. Transfer learning operates intuitively on the same principles. Thus, by utilising knowledge gained from one task, transfer learning overcomes the isolated learning limitation in the traditional machine and deep learning methods [20],[21] and [14].

In another work [16], investigated energy forecasting in the context of cross-building transfer with limited historical data by leveraging data from other buildings. In this work, reinforcement learning algorithms were combined with a deep belief network to improve the former's continuous state estimation capabilities.

In another research, [12] developed a transfer learning methodology for residential buildings climate control. They developed a generalized online transfer learning algorithm which leveraged forecasting knowledge from the source data to enhance the prediction of the target house. Their work utilised simulated residential houses created in EnergyPlus, as a test bed for the developed online transfer methodology and yielded positive results in the period within the first five weeks of the target dataset.

Recently, [19] proposed an inductive transfer learning algorithm that is sensitive to seasonality and trends present in electricity consumption data. The algorithm is applicable in a supervised transfer learning setting, that is, it requires limited data from the target building, and its extent of operation is bound only to similar buildings. A prediction accuracy increase of up to 11.2% using data from additional schools was reported on the target school with only one month of data.

This work is the first to explore transductive transfer learning for building energy studies, using different building types (leisure centres and office buildings) in a supervised learning setting. The most skilful of the five machine learning methods developed and evaluated on the task of building energy consumption prediction at two leisure centres and an office building is selected for transfer learning. The work investigates the application feasibility of transfer learning and how to improve its performances for improved energy consumption prediction using limited measured data. An ensemble tree based algorithm is tested in the transfer learning task to transfer knowledge amongst buildings with different energy consumption distributions. The work presents some of the initial results of the ongoing transfer learning experiment with a much broader scope.

2 Transductive Transfer Learning

Given a source domain, a target domain and a learning task, energy consumption prediction in our case, transfer learning aims to help improve the learning of the target predictive function in a new leisure center (Don Tatnell) using the knowledge in another leisure center (Waves) and an office building. In the traditional machine learning, to ensure the accuracy and high reliability of the model obtained by training, there are two basic assumptions: (1) the training sample used for learning and the new test sample satisfy the condition of independent and identical distribution; (2) There must be enough training samples available to learn a good model. While in transfer learning these assumptions are no longer necessary.

2.1 Predictive Algorithms

This section gives a brief overview of the adopted predictive algorithms for the energy consumption prediction exercise. A total of five predictive algorithms that is, decision trees, random forest(RF), lightGBM, k nearest neighbour(k-NN)and ensemble extra trees(EET) are considered for the input-output mapping task.

Random forest is an ensemble based learning algorithm used for both regression and classification problems [4].RF is an ensemble of models which uses the decision tree approach to data collection. Initially, an individual tree is trained by taking note of a random subset of observations. A random subset of the variable is then considered to split the decision thus creating a diverse set of trees essential for improving overall prediction performance of the ensemble model.

LightGBM is a recently launched algorithm, which by using histogram-based algorithms buckets continuous feature values into discrete bins. This enhances training and results in reduced memory usage. While most decision tree algorithms grow their trees by level (depth)-wise, LightGBM instead grows trees leaf-wise (best-first). The leaf with max delta loss is chosen to grow. However, when data is small. Leaf-wise often results in over-fitting. More details on the model performance are found in [15].

The k-nearest neighbour algorithm is one of the most straightforward supervised learning regression algorithms to implement but gives highly competitive results. The k-NN algorithm's primary assumption is that similar things exist nearby. The algorithm computes a distance value between the output and each item in the training data-set. The k-NN then picks the items with the k lowest distances and conduct a "majority vote" among those data points. The value of k is determined by either by trial and error or by cross-validation to find an optimal value [2].

Decision tree takes the form of a tree-like structure where numerous aspects and attributes are considered to predict the electrical energy demand. For evaluation, a recursive algorithm is used to identify the attributes with the highest information [13].

Ensemble extra trees (extremely randomised trees) implement a meta estimator that fits randomised decision trees on various dataset sub-samples and

then averages them to have improved predictive accuracy and at the same time to control over-fitting. Ensemble extra trees differentiate itself from other tree-based ensembles, in that it splits nodes by selecting cut-points entirely at random and incorporates the whole learning sample as opposed to a bootstrap replica to grow the trees. The individual tree predictions are aggregated to give the final prediction, by arithmetic average in regression type of problems and a majority vote in classification type of problems. Tree complexity and size is controlled by adjusting two parameters namely, $max_{depth}(D_{max})$ and $min_{sampleleaf}$ [11].

2.2 Evaluation Metrics

Assessment of the models' performance was done using standard evaluation metrics namely, the mean square error (MSE), mean absolute error (MAE) and R-squared (R^2). The MSE is the mean of the square of the errors. The closer the MSE is to zero the more ideal. The larger the MSE , the larger the error. The MSE 's basic value is in selecting one prediction model over another. R^2 describes the proportion of variance of the dependent variable that is explained by the regression model. A low R^2 value shows a low level of correlation, meaning a regression model that is not valid, but not in all cases. MAE gives the mean error (positive) for all test data. Note one cannot look at these metrics in isolation in sizing up the model. These performance evaluation metrics are calculated using Equations (2 to 5) as follows:

$$R^2 = 1 - \frac{\sum(y - y')^2}{\sum(y - \bar{y}')^2}, \quad (1)$$

$$RMSE = \sqrt{\frac{\sum(y' - y)^2}{N}}, \quad (2)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y' - y| \quad (3)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y' - y)^2 \quad (4)$$

where y is the measured energy consumption value, y' is the energy consumption predicted value, \bar{y} represents the average values of the corresponding variables, N is the number of data points considered, CV is the coefficient of variation, σ is the standard deviation and μ is the population mean.

3 Development of Prediction Models

3.1 Building Description

The office building and two leisure centres namely, Waves and Don Tatnell leisure centres are in Melbourne and are both run and managed by Kingston municipality. Waves leisure centre generally comprises of a standalone Aquatic and Leisure

Centre situated along longitude and latitude 145.0577 °E and 37.9516 °S respectively. The centre comprises an aquatic area, the health and fitness areas and ancillary facilities comprise of male and female toilet and change rooms, staff rooms, school change, family change, creche, retail store, kiosk with associated food and storage areas, Mezzanine floor party hire area, general administration, reception area and foyer/entry area. The building area sits on approximately 5500m² of land on a concrete slab base, with rendered masonry walls and covered with a pitched tin roof and fitted with aluminium framed windows and doors.

Don Tatnell leisure centre also houses various indoor leisure activities under one roof including, a fitness centre, a spa, indoor swimming pool, a formal pool and an occasional day-care centre. Don Tatnell leisure centre location is to the northeast corner of the site, with the main entrance facing east. It is constructed from a concrete slab base, with rendered masonry walls and covered with a pitched tin roof and fitted with aluminium framed windows and doors. The longitude and latitude of the site are approximately 145.0924 °E and 37.9911 °S respectively. Both leisure centres are open every day of the week, from 6am-9pm weekdays and 7am-6pm weekends.

The longitude and latitude of the office building site are approximately 145.0 °E and -37.8 °S respectively. True north is about -57° from the front elevation (Nepean Highway or North Eastern side) of the building. The building is rectangular and comprises approximately 10,500m² floor area with seven stories of office space. The building facade primarily comprises painted precast concrete panels of 200mm thickness. Within the panels are vision glass window and spandrel combination sets. The glazing height varies from 2450mm for the ground level to 1950mm for levels 1 to 6. The width of glazing varies from 1980mm to 3280mm, while the front and rear have feature combinations on the centre of the facade.

3.2 Dataset Description

The electrical energy consumption datasets of the three buildings do not contain missing values. All data points correspond with correct timestamp values. However, the datasets did include some outliers. The system on few instances recorded a series of zero values then suddenly sums up the total power usage for that given period (with zero readings) with a very high value which would not typically be consumed in 15 minutes. A total of 20 potential input variables were investigated to test their impact on electrical power demand prediction at the two leisure centres and the office building. These inputs are: maximum and minimum temperature (T_{max}, T_{min}), dry bulb air temperature (T °C), mean temperature (T_{mean} °C), dew temperature (DewT °C), maximum and minimum wind (U_{max}, U_{min}), prevailing wind speed (U), gusty winds (U_g), relative humidity (RH%), wind direction (U_d), average wind speed (U_{mean}), average wind direction (U_{dmean}), year, week of the year, weekday, day of the year, month of the year, hour of the day, and 15 minute. The climatic variables of a nearby weather station were obtained from the Bureau of Meteorology with

a 15-minute resolution. The two leisure centres' training and validation datasets contain data from 01/05/2017 00:00 to 24/09/2018 09:30 while the office building dataset ranges from 01/06/2011-24/03/2018. Melbourne is classified according to the Köppen climate classification as a temperate oceanic climate. The city warms up in summer with mean temperatures between 14 - 25.3°C and winters averaging between 6.5 - 14.2°C.

3.3 Statistical Description of the Data

This section provides a brief insight into the energy consumption of the three datasets used for experimentation. Table 1 gives the statistical description of the energy consumption profiles at the sites. The office building has the highest number of historical energy consumption observations followed by Waves, and then Don Tatnell has the least observations recorded. On average, Waves leisure has the highest energy consumption rate followed by Don Tatnell and then the office building.

Parameter	Waves Centre	Don Tatnell	Office building
Count	48998	48050	236268
Mean	55.702233	23.84	19.79
Standard deviation	11.71	8.71	16.53
Min	0	6.50	0.00
25%	45	14.03	7.42
50%	60	24.84	9.79
75%	65	30.63	33.64
Max	94	53.97	176.54

Table 1: Statistical summary of energy consumption profiles at the three sites

The energy consumption distribution patterns for the Don Tatnell leisure centre, Waves leisure centre and the office building are shown in Fig 2.

The two leisure centres show a seemingly similar electrical consumption distribution shape relative to the office building. Don Tatnell leisure centre has a somewhat almost symmetric distribution while Waves leisure centre shows some skewness to the left and the office building being skewed to the right. The majority of the office building energy consumption readings fall within the 0-35 kWh range while Waves leisure centre has a range between 60-70kWh. It is worth mentioning that leisure centres tend to consume more energy than office buildings; however, little building research exists in building energy performance literature regarding these leisure centres.

3.4 Selection of candidate inputs

Among the 20 potential inputs, *Temperature*, *Tmean*, *Tmax* and *Tmin* showed high correlation and the same is true with *Umean*, *Umax*, *Ugust*. This means

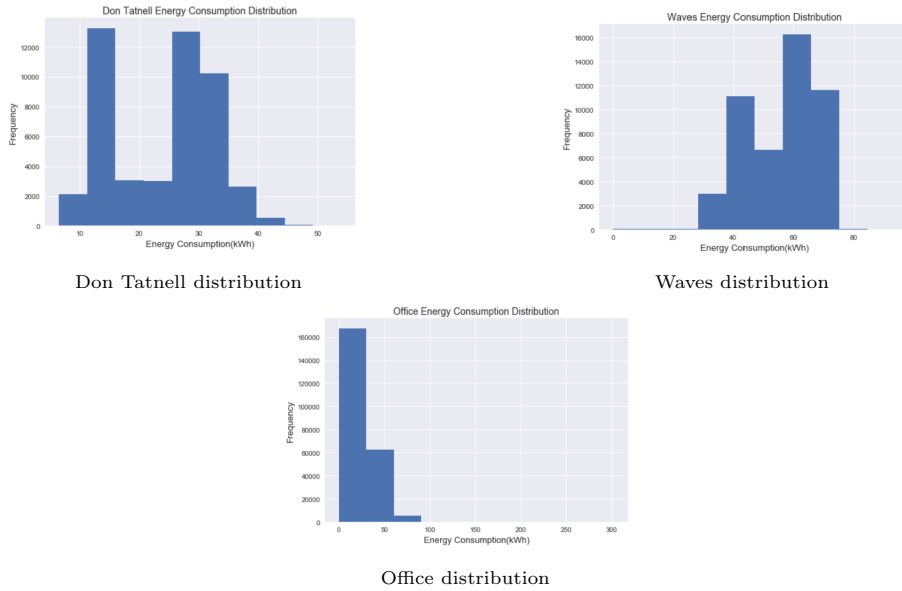


Fig. 2: Energy consumption histograms of the three buildings

that they have a similar effect on the dependent variables (energy consumption) as such, choosing only one as an input in the model is equally effective than using all. Following this explanation, T_{mean} and U_{mean} were adopted to represent temperature and wind-related inputs respectively, bringing the number of input variables down from 20 to 13 inputs.

3.5 Data Transformation

Due to the vast differences in numerical ranges between the input and output values standardisation of the inputs values was carried out. Standardisation scales each feature such that the distribution is centred around 0, with a standard deviation of 1. Standardisation allows comparability among inputs and it also enhances the training process since the numerical condition of optimisation is improved as opposed to without standardisation. As such, the mean and standard deviation for each feature is calculated, and then the feature is scaled using equation 5 :

$$z - score = (x_i - \mu) / \sigma. \quad (5)$$

where x_i is the observed value, μ is the population mean and σ is the population standard deviation.

3.6 Transfer Learning Experiment Set-up

The most skilful of the five machine learning algorithm described in section 2.1 are selected in the transfer learning investigation. Initially, the models are developed and fine-tuned from scratch on the office building (D_O), Don Tatnell (D_C) and Waves leisure centres (D_W). For comparison purposes models developed using Don Tatnell dataset will act as the baseline models to test transfer learning effect. During the development of all models, the training set sizes are varied gradually from 1% (to simulate the lack of data) to 80% (enough data). The earlier developed office building (D_O) and Waves leisure centre (D_W) models are then retrained using data from Don Tatnell for predicting energy consumption at Don Tatnell centre using a similarly sized training dataset of between 1% to 80%. The results are then compared using the evaluation metrics already described in section 2.2.

All learning algorithms were implemented using Python programming language. The development of machine learning models was done using the scikit-learn (Python programming language library) [17]. All model development and experimental tasks were conducted on a Windows machine (Intel Core i5 2.40GHz 8GB RAM).

4 Results and Discussion

4.1 Model Selection for Transfer Learning

Four evaluation metrics that are RMSE, MAE, MSE and R^2 are used for evaluation of the skill of prediction models. The performance of the five models in energy consumption prediction for Waves Leisure Centre is summarized in Table 2. All ensemble based tree models had equally good performance with slight variations amongst themselves. The decision tree algorithm has the least performance followed by the k-NN algorithm. The EET model with the least amount of error is adopted for the transfer learning task having MSE, MAE, RMSE and R^2 values of 12.89, 2.52, 3.59 and 0.913 respectively. Following this result, the EET model became the model of choice in the transfer learning experimentation.

Model	MSE	MAE	RMSE	R^2
Decision tree	28.83	3.72	5.37	0.805
K-Nearest Neighbour	22.87	3.52	4.78	0.85
Random forest	14.41	2.67	3.79	0.903
Ensemble Extra trees	12.89	2.52	3.59	0.913
LightGBM	13.58	2.64	3.68	0.908

Table 2: Waves leisure centre energy consumption prediction results

The optimum number of trees (M), maximum tree depth (D_{\max}), the minimum number of samples needed for splitting a node (n_{\min}) and the attribute

selection strength parameter (K) are 260, 8, 10 and 9 respectively for the best performing model. The (EET) algorithm was again selected and fine-tuned for office building energy consumption prediction. The EET model obtained MSE, MAE, RMSE and R^2 values of 0.48, 0.43, 0.69 and 0.97 respectively. Overall the EET model demonstrated better performance on office building relative to the leisure centre. This result is as expected and demonstrates the complexity of the leisure centre prediction exercise and also the larger dataset available at the office building.

4.2 Transfer learning results

Following a series of investigations, this section outlines the finding and gives discussion around the results. The performances of the developed EET models in the transfer learning exercise are summarised in Table 3 and 4. Initially, the EET models are developed using historical data and fine-tuned for the energy consumption prediction of the three buildings before selection for transfer learning. After that, the most skilful models according to the discussed evaluation metrics, are then set aside for the task of transfer learning.

Office building prediction has seven years worth of historical data, while both leisure centres have only sixteen months worth of available historical data. To test the effect of transfer learning on models the training data for all models is varied between 0.1% (simulating data shortage scenario) to 80% (simulating enough data situation) and the performance of the models is monitored consequently by tracking the evaluation metrics adopted. Particular importance is given to the instances where models are trained with few data as that represents the primary motivation for the investigation.

As seen in Table 3, it is evident that both instances of pre-trained models (with transfer learning) do have superior performances relative to the cases where the models are trained from scratch (no transfer learning). This phenomenon is observable at all training data sizes under consideration with the weakest performance being for training data sizes less than 10%. It is also noted that while pre-trained models do perform better than training from scratch, models pre-trained on Waves centre dataset have lower error metrics as opposed to those pre-trained on office buildings. This may be because of the similarities in building operations, location and building form. Overall, on average, models pre-trained on Waves centre relative to those that are trained from scratch, have lower MSE, MAE and RMSE values by 19%, 14% and 10% respectively. On the other hand, those models pre-trained on the office building have lower MSE, MAE and RMSE values by 7%, 5% and 4% respectively compared to Don Tatnell models. This shows that pre-trained models did have better performance than models developed from scratch with the models pre-trained on waves having the most superior performance.

In the last section of the investigation, differenced (lag=1) time-series data is considered under the same investigation conditions above, and the results do show a similar trend as observed earlier. As expected, pre-trained models show superior performance relative to the one trained from scratch. While in the earlier

Training %	With transfer				Without transfer (Direct don)				With transfer			
	Pre-trained with Office (D _O)				Direct Don (D _C)				Pre-trained with Waves (D _W)			
	MSE	MAE	RMSE	R ²	MSE	MAE	RMSE	R ²	MSE	MAE	RMSE	R ²
80	7.45	1.87	2.73	0.89	8.55	2.0	2.92	0.87	6.85	1.76	2.62	0.9
50	8.62	2.1	2.94	0.88	9.55	2.2	3.1	0.87	7.76	1.92	2.79	0.89
20	9.91	2.25	3.15	0.87	11.19	2.41	3.35	0.85	8.62	2.06	2.94	0.89
10	10.6	2.4	3.26	0.86	11.56	2.5	3.4	0.85	8.96	2.13	2.99	0.88
8	10.38	2.35	3.22	0.86	11.59	2.49	3.41	0.85	9.08	2.14	3.01	0.88
5	10.24	2.27	3.2	0.87	11.67	2.45	3.42	0.84	9.34	2.11	3.06	0.88
3	10.43	2.3	3.23	0.86	11.49	2.45	3.39	0.85	9.68	2.16	3.11	0.87
1	13.13	2.66	3.62	0.83	13.92	2.78	3.73	0.82	11.4	2.42	3.38	0.85
0.8	11.78	2.53	3.43	0.84	12.77	2.6	3.57	0.83	10.65	2.34	3.26	0.86
0.5	12.4	2.59	3.52	0.84	14.01	2.81	3.74	0.82	11.02	2.35	3.32	0.85
0.3	12.28	2.57	3.5	0.84	14.56	2.86	3.82	0.81	11.49	2.38	3.39	0.85
0.1	33.97	4.57	5.82	0.55	31.03	4.44	5.57	0.59	25.6	3.9	5.06	0.66

Table 3: Transfer learning results using an undifferenced time-series

example models pre-trained on waves leisure centre had superior performance to that training on the office building, results show almost similar performance on training set sizes greater and 10%. On average, both pre-trained models(D_O and D_W) show lower MSE, MAE and RMSE values of 19%, 16% and 10% as compared to those trained from scratch (D_C) models. However, of particular note is the rather seemingly equal performance by pre-trained models from Waves centre and the office building.

Training %	With transfer			Without transfer (Direct don)			With transfer		
	Pre-trained with Office (D _O)			Direct Don (D _C)			Pre-trained with Waves (D _W)		
	MSE	MAE	RMSE	MSE	MAE	RMSE	MSE	MAE	RMSE
80	8.75	1.8	2.96	11.03	2.18	3.32	8.73	1.8	2.95
50	8.83	1.81	2.97	11.14	2.19	3.38	8.82	1.81	2.97
20	9.5	1.88	3.08	12.14	2.32	3.48	9.54	1.88	3.09
10	9.56	1.91	3.09	12.25	2.33	3.5	9.6	1.92	3.1
8	9.61	1.92	3.1	12.12	2.32	3.48	9.66	1.92	3.11
5	9.72	1.93	3.12	12.08	2.32	3.48	9.78	1.94	3.13
3	9.75	1.94	3.12	11.8	2.31	3.44	9.81	1.95	3.13
1	10.19	2.01	3.19	12.5	2.39	3.54	10.33	2.04	3.21
0.8	10.3	2.03	3.21	12.95	2.44	3.6	10.38	2.06	3.22
0.5	10.22	2.03	3.2	12.32	2.38	3.51	10.34	2.05	3.22
0.3	10.42	2.05	3.23	12.77	2.41	3.57	10.51	2.08	3.42
0.1	11.15	2.25	3.34	13.37	2.55	3.66	11.15	2.25	3.34

Table 4: Transfer learning results using a differenced time-series

Nonetheless models trained on differenced time-series show lower error values in general across all training data sizes and models (pre-trained and direct). Thus, transfer learning with differenced time series recorded lower error values (Table 4) as compared to instances where the time-series is undifferenced (Table 3).

5 CONCLUSION

The need for fast and accurate models for building energy consumption prediction continues to increase particularly with the rise in the need for renewable energy sources and ever-changing smart grid networks. Data acquisition is a costly exercise concerning both time and financial resources. Due to the general problem of inadequate training data, the authors try to show the benefits of the well-known transfer learning paradigm. The knowledge learnt from buildings with vast amount of data was leveraged for use in the building with little data. This study investigated the applicability of transfer learning and ways of maximising its benefits in the task of building energy consumption prediction. To exemplify the transfer learning problem case three buildings comprising an office building and two leisure centres, are analysed and machine learning models, based on mainly ensemble decision trees, are developed and tested. The benefits of this approach are evident in our experimental results. Transfer learning demonstrated advantage, in terms of prediction accuracy, even comparing with models with adequate training data sets. The study also concluded that differencing a time series improves transfer learning. This advantage is independent of underlying learning methods. Hence we conclude that transfer learning is valid and a useful method for building energy consumption prediction in complex facilities like leisure centres. With this approach, extensive data collection prior to the learning becomes less essential.

The authors are currently investigating ways of improving transfer learning on these time-series related problems using deep learning models as an extension of this current research work. Unlike classical machine learning models, deep learning techniques have abilities to learn the temporal dependence automatically and naturally handle temporal structures found within time series data.

References

1. Ahmad, M.W., Mourshed, M., Yuce, B., Rezgui, Y.: Computational intelligence techniques for hvac systems: A review. In: *Building Simulation*. vol. 9, pp. 359–398. Springer (2016)
2. Altman, N.S.: An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician* **46**(3), 175–185 (1992)
3. Birdsall, B., Buhl, W.F., Ellington, K.L., Erdem, A.E., Winkelmann, F.C.: Overview of the doe-2 building energy analysis program, version 2.1 d (1990)
4. Breiman, L.: Random forests. *Machine learning* **45**(1), 5–32 (2001)
5. Chen, H., Lee, W., Wang, X.: Energy assessment of office buildings in china using china building energy codes and leed 2.2. *Energy and buildings* **86**, 514–524 (2015)
6. Costa, A., Keane, M.M., Torrens, J.I., Corry, E.: Building operation and energy performance: Monitoring, analysis and optimisation toolkit. *Applied Energy* **101**, 310–316 (2013)
7. Crawley, D.B., Lawrie, L.K., Pedersen, C.O., Winkelmann, F.C.: Energy plus: energy simulation program. *ASHRAE journal* **42**(4), 49–56 (2000)

8. Eurostat, L., et al.: Energy indicators for sustainable development: Guidelines and methodologies (2008)
9. Foucquier, A., Robert, S., Suard, F., Stéphan, L., Jay, A.: State of the art in building modelling and energy performances prediction: A review. *Renewable and Sustainable Energy Reviews* **23**, 272–288 (2013)
10. Fumo, N.: A review on the basics of building energy estimation. *Renewable and Sustainable Energy Reviews* **31**, 53–60 (2014)
11. Geurts, P., Ernst, D., Wehenkel, L.: Extremely randomized trees. *Machine learning* **63**(1), 3–42 (2006)
12. Grubinger, T., Chasparis, G.C., Natschläger, T.: Generalized online transfer learning for climate control in residential buildings. *Energy and Buildings* **139**, 63–71 (2017)
13. Hastie, T., Tibshirani, R., Friedman, J.: Unsupervised learning. In: *The elements of statistical learning*, pp. 485–585. Springer (2009)
14. Huang, J.T., Li, J., Yu, D., Deng, L., Gong, Y.: Cross-language knowledge transfer using multilingual deep neural network with shared hidden layers. In: *Acoustics, Speech and Signal Processing (ICASSP)*, 2013 IEEE International Conference on. pp. 7304–7308. IEEE (2013)
15. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.Y.: Lightgbm: A highly efficient gradient boosting decision tree. In: *Advances in Neural Information Processing Systems*. pp. 3146–3154 (2017)
16. Mocanu, E., Nguyen, P.H., Kling, W.L., Gibescu, M.: Unsupervised energy prediction in a smart grid context using reinforcement cross-building transfer learning. *Energy and Buildings* **116**, 646–655 (2016)
17. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al.: Scikit-learn: Machine learning in python. *Journal of machine learning research* **12**(Oct), 2825–2830 (2011)
18. Recast, E.: Directive 2010/31/eu of the european parliament and of the council of 19 may 2010 on the energy performance of buildings (recast). *Official Journal of the European Union* **18**(06), 2010 (2010)
19. Ribeiro, M., Grolinger, K., Elyamany, H.F., Higashino, W.A., Capretz, M.A.: Transfer learning with seasonal and trend adjustment for cross-building energy forecasting. *Energy and Buildings* **165**, 352–363 (2018)
20. Thomas, S., Ganapathy, S., Hermansky, H.: Multilingual mlp features for low-resource lvcsr systems. In: *Acoustics, Speech and Signal Processing (ICASSP)*, 2012 IEEE International Conference on. pp. 4269–4272. IEEE (2012)
21. Vu, N.T., Imseng, D., Povey, D., Motlicek, P., Schultz, T., Bourlard, H.: Multilingual deep neural network based acoustic modeling for rapid language adaptation. In: *Acoustics, Speech and Signal Processing (ICASSP)*, 2014 IEEE International Conference on. pp. 7639–7643. IEEE (2014)
22. Wei, Y., Zhang, X., Shi, Y., Xia, L., Pan, S., Wu, J., Han, M., Zhao, X.: A review of data-driven approaches for prediction and classification of building energy consumption. *Renewable and Sustainable Energy Reviews* **82**, 1027–1047 (2018)
23. Zuo, J., Zhao, Z.Y.: Green building research—current status and future agenda: A review. *Renewable and sustainable energy reviews* **30**, 271–281 (2014)