# A Transfer Learning Approach to Create Energy Forecasting Models for Building Fleets

Monica Vasquez Torres, Zahraa Shahid, Karan Mitra, Saguna Saguna, Christer Åhlund
Department of Computer Science, Electrical and Space Engineering
Luleå University of Technology, Skellefteå, Sweden
{firstname.lastname}@ltu.se

***Abstract*—The development of accurate energy prediction models plays a significant role in achieving sustainability in smart cities. However, stakeholders such as municipalities face the problem of creating individual energy forecasting models for multiple building fleets which leads to an increased amount of computational resources and time spent to prepare each model. This research proposes a method using Hierarchical clustering with dynamic time warping to group similar buildings according to their consumption values and the integration of transfer Learning (TL) to share the model weights from a source building to other target buildings. Different TL models using only 20%, 40%, and 60% of the target data were tested against a standard workflow without TL for predicting electricity and district heating for several school buildings using a Multivariate LSTM model. The results show a small variation between the TL and the standard models; when trained on only 40% of the data, the models achieved an average of 0.24% RMSE improvement for district heating and a 1.23% for electricity, indicating a potential for reduced data requirements without sacrificing predictive accuracy and demonstrating TL's efficiency to streamline the energy forecasting process for building fleets.**

***Keywords*—** Building fleet, Energy consumption, Transfer learning, LSTM, DTW, Hierarchical clustering, Time series forecasting

## I. Introduction

Municipalities own, maintain, and manage extensive building fleets and given the current energy crisis, there is a demand for environmentally friendly ICT solutions that contribute to the enhancement and reduction of energy consumption in buildings. Accurate predictions of energy consumption for future periods aid significantly in cost-cutting and energy savings [1]. However, this involves the creation of individual models for each building resource, and, in the long term, this represents a significant challenge to scale to a larger set of buildings as each model is trained independently without leveraging any past knowledge. It is also computationally expensive and requires a large amount of data to achieve high performance [2]. In the context of building fleets, optimizing energy usage and refining operations in existing and future buildings is necessary. A solution to this problem involves using Transfer Learning (TL) by leveraging knowledge from one domain to improve performance in another similar domain. The integration of TL is a promising technique that has already demonstrated its potential to enhance data efficiency, accelerate training speed, and increase model accuracy [3].

The applications of TL have been widely studied in the fields of computer vision and natural language processing [4]. Nevertheless, applications within the research area of time series analysis are still rare; for instance, one common problem is finding a domain with a high degree of similarity with the target domain, as the similarity between domains is a crucial factor influencing the effectiveness of TL in time series analysis. Different approaches to selecting the right source domain have been studied. Among them, source selection is an interesting procedure in reusing knowledge for energy consumption models in buildings, as the main idea is to only reuse knowledge from domains with reasonable similarity to the target. This degree of similarity between the time series has been studied using different techniques, such as selecting source domains based on performance testing in the target domain using labeled data and combining top-performing models into an ensemble. Common similarity measures include Dynamic Time Warping (DTW), Euclidean distance, Jensen–Shannon (JS) divergence, Pearson correlation, and Maximum Mean Discrepancy (MMD). Additionally, selecting pre-trained source models based on the similarity of target encodings using cosine distance can assist in domain selection [5]. Although the applications of TL in the realm of time series are not as prevalent as those in computer vision, it has the potential to speed up the creation and development of new forecasting models. The integration of TL is advantageous for time series forecasting in buildings to save time and computational resources by leveraging pre-trained models like RNN-LSTM (Recurrent Neural Network - Long Short-Term Memory). This approach enhances prediction accuracy efficiently [6].

School buildings constitute Sweden's largest segment of public properties, presenting a significant opportunity for sustainable development. Among these, approximately 11,567 school buildings covering 30 million square meters are heated through district heating systems, accounting for 28% of all public properties. Schools surpass all other non-residential buildings in energy consumption. In 2020, heating demands reached 4,222 GWh through district heating systems and 573 GWh via electricity, with an additional 2,534 GWh used for supplementary purposes such as lighting and ventilation. Within the service sector, school buildings stand out with energy usage occupying 27% of the total heated area in 2020, averaging 113 kWh/m$^2$ and 125 kWh/m$^2$ for district heating and electricity respectively [7]. To address these concerns, the European Union aims for a 32.5% increase in energy efficiency by 2030 compared to projected consumption, while Sweden targets a 50% improvement from 2005 levels. One of the main current challenges is that school buildings consume more energy than any other non-residential buildings. In this context, energy forecasting systems and TL are explored to advance energy efficiency in school buildings' infrastructure. Leveraging accurate forecasting of energy usage and integrating the TL approach for the existing buildings and the buildings at the initial phase of the planning could allow all stakeholders to conduct their economic analysis and optimize decision-making.

This work aims to investigate the integration of TL in the creation of energy forecasting models for building fleets located in Skellefteå municipality, in the North of Sweden. By focusing on this region, the study aims to provide insights into the applicability and effectiveness of TL within a distinct geographical and environmental setting. Based on the information provided, the contributions of this work are the following:

- A proposed method for grouping similar buildings and selecting the source building by employing a similarity measurement for time series data and a clustering technique.
- An experiment that integrates TL with a subset of target building data and a standard model with no TL to analyze and compare the performance results. This analysis evaluates and contrasts the performance outcomes between the two methods.

## II. Related Work

Several authors have stated the use of TL for energy prediction of buildings; Kim et al. [8] implemented a Long Short-Term Memory TL (LSTM-TL) model for building energy demand forecasting. TL is used to enhance the accuracy of the LSTM prediction model under different weather conditions. Moreover, Gonzalez-Vidal et al. [9] proposed a TL framework for smart buildings to address energy-related problems. It utilizes a clustering algorithm for mixed data and clustering of the image-based representation of time series to create a network that groups buildings sharing characteristics. The framework was tested with several rooms/buildings and two energy efficiency domains. It reduces the coefficient of variation of the root mean squared error (CVRMSE) in energy consumption prediction by 21.6% and in air conditioning usage prediction by a significant margin. Furthermore, Genkin et al. [10] proposed a TL approach to reduce the impact of the reinforcement learning agent's warm-up period on the building's energy efficiency by transferring knowledge from an existing, optimized smart building to a newly commissioned building, the warm-up period efficiency is enhanced by up to 6.2 times, and the prediction variance is reduced by 132 times.

On the other hand, Zhou et al. [11] presented a use case for TL to predict the electric power of a primary school building in China and leveraged the use of data from different similar buildings located in the same area to improve the model due to the school's insufficient historical data to create a robust prediction model, the results are considered satisfactory as the error rate is relatively low by using this new approach. In addition, different authors have developed similar approaches when dealing with data scarcity [12] [13] [14] [15] where a TL model is used to approach different problems such as load forecasting, occupancy prediction, energy consumption and predictive control of HVAC.

The different works show that TL has been widely implemented to approach different problems related to buildings, mainly focused on developing models with limited data to leverage the knowledge from rich pre-trained source models. However, existing literature has not thoroughly addressed the application of TL models across multiple building fleets. Therefore, this work seeks to introduce a methodology tailored to bridge this gap.

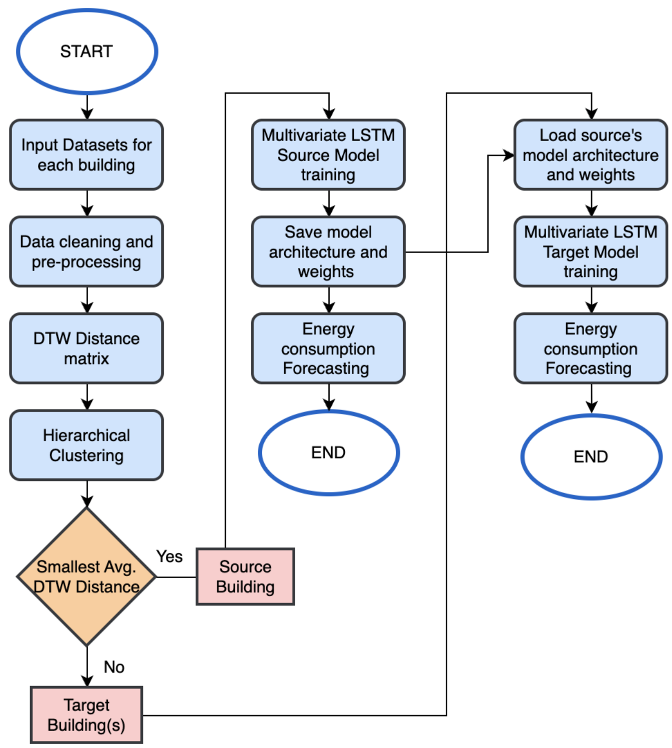



Fig. 1. Proposed Transfer Learning method Flow chart

## III. Methodology

Pan et al. [16] defined TL as the process of enhancing the predictive model of a target domain $D_T$ leveraging the patterns and knowledge from a source domain $D_S$, even when the source and target domains $D_S \neq D_T$ or tasks $T_S \neq T_T$ are not the same. For this scenario, TL allows a predictive energy consumption model for a target building to be improved by using data from a source building with similar energy consumption patterns. This similarity is calculated using the DTW distance between their energy usage time series. Essentially, a model developed for one source building is adapted by a target building, accelerating model training and potentially reducing the need for extensive data from target buildings.

A high-level overview of the proposed method is illustrated in Figure 1 with a flowchart. A more detailed explanation of the different processes of this method is presented below:

### A. Data Pre-processing:

This step involves the pre-processing and the cleaning of the data with different techniques for the datasets used in this work (electricity and district heating data). The data cleaning process included removing outliers using a Z-Score threshold of 2.0 to 2.5 and eliminating zero values. Missing values were addressed through linear interpolation, and data normalization was achieved using the Z-score method. Additionally, data exploration techniques such as stationarity analysis, descriptive statistics, and visualization were implemented to examine the data. The different methods for data cleaning used for this part were obtained from the previous work which focused on developing forecasting models for the same datasets [17].

### B. DTW Distance Matrix:

Since we are dealing with time series datasets that don't share the same length, DTW is an ideal distance metric to calculate the similarity between time series, unlike Euclidean distance that forces both time series to be the same length. The basic idea is that given two sequences, the objective of DTW is to temporally align these sequences in some optimal sense under certain constraints [18]. For this step, DTW between multiple time series is calculated using the distance matrix method from the DTAI Distance library [19]. To compute the DTW distance matrix for multiple time series (on this case, the different datasets for each building) $\{T_1, T_2, \ldots, T_k\}$, all pairs are iterated $(i, j)$, and then the DTW distance is calculated $DTW(T_i, T_j)$. The distance matrix $D$ can be represented as $D_{ij} = DTW(T_i, T_j)$ for all $i, j$, where $D$ is a symmetric matrix with zeros on the diagonal.

### C. Hierarchical clustering:

Agglomerative Hierarchical Clustering (AHC) is a method for grouping similar objects based on a measure of distance or similarity and the clusters can be visualized in a hierarchical tree called a dendrogram which also offers the advantage of not having to specify a pre-defined number of clusters, instead, the number of clusters can be decided by analyzing the dendrogram [20]. Using DTW as a similarity distance with AHC has already improved clustering results compared to other distance measures such as Euclidean, Manhattan, and Cosine [21]. Based on the aforementioned, this step groups similar buildings in terms of energy consumption values. This is achieved by using AHC, and the linkage criterion uses the DTW distance matrix generated in the previous step with a ward method to produce clusters with small variances and similar sizes.

### D. Source Building selection:

For each cluster, the building with the smallest average DTW distance to all other buildings in the cluster is selected as the source building where the knowledge sharing will come from. Furthermore, the length of the time series is also an important point for selecting the right source dataset since an abundance of data will have better and more generalizable feature representation that will contribute to better knowledge sharing [22]. Thus, the average DTW distances between buildings within each cluster are calculated and normalized by the length of each series. Afterward, the building with the smallest value is considered the source building, and the other buildings in the cluster will be considered targets where the knowledge will be shared.

### E. Source Building model training:

The source building model is trained with the preferred time series model architecture (Multivariate LSTM for this case), and the entire model is saved (containing weights and training configuration) in h5 format. After this, the forecasting is performed.

### F. Transfer Learning

*1) Feature extraction and Fine tuning:* The TL approach usually involves using 2 common techniques; Feature extraction and Fine-tuning. Feature extraction basically comprises the removal of the layers responsible for classification in the source model and the incorporation of new layers specific to the target task. The weights of the source model are frozen, meaning they are not updated during training, and only the weights of the newly added layers are trained using the target dataset. On the other hand, Fine-tuning involves unfreezing some or all of the layers from the source model with a low learning rate. This allows these layers to be updated with the new dataset, enabling the model to adapt and learn more specific features relevant to the new task or domain. Fine-tuning helps refine the performance of the target model by adjusting its parameters to better suit the specific tasks [23].

*2) TL workflow:* The standard TL workflow described on the Keras TL & fine-tuning documentation [24] is utilized for the experiments on this work. A detailed explanation of this workflow is given below:

1) The pre-trained weights of the source model are loaded into the new target model, and the original output layer of the pre-trained model is replaced by adding a new output layer specific to the target on top of the pre-trained model [25].
2) The other layers of the source model are frozen by deactivating the trainable setting of this model.
3) A small portion of the target dataset is taken for the TL models (for this case, 20%, 40%, and 60% of the data). This data portion is then divided into training and testing sets (last 16 months and last 6 months for large and small datasets, respectively).

4) The target model is trained on the data. The new output layer is initialized with random weights and optimized with the selected hyperparameters (after a random search algorithm). This is the feature extraction process described before.
5) After feature extraction where the training starts with a high learning rate (0.01) and 20 epochs, and once the model has converged on the new data, a fine-tuning process is then performed where the layers of the base model can be unfrozen and retrained with a lower learning rate (0.0001) and a higher number of epochs. This step can lead to overfitting, thus a small value for the patience parameter is assigned to mitigate this potential issue.

### G. Target Building model training:

The target buildings from the same cluster get initialized with the source building weights and training configuration, i.e., the h5 file is loaded into the target models. Subsequently, a specified segment of the target data is utilized and the model is trained using a TL workflow (described above), leading to the execution of the forecasting process.

## IV. Experiments

To evaluate the functionality of the proposed method, we utilized several datasets for electricity and district heating, covering 6 and 9 school buildings, respectively. Additional details on the experiments are provided in the following subsections.

### A. Datasets and selected model architecture

The school buildings are located in Skellefteå municipality, in Sweden. The electricity and district heating datasets are daily consumption data, and both types of energy are represented by kWh/m2. The dataset information is described in Table I. Furthermore, Shahid et al. [17] developed several forecasting models for these two types of energy using the same datasets provided for this work. The study involved the development of short-term forecasting of electricity and district heating using Multivariate RNNs, LSTM, CNNs, and autoencoders. The hybrid CNN-LSTM and Multivariate LSTM achieved the best accuracy compared to other models. For simplicity, the selected model architecture for this work is the Multivariate LSTM model with 7-time steps and 3-day forecasting. In addition, the study identified features that improve predictions for both types of energy using Pearson correlation and these are also used for this work. For district heating; the cyclic features for month day, and day of the year, as well as the actual degree day. For electricity; cyclic features for weekday.

### B. Clustering

The datasets for each type of energy are pre-processed and cleaned using the methods mentioned in the previous section. Subsequently, the DTW distance matrix is calculated for electricity and district heating using all datasets available.

TABLE I
Datasets information

| Energy type | Name | Building Type | Years |
|---|---|---|---|
| **District Heating** | Bureskolan | Primary school | 2012-2022 |
| | Björnåkersskolan | Primary school | 2012-2022 |
| | Byskeskolan | Primary school | 2012-2022 |
| | Bureskolan bath | Bathhouse | 2012-2022 |
| | Byskeskolan bath | Bathhouse | 2012-2022 |
| | Byskeskolan Förskola | Pre-school | 2014-2022 |
| | Tallbacka | Pre-school | 2017-2022 |
| | Morohojdens | Pre-school | 2017-2022 |
| | Norrbacka | Pre-school | 2015-2022 |
| **Electricity** | Bureskolan | Primary school | 2011-2023 |
| | Björnåkersskolan | PrImary school | 2011-2023 |
| | Byskeskolan | PrImary school | 2011-2023 |
| | Tallbacka | Pre-school | 2017-2023 |
| | Morohojdens | Pre-school | 2017-2023 |
| | Norrbacka | Pre-school | 2015-2023 |

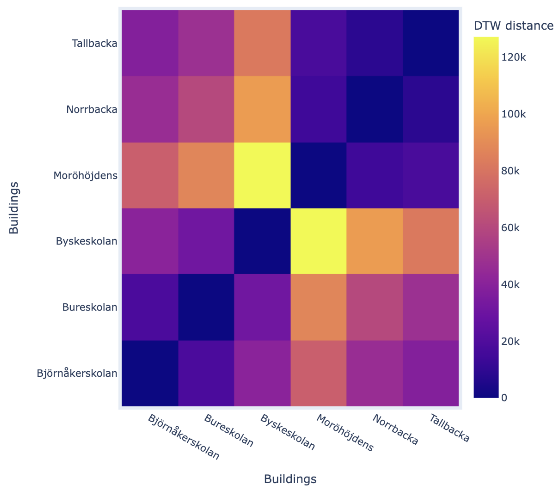
Figure 2 represents a heatmap of the raw DTW values for each type of energy. The lower the values, the higher the similarity between time series. As observed, the DTW distances are highly influenced by the type of building in both cases.

The dendrogram generated with AHC and the DTW distance matrix created in the previous step are used to group similar buildings into clusters for each type of energy. This clustering is then validated using a Silhouette Score, which is a value between -1 and 1, where higher values indicate a better clustering [26]. Figure 3 depicts the dendrogram generated for electricity and district heating. In the case of electricity, there is a clear grouping with a total of 2 clusters, and the Silhouette coefficient shows the highest value for this number, whereas, for district heating, the optimal value is around 3 clusters with a Silhouette coefficient higher than 0.6.

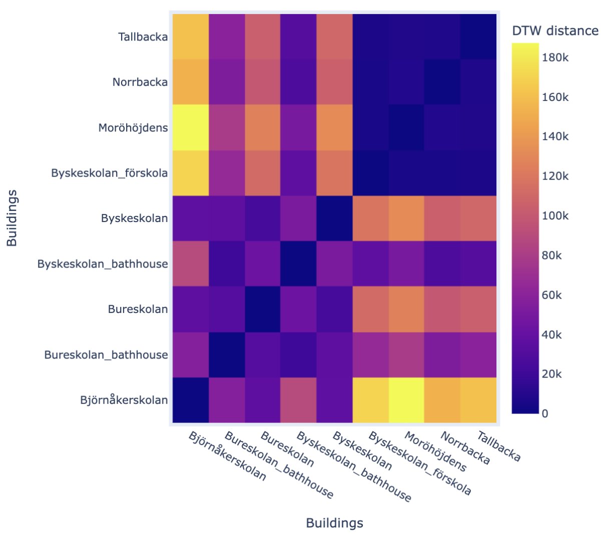
After clustering, the normalized average distances for each building to all other buildings in the cluster are calculated to select the source building. The buildings with the smallest values are categorized as the source buildings, and the others are labeled as target buildings. Table II presents the generated clusters and the selected source building.

### C. Source Models

Once the source buildings are located on each cluster, the Multivariate Multistep LSTM model [17] is used to train and forecast electricity and district heating consumption values using the data from these source buildings. To find the right hyperparameters (such as learning rate, batch size, and the number of layers and epochs), a random search algorithm is implemented with a total of 10 combinations where the combination with the highest accuracy is selected, and those hyperparameters are then used to train the final model. For this source model, the data split is 80/20 for training and testing. Finally, the save model option supported by Keras is utilized to store the weights, biases, model architecture, and training configuration in an h5 file.




(a) Electricity

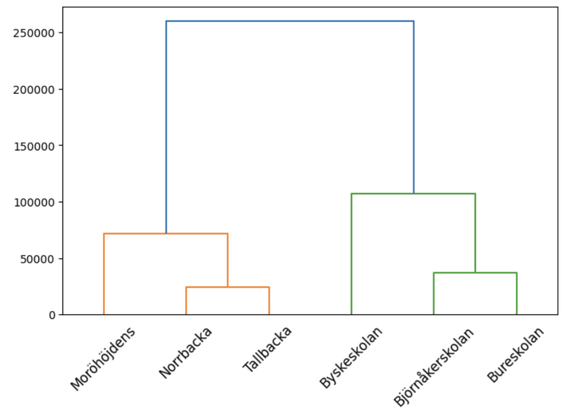



(b) District Heating

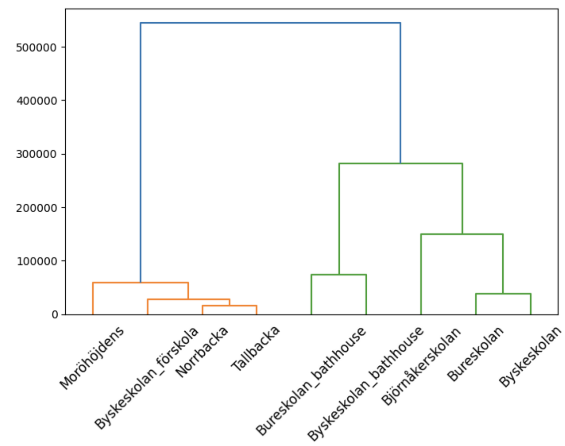
Fig. 2. DTW Distance Matrix values for each building pairwise (kWh)

TABLE II
CLUSTERING FOR ELECTRICITY AND DISTRICT HEATING WITH SOURCE BUILDINGS

| Energy Type | Cluster | Buildings | Scaled Avg. Distance | Source |
|---|---|---|---|---|
| District Heating | Cluster 1 | Bureskolan | 18.887 | Bureskolan |
| | | Björnåkersskolan | 21.987 | |
| | | Byskeskolan | 20.023 | |
| | Cluster 2 | Bureskolan bath | 6.106 | Bureskolan bath |
| | | Byskeskolan bath | 6.240 | |
| | Cluster 3 | Byskeskolan Förskola | 7.530 | Byskeskolan Förskola |
| | | Tallbacka | 13.474 | |
| | | Morohojdens | 14.218 | |
| | | Norrbacka | 10.281 | |
| Electricity | Cluster 1 | Bureskolan | 11.060 | Bureskolan |
| | | Björnåkersskolan | 13.069 | |
| | | Byskeskolan | 15.936 | |
| | Cluster 2 | Tallbacka | 11.046 | Norrbacka |
| | | Morohojdens | 12.910 | |
| | | Norrbacka | 7.020 | |




(a) Electricity




(b) District Heating

Fig. 3. Agglomerative Hierarchical Clustering dendrograms

### *D. Target Models*

To test the efficiency of the proposed TL method, the target models are created as follows:

- A standard model without TL (for comparison purposes) and using 100% of the data.
- A model using a small portion of the data (the last 20%, 40%, and 60%) and the TL workflow described in the previous section involving feature extraction and fine-tuning.

To ensure a fair comparison, all models (TL and Non-TL) use the same testing set consisting of the last 16 months for larger datasets (Primary schools and Bathhouses) and the last 6 months for smaller datasets (kindergarten). Afterwards, forecasting is performed, and common metrics such as RMSE, MAE, and R-squared are calculated for each model.

## V. RESULTS AND DISCUSSION

The performance metrics for the target models for both district heating and electricity are presented in Table III. The last column of the tables also presents the percentage of RMSE improvement between the standard Non-TL model (NO_TL) and the models using TL (TL_20, TL_40, TL_60). In general, the results for all models using TL generate similar results to those when creating individual models without TL. Figure 4 shows the electricity consumption real values and the Non-TL and TL predictions for Byskeskolan in the month of September 2023. The prediction results exhibit high similarity across both the standard model and the TL model.

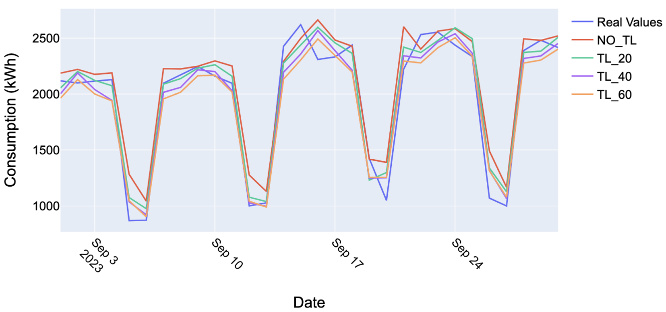



Fig. 4. Electricity consumption for Byskeskolan Actual values vs. Predicted values with the standard model and TL models for September 2023.

For district heating, the TL models do not consistently match the performance of the NO_TL models across all clusters. Nevertheless, the models using 40% and 60% of the data show an average RMSE improvement of 1.50% and 0.23%, respectively; the models using 20% of the data generally show negative improvement, indicating they do not achieve similar accuracy with less data, with the exception of Björnåkersskolan, where a lower amount of the data generated better results. For electricity, the results show a better RMSE improvement when using 40% of the data. On average, there is a 1.23% improvement, whereas there is a 0.68% improvement for 60% of the data. For this type of energy, the performance results are significantly affected when using a small portion of the data. On average, for both types of energy, the TL models using 40% of the data tend to come closer to the Non-TL models using the whole dataset, and it also comes with a small improvement on the RMSE. This suggests that, by using TL and only using 40% of the target data, we can reach similar results compared to utilizing the whole dataset and not leveraging learning from a source model. These results emphasize how TL can improve the resilience and predictive precision of models, even when faced with limited data in the target building.

## VI. Conclusion and Future work

In conclusion, this research proposes a method for creating multiple energy forecasting models for building fleets using common clustering methods and TL. Different TL models were tested against standard models without TL, and the results indicate similar accuracy. This suggests that equivalent performance can be achieved by reusing the knowledge from a pre-trained model that has been trained on a similar dataset. The integration of TL techniques in the development of forecasting models can significantly save time and computational resources. Municipalities can benefit from this approach by training a few models for representative buildings and then using the trained weights for other target buildings. This avoids the need to develop individual models for each building, which involves several tasks such as finding the right model architecture, training all layers, and fine-tuning. Furthermore, TL can also be beneficial in cases where data scarcity is present. Pre-trained models can be used to capture general patterns and representations from a rich source domain and then share this knowledge with the target with limited data, leading to a more robust and accurate model. For future work, exploring this method with other types of public buildings could be valuable as the experiments presented are only focused on different types of school buildings. Additionally, this method can be tested on other model architectures, such as Multilayer Perceptron (MLP) and CNN-LSTM.

## VII. Acknowledgment

This research was supported and funded by the European Commission grant number 610619-EPP-1-2019-1-FR-EPPKA1-JMD-MOB (EMJMD GENIAL Project).

## References

[1] "Time-series based household electricity consumption forecasting," in *2023 11th International Conference on Smart Grid (icSmartGrid)*, 2023, pp. 1–15.

[2] P. Baheti, "What is transfer learning? [examples newbie-friendly guide]," Oct 2021. [Online]. Available: https://www.v7labs.com/blog/transfer-learning-guide

[3] G. Pinto, Z. Wang, A. Roy, T. Hong, and A. Capozzoli, "Transfer learning for smart buildings: A critical review of algorithms, applications, and future perspectives," *Advances in Applied Energy*, vol. 5, p. 100084, Feb. 2022. [Online]. Available: http://dx.doi.org/10.1016/j.adapen.2022.100084

[4] S. Ruder, "Transfer learning - machine learning's next frontier," Sep 2022. [Online]. Available: https://www.ruder.io/transfer-learning/

[5] M. Weber, M. Auch, C. Doblander, P. Mandl, and H.-A. Jacobsen, "Transfer learning with time series data: A systematic mapping study," *IEEE Access*, vol. 9, p. 165409–165432, 2021. [Online]. Available: http://dx.doi.org/10.1109/ACCESS.2021.3134628

[6] S. H. Faru, A. Waititu, and L. Nderu, "A hybrid neural network model based on transfer learning for forecasting forex market," *Journal of Data Analysis and Information Processing*, vol. 11, no. 02, p. 103–120, 2023. [Online]. Available: http://dx.doi.org/10.4236/jdaip.2023.112007

[7] Energimyndigheten, "Energistatistik för lokaler, 2020," 2020, https://www.energimyndigheten.se/statistik/den-officiella-statistiken/statistikprodukter/energistatistik-for-lokaler/, Last accessed on 2022-12-01.

[8] D. Kim, Y. Lee, K. Chin, P. J. Mago, H. Cho, and J. Zhang, "Implementation of a long short-term memory transfer learning (lstm-tl)-based data-driven model for building energy demand forecasting," *Sustainability*, vol. 15, no. 3, p. 2340, Jan. 2023. [Online]. Available: http://dx.doi.org/10.3390/su15032340

[9] A. Gonzalez-Vidal, J. Mendoza-Bernal, S. Niu, A. F. Skarmeta, and H. Song, "A transfer learning framework for predictive energy-related scenarios in smart buildings," *IEEE Transactions on Industry Applications*, vol. 59, no. 1, p. 26–37, Jan. 2023. [Online]. Available: http://dx.doi.org/10.1109/TIA.2022.3179222

[10] M. Genkin and J. J. McArthur, "A transfer learning approach to minimize reinforcement learning risks in energy optimization for smart buildings," 2023. [Online]. Available: https://arxiv.org/abs/2305.00365

[11] M. Zhou, J. Yu, F. Sun, and M. Wang, "Forecasting of short term electric power consumption for different types buildings using improved transfer learning: A case study of primary school in china," *Journal of Building Engineering*, vol. 78, p. 107618, Nov. 2023. [Online]. Available: http://dx.doi.org/10.1016/j.jobe.2023.107618

TABLE III
DISTRICT HEATING AND ELECTRICITY TARGET MODELS RESULTS FOR NO_TL AND TL WITH 20%, 40%, AND 60% OF THE DATA

| Energy Type | Cluster | Building | Strategy | Metric | | | RMSE % improvement |
|---|---|---|---|---|---|---|---|
| | | | | RMSE | MAE | R2 | |
| District Heating | Cluster 1 | Björnåkersskolan | NO_TL | 581.362 | 408.661 | 0.908 | - |
| | | | TL_20 | 582.034 | 420.878 | 0.908 | **-0.116** |
| | | | TL_40 | 587.992 | 417.435 | 0.906 | -1.140 |
| | | | TL_60 | 594.302 | 422.126 | 0.904 | -2.226 |
| | | Byskeskolan | NO_TL | 565.505 | 315.616 | 0.842 | - |
| | | | TL_20 | 529.045 | 270.005 | 0.859 | 6.447 |
| | | | TL_40 | 518.956 | 257.228 | 0.866 | 8.231 |
| | | | TL_60 | 516.064 | 250.960 | 0.868 | **8.743** |
| | Cluster 2 | Byskeskolan Bath | NO_TL | 287.572 | 216.165 | 0.750 | - |
| | | | TL_20 | 312.993 | 236.553 | 0.695 | -8.840 |
| | | | TL_40 | 298.615 | 223.724 | 0.729 | **-3.840** |
| | | | TL_60 | 300.074 | 225.593 | 0.728 | -4.347 |
| | Cluster 3 | Morohojdens | NO_TL | 92.509 | 69.407 | 0.664 | - |
| | | | TL_20 | 112.027 | 85.828 | 0.490 | -21.098 |
| | | | TL_40 | 106.385 | 78.178 | 0.553 | -14.999 |
| | | | TL_60 | 95.780 | 69.381 | 0.641 | **-3.536** |
| | | Norrbacka | NO_TL | 78.726 | 52.159 | 0.805 | - |
| | | | TL_20 | 78.142 | 51.400 | 0.810 | 0.742 |
| | | | TL_40 | 74.960 | 51.071 | 0.824 | 4.783 |
| | | | TL_60 | 74.483 | 49.726 | 0.825 | **5.389** |
| | | Tallbacka | NO_TL | 73.749 | 60.394 | 0.869 | - |
| | | | TL_20 | 76.877 | 60.018 | 0.860 | -4.242 |
| | | | TL_40 | 67.560 | 49.708 | 0.891 | **8.391** |
| | | | TL_60 | 70.070 | 51.673 | 0.883 | 4.988 |
| Electricity | Cluster 1 | Björnåkersskolan | NO_TL | 156.561 | 102.930 | 0.888 | - |
| | | | TL_20 | 150.836 | 102.716 | 0.897 | 3.656 |
| | | | TL_40 | 148.180 | 101.331 | 0.900 | **5.353** |
| | | | TL_60 | 149.936 | 104.454 | 0.898 | 4.232 |
| | | Byskeskolan | NO_TL | 189.331 | 133.811 | 0.919 | - |
| | | | TL_20 | 183.480 | 131.021 | 0.924 | 3.090 |
| | | | TL_40 | 182.644 | 131.720 | 0.925 | **3.532** |
| | | | TL_60 | 183.938 | 133.475 | 0.923 | 2.848 |
| | Cluster 2 | Morohojdens | NO_TL | 36.596 | 25.239 | 0.850 | - |
| | | | TL_20 | 57.578 | 37.543 | 0.634 | -57.334 |
| | | | TL_40 | 38.312 | 26.758 | 0.837 | **-4.689** |
| | | | TL_60 | 38.357 | 26.307 | 0.836 | -4.812 |
| | | Tallbacka | NO_TL | 75.530 | 55.823 | 0.912 | - |
| | | | TL_20 | 94.926 | 74.139 | 0.861 | -25.679 |
| | | | TL_40 | 74.974 | 55.389 | 0.914 | **0.737** |
| | | | TL_60 | 75.180 | 55.809 | 0.913 | 0.464 |


[12] M. Nawar, M. Shomer, S. Faddel, and H. Gong, "Transfer learning in deep learning models for building load forecasting: Case of limited data," in *SoutheastCon 2023*. IEEE, Apr. 2023. [Online]. Available: http://dx.doi.org/10.1109/SoutheastCon51012.2023.10115128

[13] P. Banda, M. A. Bhuiyan, K. Zhang, and A. Song, *Transfer Learning for Leisure Centre Energy Consumption Prediction*. Springer International Publishing, 2019, p. 112–123. [Online]. Available: http://dx.doi.org/10.1007/978-3-030-22734-0_9

[14] M. Khalil, S. McGough, Z. Pourmirza, M. Pazhoohesh, and S. Walker, "Transfer learning approach for occupancy prediction in smart buildings," in *2021 12th International Renewable Engineering Conference (IREC)*. IEEE, Apr. 2021. [Online]. Available: http://dx.doi.org/10.1109/IREC51415.2021.9427869

[15] Y. Chen, Y. Zheng, and H. Samuelson, "Fast adaptation of thermal dynamics model for predictive control of hvac and natural ventilation using transfer learning with deep neural networks," in *2020 American Control Conference (ACC)*. IEEE, Jul. 2020. [Online]. Available: http://dx.doi.org/10.23919/ACC45564.2020.9147321

[16] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, p. 1345–1359, Oct. 2010. [Online]. Available: http://dx.doi.org/10.1109/TKDE.2009.191

[17] Z. K. Shahid, S. Saguna, and C. Åhlund, "Forecasting electricity and district heating consumption: A case study in schools in sweden," in *2023 IEEE Green Technologies Conference (GreenTech)*. IEEE, Apr. 2023. [Online]. Available: http://dx.doi.org/10.1109/GreenTech56823.2023.10173792

[18] M. Müller, *Information Retrieval for Music and Motion*. Springer Verlag, 2007.

[19] Wannesm, Khendrickx, A. Yurtman, P. Robberechts, D. Vohl, E. Ma, G. Verbruggen, M. Rossi, M. Shaikh, M. Yasirroni, , Todd, W. Zieliński, T. Van Craenendonck, and S. Wu, "wannesm/dtaidistance: v2.3.5," 2022. [Online]. Available: https://zenodo.org/record/5901139

[20] F. Karabiber, "Hierarchical clustering." [Online]. Available: https://www.learndatasci.com/glossary/hierarchical-clustering/

[21] F. AlMahamid and K. Grolinger, "Agglomerative hierarchical clustering with dynamic time warping for household load curve clustering," in *2022 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*. IEEE, Sep. 2022. [Online]. Available: http://dx.doi.org/10.1109/CCECE49351.2022.9918481

[22] J. Brownlee, "A gentle introduction to transfer learning for deep learning," Sep 2019. [Online]. Available: https://machinelearningmastery.com/transfer-learning-for-deep-learning/

[23] V. Chaba, "Understanding the differences: Fine-tuning vs. transfer learning," Aug 2023. [Online]. Available: https://dev.to/luxacademy/understanding-the-differences-fine-tuning-vs-transfer-learning-370

[24] F. Chollet, "Keras documentation: Transfer learning fine-tuning," Jun 2023. [Online]. Available: https://keras.io/guides/transfer_learning/#the-typical-transferlearning-workflow

[25] S. Wei, "Practical process of transfer learning for time series forecasting," Jul 2023. [Online]. Available: https://medium.com/@shouke.wei/a-practical-example-of-transfer-learning-for-time-series-forecasting-9519f3e94087

[26] S. Koli, "How to evaluate the performance of clustering algorithms using silhouette coefficient," Mar 2023. [Online]. Available: https://medium.com/@MrBam44/how-to-evaluate-the-performance-of-clustering-algorithms-3ba29cad8c03