# A Transfer Learning Framework for predictive energy-related scenarios in Smart Buildings

Aurora González-Vidal, José Mendoza-Bernal, Shuteng Niu, Antonio F. Skarmeta, *Member, IEEE* and Houbing Song, *Senior Member, IEEE*

***Abstract*—Human activities and city routines follow patterns. Transfer learning can help achieve scalable solutions towards the realisation of smart cities accounting for similarities between regions, domains, and activities. In this study, we propose a Transfer Learning-based framework for smart buildings to test this hypothesis in energy-related problems. Our framework has two major components: the network creation and the transferable predictive model. In order to create the network that groups buildings sharing characteristics, we evaluated two strategies: a novel clustering algorithm for mixed data, k-prod, and clustering the image-based representation of time series. Then, a combination of Long Short Term Memory and Convolutional Neural Network was trained on the centroids of the clusters for energy consumption prediction. The Coefficient of Variation of the Root Mean Squared Error (CVRMSE) of the predictions in such clusters vary between 3.85 and 58.85 %. The obtained parameters were transferred to the rest of the buildings for predictive purposes, finding accurate results in buildings with little data. Our framework deals with insufficient training data since parameters from scenarios with more sensors can be received. It also carries out state-of-the-art performance on 3 datasets from different sources having in total 533 rooms/buildings and two energy efficiency domains: consumption prediction reducing the CVRMSE in a 21.6 %, and air conditioning usage prediction moving from a 4.18 % to a 0.28% CVRMSE. Our framework extracts more knowledge from available IoT deployments, so that smartness could be spread between environments at a fewer cost given that less individual effort will be needed.**

***Index Terms*—transfer learning, IoT, smart cities, smart buildings, k-prod, CNN, LSTM**

## I. Introduction

In cities, resources are scarce, and the new paradigm of electricity generation and consumption forces the realization of smart urban environments. For addressing this, one of the main courses of action is energy management towards its efficient use [1], [2]. That is why smart meters are on the rise. In Europe, at the end of 2018, around 44% of customers had a smart electricity meter and the penetration rate is expected to reach 58%–71% by 2023 [3]. As of 2020, the penetration of smart electricity meters in North America reached 68 % [4].

Sensor provision, installation and maintenance are essential for functionally deploying IoT solutions to feed algorithms that provide intelligence to cities and buildings [5]. However, despite the reduction of their Energy Consumption (EC) and their cost, sensors are sufficiently intrusive to make it unfeasible to place them on every building for every application. In that sense, we can benefit from comprehensive IoT roll-outs that extract initial knowledge and then use transfer learning techniques in order to extrapolate the relevant information to partially monitored environments, applying higher level knowledge about the similarities of the sites. As it happens with the hardware, complete datasets are not always available under the fast emerging smart cities. Data could be non-canonical, expensive to collect and label, inaccessible [6], [7] or simply non-existent in many cases. For these two reasons, we consider that leveraging transfer learning will help in the realisation of the smart city.

Our work identifies representative buildings and how they relate to others. For finding such archetypes, we have developed two approaches: a new algorithm for clustering mixed sparse data and a visual representation of time series for its clustering. The properties that are considered include physical characteristics, geospatial information and monitored values. Once clusters are created using these characteristics, we select the building that is closer to the centroid of each cluster as the representative of the rest. The transfer learning framework for time series analysis was constructed using Convolutional Neural Networks (CNN) and Long-Short Term Memory (LSTM) Networks under Big Data requirements.

We provide experiments of forecasting EC time series and Heating Ventilating and Air Conditioning (HVAC) behavior. These forecasts are needed for energy-efficiency strategies such as buying green energy and storing it, preventing power peaks and profiling HVAC users for efficiency recommendations. In addition to that, the framework offers deep insights into how to efficiently leverage transfer learning to enable a range of IoT services. The novelty of this article stems from the fact that this is the first effort towards the generalisation of the methods that are created for the emergence of intelligence in specific buildings and therefore, it is a first step towards the sublimation of the smart city. We do this by applying transfer learning techniques to a great variety of buildings for solving two different problems related to energy efficiency. The specific contributions are as follows:

- Creation of a clustering algorithm for mixed static data able to handle missing values,
- Creation of a multivariate time series clustering methodology based on images,
- Development of a transfer learning framework based on CNN-LSTM neural networks,

A. González-Vidal, J. Mendoza-Bernal and A. Skarmeta are with the Information and Communication Engineering Dept, Computer Science Faculty, University of Murcia, 30100, Spain. e-mail: aurora.gonzalez2@um.es.

H. Song are with the Security and Optimization for Networked Globe Laboratory (SONG Lab), Embry-Riddle Aeronautical University, Daytona Beach, FL 32114 USA

Shuteng Niu is with the Dept of Computer Science, Bowling Green State University, Bowling Green, OH 43403 USA

- Evaluation of our framework to transfer knowledge in similar-domain problems: EC prediction to EC prediction and EC prediction to HVAC setpoint prediction for a set of more than 500 buildings.

The paper is structured as follows: Section II outlines the background and related work with regards to transfer learning principles, highlighting CNN and LSTM networks, the applications of Transfer Learning in smart buildings and describes k-pod and k-prototypes clustering, which are the main needed concepts for the development of our framework.

Section III describes our framework for creating a network of buildings and for transferring the learning between predictive models. Section IV presents the experiments, results and discussion and finally, the conclusions and future work are enumerated in Section V.

## II. Background and Related Work

In this section, we define the needed terms and background to build our transfer-learning methodology and algorithms.

### A. Transfer Learning Principles and Techniques

Transfer learning appears under the need to create a high-performance learner for a target domain trained from a related source domain. This can improve the learning efficiency by leveraging the knowledge learnt from related domains. In practical scenarios in which the provision of IoT devices is not feasible in terms of resources, it becomes very useful. The transfer learning framework can be formally defined as follows: Let $D$ be a domain consisting of two components as $D = \{X, P(X)\}$, where $X = \{x_1, ..., x_n\}$ represents the features and $P(X)$ represents their marginal probability.

For a given domain $D$, a task $T$ is defined by two components as $T = \{Y, f(.)\}$, where $Y$ is a label space, and $f()$ is a predictive function learned from the feature vector and label pairs $\{x_i, y_i\}$, where $x_i \in X$ and $y_i \in Y$. The task can be rewritten as $T = \{Y, P(Y|X)\}$. Now, $D_S$ is defined as the source domain data considering $D_S = \{(x_{S_1}, y_{S_1}), ..., (x_{S_n}, y_{S_n})\}$ , where $x_{S_i} \in X_S$ is the i-th data instance of $D_S$ and $y_{S_i} \in Y_S$ is the corresponding class label for $x_{S_i}$. Similarly, $D_T$ is defined as the target domain. Having this stated we can formally define transfer learning as the process of improving the target predictive function $f_T(.)$ from the target domain $D_T$ by using the related information from $D_S$ and $T_S$, where $D_S \neq D_T$ or $T_S \neq T_T$. Depending on which condition is met previously, there are many classifications for transfer learning:

- Heterogeneous transfer learning: $X_S \neq X_T$ which means that the source features differ from the target features.
- Homogeneous transfer learning: $X_S = X_T$, which means that the features are the same in both source and target.
- $P(X_S) \neq P(X_T)$: meaning that the marginal distributions are different.
- $P(Y_S|X_S) \neq P(Y_T|X_T)$: meaning that the conditional probability distributions are different.
- $Y_S \neq Y_T$ means a mismatch in the class space.
- $P(Y_S) \neq P(Y_T)$ is caused by imbalanced labeled data set between the source and target domains.
- Negative transfer learning: when transferring knowledge degrades the target performance.

Another way to classify transfer learning approaches answers the question of "***what*** to transfer?" and consists of: sample instances, feature mapping, model parameters, and association rules [8]. Instance transfer learning is used to improve models' accuracy by finding training samples that have a strong correlation with the target domain and re-weighting them so that they can be used as input for training in the target domain [9]. Feature-based methodologies generate a different feature representation of the source features with the objective of reducing the differences between the source and the target. There are two subcategories, that is, asymmetric - re-weighting the features [10] - and symmetric transfer - finding a common latent feature space [11]. The parameter-based category transfers knowledge through the shared parameters of source and target domain learner models or by creating multiple source learner models and optimally combining the reweighted learners (ensemble learners) to form an improved target learner [12]. Typically, when dealing with Neural Networks (NN), the model begins with random weights near zero and adjusts them in the process. Preparing a deep NN requires a large amount of labelled data. Using transfer learning we can start the network with prepared weights from a comparable domain and adjust them to the explicit source. This reduces the time and requirements for creating an accurate network. The last and least used approach is known as relational-based. It is based on some defined relationship between the source and target domains [13], [14].

Much research effort has been devoted to transfer learning in applications such as image classification [15], natural language processing [16] and product recommendation [17], while transfer learning in smart cities and buildings is still less studied. However, smart buildings applications that are using data are sometimes even more crucial since decisions at the city level that involve infrastructure are more definitive than others. Transfer learning can play a crucial role in this decision-making process [18].

In this paper, we have used CNNs and LSTM Networks for our transfer learning strategy and therefore, we review some basic principles in the following.

*1) CNN and LSTM Networks:* A CNN is a class of deep neural networks, most commonly applied to analyzing visual images that use convolution instead of general matrix multiplication in at least one of their layers. Convolution can be seen as a way of looking at a function's surroundings to make accurate predictions of its outcome. Even though they are known to be appropriate for image processing, some other areas where CNN are used are natural language processing, time series, proteomics and audio data classification [19], [20], [21]. The layers of a CNN have the so-called neurons arranged in three dimensions: width, height and depth, as nodes connecting input and outputs; and the neurons in a layer will only be connected to a small region of the previous layer. On CNN, the network *learns* the filters that in traditional algorithms were hand-engineered. This independence from prior knowledge in feature design is a major advantage for this method.

Also in the family of Artificial Neural Networks (ANNs), a Long Short-Term Memory network (LSTM) is a type of Recurrent Neural Network (RNN) that retain information over longer periods of time. Unlike CNN, LSTM networks have feedback connections, implying that for making a decision, they selectively consider the previous input(s) and the output(s) that they have previously learned, therefore solving the vanishing gradient problem. On LSTMs, the model is trained using back-propagation, meaning that information about the errors are sent in reverse through the network, so that it can alter the weights. This kind of network is well-suited to time series data. They can not only process single data points (such as images), but also entire sequences of data (such as videos).

The architecture of an LSTM network has the form of a chain of repeating modules or blocks. Each block has typically a memory cell and 3 gates: an input gate, an output gate, and the so-called forget gate. The memory cell saves the state of the LSTM unit, while the gates are a way to optionally let information through. The gates are composed out of a sigmoid function and a pointwise multiplication operation. The sigmoid function outputs numbers between zero and one so that it controls the extent to which a new value flows into the cell. Similarly, the forget gate and output gate control the extent to which a value remains in the cell and its usage to compute the output activation of the LSTM block, respectively.

In recent years the LSTM approach has been combined with the convolution mechanisms. This has led to the ConvLSTM, which advantages have already been seen as superior in several recent applications [22], [23], [24]. The ConvLSTM attributes derivate from the Convolution and the LSTM layer.

Let $X_1, ..., X_t$ be the inputs, $C_1, ...., C_t$ the cell outputs, $H_1, ...., H_t$ the hidden states and gates $i_t, f_t, o_t$. The key equations of ConLSTM are shown below, where '*' denotes the convolution operator and '$\circ$' denotes the Hadamard product:

$$
\begin{aligned}
i_t &= \sigma(W_{xi} * X_t + W_{hi} * H_{t-1} + W_{ci} \circ C_{t-1} + b_i) \\
f_t &= \sigma(W_{xf} * X_t + W_{hf} * H_{t-1} + W_{cf} \circ C_{t-1} + b_f) \\
C_t &= f_t \circ C_{t-1} + i_t \circ \tanh(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c) \\
o_t &= \sigma(W_{xo} * X_t + W_{ho} * H_{t-1} + W_{co} \circ C_t + b_o) \\
H_t &= o_t \circ tanh(C_t)
\end{aligned}
$$

### B. K-pod and k-prototypes for clustering

In transfer learning, it remains an open question how to choose the main subjects from which knowledge is going to be obtained. For this, we propose a clustering approach. A cluster refers to a collection of data points put together because of certain similarities, and the representation of their center is the centroid. The K-means algorithm starts with a group of randomly selected centroids and then performs iterative calculations to optimize their positions until they stabilize. K-means is useful for numerical data. When a problem presents categorical data, k-medoids [25] is used. K-medoids computations are based on medians instead of means. The k-prototypes algorithm, through the definition of a combined dissimilarity measure, integrates both [26]. K-Pod is the adaption of K-means to missing data [27]. In K-Pod the missing features are replaced by those of the centroids, given that the underlying assumption of this kind of methods is that each observation of the dataset is a noisy realization of a cluster centroid.

### C. Energy-related problems in Smart Buildings

Consumption prediction methods can be classified into engineering, statistical, and machine learning methods [28].

Engineering models, also known as white-box models, use numerical equations by considering the physical properties of building characteristics [29]. Such information is often very difficult to obtain. Some researchers have tried to simplify engineering models to effectively predict energy consumption [30], yet those models' results may not have been completely accurate [31]. Statistical methods use mathematical formulas to correlate energy consumption data with influencing factors. The calculation processes are straightforward, but they often cannot handle complex interactions between factors, so they are not flexible and often have poor prediction accuracy [32].

Several machine learning techniques have been applied in the context of smart buildings management in independent efforts. We claim that cooperation between domains is key in order to fully unlock the potential of IoT research. In the survey [33], we can see a comparison of the four main categories of machine learning applied to buildings: supervised, unsupervised, semi-supervised and reinforcement learning algorithm. Regarding supervised learning energy consumption prediction is targeted in [34] using SVM, in [35] a hybrid neuro-fuzzy inference system and a multivariate feature selection strategy for the same purpose is proposed in [36]. Other works present hybrid approaches (ANN and K-pattern clustering) for predicting user activities in the smart environment, including but not restricting to those related to energy consumption [37]. It is also interesting to mention that many works combine LSTM with other strategies such as decomposition mechanisms or data augmentation [38] for energy consumption prediction [39].

The design of a building energy consumption data-driven model consists of four steps: data collection of historical and current data such as baseline consumption and outdoor weather conditions, data preprocessing (cleaning, transformation, and/or data reduction), model training, that aims to fit the model to the train data and model testing in order to verify that the model provides good results in unseen data.

The main design decisions are normally constrained to the type of building (residential and non-residential), the temporal granularity of data collection and the predictive horizon (from sub-hourly to yearly) and type of energy consumption predicted, the features to be used in the analysis and different data sizes between scenarios. For new buildings and most existing buildings without advance IoT deployments, there is a lack of sufficient data to train data-driven predictive models [40]. Our method's novelty and value stem from the fact that we develop a strategy that can be used to improve accuracy in buildings with limited training data. As previously stated, the collection of data is one of the main design challenges in smart building problems, and we are providing a novel method to overcome such an issue with remarkable results.

Transfer learning can be leveraged to accelerate the smart city development and this was termed in [41] as the urban

transfer learning paradigm. The literature on transfer learning in buildings is extensively reviewed in [42]. A recent study [43] uses feature extraction and weight initialization as transfer learning strategies for the prediction of a 24h horizon of EC in buildings. This study serves us as a baseline to develop transfer learning-based methods for building energy management in general. Feature extracting and domain adaptation combined in one model training process for transfer learning in consumption prediction scenarios between 3 buildings has been investigated in [44], and LSTMs for the transference between 1 office buildings and 20 offices [45]. The goals on these 3 studies are similar to ours, however, they do not find the relationship between buildings so as to share other kinds of models. At the same time, they only considered 1 database for their, and the fact that they chose a subset of buildings hinders the generalisation of the study.

## III. Framework

### A. Creation of a network of buildings

We aim to identify representative buildings or their stereotype characteristics and how they relate to other buildings considering factors that should include but also go beyond physical characteristics and geospatial information. In that sense, we could either select those who best represent the rest for their sensorization or know how much a sensorized building relates to others in order to transfer the knowledge between them. For this purpose, we have created two methodologies:

*1) Creation of k-prod for meta-data / static clustering:* When analysing the static characteristics of a building we can find numerical variables (dimensions) and categorical variables (industrial, educational). In addition, many variables/values may be missing in real-world data since they are usually manually annotated.

Our proposed algorithm, k-prod is an adaption of K-prototypes so that it deals with missing data. We have reused the idea of the k-pod, in which k-means missing features are replaced by those of the centroids, given that the underlying assumption of this kind of methods is that each observation of the dataset is a noisy realization of a cluster centroid.

We first compute the mode or mean of each feature of the matrix, depending on the kind of data and introduce such values in the missing points of the matrix. The column indexes for categorical data are specified as inputs of the algorithm. We first fit the K-prototypes model with the well-known method in [46], using the implementation of kmodes python package[1] to the filled data $\hat{X}$ . After that, we iteratively change the missing points by the computed centroids (distinguish between numerical and categorical) and re-fit the K-prototypes algorithm using the current centroids as initialization. The cost function is the Euclidean dissimilarity function for the numerical variables and the matching dissimilarity function [47] for the categorical ones. Once the labels do not change or the maximum number of iterations is achieved, the algorithm finishes. This process is summarised in Algorithm 1.

Given that, we have created an algorithm that helped us clustering the building by using solely their meta-data.

[1]https://github.com/nicodv/kmodes

**Algorithm 1** k-prod: Adaptation of K-prototypes to missing values

**Inputs:** $X$: a $n \times m$ matrix
`n_cl`: an integer - number of clusters
`cat_index`: an array with the ids of categorical columns
`max_iter`: an integer - max. iterations
**Output:** `labels`: an array with the assigned label per row
`centroids`: a centroid per cluster
$\hat{X}$ a matrix with imputed missing values
$\gamma$: an integer - the weighted sum of costs
**The algorithm starts:**
labels$_0$ = NULL
$\hat{X}$ = X where missing values are filled with the mode or mean of the column
cls = apply KPrototypes for `n_cl` to $\hat{X}$
**for** i **in** 0 **to** max_iter **do**
`labels` = current cls labels
`centroids` = current cls centroids
$\gamma$ = current cls gamma
$\hat{X}$ = X where missing values are filled with the `centroids`
**if** `labels == labels`$_0$ **then**
**break**
**end if**
`labels`$_0$ `= labels`
cls = apply KPrototypes for `n_cl` using `centroids` as initialization to $\hat{X}$
**end for**
**return** `labels`, `centroids`, $\hat{X}$, $\gamma$

*2) Clustering time-series data:* The data collected from sensors in smart buildings is in time series form. For each building, we will have some static data, as above specified, and several time series that are generated, including but not limited to EC, indoor temperature, occupation. The clustering of multivariate time series is always a complicated task and we have decided to explore a new way to perform it. One of the problems where deep learning excels is image classification, so we propose to convert time series data into heatmap images in order to get an image representation of each building. Considering buildings EC and outdoor temperature in order to represent each building we would obtain the images shown in Fig. 1 for a single building. This strategy can incorporate any other kind of time series data into the representation.

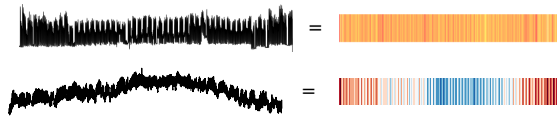


Fig. 1: EC and outdoor temperature of a building time series transformed into heatmaps.

From a deep learning perspective, it is common to apply transfer learning to reduce the training cost in image classification. In addition, fine-tune existing models pre-trained on massive data sets is a common transfer learning method. ImageNet [48] is a project which aims to provide a large image

database for research purposes. The ImageNet Large Scale Visual Recognition Challenge (ILSVRC)[2] is an annual competition organized by the ImageNet team since 2010, where research teams evaluate their computer vision algorithms for visual recognition tasks such as object detection in images. The training data used for this purpose is a subset of ImageNet with 1.2 million images belonging to 1000 classes. We will use ImageNet weights for initialization. The winners of ILSVRC have released their models to the open-source community. There are many accessible models such as AlexNet, VGGNet, Inception, ResNet, Xception. Given that we are interested in extracting the encoded features from the images we will remove the top dense layers, which are intended for classification problems. In our methodology, we have considered vgg16, vgg19 and resnet50 models. Those CNN models will provide 3D vectors that represent the image. We need to flatten these for the clustering algorithms. We can also use Principal Component Analysis (PCA) for dimensionality reduction so that we ease the clustering algorithms by reducing the features.

The vgg16, vgg19 and ResNet50 models were compiled using the cross entropy loss function , the stochastic gradient descent optimization function [49] and Accuracy (Acc) and Mean Square Error (MSE) metrics. The categorical cross entropy formally is designed to quantify the difference between two probability distributions. It is well suited to classification tasks, since one example can be considered to belong to a specific category with probability 1, and to other categories with probability 0. The stochastic gradient descent optimization algorithm estimates the error gradient for the current state of the model using examples from the training dataset, then updates the weights of the model using backpropagation. The metrics are defined as $Acc = \frac{\#\text{correct predictions}}{n}$ and $MSE = \frac{1}{n}\sum_{i=0}^{n}(y_i - \hat{y}_i)^2$, where $n$ is the number of predictions, $y_i$ are the real values and $\hat{y}_i$ are the predicted values.

After compiling, the networks outputs 3D vectors that represent the image. We then flatten these for the clustering algorithms to start working with them. Before that, we also create PCA instances for each covnet output. This process is depicted in Fig. 2.

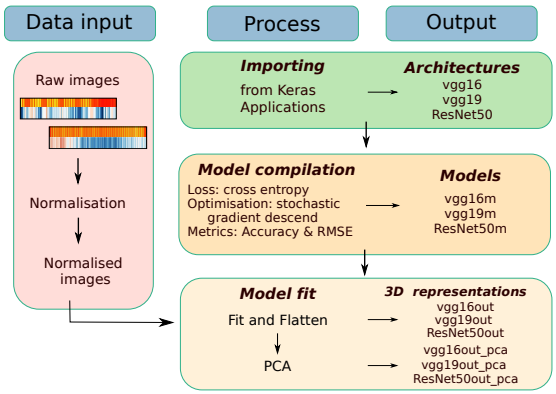



Fig. 2: Feature extraction from raw images. The inputs are the images and the output are their representations, ready to be introduced to the clustering algorithms.

[2]http://www.image-net.org/challenges/LSVRC/

Next, we use K-means and Gaussian Mixture algorithms to create and fit models that group images that look alike together. Both models return as many centroids as was indicated in the number of clusters. We will compute the distances between all the members of the clusters and the centroids in order to choose representative buildings for each cluster.

### *B. Transfer learning algorithm for time series prediction*

We have adopted instance-based transfer with the strategy of using pre-trained models for weight initialization. The application of re-weighting was used to incorporate source domain data into the target task training process. The predictive models are based on the combination of 1D CNN and LSTM networks. The conventional deep CNNs are designed to operate exclusively on 2D data such as images and videos. As an alternative, a modified version of 2D CNNs called 1D CNNs have recently been used [50]. 1D convolutional layers can automatically extract local temporal features from time series and they are advantageous in many ways when compared to 2D ones: less computational complexity, configurations with networks having10K, can run on a standard computer and are well-suited for real-time and low-cost applications [51]. They serve as a data preprocessing step and has been reported to be useful in reducing computational costs and enhancing model performance. The number of 1D convolutional layers, the number of filters, kernel sizes, strides and activation function types are arbitrarily chosen and future work should address the optimisation of the parameters.

After that, recurrent layers are adopted to capture interactions among temporal features for accurate predictions. The recurrent units are LSTM units, considering their excellence in modeling long-term dependency and tackling the vanishing or exploding gradient problem. A further extension of the combination of CNN and LSTM is to perform the convolutions of the CNN (e.g. how the CNN reads the input sequence data) as part of the LSTM for each time step. The combination that we used is called a Convolutional LSTM, or ConvLSTM for short, and it is also used for spatio-temporal data [52]. Unlike an LSTM that reads the data indirectly in order to calculate internal state and state transitions, the ConvLSTM is using convolutions directly as part of reading input into the LSTM units themselves.

Regarding the implementation, we have used the ConvLSTM2D class provided by the Keras Library. The layer expects a sequence of 2D images, having input shape [samples, timesteps, rows, columns, features]. For our purposes, we can split each sample into subsequences that we chose to be 7 (a week), and columns will be the number of time steps for each sequence, that we chose to be 24 (a day). The number of rows is fixed at 1 as we are working with one-dimensional data. Finally, we had 2 features (consumption/hvac use, temperature). Therefore, our input shape configuration is [samples, 7, 1, 24, 2]. The methodology consists of, for every cluster, train a ConvLSTM network using the representative buildings' data. Then, the weights of such network are used for initializing the ConvLSTM networks of the rest of the buildings from the same cluster. Our transfer learning is safe in the sense

that, since we are only sharing the weights of the models, it could be considered as a Federated Learning approach, where each building is a client [53]. In that sense, our approach potentially preserves the privacy of the buildings. Further investigation with privacy-preserving techniques should be done in that sense in order to diminish the security risks associated with data sharing [54]. Some works are already investigating how to incorporate Federated Learning to the smart buildings scenarios in order to keep privacy [55].

## IV. Results and Discussion

In this section, we describe the results from our framework. Data and scripts to reproduce the experiments are available at a GitHub repository [3].

### A. Data description and preparation

Open data favours science. Open and reproducible research practices enable scientific reuse, accelerating future projects and discoveries in any discipline when accompanied by workflows and explanations [56]. Many studies suggested in the past the need for a set of time series benchmarks so that contributions are generalizable enough [57].

In recent years, some initiatives of great importance have been carried out for the acquisition of data and its publication as a basis for the creation of open and shareable data repositories for building performance research.

In this work, we have used 3 different open databases on buildings operation.

The first one is *The Building non-residential Data Genome Project* [58] (DGP), which consists of one year hourly EC of 507 non-residential buildings, 38 weather condition datasets that can be related to them and building meta-data. The building meta-data provides the general building information, such as the building location, primary usage type, the total floor area etc. for the majority of buildings. Buildings are mainly located in America and Europe, while only 5 buildings in Asia are included. The buildings are categorized into five primary usage types: there are 156 Offices, 105 primary or secondary school classrooms, 95 university laboratories, 81 university classrooms and 70 university dormitories.

Given that there are very few buildings in Asia in this dataset, we decided to consider as the second database the *Indian BuiLdings Energy coNsumption Dataset (I-BLEND)* [59]. This database contains minute-based energy-related information for 52 months on 7 different buildings on an autonomous research institute in Delhi, India. We aggregated the data for having an hourly sampling. The database also includes datasets about occupancy, institute calendar, building architecture details, and four months of local weather (temperature, humidity). I-BLEND consumption dataset is vaster than the previous but it also presents more missing values.

An extra benefit from adding data from several databases is that we are covering more of the semantics that are used for buildings' metadata, which can help enable interoperability across the different categories of buildings.

The last database that we are considering will be referred to in this work as the Langevin-HVAC database [60]. It delivers occupants' energy behavior data collected through a year within 15 offices and amongst 24 workers at the Friends Center office building in Philadelphia, USA. It consists of a longitudinal comfort survey data collection together with continuous measurements of the weather, the local indoor environment around each subject, and certain of the subjects' behavioral actions. To the authors' knowledge, this is the first time that the dataset is used for predictive purposes. The preprocessing steps consists of the assignation of the rows to specific rooms and participants (subjects), select some continuous variables regarding thermostat information, weather and room conditions, and aggregate them in an hourly manner. This preprocessing can be reproduce using the code within the data folder at the already-mentioned repository[3].

All the databases contain non residential buildings, since they are offices, schools, laboratories, etc. Consumption patterns between residential and non residential buildings are complementary and the authors will delve into it in the future using available open datasources such as the ones in [61], [62].

### B. Clustering results

Firstly, we applied the novel k-prod clustering to all databases. In order to select the number of clusters we used the elbow method heuristic, that consists of plotting the explained variation (cost in our case) as a function of the number of clusters, and picking the elbow of the curve as the number of clusters. It selected 5 clusters as Fig. 3 shows.

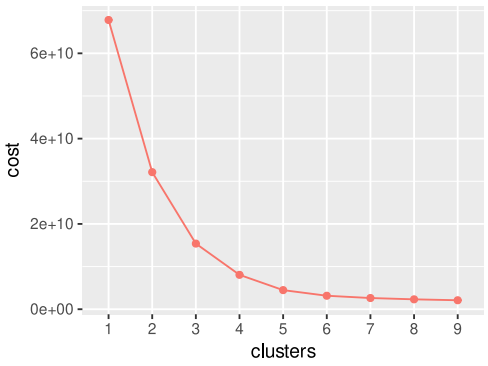



Fig. 3: Elbow method heuristic for number of clusters choice.

Secondly, we applied our image time series classification. It is important to remark that for performing the image clustering, we are only using 2000 data points as it is assumed for the further experiments that buildings started only recently to gather data, having insufficient data to train a neural network successfully by themselves. For easing access we saved the images in a public repository[4] and we built an image downloader with multiprocessing for making it faster. We have used the deep learning framework Keras [63] to define an architecture capable of accepting multiple inputs of building information, and then train a single end-to-end network with this data.

[3]https://github.com/auroragonzalez/TLIAS

[4]https://bitbucket.org/aurorax/datangi

| | Kvgg16 | Kvgg19 | KResNet50 |
|---|---|---|---|
| Kvgg16 | 1 | 0.478 | 0.887 |
| Kvgg19 | 0.478 | 1 | 0.563 |
| KResNet50 | 0.947 | 0.847 | 1 |

TABLE I: Adjusted Rand Index for test set using all clustering models.

We randomly selected a train set of buildings, that constitute 70 % of the total. For comparison between models, we have also chosen 5 clusters in this case, given the results shown in Fig. 3. The data of those buildings was used for obtaining images and then, 12 combinations of clustering algorithms were tested (vgg16, vgg19, ResNet50, vgg16-pca, vgg19-pca and ResNet50-pca 2 times, one for K-means and another one for GMM). With and without PCA, the centroids are the same. Then, we obtained the features for the test buildings' images and fit them with the previously created clustering models in order to obtain their labels. Given that there are 12 different options, we have computed the Adjusted Rand Index in order to find the most common groupings for making a decision. The Rand Index is one of the most popular alternatives for comparing partitions and has been rediscovered and/or modified by different authors [64]. Let $n$ be the number of elements (in our case, the number of buildings included in the test set). There are $\binom{n}{2}$ distinct pairs that could be found. In between two partitions $U$ and $V$, we can find:

- Agreement between partitions: objects in the pair are in the same class (in different classes) in $U$ and also in the same class (in different classes) in $V$
- Disagreement between partitions: objects in the pair are in different classes (in the same class) in $U$ and in the same class (in different classes) in $V$ (disagreement).

$$RandIndex = \frac{\#\text{pairs in agreement}}{\text{total number of pairs}} = \frac{\#\text{pairs in agreement}}{\binom{n}{2}}$$

RI is not able to take into consideration the effects of random groupings. When the number of clusters is increased, more and more example pairs will be in agreement because they are more likely to be not grouped together. To counter this drawback, an adjustment is made to the calculations by taking into consideration grouping by chance through the generalized hyper-geometric distribution, creating the Adjusted Rand Index, whose formula and further theoretical explanation can be seen in [64]. The ARI is recommended as the index of choice for measuring agreement between two partitions in clustering analysis [65] . A greater ARI indicates a greater agreement between clustering strategies.

Given that using PCA would not alter our results, we have reduced our models to 6. Table I shows that the greater metric is obtained when comparing the partitions from Kvgg16 and kResNet50: 0.947. That means that both clustering methods render almost the same results. For choosing one between them, we then seek which of them has a better agreement with the others. Kvgg16-Kvgg19 is 0.478 and kResNet50-Kvgg19 is, in average, 0.705. Therefore, we choose Kvgg16 as the most representative model.

| Method | Centroid | RMSE | CVRMSE | MAE | MAPE |
|---|---|---|---|---|---|
| k-prod | 1:UnivLab-Louie | 13.46 | 12.14 | 10.53 | 10.07 |
| k-prod | 2:UnivDorm-Leann | 12.46 | 5.25 | 49.55 | 9.37 |
| k-prod | 3:Office-Benthe | 31.21 | 17.47 | 21.7 | 13.22 |
| k-prod | 4:PrimClass-Johnathan | 2.41 | 30.58 | 1.62 | 224.45 |
| k-prod | 5:PrimClass-Julianna | 4.5 | 34.58 | 2.57 | 452.2 |
| IC | 1:UnivClass-Conor | 0.8 | 16.09 | 0.6 | 12 |
| IC | 2:PrimClass-Jimmien | 8.05 | 43.9 | 4.61 | 268.97 |
| IC | 3:Office-Madeleine | 5.86 | 8.96 | 4.5 | 6.98 |
| IC | 4:PrimClass-Jacob | 10.46 | 22.5 | 6.96 | 187.42 |
| IC | 5:UnivDorm-Carter | 3.67 | 4.23 | 2.81 | 3.21 |

TABLE II: EC prediction metrics for the 5 centroids of k-prod and Image Clustering (IC).

### C. Transfer Learning Results

For each clustering centroid that was found with k-prod and with image clustering, the closer building in terms of their characteristics are selected since from our algorithm, centroids do not have to coincide with any observation, meaning with any particular building. The distance between each building and the centroid is computed as explained in subsection III-A1, that is using the cost function. We have decided to test our approach not only with 5 centroids, as the elbow method in Fig. 3 suggested but also with 15 centroids because it seemed more appropriate considering that we have more than 500 buildings available for study. The latter was an arbitrary choice that could depend on each of the set-ups.

We created EC predictive models using the data from the 5 and 15 centroids of each of the approaches respectively and also we stored the weights. The metrics we used include Root Mean Square Error (RMSE) and its coefficient of variation (CVRMSE), Mean Absolute Error (MAE) and Mean Average Percentage Error (MAPE). The performance on the centroids can be seen in Tables II and III respectively, including also the names of the reference buildings. In k-prod we were able to use all buildings in order to find the centroids because its input does not include energy consumption. In the case of image clustering, we have used only a 70 % of the buildings as train and a 30 % as test. The test set only contains 2000 values of temperature and consumption, to be aligned with the later prediction strategy. We can see that having a greater number of centroids is translated into a greater variety in the metrics for our models. It creates very accurate models such as the one for centroid 10 (3.85 % of CVRMSE) and very inaccurate ones such as the model for centroid 4 ( 49.74 % of CVRMSE). It is not the goal of our research to create the best models for each of the buildings, however, we would expect a correlation between the accuracy of the model in the original building and the models that use it as "initialisation" in our transfer learning strategy, meaning the models for the buildings in the same clusters.

*1) Same-Domain Transfer Learning with k-prod and image clustering: EC to EC:* In this first scenario, we have assumed that we only had 2000 values of EC and weather from the rest of the buildings and used our neural networks algorithm in two ways.

- Without Transfer Learning: using only the prior 2000 data points for estimating next 24h EC

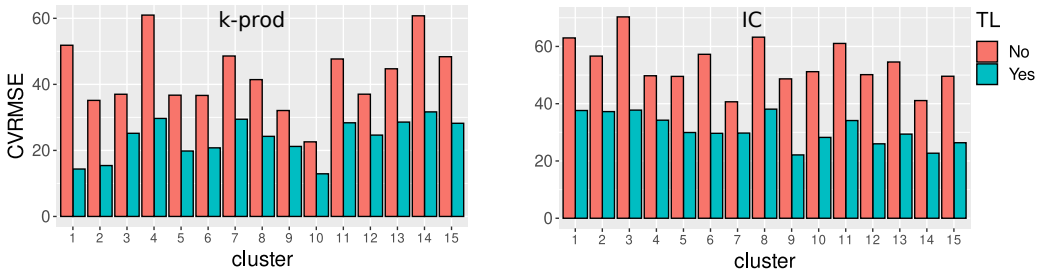



Fig. 4: CVRMSE comparison between using the pre-trained models as initialization and not using it in the DGP dataset for 15 clusters obtained using k-prod (left) and Image Clustering (right).

| Centroid | RMSE | CVRMSE |
|---|---|---|
| 1:PrimClass-Julianna | 4.55 | 34.98 |
| 2:UnivClass-Anya | 2.67 | 28.19 |
| 3:Office-Gustavo | 40.79 | 18.15 |
| 4:PrimClass-Jaylin | 1.19 | 49.73 |
| 5:UnivDorm-Marquis | 18.67 | 7.94 |
| 6:PrimClass-Ervin | 4.23 | 35.7 |
| 7:PrimClass-Jacquelyn | 3.37 | 34.66 |
| 8:Office-Maximus | 10.14 | 11.44 |
| 9:PrimClass-Johnathan | 2.62 | 33.17 |
| 10:UnivLab-Marie | 12.81 | 3.85 |
| 11:Office-Erik | 3.43 | 30.6 |
| 12:PrimClass-Johnathon | 5.61 | 31.55 |
| 13:UnivDorm-Alka | 28.79 | 16.13 |
| 14:Office-Benthe | 36.78 | 20.59 |
| 15:Office-Jude | 58.89 | 15.01 |

TABLE III: EC prediction metrics (reduced) for the 15 centroids of k-prod.

- With Transfer Learning: using the prior 2000 data points and the weights for their representative building in order to estimate the next 24h EC

In Fig. 4 (left) we can see that using our Transfer Learning strategy reduces the error in all groups when using as initialization the prototypes obtained from k-prod clustering. On average, the CVRMSE was improved a 19 % when using transfer learning (23.64%) compared to not using it (42.78 %). We have not depicted the case of 5 clusters, but the average improvement is 21.6 % in CVRMSE.

In Fig. 4 (right) we can see that using our Transfer Learning strategy reduces the error in all groups also when using as initialization the prototypes obtained from the image clustering. In average, not using the transfer learning strategy produces a CVRMSE of 53.77 %, while using it gives a 30.88 % , implying an improvement of 22.9 % in this metric. In the case of RMSE, the average for all buildings without transfer learning is 27.1 and with transfer learning, it is reduced to 16.2, implying a reduction of the error in 10.9 KWh. We have not depicted the case of 5 clusters, but the average improvement is 25 % in CVRMSE and 10.5 KWh with regards to the RMSE.

We can see that, in general, the results are better with the k-prod approach, which means, using the meta-data of the building for the construction of our relationship network.

We have separated the results for the I-BLEND dataset, given that the source is different in order to show how our strategy adapts to a different source. As it can be seen in Table IV, where the representative for each of the buildings (the prototype) is between brackets and where we can also see how our TL approach improves the CVRMSE in several points for the 3 buildings.

| Building | No TL | k-prod clustering | | Image clustering | |
|---|---|---|---|---|---|
| | | 15 cl | 5 cl | 15 cl | 5 cl |
| Academic Building | 40 %<br>- | 17.7 %<br>(5) | 17.88 %<br>(3) | 19.38 %<br>(6) | 20%<br>(3) |
| Boys Dorm | 27.5<br>- | 20.7 %<br>(6) | 20.6 %<br>(1) | 21.44<br>(5) | 23.48 %<br>(1) |
| Girls Dorm | 17.5%<br>- | 13.1 %<br>(15) | 10.64 %<br>(1) | 11 %<br>(15) | 10.66%<br>(1) |

TABLE IV: CVRMSE comparison between NO TL and TL strategies in the I-BLEND dataset for 5 and 15 clusters using both k-prod and image clustering.

*2) Similar-Domain Transfer Learning: EC to HVAC preferences:* Finally, we analysed the goodness of our transfer learning approach between different domains. We have used a dataset that collects the HVAC setting points and outdoor temperature. We have tried it in 3 different rooms. Such a study was not providing a lot of meta-data so we were not able to find which group was better to be assigned to. Given that, as a preliminary approach, we defined an experiment to predict the next 24h HVAC setting point for the 3 rooms using the weights of all the 15 buildings selected as centroids when using our k-prod algorithm. Those buildings are completely unrelated to the rooms from which we have collected the HVAC setting points, however, we assume that EC and HVAC usage are related to a certain extent. We have used our predictive algorithm again with 2000 datapoints with and without transfer learning and the results are provided in Fig 5.

Fig. 5 shows the CVRMSE of the prediction of HVAC setpoint for 3 particular rooms. First of all, the scenario without transfer learning is a straight line because we are not using the centroids for transferring anything. We can also see that when using the model trained on the centroid of cluster 3, all rooms present a lower CVRMSE when using transfer learning. Therefore, it is positive to use the weights of a neural network trained in EC data in order to initialize a network with the goal of predicting HVAC setting points. We can find in Fig. 6 a more extensive evaluation, where each rooms' model is initialized using the 15 possible centroids from the clustering analysis and the results are portrayed in the form of boxplots.

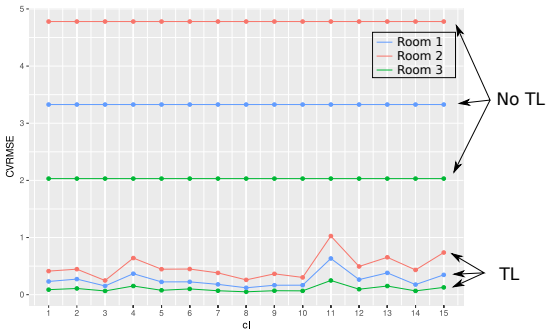



Fig. 5: CVRMSE comparison between not using our TL strategy (straight horizontal lines) and using it for 3 rooms in the Langevin-HVAC dataset.

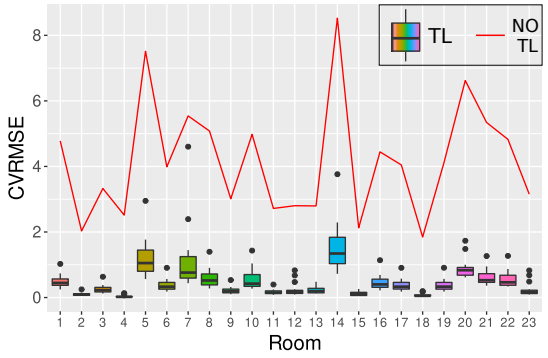



Fig. 6: CVRMSE comparison between not using our TL strategy (red line) and using it for all rooms in the Langevin-HVAC dataset. The boxplots show the CVRMSE for each room using as a representative all 15 possible buildings.

Also, we can see that the line that shows the CVRMSE when not using any transfer learning strategy stays always higher than every scenario with transfer learning.

Since there is not enough meta-data to perform the k-prod clustering, we have only tried the image clustering approach. Using it, all the 23 tested rooms belonged to cluster 3. Using the building that was chosen as representative from cluster 3, Office-Gustavo, we obtained in average an improvement of 4 % in CVRMSE, because it went from 4.18 % to 0.28%.

These preliminary results open up a wealth of opportunities for further research.

*3) Comparison with previous studies on the same datasets:* To the authors' knowledge, this is the first time that I-Blend dataset and the Langevin-HVAC database are used for transfer learning and predictive purposes. The I-Blend dataset was only aggregated and used as such, however the Langevin-HVAC was transformed so as to generate the HVAC set-point prediction scenario. This makes the later incomparable with previous studies, but provides the possibility to realise new studies with a different perspective on the same dataset.

The DGP dataset has been used in the literature for transfer learning purposes in 5 occasions as can be seen in Table V. 4 of these studies do not present a grouping strategy, they either do it manually or randomly. The general improvements on CVRMSE are also lower than with our strategy and finally, they perform their experiments in less buildings, which makes it less robust and reliable in terms of generalisation. We consider that our methodology is more general and can be easily adopted and improved in comparison to theirs. It is also important to note that none of the other transfer learning studies have investigated the transference of knowledge between different domains as we do.

With regards to predictive purposes in general, the DGP dataset has been used several times as it can be seen in Table VI. Studies that realised only 1 step prediction show better metrics as expected compared to ours, that performed 24 steps prediction. The work in [66] focused on 3 step prediction and deliver metrics that are more comparable to ours. It is important to mention that our objective was focused on scenarios with less data, and therefore is is complicated to make a fair comparison between works.

## V. Conclusion and Future Work

In this paper, we have proposed a framework that can transfer the knowledge from sensorized buildings to a wide amount of buildings that have fewer data from similar domains. The framework includes 2 methodologies that select the best buildings to transfer the knowledge from. We have performed experiments using as source domain the prediction of EC in 5 and 15 main buildings. There were 3 target domains with different challenges: 1) EC prediction on other buildings from the same database using less data 2) EC prediction on other buildings from a different database using less data 3) HVAC setpoint prediction (related domain) with few data.

In all these scenarios, our framework was successful when compared to developing a predictive strategy on the target domain without using it. When having a very reduced dataset (2000 points), our approach improved the CVRMSE of prediction a 21.6 % (k-prod) and a 25 % (IC) on average for all buildings when using data from the same domain (EC-EC). When using data from a different domain (EC-HVAC) we went from a 4.18 % to a 0.28 % CVRMSE.

This work opens up a great number of research lines:

- To investigate how to include knowledge from more than one building simultaneously, and how the distance to the centroid influence the accuracy,
- To investigate knowledge transfer between residential and non residential buildings,
- To investigate how to adapt the framework to a Federated Learning scenario for data privacy,
- To develop an optimisation strategy for the deployment of sensors in buildings and perform a cost-benefit analysis,
- To find the optimal number of buildings to be sensorised in order to achieve better accuracy and the cut-off point for the amount of data to be used in the target domain,
- To optimize the parameters of Conv2DLSTM,
- To account for inconsistent features between datasets.

## Acknowledgment

This work has been sponsored by the European Commission through the H2020 PHOENIX (g.a. 893079) and NGI explorers (g.a. 825183) projects and by

| Year and Reference | Predictive Approach | Number of buildings | Time range | Output horizon | Best results (metrics) |
|---|---|---|---|---|---|
| 2020 [66] | Random Forest, M5 model tree and Random Tree | 5 | 4 scenarios: 3,6,9 and 12-month datasets | 3 steps: 1, 12, 24h | RF- MAPEs between 11.3 and 52.65 % |
| 2020 [67] | Arquitecture: 2 convolutional layers, 2 LSTM layers and 3 embedding layers | 407 randomly selected | 1 year | 24 steps: 1-24h | Performance Improvement Ratio (PIR) reduced at least 67% of RMSE |
| 2021 [68] | Random Forest, Gradient Boosting, Linear and Extreme Gradient Boosting Regression | 507 | NA (we assume 6 years) | NA (we assume 1 step: 1h) | LGB best model with RMSE = 50.0396 |
| 2021 [69] | Ensemble learning, SVR, ANN and MLR | 1 office | 12 weeks | 1 step: 1 h | Ensemble learning with 17.7%, 16.1%, 15.4%, 15.8%, 15.6% of CVRMSE under 20%, 40%, 60%, 80% and 100% data availability |
| 2022 [70] | Hybrid RF-LSTM based on CEEMDAN | 5 (one of each type) | 3 months | 1 step: 1h | Average MAPE = 4.23 % |
| 2022 [38] | Decomposition mechanism (EEDM andWT) + lstm | 16 in 6 timezones | 1 year | 1 step: 1h | Average MAPE = 15.3% |
| 2022 [39] | Data augmentation + bi-LSTM | 52 | 276 days 9 months | 1 step: 24h | Average CV-RMSE = 13.86% |
| 2022 Our proposal | CNN-LSTM | 517 | 2000 datapoints 2.5 months | 24 steps: 1-24h | 30.88 % CVRMSE and RMSE = 16.2 |

TABLE V: Comparison of predictive results with other works that use the DGP dataset.

| Year and Reference | Predictive Approach | Source buildings | Grouping strategy | Target buildings | Output horizon | Best results (metrics) |
|---|---|---|---|---|---|---|
| 2019 [71] | RNN | 5/6 centroids (cosine distance) | K-means on the load profiles | 450/449 (80% train) | NA (we assume 1 step: 1h) | Average MAPE reduced from 30.1% to 22.8 % |
| 2020 [43] | LSTM | 80 % buildings | None | 20 % buildings | 24 steps: 1-24h | Average CVRMSE improvement between 5.62 and 9.22% |
| 2021 [44] | LSTM-DANN FC-DANN and CNN-DANN | 119 days of data from 1 office in America | None | 10 days of data from 2 offices in America | May 1–11 energy power prediction | LSTM-DANN with MAE: 5.3732 KW and MAPE: 6.35% |
| 2022 [55] | ANN | 13 offices | "Similar building" (manual selection) | 1-100 % data | 1 step: 1h | CVRMSE stabilises 17.5 % when 10% of data is available. Improvement: 8 % |
| 2022 [45] | LSTM | Simulated data building in Korea (6 months) | None/Timezones | 20 offices from DGP | 24 steps: 1-24h | CVRMSE improvement 5 - 18.4% |
| 2022 Our proposal | CNN-LSTM | 5/15 centroids (2000 datapoints 2.5 months) | k-prod and Image clustering | 512/502 | 24 steps: 1-24h | Average CVRMSE improvement of 25 % |

TABLE VI: Comparison of transfer learning approaches with other works that use the DGP dataset.

ONOFRE Project PID2020-112675RB-C44 funded by MCIN/AEI/10.13039/501100011033. It was also co-financed by the Spanish Ministry of Universities by means of the Margarita Salas linked to the European Union through the NextGenerationEU program and partially supported by the National Science Foundation under Grant No. 2150213 and the Air Force Office of Scientific Research SFFP Program.

## REFERENCES

[1] M. V. Moreno, F. Terroso-Sáenz, A. González-Vidal, M. Valdés-Vela, A. F. Skarmeta, M. A. Zamora, and V. Chang, "Applicability of big data techniques to smart cities deployments," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 2, pp. 800–809, 2016.

[2] H. Song, R. Srinivasan, T. Sookoor, and S. Jeschke, *Smart cities: foundations, principles, and applications*. John Wiley & Sons, 2017.

[3] M. Kochański, K. Korczak, and T. Skoczkowski, "Technology innovation system analysis of electricity smart metering in the european union," *Energies*, vol. 13, no. 4, p. 916, 2020.

[4] B. Insight, "Smart metering in north america and asia-pacific - 4th edition," 2021.

[5] Y. Sun, H. Song, A. J. Jara, and R. Bie, "Internet of things and big data analytics for smart and connected communities," *IEEE Access*, vol. 4, pp. 766–773, 2016.

[6] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *Journal of Big data*, vol. 3, no. 1, p. 9, 2016.

[7] S. Niu, Y. Liu, J. Wang, and H. Song, "A decade survey of transfer learning (2010–2020)," *IEEE Transactions on Artificial Intelligence*, vol. 1, no. 2, pp. 151–166, 2020.

[8] Q. Zhang, H. Li, Y. Zhang, and M. Li, "Instance transfer learning with multisource dynamic tradaboost," *The scientific world journal*, vol. 2014, 2014.

[9] Y. Yao and G. Doretto, "Boosting for transfer learning with multiple sources," in *2010 IEEE computer society conference on computer vision and pattern recognition*. IEEE, 2010, pp. 1855–1862.

[10] J. Hoffman, E. Rodner, J. Donahue, B. Kulis, and K. Saenko, "Asymmetric and category invariant feature transformations for domain adaptation," *International journal of computer vision*, vol. 109, no. 1-2, pp. 28–41, 2014.

[11] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Transactions on Neural Networks*, vol. 22, no. 2, pp. 199–210, 2010.

[12] J. Gao, W. Fan, J. Jiang, and J. Han, "Knowledge transfer via multiple model local structure mapping," in *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2008, pp. 283–291.

[13] R. Wu, Y. Yao, X. Han, R. Xie, Z. Liu, F. Lin, L. Lin, and M. Sun, "Open relation extraction: Relational knowledge transfer from supervised data to unsupervised data," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 2019, pp. 219–228.

[14] F. Li, S. J. Pan, O. Jin, Q. Yang, and X. Zhu, "Cross-domain co-extraction of sentiment and topic lexicons," in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2012, pp. 410–419.

[15] L. Shao, F. Zhu, and X. Li, "Transfer learning for visual categorization: A survey," *IEEE transactions on neural networks and learning systems*, vol. 26, no. 5, pp. 1019–1034, 2014.

[16] S. Ruder, M. E. Peters, S. Swayamdipta, and T. Wolf, "Transfer learning in natural language processing," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Tutorials*, 2019, pp. 15–18.

[17] M. M. Khan, R. Ibrahim, and I. Ghani, "Cross domain recommender systems: a systematic literature review," *ACM Computing Surveys (CSUR)*, vol. 50, no. 3, pp. 1–34, 2017.

[18] B. Guo, J. Li, V. W. Zheng, Z. Wang, and Z. Yu, "Citytransfer: Transferring inter-and intra-city knowledge for chain store site recommendation based on multi-source urban data," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 4, pp. 1–23, 2018.

[19] S. Zhao, Z. Xu, L. Liu, M. Guo, and J. Yun, "Towards accurate deceptive opinions detection based on word order-preserving cnn," *Mathematical Problems in Engineering*, vol. 2018, 2018.

[20] R. Wan, S. Mei, J. Wang, M. Liu, and F. Yang, "Multivariate temporal convolutional network: A deep neural networks approach for multivariate time series forecasting," *Electronics*, vol. 8, no. 8, p. 876, 2019.

[21] A. P. Ramallo-González, A. González-Vidal, and A. F. Skarmeta, "Ciotvid: Towards an open iot-platform for infective pandemic diseases such as covid-19," *Sensors*, vol. 21, no. 2, p. 484, 2021.

[22] A. Kelotra and P. Pandey, "Stock market prediction using optimized deep-convlstm model," *Big Data*, vol. 8, no. 1, pp. 5–24, 2020.

[23] D. Wang, Y. Yang, and S. Ning, "Deepstcl: A deep spatio-temporal convlstm for travel demand prediction," in *2018 international joint conference on neural networks (IJCNN)*. IEEE, 2018, pp. 1–8.

[24] W.-S. Hu, H.-C. Li, L. Pan, W. Li, R. Tao, and Q. Du, "Spatial–spectral feature extraction via deep convlstm neural networks for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 6, pp. 4237–4250, 2020.

[25] L. K. P. J. RDUSSEEUN and P. KAUFMAN, "Clustering by means of medoids," 1987.

[26] Z. Huang, "Extensions to the k-means algorithm for clustering large data sets with categorical values," *Data mining and knowledge discovery*, vol. 2, no. 3, pp. 283–304, 1998.

[27] J. T. Chi, E. C. Chi, and R. G. Baraniuk, "k-pod: A method for k-means clustering of missing data," *The American Statistician*, vol. 70, no. 1, pp. 91–99, 2016.

[28] H.-x. Zhao and F. Magoulès, "A review on the prediction of building energy consumption," *Renewable and Sustainable Energy Reviews*, vol. 16, no. 6, pp. 3586–3592, 2012.

[29] F. Amara, K. Agbossou, A. Cardenas, Y. Dubé, S. Kelouwani *et al.*, "Comparison and simulation of building thermal models for effective energy management," *Smart Grid and renewable energy*, vol. 6, no. 04, p. 95, 2015.

[30] R. Kramer, J. Van Schijndel, and H. Schellen, "Simplified thermal and hygric building models: A literature review," *Frontiers of architectural research*, vol. 1, no. 4, pp. 318–325, 2012.

[31] Y. Wei, X. Zhang, Y. Shi, L. Xia, S. Pan, J. Wu, M. Han, and X. Zhao, "A review of data-driven approaches for prediction and classification of building energy consumption," *Renewable and Sustainable Energy Reviews*, vol. 82, pp. 1027–1047, 2018.

[32] A. S. Ahmad, M. Y. Hassan, M. P. Abdullah, H. A. Rahman, F. Hussin, H. Abdullah, and R. Saidur, "A review on applications of ann and svm for building electrical energy consumption forecasting," *Renewable and Sustainable Energy Reviews*, vol. 33, pp. 102–109, 2014.

[33] B. Qolomany, A. Al-Fuqaha, A. Gupta, D. Benhaddou, S. Alwajidi, J. Qadir, and A. C. Fong, "Leveraging machine learning and big data for smart buildings: A comprehensive survey," *IEEE Access*, vol. 7, pp. 90 316–90 356, 2019.

[34] Y. Fu, Z. Li, H. Zhang, and P. Xu, "Using support vector machine to predict next day electricity load of public buildings with submetering devices," vol. 121, pp. 1016–1022, 2015.

[35] M. A. Jallal, A. Gonzalez-Vidal, A. F. Skarmeta, S. Chabaa, and A. Zeroual, "A hybrid neuro-fuzzy inference system-based algorithm for time series forecasting applied to energy consumption prediction," *Applied Energy*, vol. 268, p. 114977, 2020.

[36] A. Gonzalez-Vidal, F. Jimenez, and A. F. Gomez-Skarmeta, "A methodology for energy multivariate time series forecasting in smart buildings based on feature selection," *Energy and Buildings*, vol. 196, pp. 71–82, 2019.

[37] S. T. M. Bourobou and Y. Yoo, "User activity recognition in smart homes using pattern clustering applied to temporal ann algorithm," vol. 15, no. 5, pp. 11 953–11 971, 2015.

[38] S.-Y. Chou, A. Dewabharata, F. E. Zulvia, and M. Fadil, "Forecasting building energy consumption using ensemble empirical mode decomposition, wavelet transformation, and long short-term memory algorithms," *Energies*, vol. 15, no. 3, p. 1035, 2022.

[39] C. Fan, M. Chen, R. Tang, and J. Wang, "A novel deep generative modeling-based data augmentation strategy for improving short-term building energy predictions," in *Building Simulation*, vol. 15, no. 2. Springer, 2022, pp. 197–211.

[40] A. Li, F. Xiao, C. Fan, and M. Hu, "Development of an ann-based building energy model for information-poor buildings using transfer learning," in *Building Simulation*, vol. 14, no. 1. Springer, 2021, pp. 89–101.

[41] L. Wang, B. Guo, and Q. Yang, "Smart city development with urban transfer learning," *Computer*, vol. 51, no. 12, pp. 32–41, 2018.

[42] 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*, p. 100084, 2022.

[43] C. Fan, Y. Sun, F. Xiao, J. Ma, D. Lee, J. Wang, and Y. C. Tseng, "Statistical investigations of transfer learning-based methodology for short-term building energy predictions," *Applied Energy*, vol. 262, p. 114499, 2020.

[44] X. Fang, G. Gong, G. Li, L. Chun, W. Li, and P. Peng, "A hybrid deep transfer learning strategy for short term cross-building energy prediction," *Energy*, vol. 215, p. 119208, 2021.

[45] Y. Ahn and B. S. Kim, "Prediction of building power consumption using transfer learning-based reference building and simulation dataset," *Energy and Buildings*, vol. 258, p. 111717, 2022.

[46] Z. Huang, "Clustering large data sets with mixed numeric and categorical values," in *Proceedings of the 1st pacific-asia conference on knowledge discovery and data mining,(PAKDD)*. Singapore, 1997, pp. 21–34.

[47] M. K. Ng, M. J. Li, J. Z. Huang, and Z. He, "On the impact of dissimilarity measure in k-modes clustering algorithm," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 3, pp. 503–507, 2007.

[48] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.

[49] I. Goodfelow, Y. Bengio, and A. Courville, "Deep learning (adaptive computation and machine learning series)," 2016.

[50] M. Qiao, S. Yan, X. Tang, and C. Xu, "Deep convolutional and lstm recurrent neural networks for rolling bearing fault diagnosis under strong noises and variable loads," *IEEE Access*, vol. 8, pp. 66 257–66 269, 2020.

[51] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, "1d convolutional neural networks and applications: A survey," *Mechanical Systems and Signal Processing*, vol. 151, p. 107398, 2021.

[52] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-c. Woo, "Convolutional lstm network: A machine learning approach for precipitation nowcasting," *arXiv preprint arXiv:1506.04214*, 2015.

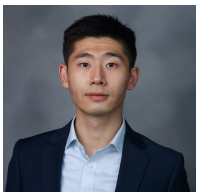
[53] E. M. Campos, P. F. Saura, A. González-Vidal, J. L. Hernández-Ramos, J. B. Bernabe, G. Baldini, and A. Skarmeta, "Evaluating federated learning for intrusion detection in internet of things: Review and challenges," *Computer Networks*, p. 108661, 2021.

[54] P. Ruzafa-Alcazar, P. Fernandez-Saura, E. Marmol-Campos, A. Gonzalez-Vidal, J. L. H. Ramos, J. Bernal, and A. F. Skarmeta,

"Intrusion detection based on privacy-preserving federated learning for the industrial iot," *IEEE Transactions on Industrial Informatics*, 2021.
[55] J. Li, C. Zhang, Y. Zhao, W. Qiu, Q. Chen, and X. Zhang, "Federated learning-based short-term building energy consumption prediction method for solving the data silos problem," in *Building Simulation*, vol. 15, no. 6. Springer, 2022, pp. 1145–1159.
[56] X. Chen, S. Dallmeier-Tiessen, R. Dasler, S. Feger, P. Fokianos, J. B. Gonzalez, H. Hirvonsalo, D. Kousidis, A. Lavasa, S. Mele *et al.*, "Open is not enough," *Nature Physics*, vol. 15, no. 2, pp. 113–119, 2019.
[57] E. Keogh and S. Kasetty, "On the need for time series data mining benchmarks: a survey and empirical demonstration," *Data Mining and knowledge discovery*, vol. 7, no. 4, pp. 349–371, 2003.
[58] C. Miller and F. Meggers, "The building data genome project: An open, public data set from non-residential building electrical meters," *Energy Procedia*, vol. 122, pp. 439–444, 2017.
[59] H. Rashid, P. Singh, and A. Singh, "I-blend, a campus-scale commercial and residential buildings electrical energy dataset," *Scientific data*, vol. 6, p. 190015, 2019.
[60] J. Langevin, P. L. Gurian, and J. Wen, "Tracking the human-building interaction: A longitudinal field study of occupant behavior in air-conditioned offices," *Journal of Environmental Psychology*, vol. 42, pp. 94–115, 2015.
[61] D. Murray, L. Stankovic, and V. Stankovic, "An electrical load measurements dataset of united kingdom households from a two-year longitudinal study," *Scientific data*, vol. 4, no. 1, pp. 1–12, 2017.
[62] M. A. Ahajjam, D. Bonilla Licea, C. Essayeh, M. Ghogho, and A. Kobbane, "Mored: A moroccan buildings' electricity consumption dataset," *Energies*, vol. 13, no. 24, p. 6737, 2020.
[63] A. Gulli and S. Pal, *Deep learning with Keras*. Packt Publishing Ltd, 2017.
[64] L. Hubert and P. Arabie, "Comparing partitions," *Journal of classification*, vol. 2, no. 1, pp. 193–218, 1985.
[65] G. W. Milligan and M. C. Cooper, "A study of the comparability of external criteria for hierarchical cluster analysis," *Multivariate behavioral research*, vol. 21, no. 4, pp. 441–458, 1986.
[66] A.-D. Pham, N.-T. Ngo, T. T. Ha Truong, N.-T. Huynh, and N.-S. Truong, "Predicting energy consumption in multiple buildings using machine learning for improving energy efficiency and sustainability," *Journal of Cleaner Production*, vol. 260, p. 121082, 2020.
[67] C. Fan, Y. Sun, F. Xiao, J. Ma, D. Lee, J. Wang, and Y. C. Tseng, "Statistical investigations of transfer learning-based methodology for short-term building energy predictions," *Applied Energy*, vol. 262, p. 114499, 2020.
[68] V. Papanikolaou, "Data mining for smart cities: Energy prediction for public buildings," 2021.
[69] Z. Dong, J. Liu, B. Liu, K. Li, and X. Li, "Hourly energy consumption prediction of an office building based on ensemble learning and energy consumption pattern classification," *Energy and Buildings*, vol. 241, p. 110929, 2021.
[70] I. Karijadi and S.-Y. Chou, "A hybrid rf-lstm based on ceemdan for improving the accuracy of building energy consumption prediction," *Energy and Buildings*, p. 111908, 2022.
[71] Y. Tian, "Cluster-based chained transfer learning for energy forecasting with big data," 2019.

**Aurora González-Vidal** graduated in Mathematics from the University of Murcia in 2014. In 2015 she got a fellowship to work in the Statistical Division of the Research Support Service, specializing in Statistics and Data Analysis. Afterward, she studied a Big Data Master. In 2019, she got a Ph.D. in Computer Science. Currently, she is a postdoctoral researcher at ITI-CERTH under the Margarita Salas program. She has collaborated in several national and European projects such as ENTROPY, IoTCrawler, PHOENIX and DEMETER. Her research covers machine learning in IoT-based environments, missing values imputation, and time-series segmentation. She is the president of the R Users Association UMUR.

**José Mendoza-Bernal** graduated in Official Program of Simultaneous Studies in Mathematics and Computer Science from the University of Murcia (Spain) in 2020. He qualified to the Southwestern Europe Regional Contest of Programming (SWERC) in 2016. He studied a Big Data Master and he is as research assistant at the same university since 2019. He has collaborated in national and European projects. His research covers machine learning in IoT-based environments, missing values imputation, and virtual reality application development.



**Shuteng Niu** received the Ph.D. degree in Electrical Engineering and Computer Science from Embry-Riddle Aeronautical University, Daytona Beach, Florida, in May 2021. He was a postdoctoral research fellow at UTHealth School of Biomedical Informatics. Now, he is an assistant professor of computer science at Bowling Green State University. His major research interests include machine learning, computer vision, and biomedical informatics.

**Antonio F. Skarmeta** received the M.S. degree in Computer Science from the University of Granada and B.S. (Hons.) and the Ph.D. degrees in Computer Science from the University of Murcia Spain. Since 2009 he is Full Professor at the same department and University. Antonio F. Skarmeta has worked on national and international research projects in networking, security and IoT area, like SEMIRAMIS, SMARTIE, SOCIOTAL, IoT6 ANASTACIA, CyberSec4Europe. His main interested is in the integration of security services, identity, IoT and Smart Cities. He has been heading of the research group ANTS since its creation on 1995. He has published over 200 international papers.

**Houbing Song** (M'12–SM'14) received the Ph.D. degree in electrical engineering from the University of Virginia, Charlottesville, VA, in 2012. In 2017, he joined the Department of Electrical Engineering and Computer Science, Embry-Riddle Aeronautical University, Daytona Beach, FL, where he is currently a Tenured Associate Professor and the Director of the SONG Lab. He has received the Best Paper Award in 5 conferences and he is the author of more than 100 articles and the inventor of 1 patent issued by WIPO. His research interests include cyber-physical systems, cybersecurity and privacy, internet of things, edge computing, AI/machine learning, big data analytics, unmanned aircraft systems, connected vehicle, smart health, and wireless communications and networking.