Building energy retrofit index for policy making and decision support at regional and national scales

Fazel Khayatian*, Luca Sarto, Giuliano Dall'O'

Department of Architecture, Built Environment and Construction Engineering, Polytechnic of Milan, Milan, Italy

The vast data collected since the enforcement of building energy labelling in Italy has provided valuable in-formation that is useful for planning the future of building energy efficiency. However, the indicators provided through energy certificates are not suitable to support decisions, which target building energy retrofit in a regional scale. Considering the bias of the energy performance index toward a building's shape, decisions based on this index will favor buildings with a specific geometric characteristics. This study tends to overcome this issue by introducing a new indicator, tailored to rank buildings based on retrofitable characteristics. The pro-posed framework is validated by a case study, in which a large dataset of office buildings are assigned with the new index. Results indicate that the proposed indicator succeeds to extract a single index, which is representative of all building characteristics subject to energy retrofit. A new labeling procedure is also compared with the conventional classification of buildings. It is observed that the proposed labels properly partitions the dataset, according to buildings' potential to undergo energy retrofit.

1. Introduction

As part of the climate change counterfeiting objectives of the European Union, the Energy Performance for Buildings Directive (EPBD recast) mandates EU Member States to promote building energy efficiency [1]. Even though building energy efficiency was already enforced in some EU Member States [2]; the implementation of the EPBD resulted in national and regional legislations that persuade public and private bodies to undergo building energy audit, often by means of an Energy Performance Certificate (EPC) [3]. The outcomes of implementing EPCs has been widely investigated from various perspectives: reliability and credibility of the certificates [4,5], Socio-economic impacts [6], energy retrofit scenarios [7,8], renting and trading properties [9–11], social participation [12,13], and the impacts of EPCs on other policies that promote building energy efficiency [14].

It has been argued that alongside other functionalities, EPCs should also serve decision makers and energy planners at regional and national scales [15]. As a result, studies have approached EPC outcomes from an energy planning perspective [16–20], and underlined the necessity of better energy efficiency policies for the building sector [21,22]. Meanwhile, it has been reasoned that relying on the wrong indicators may lead to misconceptions on the actual status of the building stock [23] and result in suboptimal policies [24]. The literature outlines this issue by indicating that: the geometry of the building may have a too strong effect on the estimated energy need [25], the relation between the building geometry and the estimated energy consumption is non-linear [26], and energy efficient renovation may render unrealistic as some pivotal information could be missing from the indicators [27]. For instance, a statistical analysis on the covariation between building properties and the energy use intensity has revealed that the thermal characteristics of walls can have a strong influence on the energy use intensity, while the correlation between window characteristics and the

---

* Corresponding author.
E-mail address: fazel.khayatian@polimi.it (F. Khayatian).

Nomenclature

Abbreviations

EPBD	Energy Performance of Buildings Directive
EPC	Energy Performance Certificate
CENED	Certificazione Energetica degli Edifici
KL	Kullback-Leibler divergence
MAPE	Mean Absolute Percentage of Error
MLP	Multi-Layer Perceptron
MSE	Mean Squared Error
MSEsparse	Mean Squared Error of sparsity
PCA	Principal Component Analysis
R^2	Squared correlation coefficient

S/V	Surface to Volume ratio
SoD	Sum of Distances
SSE	Sum Squared Error
tMLP	truncated Multi-Layer Perceptron

Variables

Eff_G	global efficiency of system [–]
EP_i	energy performance index [kW h/m^2 y]
ERi	energy retrofit index [–]
U_b	average U-value of basement [W/m^2 K]
U_e	average U-value of walls [W/m^2 K]
U_r	average U-value of roof [W/m^2 K]
U_w	average U-value of windows [W/m^2 K]

energy use intensity may be difficult to comprehend [28].

Studies have previously demonstrated the positive effects of optimal building energy retrofit on energy savings at municipal scales [29]. However, building energy retrofit planning at regional scales may face challenges due to the misconception of EPCs as the indicators often present the performance of the building by means of energy use intensity [30]. Consequently, energy performance indicators may be biased toward geometric characteristics of buildings [31,32], and fail to render the actual potential of a building to undergo an energy retrofit. This issue is particularly important in the energy retrofit planning process, during which decision makers have to prioritize a number of buildings that merit financial assistance while handling subsets of multidimensional data [33]. The challenge of a ranking system dedicated to retrofit potential is worth deepening since grants, tax deductions, loans and similar financial promotions have proved to be effective strategies for encouraging the public toward energy efficiency [34–36]. Therefore, it is necessary to support energy planning with a reliable indicator that does not merely present the energy consumption of the building, but also maps the characteristics of one building compared to others. Moreover, a new classification of buildings' fea-tures is inevitable since the conventional energy labels (classes of en-ergy) do not necessarily reflect the thermal characteristics of a build-ing's envelope. To counter the bias of energy use intensity towards a unit's geometry, studies have proposed the application of energy benchmarking [37]. Sun and Price introduced a classification strategy by resorting to prototype building characteristics as the basis of retrofit analysis [38]. This approach succeeds to rank buildings based on their retrofitbale characteristics and overcomes the bias towards geometry. However, a predefined database of reference buildings (similar to that of the USEDOE [39]) is an essential part of the framework, and may not be available for all EPC databases.

Clustering is a suitable alternative to the traditional frequency-based classification of buildings as it can return more robust subsets [40,41]. Building energy rating has been tackled by opting for various clustering techniques, namely, decision tree [42], fuzzy [43], k-means [44], as well as Gaussian, hierarchical and self-organizing maps [45]. While much progress has been achieved in obtaining better partitions between the subset clusters, there has been no attempts to tailor Ma-chine-Learning-based ranking techniques with building energy retrofit in mind. Since an EPC has a high correlation with a unit's floor area [46], it cannot explain whether a low energy performance is related to poor envelope characteristics, or due to high surface to volume (S/V) ratio. Such level of information is critical for decision making in a re-gional scale: according to the EU policies aimed at buildings' energy consumption, the priority in energy efficiency is to initially prevent buildings from excessive energy use, and then promote the application of renewable energies [47]. Therefore, a building with suitable en-velope characteristics and a high S/V ratio should have higher priority to receive financial aid for installing renewable systems, compared to a

competitor who has worse envelope characteristics but stands higher in the energy efficiency rankings due to a lower S/V ratio. This issue is significantly important since there are other parameters similar to S/V which affect a building's energy consumption, but cannot be subject to retrofit. This paper tends to overcome the described challenge by introducing a novel framework for ranking buildings. This ranking system is based on a new indicator that is specifically designed to target retrofitable building characteristics. Therefore, decision makers can easily distinguish buildings which have higher merit to undergo an energy retrofit. The proposed indicators can assist administrations in aiming policies at specific subsets of buildings. This is essential for assessing hypothetical funding policies, as well as creating trajectories of possible updates in the EPC database [48]. The original scientific contributions of this study include:

- Contrasting the ineffectiveness of the "Energy Performance index" for ranking buildings according to retrofitable properties.
- Presenting a robust and replicable Machine-Learning-based pipeline for extracting weighted nonlinear features from building characteristics.
- Introducing the "Energy Retrofit index"; a new measure specifically tailored to support the allocation of financial aids for boosting building energy retrofit at regional scales.

The rest of the paper is structured in the following manner: Section 2 provides an overview on the machine learning techniques applied in the study i.e. Multi-Layer Perceptrons, nonlinear Principal Component Analysis and k-means clustering. Section 3 applies the framework to a dataset of building energy certificates extracted from the Lombardy region in Italy, and compares the results of the proposed method with available indexes and conventional clustering techniques. Section 4 concludes the paper by summarizing the advantages of the proposed method when compared to its predecessors, as well as applications and possible expansions for future studies.

2. Methodology

The proposed framework resorts to three tools, which are adopted from the field of machine learning i.e. neural networks, autoencoders and k-means clustering. Based on the type of learning, the exploited tools can be divided into two main categories of supervised (neural networks) and unsupervised learning (autoencoders, k-means). The supervised learning of the framework tends to find covariations between some input data and the targets. Unsupervised learning on the other hand, seek structures and covariations in the input data itself. To ensure that the study provides a clear description of the framework, initially, each tool is briefly described.

2.1. Multi-layer perceptron (neural network)

A multi-layer perceptron is a feed forward neural network that maps approximations of target data by finding nonlinear dependencies among the inputs [49]. Neural networks are mostly composed of 3 layers, i.e. input layer, hidden layer, output layer, where each layer consist of one or more nodes. In this study, a simple back-propagated (vanilla) neural network refers to a three layer perceptron (Fig. 1) with two transformation layers, i.e. a hidden layer (consisting of nonlinear transfer functions) and an output layer (with a linear transfer function) [50]. The cost function of a vanilla neural network can be presented as:

$$J(W,b) = \frac{1}{m} \sum_{i=1}^{m} \left(\frac{1}{2} \| h_{W,b}(x^{(i)}) - y^{(i)} \|^2 \right) + (\lambda . \Omega_w) \tag{1}$$

where $J(W,b)$ is the prediction error, m is the number of samples, $h_{W,b}(x)$ is the hypothetical weight and bias values, y is the target of the network, λ is the is the weight decay coefficient and Ω_w is the weight regularization term. The weight regularization term can be defined as:

$$\Omega_w = \frac{1}{2} \sum_{l}^{n_l-1} \sum_{i}^{s_l} \sum_{j}^{s_l+1} (w_{j,i}^{(l)})^2 \tag{2}$$

where n_l is the number of layers, s_l is the number of units of layer l and $w^{(l)}$ is the weight of the hidden unit at layer l. In this study we resort to the Levenberg-Marquardt algorithm for training the neural network. Since the algorithm uses the Jacobian matrix instead of the more computationally intensive Hessian matrix for calculating the MSE, it is much faster than most algorithms and comparably powerful for training vanilla neural networks [51]. The applicability of neural networks in estimating primary energy consumption of buildings for retrofit purposes has been previously validated in [52].

2.2. Nonlinear principal component analysis (autoencoder)

Principal component analysis (PCA) is a statistical technique to describe correlated variation in a dataset through linear uncorrelated variables (i.e. principal components). The procedure is mainly used for dimensionality reduction or visualization purposes. PCA is particularly useful for analyzing high dimensional data, and therefore, has been adopted to cluster buildings based on their envelope energy efficiency [53]. Meanwhile, exploring only linear components can affect our understanding of the whole dataset, since it has been argued that the correlation between a building's geometry and the calculated energy consumption is nonlinear [26]. Therefore, in this study we resort to nonlinear PCA by means of unsupervised learning, namely, auto associative neural networks (autoencoders). The autoencoder is multi-layer perceptron, were the input and target data are identical [54]. The objective of an autoencoder is to identically recreate the input data by passing it through an encoder and reconstructing it by a decoder. Although the process of reconstructing the input data seems pointless, it is observed that applying constraints on the network will force the neurons' weights to return a different representation of the original data [55]. Such approach has proved to outperform linear PCA in a wide range of AI applications [56].

An autoencoder is composed of two layers, namely, an encoder and a decoder (Fig. 2). The two layers may consist of similar or various number of neurons, based on the intended application. Similar to neural networks, autoencoders can benefit from backpropagation of error to fine-tune their weights. The iterative process of training an autoencoder can be terminated by defining one or more constraints on the training algorithm (i.e. maximum number of iterations, maximum training time, minimum performance and minimum gradient). The cost function of training an autoencoder can be defined based on the MSE and a regularization term as described in Eq. (1). However, the loss function of an autoencoder (MSEsparse) is based on the difference between the inputs and the reconstructed inputs. Also, an additional regularization

term is incorporated in the cost function. The cost function of the MSEsparse can be presented as:

$$J(W,b) = \frac{1}{m} \sum_{i=1}^{m} \left(\frac{1}{2} \| h_{W,b}(x^{(i)}) - \hat{x}^{(i)} \|^2 \right) + (\lambda . \Omega_w) + (\beta . \Omega_s) \tag{3}$$

where x is the input and x is its reconstructed output, λ is the weight decay coefficient of the second layer (L_2) regularization, Ω_w is the L_2 regularization term, β is the penalty coefficient of sparsity regularization, and Ω_s is the sparsity regularization term. The weight decay regularization term is similar to the neural network regularizer as provided in Eq. (2). The sparsity regularization term Ω_s, is defined as:

$$\Omega_s = \sum_{j=1}^{S2} KL(\rho \| \hat{\rho}_j) = \sum_{j=1}^{S2} \rho \log\left(\frac{\rho}{\hat{\rho}_j} \right) + (1-p) \log\left(\frac{1-\rho}{1-\hat{\rho}_j} \right) \tag{4}$$

where KL is the Kullback-Leibler divergence function that measures the similarity of two distributions, while ρ and ρ̂ represent the desired and actual average activation value of a hidden unit respectively. The average output activation value of unit j can be presented as:

$$\hat{\rho}_j = \frac{1}{m} \sum_{i=1}^{m} a_j^{(2)}(x^{(i)}) \tag{5}$$

where x is a training sample and $a_j^{(2)}$ is the activation of the hidden unit. In this study we opt the scaled conjugate gradient training algorithm, as it converges faster when compared to steepest gradient decent methods [57].

It is possible to stack nonlinear PCAs and train a multi-layer autoencoder commonly referred to as deep autoencoder. Deep autoencoders are usually trained through a greedy layer-wise training algorithm, were shallow autoencoders are trained one at a time and the outputs of one encoder is fed to the next autoencoder as inputs [58]. Stacking autoencoders to form a deep network is useful for extracting representations of noisy data, eventually resulting in a more robust set of features [59].

2.3. K-means clustering

K-means algorithm is a centroid based clustering technique, which can partition data into coherent subsets. The algorithm returns an index for each sample of the data, where the number of indexes (k) is predefined before executing the algorithm. Since the k-means algorithm is NP-hard, it is possible to approximate the algorithm by using a heuristic. This careful seeding commonly known as k-means + + results in faster convergence to smaller intra-cluster SSE, when compared to standard k-means [60]. The cost function of k-means algorithm can be
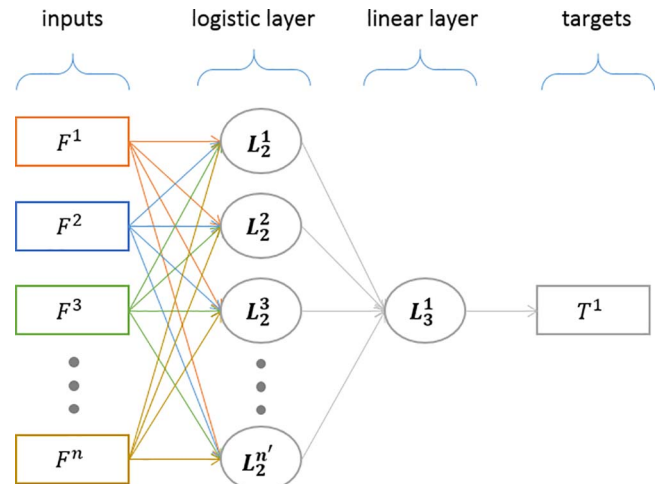

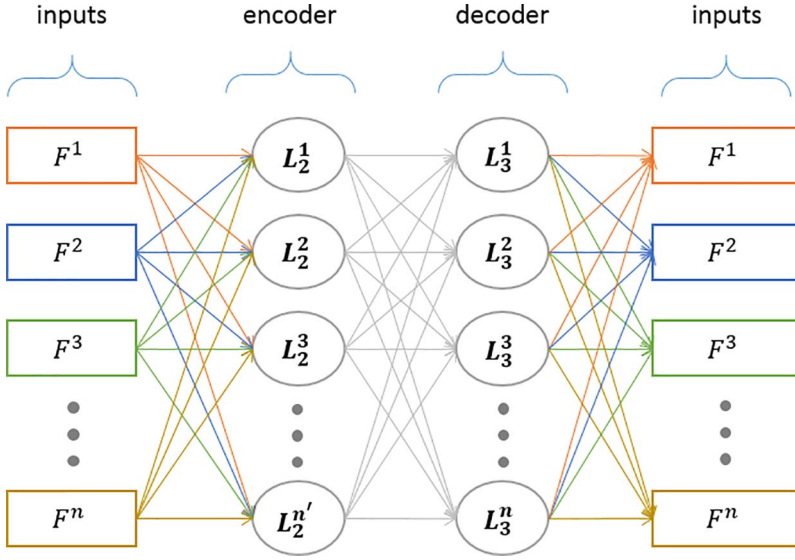**Fig. 1.** Illustration of a vanilla neural network with n input features, n′ hidden units, one linear unit and one target.

defined based on the squared Euclidean distance, presented as:

$$J(c,\mu) = \sum_{i=1}^{m} \|x^{(i)} - \mu_{c^{(i)}}\|^2 \tag{6}$$

where $J(c,\mu)$ is the sum squared distance between sample i denoted as $x^{(i)}$, and the specific centroid $\mu_{c^{(i)}}$ that is assigned to that sample. Choosing the correct number of clusters may be difficult and time consuming, since the goodness of a cluster may be defined by various measures. In this study we evaluate the number of clusters by resorting to 4 measures, i.e. Calinski-Harabasz, Davies-Bouldin, Silhouette and Gap.

The Calinski-Harabasz criterion is a function of the overall inter-cluster variance and the overall intra-cluster variance such that [61]:

$$CH = \frac{\sum_{i=1}^{k} n_i \|m_i - m\|^2}{\sum_{i=1}^{k} \sum_{x \in c_i} \|x - m_i\|^2} \cdot \frac{(N-k)}{(k-1)} \tag{7}$$

where k is the number of clusters, N is the number of samples, n_i is the number of samples in the cluster, m_i is the cluster centroid, m is the average of the cluster, x is a sample in the cluster, and c_i is the ith

cluster. In this measure, larger values correspond to better partitions. The Davies-Bouldin criterion is also calculated by means of the inter-cluster and the intra-cluster distances [62]:
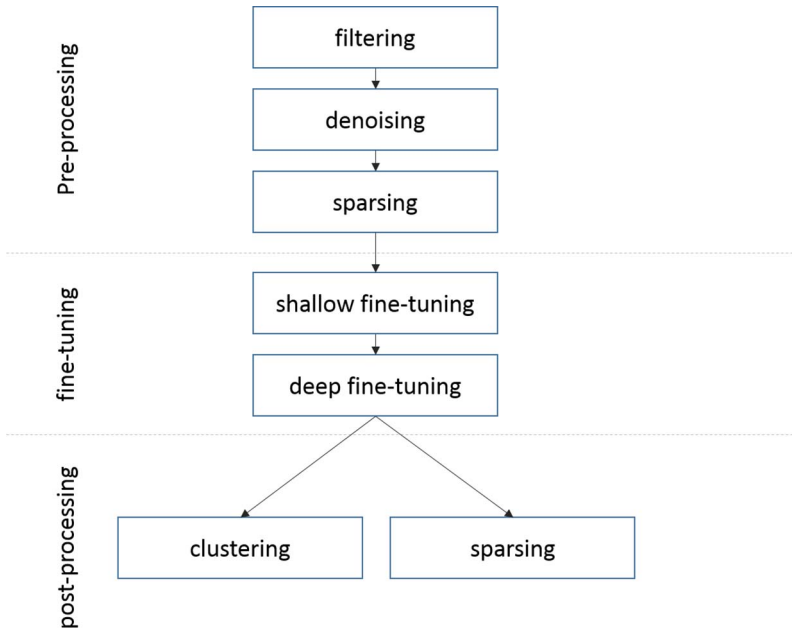
$$DB = \frac{1}{n} \sum_{i=1}^{n} max_{j \neq i} \left(\frac{\sigma_i + \sigma_j}{\|c_i - c_j\|} \right) \tag{8}$$

where n is the number of clusters, c is the cluster centroid, and σ is the average distance between all samples in a cluster and the cluster centroid. The Davies-Bouldin measure returns smaller values for better partitioning. The Silhouette criterion measures the similarity of samples inside a cluster. For each sample the similarity value is defined as [63]:

$$S = \frac{(b_i - a_i)}{max(b_i - a_i)} \tag{9}$$

where a_i is the average distances between sample i and other samples in the cluster, b_i is the average of the smallest distances between sample i and samples in other clusters. In this criterion, high values correspond to better partitioning. The Gap statistic defines the optimal number of clusters where the gap value is maximized. The gap value is defined as [64]:

$$GAP = E_n^*\{log(W_k)\} - log(W_k) \tag{10}$$

where n is the sample size, k is the number of clusters, and E_n^* is expectation for a vector of size n sampled from a reference distribution, and W_k is the pooled within cluster dispersion measure, defined as:

$$W_k = \sum_{r=1}^{k} \frac{1}{2n_r} \sum_{i,i' \in C_r} d_{i,i'} \tag{11}$$

where C_r denotes cluster r, n_r is the number of samples in cluster r, and $d_{i,i'}$ is the squared Euclidean distance between samples i and i'.

2.4. Model description

The following model consists of 3 main stages i.e. preprocessing, fine-tuning and post-processing. Fig. 3 illustrates the outline of the model, as well as the sequence of activities. The first stage is composed of 3 steps, namely, filtering, denoising and sparsing the data. The second stage has two steps i.e. shallow fine-tuning and deep fine-tuning. The final stage is responsible for extracting the desired output from the model (labels or indexes) through clustering or sparsing.

2.4.1. Data preprocessing

The first step of the preprocessing stage is to clean the data from obvious defections. A thorough study on suitable filters for cleaning similar datasets has been performed in [17]. After clearing the data from defections, the dataset is divided into two sets i.e. inputs and targets. The input dataset is a matrix composed of building characteristics that may be subject to retrofit, let it be properties of the envelope, systems or both. The target dataset is single array, containing the energy performance indicator of each building.

Since autoencoders are trained through unsupervised learning, the network only consists of inputs without any targets. Therefore, building characteristics (inputs) are fed into an autoencoder for denoising the data and extracting nonlinear principal components from the dataset (Table 1). The autoencoder is trained with the input data (m buildings and n features) and the desired number of hidden units (n' nodes) in the L_2 layer. Generally, it is a good idea to increase the dimensionality of the data before sparing the number of hidden units (i.e. $n' = n + \varepsilon$). Such approach will force the autoencoder to learn over-complete hidden representations of the input data, eventually resulting in better feature extraction [55]. After the training is terminated based on the stopping criteria (e.g. maximum iterations, maximum time, minimum performance, etc.), inputs are fed into the encoding section of autoencoder, returning m buildings each consisting of n' features. These new feature are representations of the building characteristics which are mapped onto other dimensions, and do not have a unit.

The denoised dataset extracted from the first encoder ($m*n'$) is fed into the second autoencoder to reduce the dimensionality of the data to n'', where the size of the hidden units are equal or smaller than the size of the original inputs (i.e. $n \geqslant n''$). After the second autoencoder is trained, sparsed features are extracted from the second encoder, resulting in a dataset of $m*n''$ dimensions.

2.4.2. Fine-tuning

The sparsed data derived from the second autoencoder are used as inputs for a vanilla neural network consisting of a hidden layer with n''' units ($n''' \geqslant n''$) and a linear layer with a single node (shallow fine-tuning). The target of this network is the EPi value, which represents building energy consumption by means of kWh/m^2.y. Commonly, the training process of the neural network is terminated by exceeding the number of validation checks, indicating that the network is starting to overfit onto the training data. At this stage, both encoders and the neural network are stacked on top of each other, creating a four layer perceptron (Fig. 4). This process enables the errors to be back-propagated all the way to the hidden units of the 1st encoder, eventually fine-tuning the weights of the whole MLP altogether (deep fine-

tuning). During this process, the outputs from the second encoder are weighted based on the magnitude of correlation they have with the EPi values (targets). After the training is complete, the linear layer (L_5) of the MLP is removed from the network, connecting the output node directly to the last hidden layer (L_4). As a result, the network will produce a vector of n''' features for every building ($m*n'''$ matrix).

2.4.3. Data post-processing

Since the output of the MLP is a n'''-dimensional vector, it cannot be easily perceived. However, it is possible to either extract an explicit index from the multi-dimensional dataset, or label the samples by clustering them into a number of subsets.

To derive a processable index from the outputs, we resort to an autoencoder, by sparsing the n''' features into one hidden unit at the L_2 layer (Fig. 2). After training is complete, the encoder produces a single value for each sample, namely the ERi index.

To label the samples by clustering, the n'''-dimensional dataset extracted from the MLP is fed into a k-means clustering tool. The optimum number of subsets is obtained by evaluating the performance of four clustering algorithms described in Section 2.3. Since different evalua-tion measures of the algorithms consist of dissimilar scales, the mea-sures are mean and range normalized. It is important to note that contrary to the other three criterions, smaller values in the Davies-Bouldin criterion correspond to better clusters. Therefore, the evalua-tions from the Davies-Bouldin criterion are inverted before processing. The output of the clustering dubbed the energy retrofit label assigns each building with a group using Latin numerals.

3. Numerical case study

3.1. CENED database

Studies have classified building energy benchmarking procedures based on the calculation engine (i.e. whitebox, greybox or blackbox)[65] as well as the data collection methodology (i.e. simulation, mea-sured, hybrid) [66]. In this section, we numerically evaluate the pro-posed blackbox framework by means of simulated EPCs that are col-lected by the Lombardy Region in northern Italy [67,68]. The energy certificates are issued through the Certificazione Energetica degli Edi-fici (CENED) tool, using a quasi-steady-state approximation of the en-ergy consumption [69,70]. Certificates are issued upon the request of a property owner, whereas the energy audit can be only performed by a certified energy assessor. Effective since 2006, CENED was the first implementation of the EPBD into the Italian context while containing over a million certificates by 2016. An energy label (ranging from A+ to G) is assigned to each property based on the primary energy index (EPi). Since heating energy need is dominant in the Lombardy Region, the EPi value is set to approximate the annual heating energy con-sumption. A recent update on the CENED software does provide a more complete approximation of the energy consumption by including cooling, lighting and domestic hot water energy consumption in the evaluate EPi value. However, more than 90% of the issued energy certificates are according to the old calculation procedure and will re-main valid till 2020. Therefore, it is necessary to extract as much in-formation as possible from the existing database, specifically,

Table 1
Building characteristics subject to retrofit and the filters applied for removing defected samples.

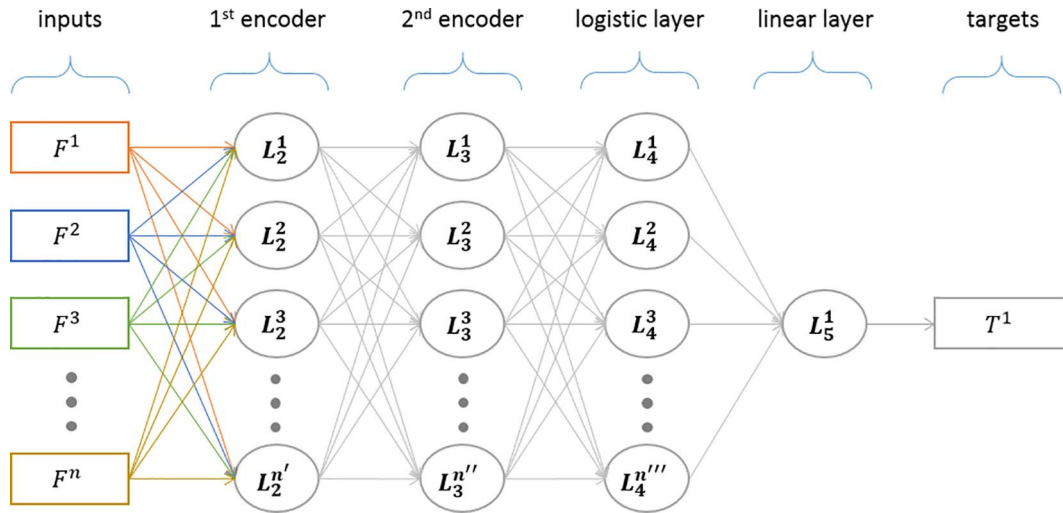| Variable	Property	Unit	Filter
U_e	Thermal conductivity	W/m^2 K	0.1 < value < 5.0
U_r		W/m^2 K	0.1 < value < 4.0
U_b		W/m^2 K	0.1 < value < 4.0
U_w		W/m^2 K	0.3 < value < 6.0
Eff_G	efficiency	–	0.5 < value < 0.95

Fig. 4. Illustration of a deep MLP by stacking autoencoders and a vanilla neural network. The network is composed of n input features, n' units for denoising, n'' units for sparsing, n''' logistic units, one linear unit and one target.

assumptions which can support decision making in a regional scale. The EPi is calculated based on building characteristic and climatic data, where building characteristics include: destination of use, geometry, dispersant surfaces and their properties, as well as the age and thermal mass of the structure. Similar to other energy certification tools, CENED's estimation of the EPi is highly correlated to geometric characteristics of the building. Although this attitude is realistic when considering the annual heating consumption, it does not provide much information from an energy retrofit point of view. The energy label assigned to each building also suffers from the same issue, where obtaining a good energy class does not necessarily imply high performance envelope properties, and vice versa. The current study will focus on the energy certificates issued for office buildings in the Lombardy Region. It is important to note that the case study will only focus on certificates issued for whole buildings, as single office units are less likely to undergo a thorough energy retrofit.

3.2. Implementing the framework

The dataset extracted from the Lombardy Region database is composed of several building characteristics (e.g. period of construction, opaque area, glazed surface, occupied floor area and etc). However, only a handful of the building properties may be subject to an energy retrofit (Table 1). In this study we consider the thermal conductivity of the envelope and the efficiency of the heating system as modifiable elements. Therefore, the input dataset is comprised of 5 features, i.e. U-value of walls, U-value of windows, U-value of roof, U-value of basement, and the global efficiency of the heating system.

Before feeding the data into the model, a series of filters are applied to the dataset (Table 1). This process ensures a more reliable and polished input data consisting of 4767 buildings. Moreover, the input dataset is mean and range normalized before implementing into the model. Such preprocessing will result in more refined cost functions, and therefore a smoother gradient decent. All calculations including supervised and unsupervised trainings were carried out within the MATLAB® environment version R2016b. The sequence of steps for implementing the framework is described through an enumerated list (Fig. 5), and the properties of each step's tool is provided in Table 2. The process of optimizing the hyper-parameters of the networks (either autoencoders or neural networks) was carried out through a trial and

```
X=inputs';                          % inputs have 4767 rows (samples) & 5 columns (features)
T=targets';                         % targets have 4767 rows (samples) & 1 column
AE1 = trainAutoencoder(X,8, ...);   % 1) train autoencoder for denoising X
X1 = encode(AE1,X);                 % 2) extract 8 new features (X1)
AE1 = trainAutoencoder(X1,5, ...);  % 3) train autoencoder for sparsing X1
X2 = encode(AE1,X1);                % 4) extract 5 new features (X2)
net = feedforwardnet(10);           % 5) set the vanilla neural network with 10 hidden units
NN = trainlm(net,X2,T);             % train the network with X2 as input and T as target
MLP = stack(AE1,AE2,NN);            % 6) connect encoder1 and encoder2 and neural network
MLP = train(MLP,X,T);               % train a MLP with X as input and T as Target
tMLP=MLP;                           % 7) create a copy of the MLP (tMLP)
tMLP.layerConnect(5,4) = 0;         % disconnect the linear layer (L5) from the logistic layer (L4)
tMLP.outputConnect(1,4) = 1;        % connect the logistic (L4) layer to the output node
tMLP.outputConnect(1,5) = 0;        % disconnect the linear layer (L5) from output node
Y=tMLP(X);                          % extract the 10 weighted features (Y)
AE3 = trainAutoencoder(Y,1, ...);   % 8) train autoencoder for sparsing Y to as single dimension
ERi=encode(AE3,Y); ERi=ERi';        % 9) extract the representative feature (ERi) from Y
Y=Y'; Label = kmeans(Y,9, ...);     % 10) cluster Y into 9 subsets
```

Fig. 5. MATLAB® script utilized in the case study (VERSION R2016B). Detailed properties of the steps are provided in Table 2.

Table 2
Models and their inputs, targets, and outputs at each step of the pipeline.

	Input	Target	Output	Task
Autoencoder I	Building characteristics	–	–	Unsupervised training
Encoder I	Building characteristics		Denoised features	Encoding
Autoencoder II	Denoised features	–	–	Unsupervised training
Encoder II	Denoised features		Sparsed features	Encoding
vanilla Neural Network	Sparsed features	Energy Performance index	–	Supervised training
Multi-Layer Perceptron	Building characteristics	Energy Performance index	–	Supervised training
truncated Multi-Layer Perceptron	Building characteristics	–	Weighted features	Encoding
Autoencoder III	Weighted features	–	–	Unsupervised training
Encoder III	weighted features	–	Energy Retrofit index	Encoding
k-means	Weighted features	–	Energy Retrofit label	Unsupervised training

error exhaustive search.

In the preprocessing step, input data is fed into two autoencoders for denoising and sparsing purposes respectively. Both autoencoders use the same cost function (Eq. (3)) for calculating the reconstruction error, yet consist of different hidden units according to their objective. Recall that the number of hidden units (denoted $L_2^{n'}$ in Fig. 2) defines the dimensionality of the encoder, and therefore the size of the output. The denoising autoencoder is assigned with more hidden units ($L_2^{n'}$) than the number of input features (F^n), so that the network can learn over-complete representations of the input data. On the other hand, the sparse autoencoder has an inverse functionality, namely, to reduce the dimensionality of the data by using less hidden units. Consequently, the hidden layer of the sparse autoencoder works as a bottleneck that squashes the features learnt by the denoising autoencoder, into a smaller dimension. Although both autoencoders generally display better performances with more hidden units, their performance merely improves after the number of hidden units are saturated. The performance of the first (denoising) autoencoder with 5 inputs flattens after 8 hidden units with a reconstruction MSE of 7.15 e−04. The second (sparse) autoencoder with 8 inputs reaches performance saturation at 5 hidden units while returning a reconstruction MSE of 7.66 e−06. The outcomes of the preprocessing stage are two encoders (denoiser and sparser) that are responsible for changing the dimensionality of the input features from 5 to 8, and then from 8 to 5, respectively. In other words, the objective of unsupervised learning through autoencoders is to extract hidden representations from the input features while minimizing the reconstruction error, implying that the learnt representations contain nearly as much information as the actual building characteristics, yet are presented in a more polished and refined manner. At this stage, the refined features are ready to be fed into the perceptron for weight assignment and fine-tuning.

The weight assignment stage is composed of two steps, i.e. shallow fine-tuning with a single hidden layer perceptron, and deep fine-tuning with a MLP. Shallow fine-tuning is performed by means of a vanilla neural network (Fig. 1), where inputs (F^n) are the refined features encoded from the preprocessing stage through autoencoders. The cost function of the vanilla neural network is to reduce the error between the targets and the hypothetical estimations (Eq. (1)) by assigning weights to the hidden units ($L_2^{n'}$). The target of this network (T) is the EPi value obtained from the energy certificate tool. Since overfitting is observed with excessive hidden units, the number of neurons is set to 10 through a trial and error exhaustive search. The process of training the network is repeated 50 times to ensure optimal initialization of weights. The shallow fine-tuning process reaches its peak performance at a MSE of 477.32 ($R^2 = 0.38$).

At this stage, three pre-trained models are available: (1) denoising autoencoder, (2) sparse autoencoder and (3) single layer neural network; nevertheless, each model has been trained separately. Stacking the three models will create a multi-layer network where each layer is responsible for a specific processing task, and the error is back-propagated continuously through all layers (Fig. 4). Therefore, training

the stacked MLP fine-tunes the three pre-trained models altogether (deep fine-tuning), eventually reducing the overall error of the model. The MLP is in fact a network composed of the previously pretrianed layers, i.e. the denoising autoencoder with a hidden size ($L_2^{n'}$) of 8, the sparse autoencoder with a hidden size $L_3^{n''}$ of 5, and the vanilla neural network with a hidden size ($L_4^{n'''}$) of 10. Similar to the vanilla neural network, the cost function of the MLP is to reduce the error between targets and the hypothetical estimation (Eq. (1)). Since the MLP is a pipeline of all pretrained models, the input of the MLP ($F^1, F^2, ..., F^5$) is the exact same input fed into the denoising autoencoder (building characteristics), while the target (T) is the same target assigned to the vanilla neural network (EPi). The MLP displays similar performances in all repetitions of the training process, with the best trial returning a MSE of 414.08. The reported error (MAPE ~ 23) may not be appealing for prediction purposes. However, it is important to note that the objective of the pipeline is not training the perfect predictor, but rather, maximizing the models capability to explore nonlinear correlations between inputs and the target. The latter objective is in fact fulfilled as a reasonable correlation coefficient ($R^2 = 0.42$) between outputs and targets is obtained.

It is also important to ensure that the trained MLP is well generalized and not overfitted onto the training dataset. Therefore, the obtained MSE from the training set and the cross-validation set was continuously monitored, using early stopping to prevent the MLP from overfitting. The training process was terminated when the error of the cross validation set increased for 10 consecutive iterations. Perceiving a MSE of 413.72 from the training set and a MSE of 414.08 from the cross-validation implies a good generalization over the entire dataset. To extract weighted features from the model, it is necessary to disconnect the final layer (L_5) from the entire model and rewire the logistic layer (L_4) directly to the output node. Applying such modification to the network is essential for preventing the linear layer from multiplying weighted features into a single vector. This truncated version of the MLP (tMLP) preserves the weights and biases of the trained network, and therefore, performs the same denoising, sparsing and weighting tasks on the inputs. However, the output of the tMLP is not a one-dimensional approximation of the EPi, but rather a 10-dimensional representation of inputs that are weighted accordingly. In other words, the tMLP encodes a vector of 10 values (one obtained from each hidden logistic unit), as the output for every building.

Since conceiving a multi-dimensional output can be challenging, two post-processing techniques are used to facilitate the decision making process, i.e. dimensionality reduction and clustering. To reduce the multi-dimensional output into a single-dimensional vector, once again we resort to a sparse autoencoder. Similar to the previous implementations, the cost function of this sparse autoencoder is to reduce the reconstruction error of the input data (Eq. (3)). Since a single-dimensional output is intended for each building, only one hidden unit is assumed for the autoencoder. The process of sparsing a 10-dimensional output into a single dimension returned a small reconstruction MSE of 3.75 e−04, indicating trivial information loss. The output of the

autoencoder dubbed the Energy Retrofit index (ERi) assigns a single value to each sample ranging from zero to one. An ERi value close to one designates high performance building characteristics, while a low ERi value indicates merit for energy retrofit. An alternative to dimensionality reduction through autoencoders is assigning Energy Retrofit labels by means of clustering (Eq. (6)). The multi-dimensional outputs of the tMLP is fed to a k-means clustering algorithm. The optimum number of clusters is determined by performing a step by step search from 5 to 15 centroids. Since k-means algorithm is randomly initialized, the clustering process was repeated 100 times to ensure that the smallest SoD is achieved. The goodness of the clusters were evaluated by four criterions (Eqs. (7)–(10)), which are mean and range normalized. Fig. 6 displays the goodness of various clusters and their respective performance according to each criterion. The optimum number of clusters was set to 9, while observing that any choice between 7 and 10 centroids would return similar performances.

3.3. Results and validation

3.3.1. Energy retrofit index

The ERi should simultaneously comply with two objectives: (1) demonstrate a correlation with the EPi and (2) rank buildings based on retrofitable characteristics.

To contrast the capabilities of the new index, a comparison between the ERi and EPi measures is provided. Fig. 7 represents each building by means of two values, i.e. the EPi (plotted by a blue dot), and the ERi (plotted with a red square). To facilitate the comparison, samples are sorted from small to large based on the EPi measure. Although the polynomial trendline fitted over the ERi scatter highlights the correlation between the two indicators, the ERi is not an exact inverse replica of the EPi. This is evident as some buildings return low values at both EPi and ERi measures. These buildings here dubbed "outlier samples", simultaneously display high energy performance and poor characteristics in the envelope or the heating system. The outliers have at least one property which is notably different from that of similar buildings with comparable EPi. Table 3 provides an in-depth view into the properties of 10 outlier buildings highlighted with cross-shaped marks in Fig. 8.

As mentioned before, the EPi is a function of both retrofitable (e.g. envelope) and non-retrofitable (e.g. S/V ratio) building characteristics. Consequently, although a low EPi indicates low energy use intensity, it does not necessarily point to high performance envelope properties. This is evident as the existing certification tool (EPi measure) may assign low energy use intensity labels to buildings with poor

characteristics. Take for example sample 129 with a window U-value of 4.51 W/m^2 K (Table 3). Since the EPC is highly dependent on geometry characteristics, the poor thermal conductivity of the window is not reflected in the EPi assigned to the building. Consequently, it is impossible to differentiate samples with poor characteristics, solely based on the EPi value.

The ERi measure on the other hand, succeeds to capture such poor performances in the building characteristics as displayed in Fig. 8. Recall that sample 129 suffers from a poor thermal conductivity in the window component. The ERi value assigned to sample 129 is con-siderably lower than the average ERi allocated to samples with high performance windows (Table 3). Consequently, it can be argued that the ERi measure is self-sufficient for ranking buildings based on char-acteristics that are subject to retrofit. This is also evident as the ERi index displays high correlation with the EPi index as well as the building characteristics (Table 4). The only exception is the correlation between the EPi and the Eff_G. However, it can be argued that the ob-served low correlation between ERi and Eff_G is a result of CENED's calculation procedure, as a similar coefficient of covariation is per-ceived between the EPi and Eff_G Table 5.

Observing a high correlation between the ERi and envelope char-acteristics is the result of data preprocessing where nonlinear principal components are extracted from the inputs. To highlight the effective-ness of the unsupervised preprocessing stage, a comparison between the outputs of two neural networks with different architectures is provided. The first network is a vanilla neural network with a single logistic layer and no preprocessing. The second network is the tMLP with two pre-processing stages and one logistic layer. Since the input data of the vanilla neural network is not resilient to noisy data, the outputs display higher dispersion when compared to the ERi. This is also evident from the highly skewed distribution of outputs obtained from the vanilla neural network. Moreover the outputs of the vanilla neural network display smaller covariation ($R^2 = 0.12$) with the EPi values, when compared to that of the sparsed tMLP ($R^2 = 0.42$) as demonstrated in Fig. 9.

3.3.2. Energy retrofit label

Similar to the ERi measure, the Energy Retrofit label should also display reasonable correlation with building characteristics. This is in fact the main advantage of the Energy Retrofit labels, over the tradi-tional EPC energy classes. To contrast the effectiveness of the frame-work in capturing nonlinear covariations, a comparison between the conventional EPC classes and the proposed Energy Retrofit labels is provided (Fig. 11). Each building is represented by the 1st and 2nd
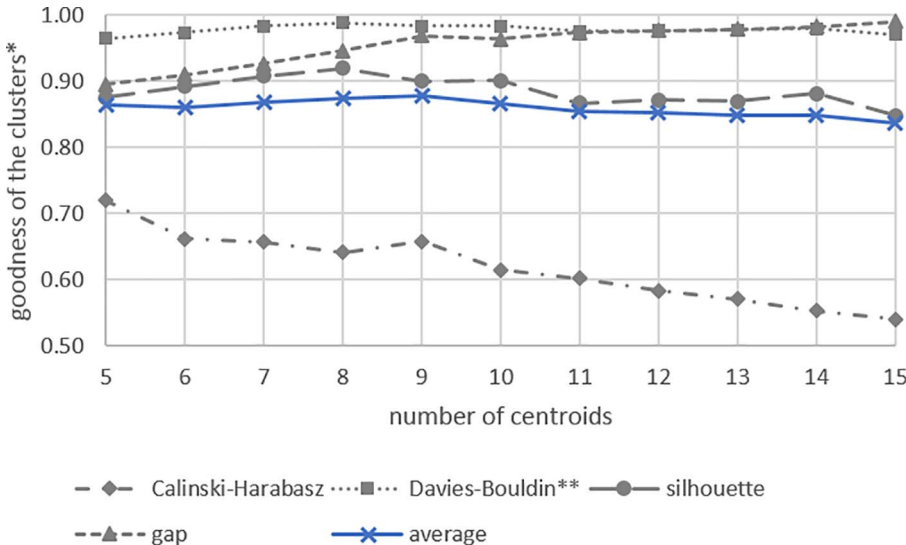

Fig. 6. Optimizing the number of clusters (5–15) according to 4 criterions. *The criterions are mean and range normil-zied, **The Davies-Bouldin criterion is inverted to facilitate the comparison.
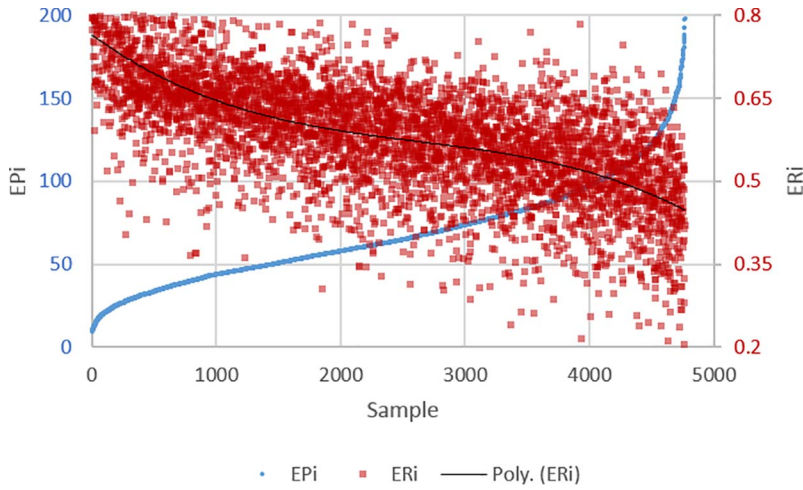
Fig. 7. Comparison of EPi and ERi indexes. To facilitate the comparison, samples are sorted from small to large based on the EPi measure.

Table 3
Properties of tools used in the casestudy. For a detailed description of the steps refer to Fig. 5.

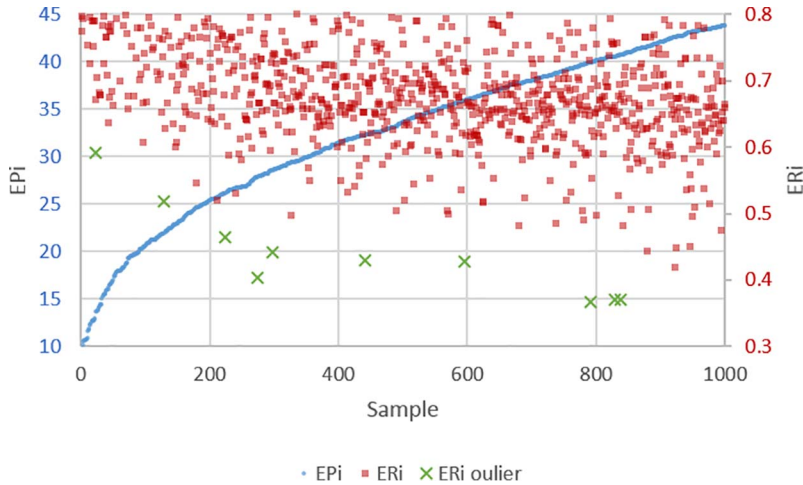| Step	Tool	Input size	Target size	Network layout	Output size	Objective	Error	
1	AE1	[4767,5]	–	{5,8,5,5}	–	Training	7.15E − 04	MSEsparse
2	AE1	[4767,5]	–	{5,8}	[4767,8]	Encoding		
3	AE2	[4767,8]	–	{8,5,8,8}	–	Training	7.66E − 06	MSEsparse
4	AE2	[4767,8]	–	{8,5}	[4767,5]	Encoding		
5	NN	[4767,5]	[4767,1]	{5,10,1}	–	Training	477.32	MSE
6	MLP	[4767,5]	[4767,1]	{5,8,5,10,1}	–	Training	414.08	MSE
7	tMLP	[4767,5]	–	{5,8,5,10}	[4767,10]	Encoding		
8	AE3	[4767,10]	–	{10,1,10,10}	–	Training	3.75E − 04	MSEsparse
9	AE3	[4767,10]	–	{10,1}	[4767,1]	Encoding		
10	k-means	[4767,10]	–	–	[4767,1]	Training	42.1271	SoD

Fig. 8. Illustration of 10 outliers from the first 1000 samples. Additional information on the performance of ERi outliers is provided in Table 3.

Table 4
The covariation between building characteristics, EPi and ERi measures. Values within the white background display the coefficient of covariation (R^2). Values inside the shaded cells refer to the SSE of residuals.

	U_e	U_r	U_b	U_w	Eff_G	EPi	ERi
U_e	-	1720	931.9	4178	28.03	3.07E+06	26.32
U_r	0.1326	-	967.3	4300	28.02	3.41E+06	37.97
U_b	0.1745	0.1432	-	4135	27.9	3.51E+06	36.74
U_w	0.095	0.0686	0.1044	-	27.79	3.78E+06	25.27
Eff_G	0.0001	0.0005	0.0048	0.0088	-	4.24E+06	51.18
EPi	0.3007	0.2232	0.2001	0.1384	0.034	-	31.46
ERi	0.5182	0.305	0.3275	0.5375	0.0632	0.4242	-

Table 5
Properties of ERi outliers. The numbers in <u>bold</u> are the outliers highlighted with a cross in Fig. 8. The numbers in <u>italic</u> are the samples close to the outlier in Fig. 8.

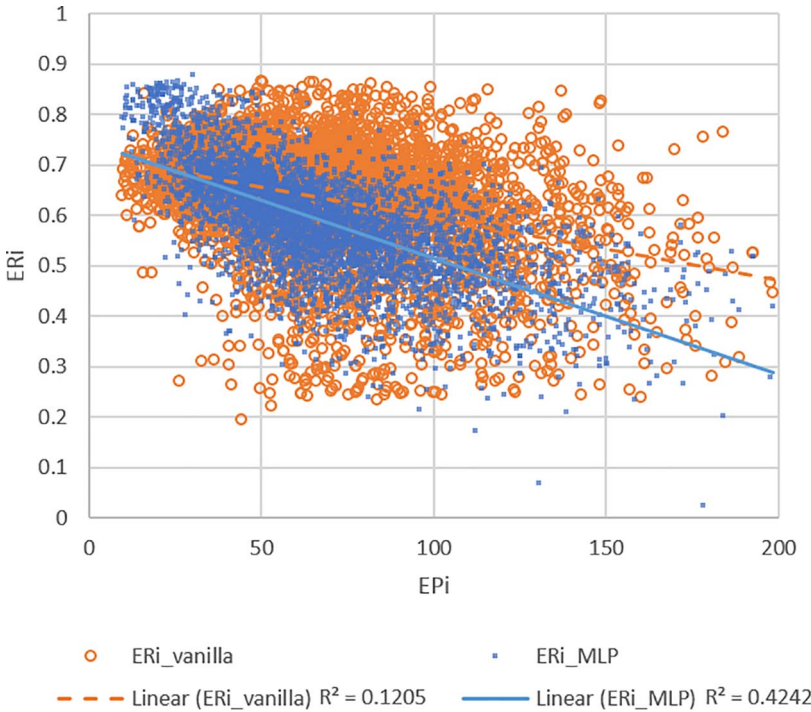|	U-value wall	U-value roof	U-value ground floor	U-value window	Global thermal efficiency	EPi	ERi
Sample **129**	1.36	0.79	1.55	4.51	0.74	21.99	0.52
Average of *119–139*	0.57	0.48	0.61	2.15	0.82	22.02	0.77
sample **225**	3.02	1.5	1.5	3.79	0.75	26.21	0.46
Average of *215–235*	0.59	0.53	0.81	2.22	0.76	26.25	0.75
Sample **275**	1.53	1.43	1.25	5.66	0.69	27.94	0.4
Average of *265–285*	0.57	0.76	0.81	2.85	0.77	27.86	0.7
Sample **298**	1.78	1.7	1.65	3.78	0.71	28.58	0.44
Average of *288–308*	0.58	0.62	0.74	2.89	0.78	28.56	0.7
Sample **326**	1.05	1.7	1.45	5.6	0.76	29.38	0.5
Average of *316–336*	0.57	0.57	0.85	3.16	0.78	29.35	0.7
Sample **441**	1.54	1.8	1.3	5.66	0.8	32.26	0.43
Average of *431–451*	0.77	0.67	1.01	3.09	0.77	32.3	0.67
Sample **596**	0.92	3.61	1.58	5.21	0.71	35.96	0.43
Average of *586–606*	0.82	0.57	1.06	3.29	0.74	35.94	0.67
Sample **792**	2.05	1.5	1.65	5.66	0.84	39.95	0.37
Average of *782–802*	0.7	0.97	1.03	3.28	0.74	39.98	0.65
Sample **839**	2.05	1.5	1.65	5.66	0.85	40.79	0.37
Average of *829–849*	0.96	0.93	1	3.02	0.77	40.76	0.63

Fig. 9. The correlation of EPi values with: (1) ERi measure and (2) outputs of a vanilla neural network.

principal components of the input data, while the energy labels are differentiated through colors. It is observed that EPC class G covers the entire length of the 1st principal component, implying a small correlation between the principal components and the EPC classes. On the contrary, the Energy Retrofit labels gradually changes alongside both axes, indicating a high correlation between the principal components and the Energy Retrofit labels. This issue can be further stressed by comparing the two labelling approaches based on the thermal conductivity of building components. It is observed that the EPC class G covers the entire range of both axes, implying that there is no specific correlation between the thermal conductivity of the walls and windows with the EPC label. This clearly underlines the inapplicability of EPCs in classifying buildings based on retrofitable characteristics. On the other hand, each Energy Retrofit label only covers a limited range on both axes, implying a strong correlation between the introduced labels and the thermal conductivity of walls and windows.

The Energy Retrofit labels should also demonstrate a covariation with the EPi, as components with larger contribution in the energy

consumption should consist of higher weights. Therefore, comparing the range of EPi in both labelling approaches is necessary (Fig. 12). The comparison indicates that the EPC classes do not have any overlap regarding the EPi values in each class, yet, different Energy Retrofit labels may share similar EPi values. Nevertheless, the sequence of Energy Retrofit labels (I to IX) increases based to the magnitude of the EPi values. In other words, higher EPi values are generally assigned with inferior Energy Retrofit labels. Therefore, it is safe to assume that there is a strong correlation between the EPi measure and the proposed Energy Retrofit labels.

It is important to note that the clusters provided by k-means algorithm can be merged according to the objective of the decision makers. For instance, buildings contained in label XIII and IX may be merged in order to create a larger group, which has the highest priority to undergo an energy retrofit. Moreover, each number of clusters may be fit for a specific application. Therefore, the desired number of clusters for prioritizing buildings which merit financial aid for installing renewable sources may be different from the intended number of groups to prioritize loans for energy retrofit.
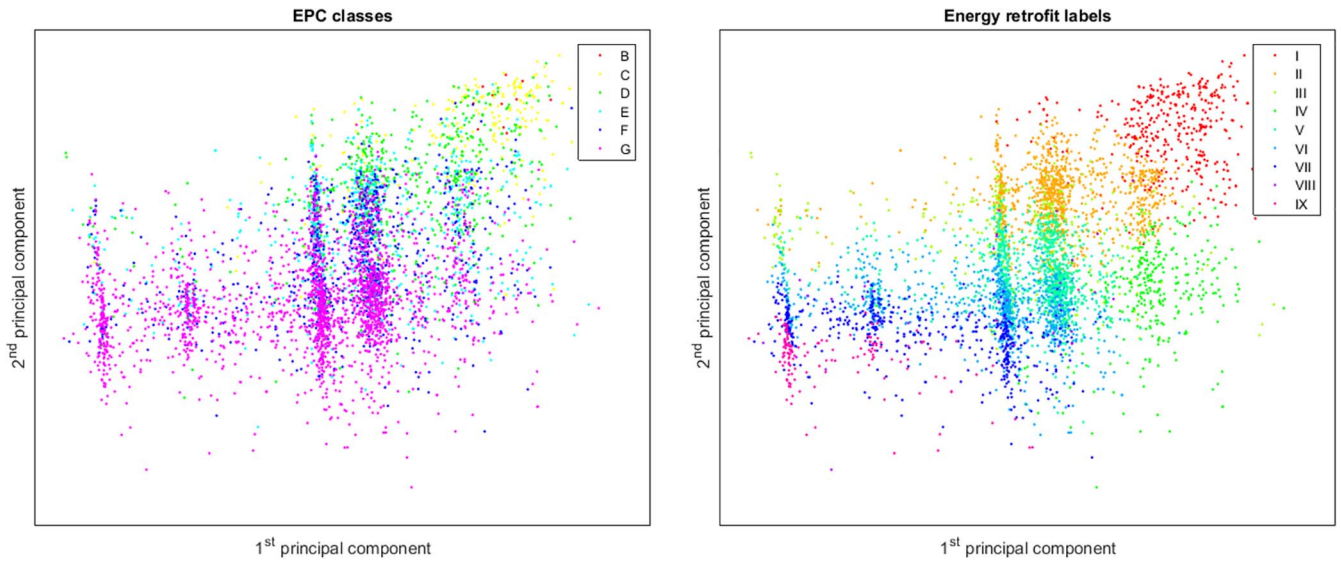
Fig. 10. Heat map of EPC classes (left) and energy retrofit labels (right) on first two principal components of building characteristics.
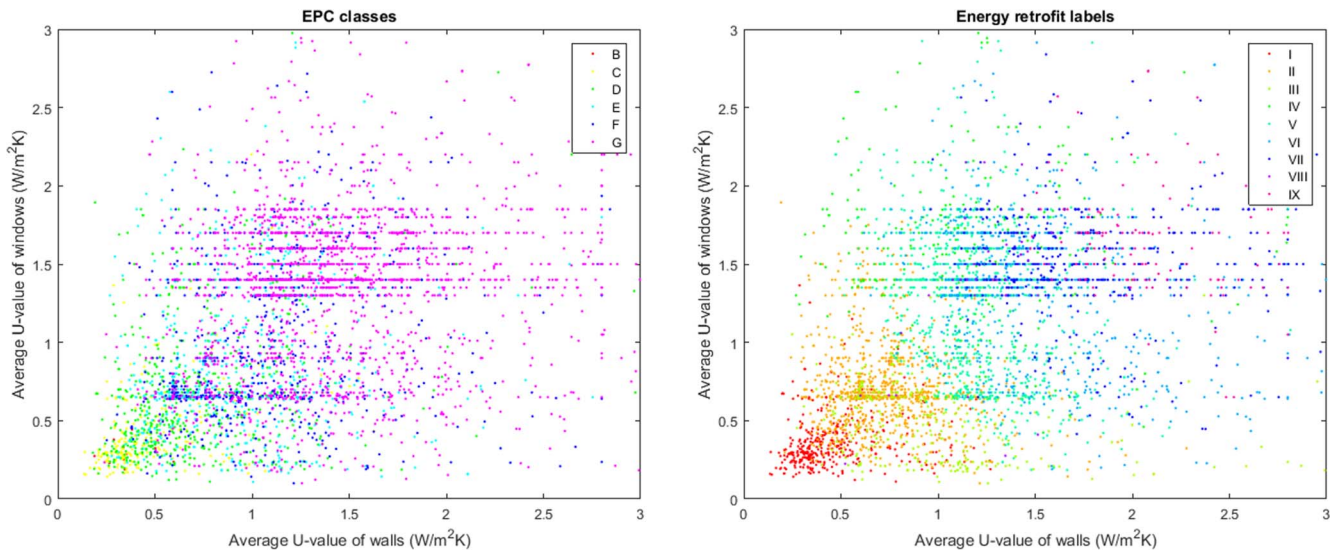

Fig. 11. Heat map of EPC classes (left) and energy retrofit labels (right) on thermal conductivity of walls and windows.

3.3.3. Application in decision support

As mentioned earlier, the energy class provided by the EPC is based on the EPi index, and therefore, underestimates the share of building components from an energy retrofit perspective. On the other hand, the Energy Retrofit label groups buildings based on their retrofitable characteristics. Consequently, one cannot assure that every building in EPC class G has poor characteristics, and vice versa. However, it is possible to guarantee poor envelope properties or heating systems, by resorting to lower Energy Retrofit labels i.e. VIII and IX. Fig. 13 displays a comparison between the quantities of buildings that each labelling strategy shares, as well as the propagation of the certificates from CENED energy classes into the proposed Energy Retrofit labels. It is observed that buildings from EPC classes "C" to "G" are distributed over the whole 9 categories of the Energy Retrofit labels. Therefore, a building previously categorized in a low energy performance group according to the EPC classification may in fact consist of a high performance envelope and heating system.

Let us assume that the administration of Lombardy region in Italy tends to provide financial assistance for incorporating one megawatts of photovoltaic panels in office buildings. At this point, a decision maker has to cherry pick a number of buildings that merit the financial aid. To avoid energy waste, priority is given to buildings which already consist of high performance envelope and systems. Therefore, one may assume that lower EPi values is consistent with better envelope properties. Assuming that each building can handle a 10 kilowatt photovoltaic plant, the first 100 buildings with the smallest EPi values are selected. The properties of the selected buildings is displayed in Table 6. This selection contains buildings with a wall U-value of 1.5 W/m^2 K, and a window U-value of 3.82 W/m^2 K. Meanwhile, by resorting to the ERi criteria, another group of 100 buildings can be selected, in which the maximum thermal conductivity of the walls and windows do not exceed 1.0 W/m^2 K and 2.2 W/m^2 K respectively. A deeper study into Table 6 reveals that the group of buildings selected according to the EPi have higher medians and larger variances in all building characteristics. Therefore, not only the average performance of the selected buildings are lower when one uses the EPi criteria, but also buildings belonging to a completely different class due to poor characteristics may be short-listed as a suitable nominee.

The application of the ERi is not only limited to after the policy making process, as allocating the right amount of financial aid at each step of the planning process may be challenging. Therefore, policy makers can resort to the ERi for selecting the right financial aid for each
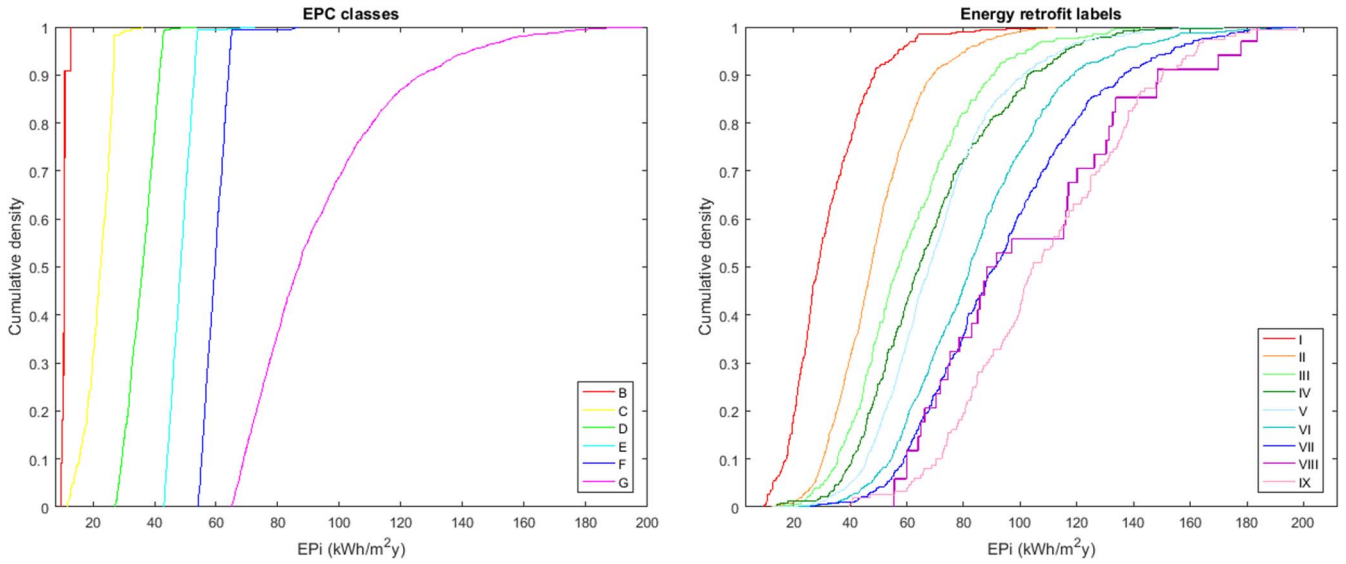
Fig. 12. The cumulative distribution of EPi values in EPC classes (left) and energy retrofit labels (right).
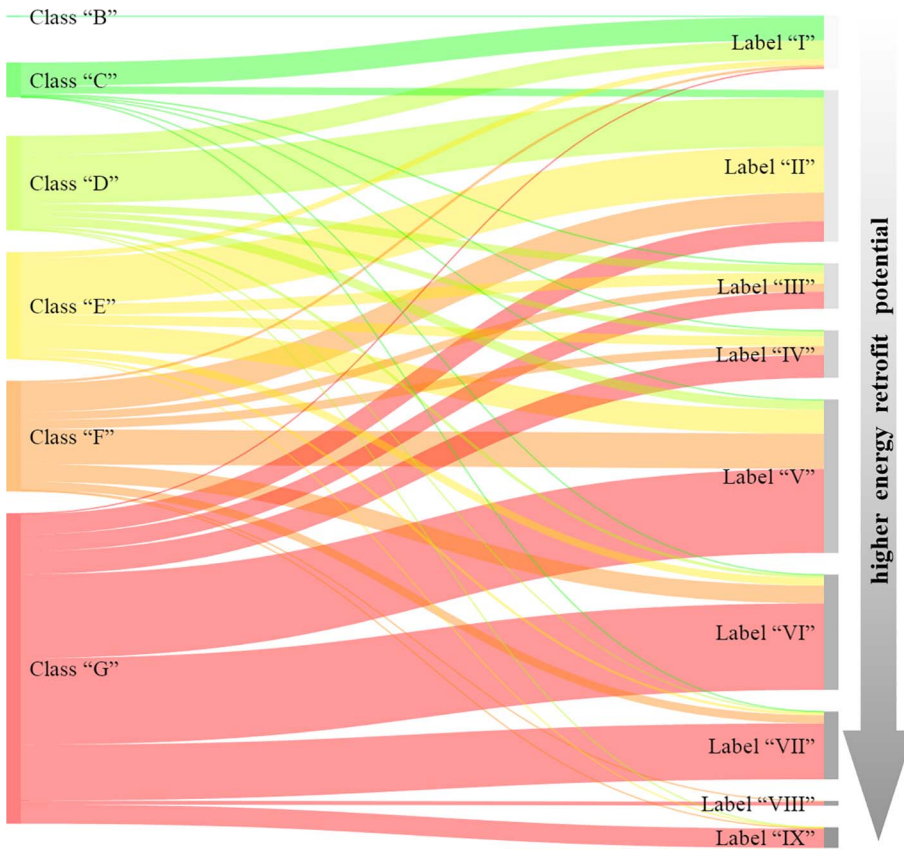

Fig. 13. A Sankey diagram mapping the flow of buildings from EPC classes (left) to energy retrofit labels (right).

building, then create trajectories of the possible updates in the building stock according to the planned policies, and eventually fine-tune the policies in order to meet energy efficiency goals.

4. Conclusion

In this study a new framework was introduced to support decisions related to in situ conditions of building characteristics. The study opted for two main machine learning approaches, namely, supervised learning and unsupervised learning. Unsupervised learning was used for preprocessing the data with the objective of either denoising,

sparsing or clustering. Supervised learning was used for assigning weights to building characteristics, based on the magnitude of their contribution in the energy consumption. The introduced framework facilitates the process of ranking buildings based on energy use intensity and retrofitable characteristics through two labelling approaches. The ERi was presented as a new indicator which assigns higher values to buildings with better performances. An energy retrofit label was also introduced to cluster buildings in a predefined number of partitions.

The framework was validated through a case study of 4767 office buildings, registered in the EPC database of Lombardy region in Italy.

Table 6
Distributions of building properties for 100 shortlisted samples. A comparison on decision making based on EPi and ERi measures.

	EPi			ERi		
	5th percentile	Median	95th percentile	5th percentile	Median	95th percentile
U_e	0.19	0.34	1.07	0.19	0.33	0.68
U_r	0.19	0.32	0.92	0.16	0.28	0.43
U_b	0.21	0.45	1.52	0.20	0.39	1.44
U_w	1.33	1.95	3.47	1.12	1.61	1.91
Eff_G	0.70	0.82	0.92	0.70	0.84	0.94

Thermal conductivity of envelope components and the global efficiency of the heating systems were considered as characteristics that can be subject to energy retrofit. A comparative study between the EPi and ERi measures displayed that the EPi is prone to ignore poor envelope characteristics, as the indicator is highly biased towards geometric characteristics. Also, it was showed that the newly introduced ERi index was able to correctly rank buildings based on the intended criteria. This was evident from the overall correlation of the energy retrofit indicator and the primary energy index, as well as the capacity to single out buildings with poor characteristics. It was highlighted that since the traditional energy classification labels can be insensitive to the building characteristics, the EPC indicators may render unsuitable for supporting decision making strategies aimed at large scale retrofits.

Since the objective of the study was to support decisions aimed at regional scales, the economical feature of energy retrofit was not included in the ERi measure. This is due to the fact that the financial aspect of building energy retrofit is highly reliant on each owners' perception of long term investment. However, the ERi can be merged with the probability theory (e.g. Bayes' theorem) to create trajectories of possible retrofit scenarios, and roughly include the economical aspect into the framework. The proposed framework can be further validated with other EPC databases, as different regions/countries resort to diverse methodologies for ranking building energy performance. Considering that the process of ranking buildings becomes more difficult when dealing with high-dimensional data, a possible extension to the current study would be to reconfigure the introduced labels for covering cooling, lighting and domestic hot water energy consumption.

References

[1] Ekins P, Lees E. The impact of EU policies on energy use in and the evolution of the UK built environment. Energy Policy 2008;36(12):4580–3.
[2] Lee W, Yik F. Regulatory and voluntary approaches for enhancing building energy efficiency. Prog Energy Combust Sci 2004;30(5):477–99.
[3] Pérez-Lombard L, Ortiz J, González R, Maestre IR. A review of benchmarking, rating and labelling concepts within the framework of building energy certification schemes. Energy Build 2009;41(3):272–8.
[4] Burman E, Mumovic D, Kimpian J. Towards measurement and verification of energy performance under the framework of the European directive for energy performance of buildings. Energy 2014;77:153–63.
[5] Ahern C, Norton B, Enright B. The statistical relevance and effect of assuming pessimistic default overall thermal transmittance coefficients on dwelling energy performance certification quality in Ireland. Energy Build 2016;127:268–78.
[6] Mangold M, Österbring M, Wallbaum H, Thuvander L, Femenias P. Socio-economic impact of renovation and energy retrofitting of the Gothenburg building stock. Energy Build 2016;123:41–9.
[7] Jakob M. Marginal costs and co-benefits of energy efficiency investments: The case of the Swiss residential sector. Energy Policy 2006;34(2):172–87.
[8] Dascalaki EG, Balaras CA, Kontoyiannidis S, Droutsa KG. Modeling energy refurbishment scenarios for the Hellenic residential building stock towards the 2020 & 2030 targets. Energy Build 2016;132:74–90.
[9] Hyland M, Lyons RC, Lyons S. The value of domestic building energy effi-ciency—evidence from Ireland. Energy Econ 2013;40:943–52.
[10] Fuerst F, McAllister P. The impact of energy performance certificates on the rental and capital values of commercial property assets. Energy Policy 2011;39(10):6608–14.
[11] Amecke H. The impact of energy performance certificates: a survey of German home owners. Energy Policy 2012;46:4–14.
[12] Banfi S, Farsi M, Filippini M, Jakob M. Willingness to pay for energy-saving measures in residential buildings. Energy Econ 2008;30(2):503–16.
[13] Farsi M. Risk aversion and willingness to pay for energy efficient systems in rental apartments. Energy Policy 2010;38(6):3078–88.
[14] Rosenow J, Fawcett T, Eyre N, Oikonomou V. Energy efficiency and the policy mix. Build Res Inform 2016;44(5–6):562–74.
[15] Bull R, Chang N, Fleming P. The use of building energy certificates to reduce energy consumption in European public buildings. Energy Build 2012;50:103–10.
[16] Majcen D, Itard L, Visscher H. Theoretical vs. actual energy consumption of labelled dwellings in the Netherlands: discrepancies and policy implications. Energy Policy 2013;54:125–36.
[17] Sarto L, Sanna N, Tonetti V, Ventura M. On the use of an energy certification database to create indicators for energy planning purposes: Application in northern Italy. Energy Policy 2015;85:207–17.
[18] Droutsa KG, Kontoyiannidis S, Dascalaki EG, Balaras CA. Mapping the energy performance of hellenic residential buildings from EPC (energy performance certificate) data. Energy 2016;98:284–95.
[19] Galante A, Pasetti G. A methodology for evaluating the potential energy savings of retrofitting residential building stocks. Sustain Cities Soc 2012;4:12–21.
[20] Galante A, Torri M. A methodology for the energy performance classification of residential building stock on an urban scale. Energy Build 2012;48:211–9.
[21] Ryghaug M, Sørensen KH. How energy efficiency fails in the building industry. Energy Policy 2009;37(3):984–91.
[22] López-González LM, López-Ochoa LM, Las-Heras-Casas J, García-Lozano C. Energy performance certificates as tools for energy planning in the residential sector. The case of La Rioja (Spain). J Clean Prod 2016;137:1280–92.
[23] Pampuri L, Cereghetti N, Bianchi PG, Caputo P. Evaluation of the space heating need in residential buildings at territorial scale: the case of Canton Ticino (CH). Energy Build 2017;148:218–27.
[24] Casals XG. Analysis of building energy regulation and certification in Europe: Their role, limitations and differences. Energy Build 2006;38(5):381–92.
[25] Ballarini I, Corrado V. Application of energy rating methods to the existing building stock: analysis of some residential buildings in Turin. Energy Build 2009;41(7):790–800.
[26] Fabbri K. Planning a regional energy system in association with the creation of energy performance certificates (EPCs), statistical analysis and energy efficiency measures: an Italian case study. Buildings 2013;3(3):545–69.
[27] Friege J, Chappin E. Modelling decisions on energy-efficient renovations: a review. Renew Sustain Energy Rev 2014;39:196–208.
[28] Walter T, Sohn MD. A regression-based approach to estimating retrofit savings using the building performance database. Appl Energy 2016;179:996–1005.
[29] Wu R, Mavromatidis G, Orehounig K, Carmeliet J. Multiobjective optimisation of energy systems and building envelope retrofit in a residential community. Appl Energy 2017;190:634–49.
[30] Chung W. Review of building energy-use performance benchmarking methodologies. Appl Energy 2011;88(5):1470–9.
[31] Clune S, Morrissey J, Moore T. Size matters: House size and thermal efficiency as policy strategies to reduce net emissions of new developments. Energy Policy 2012;48:657–67.
[32] Aksoezen M, Daniel M, Hassler U, Kohler N. Building age as an indicator for energy consumption. Energy Build 2015;87:74–86.
[33] Johansson T, Olofsson T, Mangold M. Development of an energy atlas for renovation of the multifamily building stock in Sweden. Appl Energy 2017;203:723–36.
[34] Amstalden RW, Kost M, Nathani C, Imboden DM. Economic potential of energy-efficient retrofitting in the Swiss residential building sector: The effects of policy instruments and energy price expectations. Energy Policy 2007;35(3):1819–29.
[35] Zhou N, Levine MD, Price L. Overview of current energy-efficiency policies in China. Energy Policy 2010;38(11):6439–52.
[36] Xu P, Chan EH-W, Qian QK. Success factors of energy performance contracting (EPC) for sustainable building energy efficiency retrofit (BEER) of hotel buildings in China. Energy Policy 2011;39(11):7389–98.
[37] Mathew PA, Dunn LN, Sohn MD, Mercado A, Custudio C, Walter T. Big-data for building energy performance: Lessons from assembling a very large national database of building energy use. Appl Energy 2015;140:85–93.
[38] Hong T, et al. Commercial building energy saver: an energy retrofit analysis toolkit. Appl Energy 2015;159:298–309.
[39] Deru M et al. US Department of Energy commercial reference building models of the national building stock; 2011.
[40] Rajagopalan P, Leung Tony C. Progress on building energy labelling techniques. Adv Build Energy Res 2012;6(1):61–80.
[41] Nikolaou T, Kolokotsa D, Stavrakakis G. Review on methodologies for energy benchmarking, rating and classification of buildings. Adv Build Energy Res 2011;5(1):53–70.
[42] Koo C, Hong T, Lee M, Park HS. Development of a new energy efficiency rating system for existing residential buildings. Energy Policy 2014;68:218–31.
[43] Santamouris M, et al. Using intelligent clustering techniques to classify the energy performance of school buildings. Energy Build 2007;39(1):45–51.
[44] Gao X, Malkawi A. A new methodology for building energy performance benchmarking: an approach based on intelligent clustering algorithm. Energy Build 2014;84:607–16.
[45] Nikolaou TG, Kolokotsa DS, Stavrakakis GS, Skias ID. On the application of clustering techniques for office buildings' energy and thermal comfort classification. IEEE Trans Smart Grid 2012;3(4):2196–210.
[46] Koo C, Hong T. Development of a dynamic operational rating system in energy performance certificates for existing buildings: geostatistical approach and data-mining technique. Appl Energy 2015;154:254–70.
[47] Pérez-Lombard L, Ortiz J, Velázquez D. Revisiting energy efficiency fundamentals.

Energ Effi 2013;6(2):239–54.

[48] López-González LM, López-Ochoa LM, Las-Heras-Casas J, García-Lozano C. Update of energy performance certificates in the residential sector and scenarios that consider the impact of automation, control and management systems: a case study of La Rioja. Appl Energy 2016;178:308–22.

[49] Haykin SS. Neural networks and learning machines. NJ, USA: Pearson Upper Saddle River; 2009.

[50] Friedman J, Hastie T, Tibshirani R. The elements of statistical learning. Springer series in statistics. Berlin: Springer; 2001.

[51] Demuth HB, Beale MH, De Jess O, Hagan MT. Neural network design. Martin Hagan 2014.

[52] Ascione F, Bianco N, De Stasio C, Mauro GM, Vanoli GP. Artificial neural networks to predict energy performance and retrofit scenarios for any member of a building category: a novel approach. Energy 2017;118:999–1017.

[53] Wang E, Shen Z, Grosskopf K. Benchmarking energy performance of building envelopes through a selective residual-clustering approach using high dimensional dataset. Energy Build 2014;75:10–22.

[54] Bourlard H, Kamp Y. Auto-association by multilayer perceptrons and singular value decomposition. Biol Cybern 1988;59(4):291–4.

[55] Vincent P, Larochelle H, Bengio Y, Manzagol P-A. Extracting and composing robust features with denoising autoencoders. In: Proceedings of the 25th international conference on Machine learning. ACM; 2008. p. 1096–1103.

[56] Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. Science 2006;313(5786):504–7.

[57] Møller MF. A scaled conjugate gradient algorithm for fast supervised learning. Neural Networks 1993;6(4):525–33.

[58] Bengio Y, Lamblin P, Popovici D, Larochelle H. Greedy layer-wise training of deep networks. In: Advances in neural information processing systems, vol. 19; 2007, p. 153.

[59] Vincent P, Larochelle H, Lajoie I, Bengio Y, Manzagol P-A. Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. J Mach Learn Res 2010;11(Dec):3371–408.

[60] Arthur D, Vassilvitskii S. k-means + +: The advantages of careful seeding. In: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms. Society for Industrial and Applied Mathematics; 2007. p. 1027–1035.

[61] Caliński T, Harabasz J. A dendrite method for cluster analysis. Commun Stat-theory Meth 1974;3(1):1–27.

[62] Davies DL, Bouldin DW. A cluster separation measure. IEEE Trans Pattern Anal Mach Intell 1979;2:224–7.

[63] Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math 1987;20:53–65.

[64] Tibshirani R, Walther G, Hastie T. Estimating the number of clusters in a data set via the gap statistic. J Roy Stat Soc: Series B (Stat Methodol) 2001;63(2):411–23.

[65] Li Z, Han Y, Xu P. Methods for benchmarking building energy consumption against its past or intended performance: an overview. Appl Energy 2014;124:325–34.

[66] Wang S, Yan C, Xiao F. Quantitative energy performance assessment methods for existing buildings. Energy Build 2012;55:873–88.

[67] Regional Council Decree No. VIII/5018 of 06/26/07, Determinations Regarding the Buildings Energy Certification. In: Implementation of Legislative Decree No. 192/05 and Art.9:25 of Regional Law No. 24/06; 2007.

[68] Regional Council Decree No. VIII/5773 of 10/31/07, Buildings Energy Certification – Amendments and Additions to Regional Council Decree No. VIII/5018; 2007.

[69] Director General Decree No. 5796 of 06/11/09, Updating of the Procedure for Calculating the Energy Certification of Buildings.; 2009.

[70] Regional Council Decree No. IX/2554 of 11/24/11, Policy Address, Method of Investigation of Offenses and Sanctions Imposition of Regional Competence, provided by Art.27 of Law No.24, in the Field of Energy Certification; 2011.