Journal of Cleaner Production 355 (2022) 131626

Contents lists available at ScienceDirect

Journal of Cleaner Production

journal homepage: www.elsevier.com/locate/jclepro

Explainable artificial intelligence for building energy performance certificate labelling classification

Thamsanqa Tsoka [a], Xianming Ye [a,*], YangQuan Chen [b], Dunwei Gong [c,*], Xiaohua Xia [a]

[a] Department of Electrical, Electronic and Computer Engineering, University of Pretoria, Pretoria 0002, South Africa
[b] Mechanical Engineering Department, University of California, Merced, CA 95340, USA
[c] School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China

ARTICLE INFO

Handling Editor: M.T. Moreira

Keywords:
Building EPC
ANN
Artificial intelligence
XAI
Machine learning

ABSTRACT

The building energy performance certificates (EPC) are widely adopted for sustainable development and improvement in building energy efficiency. Different from the conventional direct measurement based approach of acquiring a building's EPC label, this study proposes a novel and alternative approach to classify a building's EPC label using artificial neural network (ANN) models. Given the extensive best building EPC practices in developed countries, historical building EPC data and experiences can expedite the development and improvement of this procedure in developing countries. This study first develops the ANN classification model to attain the building EPC label. The classification result shows that the building EPC classification can achieve a 99% precision with sufficient input data. With the assistance of explainable artificial intelligence (XAI) tools such as the Local Interpretable Model-Agnostic Explanation (LIME) and SHapley Additive exPlanation (SHAP), some less important input features for the ANN classification models can be removed without severely influencing the ANN model's accuracy. In the case studies, the EPC best practices historical registry data from Lombardy, Italy are used in training the ANN model. The ANN models' accuracy for the case study 1 is 93% with 14 input features where CO_2 emissions and net surface area are the two most influential features. The most influential input feature for case study 2 is the winter AC non-renewable energy performance, and the accuracy of the case study 2 ANN model is 89% with 26 input features.

1. Introduction

The building sector contributes to 40% of the global energy consumption (Zhao and Magoulès, 2012) and 36% of greenhouse gas emissions (Bagheri et al., 2013). The building sector hence embodies great opportunities for global energy use reduction through energy efficiency measures and policies. The building energy performance certificate (EPC) programme is one of the most popular energy policy instruments that regulates the large-scale building energy efficiency and carbon emissions. Information acquired from the EPC assists tenants, owners, and facility managers to adopt energy efficiency measures to achieve lower carbon dioxide (CO_2) emissions and energy consumption while reducing building operation costs. In many countries, the building EPC label is determined by the energy intensity in kWh per m^2 per annum, where the kWh refers to the total energy consumption over a year, and m^2 refers to the effective floor area of the building under assessment (Wang et al., 2012). The EPC is usually supplemented with an advisory report, which suggests energy efficiency measures that are technically and economically viable for the EPC label improvements.

The general procedure to determine a building EPC label requires four key parameters, namely the annual building energy usage, effective floor areas, relevant climate zone, and a building energy performance reference table (Fan and Xia, 2018a). In practice, accurately quantifying the annual building energy usage is usually a challenging task. The most commonly used method to determine the annual energy usage of a building is through direct measurements. This measurement approach ranges from energy bill readings to complex methods that include building a management monitoring system or sub-metring. Energy bills are valuable documentary evidence used to obtain the annual energy usage. However, there are drawbacks to rely on energy bills for the annual energy usage. Firstly, collecting the full historical energy bills can be tedious, especially for buildings with multiple owners, buildings consuming various forms of energy, or owners who occupy numerous buildings; and secondly, sometimes utility bills are derived based on model predictions instead of actual readings. Even when the actual bills are available through the whole facility measurement, they cannot provide a complete picture of the energy use at the building

* Corresponding authors.
E-mail addresses: xianming.ye@up.ac.za (X. Ye), dwgong@vip.163.com (D. Gong).

https://doi.org/10.1016/j.jclepro.2022.131626
Received 9 December 2021; Received in revised form 15 March 2022; Accepted 31 March 2022
Available online 9 April 2022
0959-6526/

Nomenclature

Abbreviation

AI	Artificial intelligence
ANN	Artificial neural network
CENED	Certificazione Energitica degli edifici
EPC	Energy performance certificates
LIME	Local Interpretable Model-Agnostic Explanation
POET	Performance, operation, equipment, and technology energy efficiencies
SHAP	SHapley Additive exPlanations
XAI	Explainable artificial intelligence

Symbol

AC	Air conditioning
AU_b	Average U-value of basement
AU_r	Average U-value of roof
AU_w	Average U-value of walls
AU_{wn}	Average U-value of windows
CO_2	CO_2 emissions
DC	Diesel consumption
DD	Degree day
DHC	District heating consumption
EV	Energy vector
FT	Fuel type
G_s	Glazed surface
GHE	Global heating efficiency
GHWE	Global heating water efficiency
HDE	Heating and domestic hot water efficiency
LBC	Liquid biomass consumption
NGC	Natural gas consumption
O_s	Opaque surface
RBC	Reference building consumption
RC	Reference class
SA	Useful heated surface area
SAC	Summer air conditioning
SACM	Summer AC medium efficiency
SAPER	Summer AC energy performance renewable energy
SAPNER	Summer AC energy performance non-renewable energy
SBC	Solid biomass consumption
SESA	Summer equivalent solar area
SPVC	Solar photo-voltaic consumption
ST	System type
STC	Solar thermal consumption
SV	Surface area/volume ratio
TT	Thermal transmittance
V	Gross volume heated
WACE	Winter AC medium efficiency
WAEPNR	Winter AC energy performance non-Renewable energy
WAEPR	Winter AC energy performance renewable energy
WANP	Winter AC nominal power
WHT	Winter heating thermal reference

energy system level, which is critical information to identify the energy savings opportunities to improve the EPC label levels. The sub-metring approach has long been proven as an effective and accurate procedure. It is however a lengthy and costly process for metre procurement, installation, metre reading, calibration, and maintenance (Parker et al., 2015). Worse still, waiting for the full interval readings over the desired measurement period, say at least a whole year for the national building EPC programme, may cause significant delays to the EPC roll out at the national level.

The annual energy usage can also be characterised by energy models such as computer simulation models. Some common dynamic simulation tools are DOE-2 (Amasyali and El-Gohary, 2018), TRNSYS (Wang et al., 2012), and EnergyPlus (Pham et al., 2020), which take into consideration features such as weather conditions, building envelopes, building appliances, and building operations. The computer simulation process is precise, but it necessitates a large amount of data, time, and knowledge, along with highly skilled individuals to run the computer simulations. The input data at certain instances, may not be available at the time of simulation, which further increases its complexity. When numerous buildings in a city or region have to be evaluated, the intricacy levels are high and a great deal of effort is required.

The recent inception of the building EPC programme in South Africa demonstrates urgent needs to accurately determine the EPC label and the labelling improvement opportunities for the national buildings that are participating in the programme. The intention of the policy is to classify buildings with reference to their energy intensity value into grades. These grades are defined through a label range from A to G with 'G' symbolising the worst efficiency rating and 'A' being a highly efficient rating for a building. A comparison between reference consumption of that building and its energy intensity provides the grading. The reference consumption of each building is given in SABS (2011) and provides the expected energy intensity consumption for different buildings with different occupancy classes. The policy is deemed to apply to all government owned and occupied buildings greater than 1 000 m^2 and private sector buildings greater than 2 000 m^2 with a required rating of Grade D firstly. The mandatory assessed building categories are offices, places of instruction, entertainment and public assembly. A roll out to other building categories is expected in the near future. Due to the oldness of these buildings they tend to be inefficient and therefore considerable investments are required to improve their rating. These financial implications result in a need for tools that will assist in providing explanations as to what measures may result in the quickest and best way to improve their EPC labels (Fan and Xia, 2018b).

The knowledge of the challenges experienced with acquiring the annual building energy usage via bill reading, sub-metring, or computer based simulation, and recognising the extensive experiences and massive data records available from countries that have implemented the building EPC policies, leads to the impression of whether alternative procedures maybe used. Possibly, the building EPC labels maybe identified via machine learning and big data techniques, which can determine a building's EPC rating level by the most easily attainable building features, characteristics and building energy efficiency classifications such as those in the performance, operation, equipment, and technology energy efficiencies (POET) framework (Xia and Zhang, 2010) of a particular national interest, instead of only relying on the time series numeric interval readings from the energy metres. Once the machine learning approach is available, it will offer the building owners, third-party EPC rating practitioners, researchers, and government an alternative solution, which is fast, easily implementable, and cost-effective, to determine the buildings EPC label.

Implementation of data driven models requires data to enact numerous decisions and perform predictions (Fathi et al., 2020). Random tree forests (RF) (Walker et al., 2020), support vector machines (SVM) (Guo et al., 2018), Extreme Gradient Boosting (Lu et al., 2020), and artificial neural networks (ANN) are prime examples of data driven models (Ascione et al., 2017). Unsupervised, semi-supervised, supervised, and reinforcement learning are four major categories that form part of machine learning techniques. The unsupervised learning technique is used in pre-processing data and analysing the characteristics of the data as the input data are known however, the output data are unknown. Semi-supervised learning incorporates input data with both known and unknown output labels or values making it a hybrid approach of supervised and unsupervised learning (Al-Azzam and Shatnawi, 2021). Reinforcement learning employs sets of input, some output, and grade as training data. It is generally used when optimal interaction is required, such as control and game plays. Supervised learning requires the data to have predetermined input and output pairs to further define a relationship between these inputs and outputs. In supervised learning the training process adjusts the model to limit the error between the correct output and predicted output. There are two distinct procedures in supervised learning namely regression and classification. Regression determines a value mapped from the input data. Classification, categorises the group that the input data is attributed to from the identified groups in the training dataset (Kim, 2017; Fathi et al., 2020). Potential classification limitations are data bias, the need for a substantial amount of historical data, and that predictions of unrepresented classes in the dataset cannot be determined. The implementation of this alternative procedure for classifying buildings' energy performance into the correct EPC label category instead of the direct measurement approach requires the careful avoidance of these three limitations.

Various studies show the implementation of machine learning models in building energy estimations with ANN, random forest, and SVM being the popular choices (Fathi et al., 2020; Zhao et al., 2020). The ANN and SVM may outperform each other in different scenarios. Building energy estimations in Fathi et al. (2020) include but are not limited to the prediction of electricity, cooling/heating loads, natural gas and total energy consumption of various building energy systems and building environment. Building retrofit opportunities are essential solutions to improve the building EPC labels. Studies have shown the use of ANN in addressing retrofit scenarios by the pre-retrofit and post-retrofit analyses (Yalcintas, 2008). Sensitivity analysis has also been widely applied to determine input parameters for the ANN model with the application of particular energy retrofit measures, with the capability of determining the post-retrofit energy consumption values (Ascione et al., 2017). According to Seyedzadeh et al. (2019), the most accurate predictions are provided by ensemble models (Seyedzadeh et al., 2020) such as gradient boost random trees. These ensemble models tend to be a combination of two or more base models (Amasyali and El-Gohary, 2021; Dong et al., 2021) and have been popularised in recent years. Conversely, the support vector machine is generally well suited for instances with smaller data-sets and generic input variables due its simplicity and speed of calculations. However, for complex data-sets it is better to make use of deep neural networks as they perform well in these circumstances in comparison to the other machine learning approaches. The study (Olu-Ajayi et al., 2022) illustrates that deep neural networks perform the best for energy use predictions as compared to SVM, RF, Gradient Boosting, K-Nearest Neighbours, and decision trees. The major drawbacks of deep neural networks are its complexity which require appropriate tuning for good performance and that it is considered an opaque model. Highly performing algorithms such as these neural networks tend to be opaque and require some explanations in order to understand them. These are considered post-hoc explanations implemented through explainable artificial intelligence (XAI) (Alicioglu and Sun, 2021; Belle and Papantonis, 2021).

The ANN models have proven to be accurate and robust in previous studies (Zhao and Magoulès, 2012; Amasyali and El-Gohary, 2018). The input parameters include those derived from measured and computer simulations. The ANN models consequently enable the flexibility to acquire and determine the most essential parameters with respect to time and finances for energy quantification (Yezioro et al., 2008). Through the acquisition of relevant and optimal information, assessment time is reduced as parameters that are acquired over a whole year period maybe reduced. In this study, the building EPC ANN classification model is proposed as a verification procedure and processes where other approaches may not be necessary, resulting in improving EPC labelling procedures. Through the oversight of the actual measurement of the annual energy consumption, new building EPC labelling is enabled.

Artificial Intelligence (AI) models are considered to be black box models as their internal functioning mechanisms are indistinct (Arrieta et al., 2020). The design and deployment of many AI systems require limited human interaction. Interpretability of these black box models to understand how they reach certain decisions is evidently important to enable their applicability in key areas such as law or medicine (Adadi and Berrada, 2018; Goodman and Flaxman, 2017). The XAI has been at the forefront of various research initiatives for this reason. The XAI tools help to identify how each input influences the output in the AI models. The AI models become robust as essential input elements that guarantee accurate decision making are identified (Arrieta et al., 2020). Model operation and post modelling explanations are rendered by the XAI which gives greater insight and understanding (Adadi and Berrada, 2018; Arrieta et al., 2020). The psychology of explanation, a social science field is the major driver of the XAI principles (Miller, 2019). The XAI invokes trust in AI models through explainability, producing good accuracy, and insight on buildings' energy performance and how to operate them. In this study two XAI techniques are adopted to explain the machine learning based building energy classification model, which are the Local Interpretable Model-Agnostic Explanation (LIME) and SHapley Additive exPlanations (SHAP) (Ribeiro et al., 2016), respectively. The LIME and SHAP assist in the explanation and determination of what influence each input feature in the building EPC ANN classification model has for each particular EPC classification output. Moreover, LIME and SHAP with the combination of feature engineering techniques provide and define the essential input features.

This study contributes to both the building EPC labelling and the EPC label improvements. Its objectives are presented in the following aspects:

- Building EPC ANN classification models are developed based on historical building EPC labelling data, which directly classify a building's energy performance to a building EPC label with up to a weighted classification precision of 91.49% for case study 1 and 88.69% for case study 2.
- The proposed building EPC ANN classification approach is superior compared to the traditional measurement based building EPC labelling approaches in following aspects: (1) it does not require long duration of interval measurements of the annual building energy consumption; (2) it enables the new building EPC labelling. However, it is obvious that the proposed building EPC classification approach relies on the availability of historical data.
- Explainable AI techniques such as the LIME and SHAP are adopted to explain the internal classification mechanism with respect to numerous input features in a quantitative manner. The XAI explanations help to identify the key influential factors to the building EPC label. Such findings bring multiple benefits: (1) less important input features can be removed from the original building EPC ANN classification model while maintaining the EPC classification accuracy; (2) providing guidance on the building EPC labelling determinant factors; and (3) reveal the key factors to focus on for one to improve the building EPC label for a given building.

This study is distinct from existing literature in twofold. First, some existing literature (Khayatian et al., 2016) applies the neural networks to estimate the building EPC labelling through regression analysis of the annual building energy usage, while our study develops the ANN model to directly classify the building EPC labels. Moreover, since the ANN models are black box models, the internal processes and reasons for their classifications are unknown. As a result, application reluctance and distrust are increased. The explainable artificial intelligence techniques attempt to address the constraints of this method by offering explanations for the ANN classifications. To the best of our knowledge, this is the first study trying to explain the ANN classifications for building EPC applications. Effectiveness of the proposed building EPC ANN classification models are demonstrated via two case studies based on the Certificazione Energitica degli edifici (CENED) tool derived energy database (Anon, 0000), which comprises of data pertaining to over 500,000 residential building energy certificates. The database splits itself into building data with the old and new energy building EPC labelling methodologies (Anon, 0000). This study analyses both datasets. The new building EPC methodology post 2015 is very similar to that of the South African building EPC policy. The database offers input data such as technical and human influenced building factors, which enables the training, validation, and testing of the building EPC ANN classification model. The results depicted by the EPC classification are analysed through the XAI tools and principles, making the building EPC classification models for both datasets trustable. The key input features that positively or negatively influence the EPC classification are revealed, which are able to assist building owners, tenants, facility managers, and energy auditors in knowing the factors that mainly contribute to the building EPC labels.

The rest of the paper is organised as follows: Section 2 presents the ANN system modelling with feature selection and the machine learning processes for model development and analyses using two XAI tools to explain building EPC classification. Section 3 describes the case studies with Section 4 giving the discussion and results. Section 5 covers the conclusion of the paper.

2. Methodology

The research methodology for this study is shown in Fig. 1. We start with the development of the ANN model for the building EPC label classification. The modelling process mainly includes the feature engineering, model training, and the model performance testing. Defective data of the operational and historical building characteristics data acquired from the CENED database are removed through filters and conditioned through feature engineering principles that improve the performance of the ANN model (Khayatian et al., 2016). The screened data are used for the ANN modelling. The ANN model structure and its parameters are tuned to produce an optimally performing model. Section 2.1 expands on the ANN modelling procedures whereby the model is validated and the model accuracy is analysed. The ANN model's accuracy, transparency and robustness are improved through the guidance of the XAI. The selected XAI tools, SHAP and LIME are capable of explaining the internal functioning mechanisms of the ANN for the building EPC classifications.

2.1. The building EPC ANN classification model

An ANN model consists of three layers that are the input, output and hidden layers, which resemble the concept of a biological neuron in the brain. The three layers are interconnected with the connections leading to the next adjacent layer as depicted in Fig. 1. The interconnected neurons of the activation and transfer functions form the integral part of the ANN. The transfer function propagates the calculated values to the next layer. The activation function consists of weight and bias values that estimate the output values (Buratti et al., 2014; Dayhoff, 1990; Khayatian et al., 2016). The optimisation technique used for the ANN modelling procedures for fine-tuning and updating the weights of the hidden neurons is the adaptive momentum estimation (Ketkar and Santana, 2017; Kingma and Ba, 2015). It provides the perfect balance of speed and accuracy, which is important, given the large size of the dataset for model training (Van Le et al., 2021). The loss function is used to determine the deviation of the model from the predicted and output value. The ANN model implemented in this study utilised the categorical cross entropy for the loss determination in the classification process.

The classic ANN model takes the form of Ketkar and Santana (2017)

$$y = f_x\left(\sum_{i=1}^{n}(w_i X_i + b)\right), \tag{1}$$

where y represents the model output. The weights w_i are multiplied by the input data X_i. The dummy neuron weight value is b, which is the threshold value or bias. The sigmoid, rectified linear, step, or hyperbolic function are application options for the transfer function f_x (Buratti et al., 2014).

In this study, there are two transfer functions utilised, which are the rectified linear function and softmax function, implemented for the hidden layers and output layer, respectively. The function is expressed as (Ketkar and Santana, 2017)

$$h(a) = \begin{cases} 0, & \text{if } a \le 0, \\ a, & \text{if } a > 0, \end{cases} \tag{2}$$

where $h(a)$ represents the transfer function, and a as the input values (bias values and the input weights). The output values from the classification model are usually un-normalised score values pertaining to particular classes. Therefore, N real values taken from the preceding layer are normalised through the softmax function into N real values that add up to 1. The normalisation results are values between 0 and 1, which are defined as the probability scores. This probability score is directly proportional to its attributed value from the deep neural network model output layer and remains between 0 and 1. This process is ideal for multi-classification models where the classes are mutually exclusive. The softmax function is expressed as (Ketkar and Santana, 2017)

$$\sigma(\vec{y}_i) = \frac{e^{x_i}}{\sum_{i=1}^{N} e^{x_i}}, \tag{3}$$

where $\vec{y}$ denotes the input vector $(y_1, \ldots, y_N)$ of the softmax function. The previous layers' input values are denoted by x_i. The equation further incorporates an exponential term e^{x_i}. Input values less than one are excluded by the exponential. Moreover, the normalisation term that governs the probability distribution is given by the output vector summation at the denominator in Eq. (3).

2.2. The ANN model training, testing, and evaluation

The ANN model characteristics and parameters are adjusted suitably for the building EPC classification. Moreover, the historical dataset is split into three groups to minimise over-fitting, namely a training set, a validation set, and a test set. The training and validation sets are given a 80/20 percentage split, respectively. This process is performed after the initial 20% split of the whole dataset to establish the test data. The test data are not used in the training process as this group is used to determine the practical performance of the ANN model. Performance evaluations of the model are executed through the test data. They take the form of accuracy measures that include but are not limited to the overall accuracy, probability of detection, and precision. XAI tools and the confusion matrix are other procedures and processes used to further evaluate the performance of the building EPC ANN classification model.

The performance indicators such as the model accuracy, sensitivity and precision defined in equations Eqs. (4)–(6) are adopted to facilitate the model training. The model accuracy in Eq. (4) represents the total test data samples correctly classified as a percentage of the total

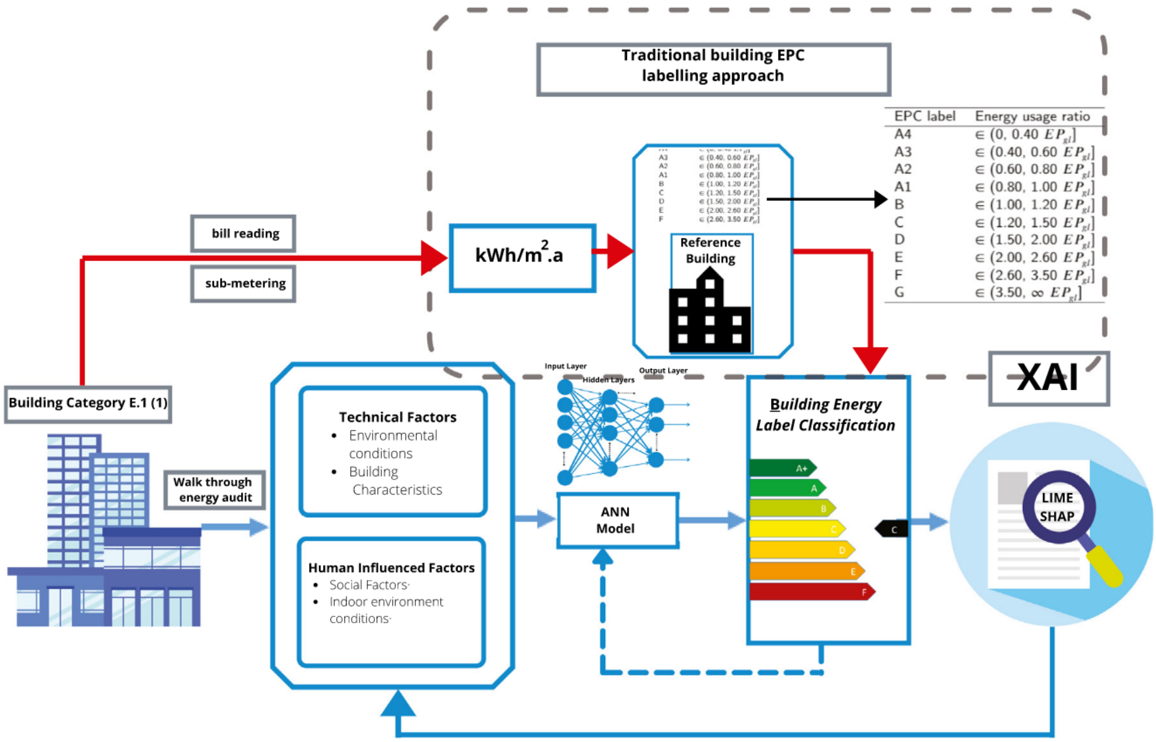

Fig. 1. Building EPC ANN classification and explainable artificial intelligence.

test data samples supplied to the ANN model for the building EPC classification.

$$Accuracy = \frac{CP + CN}{CP + CN + WP + WN}, \tag{4}$$

where CP denotes the correct positive classifications, CN denotes the correct negative classifications, WP denotes the wrong positive classifications, and WN denotes the wrong negative classifications. The sensitivity in Eqs. (5) is the fraction of the total correct positive classifications against the total of correct positive and wrong negative classifications by the ANN model. Whereas, the precision in Eqs. (6) is the ratio between the correct positive classifications against the total of correct positive and wrong positive classifications by the ANN model classification.

$$Sensitivity = \frac{CP}{CP + WN}, \tag{5}$$

and

$$Precision = \frac{CP}{CP + WP}. \tag{6}$$

2.3. Explanation of the building EPC ANN classification model

The building EPC ANN classification models are "blackbox" machine learning models. In order to demonstrate their internal classification mechanisms and increase public acceptance of the model, this study deploys the XAI tools to explain the underlying model and the influence of each input feature on the classification outcomes. LIME and SHAP are popular XAI tools used for model interpretation and making machine learning models comprehensible, which employ a simplified explanation model that is regarded as the interpretable approximation of the obtained building EPC ANN classification model. The LIME and SHAP are considered additive feature attribution methods, which takes the form represented in Eq. (7) (Lundberg and Lee, 2017).

$$g(z') = \phi_0 + \sum_{j=1}^{M} \phi_j z'_j, \tag{7}$$

where $z' \in \{0,1\}^M$, M is the number of simplified input features, $g(\cdot)$ denotes the explanation of the machine learning model. Each feature attribute, represented by j, is assigned a weight ϕ_j. These weights are summed to determine an approximation of f the original machine learning model. $z'_j \in \{0,1\}$, the coalition vector, represents the features presence or non presence in the summation, respectively.

2.3.1. Local interpretable model-agnostic explanation

Model agnostic interpretation is achieved through Local Interpretable Model-agnostic Explanations (LIME) (Molnar, 2022). This procedure produces an approximation model (Ribeiro et al., 2016). Through the locally executed procedure a globally understood model is achieved (Kuzlu et al., 2020). The LIME aims to minimise the following objective function: (Ribeiro et al., 2016):

$$\xi(z) = \arg\min_{g \in G} L(f, g, \pi_z) + \Omega(g), \tag{8}$$

where $g \in G$ is the explanation model, G represents a variety of interpretable models namely linear models, rule lists, and decision trees. For the black-box building EPC ANN classification models represented by f, the probability that z belongs to a specific class is given by $f(z)$. The establishment of the locality around z is realised through π_z, the proximity measure. The minimisation of the loss function $L(f, g, \pi_z)$ (Ribeiro et al., 2016) is the operation prescribed in an attempt to establish the performance of $g(\cdot)$ in Eq. (7), in explaining f. Lastly, with the derived $g \in G$ model not always being simple to explain, $\Omega(g)$ a complexity factor is introduced. In order to allow for interpretations by humans, this factor is kept low. Examples include non-zero weights in linear models and tree depth in decision trees. Through LIME, locally produced models estimate f globally (Adadi and Berrada, 2018; Kuzlu et al., 2020).

2.3.2. SHapley Additive exPlanation

SHapley Additive exPlanation (SHAP) is a technique with the ability to interpret and determine feature importance scores from black-box models. In this algorithm, SHAP values ϕ_j allocated to the jth feature are derived through the support of game theory. This value provides a score that determines the effect the feature has in the classification process of the machine learning models (Kuzlu et al., 2020; Bi et al., 2020).

The SHAP values are determined through the equation (Lundberg and Lee, 2017)

$$\phi_j = \sum_{S \subseteq Z \setminus \{j\}} \frac{|S|!(Z - |S| - 1)!}{Z!} [f_{S \cup \{j\}}(x_{S \cup \{j\}}) - f_S(x_S)], \tag{9}$$

where Z is a set of the input features, S is a subset representation of the features extracted from Z. The building EPC ANN classification model is trained on all the feature subsets and the importance of each feature is evaluated to determine its effect on the classification. The analysis results in a value being allocated to each feature. In order to realise this effect, two models are analysed, one with the feature present and one without the feature present, which are represented by $f_{S \cup \{j\}}(x_{S \cup \{j\}})$ and $f_S(x_S)$, respectively (Lundberg and Lee, 2017). x_S is set S's input features' actual values.

2.4. Applications of the explained ANN model

The SHAP and LIME algorithms are performed to explain the building EPC ANN classification model. The explanation outcomes can be taken as guidance to eliminate redundant and negatively influencing features during the classification. In addition, the XAI algorithms help to identify the key determinant features of the building EPC labels. Such explanations consequently offer guidance for building EPC label level improvements. For instance, if the XAI explains that the thickness of the building window is one of the determinants of the building EPC's label rating level, then it is reasonable for building owners to modify the thickness of the building window as a prior building energy efficiency solution.

3. Case studies

In this section, two case studies are presented to demonstrate the effectiveness of the proposed building EPC ANN classification models and AI explanations of the ANN models. The neural networks are constructed using Python 3.7. Feature importance calculations are performed through the XAI algorithms that explain the reasons for a particular building classification. The SHAP and LIME explanations are used to identify the key features while recognising features to be excluded that portray little or insignificant influences on the classification.

3.1. Case study 1: Italian EPC labelling pre-2015

In the first case study, buildings are classified into eight distinct classifications, namely A+, A, B, C, D, E, F and G using the CENED software (Khayatian et al., 2016), which is capable of calculating the required energy for heating, cooling, lighting, and domestic hot water independently. The EPC energy label is exclusively decided by the heating energy consumption EP_H. The criteria for assigning an EPC label differ between residential and non-residential structures, as well as their corresponding climate zones. The EP_H values are the key determinants of the EPC label in a certain climate zone as depicted in Table 1. This concern of only using EP_H for labelling purposes has been resolved in a current upgrade to the CENED software, which takes into account all of the aforementioned energy loads when calculating the energy performance index. In the second case study, buildings are therefore classified into ten distinct groups, namely A4, A3, A2, A1, B, C, D, E, F, and G, determined by the current CENED tool (Anon, 0000).

3.1.1. Data description and processing

The open source CENED database that contains numerous buildings' energy information in the Italian Lombardy region is utilised to train the building EPC ANN classification model. For the first case study, the building EPC labelling is mainly dependent on EP_H, which is the heating energy consumption of the building as depicted in Table 1. The input parameters for the classification process of building EPC labels is described in Fig. 1. In Fig. 1, residential buildings are the biggest group, which are labelled with the code E.1(1) on the open source CENED database. As part of the filters identified in Table 2, zero entries in particular are removed. In order to acquire a reliable neural network, defective data must be removed as the errors experienced in the weights are transferred from one layer to the next resulting in a poorly performing model (Khayatian et al., 2016). The input features go through various filters and various feature engineering techniques that are adopted from studies on the same dataset, e.g. removal of zero inputs of thermal conductivity data and label encoding categorical data (Dall'O' et al., 2015; Khayatian et al., 2016).

Table 1
Annual kWh/m^2 for EPC labelling of Italy's Lombardy region pre-2015 (Anon, 0000).

EPC label	kWh/m^2
A+	$\in$ (0, 14)
A	$\in$ [14, 29)
B	$\in$ [29, 58)
C	$\in$ [58, 87)
D	$\in$ [87, 116)
E	$\in$ [116, 145)
F	$\in$ [145, 175)
G	$\in$ [175, ∞)

Table 2
Data processing filters from the CENED database (Khayatian et al., 2016; Buratti et al., 2014).

Parameter	Filter	Unit
Net floor area	(50, ∞)	m^2
Glass opaque surface ratio	(0, 0.9]	–
Net volume	(130, ∞)	m^3
U-value of walls, roof and floor	(0.15, 4)	W/m^{2}K
U-value of windows	(0.8, 6)	W/m^{2}K
Glazed surface	(1, ∞)	m^2

The input data used to train the building EPC ANN classification model can be classified into three categories, namely the building characteristics, environmental conditions, and the social factors. The building characteristics refer to the building envelope and building appliances. The environmental conditions include both the indoor and outdoor climatic conditions, and the social factors relate to occupant behaviours, building operations and maintenance. The influence of each input parameter on the model is established and is known as a causal strength. The input feature values are weighted to give different output classes and together with the determined causal strengths, which enable the development of a more reliable ANN model that is able to classify their EPC label levels (Buratti et al., 2014). Additionally to these three categories in South Africa, for instance, we could classify the building characteristic input features into the POET categories (Xia and Zhang, 2010). The POET framework is a well-established grouping strategy to capture the key driven factors of energy systems.

Fig. 2 shows the distribution of the recorded building EPC labels, where 'G' rated buildings occupy 52.03%. 'A+' and 'A' labelled buildings occupy 0.6% and 3.3% respectively in case study 1. The filters and data cleaning concepts applied are extracted from studies performed on Italian buildings (Dall'O' et al., 2015; Khayatian et al., 2016), which results in 254 128 buildings with 129 682 'G' rated buildings.

3.1.2. The building EPC ANN classification model training and analysis

The ANN model is initially trained by fourteen inputs depicted in Table 3. These input features are numerical features except for the "FT". The obtained model exhibits an accuracy of 93.10%, which is shown by Fig. 3. The optimum time as well as epochs are established in order to attain the highest possible accuracy efficiently by stopping the training after no improvement for five epochs.

A summary of the model's performance is given by the confusion matrix. In Fig. 4, the confusion matrix shows a matrix of certified building EPC labels against the classified building EPC labels. The overall

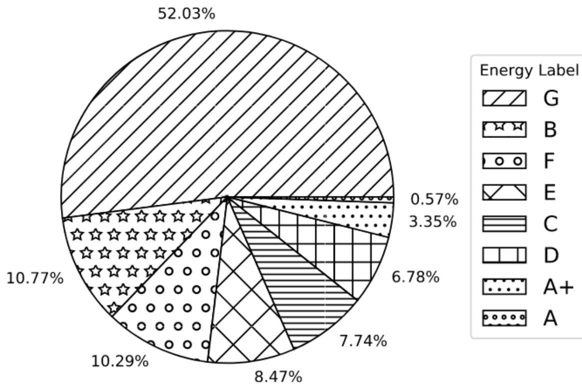

Fig. 2. Building EPC label distributions in the source data.

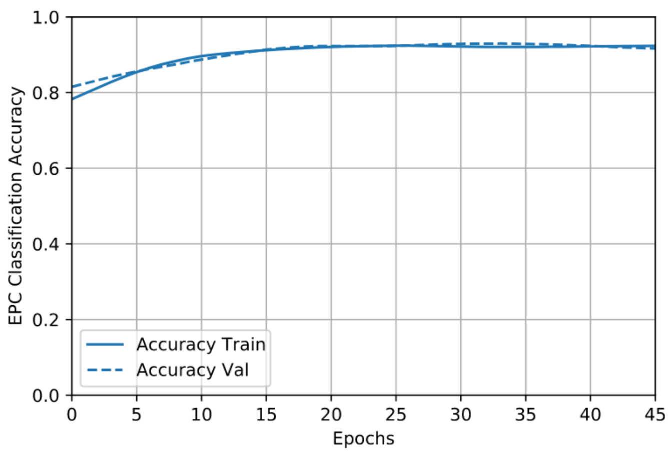

Fig. 3. EPC classification accuracy during the ANN model training for Italy's Lombardy region pre-2015.

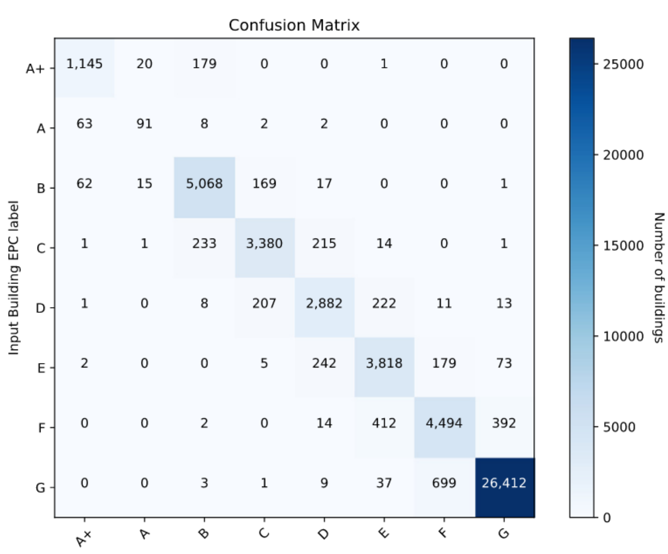

Fig. 4. A confusion matrix: input building EPC label vs. classified EPC labels (Case study 1).

accuracy achieved from the building EPC ANN classification models on the test dataset is 93.10%. The diagonal of the confusion matrix represents correctly classified building EPC labels that are determined as true by the building EPC ANN classification model. The acquisition of

Table 3
The building EPC ANN classification modelling characteristics.

Characteristic	Case study 1	Case study 2
EPC input features	14	26
First layer neurons	64	64
Second layer neurons	128	128
Output layer nodes	8	10
Test data ratio	0.2	0.2
Validation data ratio	0.16	0.16
Training data ratio	0.64	0.64
Epochs	45	144

the performance of the classification of each class is illustrated through Eqs. (4)–(5). These results are further summarised in Table 4. The buildings labelled with an 'A' EPC label occupy 0.33%. This low representation in the test and training data-sets results in the lowest values of precision (0.4270) and probability of detection (0.6867). Buildings labelled 'G' present the highest probability of detection (0.9606) and precision (0.9911) values as they constitute 53.43% of the dataset.

Table 4 addresses the performance overview numerically of the building EPC classification ANN model. Through it, the unequal distribution of various building EPC energy labels and their effect on performance is apparent as depicted in Fig. 2.

3.1.3. LIME XAI

The local explainability is achieved through the LIME algorithm. Its application in the building EPC programme results in feature importance identifications and probabilities of the classified classes. Figs. 5–6 illustrate the findings from the LIME algorithm. Fig. 5 presents a list of explanations that reflect the key features contributing to the EPC 'G' classification, consisting of three parts: the prediction probabilities, the feature probabilities, and the feature value table. The prediction probabilities graph shows the building EPC ANN model's classification decision, which clearly indicates a probability of 1, labelled 'G' and represented by the grey bar, to classify this building to an EPC label 'G'. The feature probabilities graph shares more details about how each of the 14 input features contribute to the given classification decision. In this case, detailed feature probabilities reveal informative details regarding what input feature supports the classification to the 'F' and 'G' EPC label, together with the details about what input feature contradicts classification to the 'F' or 'G' EPC label. The feature value graph presents the actual value that one feature possesses in this observation, and also sorts the features by importance. The top 13 input features are illustrated in the feature value table. The feature value table helps the interpretations of the results in the feature probabilities graph. For instance, in this explanation, the "CO_2" value is 111.02 kg/m^2a, which is greater than 60.37 kg/m^2a. The feature "$CO_2 > 60.37$" supports the classifications to an EPC label of 'G' and also 'NOT F'.

LIME also gives numeric thresholds in inequalities with a clear ranking of the feature importance as shown in Fig. 6, where one can observe that features with the highest impact on the classification are the "CO_2" and the "GHE", and the "AU_b" has the lowest impact. The weight values depicted in the figure are derived through the determination of the distances of the data points, that are generated through perturbation, in the neighbourhood of the original data point. These perturbed values are kept within a predetermined proximity which keeps the model locally faithful. These LIME explanations offer essential guidance on the possible solutions to improve the EPC label ratings. For instance, Fig. 6 illustrates that the "CO_2" feature can be reduced to an input value lower than 60.37 kg/m^2a, which may result in the building's EPC label to improve from 'G' to a better rating of 'F' or 'E'. When the input feature values are altered in the manner described in Table 5, it results in an upgrade of the building EPC label 'G' TO 'E' when the "AU_w" is changed to 0.3 and the "CO_2" is changed to 20 kg/m^2a.

Table 4
ANN model classification performance.

EPC label	Probability of detection	Precision	Number of input testing labels	% of input testing labels
A+	0.7346	0.8316	1345	2.64%
A	0.6867	0.4270	166	0.33%
B	0.9467	0.8845	5332	10.49%
C	0.7982	0.9388	3845	7.57%
D	0.9088	0.7731	3344	6.58%
E	0.8328	0.7483	4319	8.50%
F	0.8026	0.8000	5314	10.46%
G	0.9606	0.9911	27161	53.43%
Weighted average precision		**0.9149**		

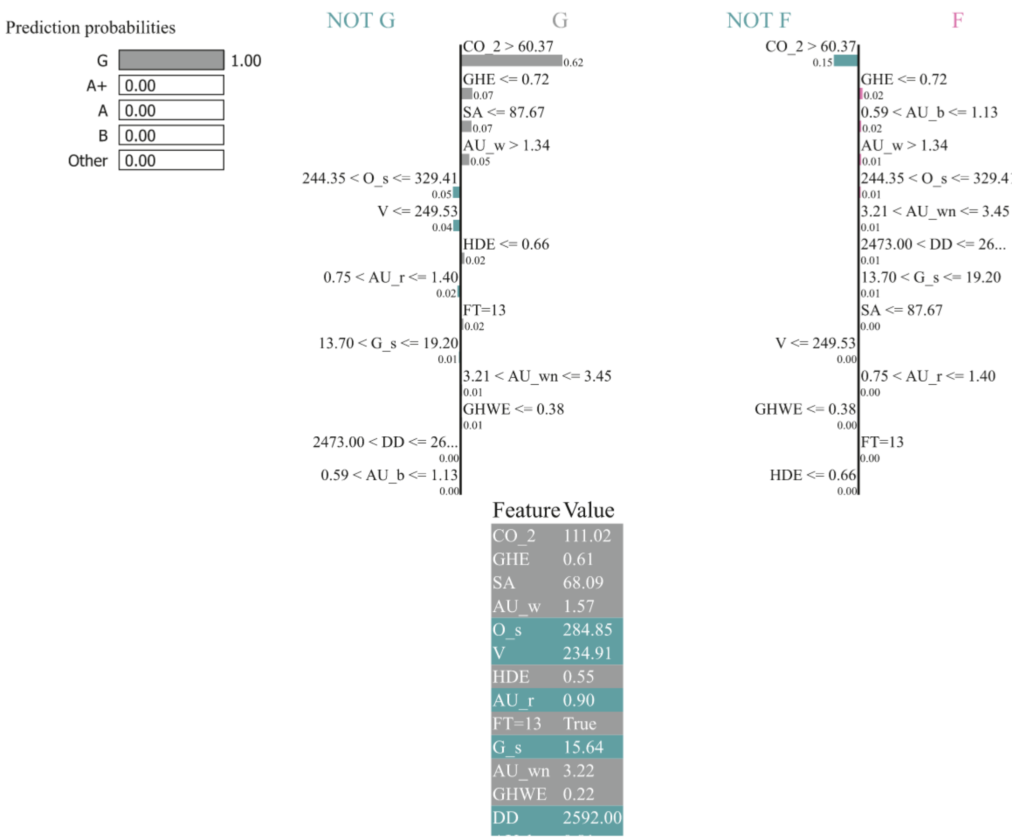

Fig. 5. LIME explanations for a building EPC label ‘G’ classification.

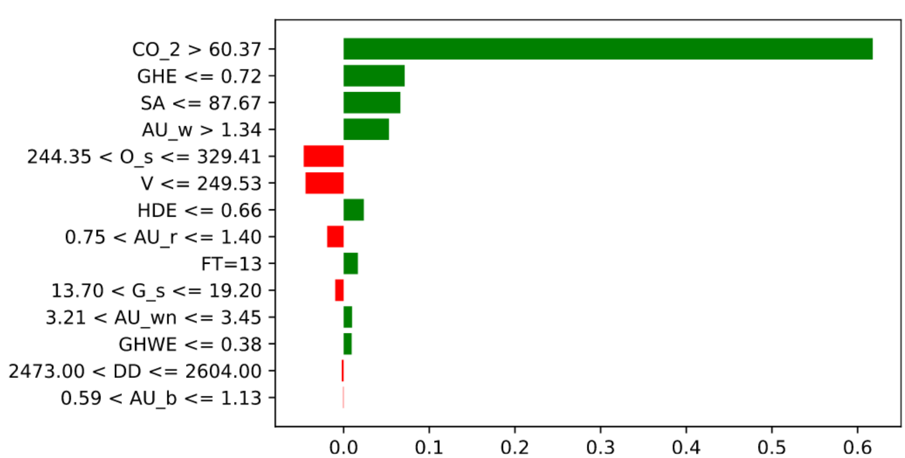

Fig. 6. Input feature importance ranking via LIME explanations for a building EPC ‘G’ classification.

Table 5
EPC classification performance with altered input features.

Input feature	EPC label G	Target EPC label E	Unit
DD	2592	2592	°C
SA	68.09	68.09	m^2
V	234.91	234.91	m^3
O_s	284.85	284.85	m^2
G_s	15.64	15.64	m^2
AU_w	**1.57**	**0.3**	W/m^2K
AU_b	0.81	0.81	W/m^2K
AU_r	0.90	0.90	W/m^2K
AU_{wn}	3.22	3.22	W/m^2K
CO_2	**111.02**	**20**	kg/m^2a
GHE	0.61	0.61	–
GHWE	0.22	0.22	–
FT	13	13	–
HDE	0.55	0.55	–

Table 6
Model performance with different number of input features.

Model Id	Input features	Feature names	Accuracy
1	14	AU_b, AU_r, AU_w, AU_{wn}, DD, FT,GHE, GHWE, HDE, O_s, G_s, V, SA, CO_2	93.10%
2	13	AU_b, AU_r, AU_w, AU_{wn}, DD, FT,GHE, GHWE, HDE, O_s, G_s, V, SA	76.10%
3	12	AU_b, AU_r, AU_w, AU_{wn}, DD, FT,GHE,GHWE, O_s, G_s, V, SA ,	75.92%
4	11	AU_b, AU_r, AU_w, AU_{wn}, DD, FT,GHE, O_s, G_s, V, SA	75.37%

3.1.4. SHAP XAI

When SHAP is applied to explain the building EPC ANN classification model in the case study 1, it yields the global influence of each feature across all classes and buildings in the dataset. The average magnitude of SHAP values across the dataset is derived to give the feature importance. Fig. 7 displays and ranks the overall feature importance from the highest to the lowest. The most influential feature across all building EPC label classifications is the "CO_2", with a SHAP value of 0.62, followed by the "SA". The "HDE" and the "GHWE" are features with the least effect on the model classifications. Removal of these two least influential features according to the SHAP explanation does not result in a drastic change in the model accuracy. Table 6 illustrates as to what extent the removal of other features has on the overall accuracy of the model. Model 1 represents the initial input features and these are depicted on the vertical axis of Fig. 7. Further investigation shows that with the exclusion of the feature "CO_2" as the input feature, the only input feature measured over a yearly period, the training process yields a model with an overall accuracy of 76.10%. SHAP is also able to perform local interpretability of the ANN classification model. Local explainability of one specific building to be classed into an EPC label 'G' through SHAP is summarised in Table 7. The classification probability of the building EPC as 'G' is 1.00. The most influential input feature supporting this classification is the "CO_2" while the most influential input feature against this classification is the "DD". In addition, the input feature of the "CO_2" is strongly against this building to be classified into any other EPC labels other than the EPC label 'G'. The observations from Table 7 is also supported by the absolute SHAP values in Fig. 8. In Fig. 8, the bar values in red indicate those input features who are positively contributing to the classification to the EPC label 'G', while the bar values in blue symbolise the input features who are against the building to be classified into a label 'G'. The magnitude of bar values indicates the strength of the contribution to each classification.

3.1.5. Comments on the XAI outcomes

The SHAP is able to perform both local and global explanations while the LIME only offers local explanations. The two XAI tools show that the most influential feature of the building EPC classification is the "CO_2", whether explained locally or globally. The LIME local explanations show that the "GHE" and the "SA" are the next most influential parameters, with the "AU_b" being the least influential. For the SHAP XAI, the "SA" and the "V" are the next most influential, with the "AU_b" being the least influential. The local explanations only pertain to a particular classification being explained locally and the explanations may not always hold true globally. The LIME explanations, with its revealed threshold given by the inequalities, allow one to determine the quantifiable adjustments to enact the building EPC label improvement while the SHAP only explains the magnitude of the feature impact through SHAP values. The SHAP offers explanations that determine feature importance globally hence enabling to avoid the features of no influence for the building EPC ANN classification.

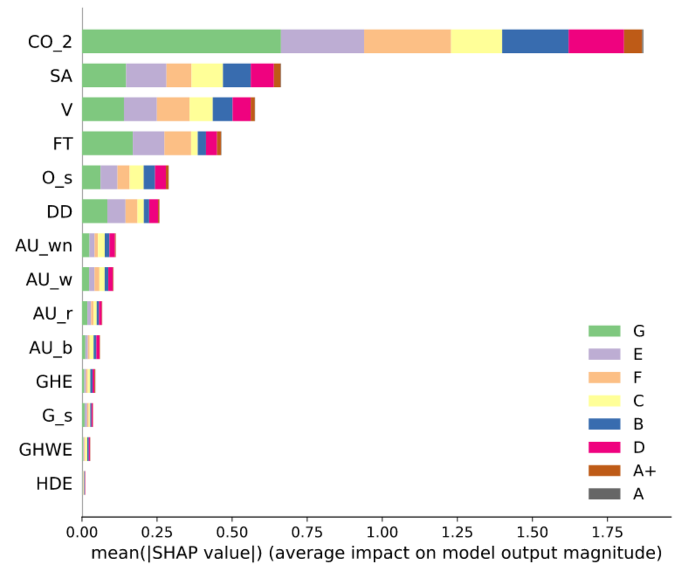

Fig. 7. Input feature importance ranking via SHAP global explanations for building EPC classification.

Table 7
SHAP classification explanations for a building EPC label G.

EPC label	Classification probability	Top supporting feature	Top contradicting feature
A+	0.00	FT	CO_2
A	0.00	FT	CO_2
B	0.00	DD	CO_2
C	0.00	DD	CO_2
D	0.00	DD	CO_2
E	0.00	DD	CO_2
F	0.00	DD	CO_2
G	1.00	CO_2	DD

3.2. Case study 2: Italian EPC labelling post-2015

3.2.1. Data description and processing

The procedure applied in the building EPC labelling of case study 2 is similar to that of the South African building EPC programme. The general process to perform the EPC labelling includes two steps. Firstly, the global energy performance index of the non-renewable energy value EP_{gl} for the national reference building should be determined, see the Annex 1, Chapter 3 of the minimum requirements decree (Costanzo et al., 0000). Secondly, the property that is subject to the EPC evaluation must have its $\bar{EP}_{gl}$ calculated. The building EPC label is determined by comparing the ratios against the EP_{gl} value of the reference building as depicted in Table 8. The minimum expected

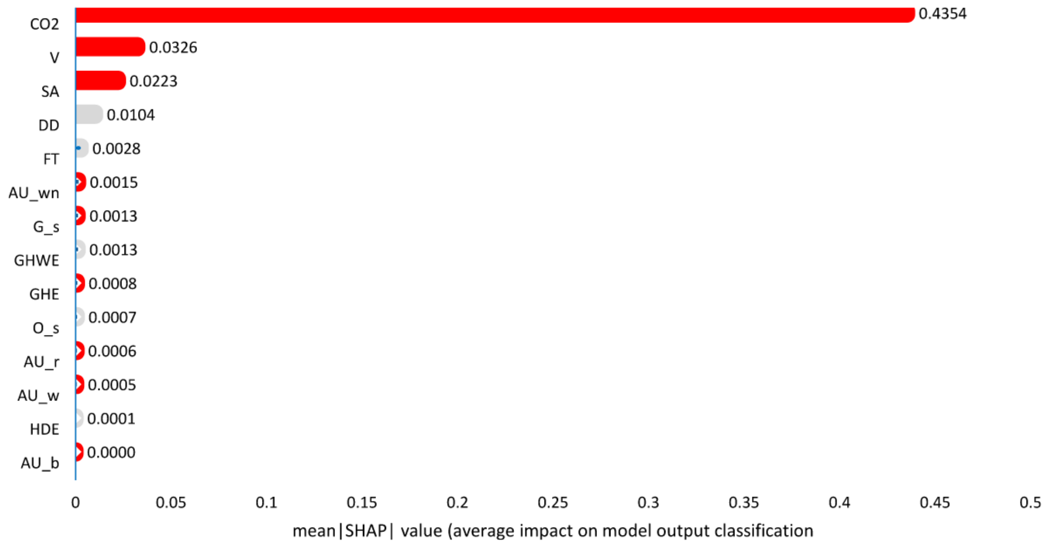

Fig. 8. Input feature importance ranking via SHAP explanations for a building EPC G classification.

Table 8
Annual energy usage ratio for different EPC labels in Italy's Lombardy region.

EPC label	Energy usage ratio
A4	$\in (0EP_{gl},\ 0.40EP_{gl}]$
A3	$\in (0.40EP_{gl},\ 0.60EP_{gl}]$
A2	$\in (0.60EP_{gl},\ 0.80EP_{gl}]$
A1	$\in (0.80EP_{gl},\ 1.00EP_{gl}]$
B	$\in (1.00EP_{gl},\ 1.20EP_{gl}]$
C	$\in (1.20EP_{gl},\ 1.50EP_{gl}]$
D	$\in (1.50EP_{gl},\ 2.00EP_{gl}]$
E	$\in (2.00EP_{gl},\ 2.60EP_{gl}]$
F	$\in (2.60EP_{gl},\ 3.50EP_{gl}]$
G	$\in (3.50EP_{gl},\ \infty\ EP_{gl}]$

Table 9
Data processing filters for the CENED database.

Parameter	Filter	Unit
Net floor area	$(50,\ \infty)$	m^2
Net volume	$(130,\ \infty)$	m^3

building EPC label is B. The $\bar{EP}_{gl}$ calculation is depicted in Eq. (10),

$$\bar{EP}_{gl} = EP_H + EP_C + EP_W + EP_V, \tag{10}$$

where EP_H, EP_C, EP_W, and EP_V denote the energy intensities of heating, cooling, domestic, and hot water and ventilation, respectively in kWh/m^2a.

Residential buildings make up the largest category in case study 2 and are assigned the code E.1(1) in the CENED database. Zero entries in specific input features are eliminated as part of the filters described in Table 9. The input features are categorical features such as the "RC", "ST", "EV", and "SAC", and the other input features are all numerical features that are displayed in the vertical axis of Fig. 14.

Fig. 9 illustrates the percentage of buildings EPC label counts in the CENED database, where 'G' rated buildings occupy 35.94%, and EPC labels 'A1' - 'A4' account for 7.21% of the labelled buildings. The filters and data cleaning procedures applied are extracted from studies performed on Italian buildings (Dall'O' et al., 2015; Khayatian et al., 2016). As a consequence, there are 174 793 'G' rated buildings out of a total building entry of 486 347 in this case study.

Table 10
Input feature classification for training purposes.

Feature names	Feature type
DD, LBC, NGC, SBC, DC, SPVC, STC, DHC, SACM, SAPER, SAPNER, WANP, WACE, WAPEPR, WAPNER, SA, V, SV, WHT, SESA, TT, RC,	Numerical
RBC, ST, EV, SAC	Categorical

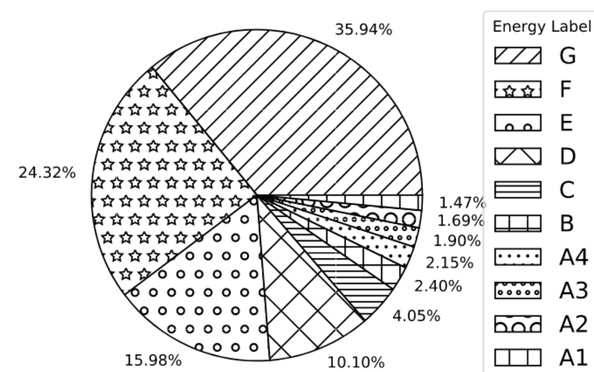

Fig. 9. Building EPC label distributions in the source data.

3.2.2. The building EPC ANN classification model training and analysis

The building EPC ANN classification model is trained with 26 input features as depicted in Table 10. For classification, the model exhibits an accuracy of 88.49% as shown by Fig. 10.

The XAI tools are used to evaluate the feature significance, to explain why a specific building's EPC label is derived during the classification. The XAI explanations provide technical guidance to exclude characteristics that have a minor impact on the classification.

Fig. 11 depicts a matrix of certified building EPC labels against the classified building EPC labels. The overall accuracy of the classification is 89.49%. The performance of the model on classifying each class is illustrated through Eqs. (4)–(5), with the results depicted in Table 11. The 'G' labelled buildings account for 34.36%, which shows an associated probability of detection of 0.9638 and precision of 0.9756. Buildings having EPC labels of 'A1-A4' account for 1.69%, 1.86%, 2.08% and 2.22% in the test dataset, respectively. They account for some of the lowest precision and probability of detection values due to their low representations in the dataset. Table 11 addresses

Table 11
ANN classification model performance.

EPC label	Probability of detection	Precision	Number of input testing labels	% of input testing labels
A4	0.5178	0.6279	2161	2.22%
A3	0.5059	0.5295	2024	2.08%
A2	0.7117	0.5225	1814	1.86%
A1	0.6636	0.8136	1644	1.69%
B	0.5759	0.7416	2377	2.44%
C	0.8086	0.6811	3903	4.01%
D	0.8646	0.8451	9972	10.25%
E	0.8643	0.9058	15958	16.40%
F	0.9335	0.9026	23997	24.67%
G	0.9638	0.9756	33419	34.36%
Weighted average precision		**0.8869**		

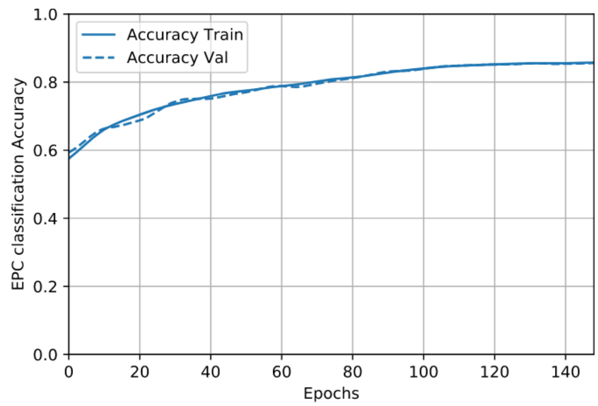

Fig. 10. EPC classification accuracy during the ANN model training for Italy's Lombardy region post-2015.

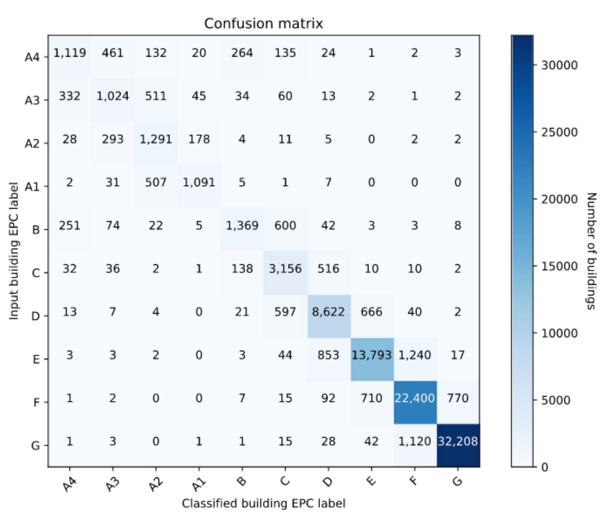

Fig. 11. A confusion matrix: input building EPC label vs. classified EPC labels (Case study 2).

the performance overview numerically of the building EPC labelling classification. The uneven distribution of various building EPC labels is shown in Fig. 9. The confusion matrix encapsulates the performance of the EPC labelling classification.

3.2.3. LIME XAI

In Fig. 12, LIME explanations for a building EPC label 'E' classification are presented. The prediction probabilities graph indicates that the building EPC label has a probability of 0.84 to be an EPC label 'E', represented by the orange bar, while a probability of 0.16 to be an EPC label 'F', represented by the purple bar. The feature probabilities graph categorises the 26 input features in terms of the supporting features and contradicting features for the prediction outcomes, the EPC 'E' and 'F' labels, respectively. It is observed that the top feature against the classification 'E' is the "WACE". The top three input features contributing towards the 'E' classification are "DHC", "DC", and the "NGC". There are 13 selected feature values given in the feature value table. Fig. 13 illustrates the feature importance ranking for the given EPC label 'E' classification, where the "WACE" is the most influential feature for the classified building EPC label. To enact an improvement in the building EPC label, "WACE" may be adjusted to an input value greater than 0.79.

Table 12
SHAP classification explanations for a building EPC label E.

EPC label	Classification probability	Top supporting feature	Top contradicting feature
A4	0.00	NGC	SPVC
A3	0.00	SPVC	NGC
A2	0.00	SPVC	NGC
A1	0.00	NGC	SPVC
B	0.00	STC	NGC
C	0.00	WAEPR	RBC
D	0.00	RBC	WAEPR
E	0.84	WAEPR	RBC
F	0.16	SA	NGC
G	0.00	RBC	WAEPNR

3.2.4. SHAP XAI

Fig. 14 shows a global view of the most influential features during the building EPC labelling classifications. The "WAEPR" is the most influential, with a SHAP value of 0.75, followed by the "NGC". The "LBC" and "TT" are features with the least effect on the classifications. Removal of the 10 least influential features according to the SHAP explanation shown in Fig. 14 does not result in a drastic change in the model accuracy. SHAP XAI is also capable of revealing the input feature importance ranking for a local classification. A SHAP explation for a building EPC E classification is shown in Fig. 15. Model 1 in Table 13 represents the initial input features that are depicted on the vertical axis of Fig. 14. Table 13 further illustrates the ANN model accuracy performance with insignificant input features removed. It is observed that a model accuracy of 80.50% is achieved with only 10 essential inputs as advised by the XAI explanations. The Model 5 listed in Table 13 is trained by the input features of "RC", "RBC", "V", "WHT", "DHC", "SA", "DD", "WANP", "WAEPR", and "WAEPNR", which are mostly the technical specifications of building appliances in the winter season. This implies that information acquired from the winter season is adequate to estimate the building's EPC label for this case.

The SHAP classification probability for an 'E' labelled building is depicted in Table 12. The probability of classifying this building as an EPC label 'E' building is 0.84, which is strongly supported by the input feature "WAEPR" while contradicted by the input feature "RBC".

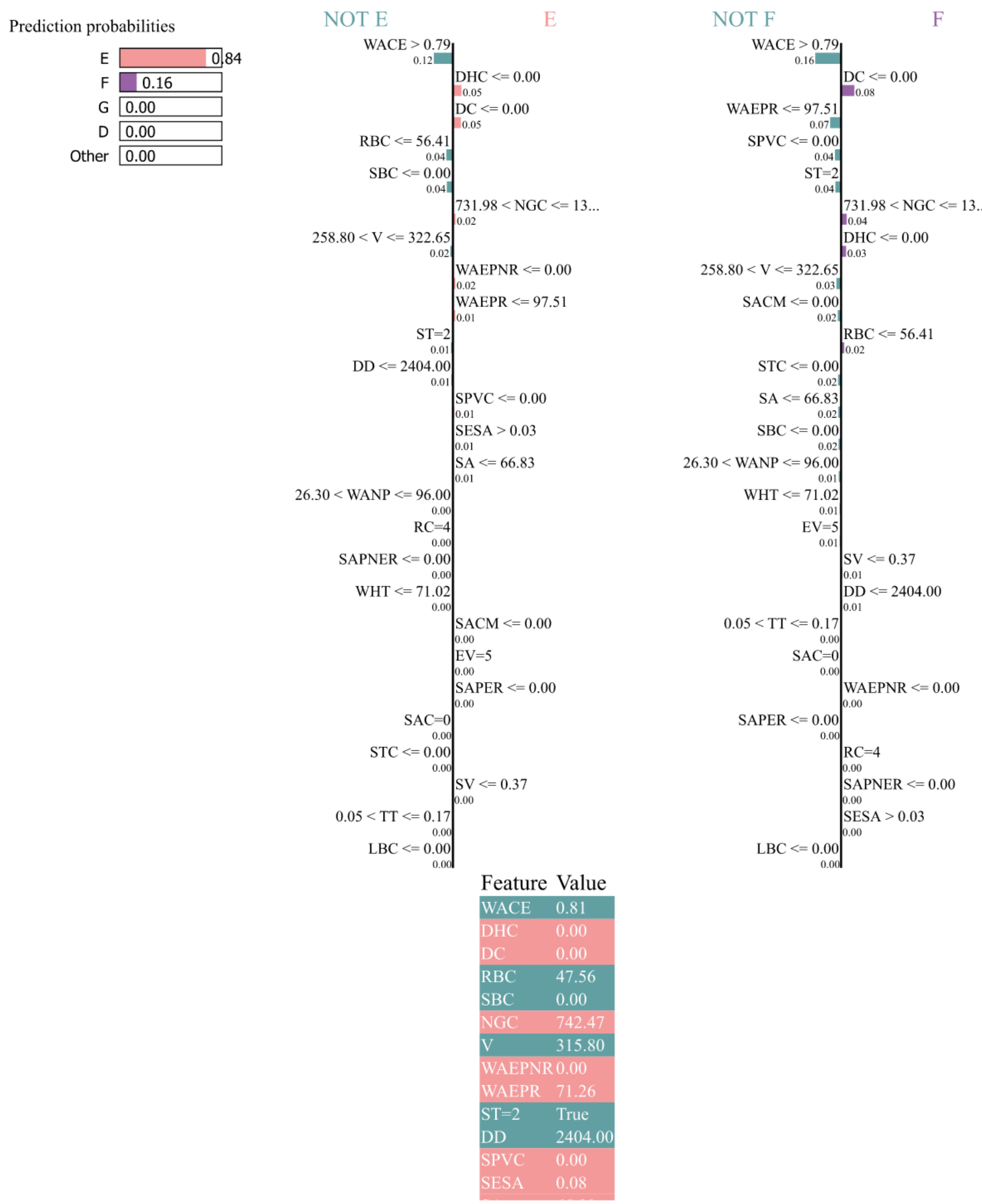

Fig. 12. LIME explanations for a building EPC 'E' label classification.

3.2.5. Comments on the XAI outcomes

In this case study, the LIME XAI tool reveals that the "WACE" is the most influential input feature for a specific building EPC classification. The SHAP explains that the "WAEPR" is the most influential input feature across the entire building EPC classification exercise in the database. The "DHC", "DC", "RBC", and "SBC" are the next most significant features, according to the LIME local explanations. For that particular building EPC label classification, the "TT" and "LBC" have the least impact. The "WHT" and "NGC" are the next most influential features explained by SHAP, with the "STC" being the least influential input feature. The LIME reveals the lower bound or upper bound of the evaluated input features in its explanation output, which guides one to conduct necessary modifications for the EPC energy label improvement, while the SHAP values are valuable indicators to point out the most impactful features to retrofit for the building EPC label improvement.

4. Discussions

This study provides alternative solutions to the building EPC labelling, by ANN classification and the XAI outcomes further advised the building EPC label improvement procedures incorporating both indoor building appliances and building envelope features.

From the two case studies, it is evident that the building EPC ANN classification models are able to classify the building EPC labels. The first case study mainly considers the building envelope features that are the thermal properties of buildings. There are particular energy efficiency measures available to improve these thermal resistivity values. This allows stakeholders to make preferable decisions to improve the building through an optimal and efficient manner. The second case study adds to this phenomenon through showing how one may limit the amount of features required for the building EPC labelling process. This effectively reduces the procedures, data and information required for the building EPC labelling exercise. The features in case study 2 are mainly the energy records of the indoor building appliances.

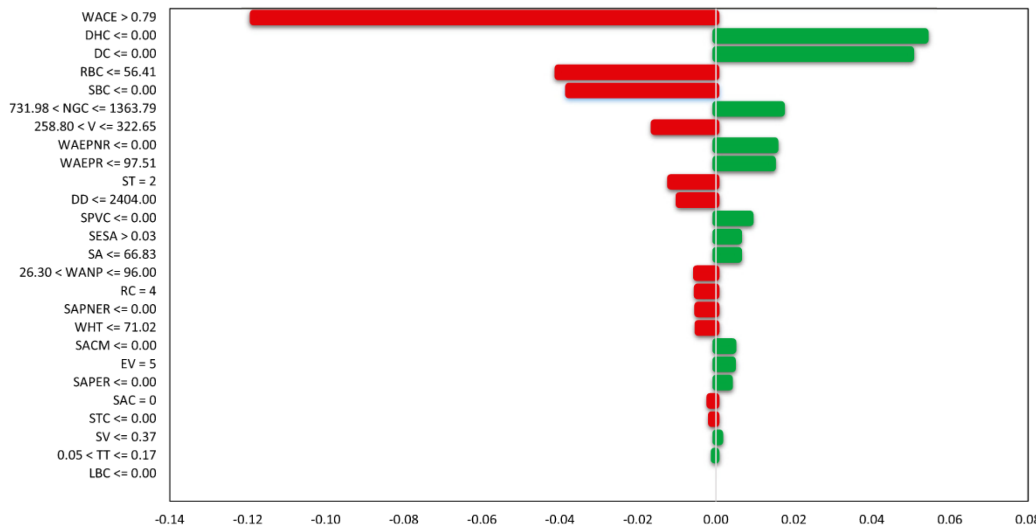

Fig. 13. Input feature importance ranking via LIME explanations for a building EPC E classification.

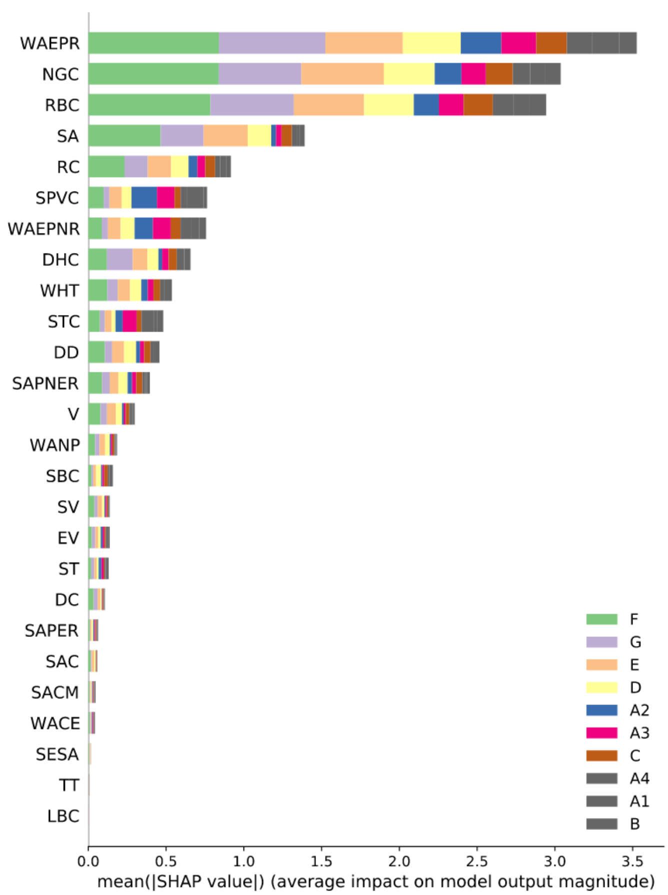

Fig. 14. Input feature importance ranking via SHAP global explanations for the building EPC classification.

With the technical support of the XAI explanations, the ANN model training process can be improved iteratively towards the simplified and minimalist input of 10 input features that can still maintain an overall model accuracy of 80.5%.

Clear evidence from the two case studies shows that a larger number of entries of the certified building EPC labels results in better trained ANN models, hence this leads to a higher probability of detection and

Table 13
Model performance with different numbers of features.

Model Id	Input features	Feature names	Accuracy
1	26	DD, LBC, NGC, SBC, DC, SPVC, STC, DHC, SACM, SAPER, SAPNER, WANP, WACE, WAPEPR, WAPNER, SA, V, SV, WHT, SESA, TT, ST, EV, SAC, RC, RBC	89.62%
2	20	DD, NGC, DC, SPVC, STC, DHC, SAPNER, WANP, WACE, WAEPR, WAEPNR, SA, V, SV, WHT, TT, ST, EV, RC, RBC	87.52%
3	14	DD, NGC, SPVC, STC, DHC, SAPNER, WANP, WAEPR, WAEPNR, SA, V, WHT,RC, RBC	87.48%
4	11	DD, NGC, DHC, WANP, WAEPR, WAEPNR, SA, V, WHT, RC, RBC	83.58%
5	10	DD, DHC, WANP, WAEPR, WAEPNR, SA, V, WHT, RC, RBC	80.50%

precision of the EPC ANN classification outcomes. A significant rise in Net Zero Energy Buildings (NZEB) is expected in the coming years due to the current regulations that enforce these NZEB guidelines being a reality in the European Union. As a result, we may expect more feature inputs of buildings at levels 'A' or 'B' to be available to calibrate the training of our proposed building EPC ANN classification models.

The XAI tools adopted in this study share insight of the inner working mechanism of the ANN models and reasons for the building EPC label classifications. The LIME explanations are local and prove helpful in determining the adjustable features in a quantifiable manner to improve a building EPC label. The SHAP explanations give global and local feature importance evaluations. The global feature importance identification allows insignificant features to be excluded from the ANN models. When the local explanations of SHAP are compared to those by LIME, the most influential features are the same. The proposed building EPC ANN classification models and the XAI tools are ideal and beneficial for the building EPC programmes. This study is able to inform the key decision makers to take quantifiable effective measures to improve the building EPC label. Unlike the traditional building EPC labelling process, the machine learning classification process renders a building EPC label as its direct output without estimating the numeric energy consumption, which might lead to doubts to the legitimacy of the building EPC label. The XAI explanations curtail these doubts.

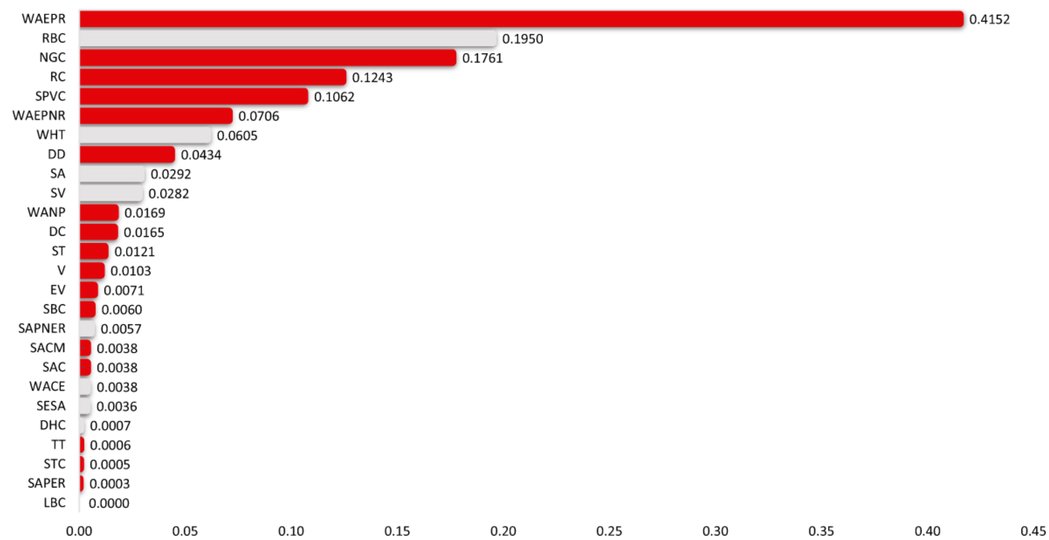

Fig. 15. Input feature importance ranking via SHAP explanations for a building EPC E classification.

5. Conclusion

This study proposes an artificial intelligence based classification method for the building EPC programme, which aims to acquire the building EPC through a more efficient and cost-effective process. The building EPC ANN classification model is capable of classifying the building EPC label by analysing input features of the technical and human related building energy driving factors. After the ANN model classification, two XAI tools help to identify the minimal input features that are most influential during the classifications. The minimum input feature identification, for the building EPC classification, helps reduce the data and information required during the EPC labelling process. In addition, the integral application of the building EPC ANN classification model and the XAI tools also reveal the quantifiable threshold of each input feature with respect to the numeric contribution to the final classification decision. This offers clear guidance for the EPC label improvement solutions, when parties involved are cognisant of which input feature can be modified, and by how much, such that a particular building EPC label can be improved to a better level.

The results from the case study 1 show that before feature removal the models accuracy was 93.10% at 14 input features, and 75.37% at 11 input features without yearly calculated data. The two most influential features are the “CO_2” and “SA”. For the case study 2, the ANN model classification accuracy is 89.62% with 26 input features, and 80.50% with 10 features acquired over the winter season. The most influential features are the “RBC" and “WACE”.

6. Future work and recommendations

This study shares our initial investigations of applying the ANN models to classify building EPC labels, using LIME and SHAP XAI tools to explain the building EPC ANN classification models. Our ongoing research focus is in the following areas: (1) develop new artificial intelligence models for the building EPC classification, and try to optimise the model classification accuracy with minimum input features through explainable AI techniques; (2) calibrate the existing model by using the local data from the South African building EPC programme, with the aid of the existing and newly developed XAI tools; and (3) based on the revealed building features and determinants of the existing, we will try to prioritise the optimal building EPC label improvement solutions.

CRediT authorship contribution statement

Thamsanqa Tsoka: Writing – original draft, Software, Data curation, Investigation. **Xianming Ye:** Writing – review & editing, Supervision, Funding acquisition, Methodology, Validation. **YangQuan Chen:** Conceptualisation, Formal analysis. **Dunwei Gong:** Writing – review & editing, Validation. **Xiaohua Xia:** Writing – review & editing, Supervision, Funding acquisition, Validation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research is supported by the National Key R&D Program of China (grant No. 2021YFE0199000), National Natural Science Foundation of China (Grant no.: 61803162), the National Research Foundation Competitive Support for Unrated Researchers (CSUR) programme with grant no.: 116309, and the Royal Academy of Engineering Transforming Systems through Partnership grant scheme with reference no.: TSP1020.

References

Adadi, A., Berrada, M., 2018. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE Access 6, 52,138–52,160.

Al-Azzam, N., Shatnawi, I., 2021. Comparing supervised and semi-supervised machine learning models on diagnosing breast cancer. Ann. Med. Surg. 62, 53–64.

Alicioglu, G., Sun, B., 2021. A survey of visual analytics for explainable artificial intelligence methods. Comput. Graph..

Amasyali, K., El-Gohary, N.M., 2018. A review of data-driven building energy consumption prediction studies. Renew. Sustain. Energy Rev. 81, 1192–1205.

Amasyali, K., El-Gohary, N., 2021. Machine learning for occupant-behavior-sensitive cooling energy consumption prediction in office buildings. Renew. Sustain. Energy Rev. 142, 110714.

Anon, 0000. A. regionale per l'ionnovazione e gli acquisti, EPC model based on dgr x / 3868. [Online]. Available: https://www.cened.it/modelli-ape.

Arrieta, A.B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., et al., 2020. Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Inf. Fusion 58, 82–115.

Ascione, F., Bianco, N., De Stasio, C., Mauro, G.M., Vanoli, G.P., 2017. Artificial neural networks to predict energy performance and retrofit scenarios for any member of a building category: A novel approach. Energy 118, 999–1017.

Bagheri, F., Mokarizadeh, V., Jabbar, M., 2013. Developing energy performance label for office buildings in Iran. Energy Build. 61, 116–124.
Belle, V., Papantonis, I., 2021. Principles and practice of explainable machine learning. Front. Big Data.
Bi, Y., Xiang, D., Ge, Z., Li, F., Jia, C., Song, J., 2020. An interpretable prediction model for identifying N7-methylguanosine sites based on XGBoost and SHAP. Mol. Therapy-Nucleic Acids 22, 362–372.
Buratti, C., Barbanera, M., Palladino, D., 2014. An original tool for checking energy performance and certification of buildings by means of artificial neural networks. Appl. Energy 120, 125–132.
Costanzo, E., Martino, A., Varalda, G., Antinucci, M., Federici, A., 0000. EPBD implementation in Italy. [Online]. Available: https://www.epbd-ca.eu/wp-content/uploads/2018/08/CA-EPBD-IV-Italy-2018.pdf.
Dall'O', G., Sarto, L., Sanna, N., Tonetti, V., Ventura, M., 2015. On the use of an energy certification database to create indicators for energy planning purposes: Application in northern Italy. Energy Policy 85, 207–217.
Dayhoff, J.E., 1990. Neural Network Architectures: An Introduction. Van Nostrand Reinhold Co..
Dong, Z., Liu, J., Liu, B., Li, K., Li, X., 2021. Hourly energy consumption prediction of an office building based on ensemble learning and energy consumption pattern classification. Energy Build. 241, 110929.
Fan, Y., Xia, X., 2018a. Building retrofit optimization models using notch test data considering energy performance certificate compliance. Appl. Energy 228, 2140–2152.
Fan, Y., Xia, X., 2018b. Energy-efficiency building retrofit planning for green building compliance. Build. Environ. 136, 312–321.
Fathi, S., Srinivasan, R., Fenner, A., Fathi, S., 2020. Machine learning applications in urban building energy performance forecasting: A systematic review. Renew. Sustain. Energy Rev. 133, 110287.
Goodman, B., Flaxman, S., 2017. European union regulations on algorithmic decision-making and a right to explanation. AI Mag. 38 (3), 50–57.
Guo, Y., Wang, J., Chen, H., Li, G., Liu, J., Xu, C., Huang, R., Huang, Y., 2018. Machine learning-based thermal response time ahead energy demand prediction for building heating systems. Appl. Energy 221, 16–27.
Ketkar, N., Santana, E., 2017. Deep Learning with Python. Springer, p. 1.
Khayatian, F., Sarto, L., Dall'O', G., 2016. Application of neural networks for evaluating energy performance certificates of residential buildings. Energy Build. 125, 45–54.
Kim, P., 2017. Matlab deep learning, with machine learning. Neural Netw. Artif. Intell. 130, 21.
Kingma, D., Ba, J., 2015. Adam: A method for stochastic optimization. In: Proceedings of the 3rd International Conference for Learning Representations. San Diego.
Kuzlu, M., Cali, U., Sharma, V., Güler, Ö., 2020. Gaining insight into solar photovoltaic power generation forecasting utilizing explainable artificial intelligence tools. IEEE Access 8, 187,814–187,823.
Lu, H., Cheng, F., Ma, X., Hu, G., 2020. Short-term prediction of building energy consumption employing an improved extreme gradient boosting model: a case study of an intake tower. Energy 203, 117756.
Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. Adv. Neural Inf. Process. Syst. 30.
Miller, T., 2019. Explanation in artificial intelligence: Insights from the social sciences. Artificial Intelligence 267, 1–38.
Molnar, C., 2022. Interpretable machine learning: A guide for making black box models explainable. Lulu. com.
Olu-Ajayi, R., Alaka, H., Sulaimon, I., Sunmola, F., Ajayi, S., 2022. Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques. J. Build. Eng. 45, 103406.
Parker, S.A., Hunt, W., McMordie Stoughton, K., Boyd, B.K., Fowler, K.M., Koehler, T.M., Sandusky, W.F., Sullivan, G.P., Pugh, R., 2015. Metering Best Practices: A Guide To Achieving Utility Resource Efficiency, Release 3.0. Tech. Rep., Pacific Northwest National Lab.(PNNL), Richland, WA (United States).
Pham, A.-D., Ngo, N.-T., Truong, T.T.H., Huynh, N.-T., Truong, N.-S., 2020. Predicting energy consumption in multiple buildings using machine learning for improving energy efficiency and sustainability. J. Cleaner Prod. 260, 121082.
Ribeiro, M.T., Singh, S., Guestrin, C., 2016. Why should I trust you? explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 1135–1144.
SABS, 2011. The Application of the National Building Regulations Part XA: Energy Usage in Buildings, first ed. South African Bureau of Standards, SANS10400-XA.
Seyedzadeh, S., Rahimian, F.P., Oliver, S., Rodriguez, S., Glesk, I., 2020. Machine learning modelling for predicting non-domestic buildings energy performance: A model to support deep energy retrofit decision-making. Appl. Energy 279, 115908.
Seyedzadeh, S., Rahimian, F.P., Rastogi, P., Glesk, I., 2019. Tuning machine learning models for prediction of building energy loads. Sustainable Cities Soc. 47, 101484.
Van Le, H., Hoang, D.A., Tran, C.T., Nguyen, P.Q., Hoang, N.D., Amiri, M., Ngo, T.P.T., Nhu, H.V., Van Hoang, T., Bui, D.T., et al., 2021. A new approach of deep neural computing for spatial prediction of wildfire danger at tropical climate areas. Ecol. Inform. 63, 101300.
Walker, S., Khan, W., Katic, K., Maassen, W., Zeiler, W., 2020. Accuracy of different machine learning algorithms and added-value of predicting aggregated-level energy performance of commercial buildings. Energy Build. 209, 109705.
Wang, S., Yan, C., Xiao, F., 2012. Quantitative energy performance assessment methods for existing buildings. Energy Build. 55, 873–888.
Xia, X., Zhang, J., 2010. Energy efficiency and control systems–from a poet perspective. IFAC Proc. Vol. 43 (1), 255–260.
Yalcintas, M., 2008. Energy-savings predictions for building-equipment retrofits. Energy Build. 40 (12), 2111–2120.
Yezioro, A., Dong, B., Leite, F., 2008. An applied artificial intelligence approach towards assessing building performance simulation tools. Energy Build. 40 (4), 612–620.
Zhao, H., Magoulès, F., 2012. A review on the prediction of building energy consumption. Ren. Sustain. Energy Rev. 16 (6), 3586–3592.
Zhao, Y., Zhang, C., Zhang, Y., Wang, Z., Li, J., 2020. A review of data mining technologies in building energy systems: Load prediction,, pattern identification, fault detection and diagnosis. Energy Built Environ. 1 (2), 149–164.