



Article

Using Machine Learning to Enrich Building Databases—Methods for Tailored Energy Retrofits

Jenny von Platten ^{1,2,*} , Claes Sandels ³, Kajsa Jörgensson ⁴, Viktor Karlsson ⁴,
Mikael Mangold ¹ and Kristina Mjörnell ^{2,5} 

¹ Division of Built Environment, RISE Research Institutes of Sweden, Sven Hultins plats 5, 412 58 Gothenburg, Sweden; mikael.mangold@ri.se

² Department of Building and Environmental Technology, Faculty of Engineering, Lund University, Ole Römers väg 1, Box 118, 221 00 Lund, Sweden; kristina.mjornell@ri.se

³ Division of Safety and Transport, RISE Research Institutes of Sweden, Sven Hultins plats 5, 412 58 Gothenburg, Sweden; claes.sandels@ri.se

⁴ Department of Energy Sciences, Faculty of Engineering, Lund University, Ole Römers väg 1, Box 118, 221 00 Lund, Sweden; kajsajorgensson@gmail.com (K.J.); viktor.skanor@gmail.com (V.K.)

⁵ Sustainable Cities and Communities, RISE Research Institutes of Sweden, Sven Hultins plats 5, 412 58 Gothenburg, Sweden

* Correspondence: jenny.von.platten@ri.se

Received: 20 April 2020; Accepted: 15 May 2020; Published: 19 May 2020



Abstract: Building databases are important assets when estimating and planning for national energy savings from energy retrofitting. However, databases often lack information on building characteristics needed to determine the feasibility of specific energy conservation measures. In this paper, machine learning methods are used to enrich the Swedish database of Energy Performance Certificates with building characteristics relevant for a chosen set of energy retrofitting packages. The study is limited to the Swedish multifamily building stock constructed between 1945 and 1975, as these buildings are facing refurbishment needs that advantageously can be combined with energy retrofitting. In total, 514 ocular observations were conducted in Google Street View of two building characteristics that were needed to determine the feasibility of the chosen energy retrofitting packages: (i) building type and (ii) suitability for additional façade insulation. Results showed that these building characteristics could be predicted with an accuracy of 88.9% and 72.5% respectively. It could be concluded that machine learning methods show promising potential to enrich building databases with building characteristics relevant for energy retrofitting, which in turn can improve estimations of national energy savings potential.

Keywords: building database enrichment; machine learning; artificial intelligence; Google Street View; energy performance certificate; support vector machine; energy retrofitting; energy transition; building-specific information; long-term renovation strategy

1. Introduction

Energy used in buildings accounts for 40% of the total energy use in the European Union (EU) [1] and improving the energy efficiency of buildings is thus an important undertaking in the EU's energy transition. It has been widely argued that an increased refurbishment rate is needed in order to intensify the rate of energy efficiency in the building stock [2], and it is thus favourable to prioritise energy conservation measures in buildings that, due to technical deficiencies, must be refurbished either way. To precipitate this process, the EU requires all member states to have a long-term renovation strategy with focus on energy efficiency in order to comply with directive 2018/844 on the energy performance

of buildings [3]. One of the main objectives of the long-term renovation strategy is to facilitate the transformation of existing buildings into nearly zero-energy buildings [3].

In Sweden, a significant part of the multifamily building stock was constructed between 1945 and 1975 [4], and many of these buildings are today facing significant needs for refurbishment [5]. Consequently, there is a window of opportunity to incorporate energy conservation measures in the refurbishment of this part of the multifamily building stock [6]. One advantage with doing so is that the rapid construction of new buildings between 1945 and 1975 was partly facilitated by standardised building methods and building types, which now allows for standardised methods for refurbishment and energy conservation measures for these different building types [7]. By utilising knowledge on what energy conservation measures are suitable for the different building types, energy retrofits can be tailored for each individual building in the stock 1945–1975. This in turn makes it possible to generate more accurate estimations of the national energy savings potential in the multifamily building stock and the associated costs, which is important information for the long-term renovation strategy.

However, in order to successfully estimate the energy savings potential, detailed information about this part of the multifamily building stock is required. Today, building-specific data about the multifamily building stock 1945–1975 is available through the energy performance certificates (EPCs), which cover more than 90% of all multifamily buildings in Sweden [8], but the information of building characteristics in the EPCs is limited.

One data source that could complement the EPCs with lacking building-specific information is Google Street View, which provides 360-degree panorama imagery of streets and their surroundings, with a high coverage of both urban and rural areas across the globe [9]. Google Street View is used to collect information in many scientific disciplines [10–13]. As building address is provided in Swedish EPCs, it is possible to search for and visually collect building-specific data of characteristics such as building type, façade material, and eaves shape from Google Street View.

To facilitate such data collection, machine learning (ML) methods can be used. A broad body of research has explored ML applications in the field of building research [14]. Many studies use ML to predict energy use [15–17], to conduct occupant sensing [18], and for various applications in smart buildings [19]. Studies have also explored ML applications concerning building retrofitting potential. Re Cecconi et al. [20] used artificial neural networks and geographic information software to cluster school buildings in northern Italy and define appropriate retrofit scenarios for the homogeneous clusters. Similarly, Marasco and Kontokosta [21] used energy audit data from New York City and ML classification to identify eligible energy conservation measures based on buildings' technical characteristics. Closer to the aim of this paper, other studies have focused on using ML to enrich building databases with various building characteristics [22–28], and Google Street View has been used to collect data in some of these studies. For example, Zeppelzauer et al. [29] and Li et al. [30] used artificial neural networks for image feature extraction to estimate building age from Google Street View, and Doersch et al. [31] as well as Lee et al. [32] used Google Street View to estimate architectural epochs.

However, studies utilising Google Street View and ML for predictions of building-specific suitability for specific energy conservation measures remain a rather unexplored area of research. More so, as most studies use artificial neural networks and image recognition, there is limited knowledge on the benefits of expert influence in the generation of ML models outside the sphere of deep learning. The contribution of this study is thus to investigate the prospects for such methods by utilising a limited number of expert observations in Google Street View and more transparent ML methods to enrich EPC data for the Swedish multifamily building stock 1945–1975. By predicting unknown building characteristics relevant for energy retrofitting on this part of the multifamily building stock, suitable combinations of energy conservation measures can be tailored to each specific building based on the building type and other hitherto unknown characteristics. Refining methods for such analyses has the potential to enable unparalleled national estimations and strategies for improved energy efficiency in the multifamily building stock 1945–1975, which can be of advantage in the national long-term

renovation strategy. This paper summarises and adds to work conducted to support the Swedish long-term renovation strategy and the master thesis project conducted by Karlsson and Jörgensson [33].

2. Materials

In this section, the materials used in this paper are described. First, characteristics of the Swedish multifamily building stock 1945–1975 are presented. Second, the EPC data, which acts as the representation of the building stock in the analyses, are described. Third, tailored energy retrofitting packages that are used as cases in this study are presented. These energy retrofitting packages will determine which building characteristics that need to be predicted in order to assess the feasibility of each energy retrofitting package for each individual multifamily building 1945–1975. Finally, Google Street View is briefly described, which is where the new building characteristics are collected.

2.1. The Swedish Multifamily Building Stock from 1945–1975

At the beginning of the 20th century, Sweden had among the poorest housing standards in Europe [34]. As a remedy, construction of new buildings of high quality was initiated by the government in the 1940s with the aim to provide adequate housing for all [34]. This was the start of three decades of extensive and standardised construction of new buildings in Sweden. The construction peaked between 1965–1975 when more than 100,000 dwellings, apartments as well as single-family houses, were built per year due to a government decision in 1964 to build one million dwellings in ten years, the so called Million Homes Programme [4]. To date, the high rate of construction during the Million Homes Programme remains unparalleled in Sweden. The effort put into construction of new buildings from the government ultimately contributed to a high housing standard in Sweden that today is one of the best across Europe. However, buildings from this period are now facing increasing needs for refurbishment and improvements in energy efficiency [5]. The EPC ratings of the multifamily building stock 1945–1975 can be seen in Figure 1, which shows that the majority of the buildings have an EPC rating of E or F.

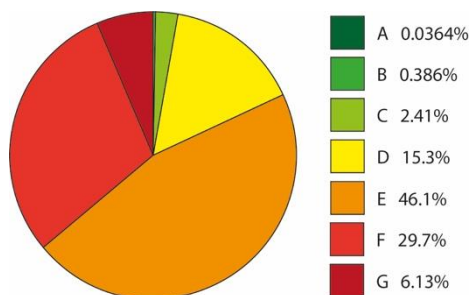


Figure 1. Share of the multifamily building stock 1945–1975 in each EPC rating A–G.

Due to the high intensity of standardised construction, recurring patterns and properties can be found in buildings built between 1945 and 1975. The main sectioning of multifamily buildings from this period is based on the appearance and the size of the building, where three main building types are usually considered: slab blocks, panel blocks, and tower blocks [35]. Representative images of these building types can be seen in Figure 2a–d.

Slab blocks. The vast majority of multifamily buildings built between 1945 and 1975 belonged to the category slab blocks. Slab blocks are rectangular, detached buildings often found in smaller groups and with 3 storeys, but can vary between 2 and 4 storeys. Most often, slab blocks have 2 or 3 stairwells, and a building depth of 10–11 m. Slab blocks were slightly differently built before and after 1960 [35]. Representative slab blocks 1945–1960 and 1960–1975 can be seen in Figure 2a,b respectively.

Panel blocks. Towards the end of the period 1945–1975, panel blocks were built on the outskirts of urban areas. Panel blocks are long, rectangular multifamily buildings with more storeys than

slab blocks. They are usually found in parallel formations and often have 2 or 3 stairwells [35]. A representative panel block can be seen in Figure 2c.

Tower blocks. Tower blocks are square-shaped multifamily buildings with one stairwell in the middle. During the 1940s, tower blocks were usually built with 3–8 storeys, but during the 1950s and 1960s they were built with 8–11 storeys. Tower blocks are usually found in groups [35]. A representative tower block can be seen in Figure 2d.



Figure 2. Representative building types 1945–1975: (a) slab block from 1945–1960; (b) slab block from 1960–1975; (c) panel block and (d) tower block.

2.2. Data from Energy Performance Certificates

In Sweden, EPCs constitute the most comprehensive national register of energy-related building-specific data for the multifamily building stock [36]. At the end of 2008, all Swedish multifamily buildings had to have a registered EPC [37]. Consequently, the coverage of the multifamily building stock is high, with over 90% of all multifamily buildings having a registered EPC [8]. Although critique has been directed towards the data quality in Swedish EPCs [38–41], one advantage over the EPCs of other EU member states is that values for building energy use are measured in Sweden, as opposed to calculating energy use which is the most common approach across the EU [42].

In Sweden, the Board of Housing, Building and Planning are responsible for monitoring the EPCs. EPCs are collected in their database Gripen which offers public access to limited EPC data, but researchers can access full EPC records via special snapshots from the database that are downloaded every six months.

In this paper, a GRIPEN snapshot from 1st of July 2015 is used. The snapshot contains approximately 130,000 unique EPCs out of which approximately 50,000 are from the period 1945–1975. The EPC data have previously been enriched with data from the Swedish Land Survey according to methods described by Johansson et al. [8]. Table 1 provides an overview of the enriched data as well as the EPC data.

Table 1. An overview of the available data in Swedish EPCs as described by Mangold et al. [43].

Data Source	Value Category	Data Specification	Measurement Type
Previously enriched data	Data from the Swedish Land Survey	Coordinates	Scale variable (m ²)
		Year of re-construction	Scale variable (year)
		Value year	Scale variable (year)
EPC data	Matching, keys, and sorting	National real estate number and index	-
		Address, area code, post code	-
		EPC index	-
	Building characteristics	Year of construction ¹	Scale variable (year)
		Complexity ¹	Binary (complex, non-complex)
		Shared walls with other buildings ¹	Ordinal (detached, semi-attached, attached)
		Recognition of heritage value	Binary (heritage value, no heritage value)
		Number of storeys	Ordinal
		Number of stairwells	Ordinal
		Number of apartments	Ordinal
	Building usage	Number of floors below ground	Ordinal
		National registration of building usage type code	Nominal
	Building area	Detailed usage of building ¹	Share (% area) of building used for the 12 most common usages
		Interior areas ¹	Scale variable (m ²)
	Heating	Heated garage area	Scale variable (m ²)
		Energy use for heating divided in 13 energy sources ¹	Scale variable (kWh/year)
		Tic box for how energy use is measured	Binary (measured, distributed)
	Household electricity and water	Period of energy use measurement	Interval (year and month)
		Energy use for cooling ¹	Scale variable (kWh/year) and nominal (measured, distributed)
		Energy use for tap water ¹	Scale variable (kWh/year) and nominal (measured, distributed)
Electricity use divided in domestic, shared, and non-domestic use ¹		Scale variable (kWh/year) and nominal (measured, distributed)	
Ventilation	Type of ventilation system ¹	Nominal (exhaust, balanced, balanced with heat exchanger, exhaust with heat pump, natural ventilation)	
	Tic box for conducted/not conducted ventilation control ¹	Nominal (yes, no, partially)	
Recommended energy conservation measures	Tic box for 28 common energy conservation measures	Nominal	
	Estimated energy savings ¹	Scale variable (kWh/year)	
	Estimated cost per saved kWh ¹	Scale variable (SEK/kWh)	

¹ Required by the EU.

2.3. Case: Tailored Energy Retrofitting Packages

For this study, a reference with tailored energy retrofitting packages for the Swedish multifamily building stock 1945–1975 have been used (see Reference [44]). These energy retrofitting packages will be used to exemplify how building database enrichment can enable more accurate national estimations of energy savings and costs, as a certain number of building characteristics are needed in order to allocate appropriate energy retrofitting packages to buildings. For each building type described in Section 2.1, there are three available packages (1–3) which entail different costs and energy savings (low to high). The packages must be applied in successive order, meaning that Package 2 requires Package 1 to have been conducted, and Package 3 requires both Package 1 and Package 2 to have been conducted.

- In Package 1, a number of measures that aim at optimising the operation of the building are undertaken [44]. Apart from building type, no building characteristics must be known in order to determine the feasibility of the measures in Package 1.

- In Package 2, components such as pumps and fans are changed to more effective counterparts, and additional insulation is added in the attic and to existing windows [44]. As for Package 1, building type is the only characteristic that needs to be known in order to determine the feasibility of the measures in Package 2.
- Package 3 contains the most extensive measures, including a new ventilation system with heat exchange from exhaust air, a change of windows, and 10 cm additional insulation on the building envelope [44]. To determine the feasibility of Package 3, two building characteristics apart from building type are of advantage to know. The first characteristic is the façade material; more specifically, it is of advantage to know whether the building has a brick façade or not, as brick facades often must be preserved due to cultural and historical values. Additional insulation on a brick façade is thus not always a feasible option. More so, the shape of the roof and length of the eaves determines whether there is room for additional façade insulation or not, and additional façade insulation on buildings with an existing eaves overhang thus involves less extensive inventions than when the existing roof must be adjusted to a thicker façade. Consequently, eaves overhang is a necessary building characteristic to know to determine the feasibility of Package 3.

The energy savings and the associated costs for each of the energy retrofitting packages can be seen in Tables 2 and 3 respectively. The costs in Table 3 are marginal costs for carrying out the energy conservation measures in each energy retrofitting package in conjunction with planned refurbishment, meaning that the costs are in most cases lower than what they would have been if the energy retrofitting packages were carried out independent of other refurbishment. A detailed list of the energy conservation measures in each of the energy retrofitting packages can be seen in Table A1 in the Appendix A. It should be noted that slab blocks have been divided into two categories based on their construction year, as they were constructed differently during 1945–1960 compared to 1960–1975, as mentioned in Section 2.1. Finally, the energy savings and costs in Tables 2 and 3 have been slightly altered from the reference [44] as the reference contained more detailed differentiation between building types than was considered necessary for this study.

Table 2. Percental energy savings for each building type and energy retrofitting package according to the reference study [44].

Building Type	Package 1 (%)	Package 2 (%)	Package 3 (%)
Slab block, 1945–1960	14.2	25.2	63.8
Slab block, 1960–1975	9.7	25.6	59.1
Tower block	17.6	25.4	63.6
Panel block	8.5	23.7	54.6

Table 3. Marginal costs for each building type and energy retrofitting package according to the reference study [44].

Building Type	Package 1 (€/m ²)	Package 2 (€/m ²)	Package 3 (€/m ²)
Slab block, 1945–1960	6.0	115	398
Slab block, 1960–1975	5.1	147	426
Tower block	3.9	112	435
Panel block	2.5	120	437

2.4. Google Street View

Google Street View is a technology developed from what was initially called the Stanford CityBlocks Project and started as a collaboration between Google and Stanford University [9]. The aim was to utilise the massive amount of information in street-level imagery to “organize the world’s information and make it universally accessible and useful” [9]. Since 2007, Google Street View is

featured in Google Maps and Google Earth and allows users to visually interact with streets (and an increasing amount of off-road sites as well) across the globe that have been captured in Google's 360-degree panoramic imagery.

Today, Google Street View is widely used for research purposes, and a search for the technology in Google Scholar search engine generates close to 20,000 hits (January 2020). It is used for collection of information in fields such as urban forestry [10,45], health research [11], and pedestrian behaviour [12,13], to name a few.

Limitations of making observations in Google Street View include limited coverage of certain areas (especially rural areas) as well as a limitation of observations to morphological building characteristics that are apparent to an external observer.

3. Methods

In this section, the methods of collecting data, developing ML models, and applying these algorithms on the entire multifamily building stock 1945–1975 are described.

3.1. Observations in Google Street View

Ocular observations in Google Street View were conducted for 476 EPCs that were sampled from the total of 50,000 EPCs 1945–1975. The sampling was performed as weighted random sampling, where the probability of each EPC being selected was determined by the area of the building the EPC represented. The reason for the weighted random sampling was to gain a sample that was representative of the building stock in respect to area rather than in respect to the individual EPCs. However, due to a low representation of certain building types (tower blocks) in the sampled data, observations were conducted for another 41 manually selected EPCs, resulting in a total of 517 observations. The manual selection of complementing EPCs was based on number of storeys, as tower blocks usually are higher than slab blocks.

For the sample of 517 EPCs, observations were conducted in Google Street View using the registered address in the EPC. Observations were conducted by all of the authors, and methods to ensure that observations were conducted uniformly were undertaken. The quality of the observations was ensured by first letting all authors make observations guided by a senior researcher. After that, a control matrix of 13 observations was constructed to ensure that all authors' classifications agreed. After corrections, authors conducted observations individually. Any ambiguous cases were discussed with a senior researcher before classification, or rejection as valid observation. In three cases, observations were not possible due to lack of coverage in Google Street View. These EPCs were thus removed which resulted in a total of 514 observations. This was considered a sufficient number of observations as iterative testing of ML models starting at 200 observations showed no significant improvement in accuracy after 400 observations. The building characteristics and the respective classes that were observed are listed in Table 4. The choice of which building characteristics to observe was based on the gap between available data in the EPCs and data needed in order to assess the feasibility of energy retrofitting packages from the case presented in Section 2.3. It was found that the characteristics building type, whether or not the building has a brick façade, and whether or not the building has eaves overhang were needed. As seen in Table 4, rowhouses are included as a building type to be observed despite them not being introduced in Section 2.1. Rowhouses are not multifamily buildings per se, but the way they are owned determine whether their EPCs end up in the category for multifamily buildings or not. Rowhouses that are owned in similar ways as multifamily buildings (rental housing, resident-cooperatives) are classified as multifamily buildings in the EPC database, and they will thus be necessary to identify when using the EPCs to study the multifamily buildings 1945–1975. They will however be excluded from analyses of the energy savings potential in the multifamily building stock 1945–1975.

Table 4. Building characteristics that were observed in Google Street View and their respective measurement type.

Building Characteristic	Measurement Type and Classes
Building type	Nominal [slab block, panel block, tower block, rowhouse, other]
Façade material	Binary [brick, not brick]
Eaves overhang	Binary [overhang, no overhang]

Figure 3 shows an example of an observation of a building with eaves overhang and brick façade and a break-down of the classifications in the 514 observations is shown in Table 5. It should be noted that part of the methods development included dropping several observed building characteristics due to difficulties in assuring data quality; the observations presented in this paper are features for which we could reach a satisfactory data quality.



Figure 3. Example of observation in Google Street View of a building with eaves overhang and without brick façade, i.e., a building that is suitable for additional façade insulation.

Table 5. A specification of the classifications of the 514 observations conducted in Google Street View.

Observed Building Characteristic	Number of Observations	Share of Observations
Building type		
Slab block	342	63.0%
Panel block	81	15.8%
Tower block	36	7.00%
Rowhouse	32	6.23%
Other	23	4.47%
Total	514	100%
Not brick façade		
Eaves overhang	215	41.8%
Eaves overhang and not brick façade	117	22.8%

Finally, the geographical distribution of the 514 observations is shown in Figure 4. The light dots in the map show all multifamily buildings constructed between 1945 and 1975, whereas the black crosses mark the multifamily buildings that were observed in Google Street View. It can be seen that the studied multifamily building stock is distributed all across Sweden in a way that reflects the population density of the country. The observations show a similar pattern, indicating that they constitute a geographically representative sample.

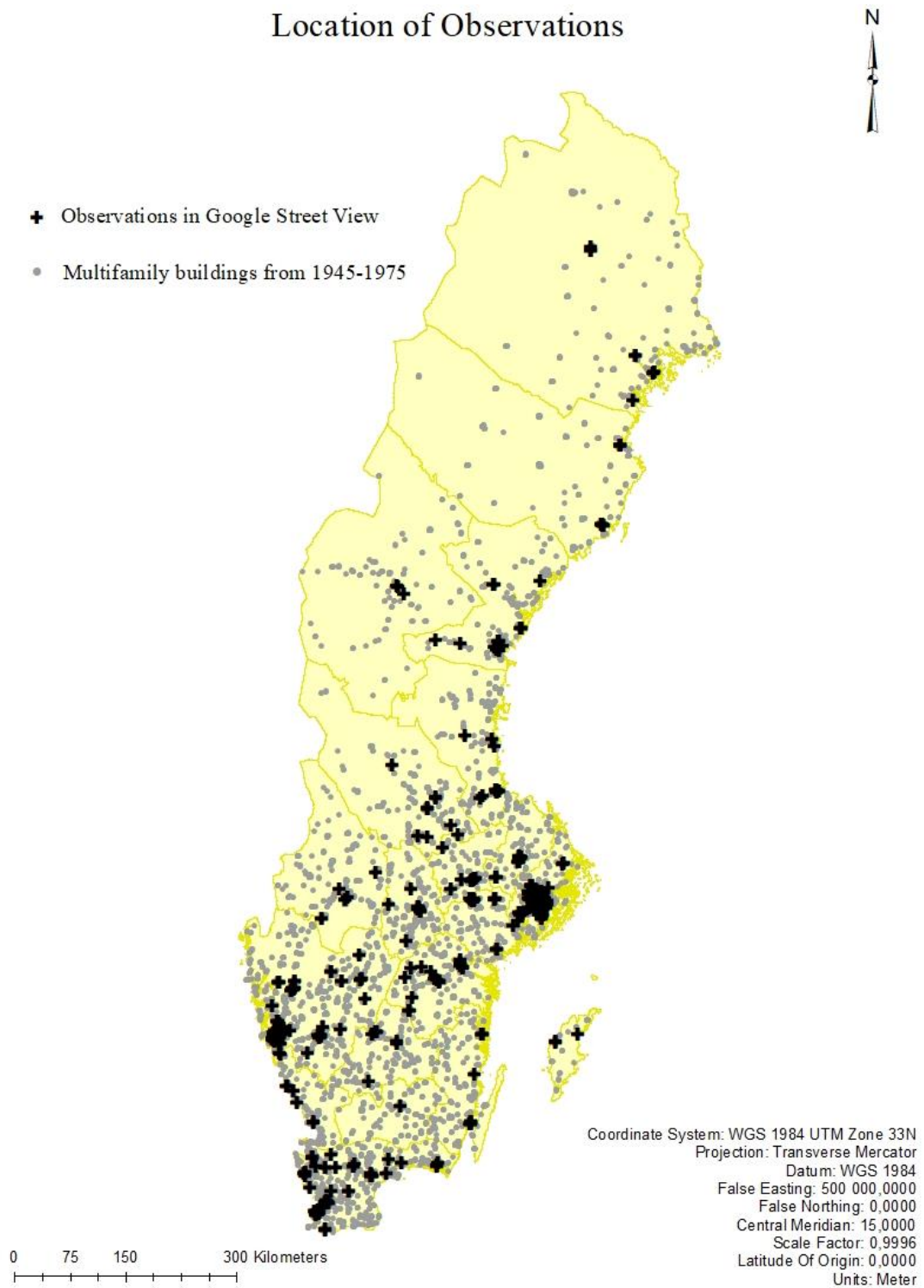


Figure 4. A map of Sweden showing the distribution of the multifamily building stock constructed between 1945 and 1975 (light dots) and the buildings that were observed in Google Street View (black crosses).

3.2. Selection of Possible Features

After the observations had been conducted, the features that should be considered in the development of prediction models were chosen with a two-step approach: (i) relevant features were highlighted by research experts, (ii) the features from step (i) was validated with stepwise regression which is an automatic feature selection methodology. The stepwise regression was thus mainly used to confirm features pointed out by domain experts, although it also contributed with a few additional features. When appropriate, new features were generated as ratios of two or more of the available features to better differentiate between building characteristics. The features selected to proceed with for each building characteristic to be predicted are shown in Table 6. As an example, for the prediction of building type, characteristics of the building types described in Section 2.1 along with research expertise were used to find appropriate features. It was found that number of storeys and year of construction were suitable features to separate building types from one another, as can be seen in Figure 5. Before the application of features in ML models, all numerical features were normalised with minmax.

Table 6. The selected features (from available EPC data) to consider for the prediction of each building characteristic.

Building Characteristic	Selected Possible Features from Stepwise Regression	Feature Type	Unit	Numerical Feature Representation
Building type	Number of stories	Raw feature	-	(1–15)
	Construction year	Raw feature	Year	(1945–1975)
	Heated space per story and address	Derived feature	m ²	(681–70,110)
	Number of stairwells per EPC	Raw feature	-	(0–82)
	Number of apartments per address	Derived feature	-	(1–189)
Façade material	Building type ¹	Derived feature	-	Slab block (1), panel block (2), tower block (3), other (4)
	Position longitude	Raw feature	m ²	(6,134,178.3–7,537,187)
	Position latitude	Raw feature	m ²	(279,176.1–916,455.9)
	Area code	Raw feature	-	(1–25)
	Post code	Raw feature	-	(11,111–98,492)
	EPCs per property	Derived feature	-	(1–132)
Eaves overhang	Building type ¹	Derived feature	-	Slab block (1), panel block (2), tower block (3), other (4)
	Construction year	Raw feature	Year	(1945–1975)
	Number of stories	Raw feature	-	(1–15)
	Position longitude	Raw feature	m ²	(6,134,178.3–7,537,187)
	Energy performance	Raw feature	kWh/m ²	(21–482)
	Number of stairwells per apartment	Derived feature	-	(0–6.5)
	Post code	Raw feature	-	(11,111–98,492)

¹ It should be noted that a predicted building characteristic (building type) was used as a feature in the prediction of the other building characteristics (eaves overhang and façade material).

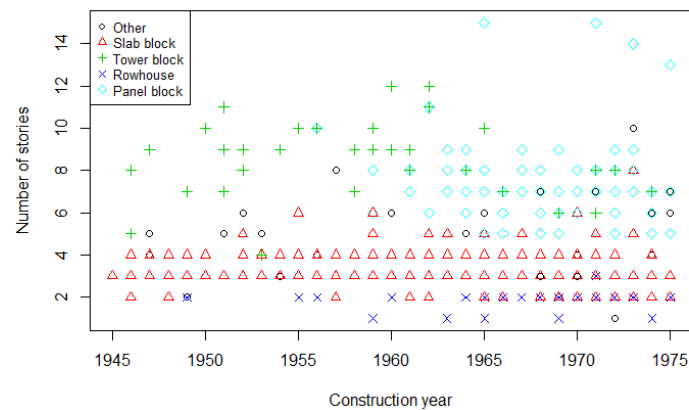


Figure 5. Visualisation of how construction year (x) and number of stories (y) correlate with building type. It can be seen that building types tend to form clusters, indicating that the x and y variables together succeed to differentiate between building types. It should be noted that one point in the figure sometimes corresponds to several identical observations, which explains why the total number of points is lower than 514.

3.3. Training and Testing of Algorithms

Three types of supervised ML models for classification problems are considered in the search of the optimal model. In supervised learning, the ML models are trained on labelled training data which provides instant feedback on whether a prediction is correct or not. Consequently, the ML models go through an iterative self-learning process to refine their respective algorithms. These models have different characteristic when it comes to the bias and variance trade-off. The model types that were considered in this study were logistic regression (LR) and support vector machines (SVM). The objective of logistic regression is to model the mean of a dependent variable with respect to a set of predictors. Logistic regression uses a logistic function to model the binary dependent variables. The aim of SVM is to find a hyperplane that separates the data point of the different classes. The optimal plane maximizes the distance between the data points of the classes. The hyperplanes will become decision boundaries that will be used for classifying new data observations. SVM can both handle linear and non-linear classifiers.

LR has high bias and low variance, meanwhile SVM has low bias and high variance. To find the best prediction model for the data, trade-offs must be made between model bias and model variance. Model bias is the risk of oversimplifying the model, e.g., choosing a linear model when the data is non-linear. Model variance is the risk of considering random noise in the training data, which could reduce the capability of predicting new examples. Balance between bias and variance can be achieved by choosing the most appropriate model type for the problem, together with proper regularizing of the model parameters. The appropriate choice can only be made by testing a wide variety of models.

By testing numerous ML model structures with various features, the search space of the optimal model increases. Note, the models can be further optimized with parameter tuning and regularization [46].

The next step is to select the features and the ML model type that gives the highest prediction accuracy. This is done with 10-fold cross-validation. Before cross-validation, the 514 observations are randomly divided into training data (80%) and testing data (20%). In cross-validation, the training data is randomly split into 10 training and test data sets. Then, in each iteration, models are trained with data from 9 folds, and tested on the 10th fold. The process is repeated until all folds have acted as test data. Finally, the resulting models are fitted with the training data, and an error rate is estimated with the test data. Various combinations of feature selection, ML model types and parameter tunings are tested.

Table 7 shows the four ML models that were considered for prediction of building type, which constitutes a 5-class classification problem. Three main attributes of the ML models were considered in

the choice of the optimal model: (i) a high overall accuracy in cross-validation with close proximity to the accuracy on the testing data, (ii) a distribution of accuracy among the model classifications that is suitable for the intended application of the ML model, and (iii) a low number of input features in order to maintain a certain level of interpretability. With respect to these attributes, the model SVM1 was chosen. This model had (i) high accuracy in both cross-validation and on the testing data, with only minor differences between the two, as well as (ii) high accuracy for slab blocks and panel blocks, which both dominate over tower blocks and rowhouses in the multifamily building stock 1945–1975. The distribution of accuracy among the classifications was thus considered appropriate for the intended application. Finally, the model had (iii) a relatively low number of input features. For the choice of ML model to predict suitability for additional façade insulation, which was a 2-class classification problem, similar reasoning was conducted. The features and the test accuracy for the chosen ML models for prediction of building type and suitability for additional façade insulation can be seen in Table 8.

Table 7. Details for four of the considered models for prediction of building type.

Model	Overall Accuracy (%)		Specific Accuracy (%)				
	Cross-Validation	Testing Data	Slab Blocks	Panel Blocks	Tower Blocks	Rowhouses	Other
SVM1 (Chosen model)	88.5	88.9	95.2	94.4	71.4	85.7	0
SVM2	89.3	87.9	98.4	88.9	71.4	57.1	0
LR1	88.0	87.9	93.7	88.9	85.7	85.7	0
LR2	87.5	87.9	95.2	88.9	85.7	71.4	0

Table 8. The chosen prediction model and its accuracy for each of the predicted building characteristics.

Building Characteristic	Features in Selected Model	Machine Learning Model	Accuracy
Building type	Number of stories	SVM	88.9
	Construction year		
	Heated space per story and address		
	Number of apartments per address		
Eaves overhang + not brick façade	Construction year	SVM	72.5
	Number of apartments		
	Number of stairwells per apartment		
	Area code		

It can be seen in Table 8 that building type can be predicted with an accuracy close to 90%. Façade material and eaves overhang could both be predicted with an accuracy of approximately 68%, but in combination, these characteristics could be predicted with an accuracy of 72.5%. The combined model was thus chosen as the two building characteristics were to be used for the common cause of determining buildings' suitability for additional façade insulation.

Finally, the ML models were validated by observing a random sample of 20 EPCs in Google Street View, for which building type and suitability for additional façade insulation had been predicted. The validation showed an accuracy of 90.0% for building type, which is close to the obtained test accuracy. For suitability for additional façade insulation, the validation showed an accuracy of 63.2%, which is lower than the obtained test accuracy. The relatively small validation sample however makes it difficult to draw major conclusions from these results.

4. Results

In this section, the distribution of the predicted characteristics of the multifamily building stock 1945–1975 are presented. The predicted characteristics are then used to estimate energy savings potential in this part of the housing stock, using the energy retrofitting packages presented in Section 2.3.

4.1. Energy Retrofitting Characteristics of the Multifamily Building Stock from 1945–1975

Based on the models shown in Table 8, building type and possibility for additional insulation could be predicted for the entire multifamily building stock 1945–1975. The results of this prediction can be seen in Figure 6 and Table 9. In Figure 6, the predicted distribution of the different building types in the multifamily building stock 1960–1975 is presented with a model accuracy of 88.9%. First of all, it can be seen that almost all of the multifamily buildings from this era can be categorised into one of the four building types that were described in Section 2.1. Figure 6 shows that approximately 3% of the buildings in the EPC database belong to the category “other” (rowhouses are included here), and less than 2% of the buildings in the EPC database lacked sufficient information to be categorised at all. Second of all, the results in Figure 6 confirm that slab blocks dominate among the multifamily buildings 1945–1975, followed by panel blocks and tower blocks.

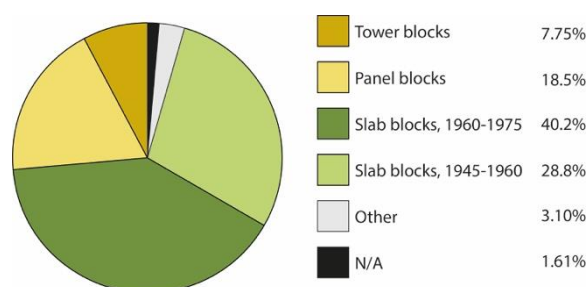


Figure 6. The predicted distribution of building types in the multifamily building stock from 1945–1975. N/A values were generated if one or more of the features in the prediction model were missing.

Table 9. The predicted possibility for additional façade insulation (and thus for energy retrofitting package 3 in the reference [44]) among the predicted building types and in the entire multifamily building stock 1945–1975.

Building Type	Eaves Overhang and Not Brick Façade [%]
Slab blocks, <1960	63.9
Slab blocks, 1960–1975	22.0
Panel blocks	6.81
Tower blocks	26.4
All building types in multifamily building stock 1945–1975	32.0

Table 9 shows the predicted distribution of multifamily buildings 1945–1975 with characteristics favourable for additional façade insulation, i.e., buildings with eaves overhang and a façade material that is not brick. With a model accuracy of 72.5%, it was predicted that 32.0% of all multifamily buildings from this era have both of the favourable characteristics for additional façade insulation which make them suitable for energy retrofitting package 3. Table 9 also shows how the suitability of energy retrofitting package 3 is distributed among the different building types, based on the share of buildings with eaves overhang and not brick façade. It can be seen that it is primarily slab blocks constructed before 1960 that have favourable characteristics for additional façade insulation, followed by tower blocks, slab blocks constructed between 1960 and 1975, and finally panel blocks. This type of information provides knowledge regarding the feasibility of certain energy conservation measures on specific building types, which can facilitate planning of energy retrofitting programmes and means for allocation of resources.

These results provide new insight into the energy savings potential of this part of the building stock. As can be seen in Table 2, energy savings above 50% are only possible with energy retrofitting package 3, as energy savings from energy retrofitting package 2 are only approximately 25%. Assuming that there would be extensive requirements to preserve brick façades, and that issues of cost-effectiveness

would restrict additional façade insulation to be added in the absence of a eaves overhang, the results in Table 9 indicate that the energy savings potential for most of the buildings 1945–1975 would be around 25% rather than the often assumed 50%. Although these are rough assumptions, the results in Table 9 provide valuable insight in the potential trade-offs that could be faced between energy savings and cultural preservation, as well as between energy savings and cost-effectiveness. Even more so, these results showcase how building databases enriched with new building-specific information can improve understanding and descriptions of the building stock.

4.2. Examples of National Strategies for Tailored Energy Retrofitting

This section will provide an example of how enriched building databases can be applied to generate more accurate energy retrofitting strategies that can be used for policy purposes such as in the long-term renovation strategy. Figure 7 showcases how building-specific information can be used in decision trees for different energy retrofitting packages, based on the example described in Section 2.3. In this decision tree, four building characteristics are used: renovation status, energy rating, suitability for additional façade insulation, and building type. Renovation status and energy rating are characteristics that were present in the building database before enrichment, whereas suitability for additional façade insulation and building type are characteristics that were predicted in this study. It is assumed that buildings with eaves overhang and without brick façade are suitable for additional façade insulation. In other words, this example illustrates the energy savings potential assuming that there are strict requirements for the preservation of brick facades, and that additional façade insulation is only relevant to buildings with an already existing eaves overhang due to limitations in retrofitting costs.

The example in Figure 7 is based on the notions that (i) energy retrofitting should be carried out along with other planned refurbishment measures, i.e., in a “window of opportunity”, and (ii) that the overall objective is to transform existing buildings into nearly zero-energy buildings, in accordance with the objective of the long-term renovation strategy [3]. In Sweden, nearly zero-energy buildings are defined as buildings with an EPC rating between A–C [47]. Based on these notions, it can be seen in Figure 7 that recently renovated buildings, i.e., cases where the window of opportunity has been missed, are excluded from energy retrofitting. Likewise, buildings that already fulfil the requirements of nearly zero-energy buildings (EPC rating A–C) are also excluded from energy retrofitting. Buildings that have not been recently renovated and with the EPC rating D (i.e., buildings close to nearly zero-energy building standard) are allocated energy retrofitting package 1. Finally, buildings that have not been recently renovated and with an EPC rating between E–G are allocated energy retrofitting package 2 or 3 depending on their suitability for additional façade insulation, which is part of energy retrofitting package 3.

Based on the decision tree in Figure 7 and the energy savings presented in the reference case in Table 2, the yearly national energy savings potential and the associated costs were estimated for two scenarios. The results can be seen in Figure 8a,b and in Figure 9a,b. In Figure 8a,b, it is assumed that only buildings that are suitable for additional façade insulation (eaves overhang and not brick façade) are allocated Package 3. This assumption represents a conservative approach where all brick facades are preserved. In Figure 9a,b, it is assumed that 50% of buildings that are not considered suitable for additional façade insulation are allocated Package 3 regardless of their unsuitability. This assumption represents a more realistic case with increased compromise between historical preservation and energy savings. The figures are based on the assumption that buildings are refurbished and energy retrofitted when they reach their expected service life of 50 years. The expected service life has been adjusted based on previous refurbishments according to methods developed by Mangold et al. [48], and pent-up needs for refurbishment have been evenly distributed between year 2020 and 2030.

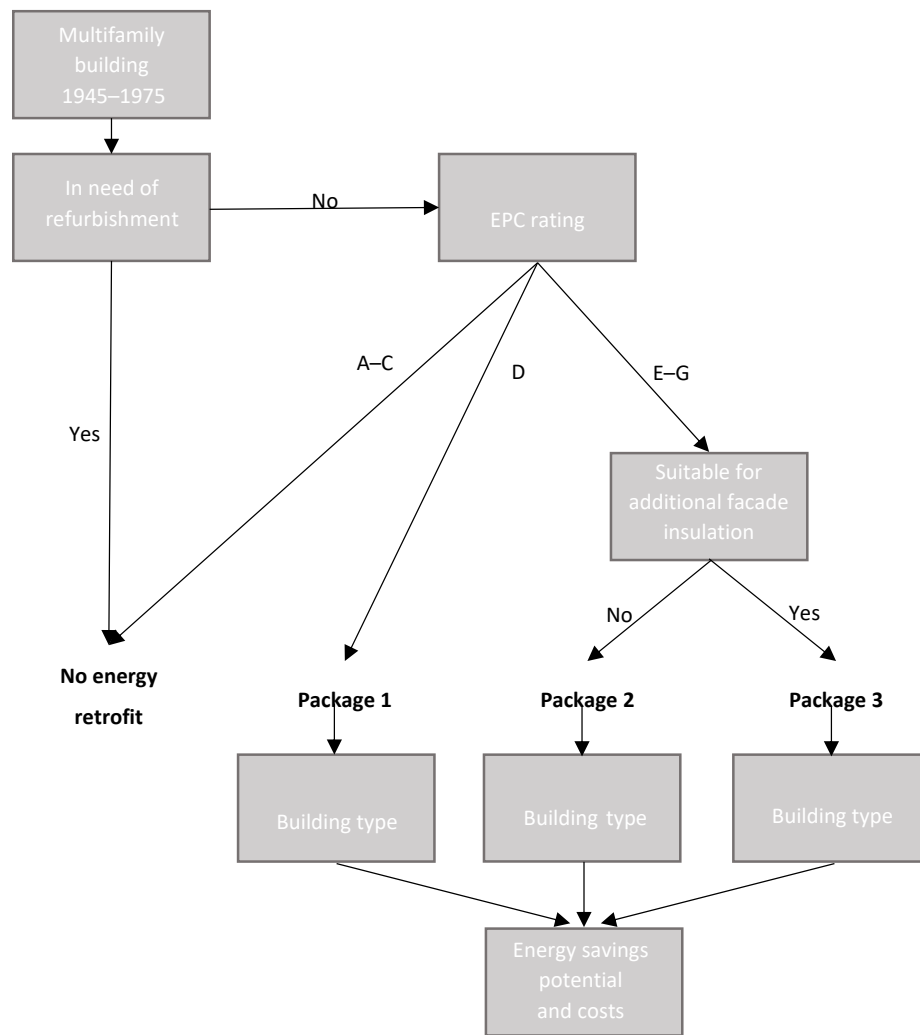


Figure 7. Decision tree showing how four building characteristics (renovation status, EPC rating, suitability for additional façade insulation, and building type) can help determine a tailored energy retrofitting package for each individual building based on the suggested energy retrofitting packages described in Section 2.3.

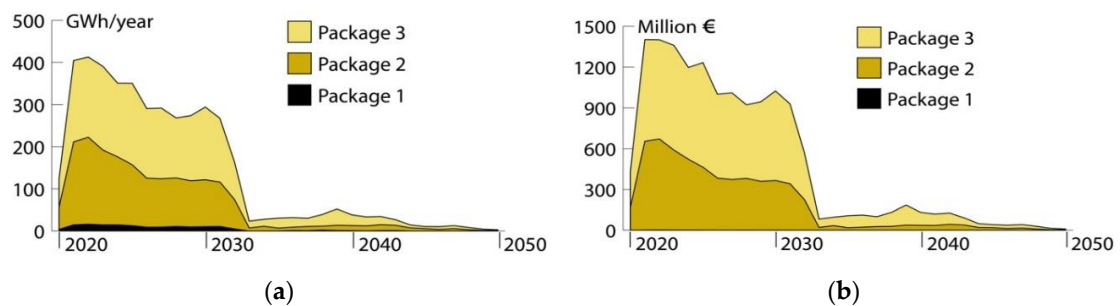


Figure 8. The figure shows yearly, cumulative: (a) Energy savings potential from the different energy retrofitting packages and (b) The associated costs. The results are based on the energy retrofitting strategy presented in Figure 7 and the referred energy retrofitting packages [44]. This figure is based on the assumption that buildings that are not suitable for additional façade insulation are not allocated Package 3, and that buildings are refurbished, and energy retrofitted when they reach their expected service life of 50 years. The expected service life has been adjusted based on previous refurbishments, and pent-up needs for refurbishment have been evenly distributed between year 2020 and 2030.

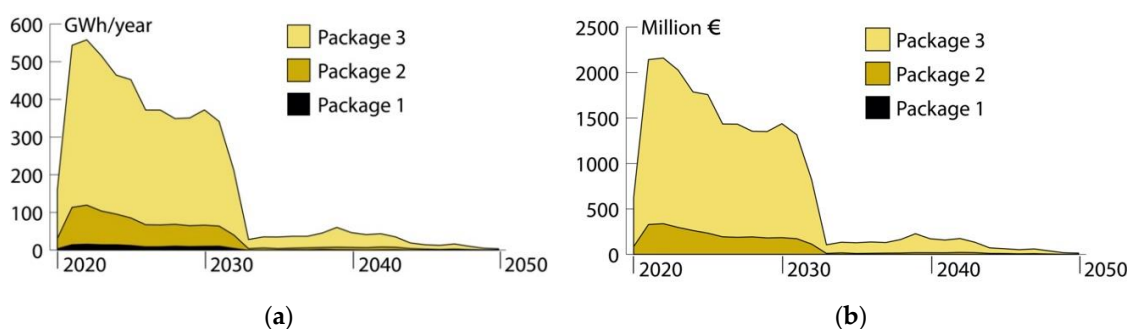


Figure 9. The figure shows yearly, cumulative: (a) Energy savings potential from the different energy retrofitting packages and (b) The associated costs. The results are based on the energy retrofitting strategy presented in Figure 7 and the referred energy retrofitting packages [44]. This figure is based on the assumption that 50% of buildings not suitable for additional façade insulation are allocated Package 3, and that buildings are refurbished and energy retrofitted when they reach their expected service life of 50 years. The expected service life has been adjusted based on previous refurbishments, and pent-up needs for refurbishment have been evenly distributed between year 2020 and 2030.

Figure 8a shows that with the more conservative approach, the yearly energy savings potential is rather evenly distributed between energy retrofitting package 2 and package 3. This is explained by the fact that the high energy savings from energy retrofitting package 3 compensate for the relatively low suitability of energy retrofitting package 3 among the multifamily buildings 1945–1975 (32%). Consequently, despite a relatively low feasibility in the concerned part of the building stock, the energy savings potential from energy retrofitting package 3 constitutes a significant part of the total energy savings potential in this part of the building stock. Yet, it can also be concluded that an equally significant part of the total energy savings potential is constituted by energy retrofitting package 2, where energy savings are approximately 25% (as seen in Table 2). As shown in Figure 8b, the higher energy savings in energy retrofitting package 3 come at a considerable cost.

Figure 9a shows that the less conservative approach markedly increases the share of energy savings coming from Package 3. Compared to the results in Figure 8a, Figure 9a show rewards in terms of energy savings for compromising the historical and cultural preservation. The cost for this reward is however significant, as can be seen in Figure 9b. Based on the strategy for energy retrofitting in Figure 7, the results in Figure 8a,b and Figure 9a,b showcase how building characteristics can be applied to estimate the national energy savings potential and the associated costs under different assumptions. More specifically, enriched building databases can help create scenarios for the national feasibility of certain energy conservation measures and pinpoint building types with high and low energy savings potential. This knowledge is useful to improve the national long-term renovation strategy and can facilitate decision-making in the area of energy policy.

5. Discussion

In this paper, Google Street View data collection together with ML were used to enrich existing building databases with characteristics relevant for investigating buildings' energy retrofitting potential. This section will first discuss the challenges and opportunities of using ML methods to enrich building databases (Section 5.1). Subsequently, the benefits of enriching databases with building characteristics relevant for energy retrofitting will be discussed (Section 5.2). Finally, Section 5.3 will provide a short discussion on the main contributions of this paper.

5.1. Using Machine Learning to Enrich Building Databases

This paper has demonstrated how ML methods can be used to enrich building databases with new information based on a small sample of expert observations. Previously unregistered building characteristics were observed in Google Street View for a representative sample of the concerned

building stock, and these observations could then be used to develop ML models that were used to predict the building characteristics for the entire multifamily building stock from 1945–1975. The building characteristics in this paper could be predicted with an accuracy of approximately 70–90%, and it was found that building type was easier to predict (accuracy 88.9%) than the more detailed characteristics eaves overhang and façade material (accuracy 72.5%).

Whether a certain prediction accuracy is enough depends on the application and purpose of the prediction, as well as the potential cost of misclassifications. In the case of the predictions in this paper, the purpose is to improve estimations of national energy savings potential. This means that a prediction accuracy above 50% (i.e., higher accuracy than a random model) will contribute to the purpose of improving estimations as the alternative would be estimations without concern regarding the feasibility of certain energy conservation measures. Owing to that, the attained model accuracies are considered sufficient for the intended purpose of the predictions. Moreover, as prediction accuracy is dependent on a simplified model of reality, e.g., that there only are four different multifamily building types constructed 1945–1975, it is rarely expected to reach an accuracy of 100%.

Although image recognition could be a more feasible method to predict more detailed building characteristics, there are benefits of using building-specific data as input features rather than images. One benefit is that using building data instead of building images allows for increased expert influence, as features for ML models can be selected by researchers with expert knowledge on the analysed building stock. In this study, a combination of expert knowledge and regression methods was used for feature selection, which thus allowed expert influence without closing the doors to unexpected features. This in turn led to generation of ML models that were not entirely of “black-box” character, which is likely to appeal to researchers who are new to applying ML in their building stock research.

The results of this paper thus imply that ML methods can be suitable for predicting building characteristics and enriching national building databases, especially for the application of improving estimations on a national level. Predictions will never constitute decision support for retrofits of specific buildings, as this requires a high level of accuracy and detail. It is thus first and foremost for planning and estimations on national and regional levels that the methods developed in this paper are useful.

Future work could explore image recognition as used in the studies by Zeppelzauer et al. [29] and Li et al. [30] to enrich building databases with more building characteristics relevant for energy retrofitting. Similar methods could also be useful to predict where energy retrofitting measures such as additional façade insulation have already been done, as such information rarely is available in national registers. Although Google Street View was used to collect observations in this study, similar studies could also be done using other types of sample observations of non-ocular building characteristics. Records from building inventories could e.g., provide information on the prevalence of certain materials or chemical substances that could be used to predict the potential occurrence of such materials and substances in the entire building stock and in specific building types.

As buildings often have been constructed with certain characteristics during different periods of time, they constitute a suitable subject for ML methods as they are likely to display distinct patterns in choice of, e.g., materials, morphology, and construction methods over time. The Swedish multifamily building stock 1945–1975 and the results of this paper constitute an evident example of this. Similar traits of building stocks worldwide motivate an increased use of ML methods to enrich building databases.

5.2. Implications of the Ability to Tailor Energy Retrofits

Enriching building databases with building characteristics relevant for energy retrofitting enables more accurate estimations of the national energy savings potential. More specifically, it can enable more accurate estimations of which energy conservation measures that are feasible and for which types of buildings. This paper has showcased how building database enrichment can be used to tailor energy retrofitting packages for the entire Swedish multifamily building stock from 1945–1975.

Such estimations elevate the discussion on energy retrofitting to a more realistic level. Although it is explicitly stated that the long-term renovation strategy should facilitate transformation of existing buildings into nearly zero-energy buildings, which in most cases requires energy performance improvements of 50% or more, the results of this paper suggest that with regards to matters of cultural preservation and cost-effectiveness, it is likely that many buildings will only achieve energy performance improvements of approximately 25%. This is based on the finding that a minority of the existing buildings 1945–1975 are readily suitable for additional façade insulation, and that a significant part of the energy savings is likely to stem from less intrusive energy conservation measures. More so, cultural and historical preservation is only one out of several values that can conflict with deep energy retrofitting. Other values such as social justice and rights to affordable housing are likely to further infringe on the potential for deep energy retrofitting [48–50].

As all EU member states are obliged to provide a long-term renovation strategy every three years in order to comply with directive 2018/844 on the energy performance of buildings [3], similar applications of building database enrichment could be adopted elsewhere. In this paper, pre-constructed energy retrofitting packages were used in order to estimate the energy savings potential, and although similar material is likely to be found in other EU member states, studies can also focus on the feasibility of specific measures rather than packages of measures. Ultimately, ML methods for building database enrichment should facilitate the upscaling of limited information to regional or national level and enable more accurate estimations of energy saving potentials. This will provide a strong foundation for improved policies and more feasible roadmaps for improved energy efficiency in the building stock.

5.3. Contribution

We would finally like to emphasise what we consider to be the main takeaways from this study. In this paper, we are making two different types of observations and predictions for each observed building. First, the building type is classified for which different renovation packages have been developed [44]. Second, specific features that might enable or restrict energy retrofitting depths were observed. The primary contribution of this paper is the addition of building type as a building-specific characteristic. This makes it possible to apply building-specific retrofitting strategies with associated costs and energy savings. As for the secondary contribution, choice of energy retrofitting depths (such as whether to add additional façade insulation or not) is a more complex matter: there are many building features that are of relevance of which few are visually observable, and energy retrofitting depth is ultimately decided by the building owner whom have a host of different parameters and uncertainties to consider. Furthermore, when the results of the building specific predictions are turned into possible long-term renovation strategies, numerous of other assumptions need to be made on a building stock level. While this research was used for the Swedish long-term renovation strategy, this paper does not explore the building stock level assumptions needed to provide decision support. The scope of the paper is instead limited to the building specific predictions using ML methods.

Finally, it is the use of ML methods and quick observations in Google Street View to enrich national building databases that is the main contribution of this paper. This showcases how access to national building stock data can generate new data that enables a wider range of analyses at the national level.

6. Conclusions

Based on 514 ocular observations collected from Google Street View, this paper has explored using machine learning methods to enrich building databases with new building characteristics relevant for estimating the energy retrofitting potential. With the aim to utilise these building characteristics to improve national estimations of energy savings potential, machine learning was used to predict the building characteristics (i) building type and (ii) suitability for additional façade insulation. This was done for all multifamily buildings in Sweden constructed between 1945 and 1975 based on the Swedish database of energy performance certificates. It was found that these building characteristics could be predicted with a model accuracy of 88.9% and 72.5% respectively, which was considered a sufficient

level of accuracy for the intended applications. These results were finally used to exemplify the national energy savings potential in the multifamily building stock 1945–1975 under different assumptions.

Two main takeaways can be concluded from this paper. First, due to the time-dependent characteristics of buildings, building stocks are suitable subjects for machine learning methods as time-dependent patterns are likely to be found. The prospects for increased use of machine learning methods for building database enrichment are thus promising in a wide range of applications. Second, machine learning methods for enriching building databases with characteristics relevant for energy retrofitting showed to offer new insights into potential scenarios for energy savings from different energy conservation measures. For example, it was found that many of the multifamily buildings 1945–1975 were not suitable for additional façade insulation when considering the cultural preservation of façades and extension of eaves overhang. There are thus great opportunities for machine learning applications in building database enrichment to offer more accurate estimations of national energy savings potential, and to provide insights in trade-offs that could occur between energy savings and other values in the building stock.

Author Contributions: Conceptualization, C.S. and M.M.; Data curation, J.v.P. and M.M.; Formal analysis, J.v.P., C.S., K.J. and V.K.; Funding acquisition, C.S.; Investigation, J.v.P., C.S., K.J., V.K., M.M. and K.M.; Methodology, C.S., K.J., V.K., M.M. and K.M.; Project administration, C.S. and M.M.; Resources, M.M.; Software, C.S.; Supervision, C.S., M.M. and K.M.; Validation, J.v.P.; Visualization, J.v.P., C.S. and M.M.; Writing—original draft, J.v.P. and C.S.; Writing—review & editing, J.v.P., C.S., K.J., V.K., M.M. and K.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the Swedish Energy Agency (Energimyndigheten) [grant number 2018-006053] within the project Artificial Intelligence for Interpretation of Retrofitting Potential, as well as by The Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning (Formas) [grant number 2017-01449] within the project National Building-Specific Information (NBI).

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. An overview of the three energy retrofitting packages (1–3) in the Reference [28] and the energy conservation measures in each of these packages.

Package 1—Operation Optimisation	Package 2—Update to more Efficient Components and Smaller Supplements	Package 3—Long-Term Sustainable Envelope and Ventilation
Check that ventilation flows are in accordance with projected flows	Change circulation pumps to effective pumps with the accurate capacity	Install heat exchange from exhaust air
Lower temperature in stairwell to 15 °C	Change to energy efficient fans	Change to windows with better U-value
Adjustment of temperature of incoming air flow (only for FTX)	Additional insulation of pipes and conduits where possible	Additional façade insulation, 10 cm
Limit ventilation flows in areas that are not constantly occupied	Upgrade laundry equipment	
Review of control systems to minimise energy losses	Complement existing windows with insulating windowpanes	
Review and update operational instructions	Additional insulation in attic, 20 cm	
Develop routines for operational statistics	Installation of individual metering and billing of domestic hot water	
Automatics control of stairwell lightning and switch to energy efficient light bulbs		
Adjustment of heating system to minimise temperature gradients in the building		
Lower indoor temperature to 21 °C		
Education of operational managers on the building's systems		

References and Notes

1. EU. Buildings. Available online: <https://ec.europa.eu/energy/en/topics/energy-efficiency/buildings> (accessed on 20 August 2019).
2. Boverket. Förslag Till Nationell Strategi för Energieffektiviserande Renovering av Byggnader. 2013. Available online: <https://www.boverket.se/sv/om-boverket/publicerat-av-boverket/publikationer/2013/forslag-till-nationell-strategi-for-energieffektiviserande-renovering-av-byggnader/> (accessed on 7 February 2020).
3. Directive (EU) 2018/844 of the European Parliament of the Council of 30 May 2018 Amending Directive 2010/31/EU on the Energy Performance of Buildings and Directive 2012/27/EU on Energy Efficiency. 2018.
4. Hall, T.; VidÉN, S. The Million Homes Programme: A review of the great Swedish planning project. *Plan. Perspect.* **2005**, *20*, 301–328. [[CrossRef](#)]
5. Mjörnell, K.; Femenías, P.; Annadotter, K. Renovation strategies for multi-residential buildings from the record years in Sweden—Profit-driven or socioeconomically responsible? *Sustainability* **2019**, *11*, 6988. [[CrossRef](#)]
6. Högberg, L.; Lind, H.; Grange, K. Incentives for improving energy efficiency when renovating large-scale housing estates: A case study of the swedish million homes programme. *Sustainability* **2009**, *1*, 1349. [[CrossRef](#)]
7. Brown, N.W.O.; Malmqvist, T.; Bai, W.; Molinari, M. Sustainability assessment of renovation packages for increased energy efficiency for multi-family buildings in Sweden. *Build. Environ.* **2013**, *61*, 140–148. [[CrossRef](#)]
8. Johansson, T.; Olofsson, T.; Mangold, M. Development of an energy atlas for renovation of the multifamily building stock in Sweden. *Appl. Energy* **2017**, *203*, 723–736. [[CrossRef](#)]
9. Anguelov, D.; Dulong, C.; Filip, D.; Frueh, C.; Lafon, S.; Lyon, R.; Ogale, A.; Vincent, L.; Weaver, J. Google street view: Capturing the world at street level. *IEEE Comput.* **2010**, *43*, 32–38. [[CrossRef](#)]
10. Li, X.; Ratti, C.; Seiferling, I. Quantifying the shade provision of street trees in urban landscape: A case study in Boston, USA, using Google Street View. *Landsc. Urban Plann.* **2018**, *169*, 81–91. [[CrossRef](#)]
11. Rzotkiewicz, A.; Pearson, A.L.; Dougherty, B.V.; Shortridge, A.; Wilson, N. Systematic review of the use of Google Street View in health research: Major themes, strengths, weaknesses and possibilities for future research. *Health Place* **2018**, *52*, 240–246. [[CrossRef](#)]
12. Yin, L.; Cheng, Q.; Wang, Z.; Shao, Z. ‘Big data’ for pedestrian volume: Exploring the use of Google Street View images for pedestrian counts. *Appl. Geogr.* **2015**, *63*, 337–345. [[CrossRef](#)]
13. Yin, L.; Wang, Z. Measuring visual enclosure for street walkability: Using machine learning algorithms and Google Street View imagery. *Appl. Geogr.* **2016**, *76*, 147–153. [[CrossRef](#)]
14. Hong, T.; Wang, Z.; Luo, X.; Zhang, W. State-of-the-art on research and applications of machine learning in the building life cycle. *Energy Build.* **2020**, *212*, 109831. [[CrossRef](#)]
15. Runge, J.; Zmeureanu, R. Forecasting energy use in buildings using artificial neural networks: A review. *Energies* **2019**, *12*, 3254. [[CrossRef](#)]
16. Bourdeau, M.; Zhai, X.; Nefzaoui, E.; Guo, X.; Chatellier, P. Modeling and forecasting building energy consumption: A review of data-driven techniques. *Sustain. Cities Soc.* **2019**, *48*, 101533. [[CrossRef](#)]
17. Seyedzadeh, S.; Rahimian, F.P.; Glesk, I.; Roper, M. Machine learning for estimation of building energy consumption and performance: A review. *Vis. Eng.* **2018**, *6*, 5. [[CrossRef](#)]
18. Saha, H.; Florita, A.R.; Henze, G.P.; Sarkar, S. Occupancy sensing in buildings: A review of data analytics approaches. *Energy Build.* **2019**, *188–189*, 278–285. [[CrossRef](#)]
19. Qolomany, B.; Al-Fuqaha, A.; Gupta, A.; Benhaddou, D.; Alwajidi, S.; Qadir, J.; Fong, A.C. Leveraging machine learning and big data for smart buildings: A comprehensive survey. *IEEE Access* **2019**, *7*, 90316–90356. [[CrossRef](#)]
20. Cecconi, F.R.; Moretti, N.; Tagliabue, L.C. Application of artificial neural network and geographic information system to evaluate retrofit potential in public school buildings. *Renew. Sustain. Energy Rev.* **2019**, *110*, 266–277. [[CrossRef](#)]
21. Marasco, D.E.; Kontokosta, C.E. Applications of machine learning methods to identifying and predicting building retrofit opportunities. *Energy Build.* **2016**, *128*, 431–441. [[CrossRef](#)]
22. Tooke, T.R.; Coops, N.C.; Webster, J. Predicting building ages from LiDAR data with random forests for building energy modelling. *Energy Build.* **2014**, *68*, 603–610. [[CrossRef](#)]

23. Henn, A.; Römer, C.; Gröger, G.; Plümer, L. Automatic classification of building types in 3D city models. *GeoInformatica* **2010**, *16*, 281–306. [CrossRef]
24. Liu, H.; Zhang, J.; Zhu, J.; Hoi, S.C.H. Deepfacade: A deep learning approach to facade parsing. In Proceedings of the 26th International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, 19–25 August 2017.
25. Jampani, V.; Gadde, R.; Gehler, P.V. Efficient facade segmentation using auto-context. In Proceedings of the 2015 IEEE Winter Conference on Applications of Computer Vision, Waikoloa Village, HI, USA, 5–9 January 2015; pp. 1038–1045. [CrossRef]
26. Yang, J.; Shi, Z.-K.; Wu, Z.-Y. Towards automatic generation of as-built BIM: 3D building facade modeling and material recognition from images. *Int. J. Autom. Comput.* **2016**, *13*, 338–349. [CrossRef]
27. Despotovic, M.; Koch, D.; Leiber, S.; Döller, M.; Sakeena, M.; Zeppelzauer, M. Prediction and analysis of heating energy demand for detached houses by computer vision. *Energy Build.* **2019**, *193*, 29–35. [CrossRef]
28. Koch, D.; Despotovic, M.; Sakeena, M.; Döller, M.; Zeppelzauer, M. Visual estimation of building condition with patch-level ConvNets. In Proceedings of the 2018 ACM Workshop on Multimedia for Real Estate Tech, Yokohama, Japan, 11 June 2018. Available online: <https://doi.org/10.1145/3210499.3210526> (accessed on 10 February 2020).
29. Zeppelzauer, M.; Despotovic, M.; Sakeena, M.; Koch, D.; Döller, M. Automatic Prediction of Building Age from Photographs. In Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval, Yokohama, Japan, Association for Computing Machinery, 11–14 June 2018; pp. 126–134.
30. Li, Y.; Chen, Y.; Rajabifard, A.; Khoshelham, K.; Aleksandrov, M. Estimating Building Age from Google Street View Images Using Deep Learning (Short Paper). In Proceedings of the 10th International Conference on Geographic Information Science (GIScience 2018), Melbourne, Australia, 28–31 August 2018.
31. Doersch, C.; Singh, S.; Gupta, A.; Sivic, J.; Efros, A. What makes Paris look like Paris? *ACM Trans. Graph.* **2012**, *31*, 103–110. [CrossRef]
32. Lee, S.; Maisonneuve, N.; Crandall, D.; Efros, A.A.; Sivic, J. Linking past to present: Discovering style in two centuries of architecture. In Proceedings of the IEEE International Conference on Computational Photography, Houston, TX, USA, 24 April 2015. Available online: <https://hal.inria.fr/hal-01152482/document> (accessed on 10 February 2020).
33. Karlsson, V.; Jörgensson, K. Energibesparande Renoveringspotential—Renoveringspotentialen för det Svenska Flerbostadshusbeståndet Uppskattad Med Maskininläring. Master’s Thesis, Department of Energy Sciences, Lund University, Lund, Sweden, 2020.
34. Nylander, O. *Svensk Bostad 1850–2000*; Studentlitteratur: Lund, Sweden, 2013.
35. Björk, C.; Reppen, L.; Kallstenius, P. *Så byggdes Husen 1880–2000: Arkitektur, Konstruktion och Material i Våra Flerbostadshus under 120 år*; Svensk Byggtjänst: Stockholm, Sweden, 2013.
36. Mangold, M. Chalmers tekniska högskola, Department of Civil and Environmental Engineering, Challenges of Renovating the Gothenburg Multi-Family Building Stock: An Analysis of Comprehensive Building-Specific Information, Including Energy Performance, Ownership and Affordability (Doktorsavhandlingar vid Chalmers Tekniska Högskola: Chalmers tekniska högskola). Ph.D. Thesis, Department of Civil and Environmental Engineering, Chalmers University of Technology, Gothenburg, Sweden, 2016.
37. Lag (2006:985) om Energideklaration för Byggnader (English Translation: The Swedish act Concerning Energy Performance Certificates). 2006.
38. Hårsman, B.; Daghbashyan, Z.; Chaudhary, P. On the quality and impact of residential energy performance certificates. *Energy Build.* **2016**, *133*, 711–723. [CrossRef]
39. Claesson, J. CERBOF Projekt no. 72: Utfall och Metodutvärdering av Energideklaration av Byggnader. 2011.
40. Göransson, A. Recalculation between BOA+LOA and Atemp for Mult-Family Dwellings: Account of Conducted Measurement Work. 2007.
41. Pasichnyi, O.; Wallin, J.; Levihn, F.; Shahrokni, H.; Kordas, O. Energy performance certificates—New opportunities for data-enabled urban energy policy instruments? *Energy Policy* **2019**, *127*, 486–499. [CrossRef]
42. Arcipowska, A.; Anagnostopoulos, F.; Mariottini, F.; Kunkel, S. *A Mapping of National Approaches: Energy Performance Certificates across the EU*; Buildings Performance Institute Europe (BPIE): Brussels, Belgium, 2014.
43. Mangold, M.; Österbring, M.; Wallbaum, H. Handling data uncertainties when using Swedish energy performance certificate data to describe energy usage in the building stock. *Energy Build.* **2015**, *102*, 328–336. [CrossRef]

44. Lönsam Energieeffektivisering: Saga Eller Verklighet? för hus Byggda 1950-75; VVS-företagen: Stockholm, Sweden, 2012.
45. Berland, A.; Lange, D.A. Google Street View shows promise for virtual street tree surveys. *Urban For. Urban Green.* **2017**, *21*, 11–15. [[CrossRef](#)]
46. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning: With Applications in R*; Springer: New York, NY, USA, 2013.
47. The National Board of Housing, Building and Planning. 2011:6, *Boverkets Byggregler (2011:6)—Föreskrifter och Allmänna Råd*; BBR: Karlskrona, Sweden, 2011.
48. 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*, 41–49. [[CrossRef](#)]
49. Grossmann, K. Using conflicts to uncover injustices in energy transitions: The case of social impacts of energy efficiency policies in the housing sector in Germany. *Glob. Trans.* **2019**, *1*, 148–156. [[CrossRef](#)]
50. Von Platten, J.; Mangold, M.; Mjörnell, K. On the significance of energy performance metrics for social (in)justice in the energy transition of the urban housing stock. **2020**, Submitted manuscript.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).