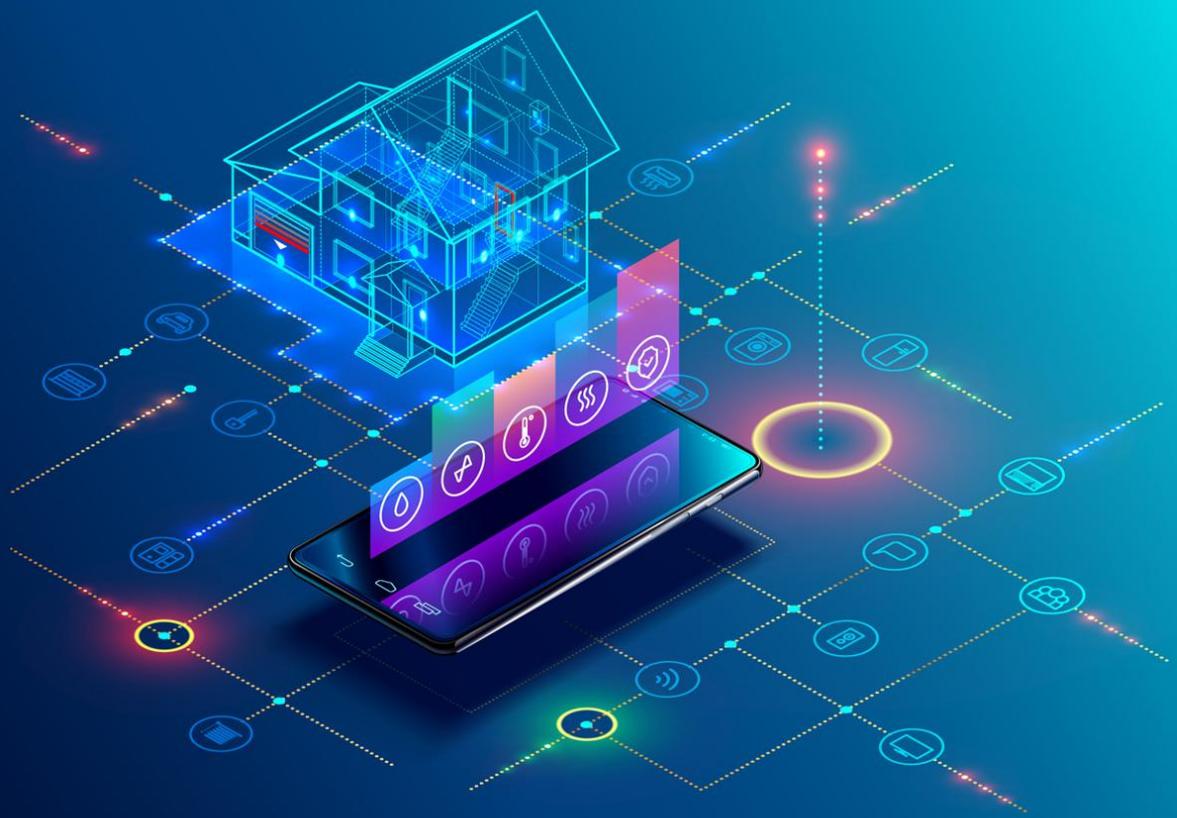


# Machine-learning for As-Built diagnostics and enrichment of design rules for deep renovation

Deliverable Report D7.2



Deliverable Report: D.7.2, issue date on 22.02.2022

BIM-SPEED

Harmonised Building Information Speedway for Energy-Efficient Renovation

This research project has received funding from the European Union's Programme H2020-NMBP-EEB-2018 under Grant Agreement no 820553.

**Disclaimer**

The contents of this report reflect only the author's view and the Agency and the Commission are not responsible for any use that may be made of the information it contains.

# Machine-learning for As-Built diagnostics and enrichment of design rules for deep renovation

## Deliverable Report D7.2

Issue Date	22 <sup>nd</sup> February 2022
Produced by	CARTIF
Main author	Sofía Mulero Palencia (CARTIF)
Co-authors	Sonia Álvarez Díaz (CARTIF), Manuel Andrés Chicote (CARTIF)
Contributors	FAS
Version:	V 2
Reviewed by	Bogdan Andrei Tofan (ARC), Dragos Nica (ARC), Paul Stupin (PB40)
Approved by	Timo Hartmann (TUB)
Dissemination	Public
Type	Report

### Colophon

Copyright © 2019 by BIM-SPEED consortium

Use of any knowledge, information or data contained in this document shall be at the user's sole risk. Neither the BIM-SPEED Consortium nor any of its members, their officers, employees or agents shall be liable or responsible, in negligence or otherwise, for any loss, damage or expense whatever sustained by any person as a result of the use, in any manner or form, of any knowledge, information or data contained in this document, or due to any inaccuracy, omission or error therein contained. If you notice information in this publication that you believe should be corrected or updated, please get in contact with the project coordinator.

The authors intended not to use any copyrighted material for the publication or, if not possible, to indicate the copyright of the respective object. The copyright for any material created by the authors is reserved. Any duplication or use of objects such as diagrams, sounds or texts in other electronic or printed publications is not permitted without the author's agreement.

This research project has received funding from the European Union's Programme H2020-NMBP-EEB-2018 under Grant Agreement no 820553. Disclaimer: The European Commission is not responsible for any use made of the information contained herein. The content does not necessarily reflect the opinion of the European Commission.



## Change log

Description	Author	Date
Initial version for internal review	Sonia Álvarez Díaz, Sofía Mulero Palencia (CARTIF)	30.04.2020
First complete version for internal review	Sofía Mulero Palencia, Sonia Álvarez Díaz (CARTIF)	09.10.2020
Contribution to the definition of the solutions catalogue for Poland.	Agnieszka Łukaszewska (FAS)	12.10.2020
Internal reviewing	Bogdan Andrei Tofan (ARC), Dragos Nica (ARC)	16.10.2020
	Paul Stupin (PB40)	16.10.2020
Addressed internal review comments	Sofía Mulero Palencia, Sonia Álvarez Díaz (CARTIF)	26.10.2020
Final technical editing	Timo Hartmann (TUB)	28.10.2020
Addressed EU revision comments: Changes in subsection 1.1	Sonia Álvarez Díaz (CARTIF)	11.10.2021
Addressed EU revision comments: Changes in executive summary	Sonia Álvarez Díaz, Sofía Mulero Palencia (CARTIF)	15.02.2022



## Publishable executive summary

The digitalisation of the AEC (Architecture, Engineering and Construction) sector is enabling improved construction process and control over the life cycle of buildings. This digitalisation, mainly focused on the generation of BIM models, combined with the use of new emerging technologies such as Artificial Intelligence (AI), is allowing for decision making in these construction processes. This improved decision making in the field of building renovation would allow for better tailoring of the measures to be applied and help to reduce the design time of projects, which is of great interest given the large number of buildings that need to be retrofitted to comply with new energy efficiency regulations. This will necessitate the collecting, recording and evaluating data, to cost/benefit analysis and to providing information to involved bodies like: regulations and code makers, designers, contractors, implementers of quality assurance systems, insurance companies, planners, etc.

This report summarizes the results of the second part of TASK 7.2- *Renovation design using building and HVAC components in BIM Libraries* of the BIM-SPEED project, focused on exploring possibilities of the optimisation algorithm using machine-learned heuristics to analyse the findings from as-built data acquisition in order to determine the critical areas to be solved in a building through deep renovation.

The goal of this report is to describe the potentials, constraints and viable solutions of the use of machine-learning / artificial intelligence approaches at the design stage of deep renovation projects, to enable technical users (architects, designers, etc.), in conjunction with the owner's chosen intervention selection, to make better decisions on the measures to be applied. Considering the acquired data from the *WP1-Collecting and understanding BIM data from existing buildings* of the BIM-SPEED project, diagnosis and optimisation processes will be supported by machine learning/ artificial intelligence and data –driven approaches. In this task, some of these methodologies will be tested to prove how machine learning may help this process. Machine-learning will be dedicated to diagnostics (building thermal pathology) and enrichment of the design rules in a deep renovation. In order to validate this approach, a prototype of a diagnostic thermal pathologies tool has been developed and the result obtained from its application with a proof-of-concept is shown in this report.



## List of acronyms and abbreviations

AI: Artificial Intelligence

API: Application Programming Interface

As-Built: drawings, models and documentation of the building as-constructed, including all deviations that have appeared in the construction phase.

BIM: Building Information Model

HVAC: Heating Ventilation Air Conditioning

IFC: Industry Foundation Classes

KBS: Knowledge Based System

ML: Machine Learning

## Definitions

**Artificial intelligence** refers to the simulation of human intelligence in machines, which are programmed to think and perform tasks that require human intelligence to be accomplished.

**Machine learning is a** scientific discipline in the field of Artificial Intelligence that creates systems with the ability to learn automatically and improve from experience.



# Contents

<b>1. INTRODUCTION</b>	<b>8</b>
1.1 Description of the deliverable content and purpose	8
1.2 Contributions of partners	9
1.3 Relation to other activities	9
<b>2. A REVIEW ON PRIOR USAGES OF MACHINE-LEARNING IN DEEP RENOVATION PROJECTS</b>	<b>10</b>
2.1 Machine learning in deep renovation	10
2.2 Classification methods overview	12
2.2.1 Decision Trees	12
2.2.2 Rule-Based Classification	15
2.2.3 Expert System	16
<b>3. BUILDING PATHOLOGIES</b>	<b>19</b>
3.1 Existing pathology databases	20
3.2 Pathologies in the context of the available information	21
<b>4. DIAGNOSIS AND OPTIMIZATION ALGORITHM USING MACHINE LEARNING</b>	<b>23</b>
4.1 Description	23
4.2 Data acquisition	24
4.3 Design of the algorithm	25
4.3.1 Pathologies definition according to algorithm design	26
4.3.2 Solutions catalogue definition	30
4.3.3 Enrichment design rules in renovation projects	33
<b>5. PROOF OF CONCEPT</b>	<b>34</b>
5.1 Analysis of the results	34
5.2 Exploitation results of T7.2	42
<b>6. CONCLUSIONS</b>	<b>43</b>
<b>7. REFERENCES</b>	<b>44</b>
<b>APPENDIX 1 – RENOVATION MEASURES</b>	<b>46</b>



## List of Figures & Tables

Figure 1. Relation to other activities. ....	9
Figure 2. Example of decision tree to determine whether to wait for the supplier to receive the required material. In addition, main components haven been highlighted in colours.....	12
Figure 3. Architecture of an expert system (own elaboration). ....	18
Figure 4. Elements of the Building Pathology domain for several types of defects (own elaboration) [15]..	19
Figure 5. Example of report format extracted from the INSITER project, where some common error in buildings are included [21]. ....	21
Figure 6. Diagram of the machine learning process for diagnosis and optimisation processes in deep renovation projects. ....	24
Figure 7. User interface of the diagnosis tool to request information and proceed with the analysis. ....	25
Figure 8. K limit values [W/m <sup>2</sup> ·K] for private residential use in Spain.....	26
Figure 9. Flow diagram for the designed machine learning algorithm.....	29
Figure 10. First building model under analysis: medium quality building prototype. ....	34
Figure 11. Second building model under analysis: low quality building prototype.....	36
Figure 12. Third building model under analysis: low quality building prototype with flat roof. ....	39
Figure 13. Sample of information provided about the decision tree algorithm. ....	41
Figure 14. Sample evaluation output provided to user. ....	42
Table 1: Contribution of partners .....	9
Table 2. Confusion matrix definition. ....	16
Table 3. Regulation limits data sources considered for each country under study.....	26
Table 4. Decision tree outputs.....	30
Table 5. Proposed material solutions and costs for walls and roof (Spain and France).....	31
Table 6. Proposed material solutions and costs for walls and roof (Germany). ....	32
Table 7. Proposed material solutions and costs for walls and roof (Poland). ....	33
Table 8. Initial values of the Medium quality building prototype. ....	35
Table 9. Final values of the Medium quality building prototype. ....	35
Table 10. Initial values of the Low quality building prototype. ....	37
Table 11. Final values of the Low quality building prototype. ....	37
Table 12. Initial values of the Low quality building prototype with flat roof. ....	39
Table 13. Final values of the Low quality building prototype with flat roof.....	39

# 1. Introduction

## 1.1 Description of the deliverable content and purpose

The work carried out and shown in this deliverable explores the potential of optimisation algorithms to analyse the findings from as-built data acquisition to determine the critical areas to be taken into account by deep renovation strategies. Taking advantage of the acquired data from the WP1, we expect that diagnosis and optimisation processes can be supported by machine-learning / artificial intelligence / data-driven approaches. In this work, some methodologies will be tested to prove how these latest technologies may help to support this process. Machine-learning will be dedicated to diagnostics (building pathology) and enrichment of the design rules. This exploration of the potential of machine learning will profit in particular from the availability of suitable data and the possibility to train applicable models with this available data.

Section 2 provides a literature review on machine learning applied to deep renovation. Additionally, the theory behind the models that proved most successful is further analysed. Section 3 focuses on identifying data sources that collect information about building pathologies. Section 4 describes the diagnosis and optimisation process carried out following a machine learning approach at the design stage of deep renovation project. Details about the design procedure of the system and the algorithm are given. Section 5 analyses the obtained results in the proof of concept. To finish, the document collects the most relevant conclusions and references.

Note: Task 7.2- *Renovation design using building and HVAC components in BIM Libraries of the BIM-SPEED project*, is divided into two distinct parts, the results of which will be shown on the one hand in D7.3- *Semantic design rules and tool for deep renovation design (M36)*, and on the other hand in D7.2- *Machine-learning for As-Built diagnostics and enrichment of design rules for deep renovation (M24)*. The results shown in this deliverable 7.2 are part of the work that has been done in task 7.2, but focus exclusively on the second part of the task, which aimed was to show the possibilities of the use of ML applied in the design phase of projects and deep renovation.



## 1.2 Contributions of partners

Partners' contribution to the deliverable can be divided as follows:

Table 1: Contribution of partners

#	Partner	Contribution
1	<b>CARTIF</b>	Leader of the deliverable, coordination of the overall task activity, general state-of-the-art overview, analysis of information sources on pathologies in buildings, theoretical analysis of algorithms used for this purpose. Characterization of thermal pathologies and definition of the solutions catalogue. Design of the algorithm for the proof of concept and the evaluation of results.
2	<b>FAS</b>	Contribution to the definition of the solutions catalogue for Poland.
3	<b>ARC</b>	Reviewing the final version of the deliverable
4	<b>PB40</b>	Reviewing the final version of the deliverable

## 1.3 Relation to other activities

Task 7.2-Renovation design using building and HVAC components in BIM Libraries is directly linked to other activities carried out in WP7-Generating EEB renovation solutions and implementation strategies. This approach serves to improve the design of a decision-making process in renovation, and the possible solutions provided in the output can be used:

- As a starting point for other tools (for example, the METABUILD platform for scenario optimization).
- To offer recommendations to a user as a preliminary step to the beginning of the design process.

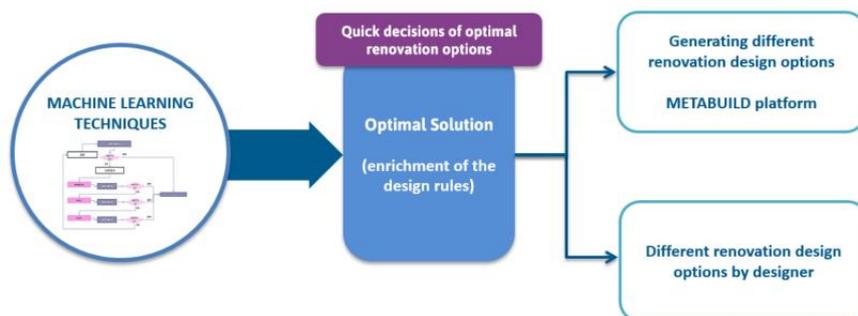


Figure 1. Relation to other activities.



## 2. A review on prior usages of machine-learning in deep renovation projects

Nowadays, the use of machine-learning in the construction field is mainly focused on the identification of elements using image recognition. This approach will, for example, also be explored in the deliverable *D1.2- Methods for surveying and diagnostics of HVAC systems in the existing buildings*. However, the aim of this task differs a little bit: the goal is to find the convergence among the use of machine-learning techniques and the ability to detect existing defects in previous constructions, which need to be solved through deep renovation. In order to automate its application, the main data source considered for the diagnosis is the IFC file, because it is one of the most widely used BIM data exchange standards. This draws a quite specific context, as there are only few examples of work that have been carried out on that issue before.

### 2.1 Machine learning in deep renovation

John McCarthy used the term AI (Artificial Intelligence) for the first time at a conference in Dartmouth in 1956. Several fields have exploited its potential since then, as it is the case of the construction sector for modelling and decision making. Some examples of research carried out in the domain of structural building diagnosis have been supported by AI techniques. Many of them are oriented to solve very specific problems. For example, Van Balen in [1] developed an expert system, which assists the user in **defining damage types and causes in ancient masonry structures, related to air pollution and traffic**. This knowledge based system (KBS) is a form of artificial intelligence that aims to capture the knowledge of human experts to support decision-making and solve complex problems. It consists of a database with information and additional knowledge to create relations according to given answers and questions. Most of the information is based on literature. Moodi and Knapton in [2] carried out another research: the main goal was to **design an effective supporting tool for experts to offer repairing solutions for damages in concrete elements**. It contains 13 decision tables, one of them for diagnosis, which constitute a dynamic resource that evolves with the help of different contributors (REPCON system). Bernat and Gil in [3] presented an Expert System for **diagnosis, finding pathologies of structural elements in buildings**, which deals with uncertainty in a useful way, aiding during the initial inspection. This system encodes knowledge using production rules and decision tables together. In [4], a research team in Lisboa developed a series of knowledge- and experience-based inspection, diagnosis and repair systems for individual non-structural building elements. The expert inspection, diagnosis and repair systems classify and **define the typical defects, probable causes, diagnosis methods and repair techniques for the considered building element**.

Since it is intended to take the IFC file as a source of information about the building, the use of machine learning techniques within this field has been analysed. In the context of IFC files, literature offers examples of the use of machine-learning techniques for detecting internal problems and incoherencies, because of the importance of IFC integrity checking. Lee [5] proposed an **inspection**



**methodology using BIM and linked data technologies for sharing defects.** This approach offers a flow of information from the detection of a pathology to those responsible for its occurrence, with the aim of avoiding future problems, and promoting dissemination actions on this matter.

Koo et al. [6] [7] explore the possibility of employing novelty detection, a machine learning approach to **detect potential misclassifications that occur during ad hoc and manual mapping.** The research explores the use of two machine learning techniques for detecting anomalies, called outlier and novelty detection. Both approaches are tested on three BIM models to evaluate their accuracy in identifying wrong classifications. Tests are carried out for different IFC classes: the algorithm is trained to learn the geometry of the BIM elements for a given IFC class. Novelty detection is implemented using one-class Support Vector Machines. One of the main limitations of the method is that it is not possible to initially know the type of outliers that may arise in unseen data, making it more difficult to comprise a training dataset with inliers and outliers. Novelty detection is superior in terms of overcoming limitations in outlier detection. Both approaches show limitations due to the lack of use of semantic relationships between elements in the BIM model.

Krijnen and Tamke [8] explore the potential of using supervised and unsupervised machine learning techniques to **detect errors in BIM models.** An unsupervised learning approach is used for anomaly detection, in order to find outliers in the geometrical attributes of elements in a model. On the other hand, a developed neural network is capable of differentiating floor plan layouts into different categories based on their geometrical appearance.

Lee et al. [9] propose a set of metrics for the **comparison between different IFC files,** a procedure that is generally performed by visual check (similarity rate, match rate, preservation rate of the global unique identifier, missing rate and rate of addition). It uses a recursive comparison algorithm to flatten distances. The algorithm records the analysis data based on the proposed metrics according to the comparison direction (from file A to B, or B to A).

Lilis et al. [10] focus their attention on **solving inaccuracies related to geometry in IFC files:** crash error, where two building solids intersect, space definition errors, where the volume of a building space is not correctly defined (gaps between volumes), and surface orientation errors, where the normal vector of some boundary surfaces is misoriented. The developed recursive orientation algorithm is tested in two IFC files.

Eftekharirad et al. [11] extend the **IFC data model for the real-time management of fire emergency, integrating sensors and occupant sensors,** by adding this information to the IfcSensor and IfcOccupant entities, respectively. After that, relationships between sensors, occupants, occupancy patterns, time series and building components are defined in the context of building evacuation. The article does not explore any machine learning techniques, but the resulting contribution is of interest to evaluate the possibility of extending the IFC file to include real-time information.

The field of building regulations has also been explored in order to verify compliance with IFC models. Regulation based BIM model verification is widely used in many disciplines, to help with the detection of potential requirements or defects. Given a building regulation, first, it is formally



described by logical formulas, then an OWL model is extracted from the IFC and SPARQL queries are generated to determine the integrity of the information. SPARQL is used to improve time response in large-scale applications. The IFC model is verified using ontologies and a prototype called BIMRuleChecker is developed for the domain experts to write rules and apply them to the process, based on the integrity verification generation algorithm.

Research in [12] also considers **health and safety issues**. This multi-domain framework is based on parametric rules and BIM-based model verification concepts (such as clash detection). The construction normative text is used, first, for rule interpretation and parametrization, and then, for rule-based code verification, to semi-automatically verify a construction site safety plan.

## 2.2 Classification methods overview

The classification procedure requires the development of models or classifiers to describe classes of data. Those data classes are assumed to be in relationship with other random variables called features, on which the classification models are trained. Among typical techniques, decision trees, rule-based classification or expert systems can be mentioned. A brief explanation for each is included in this section.

### 2.2.1 Decision Trees

This approach shows the ability to learn decision trees from tuples labelled by classes (the classes have to be categorical). It is shown as a structure similar to a flowchart where different elements appear [13]:

- the *internal nodes*, non-leaf nodes, that denote a test on an attribute,
- the *branches*, that represent the outcomes of a test,
- the *leaf nodes*, terminal nodes that hold a class label.
- the *root node*, the topmost node in the tree.

It works as follows: given a tuple  $C$ , for which the class label is unknown, the attribute values are tested against the decision tree. As a result, a path is drawn from the root to a leaf node, which holds the class prediction for that tuple (Figure 2).

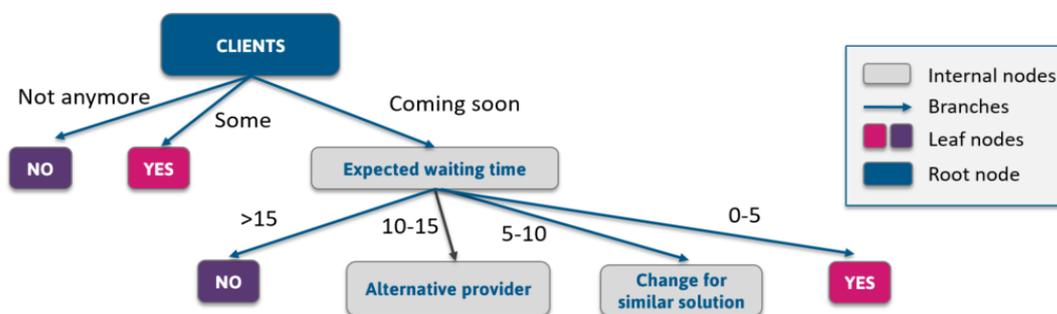


Figure 2. Example of decision tree to determine whether to wait for the supplier to receive the required material. In addition, main components have been highlighted in colours.



Among principal advantages of using them, it is not required any domain knowledge or parameter setting, they can handle multidimensional data, their representations is quite intuitive, they usually have good accuracy, and finally, they are appropriate for exploratory knowledge discovery.

### Attribute selection

This measure makes possible to separate a given data partition of class-labelled tuples into individual classes. The best splitting criteria is the one that most closely leads to a scenario. The splitting rules are the base that determine how tuples are going to be separated. The attribute selection measure offers a ranking for each attribute describing the training tuples, and the attribute with the best score for the measure is selected as the splitting attribute for the given tuples. The most popular attribute selection measures include the information gain, the gain ratio and Gini index [13]:

- Information Gain. It is based on the work of Shannon on information theory, which studied the value of “information content” of messages. The highest gain attribute is the splitting gain attribute for the node, and it minimizes the information needed to classify the tuples in the resulting partitions, in order to be the least impure as possible. The information needed to classify a tuple in A is:

$$Info(D) = - \sum_{i=1}^m p_i \log_2(p_i),$$

Where  $p_i$  is the nonzero probability that an arbitrary tuple in D belongs to class  $C_i$  and it is estimated by  $|C_{i,D}|/|D|$ . Info (D) is also known as the entropy of D. Info gain is defined as the difference between the original information requirement and the new one. It indicates how much would be gained by branching on A.

$$Gain (A) = Info(D) - Info_A(D)$$

- Gain ratio. It is defined in the following way:

$$Gain Ratio (A) = \frac{Gain (A)}{Split Information_A (D)}$$

The splitting attribute selected will be the one with the maximum gain ratio. It must be taken into account that as the divided information approaches 0, the ratio becomes unstable.

- Gini index. It measures the impurity of D, a data partition or set of training tuples as:

$$Gini(D) = 1 - \sum_{i=1}^m p_i^2,$$

Where  $p_i$  is the probability that an arbitrary tuple in D belongs to class  $C_i$  and is estimated by  $|C_{i,D}|/|D|$ . The sum is computed over m classes. Each attribute is splitted in a binary way. In this situation, a weighted sum of the impurity of each resulted partition is computed, as it can be seen in the following example.

$$Gini_A(D) = \frac{|D_1|}{|D|} Gini(D_1) + \frac{|D_2|}{|D|} Gini(D_2)$$

For a discrete-valued attribute, the subset that provides the minimum Gini index for the considered attributed is selected as the splitting subset. In case of continuous-value attributes, it is common to establish the set point in the middle of the interval values.

To sum up, each of the described indexes has its own bias: information gain tends to prefer unbalanced splits in which one partition is much smaller than the others are. On the other hand,



Gini index prefers multivalued attributes **as well, but also** has difficulty when the number of classes is large. **Gain ratio reduces the bias exhibited by the information and Gini index measures, but ultimately the best choice will often depend on the data.**

### Main issues in decision tree learning

During the process of designing a decision tree, some of the questions that need to be answered are how far to extend the tree, what are the metrics for the gradual selection of attributes, how to process training data if any information seems incomplete, or how to be able to achieve good running speed. The basic ID3 algorithm solves many of these issues, as explained below [14].

- Avoid overfitting the data: a hypothesis overfits training data in case another hypothesis that fits the training data less well performs better over the entire distribution of instances. This situation usually take place when the training examples contain random error or noise, however, it can also be possible even when the training data is noise-free, especially when small numbers of examples are associated with leaf nodes. To avoid it, there are different approaches:
  - Stop growing the tree earlier, before it reaches the point where it perfectly classifies the training partition (generally, it is difficult to find when to stop growing the tree).
  - After the tree fits the data, post-prune it (seems to be more successful in practice). In this case, the most common procedure is to separate a set of samples, other than those included in the training set, to evaluate the utility of post-pruning the nodes. This is the common technique of using a training and validation set.
- Incorporating Continuous-Valued attributes: first approach consists in restringing attributes to a discrete set of values. Target attribute whose value is predicted by the learned tree must be discrete. Additionally, the attributes tested in the decision nodes must be also discrete-valued. In order to consider continuous-valued decision attributes, new discrete attributes that partition the continuous attribute value into a discrete set of intervals must be included. The algorithm can dynamically create a new Boolean attribute that is true if  $A < B$ , or false otherwise. Threshold should be defined considering that the greatest information gain is obtained.
- Alternative measures for selecting attributes: the gain measure favours attributes with many values over those with few values. One option to avoid this problem consists in selecting decision attributes based on some measure, such as the gain ratio, which penalizes attributes in the situation described, by incorporating a term called split information, sensitive to how broadly and uniformly the attribute splits the data. The Split Information term represents the entropy of  $S$  respect to the values of attribute  $A$ , and it was defined in the previous section.
- Handle training examples with missing attribute values: sometimes data is not complete, and it is common to estimate the missing attribute value based on other examples for which the attribute has a value. One strategy for solving this is to assign the most common value among training examples. A second and more complex procedure consists of assigning a probability to each of the possible values that it can take, based on the observed frequencies for it among the examples in the node.



- Handle attributes with differing costs: in some tasks, the instance attributes have associated costs. This can be considered by introducing a cost term into the attribute selection measure, for example, by dividing the Gain by the cost of the attribute, so that the lower-cost attributes are preferred. Cost-sensitive measures do not guarantee finding for an optimal cost-sensitive decision tree, but they do work in favour of low-cost attributes.

## 2.2.2 Rule-Based Classification

In rule-based classifiers, the learned model is represented in form of IF-THEN rules. An IF-THEN rule is an expression on the form

*IF condition THEN conclusion*

As example, rule R1 can be defined as follows:

*R1: IF Isolated= false and thickness>0 THEN isolation\_possibility= true*

*R1: (Isolated= false)  $\wedge$  (thickness>0)  $\rightarrow$  (isolation\_possibility= true)*

-The IF part of the rule (left side) is known as the antecedent or precondition. If the condition holds true, the antecedent is satisfied and the rule covers the tuple.

-The THEN part of the rule (right side) is the rule consequent.

The antecedent contains one or more attribute tests, and the consequent contains a class prediction (in this scenario, it decides if an isolation measure can be applied) [13].

One of the inference strategies used more often to obtain complex conclusions is the chain of rules procedures. It can be used in case the premises of certain rules coincide with the conclusions of others. When the rules are chained, original facts can create new facts, and this is repeated successively until no further conclusions can be obtained.

### **Rule extraction from a decision tree**

Decision trees can become large and difficult to understand, that is why sometimes it can be a good idea to translate them into an IF-THEN classifier. This other strategy seems easier for humans to interpret, especially in those cases when trees become too large. To extract them, one strategy is to create a rule for each path from the root to a leaf node. Each splitting criteria along the path is logically ANDed to form the rule antecedent (IF). The rule consequence (THEN) is placed on the leaf node, which holds the class prediction.

As some rules are not mutually exclusive after a pruning process, a class-base ordering scheme can be adopted, grouping all rules for a single class, and then, determining a ranking of the class rule sets [13].

### **Metrics for performance evaluation**

In order to know how “accurate” the classifier behaves, different metrics can be calculated. Following previous sections, tuples can be seen as positive, if they represent the main class of interest, or negative (the other ones). Playing with this idea, four terms can be defined [13]:



- True positives (TP): positive tuples that were correctly labelled by the classifier. TP represents the number of true positives.
- True negatives (TN): negative tuples that were correctly labelled by the classifier. TN represents the number of true negatives.
- False positives (FP): negative tuples that were incorrectly labelled as positive by the classifier. FP represents the number of false positives.
- False negatives (FN): positive tuples that were incorrectly labelled as negative by the classifier. FN represents the number of false negatives.

These four terms define the confusion matrix, which is a useful tool for analysing how well the classifier can recognize tuples of different classes. TP and TN talk about good performance, besides, FP and FN talk about wrong behaviour.

Table 2. Confusion matrix definition.

		Predicted class		
		yes	no	Total
Actual class	yes	TP	FN	P
	no	FP	TN	N
	Total	P'	N'	P+N

Finally, different evaluation measures can be considered. Starting with accuracy, it can be calculated as the percentage of test set tuples that are correctly classified. On the other hand, the error rate can be seen as (1-accuracy).

$$accuracy = \frac{TP + TN}{P + N} \quad error\ rate = \frac{FP + FN}{P + N}$$

The precision and recall measures are also widely used in classification. Precision is related to exactness, that means the percentage of tuples labelled as positive which are actually such, whereas recall is a measure of completeness, which represents which percentage of positive tuples are labelled as such.

$$precision = \frac{TP}{TP + FP} \quad recall = \frac{TP}{TP + FN} = \frac{TP}{P}$$

An alternate way to use these two last measures is by combining them to obtain the F measure (F-score). It represents the harmonic mean of precision and recall. It gives equal weight to both of them. Its highest possible value is 1, which means perfect precision and recall, and its lowest possible value is 0.

$$F = \frac{2 \times precision \times recall}{precision + recall}$$

### 2.2.3 Expert System

The use of expert systems is of special interest in the field of building pathology. Its structure and logic is explained in the following lines. It consists in a programmed system that provides adequate



answers to questions of the user in the concerned field of knowledge, after analysing a great amount of data concerning the problem [15].

Professor Edward Feigenbaum of Stanford University, one of the pioneers of expert systems technology, defined an Expert System as “an intelligent computer program that uses knowledge and inference procedures to solve problems that are difficult enough to require significant human expertise for their solutions” [16].

That means that an expert system is an application that performs a task that, in other case, would be performed by a human expert, emulating their decision-making ability. In this context, emulate means that the expert system is intended to act in all respects like a human expert, which represents a much stronger concept than trusting on a simulation, which only acts like a human in some of them. This approach represents a branch of Artificial Intelligence where there is a need of establishing how experts would take decisions in a given situation, considering special skills that are representative of them.

Internally, an expert system consists of two main components: the knowledge base and the inference engine. The knowledge base contains the expert knowledge with which the inference engine extracts conclusions. Components included in the architecture and its principal functionality are explained below [15] [17]:

- Inference engine, which can be called the brain. It is responsible for analysing and interpreting the knowledge stored in the knowledge base, and makes inferences by deciding which rules are satisfied by facts or objects, prioritizes the satisfied rules, and executes rules according to the established priority.
- User interface, which makes it possible to establish communication between the system and the users.
- Explanation facility, which explains the reasoning of the system to a user.
- Knowledge acquisition facility, which represents an automatic way for the user to enter knowledge in the system rather than by having the knowledge engineer explicitly code the knowledge.
- Knowledge base, which contains facts, statements and relations among data
- Working memory, which is a database of facts used by the rules.

A last element called the *Agenda* can be included, which contains a prioritized list of rules created by the inference engine, whose patterns are satisfied by facts or objects in the working memory [18]. It is updated regularly in order to reduce the number of comparisons needed to be made between rules and facts in the database. This architecture is shown in Figure 3.

Among the advantages of using an expert system, it can be highlighted that it can speed up the process of diagnosis, by providing increased availability, reduced cost, reduced danger, permanence or fast response, and can also be used as an advisory tool for no experts (knowledge of multiple experts can be made available to work simultaneously and continuously to solve problems, in an increasingly reliability framework). Even more, the expert system can explicitly explain in detail the mechanisms followed to extract a conclusion. Direct benefits are evident, however, its development may suppose a long process, and sometimes it is difficult to understand their complex definition if you are not an expert [15] [16].



In this scenario a diagnosis (or classification) expert system could be developed, with the aim of advising the user. Base knowledge would be the experience, considering information is not as complete as desired to determine the best set of solutions.

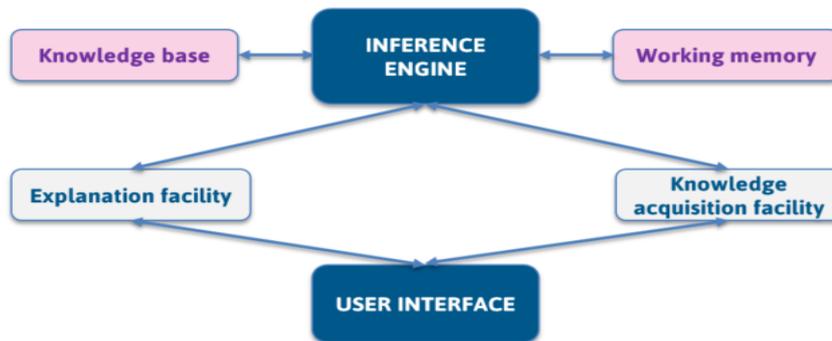


Figure 3. Architecture of an expert system (own elaboration).



### 3. Building pathologies

In this section, a number of existing databases of pathologies is presented, with the aim of having a reference on the most frequently found in buildings, and which information is available. First document analysed, CIB report (International Council for Research and Innovation in Building and Construction) [15], mentions a real need to recover information from experience and knowledge: [...] *the necessity of collecting, recording and evaluating data, to cost/benefit analysis and to providing information to involved bodies like: regulations and code makers, designers, contractors, implementers of quality assurance systems, insurance companies, planners, etc. Such output can be quite different for different users of the information. It mainly comprises: number and/or frequency of several specific defects, actual causes, characteristics of the degradation process, losses or costs involved and appropriate remedial and/or preventive measures.*

There are some examples of attempts to create a database or fault catalogue that collects this information. However, not all of them are open, and they are not presented in the same format. CIB Committee W086 (Building Pathology Commission) started this in their 1993 report, defining a possible methodology to follow, which is included in Figure 4. Their main objective was to produce information that can assist in the diagnosis and prevention of significant defects and failures in the design, construction and use of buildings, and additionally, consider technical aspects of defects [19].

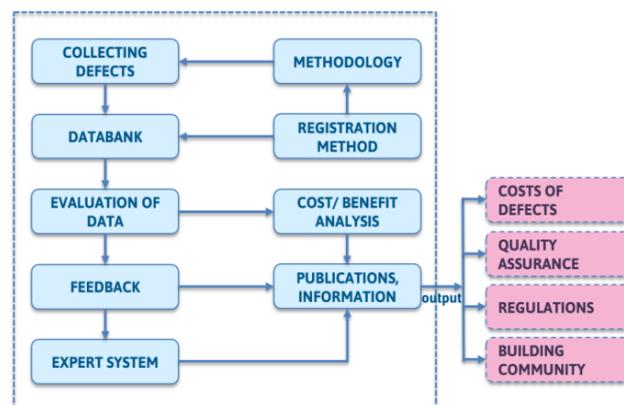


Figure 4. Elements of the Building Pathology domain for several types of defects (own elaboration) [15].

CIB defines a pathology as *the systematic treatment of building defects, their causes, their consequences and their remedies*. On the other hand, ISO13686-1:2000 (E) explains a defect is a *fault or deviation from the intended level of performance of a building or its parts*, considering the levels of performance in terms of: mechanical resistance and stability; safety in case of fire; hygiene, health and the environment; safety and accesibility in use; protection against noise; energy economy and heat retention; and sustainable use of material resources [20]. To sum up, the term pathology is the systematic study of diseases with the aim of understanding their causes, symptoms and treatment, and in the context of buildings, it



requires a detailed knowledge of how they are designed, constructed, used and changed, and how environmental conditions can affect them [19].

Otherwise, a building defect can be considered as a failing or shortcoming in the function, performance or user requirements of a building, and might manifest itself within the structure, fabric, services or other facilities of the affected building [19].

### 3.1 Existing pathology databases

The following websites offer information about pathologies in buildings. Not all are accessible for free, and the most relevant aspects related to each of them are described.

- **Technische ABC-lijst, Woningborg** (Netherlands), <http://www.technische-abc.nl/>

This database mainly contains attention points and recommendations for building designers and building constructors, but not many typical pathologies are described. The searching tool allows to look for information by a set of predefined fields (construction products, regulatory aspects, design features). A license is required to create a user account.

- **Danish Building Defects Fund** (Denmark) <https://www.bsf.dk/om-fonden/in-english/general-information/>

This website belongs to a privately owned institution established in 1986, same year of the law on public housing in that country. The Fund comprises publicly subsidised housing and also publicly subsidised renovations. The database offers information from 1 to 5 year inspections carried out since 1997 (in the Danish language), and no login is required. Among the aspects included in the inspection, the following building elements can be mentioned: excavation foundations and basement, structural and stabilizing elements, exterior walls, roof constructions, wet room, floor and building drainage, water, heat and ventilation, concrete in an aggressive environment and other elements. The search engine allows to search by type of building, but not by technical aspect, which impedes the query a bit for our purpose.

- **"Schadis – Die Datenbank zu Bauschäden" of Fraunhofer Institut IRB**, [www.irb.fraunhofer.de/schadis/](http://www.irb.fraunhofer.de/schadis/):

This site, offered by Fraunhofer-Informationszentrum Raum und Bau IRB, represents the largest collection in the field of building pathologies in German. It contains information from over 700 books, articles and reports, and can be accessed online under license.

- **The Building Pathology Study Group – PATORREB** (Portugal), [www.patorreb.com](http://www.patorreb.com):

This website provides a pathology catalogue compiled by a set of Portuguese universities. Pathologies can be consulted by location (roof, external wall, basement wall, etc.), and for each of them, a set of typical problems are listed and described. An account is required, otherwise, only an example can be accessed.

- **Agence Qualité Construction- REX BBC** (France): <https://pedagogie-rexpb.qualiteconstruction.com/>:

This agency has developed a database that comprises several pathology reports called SYCODÈS (SYstème de COLlecte des DÉsordres). This website, produced by the AQC and funded by the PRACTE program, the ADEME and the AQC, provides technical resources for trainers and players in high performance construction and renovation. Information is organized in several areas such as opaque walls, glass walls, heating and cooling, DHW, ventilation, lighting and specific electricity, steering, electricity production,



organizational aspects and regulations. Each area contains short articles of good practice examples and links to reports, guides, technical information and reference texts.

Construction errors: Errors during the building construction phase of the process	
<i>Task: on-site</i>	Onsite manufacturing in conflict with the design
	Onsite manufacturing in divergence with the construction programme
	Irregular site inspection by the construction manager
	Assembly of damaged building components
	Incorrect assembly of building components
	Improper installation of windows and doors
	Improper installation of MEP/HVAC systems
	Assembly of building components differing from the technical design
	Lack of appropriate coordination of the work
	Omission or changes of constructive elements
	Incorrect location of the building components
	Geometric problem of the building component (e.g. panels)
	Mistakes during the assembly of the building components (e.g. leakage, water infiltration, damp)
	Flawed execution by the construction workers
	Misinterpretation or incorrect use of the documentation
	Interference with other activity
	Installation of unsuitable material
	Material change
	Documentation without distribution
	Inaccurate use of materials
	Error with measurement and testing equipment
	Incorrect stoking
	Non-qualified personnel (first and subcontractor)
	Broken pipelines
	Climate interference: rain, temperature (e.g. panel faced)
	Coating defect (e.g. panel faced)
	Increased loads (e.g. slab)
Incorrectly sealed (e.g. panel faced, floor, services)	
Ongoing iteration process –model versus realized on site- between expected 100 % quality model and realised model missing	
Crisis intervention model to cover not calculable internal and external influences missing	

Figure 5. Example of report format extracted from the INSITER project, where some common error in buildings are included [21].

Information is also provided in form of reports, such as the document included above, coming from the INSITER project [21], which lists common precast construction errors that affect energy efficiency for the design, construction and maintenance phases of the process. As they are related to energy efficiency, they are the closest to the problem under analysis.

The CIB report in 1993 already highlighted the need to create databases that collect information on pathologies in buildings from the experience and knowledge of the agents involved. However, nowadays the data sources in this term have not reached a consensus in the way of displaying the information, and often they are tools for private use or not free of charge. The different accessibility options depending on the database and the variety in the way the information is presented, which is not aimed at finding information on specific characteristics, means it is still necessary to have experts involved in processes such as the one intended to be carried out in task T7.2.

### 3.2 Pathologies in the context of the available information

In case of energy renovation, three main constructive elements of the building are considered: openings (doors and windows), roofs, and external walls. Qi et al. [22] include technical specifications linked to each of these elements: actions related to the replacement of doors and windows consist in guaranteeing some values for the heat-transfer coefficient and the air ventilation rate; in case of thermal insulation of roofs, control on the kind of material, bulk density, heat-transfer coefficient, thermal conductivity of the insulation, tensile/compressive strength and dimensional stability; finally, for the thermal insulation of exterior walls, similar actions to those of roof insulation are required, by checking the kind of material, bulk density, heat-transfer coefficient, thermal conductivity of the insulation, tensile/compressive strength and dimensional stability.



This strategy fits with the building information that may be available: the data provided in the IFC file regarding materials, dimensions and characteristics of the three main constructive elements on the buildings (windows, walls and roof) can be used in order to perform a thermal analysis of the building. Moreover, regulation is different depending on the country or even in the area in the country (for example, in Spain 5 different climatic zones are considered). All these aspects allow defining the approach to be followed, which is explained in the next section.



## 4. Diagnosis and optimization algorithm using machine learning

Diagnostic aids can be provided in form of equipment, devices and test procedures, in order to collect useful information about the whole system. An alternate means for supporting a building diagnosis can go from a diagnostic chart or table with a list of anomalies, and undesirable range of values and its possible explanation, to a system that automatically makes this verification, and offers a final report. Among the possible bases for these systems, the following ones can be mentioned:

- Diagnostic charts and matrices
- Fault and diagnostic Trees
- Expert systems
- Decision trees/rule based systems

In this context, the chosen approach includes the definition and development of a decision tree that supports the determination of the optimal solution for thermal renovation of buildings, depending on their main features, the possible existing constraints to take into account and regulation in the country where it is located.

### 4.1 Description

The goal of the algorithm is to find the convergence among the use of machine learning techniques and the ability of detecting already existing defects, which need to be solved through deep renovation. The overall concept is shown in Figure 6.

Specifically, the purpose of the designed algorithm has focused its attention on thermal pathologies. Pathology has been defined as the scenario in which a building does not comply with current regulations of a country on thermal transmittance. For the definition of pathologies, the data sources available on the Internet have been verified, however, much of the information required to detect many of the listed problems is not usually provided by the IFC schema. The IFC file has been selected as the most commonly used format for open data exchange to be included in the analysis, in order to provide an automatic mechanism that can be useful and easily integrated with other more sophisticated tools.

On the other hand, it has been necessary also to define the set of Energy Conservation Measures (ECMs) that the algorithm can select to determine an optimal solution. These have been designed taking into account the information given by Tabula WebTool [23], and real market possibilities. Depending on the set of measures to be applied, the algorithm will give an optimal solution, if possible. If not, the set of constraints may be too limiting, and other alternatives should be considered instead. Details about each solution, which are structured in three pillars (windows, walls and roof), will be explained later in this section.



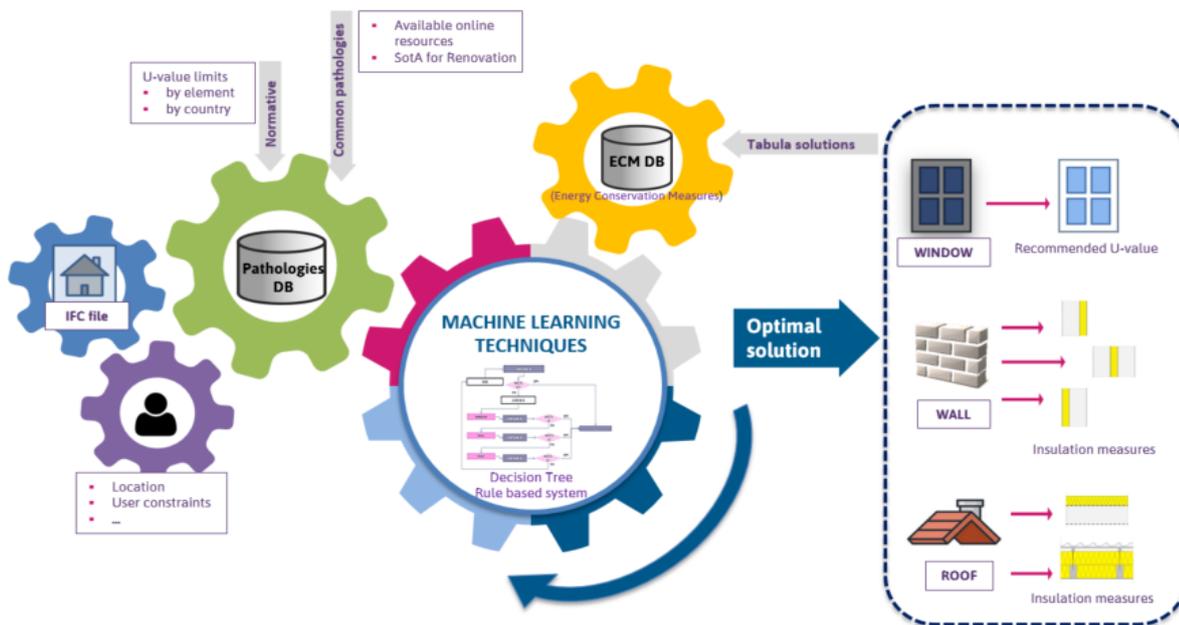


Figure 6. Diagram of the machine learning process for diagnosis and optimisation processes in deep renovation projects.

## 4.2 Data acquisition

As previously described, there will be two main data sources to define the context of the case into analysis: information about the building contained in the IFC model, and other contextual data or restrictions that must be considered. Some of these parameters are asked to be entered by the user providing a simple, user-friendly interface, whereas the rest is automatically read from the IFC file. IFC4 is supported by the diagnosis tool and future versions of IFC will be considered in next versions of the tool.

The Industry Foundation Classes (IFC) represents a data model shared among different actors in a project. It is a standardized, digital description of the built environment, including buildings and civil infrastructure. It is an open, international standard (ISO 16739-1:2018), vendor-neutral, and usable across a wide range of hardware devices, software platforms and interfaces for many different use cases. Its schema comes from the primary technical deliverable of buildingSMART International, with the aim of promoting openBIM. It codifies in a logical way: the identity and semantics, the characteristics/ attributes and relationships of objects, abstract concepts, processes and people [24].

The way to capture the IFC information is through the IfcOpenShell library. IfcOpenShell is an open source library, which helps users to work with the IFC file format. It can be used to edit, add new codes to the .ifc file and to export information. IfcOpenShell also uses Open CASCADE (the Open CASCADE Community Edition) internally to convert the implicit geometry in IFC files into explicit geometry that software CAD can understand. It is used in python language [25].

-The relevant parameters read from the IFC file have been carefully selected to perform the thermal analysis, considering them as elements commonly included in the file with the building model, with the aim of extending the use of the approach. Some of them are very basic, such as windows, walls



and roof dimensions, materials, and its relationships. The most critical one is the U-value (heat transfer coefficient) related to the element, which is considered as key factor to perform the analysis.

Specific parameters related to different IFC type elements are not explained in detail, in order to avoid overlaps with the information provided in other deliverables. D5.2 includes the Model View Definition (MVD), and there the parameters associated to walls, windows, roofs, etc. are explained.

-The information that is requested from the user is related to the constraints to take into account before selecting the measures. For example, in case of listed buildings, it is often not permitted to thermally insulate the façade. In another hypothetical case, it may be necessary to restrict wall interventions only to the external side of the building, due to the available space and to avoid overly intrusive activity for end users

Figure 7 shows the user-interface for the collection of the constraints of the case under analysis and launching the calculation process.

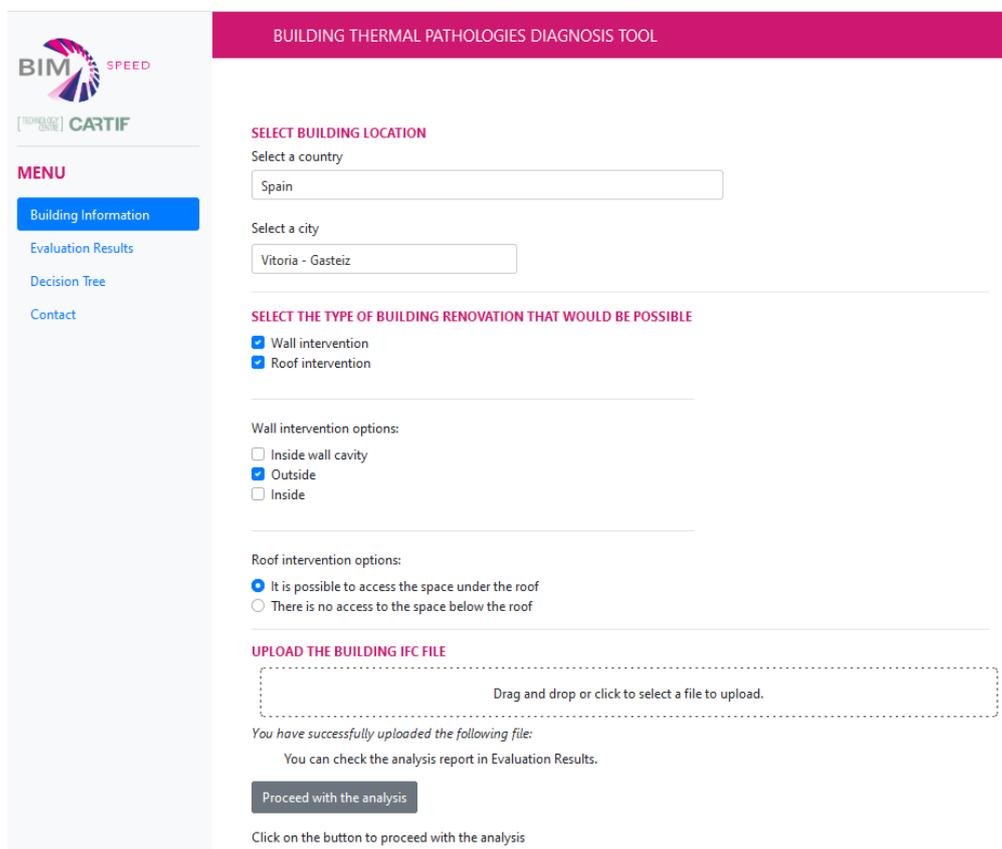


Figure 7. User interface of the diagnosis tool to request information and proceed with the analysis.

### 4.3 Design of the algorithm

Once the sources of information on pathologies have been explored and their convergence with the possibilities offered by the IFC file, the next step is to define the problems to detect and solve. After



that, measures to follow to address each situation need to be defined. Finally, how it can contribute to the enrichment of the design rules in renovation projects must be analysed.

#### 4.3.1 Pathologies definition according to algorithm design

Pathologies have been defined as the situation in which the building cannot comply with regulations in force in the country regarding its thermal properties. The algorithm has been designed to be able to check if the values for the transmittance of the thermal envelope are in the desired range. This verification is done at two different levels: for each constructive element separately (window, wall, roof), and the building as a whole.

Range values considered for each country come from national regulations. The thermal envelope of the building should comply with a set of conditions. For example, the thermal transmittance of each element that belongs to the thermal envelope should not exceed a certain value. Table 3 includes what data sources are used to define these limits.

Table 3. Regulation limits data sources considered for each country under study.

Country	Regulation
Spain	Basic document HE Energy Saving- Technical Building Code (2019)
Germany	Energy Saving Ordinance-EnEV (2013)
Poland	Implementation of the EPBD in Poland (2016)
France	Order related to the energy performance of existing buildings, when they are subject of major renovation work (2008)

The second level validation is related to the global coefficient of heat transmission through the thermal envelope of the building (K), or a part of it. Similarly, the targeted value should not exceed the limit given in regulation. This parameter can be calculated as follows:

$$K_{AVE} = \frac{\sum_i^{windows} A_i \cdot U_i + \sum_j^{walls} A_j \cdot U_j + \sum_k^{roofs} A_k \cdot U_k}{\sum_i^{windows} A_i + \sum_j^{walls} A_j + \sum_k^{roofs} A_k}$$

As an example, the Technical building Code in Spain determines a limit K value for private residential use, that depends on the compactness of the building and the climatic zone where it is located (Figure 8).

	Compacidad V/A [m³/m²]	Zona climática de invierno					
		α	A	B	C	D	E
<b>Edificios nuevos y ampliaciones</b>	V/A ≤ 1	0,67	0,60	0,58	0,53	0,48	0,43
	V/A ≥ 4	0,86	0,80	0,77	0,72	0,67	0,62
<b>Cambios de uso. Reformas en las que se renueve más del 25% de la superficie total de la envolvente térmica final del edificio</b>	V/A ≤ 1	1,00	0,87	0,83	0,73	0,63	0,54
	V/A ≥ 4	1,07	0,94	0,90	0,81	0,70	0,62

Figure 8. K limit values [W/m²·K] for private residential use in Spain.



In the case of France, the K value is not given as it may depend on the characteristics of different elements of the building. The regulation considered (*Order related to the energy performance of existing buildings, when they are subject of major renovation work*) suggests that the K value ( $U_{bât}$ ) cannot exceed the maximum coefficient loss through the walls and openings of the building ( $U_{bât-max}$ ), which is determined according to the use of the building, the coefficient of basic losses through the elements of the building ( $U_{bât-base}$ ), and the coefficient  $C_{td}$ , defined as follows:

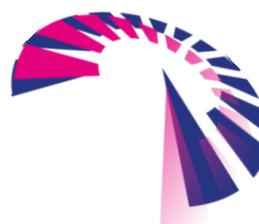
$$U_{bât-max} = U_{bât-base} \cdot 1,25 \cdot C_{td}$$

$U_{bât-base}$  is calculated as follows, where lowercase coefficients represent predefined constant values, and uppercase variables are associated to areas (A) or linear parts (L) of the building, as described in previously mentioned normative.

$$U_{bât-base} = \frac{a_1A_1 + a_2A_2 + a_3A_3 + a_4A_4 + a_5A_5 + a_6A_6 + a_7A_7 + a_8L_8 + a_9L_9 + a_{10}L_{10}}{A_1 + A_2 + A_3 + A_4 + A_5 + A_6 + A_7}$$

The algorithm performs the validation of these parameters following the flow depicted in Figure 9. The sequence of activities has been designed giving priority to the least intrusive measures. That is why it starts with the validation of windows, continuing with walls, in case no solution has been found, and finally, considering the renovation of roofs. The same criteria is applied to structure the sequence of measures proposed for the walls, starting with the possibility of inserting insulation in the air cavity, continuing with the insulation on the exterior surface, or finally, installing insulation on the inside of the building. In case of roofs, the criteria may seem different, as in case of a sloped roof it is preferable to install insulation inside. This intervention has been considered much easier to execute than the one in which the insulation is placed on the outside of the roof. Installing the insulation on the outer layer of the roof is considered for the rest of the alternatives.

- If the global coefficient of heat transmission through the thermal envelope exceeds the given limit, the heat transmission coefficient of the windows is checked. The first correction would be to use a new window element, with a heat transmission coefficient according to the respective regulation. If the problem remains unsolved, a new iteration is performed, reducing the U-value of the window by 10%. Again, if the desired limit is still not reached, a second iteration is carried out, now considering a much bigger improvement that will be explained later in the solutions catalogue. This second iteration is conditioned by real market solutions, where it can be verified that there is a significant jump in the value of the heat transmission coefficient of the windows depending on whether they incorporate double or triple glass.
- If the window renovation is not sufficient to address the problem and an intervention on the wall is possible, the algorithm continues to check the U-values of the walls. The process at this level is quite similar to previously mentioned problem existing for windows, but in this case, it is almost impossible to transform the U-values of the walls exactly to the limit value. Therefore, at this stage, two



possible measures are given for each kind of intervention. If all the solutions are applicable, the validation sequence would be to insert insulation at the wall cavity, or the outer or internal layer.

- It is possible to select only one solution of this type, two or the three of them, always maintaining the prioritization sequence. As mentioned above, there may be a number of reasons to avoid considering any of them.
  - Cavity insulation: the algorithm first checks the presence of an air cavity and its thickness (has to be between 40 and 100 mm in order to consider it viable). In case it exists, first iteration would consider filling the entire cavity with a material with certain conductivity characteristics available in the market. A second iteration would be carried out if desired levels are not achieved, this time considering a second material with a lower conductivity.
  - Insulation on the external layer: the algorithm proposes an intervention with a specific material, with specific thickness and conductivity characteristics according to the market options and the typical thickness values used in the country. If the desired limits are still not reached, a second iteration includes an improved performing solution, changing the used material, this time selecting one with a lower conductivity.
- Again, if there is still not a solution, the algorithm iterates once again, this time changing the thickness of the material: the same materials considered before are included in the third and fourth iteration.
  - Insulation on the internal side: the approach is similar to that followed for the insulation on the external layer. The algorithm proposes an intervention with a specific material, with a specific thickness and conductivity characteristics according to market options and the typical thickness values used in the country. If desired limits are still not reached, a second iteration includes an improved performing solution, changing the material used, this time selecting one with a lower conductivity.
- If there is still no solution, the algorithm iterates once again, this time changing the thickness of the material: the same materials considered before are included in the third and fourth iteration.
- If the insulation of the wall is not enough to address the problem and an intervention on the roof is possible, the algorithm continues to check the U-values of the roof. The process at this level is quite similar to the one for walls. The algorithm first checks if the building has a flat or sloped roof, whereby two different paths are drawn depending on this feature.
  - Pitched roof: in case there is a slope, the first proposed intervention is to place insulation on the internal side of the roof (this requires access under the roof). A second iteration proposes an external measure: as explained above, it can be considered more intrusive although it is performed at the external side of the building.
  - Flat roof: in case the roof is flat, only external measures are proposed, taking into account a specific material, with specific thickness and conductivity characteristics according to market options and the typical thickness used in the country. The second iteration, again offers an improvement in thermal terms.



- The algorithm has been defined so that it is not possible to continue with roof renovation without renovating the wall before.

If the set of constraints is so broad that it considerably limits the set of measures to apply, the algorithm may not find a solution. In that case, the output will encourage the user to launch the process again, considering some changes in the requirements. The solution obtained will be different depending on the country, since the range of limits in the regulation differs substantially.

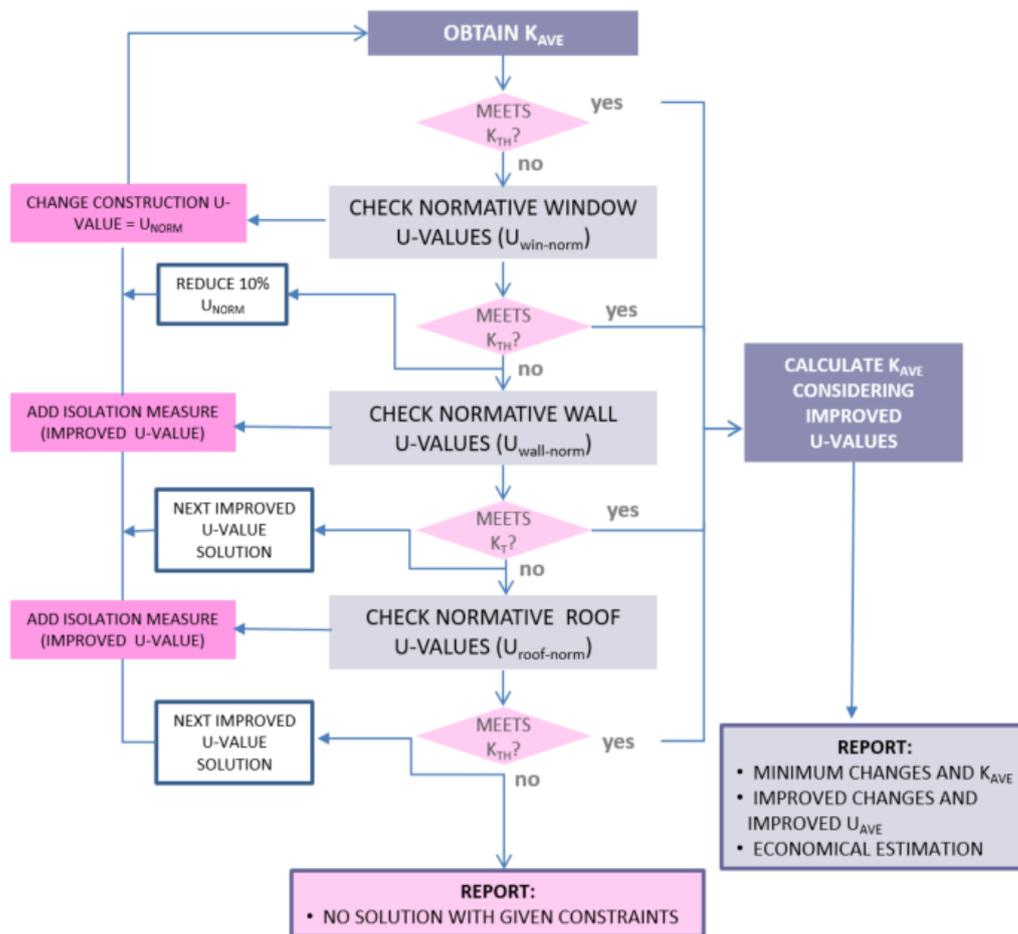


Figure 9. Flow diagram for the designed machine learning algorithm.

This flow of operations is executed through a decision tree that is able to determine the best solution according to predefined constraints and the main features of the building. Then, taking into account the steps outlined above, the tree is configurable (it depends on the constraints related to walls and roof), and the total number of possible solutions is 25 (including the *no solution* output). Possible outputs are summarized in Table 4. The data used to train the model has been created synthetically, and considers the thermal reference parameters included in the flow diagram, to follow the sequence of operations established for the decision. It is a dataset generated for a proof of concept, and the interest would be to enrich and expand it. The next steps to take starting from the prototype have not yet been decided. Its exploitation in a collaborative way could contribute to the creation of a public data set, where



users gave the information in exchange for being able to use the tool at no cost, providing new useful information to the community.

Table 4. Decision tree outputs.

Case	Number of solutions	Details about solutions
Tree ends in Window	Up to 3 + <i>no solution</i>	
Tree ends in Wall	Up to 10 + <i>no solution</i>	- 2 for insulation in cavity - 4 for external insulation - 4 for internal insulation
Tree ends in Roof	Up to 2 + <i>no solution</i>	<i>Pitched roof:</i> - 1 for internal insulation, 1 for external insulation <i>Flat roof:</i> - 2 for external insulation

#### 4.3.2 Solutions catalogue definition

A research of measure solutions applicable on walls and roofs has been carried out. The Tabula WebTool [23] has been carefully revised to provide a global overview of typical actions performed in the countries under study. A summary table for each location is included in the APPENDIX 1 – RENOVATION MEASURES. Measures have been considered for residential buildings (multi-family house and apartment block), classified by the age of the building.

This information has been adapted to real solutions, considering commonly used products on the market, whereby a cost estimate is also provided. The following lines provide some clarifications for the most common measures of each country:

- **SPAIN:** in this context, the solutions provided for walls include the possibility of adding insulation on the interior surface of the building, on the exterior surface (through ETICS), as well as the cavity inside the wall. ETICS is an easy technology, and all the works are made on site: generally, EPS (expanded polystyrene) has a good performance, but the drawback is that it has poor fire performance (E fire class). Due to this poor fire properties, it is recommended to use it on the buildings until 25 m above the ground. Moreover, the solution is relatively light and no reinforcing is needed.

Considering this information, and the material thicknesses available on the market, the range values for them have been defined as shown in Table 5. Roofing solutions for roofs are also determined by these thickness limits.

- **FRANCE:** no detailed information has been found in Tabula (incomplete). That is why the same approach as followed for Spain has been considered.
- **GERMANY:** the solutions listed in Tabula include the possibility of adding insulation on the interior or exterior surface of the building. However, cavity wall insulation is also considered as the algorithm checks in advance if there is a space available. The thicknesses of the materials for the insulation



on the exterior surface are different from those considered in Spain. They have been adapted to the solutions available in the market (the same for the roof insulation). Table 6 depicts this information.

- **POLAND:** according to Tabula solutions, it is not clear if it is typical to apply the three types of measures considered for walls (not indicated). However, most common renovation technique used in Poland is by adding insulation on the external walls with ETICS. Insulating buildings from inside is not very common, it is used mainly on the buildings that are under cultural heritage protection. When it comes to the cavity, the number of buildings where it can be applied is rather low. Therefore, all three have been included, and material thicknesses have been adapted to the range values included here, both for walls and roofs. Detailed values are included in Table 7.

Solutions for windows have also been explored. However, they are easier to characterize, as, depending on different configurations, expected U-values can be achieved. They are also accompanied by an economic estimate.

Table 5. Proposed material solutions and costs for walls and roof (Spain and France).

Country	Level	Category	Type	Insulation						Finishing layer		Total values ECM			
				Thickness insulation [mm]	Material	U-value [W/(m <sup>2</sup> ·K)]	Thermal resistance [m <sup>2</sup> ·K/W]	Conductivity [W/(m·K)]	Cost [€/m <sup>2</sup> ]	U-value finishing layer	Cost [€/m <sup>2</sup> ]	Total U-value (Insulation + Finishing layer)	Total thickness (Insulation + Finishing layer) [mm]	Total cost [€/m <sup>2</sup> ]	
SPAIN	Façade	Insulation at the exterior	ETICS (50mm Cement/mortar plaster + Insulation)	60	Extruded polystyrene	0.556	1.8	0.036	10.24 €	1.6	20.05 €	0.412	65	30.29 €	
				60	Mineral Wool	0.571	1.75	0.034	19.61 €	1.6	20.05 €	0.421	65	39.66 €	
				100	Expanded Polystyrene	0.380	2.63	0.038	23.26 €	1.6	20.05 €	0.307	105	43.31 €	
				100	Extruded polystyrene	0.357	2.8	0.036	14.14 €	1.6	20.05 €	0.292	105	34.19 €	
		Insulation in the cavity wall	Insufflating thermal insulation inside the existing air chamber	60	Polyurethane foam	0.592	1.6901408	0.0355	7.13 €	-	0.00 €	0.592	60	7.13 €	
					Mineral wool	0.575	1.7391304	0.0345	11.14 €	-	0.00 €	0.575	60	11.14 €	
		Internal insulation	Plasterboard finish (Insulation+15 mm Plasterboard)	60	Extruded polystyrene	0.571	1.75	0.034	13.89 €	5.3333	17.49 €	0.516	75	31.38 €	
					Expanded Polystyrene	0.526	1.9	0.032	8.71 €	5.3333	17.49 €	0.479	75	26.20 €	
					Extruded polystyrene	0.357	2.8	0.036	22.32 €	5.3333	17.49 €	0.335	115	39.81 €	
					Expanded Polystyrene	0.317	3.15	0.032	11.42 €	5.3333	17.49 €	0.300	115	28.91 €	
		Pitched roof	External insulation	Insulation of cavity between rafters	100	Extruded polystyrene	0.380	2.6315789	0.038	24.66 €	-	0.00 €	0.380	115	24.66 €
						Mineral wool	0.370	2.7	0.038	21.94 €	5.3333	17.49 €	0.346	115	39.43 €
	Flat roof	External insulation	60 mm of insulation, geotextil filter and 100mm gravel	60	Extruded Polystyrene	0.567	1.765	0.034	13.32 €	0.87	44.77 €	0.343	160	58.09 €	
					Extruded Polystyrene	0.360	2.778	0.036	33.30 €	0.87	44.77 €	0.255	200	78.07 €	



Table 6. Proposed material solutions and costs for walls and roof (Germany).

Country	Level	Category	Type	Insulation						Finishing layer		Total values ECM			
				Thickness insulation [mm]	Material	U-value [W/(m <sup>2</sup> .K)]	Thermal resistance [m <sup>2</sup> .K/W]	Conductivity [W/(m.K)]	Cost [€/m <sup>2</sup> ]	U-value finishing layer	Cost [€/m <sup>2</sup> ]	Total U-value (Insulation + Finishing layer)	Total thickness (Insulation + Finishing layer) [mm]	Total cost ECM [€/m <sup>2</sup> ]	
GERMANY	Façade	Insulation at the exterior	ETICS (50mm Cement/mortar plaster + insulation)	120	Expanded Polystyrene	0.317	3.158	0.038	19.61 €	1.6	20.05 €	0.264	125	39.66 €	
				120	Extruded polystyrene	0.300	3.333	0.036	10.24 €	1.6	20.05 €	0.253	125	30.29 €	
				240	Expanded Polystyrene	0.158	6.316	0.038	39.22 €	1.6	20.05 €	0.144	245	59.27 €	
				240	Extruded polystyrene	0.150	6.667	0.036	20.48 €	1.6	20.05 €	0.137	245	40.53 €	
		Insulation in the cavity wall	Insufflating thermal insulation inside the existing air chamber	60	Polyurethane foam	0.592	1.690	0.0355	7.13 €	-	0.00 €	0.592	60	7.13 €	
				60	Mineral wool	0.575	1.739	0.0345	11.14 €	-	0.00 €	0.575	60	11.14 €	
		Internal insulation	Plasterboard finish (Insulation+15 mm Plasterboard)	60	Extruded polystyrene	0.567	1.765	0.034	13.89 €	5.3333	17.49 €	0.512	75	31.38 €	
				60	Expanded Polystyrene	0.533	1.875	0.032	8.71 €	5.3333	17.49 €	0.485	75	26.20 €	
				100	Extruded polystyrene	0.360	2.778	0.036	22.32 €	5.3333	17.49 €	0.337	115	39.81 €	
				100	Expanded Polystyrene	0.320	3.125	0.032	11.42 €	5.3333	17.49 €	0.302	115	28.91 €	
		Pitched roof	External insulation	Insulation of cavity between rafters	180	Mineral wool	0.200	5.000	0.036	73.08 €	-	0.00 €	0.200	180	73.08 €
			Internal insulation	Plasterboard finish (Insulation+15 mm Plasterboard)	200	Mineral wool	0.185	5.405	0.037	41.52 €	-	17.49 €	0.185	215	59.01 €
	Flat roof	External insulation	120 mm of insulation, geotextil filter and 100mm gravel	120	Extruded Polystyrene	0.283	3.529	0.034	13.32 €	0.87	44.77 €	0.214	220	58.09 €	
			300 mm of insulation, geotextil filter and 100mm gravel	300	Extruded Polystyrene	0.120	8.333	0.036	33.30 €	0.87	44.77 €	0.105	400	78.07 €	



Table 7. Proposed material solutions and costs for walls and roof (Poland).

Country	Level	Category	Type	Insulation					Finishing layer			Total values ECM			
				Thickness insulation [mm]	Material	U-value [W/(m <sup>2</sup> .K)]	Thermal resistance [m <sup>2</sup> .K/W]	Conductivity [W/(m.K)]	Cost [€/m <sup>2</sup> ]	U-value finishing layer	Cost [€/m <sup>2</sup> ]	Total U-value (Insulation + Finishing layer)	Total thickness (Insulation + Finishing layer) [mm]	Total cost ECM [€/m <sup>2</sup> ]	
POLAND	Façade	Insulation at the exterior	ETICS (50mm Cement/mortar plaster + insulation)	120	Expanded Polystyrene	0.317	3.158	0.038	19.61 €	1.6	20.05 €	0.264	125	39.66 €	
				120	Extruded polystyrene	0.300	3.333	0.036	10.24 €	1.6	20.05 €	0.253	125	30.29 €	
				240	Expanded Polystyrene	0.158	6.316	0.038	39.22 €	1.6	20.05 €	0.144	245	59.27 €	
				240	Extruded polystyrene	0.150	6.667	0.036	20.48 €	1.6	20.05 €	0.137	245	40.53 €	
		Insulation in the cavity wall	Insufflating thermal insulation inside the existing air chamber	60	Polyurethane foam	0.592	1.690	0.0355	7.13 €	-	0.00 €	0.592	60	7.13 €	
				60	Mineral wool	0.575	1.739	0.0345	11.14 €	-	0.00 €	0.575	60	11.14 €	
		Internal insulation	Plasterboard finish (Insulation+15 mm Plasterboard)	60	Extruded polystyrene	0.571	1.750	0.034	13.89 €	5.33	17.49 €	0.516	75	31.38 €	
				60	Expanded Polystyrene	0.526	1.900	0.032	8.71 €	5.33	17.49 €	0.479	75	26.20 €	
				100	Extruded polystyrene	0.357	2.800	0.036	22.32 €	5.33	17.49 €	0.335	115	39.81 €	
				100	Expanded Polystyrene	0.317	3.150	0.032	11.42 €	5.33	17.49 €	0.300	115	28.91 €	
		Pitched roof	External insulation	Insulation of cavity between rafters	180	Mineral wool	0.200	5.000	0.036	73.08 €	-	0.00 €	0.200	180	73.08 €
			Internal insulation	Plasterboard finish (Insulation+15 mm Plasterboard)	200	Mineral wool	0.185	5.405	0.037	41.52 €	-	17.49 €	0.185	215	59.01 €
		Flat roof	External insulation	120 mm of insulation, geotextil filter and 100mm gravel	120	Extruded Polystyrene	0.283	3.529	0.034	13.32 €	0.87	44.77 €	0.214	220	58.09 €
					300	Extruded Polystyrene	0.120	8.333	0.036	33.30 €	0.87	44.77 €	0.105	400	78.07 €

### 4.3.3 Enrichment design rules in renovation projects

The measures to be applied through building renovation are very wide. The objective of the designed algorithm is to offer an easy-to-use tool that can help to evaluate different alternatives, taking into account the constraints of the real environment. As previously described, the characteristics of existing buildings vary greatly depending on the country and it is necessary to consider both the most frequent practices, environmental restrictions and the regulatory limits in thermal terms.

The IFC file is increasingly important in the building sector, it represents a shared data model between the different actors in a project. This IFC should be exploited and enriched thanks to the information provided by other data sources related to the building (regulatory data, environmental data, etc.). The evaluation carried out by the tool, makes it possible to consider different scenarios and offer a recommendation to the user as a preliminary step of the beginning of the design process for a deep renovation project. The set of solutions considered is rather limited and restricted to more common practices, since the main aim is not to be precise with this information, but to lay a foundation that can give a first overview of the options. It can be seen as a starting point for other tools focused on scenario optimization (for example, the METABUILD platform).



## 5. Proof of concept

### 5.1 Analysis of the results

In this section, the results from the diagnosis tool are analysed. Three building models have been tested in different locations, in order to extract conclusions about differences in the results.

#### **First model: medium quality building prototype**

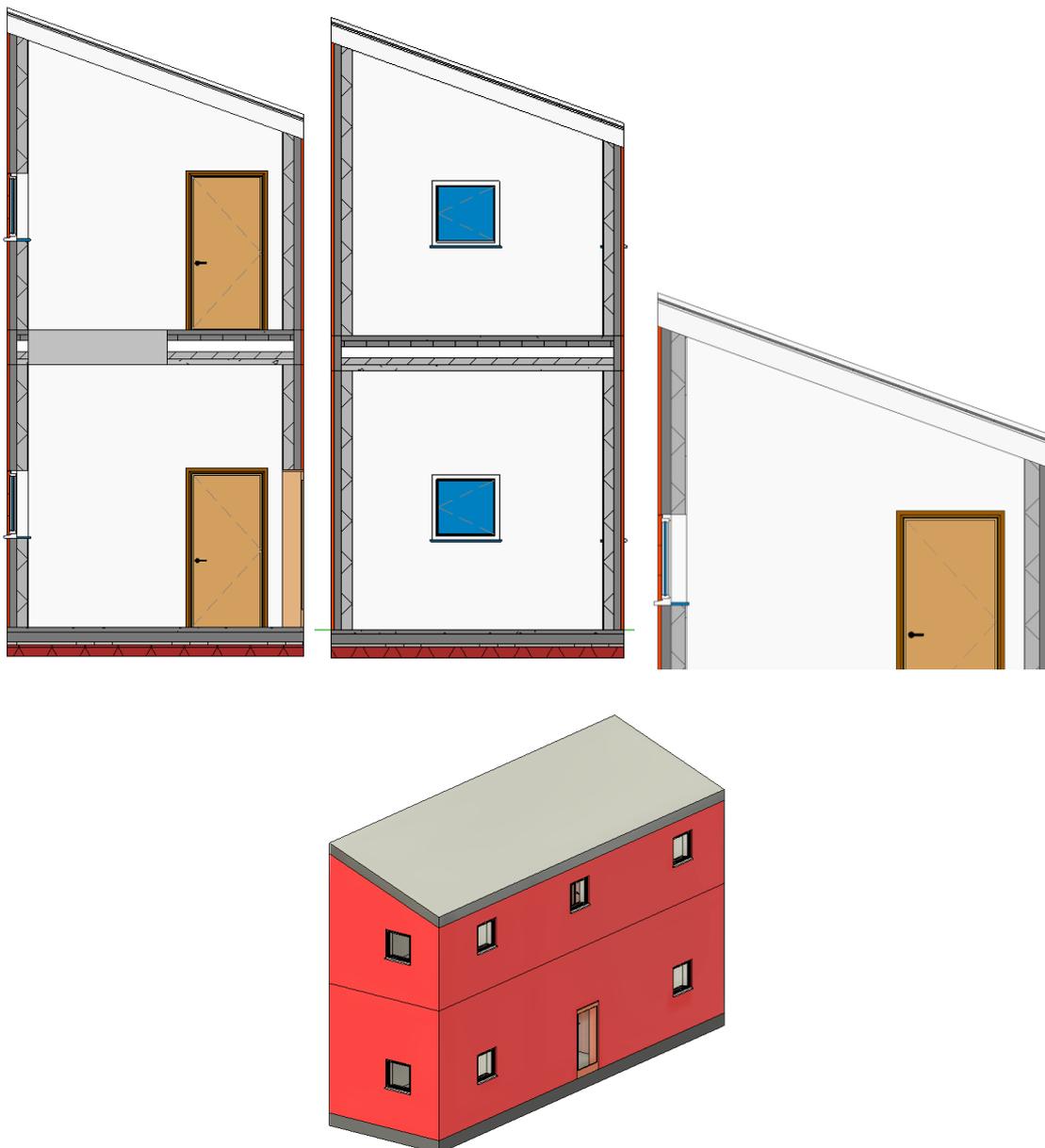


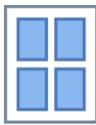
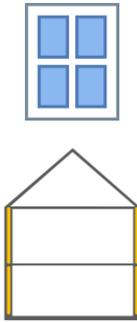
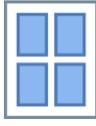
Figure 10. First building model under analysis: medium quality building prototype.



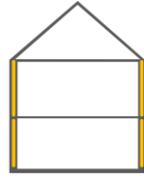
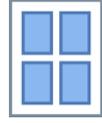
Table 8. Initial values of the Medium quality building prototype.

$K_{AVE}$ [W/(m <sup>2</sup> ·K)]	$U_{window}$ [W/(m <sup>2</sup> ·K)]	$U_{wall}$ [W/(m <sup>2</sup> ·K)]	$U_{roof}$ [W/(m <sup>2</sup> ·K)]	Type of roof	Access to space under roof
0.866	3.689	0.556	1.73	Pitched	Yes

Table 9. Final values of the Medium quality building prototype.

Location	$K_{AVE}$ [W/(m <sup>2</sup> ·K)]	$U_{window}$ [W/(m <sup>2</sup> ·K)]	$U_{wall}$ [W/(m <sup>2</sup> ·K)]	$U_{roof}$ [W/(m <sup>2</sup> ·K)]	Selected solution	Cost [€]
Spain (Málaga, zone A)	0.827	2.7			Double glazing window 	5713.89
Spain (Oviedo, zone C)	0.803	2.1			Double glazing window 	5900.21
Spain (Burgos, zone E)	0.553	1.8	0.217		Double glazing window + Insulation in cavity (0.1m) 	7970.01
Poland	0.784	1.6			Double glazing widows 	6272.86
France	0.557	1.9	0.217		Double glazing window + Insulation in cavity (0.1m) 	7970.01



							
Germany	0.46	1.3	0.123		Double glazing windows + Insulation outside the walls	 	20959.30

As the results show, the proposed outputs differ according to location. In Spain the measures to be applied are the same for the whole country in terms of materials and their thickness, however, warmer areas such as Málaga or Oviedo only require an intervention on the window, while in colder areas such as Burgos, it is also necessary to add insulation in the wall cavity (the result is equivalent for France). In Poland it is sufficient to replace the window by one with a better thermal transmittance value, and for Germany, the requirements are higher, and it is also necessary to apply insulation from the outside.

**Second model: low quality building prototype**

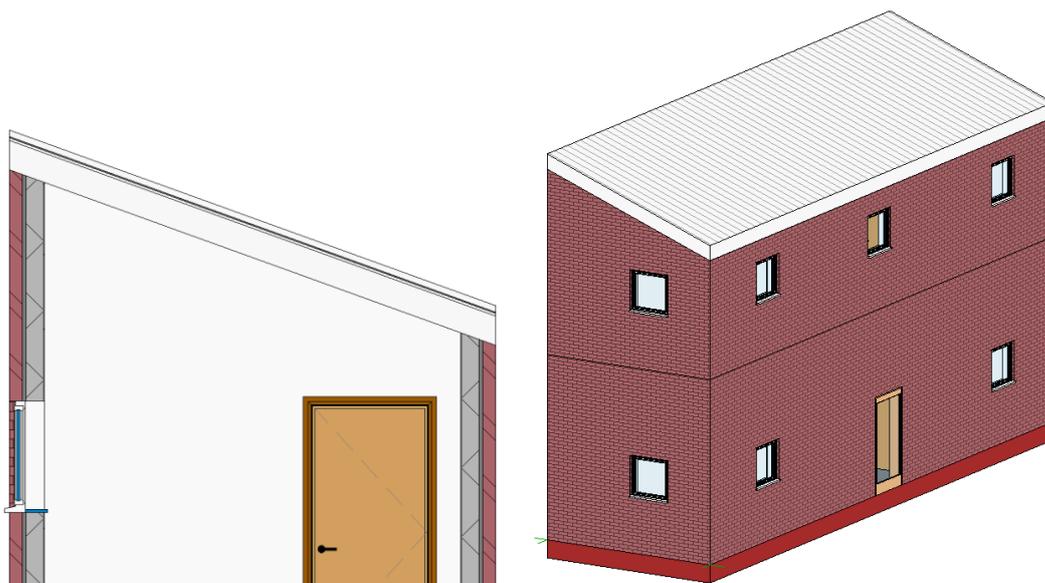


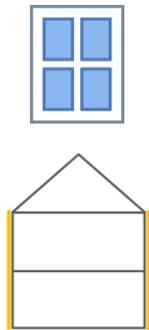
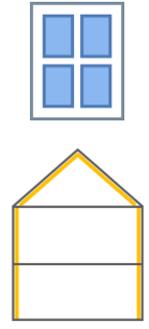
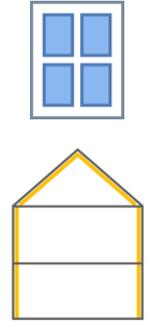
Figure 11. Second building model under analysis: low quality building prototype.



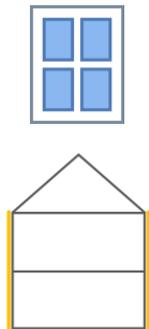
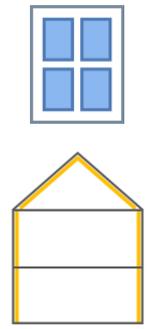
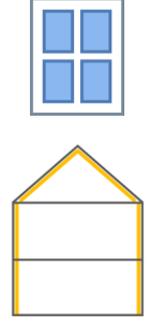
Table 10. Initial values of the Low quality building prototype.

$K_{AVE}$ [W/(m <sup>2</sup> ·K)]	$U_{window}$ [W/(m <sup>2</sup> ·K)]	$U_{wall}$ [W/(m <sup>2</sup> ·K)]	$U_{roof}$ [W/(m <sup>2</sup> ·K)]	Type of roof	Access to space under roof
1.445	5.562	0.827	3.423	Pitched	Yes

Table 11. Final values of the Low quality building prototype.

Location	K value [W/(m <sup>2</sup> ·K)]	$U_{window}$ [W/(m <sup>2</sup> ·K)]	$U_{wall}$ [W/(m <sup>2</sup> ·K)]	$U_{roof}$ [W/(m <sup>2</sup> ·K)]	Selected solution	Cost [€]
Spain (Málaga, zone A)	0.929	2.7	0.26		Double glazing windows + Insulation outside the walls 	16400.24
Spain (Oviedo, zone C)	0.407	2.1	0.336	0.342	Double glazing windows + Insulation inside the walls + Insulation inside the roof 	15055.57
Spain (Burgos, zone E)	0.395	1.8	0.336	0.342	Double glazing windows + Insulation inside the walls + Insulation inside the roof 	15366.11



Poland	0.861	1.6	0.229		Double glazing windows + Insulation outside the walls		16058.60
France	0.399	1.9	0.336	0.342	Double glazing windows + Insulation inside the walls + Insulation inside the roof		15366.11
Germany	0.347	1.3	0.336	0.189	Double glazing windows + Insulation inside the walls + Insulation inside the roof		18264.02

As can be seen in the outputs for the low quality model with the sloped roof, the proposed solutions differ depending on the location. In warm areas of Spain like Málaga, it is sufficient to replace the windows with a double window and place insulation outside the wall. However, in colder areas such as Oviedo or Burgos, the strategy to be followed includes double windows, insulation on the inside of the walls and insulation inside the roof (same solution as the one proposed for Germany, where the regulatory limits are demanding). For Poland, a solution similar to that of Málaga has been obtained, and in the case of France a solution is proposed that includes a double window, insulation on the internal layer of the wall, and on the inside of the roof (similar to those of Oviedo, Burgos and Germany).



**Third model: low quality building prototype with flat roof**

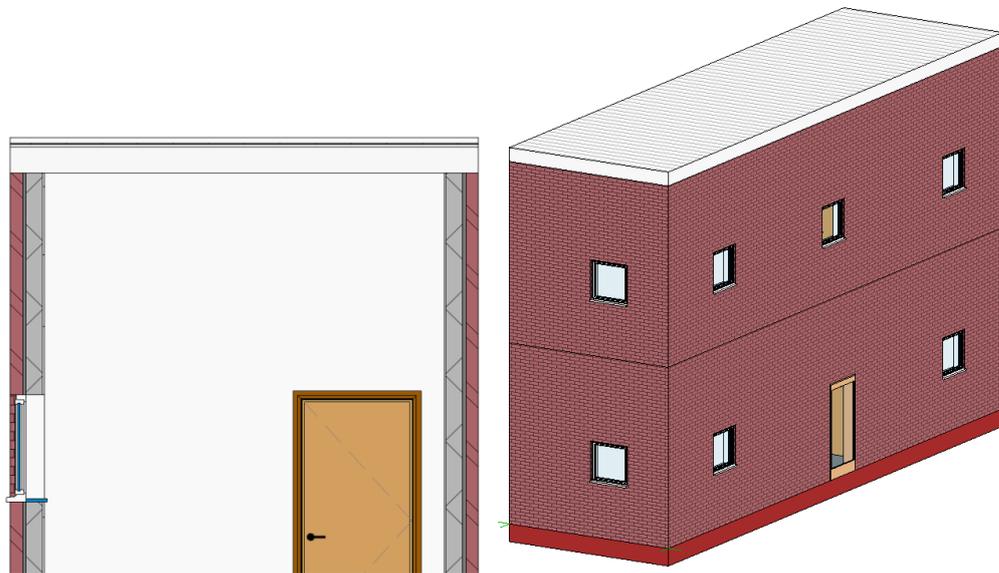
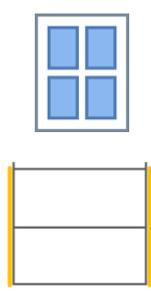


Figure 12. Third building model under analysis: low quality building prototype with flat roof.

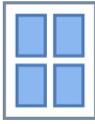
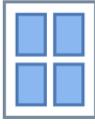
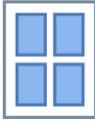
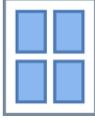
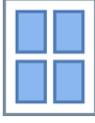
Table 12. Initial values of the Low quality building prototype with flat roof.

$K_{AVE}$ [W/(m <sup>2</sup> ·K)]	$U_{window}$ [W/(m <sup>2</sup> ·K)]	$U_{wall}$ [W/(m <sup>2</sup> ·K)]	$U_{roof}$ [W/(m <sup>2</sup> ·K)]	Type of roof
1.39	5.562	0.827	3.423	Flat

Table 13. Final values of the Low quality building prototype with flat roof.

Location	K value [W/(m <sup>2</sup> ·K)]	$U_{window}$ [W/(m <sup>2</sup> ·K)]	$U_{wall}$ [W/(m <sup>2</sup> ·K)]	$U_{roof}$ [W/(m <sup>2</sup> ·K)]	Selected solution	Cost [€]
Spain (Málaga, zone A)	0.936	2.7	0.348		Double glazing windows + Insulation outside the walls 	13737.87



Spain (Oviedo, zone C)	0.427	2.1	0.336	0.486	Double glazing windows + Insulation inside the walls + Insulation outside the roof	 	17339.90
Spain (Burgos, zone E)	0.416	1.8	0.336	0.486	Double glazing windows + Insulation inside the walls + Insulation outside the roof	 	17650.44
Poland	0.8	1.6	0.229		Double glazing windows + Insulation outside the walls	 	16779.01
France	0.419	1.9	0.336	0.486	Double glazing windows + Insulation inside the walls + Insulation outside the roof	 	17650.44
Germany	0.36	1.3	0.336	0.262	Double glazing windows + Insulation inside the walls + Insulation outside the roof	 	17774.65



As can be seen from the outputs for the low-quality flat roof model, the proposed outputs differ by location. In hot areas of Spain like Malaga, it is enough to replace the windows with a double window and place insulation outside the wall. On the other hand, in colder areas such as Oviedo or Burgos, the strategy to follow includes double windows, insulation inside the walls of the building and insulation outside the roof (same solution as the one proposed for Germany, where the regulatory limits are demanding, and the one proposed for France). For Poland, a solution similar to that of Malaga has been obtained.

The decision tree model has been evaluated according to the most used metrics, already explained in section 2. For this purpose, the data has been divided into a training set and a test set, considering 70% of the data for training and 30% to test performance. Starting with the accuracy, the model reaches a value of 0.86, therefore the value obtained for the error rate is 0.13, Precision and recall measures have already been calculated, obtaining values of 0.86 for both of them. The same value has been provided for the F score. The model shows good accuracy and can perform a good classification.

The following figures show an example of output results displayed to the user through the tool. It is also possible to view the decision tree that the algorithm has used in making decisions to select the optimal solution (Figure 13 and Figure 14).

**BUILDING THERMAL PATHOLOGIES DIAGNOSIS TOOL**

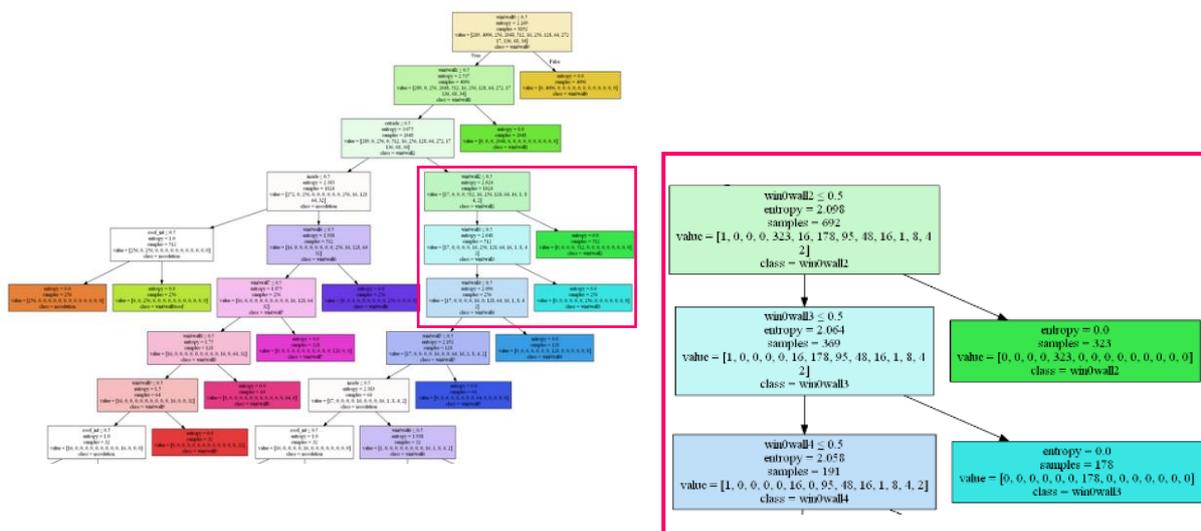


Figure 13. Sample of information provided about the decision tree algorithm.



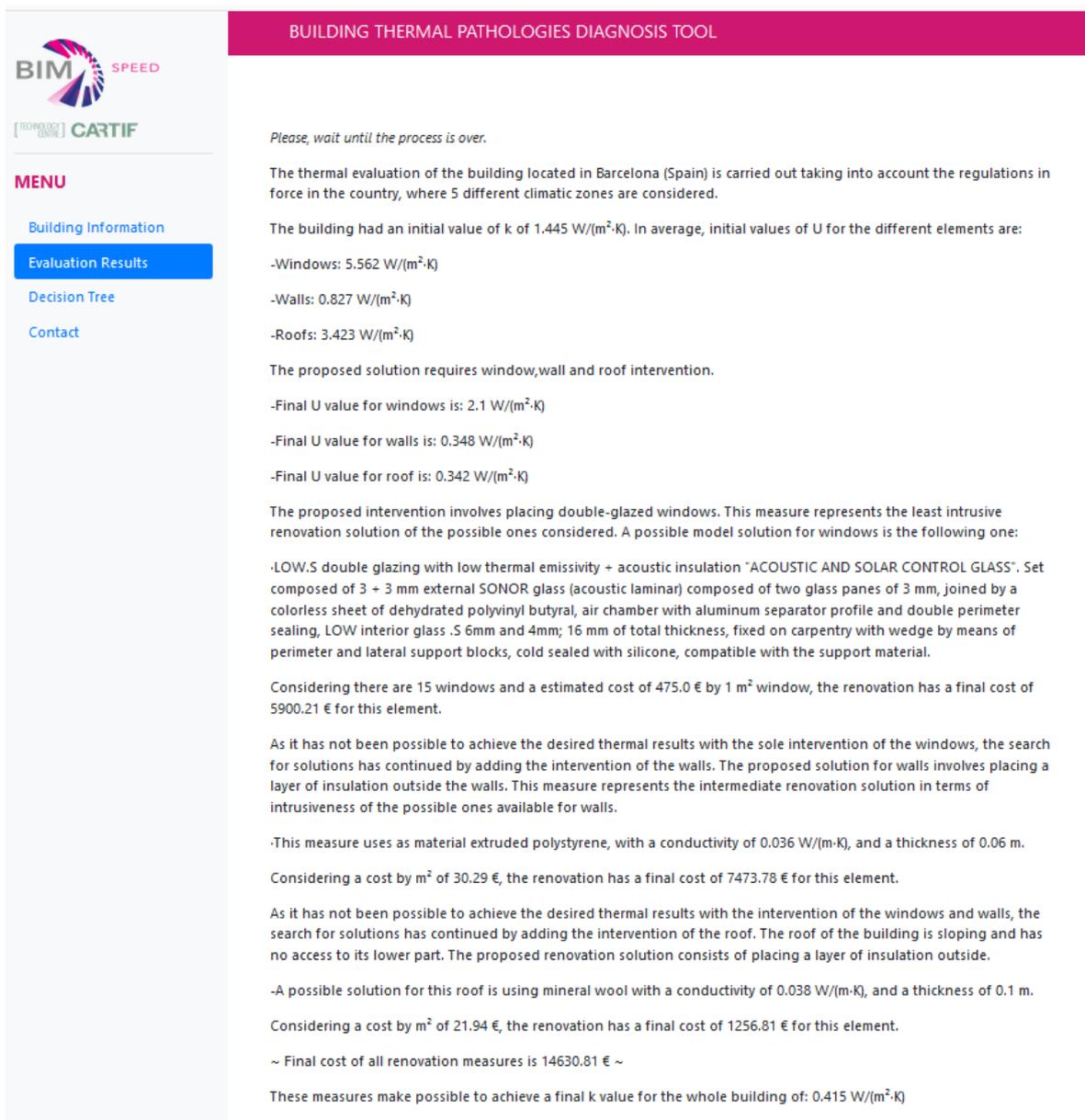


Figure 14. Sample evaluation output provided to user.

## 5.2 Exploitation results of T7.2

The diagnosis tool shown in this deliverable is the result of the developments carried out in task 7.2- *Machine-learning for As-Built diagnostics and enrichment of design rules for deep renovation* and is currently in a prototype phase. The validation status of the tool is in TRL5. We expect to reach TRL6 at the end of the BIM-SPEED project. The commercial exploitation of the software is not explored at time of writing the report as the current developments are focused on a research and development phase and will be considered at the end of the BIM-SPEED project.



## 6. Conclusions

This report has described the potentials, constraints and workable solutions of the use of machine-learning/artificial intelligence approach at the design stage of deep renovation projects. Machine learning techniques have been used for this purpose for a long time, in order to provide powerful mechanisms to deal with problems in specific elements of buildings (from decision tables, to knowledge-based systems or expert systems). Nevertheless, much of this information is not easy to find, takes long to fill in, or needs to be provided by experts. That is why it is proposed here to use information about the building elements included in the IFC data model.

On the other hand, there are examples in the literature on the use of machine learning with IFC files. Some of them focus on detecting internal problems and inconsistencies, due to the importance of IFC integrity checking. Others try to find problems in geometry. Research has also been carried out on health and safety issues, as well as verifying compliance with building regulations. No examples have been found in the field of thermal pathologies, and it was identified as a possible problem to be addressed.

Some of the most commonly used techniques for solving this kind of problem include decision trees, rule-based systems and experts systems, that is why a part of the research focuses on understanding how each of them works and is defined. Finally, due to the nature of the selected problem to be solved, the decision tree seemed to be a good technique to apply.

In relation to the information available on the Internet on building pathologies, there are some databases that provide interesting material. However, some of them are licensed and offer the information with a different structure, sometimes it is even difficult to use their search engines. Once the problem was defined, another data source identified as potentially useful was the Tabula WebTool, which provides information on different measures applied to the three main constructive elements in buildings for renovation, classified by country, type of building and year of construction.

The designed tool in the proof of concept aim is to respond to these thermal pathologies following a user-friendly approach, with the idea that it can be integrated with other tools, and that an automatic calculation mechanism is provided. Additionally, it could be used by experts or even other kinds of users who do not need to have extensive knowledge about buildings, as the algorithm does all the work by itself. The results provided by the tool show that there are significant differences regarding the requirements depending on the respective country (the regulations seem to be stricter in places like Germany), and even in different climatic zones of the same country (the case of Spain).

To conclude, further research could be done in this field. Provided that other parameters related to building thermal information are included in the IFC model, the thermal performance analysis could be enriched and more specifications could be given on the proposed solutions (for example, including information about the building's energy systems).



## 7. References

- [1] K. Van Balen, "Expert system for evaluation of deterioration of ancient brick masonry structures," *The Science of the Total Environment*, vol. 189, pp. 247-254, 1996.
- [2] F. Moodi and J. Knapton, "Research into a Management System for Diagnosis, Maintenance, and Repair of Concrete Structures," *Journal of Construction Engineering and Management*, vol. 129, pp. 555-561, 2003.
- [3] E. Bernat and L. Gil, "Aided Diagnosis of Structural Pathologies with an Expert System," *Advances in Structural Engineering*, vol. 16, pp. 379-393, 2013.
- [4] J. Brito, C. Pereira, J. Silvestre and I. Flore-Colen, *Expert Knowledge-based Inspection Systems: Inspection, Diagnosis, and Repair of the Building envelope*, Lisbon, Portugal: Springer, 2020.
- [5] D.-Y. Lee, H.-L. Chi, J. Wang, X. Wang and C.-S. Park, "A linked data system framework for sharing construction defect information using ontologies and BIM environments," *Automation in Construction*, vol. 68, pp. 102-113, 2016.
- [6] B. Koo, B. Shin and T. F. Krijnen, "Employing outlier and novelty detection for checking the integrity of BIM to IFC entity associations," in *34th International Symposium on Automation and Robotics in Construction (ISARC 2017)*, Taipei, Taiwan, 2017.
- [7] B. Koo and B. Shin, "Applying novelty detection to identify model element to IFC class misclassifications on architectural and infrastructure Building Information Models," *Journal of Computational Design and Engineering*, vol. 5, pp. 391-400, 2018.
- [8] T. Krijnen and M. Tamke, "Assessing Implicit Knowledge in BIM Models with Machine Learning," in *Design Modelling Symposium*, Copenhagen, Denmark, 2015.
- [9] G. Lee, J. Won, S. Ham and Y. Shin, "Metrics for quantifying the Similarities and Differences between IFC files," *Journal of Computing in Civil Engineering*, vol. 25, pp. 172-181, 2011.
- [10] G. Lilis, G. Giannakis and D. Rovas, "Detection and semi-automatic correction of geometric inaccuracies in IFC files," in *14th Conference of International Building Performance Simulation Association*, Hyderabad, India, 2015.
- [11] R. Eftekhari, M. Nik-Nakht and A. Hammad, "Extending IFC for Fire Emergency Real-Time Management Using Sensors and Occupant Information," in *35th International Symposium on Automation and Robotics in Construction (ISARC 2018)*, Berlin, Germany, 2018.
- [12] V. Getuli, S. Mastrolemo Ventura, P. Caponea and A. L. Ciribini, "BIM-based code checking for construction health and safety," in *Creative Construction Conference*, Primosten, Croatia, 2017.



- [13] J. Han, M. Kamber and J. Pei, *Data mining: concepts and techniques*, Waltham, USA: Morgan Kaufmann Publishers, 2012.
- [14] T. M. Mitchell, *Machine Learning*, McGraw-Hill, 1997.
- [15] A. v. d. Beukel, "Building pathology: a State-of-the-art report," CIB Report, Netherlands, 1993.
- [16] J. Giarratano and G. Riley, *Expert Systems-Principles and Programming*, Course Technology Inc., 2004.
- [17] C. Grosan and A. Abraham, "Rule-Based Expert Systems," in *Intelligent Systems*, Berlin, Springer, 2011, pp. 149-185.
- [18] S. Sahin, M. Tolun and R. Hassanpour, "Expert Systems," in *Kirk-Othmer Encyclopedia of Chemical Technology*, Wiley-Interscience, 2016.
- [19] D. S. Watt, *Building Pathology: Principles and Practice (Second Edition)*, Oxford, UK: Blackwell Publishing, 2007.
- [20] V. P. d. Freitas, "A state-of-the-art report on building pathology," CIB-Building Pathology, Porto, 2013.
- [21] P. INSITER, "Best practices and existing shortcomings. Deliverable report 1.1," 2016.
- [22] Y. Qi, Q. Qian, F. Meijer and H. Visscher, "Causes of quality failures in building energy renovation projects of northertn China: a review and empirical study," *Energies, MDPI, Open Access Journal*, vol. 13, no. 10, pp. 1-19, 2020.
- [23] TABULA and EPISCOPE, "TABULA WebTool," [Online]. Available: [www.webtool.building-typology.eu](http://www.webtool.building-typology.eu). [Accessed 01 09 2020].
- [24] I. builingSMART, "Industry Foundation Classes (IFC)- An Introduction," buildingSMART International, Ltd., 31 December 2016. [Online]. Available: <https://technical.buildingsmart.org/standards/ifc/>. [Accessed May 2020].
- [25] T. Krijnen, "IfcOpenShell," 2015. [Online]. Available: <http://ifcopenshell.org>. [Accessed May 2020].
- [26] ELIOS, "Draft final report," 2014.
- [27] J. Wu and J. Zhang, "New automated BIM Object Classification Method to Support BIM Interoperability," *Journal of Computing in Civil Engineering*, vol. 33, no. 5, 2019.
- [28] Y. Qi, Q. Qian, F. Meijer and H. Visscher, "Identification of Quality Failures in Building Energy Renovation Projects in Northern China," *Sustanability*, vol. 11, no. 4203, 2019.
- [29] J. Gu, H. Zhang and M. Gu, "Automatic Integrity Checking of IFC Models relative to building Regulations," 52-56, 2016.
- [30] S. Russel and P. Norvig, *Inteligencia Artificial: un enfoque moderno*, Madrid, Spain: Pearson Prentice Hall, 2004.



## APPENDIX 1 – RENOVATION MEASURES

This section includes a summary of the measures for Germany, Spain and Poland applied to walls and roofs, considering the year of construction of the building, and the type (multi-family house or apartment). The information has been extracted from the Tabula website (<http://webtool.building-typology.eu/>). The total number of solutions provided by Tabula is greater than the set shown here. The main objective has been to simplify the set of solutions, and make them as homogeneous as possible, even for different countries, which have elements in common. The solutions for France are not fully described on the website, therefore they are not included here, and, just to check the availability of the approach, the same measures that the ones for Spain have been considered in that country.

			Multihouse building/ Apartment		
State	Year	Number	Roof 1	Roof 2	Wall
Original	1918	1	Tilted roof with cavity between rafters		Timber frame
Usual refurbishment			Insulate cavity between rafters (heighten rafters and clear interspace if necessary), total insulation thickness 12 cm		Add 8 cm of insulation on inner surface (air-tight c) Adding, water pipes must not be conducted in outer wall)
Advanced refurbishment			Insulate cavity between rafters + insulation layer, total insulation thickness 30 cm		Add 24 cm of insulation + plaster (external insulated render system), alternative: curtain wall
Original	1919-1948	2	Tilted roof with wood fiberboard planking	Wooden beam ceiling	Brickwork
Usual refurbishment			Insulate cavity between rafters (heighten rafters and clear interspace if necessary), total insulation thickness 12 cm	Add 12 cm insulation above (+ flooring if necessary)	Add 12 cm of insulation + plaster (external insulated render system), alternative: curtain wall (e.g. cellulose between timbers)
Advanced refurbishment			Insulate cavity between rafters + insulation layer, total insulation thickness 30 cm	Add 30 cm insulation above (+ flooring if necessary)	Add 24 cm of insulation + plaster (external insulated render system), alternative: curtain wall
Original	1949-2001	3	Concrete ceiling		Masonry of hollow blocks or honeycomb bricks
Usual refurbishment			Add 12 cm insulation above (+ flooring if necessary)		Add 12 cm of insulation + plaster (external insulated render system), alternative: curtain wall (e.g. cellulose between timbers)
Advanced refurbishment			Add 30 cm insulation above (+ flooring if necessary)		Add 24 cm of insulation + plaster (external insulated render system), alternative: curtain wall
Original	2002-2009	4	Tilted roof with 18 cm insulation		Masonry with 14 cm render insulation system
Usual refurbishment			Insulate cavity between rafters (heighten rafters and clear interspace if necessary), total insulation thickness 12 cm		Add 12 cm of insulation + plaster (external insulated render system), alternative: curtain wall (e.g. cellulose between timbers)
Advanced refurbishment			Insulate cavity between rafters + insulation layer, total insulation thickness 30 cm		24 cm of external insulation brickwork + plaster (external insulated render system)
Original	2010-2015	5	18 cm insulation on ceiling + roofing		12 cm of external insulation brickwork + plaster (external insulated render system)
Usual refurbishment			35 cm insulation on ceiling + roofing		24 cm of external insulation brickwork + plaster (external insulated render system)
Advanced refurbishment			40 cm insulation (WLS 032) on ceiling + roofing		28 cm of external insulation brickwork + plaster (external insulated render system)

### Measures for Germany



State	Year	Number	Multihouse building		Apartment	
			Roof	Wall	Roof	Wall
Original	-1966	1	Flat roof nonventilated	2 brick wall		
Usual refurbishment			Insulation of cavity between rafters 12 cm + 4 cm of additional Insulation layer.	Add polystyrene 18 cm		
Advanced refurbishment			Insulation of cavity between rafters 12 cm + 10 cm of additional Insulation layer.	Add polystyrene 16 cm		
Original	1967-1992	2	Flat roof ventilated	Pref. concrete panels	Average U-value for roofs or top ceilings / single-unit housing / construction years 1995 ... 2009 (S UH III) / not refurbished	Pref. concrete panels
Usual refurbishment			Insulation of cavity between rafters 12 cm + 10 cm of additional Insulation layer.	Add polystyrene 18 cm	Insulation of cavity between rafters 12 cm + 10 cm of additional Insulation layer	Add polystyrene 16 cm
Advanced refurbishment			Insulation of cavity between rafters 12 cm + 18 cm of additional Insulation layer.	Add polystyrene 16 cm	Insulation of cavity between rafters 12 cm + 18 cm of additional Insulation layer	Add polystyrene 18 cm
Original	1993-2002	3	Flat roof ventilated	Pref. concrete panels	Flat roof nonventilated	Hollowbloks
Usual refurbishment			Insulation of cavity between rafters 12 cm + 10 cm of additional Insulation layer	Add polystyrene 16 cm	Insulation of cavity between rafters 12 cm + 10 cm of additional Insulation layer	Add polystyrene 16 cm
Advanced refurbishment			Insulation of cavity between rafters 12 cm + 18 cm of additional Insulation layer	Add polystyrene 14 cm	Insulation of cavity between rafters 12 cm + 18 cm of additional Insulation layer	Add polystyrene 18 cm
Original	2003-	4	Flat roof ventilated	Hollowblocks	Flat roof ventilated	Brickwall & internal insulation (3 layer)
Usual refurbishment			Insulation of cavity between rafters 12 cm + 10 cm of additional Insulation layer	Add polystyrene 12 cm	Insulation of cavity between rafters 12 cm + 10 cm of additional Insulation layer	Add polystyrene 14 cm
Advanced refurbishment			Insulation of cavity between rafters 12 cm + 18 cm of additional Insulation layer	Add polystyrene 14 cm	Insulation of cavity between rafters 12 cm + 18 cm of additional Insulation layer	Add polystyrene 16 cm

### Measures for Poland



State	Year	Number	Multifamily building			Apartment			
			Roof 1	Roof 2	Wall 1	Wall 2	Roof 1	Roof 2	Wall
Original	-1900	1	Ventilated wooden frame pitched roof with suspended ceiling	Ventilated flat roof with concrete horizontal framework	Masonry of coating bricks	Masonry of coating bricks			
Usual refurbishment			Add new waterproof, 20 mm of insulation and gravel	Add new waterproof, 20 mm of insulation and gravel	Adding 30 mm of insulation at the inside side of the facade, clad with plasterboard	Adding 30 mm of insulation at the inside side of the facade through ETIC			
Advanced refurbishment			Add new waterproof, 60 mm of insulation and gravel	Add new waterproof, 60 mm of insulation and gravel	Adding 50 mm of insulation at the exterior side of the facade through ETIC	Adding 50 mm of insulation at the exterior side of the facade through ETIC			
Original	1901-1936	2	Ventilated metal frame flat roof		Masonry of coating bricks				
Usual refurbishment			Add new waterproof, 20 mm of insulation and gravel		Adding 30 mm of insulation at the inside side of the facade, clad with plasterboard	Adding 30 mm of insulation at the exterior side of the facade through ETIC			
Advanced refurbishment			Add new waterproof, 60 mm of insulation and gravel		Adding 50 mm of insulation at the exterior side of the facade through ETIC	Adding 50 mm of insulation at the exterior side of the facade through ETIC			
Original	1937-1959	3	Ventilated concrete framework flat roof with horizontal framework		Masonry of coating bricks				
Usual refurbishment			Add new waterproof, 20 mm of insulation and gravel		Adding 30 mm of insulation at the inside side of the facade, clad with plasterboard	Adding 30 mm of insulation at the exterior side of the facade through ETIC			
Advanced refurbishment			Add new waterproof, 60 mm of insulation and gravel		Adding 50 mm of insulation at the exterior side of the facade through ETIC	Adding 50 mm of insulation at the exterior side of the facade through ETIC			
Original	1960-1979	4	Ventilated concrete framework flat roof with horizontal framework		Cavity wall brick/air cavity				
Usual refurbishment			Add new waterproof, 20 mm of insulation and gravel		Injection of 30 mm of thermal insulation inside the existing air chamber	Injection of 30 mm of thermal insulation inside the existing air chamber			
Advanced refurbishment			Add new waterproof, 60 mm of insulation and gravel		Adding 50 mm of insulation at the exterior side of the facade through ETIC	Adding 50 mm of insulation at the exterior side of the facade through ETIC			
Original	1980-2006	5	Ventilated concrete framework flat roof with horizontal framework		Cavity wall with insulation inside the cavity				
Usual refurbishment			Add new waterproof, 80 mm of insulation and gravel		Adding 50 mm of insulation at the inside side of the facade, clad with plasterboard	Adding 30 mm of insulation at the inside side of the facade, clad with plasterboard			
Advanced refurbishment			Add new waterproof, 100 mm of insulation and gravel		Adding 100 mm of insulation at the exterior side of the facade through ETIC	Adding 50 mm of insulation at the exterior side of the facade through ETIC			
Original	2007-	6	Flat roof: one-way framework with prestressed joint		Brick wall				
Usual refurbishment			No improvement/measure		Adding 80 mm of insulation at the exterior side of the facade through ETIC	Adding 100 mm of insulation at the exterior side of the facade through ETIC			
Advanced refurbishment			No improvement/measure		Adding 80 mm of insulation at the exterior side of the facade through ETIC	Adding 100 mm of insulation at the exterior side of the facade through ETIC			

Measures for Spain

