



**HAL**  
open science

## Deep learning models for building window-openings detection in heating season

Enguerrand de Rautlin de la Roy, Thomas Recht, Akka Zemmari, Pierre  
Bourreau, Laurent Mora

► **To cite this version:**

Enguerrand de Rautlin de la Roy, Thomas Recht, Akka Zemmari, Pierre Bourreau, Laurent Mora.  
Deep learning models for building window-openings detection in heating season. *Building and Environment*, 2023, 231, pp.110019. 10.1016/j.buildenv.2023.110019 . hal-04432021

**HAL Id: hal-04432021**

**<https://hal.science/hal-04432021>**

Submitted on 1 Feb 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# 1 Deep learning models for building window-openings detection 2 in heating season

3 Enguerrand de Rautlin de la Roy<sup>\*a</sup>, Thomas Recht<sup>a</sup>, Akka Zemhari<sup>b</sup>, Pierre  
4 Bourreau<sup>c</sup>, Laurent Mora<sup>a</sup>

5 <sup>a</sup>*Univ. Bordeaux, CNRS, Bordeaux INP, I2M, UMR 5295, F-33400, Talence, France*

6 <sup>b</sup>*Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, F-33400, Talence, France*

7 <sup>c</sup>*Nobatek/INEF4, 9 rue Jean-Paul Alaux, F-33000 Bordeaux, France*

---

## 9 Abstract

10 The increasing use of monitoring systems such as Building Management System (BMS) or connected devices bring the  
11 opportunity to better evaluate, model or control both occupants' comfort and energy consumed by an operated  
12 building thanks to the consequent amount of data provided (e.g., air temperature, CO<sub>2</sub> concentration, electricity  
13 consumption). Occupants' behavior and more specifically window-openings affect both occupants' thermal comfort  
14 and building energy consumption and are therefore key components to consider. This paper presents a comparison of  
15 machine learning models applied on window-openings detection during the heating season such as: Linear  
16 Discriminant Analysis (LDA), Support Vector Machine (SVM), Random Forest Classifier (RFC) and two Recurrent  
17 Neural Network (RNN), namely, Long Short Term Memory (LSTM) and Gated Recurrent Unit (GRU). While some  
18 applications of Artificial Intelligence (AI) methods applied on window-openings detection exist in the literature, this

---

\* enguerrand.de-rautlin-de-la-roy@u-bordeaux.fr

19 study proposes a detailed comparison of the main methods and focuses on the impact of feature engineering process  
20 considering four different data transformations based on field expertise and more than 800 different combinations  
21 built on six indoor and outdoor measurements. Results show that some of the proposed transformations and  
22 combinations positively impact all models performances. The best performances on window-openings detection are  
23 attained by using indoor temperature and CO<sub>2</sub> concentration on RNN models with an average F1-score of 0.78 while  
24 LDA, SVM and RFC models tend to provide satisfying but lower performance around 0.70-72. In addition, by using the  
25 right transformation, significant results can be achieved by detecting up to 84-88 % of window-opening times with the  
26 sole use of indoor air temperature measurements.

27 Keywords: deep learning, window-opening, recurrent neural network, support vector machine, Random forest

---

## 28 **1. Introduction**

29 The building sector accounts for approximately 40% of the final energy use in Europe [1]. In addition,  
30 life cycle analysis of buildings tends to show that most of the building life cycle energy consumption  
31 depends on its operation (80 to 90%) [2]. Thus, evaluating and defining operational loads of building are  
32 keys elements in order to reduce or comprehend buildings energy consumptions. The most common use  
33 of Building Management System (BMS) and connected devices in the building sector has led to a  
34 democratization of edifices called smart buildings. A *smart building* can be seen as an association of  
35 multiple systems, software and sensors [3] that aims to meet two main objectives: to reduce both  
36 operational and environmental costs by managing and optimizing the energy use [4, 5, 6] and to improve  
37 the comfort of the occupants [7, 8]. Hence, smart buildings generate a consequent amount of data from  
38 measurements (air temperature, CO<sub>2</sub> concentration, energy consumption, *etc.*) that can be studied in order  
39 to evaluate, model, correct or optimize specific operational loads during the building life cycle [9, 10, 11].  
40 Among these loads, occupants' behavior can have a strong influence on the operational energy  
41 consumption of buildings as well as the thermal comfort [12, 13]. A common action is often identified:

42 window-openings [14, 15, 16]. However, although window states are required to understand the  
43 functioning and performances of a building, they are rarely measured on sites or exploitable. Unlike  
44 ambient sensors which are commonly used (e.g., air temperature), window opening sensors are generally  
45 numerous to install and considered more intrusive. In addition, data collected from sensors are rarely  
46 clean and straight usable due to common errors that may occur during the measurement phase (e.g., data  
47 transmission failure, accident) and thus, often require some preprocessing [17, 18]. There is currently no  
48 consensus on how occupants interact with their building as well as all factors that may have an influence  
49 on their behaviors [19] and modeling occupants actions without dedicated measurement or by using poor  
50 quality data (anomalies, *etc.*) can rather be difficult. Thus, openings are often approximated by expert  
51 rules (e.g., ratios) or by stochastic approaches [20] that can induce significant gaps compared to real in-  
52 situ observations [13].

53 Nevertheless, window-openings impacts on other measurements (such as air temperature, CO<sub>2</sub>  
54 concentration, *etc.*) can be observed through specific patterns that tend to deviate from other  
55 observations. These patterns can be recognized and classified by using machine learning techniques in  
56 order to determine the corresponding window-status (*open or close*).

57 Nowadays, pattern recognition and classification through machine learning techniques is commonly  
58 used in various domains for multiple purposes such as financial with the fraud detection or in the security  
59 field to detect intrusion and even in the medical field to detect breast cancer [21]. In the building sector,  
60 studies covering machines learning techniques applied to window-status detection seem to rarely  
61 compare multiple models performances and tend to mainly be based on logistic regression models [15].  
62 Furthermore, these studies rarely discuss the selection of feature as well as the associated feature  
63 engineering process [15], yet considered as core keys to influence positively models performances [22].  
64 Since machine learning models show poor generalization capabilities and usually require a specific tuning  
65 for each household and building [20] the present work aims to contribute on window-opening status  
66 detection by comparing five different models to provide tendency observations on models, measurements  
67 combinations and transformations.

68

69 The main contributions of this article are therefore:

- 70 • The comparison on window-status detections for several machine learning models classifier such as  
71 Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), Random Forest Classifier (RFC)

72 and Recurrent Neural Network (RNN) such as Long Short Term Memory (LSTM) and Gated Recurrent  
73 Unit (GRU).

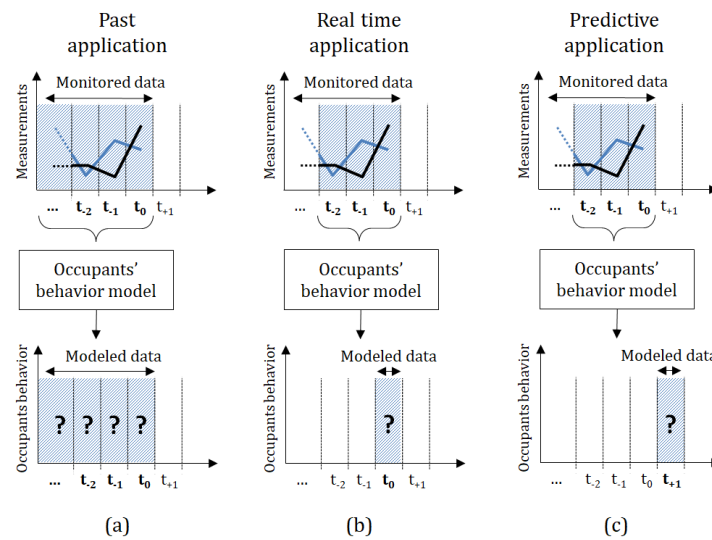
- 74 • A detailed approach on indoor and outdoor measurements combinations impact on models outputs.  
75 This study might provide a guide on the type of sensor to be preferred for in-situ sites installation in  
76 order to detect window openings.
- 77 • A detailed explanation on feature engineering followed process in order to quantify measurements  
78 transformations and association performances on models outputs.

## 79 **2. Related work**

### 80 *2.1. Occupants' behavior detection or prediction*

81 Occupants' behavior impact both occupants' thermal comfort and building energy consumption [23]  
82 with around 25% of the total energy consumption in Europe between 1990 to 2014 being dedicated to  
83 domestic uses [24]. To understand, control or minimize these impacts, machine learning techniques used  
84 to model these behaviors (e.g., occupancy, window-openings) can be applied to serve multiple purposes in  
85 the building sector including, among others, prediction, fault detection, diagnosis or control optimization  
86 [9, 15]. Following this perspective, three different applications of occupants' behavior models are  
87 presented in [Figure 1](#) **Erreur ! Source du renvoi introuvable.**: past, present (or real-time) and future (or  
88 predictive) detection models. Past detection models (1.a) can take advantage of all monitored data in  
89 order to detect or classify, a posteriori, occupants' behavior. This approach can be used, among others  
90 things, to perform fault detection or to detect and correct anomalies by comparing modeled to measured  
91 behaviors, to reduce the gap between simulated and real energy consumption for energy performance  
92 verification [25, 26] or to evaluate retrofit actions [27, 28]. Present detection models (1.b) focus on  
93 detecting real time behavior while future detection models (1.c) focuses on predicting behavior one or  
94 multiple time step ahead. Both approaches are the most commonly performed in studies [15] and can be  
95 applied for real time implementation in order to perform fault detection [29], control optimization for  
96 comfort improvements or energy savings [30, 31, 32]. Specificities of models and training process apart,  
97 these three applications can be performed by using the same data and are mainly related to intended uses  
98 (regardless of the results). Therefore and to avoid any overload, this study focuses solely on past detection  
99 but all measurements transformations can also be applied to real-time detection and future prediction.

100 Regarding these aspects, the approach extended to real-time applications is also succinctly addressed and  
 101 discussed in section 5.5.



102 (a) (b) (c)  
 103 **Figure 1** Occupants behavior detection based on monitored data: (a) past detection, (b) present detection and (c) future prediction

104 *2.2. Window-opening behavior models*

105 As presented in Xilei Dai review [15], studies on window-openings through machine learning models  
 106 can be divided into two groups. The first focuses mainly on occupants' actions toward the window, given a  
 107 specific environment (e.g., indoor and outdoor air temperature, wind speed), in order to model openings  
 108 and closings actions [33, 34]. The second, on which this study is based, mainly focuses on modeling  
 109 window-status as open or close depending on the environment [20, 35, 36]. Three main elements can be  
 110 highlighted from previous studies [15]:

- 111 • Most of studies mainly focus on one sole machine learning model at a time and rarely compare  
 112 different models on the same study.
- 113 • Logistic Regressions (LR) models are the most used to determine window-status [36, 37, 38, 39]  
 114 followed by Artificial Neural Networks (ANN) in more recent studies [20, 40, 41]. The vast majority of  
 115 presented models is based on a supervised approach and thus, uses labeled data.
- 116 • Generic models apart, machine learning models for window-status are specific to buildings, occupants  
 117 or seasons [20, 38] and their performances can be assessed regarding multiple evaluation metrics  
 118 (such as accuracy, F1-score, true positive rate, area under the curve, etc.). Thus, evaluating and  
 119 comparing different models through multiple studies is rather difficult.

120

121 Hence, the decision was made to focus this study on various models to compare their results and  
122 tendencies. Considering the amount of significant contribution realized with logistic regression models  
123 and Artificial Neural Networks, others models underrepresented for window openings detection and  
124 based on their popularity on other fields of research, towards pattern recognition, classification,  
125 prediction or anomaly detection, were selected. It includes, Recurrent Neural Network (RNN) [42, 43, 44],  
126 Linear Discriminant Analysis (LDA) [45, 46], Support Vector Machine (SVM) [44, 45] and Random Forest  
127 Classifier (RFC) [43, 45, 46]. It is important to note that SVM and RFC models have been applied on a few  
128 studies regarding window-status modeling, showing great results [35]. On the other hand RNN and LDA  
129 models appear to be unrepresented despite potentially being highly effective regarding their actual  
130 performances on similar tasks such as detecting occupancy [43, 46] or on other fields of studies [47] such  
131 as medical by detecting anomalies [48] or energetic by optimizing performances [49]. Machine learning  
132 models used in the present study are further detailed in section 3.2.

### 133 *2.3. Feature selection and transformation for window opening models*

134 Features used for window opening models can rather be divided into two groups [15] environmental  
135 and non-environmental. Environmental features are based on indoor or outdoor environment  
136 measurements such as air temperature, CO<sub>2</sub> concentration, wind speed, solar radiation, noise, *etc.* while  
137 non-environmental features are based on buildings, occupants or time characteristics such as room type,  
138 gender, age or time of the day. As analyzed in Xilei Dai review [15] and regardless of the models, the most  
139 used features come from environment measurements such as: outdoor and indoor air temperature,  
140 humidity, indoor CO<sub>2</sub> concentration and wind speed. Regarding logistic regression models, most of the  
141 studies tend to show that indoor and outdoor air temperature have the most impact [39, 50, 51]. However,  
142 concerning artificial neural network models, a detailed study highlighting, among others, measurements  
143 impact for window openings application, such as Romana Markovic [40], shows different results. Her  
144 study provides a relevant example based on an ANN model by analyzing neurons learned weights from  
145 more than twenty input features that highlight the importance of indoor environmental data and more  
146 specifically, CO<sub>2</sub> concentration. Regarding these results, a broad approach including different ambient  
147 indoor and outdoor measurements is privileged for this study. In addition, another main point should be  
148 specified regarding feature selection. Most of the features used in previous studies are measurements that  
149 are neither transformed (e.g., derivation, smoothing) nor combined (e.g., differences) despite positive

150 impacts that might be obtained on models performances [22]. Hence, this study offers to test and compare  
151 various measurements transformations that are further presented in section 4.2.1 on several models in  
152 order to evaluate their contribution and relevance.

### 153 3. Experiment methodology

154 In this section a presentation of metrics used and models selected for this study is provided. Evaluation  
155 metrics chosen and built to compare the results are discussed and a short explanation of every model  
156 specificities, architecture selection and corresponding training process is given.

#### 157 3.1. Evaluation process

##### 158 3.1.1. F1-score

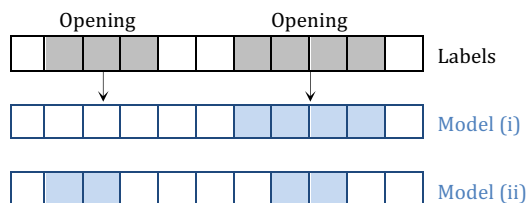
159 Window opening is a common binary classification task in machine learning. The window-status is  
160 reflected by two classes, *open* and *close*, which are underrepresented and overrepresented groups,  
161 respectively. As shown by [15], several metrics can be used to assess the classification performance on  
162 window-status. In this study the F1-score metric is firstly used to provide an overall evaluation of models  
163 results and secondly to allow a comparison with other studies. F1-score is calculated from Equation (1)  
164 where True Positive (TP) and True Negative (TN) represent the total amount of right classifications for  
165 window open and close status while False Positive (FP) and False Negative (FN) represent the total  
166 amount of wrong classifications for window open and close status respectively. F1-score values are  
167 ranged between 0 and 1, with 1 corresponding to a perfect window opening classification. An average F1-  
168 score of 0.5 means that for one TP there are two false classifications: two FP, two FN or one of both.

$$169 \quad F1\text{-score} = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \quad Eq. (1)$$

170 However, although this evaluation metric may provide a global overview on every models' performances,  
171 it alone might not be sufficient to choose which model is better especially in case of similar or identical  
172 results. As illustrated in Figure 2, a window opening state is defined, in this work, as one or successive  
173 open states (full-cells) bounded by one or successive close states (empty-cells). Both models evaluations  
174 (2.i) and (2.ii) provides the same F1-score values whereas both provide different results with only half of  
175 the openings perfectly detected for (2.i) and all openings detected but underestimated for (2.ii). Other



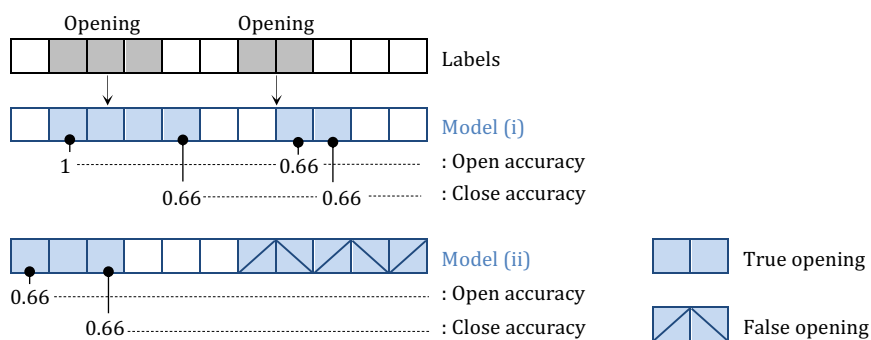
176 metrics might be useful in order to compare the results to other studies, to deepen models outputs and  
 177 guide or adapt the choice of models or measurements according to specific needs, with a focus on number  
 178 of openings detected or on the total opening time for instance. Thus, more domain oriented metrics  
 179 focusing on window opening classification and evaluation based on true and false opening detection are  
 180 introduced and discussed in this paper.  
 181



182 **Figure 2** Window-openings detection for two different models

183 **3.1.2. True and false openings**

184 A *true opening* is a modeled window opening that corresponds to a measured window opening within a  
 185 given time step limit. Therefore, a true opening is a true positive or a successive set of true positives which  
 186 may include one or more false positives. On the other hand, a modeled window opening that does not  
 187 satisfy this requirement is considered as a *false opening*. Thus, as shown in [Figure 3](#), for a time limit of two  
 188 time steps used in this study, the model results (3.i) and (3.ii) are made of two and one true openings,  
 189 respectively. Six evaluation metrics result from these definitions, the total true and false openings number,  
 190 the total true and false openings time and the average true opening accuracy score.



191 **Figure 3** True and false opening examples

192 **3.1.2.1. Total true and false openings number**

193 These metrics are used to evaluate a model capability to detect windows openings regardless of their  
 194 duration and correspond to the total number of true and false openings provided by a model.

195 *3.1.2.2. Total true and false openings time*

196 These metrics are used to evaluate a model capability to quantify windows openings duration and  
197 correspond to the total amount of time for all true and false openings provided by a model. In [Figure 3](#), the  
198 model results [\(3.i\)](#) has a six time step true opening time and no false opening time whereas the model  
199 results [\(3.ii\)](#) has a three time step true opening times for a five time step false opening times.

200 *3.1.2.3. Average true opening accuracy score*

201 A score is associated to every detected opening action (represented by the first open status of a true  
202 opening) and to every detected closing action (represented by the last open status of a true opening) in  
203 order to evaluate a model precision on window-openings detection. The score is set to 1 for a perfect  
204 match between the true opening and the measured opening and is linearly decreased by 0.33 for every  
205 time step difference. The penalty of 0.33 is chosen regarding the two time step limit set to define true and  
206 false openings. Lastly, the accuracy score, specific to each opening, is averaged for all the true opening  
207 provided by the model. In [Figure 3](#), the model results [\(3.i\)](#) has an average opening score of 0.83 and  
208 closing score of 0.66 whereas the model results [\(3.ii\)](#) has both average opening and closing score at 0.66.

209 *3.2. Models*

210 *3.2.1. SVMs*

211 Support Vector Machines (SVMs) are supervised machine learning methods commonly used for  
212 classification, regression and novelty detection. In a two-class classification problem and if the data is  
213 assumed to be separable in feature space, many boundaries that separate the classes may exists. An SVM  
214 model is therefore trained to determine the best boundary between classes (also called decision  
215 boundary) by maximizing the distance between every class sample [[52](#), [53](#)]. In this paper, a Radial Basis  
216 Function (RBF) kernel SVM classifier is trained following a 5-fold cross-validation for time series with an  
217 adaptive search algorithm to optimize the regularization and inverse of radius parameters. For every  
218 parameters combination in the range listed in [Table 1](#), F1-scores extracted from the 5-fold cross validation  
219 are averaged and the SVM model with the best performances is selected for the evaluation.

### 220 3.2.2. LDA

221 Linear Discriminant Analysis (LDA) is a supervised dimensionality reduction technique commonly used  
222 for classification. In a two-class classification problem with the same assumption as presented in section  
223 3.2.1, an LDA model is trained in order to construct a linear projection that maximizes the projected  
224 interclass variance and minimize the projected intraclass variance [52, 53]. The classification is based on  
225 Bayes' theorem to estimate a sample probability to belong to a class. Although LDA is only optimal for data  
226 with normal distribution and equal covariance matrices, its simplicity and robustness balance the loss in  
227 performances if above conditions are not fulfilled [54]. In this paper, an LDA classifier is trained following  
228 a 5-fold cross-validation for time series with an adaptive search algorithm to optimize the choice of solver  
229 and solver-dependent parameters such as shrinkage or threshold. For every solver and parameters  
230 combination in the range listed in Table 1, F1-scores extracted from the 5-fold cross validation are  
231 averaged and the LDA model with the best performances is selected for the evaluation.

### 232 3.2.3. Random Forest Classifier

233 Random Forest (RF) is a supervised machine learning method commonly used for classification (RFC)  
234 and regression that combines decision tree and ensemble methods. A decision tree is a tree-based method  
235 that divides the feature space into a set of rectangles that optimally split the data into classes. However  
236 trees tend to overfit during training and thus, have a low bias and a high variance. To be less prone to  
237 overfitting, a RF model is trained by randomly splitting the data into subsets and building a decision tree  
238 on each before aggregating their results (ensemble method) [52, 55, 56]. In this study, a Gini RF classifier  
239 is trained following a 5-fold cross-validation for time series with an adaptive search algorithm to optimize  
240 the choice of the number of trees, the minimum number of samples placed in a node before a node is split,  
241 the minimum number of samples required in a leaf node. For every combination of parameters in the  
242 range listed in Table 1, F1-scores extracted from the 5-fold cross validation are averaged and the RF  
243 model with the best performances is selected for the evaluation.

### 244 3.2.4. RNNs

245 Recurrent Neural Networks (RNNs) are a subclass of Artificial Neural Networks (ANNs). Unlike ANNs,  
246 RNNs possess an internal state memory that captures temporal order and dependencies of sequences,  
247 making them regularly used for task involving sequential data such as automatic translation, time series  
248 forecasting or classification. However, in practice, RNNs are not able to handle long-term dependencies

249 [57]. Long Short-Term Memory (LSTM) is a specific RNN that is designed to avoid this issue by selectively  
 250 forgetting long-term information [47]. On the other hand, Gated Recurrent Unit (GRU) is a variation of  
 251 LSTM that uses less training parameters and therefore consumes less memory, is faster and can  
 252 outperform LSTM on some tasks [58, 59, 60]. Unlike previous presented models, LSTMs and GRUs  
 253 hyperparameters listed in Table 1, were tuned beforehand in order to make a balance between  
 254 performance, training difficulties and computing time. Hence, both LSTM and GRU models for this study  
 255 are composed of a first layer of 16, 32 or 64 units (depending on the number of features used for training)  
 256 followed with a dense output layer of size 2 with softmax activation. An Adam optimizer is used with a  
 257 learning rate of 0.001 along with a binary cross-entropy loss function. To avoid overfitting, 25% of the  
 258 training set is selected as a validation set, the remaining training set is shuffled and a function to stop the  
 259 training if the model stop improving is used (also called early stopping).

260 **Table 1** List and range of Models' tuning parameters

Models	Tuning parameters and hyperparameters
SVM	<b>Regularization:</b> range 0.1 to 100 ; <b>Inverse radius:</b> range 0.01 to 10
LDA	<b>Solver:</b> Singular value decomposition, Least squares solution or Eigenvalue decomposition ; <b>Shrinkage:</b> Ledoit-Wolf lemma or None ; <b>Absolute threshold:</b> range 0.001 to 0.00001
RFC	<b>Number of trees:</b> range 10 to 150, <b>Minimum number of samples required to split an internal node:</b> range 2 to 10 ; <b>Minimum number of samples required to be at a leaf node:</b> range 1 to 4
LSTM & GRU	<b>Number of hidden layers:</b> range 0 to 2 ; <b>Number of units (size):</b> range 4 to 128 ; <b>Dropout:</b> range 0 to 0.5

## 261 4. Data description and data preprocessing

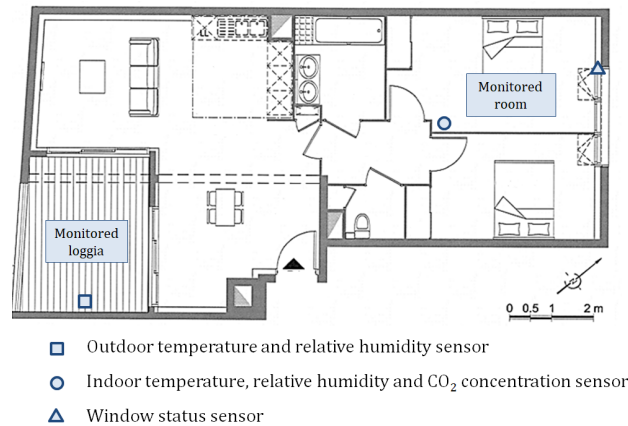
262 In this section a review of the data collected and features used for this study is presented. A short  
 263 explanation of the specificities of the train and test data is given followed by a detailed approach of the  
 264 feature engineering process conducted.

### 265 4.1. Data description

#### 266 4.1.1. Data collection and preparation

267 The raw data used for this study is made of 1 minute time step measurements collected over two years  
 268 (from July 2019 to February 2021) in a northwest bedroom of an apartment located in Bordeaux city-  
 269 center (France). The raw data include indoor climate measurements (such as temperature, relative  
 270 humidity and CO<sub>2</sub> concentration), outdoor climate measurements (such as temperature and relative

271 humidity) and window-status measurements. Sensors positions for all studied measurements are  
 272 illustrated in Figure 4.



273  
 274

**Figure 4** Studied sensors position

275 Additional data are also created from timestamp (such as hours, minutes and weekday) or original  
 276 values (such as absolute humidity) and added to the raw dataset. Over the two years of available  
 277 measurements a few months with a low anomalies rate were retained, cleaned, aggregated into a  
 278 15 minutes time step and separated into a train and test sets. The training set is composed of four months  
 279 of data, from the end of October 2019 to the beginning of March 2020. The training set consists in  
 280 12 095 data points collected during the heating season and its measurements characteristics are listed in  
 281 Table 2. Thus, one of the main limitations of this paper lies in the studied period which is characteristic of  
 282 a heating season with an average indoor air temperature usually superior to the outdoor. The test set is  
 283 made of one month of data from mid-December 2020 to mid-January 2021 for a total of 3 115 data points  
 284 that is also representative of heating seasons. This set is introduced in the following section.

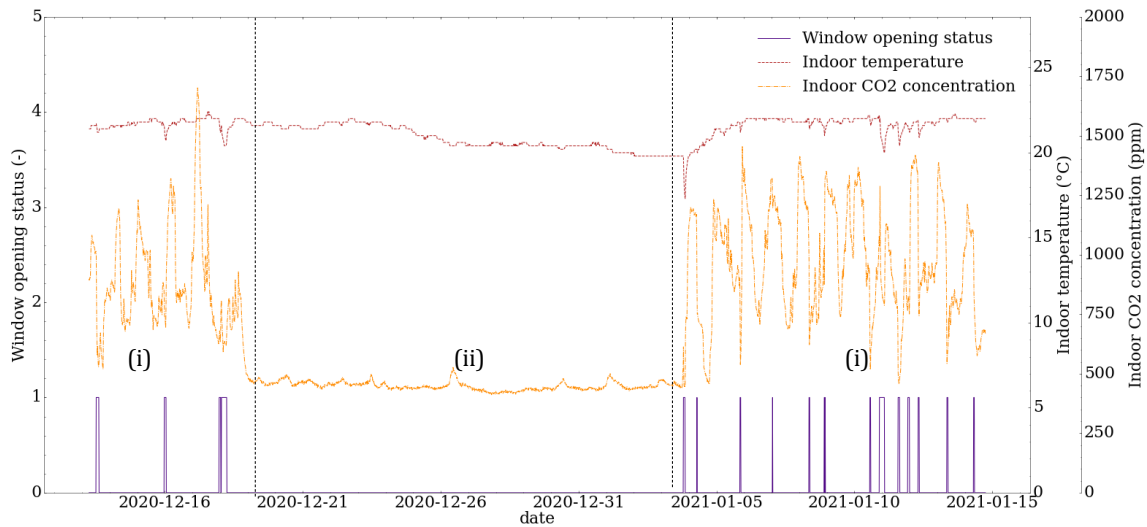
285

**Table 2** Training dataset measurements characteristics

Measurement name	Maximum value	Minimum value	Mean	Standard deviation
Indoor temperature (°C)	23.0	19.6	21.7	0.5
Indoor relative humidity (%)	69.8	29.6	49.7	6.1
Indoor CO <sub>2</sub> concentration (ppm)	2000.0	436.2	727.3	237.4
Outdoor temperature (°C)	23.5	4.8	12.3	2.9
Outdoor relative humidity (%)	93.6	36	71.2	10.2
Window-status (0-1)	1	0	-	-

286 *4.1.2. Data analysis*

287 Figure 5 provides an overview of the test data used to evaluate every model and highlights two periods  
 288 with (5.i) and without occupants (5.ii).



**Figure 5** Test dataset of indoor temperature, indoor CO<sub>2</sub> and window-status

289  
290  
291

292 The unoccupied period is characterized by a slow drift of approximately 2°C of indoor air temperature  
293 and a low stagnating CO<sub>2</sub> concentration. These differences are shown in Table 3. The occupied period is  
294 characterized by higher average indoor temperature and indoor CO<sub>2</sub> concentration. While period (5.i)  
295 presents data characteristics relatively close to the training set characteristics, those shown in  
296 period (5.ii) tend to differ. Such differences might impact the evaluation phase but may also provide useful  
297 information on models behavior and sensibility to uncommon data characteristics.

298 For all the test data, the measured open window-status time represents 25.75 hours for a total of 18  
299 openings and can be separated into two types of openings based on their impact on the indoor data.  
300 (1) Low impact openings that are characterized by a slow or inexistent fluctuation on indoor climate  
301 measurements that can be due to a window that is briefly or only slightly open, and (2) high impact  
302 openings which include all other and more impactful openings. Low impact openings represent 5 hours of  
303 open window-status time out of the 25.75 measured and are considered hard to classify for the models.  
304 On the other hand high impact openings tend to be easier to detect and classify.

305

**Table 3** Test dataset measurements characteristics by occupancy period

Measurement name	Maximum value	Minimum value	Mean	Standard deviation
(1) Indoor temperature (°C)	22.4	17.3	21.6	0.6
(2) Indoor temperature (°C)	22.0	19.8	20.7	0.7
(1) Indoor relative humidity (%)	58.1	26.5	43.4	6.3
(2) Indoor relative humidity (%)	54.0	34.0	42.8	5.8
(1) Indoor CO <sub>2</sub> concentration (ppm)	1700.7	440.7	932.9	236.9
(2) Indoor CO <sub>2</sub> concentration (ppm)	526.2	411.5	449.2	17.8

## 306 4.2. Data preprocessing

### 307 4.2.1. Feature engineering

308 Feature selection and representation tend to have a direct impact on most models performances [22,  
309 61]. Feature engineering is the process of combining or transforming existing features to create additional  
310 features that are not in the original dataset. The main idea behind feature engineering is to use domain  
311 knowledge, visualization and statistical methods to provide discriminative information from the data that,  
312 the model alone, may not or cannot extract [22]. For this study, two types of additional features are  
313 created based on the heating season specificities presented in section 4.1.1: STL-Residue and  
314 EMA-Difference. These features are built to reflect the impact created by a window opening between two  
315 different environments, indoors and outdoors. Thus, window openings influence ambient indoor  
316 measurements by creating a data point or successive data points with specific values that locally seem to  
317 be inconsistent with the rest of the data. In other words, window openings are represented by specific  
318 patterns that tend to differ from the tendency. Therefore, as shown in Table 4 and represented in , four  
319 features transformations and combinations are applied in this study with the aim of extracting  
320 information that differentiate open and close window-status in measurements:

- 321 • **Exponential Moving Average (EMA):** is a feature transformation based on a smoothing technique  
322 used to reduce measurements noises and only capture important patterns such as windows opening.  
323 The *EMA* for a measurement is calculated following Equation (2) where  $M_t$  is the value of the  
324 measurement at time  $t$ ,  $EMA_{M_t}$  is the value of the EMA for this measurement at time  $t$  and  $\alpha$  is a  
325 constant smoothing (or weight) coefficient ranged between 0 and 1. For this study, an *EMA* is applied  
326 on the data with a light tuned smoothing coefficient ( $\alpha$ ) of 0.10, previously chosen in range between  
327 0.10 and 0.25. A smoothed measurement is referred as  $EMA_{Measurement}$ .

$$EMA(\alpha)_{M_t} = \begin{cases} M_0 & t = 0 \\ \alpha \cdot M_t + (1 - \alpha) \cdot EMA_{M_{t-1}} & t > 0 \end{cases} \quad Eq. (2)$$

- 328 • **Derivation:** is a feature transformation used in order to capture sudden variations on measurements  
329 by differencing seasonal and cyclic drifts (low successive derivate values) and sudden drops such as  
330 windows opening (higher successive derivate values). The *derivation* transformation of a  
331 measurement is presented with Equation (3) where  $M_t$  is the value of the measurement at time  $t$ ,  $d_{M_t}$   
332 is the value of the derivate of this measurement at time  $t$  and  $\Delta t$  is a time step value. A *derivate*  
333 measurement is further referred as  $d_{Measurement}$ .

$$d_{M_t} = \begin{cases} 0 & t = 0 \\ \frac{M_t - M_{t-\Delta t}}{\Delta t} & t > 0 \end{cases} \quad Eq. (3)$$

334 • **STL-Residue:** is a feature transformation based on the Seasonal-Trend decomposition using LOESS  
 335 (STL). STL is a statistical method which decomposes the input data into three components: a  
 336 recurrent pattern over time (seasonality), a tendency (trend) and a residue (or noise) composed of  
 337 random or unpredictable fluctuation [62]. Hence, *STL-Residue* is used to extract unusual pattern from  
 338 measurements such as opening window impacts. The STL-Residue is calculated with Equation (4)  
 339 where  $M_t$  is the value of the measurement at time  $t$  and  $Seasonality_{M_t}$ ,  $Trend_{M_t}$  and  $Residue_{M_t}$  are the  
 340 value of the seasonality, trend and residue for a measurement at time  $t$ , respectively. The *STL-Residue*  
 341 transformation of a measurement is further referred as *residue<sub>Measurement</sub>*.

$$Residue_{M_t} = M_t - Trend_{M_t} - Seasonality_{M_t} \quad Eq. (4)$$

342 • **EMA-Difference:** is a feature combination pursuing the same goal as the *STL-Residue* transformation.  
 343 The *EMA-Difference* for a measurement is calculated using Equation (5) where  $M_t$  is the value of the  
 344 measurement at time  $t$ ,  $EMA_{M_t}$  is the value of the *EMA* for this measurement at time  $t$  and  $Diff_{M_t}$  is the  
 345 value of the *EMA-Difference* for this measurement at time  $t$ . This feature consists of a difference  
 346 between the data and the same data smoothed by an *EMA*. The *EMA* applied is composed of a strong  
 347 tuned smoothing coefficient ( $\alpha$ ), in range between 0.01 and 0.10 to extract the measurement  
 348 tendencies only. Therefore, a high (*resp.* low) value characterizes a measured point that is far from  
 349 (*resp.* close to) the tendency and that is more likely to be unusual (*resp.* usual). Several smoothing  
 350 coefficients were tested on the training set with similar observed results but the one selected for this  
 351 study corresponds to a value of 0.04. The *EMA-Difference* transformation of a measurement is further  
 352 referred as *difference<sub>Measurement</sub>*.

$$Diff_{M_t} = M_t - EMA_{M_t} \quad Eq. (5) \quad 353$$

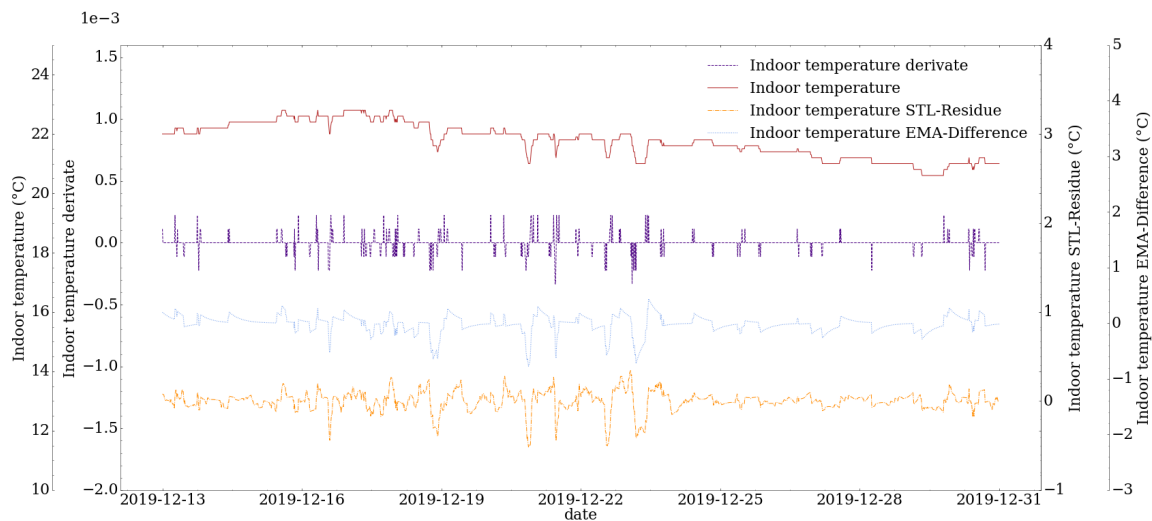
354  
 355 These transformations are applied on indoor measurements such as temperature, relative and absolute  
 356 humidity and CO<sub>2</sub> concentration. They can be associated in order to provide different information from  
 357 the data to the model. Hence, for every indoor measurement, 20 different associations are performed by  
 358 combining the measurement without modifications and the transformations presented above. As shown  
 359 in Table 4 by comparing both periods presented previously on the test set to the training set, derivation,  
 360 *STL-Residue* and *EMA-Difference* transformations appear to provide information without large variation  
 361 such as global or local seasonality by centering the data. The same effect can be observed on three weeks



362 extracted from the training set with Figure 6 where  $d_{temperature}$ ,  $residue_{temperature}$  and  $difference_{temperature}$   
 363 transformations remain centered contrary to the temperature measurement. These transformations are  
 364 intended to push models to be less sensitive to measurements values or global variations and more on  
 365 local dynamics. However some drawbacks might be observed: a lag effect can be noticed on EMA based  
 366 transformations such as EMA-Difference whereas derivation transformations can provide data with a low  
 367 variance that might be, depending on the model, delicate to exploit.

368 **Table 4** Applied data transformation on train and test set temperature measurements

Transformation	Period	Maximum	Minimum	Mean
Temperature	Train	23.0	19.6	21.7
	Test (1)	22.4	17.3	21.6
	Test (2)	22.0	19.8	20.7
EMA smoothing	Train	23.0	20.2	21.7
	Test (1)	22.2	18.7	21.5
	Test (2)	22.0	19.8	20.9
Derivation ( $10^3$ )	Train	0.5	-0.9	0.0
	Test (1)	0.5	-1.0	0.0
	Test (2)	0.2	-0.2	0.0
STL-Residue	Train	0.6	-1.2	0.0
	Test(1)	0.6	-1.5	0.0
	Test (2)	0.4	-0.1	0.0
EMA-Data difference	Train	0.9	-1.7	0.0
	Test (1)	0.6	-2.0	0.0
	Test (2)	0.2	-0.3	0.0



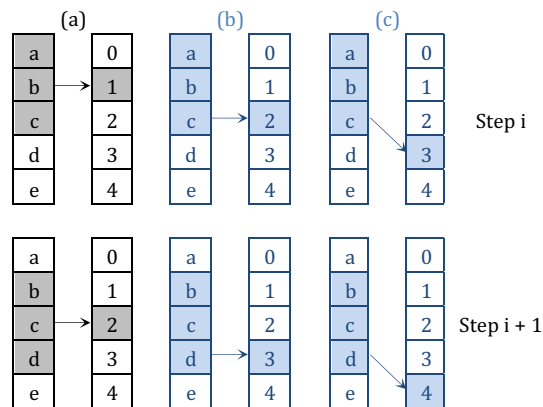
369 **Figure 6** Applied data transformation and combinations representation on 3 training weeks  
 370

371 **4.2.2. Data preparation and association**

372 For this study, more than 800 different combinations of measurements and measurements  
 373 transformations and associations were performed for every model. More specifically a base of 20 different

374 associations, as presented in section 4.2.1, was processed for every indoor measurement (temperature,  
 375 humidity and CO<sub>2</sub> concentration). Based on the previous combination results, 175 additional  
 376 combinations were performed for dual indoor measurement combinations (temperature + humidity and  
 377 temperature + CO<sub>2</sub>). By following a similar process, around 100 different combinations for triple indoor  
 378 measurements combination were created with around 300 more by adding outdoor measurements  
 379 (temperature and humidity) to all previous combinations. Except for RFC models, features are normalized  
 380 between 0 and 1 based on the training set characteristics.

381 Unlike other models, LSTM and GRU models are trained in a “many to one” way with an overlapping  
 382 sliding window that moves one step ahead. The size of the observation window is set to three hours of  
 383 data in order to find a good balance between time training and performances. As shown in Figure  
 384 7Erreur ! Source du renvoi introuvable. (7.a) on a three time step sample, the overlapping sliding  
 385 windows was centered in a way that the target corresponds to the window-status at the center of the  
 386 observation window. For reference and as presented in (7.b), the sliding window for real-time application  
 387 targets the last window-status while it targets one time step-ahead for prediction purposes (7.c).



388 **Figure 7** Training window for LSTM and GRU models for: (a) past window-openings, (b) real-time window-openings and (c)  
 389 window-openings prediction

## 390 5. Results and discussion

391 Due to the specificities of LSTM, GRU and RFC models training, each of the 800 combinations has been  
 392 performed ten times. Thus, only the model output with the best F1-score and accuracy results out of the  
 393 ten was retained for evaluation. Since LDA and SVM models provide more stable outputs, each of the 800  
 394 combinations was only performed once. To not overload this study and since relative humidity  
 395 systematically provides poorer results than absolute humidity, only absolute humidity based associations

396 are presented. Hence, absolute humidity is further referred simply as humidity. All the following results  
 397 are evaluated regarding their F1-score and alternative metrics presented in section 3.1 in order to discuss  
 398 about measurements transformations, associations and combinations that might appear to be the most  
 399 adequate to detect window openings. Models raw performances to recognize past window openings will  
 400 be taken into consideration but are not the sole focus of this study. Thus, a comparison of the observed  
 401 results between models and combination is preferred.

402 *5.1. Transformations and associations impact on one indoor measurement based on F1-score*

403 An overview of the performances of each model trained on the twenty bases transformations and  
 404 associations (4.2.1) for every indoor measurement is presented in Table 5 Average F1-score and standard  
 405 deviation of the 20 best results for each measurements combination and models. It appears that,  
 406 regardless of the transformation or association used, indoor air humidity or CO<sub>2</sub> concentration sole base  
 407 combination (referred as H<sub>in</sub> and C<sub>in</sub>, respectively) does not provide good results on window-status  
 408 detection with an average F1-score that, at best, usually does not exceed 0.39. On the other hand, indoor  
 409 air temperature sole base combinations (T<sub>in</sub>) seem to provide better and exploitable results with a higher  
 410 F1-score average for all models and especially for RNN based models with an average value close to 0.70.

411 **Table 5** Average F1-score and standard deviation of the 20 best results for each measurements combination and models

<b>F1-score:</b> <i>mean ± standard deviation</i>	<b>GRU</b>	<b>LSTM</b>	<b>LDA</b>	<b>SVM</b>	<b>RFC</b>
Indoor absolute humidity (H <sub>in</sub> )	0.35 ± 0.23	0.39 ± 0.20	0.11 ± 0.11	0.18 ± 0.16	0.21 ± 0.13
Indoor CO <sub>2</sub> (C <sub>in</sub> )	0.28 ± 0.15	0.19 ± 0.14	0.01 ± 0.01	0.04 ± 0.04	0.18 ± 0.08
Indoor temperature (T <sub>in</sub> )	<b>0.73 ± 0.07</b>	<b>0.68 ± 0.18</b>	<b>0.54 ± 0.24</b>	<b>0.38 ± 0.28</b>	<b>0.46 ± 0.21</b>

412 However these results present a high dispersion that can be explained by looking at the F1-score  
 413 results for the five base transformations without associations shown in Table 6. For air humidity  
 414 measurements, it appears that *EMA-Difference* and *STL-Residue* based transformations increase models  
 415 performances compared to other transformations or base measurement. Best results are observed with  
 416 *difference<sub>humidity</sub>* for both LSTM and GRU models with an F1-score around 0.70. Regarding air temperature,  
 417 the dispersion might be caused by the difficulty of all models to provide results when trained on  
 418 untransformed temperature, *EMA<sub>temperature</sub>* or some other measurements transformations including them.  
 419 These results might be explained by the differences between the training and testing set. Testing set  
 420 measurements reach values and dynamics unseen during the training phase, allowing less contextual

421 transformations such as the *derivate*, the *STL-Residue* or the *EMA-Difference* to perform better and  
 422 increase the F1-score from 0.60 to 0.77 for LSTM and GRU models (representing an improvement of 25 %)  
 423 and from around 0.00 to 0.6 for LDA, SVM and RFC models . Results including different measurements  
 424 combination based on the same transformations and associations are presented in the following section.

425

426

**Table 6** F1-score for indoor humidity and temperature base transformations

Transformation F1-score	GRU	LSTM	LDA	SVM	RFC
$T_{in}$	0.60	0.57	0.00	0.04	0.04
EMA. $T_{in}$	0.54	0.00	0.00	0.00	0.03
Derivate. $T_{in}$	<b>0.77</b>	<b>0.75</b>	0.25	0.25	0.50
EMA-Difference. $T_{in}$	0.75	0.72	<b>0.65</b>	<b>0.62</b>	<b>0.60</b>
STL-Residue. $T_{in}$	0.77	0.70	0.55	0.54	0.49
$H_{in}$	0.00	0.26	0.00	0.00	0.04
EMA. $H_{in}$	0.00	0.00	0.00	0.00	0.06
Derivate. $H_{in}$	0.43	0.40	0.16	0.14	0.07
EMA-Difference. $H_{in}$	<b>0.72</b>	<b>0.71</b>	0.13	0.17	<b>0.41</b>
STL-Residue. $H_{in}$	0.57	0.52	<b>0.41</b>	<b>0.40</b>	0.35

427 *5.2. Measurements selection and combination impact based on F1-score*

428 In order to have an overview of the best achievable performances of each model depending on the  
 429 measurements combination used, the best twenty models are selected based on F1-score, for each  
 430 performed combination. The average and standard deviation of those twenty best models outputs are  
 431 presented in [Table 7](#) with a total of 280 combinations out of the 800 originals for each model. Dual indoor  
 432 measurements combinations ( $T_{in} + H_{in}$  and  $T_{in} + C_{in}$ ) usually tend to provide higher opening window  
 433 detection performances with a systematic increase of the maximum and average F1-score. For all models,  
 434 indoor  $CO_2$  concentration combined with indoor temperature seems to provide significate higher  
 435 performances than humidity and temperature combinations. This performance enhancement is reflected  
 436 by a consistent improvement on F1-scores averages for all models compared to temperature and humidity  
 437 combinations. The combination of the three indoor measurements ( $T_{in} + H_{in} + C_{in}$ ) seems to provide only  
 438 slight to no improvement for all models on window opening detection compared to temperature and  $CO_2$   
 439 combination. Based on F1-scores it appears that enough information are provided during training with  
 440 this dual combination for LSTM, GRU, LDA and RFC models contrary to the SVM. Hence, LSTM and GRU

441 models output on opening window detection seem to be capped around an average F1-score of 0.76 - 0.78  
 442 whereas LDA, SVM and RFC models appear to be lower limited around a 0.72 - 0.73 average score.

443 Lastly and LDA model apart, the addition of outdoor humidity measurement ( $H_{out}$ ), outdoor  
 444 temperature measurement ( $T_{out}$ ) or both ( $T_{out} + H_{out}$ ) to all indoor measurement combination tend to  
 445 usually deteriorate all models window-opening detections with a common decrease of the maximum and  
 446 the average F1-score. However, even if indoor and outdoor temperature combination appears to slightly  
 447 deteriorate the best attainable performances with a drop of average F1-score for RNN models, it appears  
 448 to be more relevant for LDA, SVM and RFC models than the sole use of indoor air temperature.

449 Of all models, LSTM and GRU appear to be the most efficient ones in order to detect window-status with  
 450 the best average and maximum F1-scores and thus even with just one measurement. Both of these models,  
 451 including RFC, tend also to be sensitive to the addition of, what appears to be, sub-optimal measurements  
 452 and might need proper data selection or transformation. LDA and SVM seem to be more reliable with a  
 453 low repartition of results and by their tendency to improve or to maintain their performances despite the  
 454 addition of measurements that worsen other models results. On the contrary the RFC model appears to be  
 455 the less consistent and sensitive one.

456 **Table 7** Average F1-score and standard deviation of the 20 best results for each measurements combination and models

<b>F1-score:</b> <i>average ± standard deviation</i>	<b>GRU</b>	<b>LSTM</b>	<b>LDA</b>	<b>SVM</b>	<b>RFC</b>
Indoor absolute humidity ( $H_{in}$ )	0.35 ± 0.23	0.39 ± 0.20	0.11 ± 0.11	0.18 ± 0.16	0.21 ± 0.13
Indoor CO <sub>2</sub> ( $C_{in}$ )	0.28 ± 0.15	0.19 ± 0.14	0.01 ± 0.01	0.04 ± 0.04	0.18 ± 0.08
Indoor temperature ( $T_{in}$ )	0.73 ± 0.07	0.68 ± 0.18	0.54 ± 0.24	0.38 ± 0.28	0.46 ± 0.21
$T_{in} + T_{out}$	0.70 ± 0.03	0.70 ± 0.03	0.67 ± 0.03	0.66 ± 0.04	0.62 ± 0.08
$T_{in} + H_{in}$	0.76 ± 0.01	0.75 ± 0.02	0.69 ± 0.01	0.68 ± 0.03	0.65 ± 0.04
$T_{in} + H_{in} + T_{out}$	0.71 ± 0.03	0.69 ± 0.05	0.68 ± 0.02	0.68 ± 0.04	0.63 ± 0.11
$T_{in} + H_{in} + H_{out}$	0.71 ± 0.03	0.71 ± 0.02	0.69 ± 0.01	0.66 ± 0.04	0.53 ± 0.05
$T_{in} + H_{in} + T_{out} + H_{out}$	0.73 ± 0.03	0.72 ± 0.02	0.70 ± 0.01	0.68 ± 0.03	0.53 ± 0.07
$T_{in} + C_{in}$	<b>0.78 ± 0.01</b>	<b>0.76 ± 0.01</b>	<b>0.72 ± 0.01</b>	<b>0.70 ± 0.01</b>	<b>0.73 ± 0.02</b>
$T_{in} + C_{in} + T_{out}$	0.70 ± 0.03	0.65 ± 0.05	0.71 ± 0.02	0.67 ± 0.02	0.71 ± 0.04
$T_{in} + H_{in} + C_{in}$	<b>0.78 ± 0.01</b>	<b>0.76 ± 0.01</b>	0.71 ± 0.01	<b>0.72 ± 0.01</b>	<b>0.73 ± 0.01</b>
$T_{in} + H_{in} + C_{in} + T_{out}$	0.70 ± 0.05	0.64 ± 0.07	0.71 ± 0.01	0.69 ± 0.04	0.71 ± 0.04
$T_{in} + H_{in} + C_{in} + H_{out}$	0.73 ± 0.02	0.70 ± 0.02	0.72 ± 0.01	0.70 ± 0.02	0.68 ± 0.02
$T_{in} + H_{in} + C_{in} + T_{out} + H_{out}$	0.72 ± 0.01	0.69 ± 0.03	<b>0.73 ± 0.01</b>	0.69 ± 0.02	0.65 ± 0.02

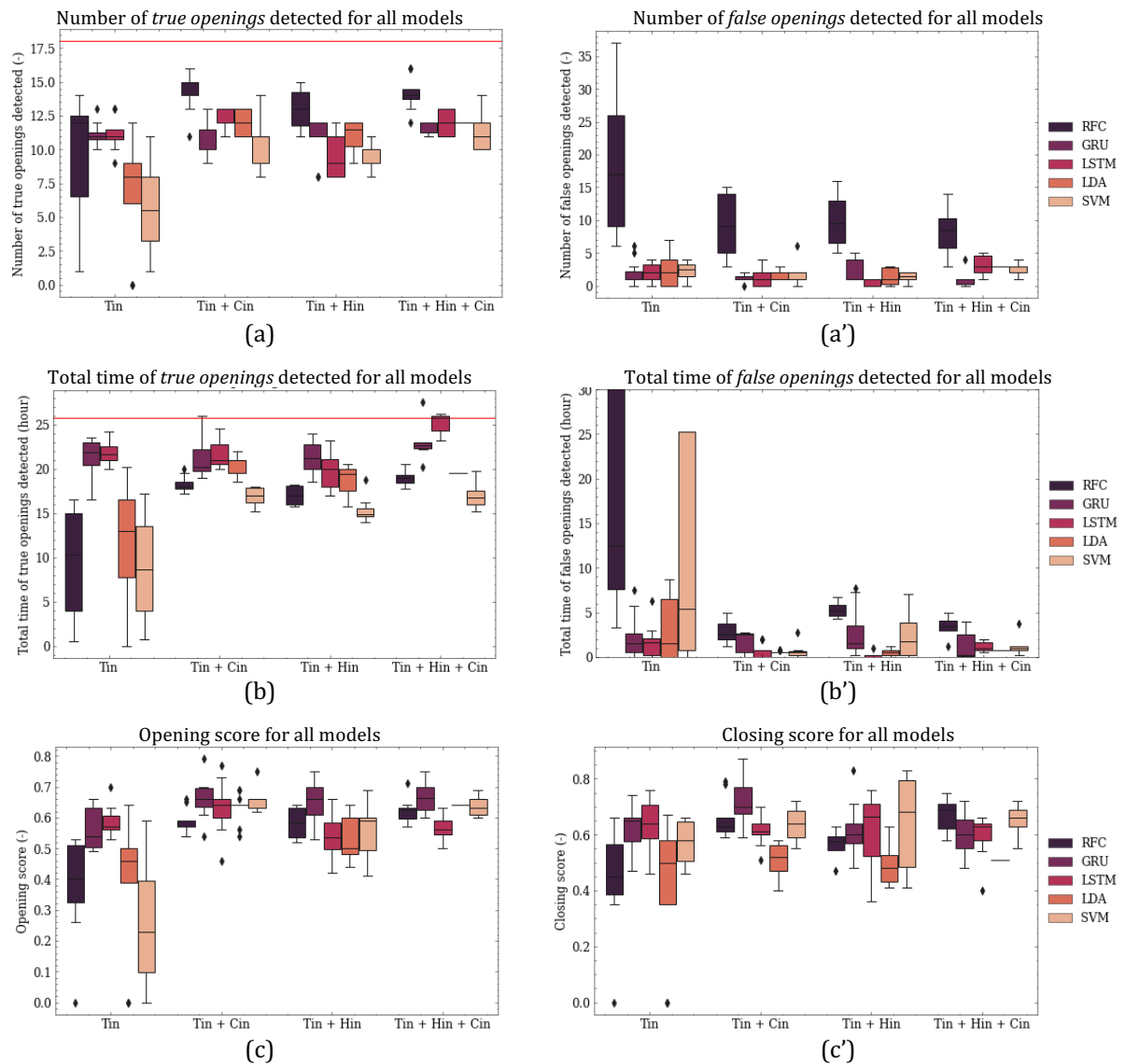
457 To conclude, even if the combination of the three indoor measurements ( $T_{in} + H_{in} + C_{in}$ ) seems to  
 458 provide the best results on opening detection regardless of the model, two indoor measurements such as  
 459  $T_{in} + C_{in}$  or even  $T_{in} + H_{in}$ , are likely to be sufficient to provide good or great results for all models. Although

460 very fluctuating with variations that are not only related to windows openings (occupancy, occupant  
461 position in the room, natural air movement) the indoor CO<sub>2</sub> concentration measurement seems to be  
462 preferable to indoor humidity. Furthermore, depending on the transformation used, the sole indoor  
463 temperature measurement proves to be consistent enough to provide opening window detection results  
464 on par with dual or triune combinations. Due to the observed tendency of outdoor measurements to  
465 decrease opening detection performance for the majority of the models, this study will further be focused  
466 on indoor measurements.

### 467 *5.3. Measurements selection and combination impact based on additional metrics*

468 The additional evaluations metrics introduced in section 0 are used in order to provide more in-depth  
469 explanations on the differences observed and described previously. The number of true and false  
470 openings, the total time (in hour) of true and false opening and the opening and closing score are recorded  
471 as a boxplot repartition in Figure 8. This figure is constructed by using the same best twenty models  
472 output, based on F1-scores, for each performed combination as in Table 7. However, a specific focus on  
473 one to three indoor measurements combinations is made with T<sub>in</sub>, T<sub>in</sub> + C<sub>in</sub>, T<sub>in</sub> + H<sub>in</sub> and T<sub>in</sub> + H<sub>in</sub> + C<sub>in</sub> that  
474 represent a total of 80 combinations out of the 470 originals for each model. Regarding the measurements  
475 combination, Figure 8 shows that for all models and for all additional metrics, indoor temperature and CO<sub>2</sub>  
476 combinations appear to perform slightly better than indoor temperature and humidity combinations. The  
477 difference seems to be mostly due to the fact that CO<sub>2</sub> based combinations tend to have a higher capacity  
478 to get better maximum results for true openings detections (8.a), opening and closing score (8.c and 8.c').  
479 This observation might be explained by the propensity of the CO<sub>2</sub> concentration to fluctuate on higher  
480 levels than the humidity and thus, with an adequate transformation, to detect or to better define a few  
481 more openings. Furthermore, for all models, aside of GRU and LSTM, the use of a combination of minimum  
482 two indoor measurements provide a clear improvement on the results compared to the sole use of the  
483 indoor temperature even if the higher score tend to be close to all other combinations.

484



486 **Figure 8** Additional evaluation metrics for every model and measurement combination

486

487

488

489

490

491

492

493

494

495

Based on F1-scores presented in Table 7, GRU and LSTM models seem to produce similar results and follow identical tendencies. Figure 8 shows that whatever the combination is, LSTM models predictions seem to detect more false openings (8.a') that are rather small with a lower average total time of false opening (8.b') whereas GRU models detections appear to be more precise in defining opening and closing window-status (8.c and 8.c'). SVM and LDA models also seem to provide rather close opening detection results but SVM models predictions appear to provide the worst rate of true opening detection (8.a) while LDA models appears to heavily underperform in closing window precision (8.c'). The RFC model seems to be the most sensitive model with the highest number of true opening (8.a) and false opening (8.a') detected from all models that, apart from the sole indoor temperature combination, appear to be short.

496 To conclude, although all models in Figure 8 present an average number of true openings of 10 to 13  
497 out of the 18 existing (represented as a red line in 8.a), this result should be balanced. As explained in  
498 4.1.2, several openings have no or little impact on the indoor environment, and thus are harder (or  
499 impossible) to detect. However, all models tend to detect an average of 18 to 22 hours of opening out of  
500 the 25.75 existing (represented as a red line in 8.b) for an average of 30 minutes to 2 hours of false  
501 opening. Additionally, LSTM and GRU models show satisfying results with the sole use of indoor  
502 temperature by detecting on average 84 to 88 % of opening time (21 to 22 hours out of 25.75) for an  
503 average of 1.5 to 2 hours of false openings while the use of a second measurement tend to be needed for  
504 other models to present an average of 66 to 77 % of opening time (17 to 20 hours out of 25.75). These  
505 results tend to show that most of the impactful openings are detected over this one month test period. The  
506 major negative point and realistic way to improve seems to be based on improving opening and closing  
507 precision that always seem to be more than 1 time step too early or late with an average score of 0.50 or  
508 0.60.

#### 509 5.4. Measurements transformations and association impact based on F1-score

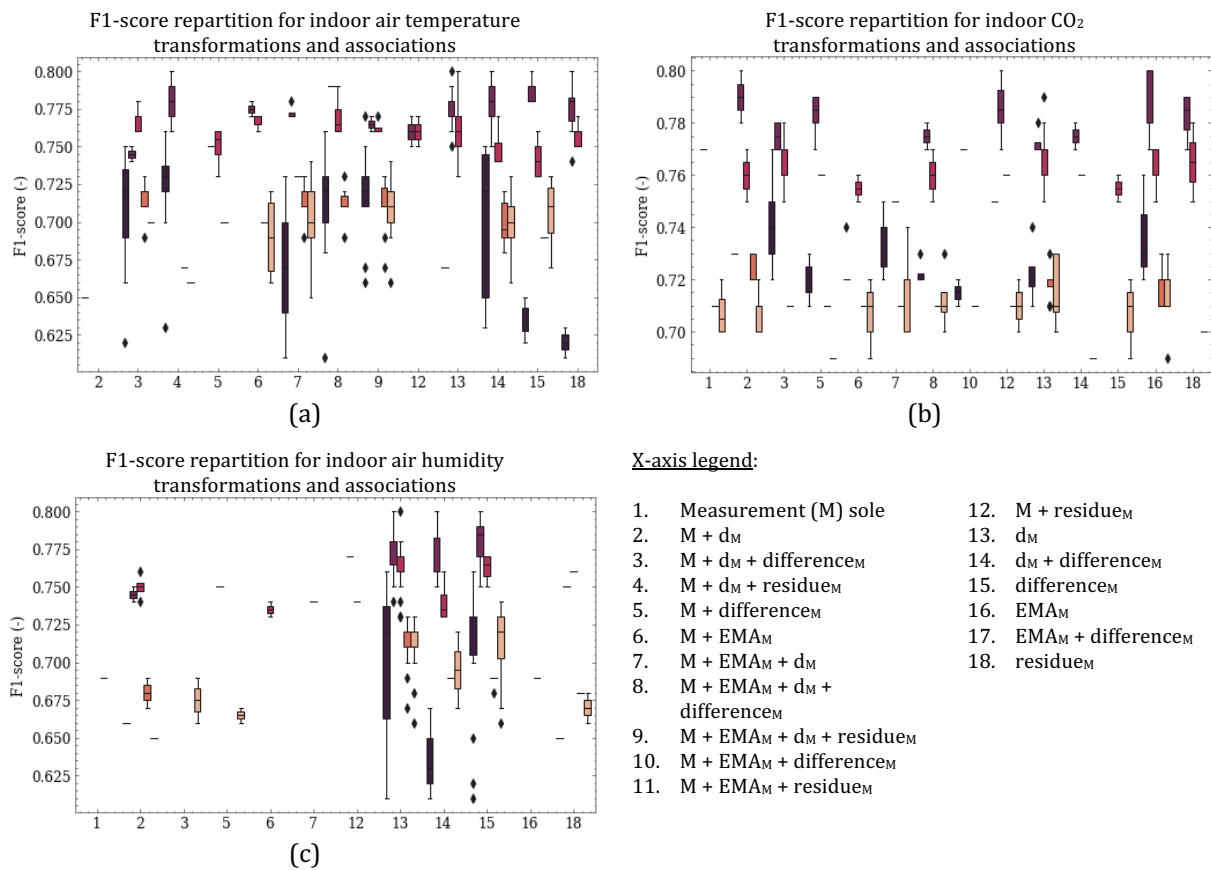
510 In order to have an overview of the best performance of each model depending on the indoor  
511 measurements transformations and associations used, the best twenty models output, based on F1-score,  
512 for two and three indoor combinations ( $T_{in} + C_{in}$ ,  $T_{in} + H_{in}$  and  $T_{in} + H_{in} + C_{in}$ ) are studied. A boxplot of these  
513 twenty best models outputs is presented in Figure 9 with a total of 60 combinations out of the 450  
514 originals for each model. All measurements (referred as M) transformations and associations that are not  
515 part of the top twenty are not represented on this figure. Similarly, those under or low represented are  
516 displayed as a box with a small repartition or a small horizontal line.

517 Regarding humidity transformations and associations the three most recurrent and efficient for all  
518 models appear to be, by far, solely composed of  $d_{humidity}$  (9.c.13),  $difference_{humidity}$  (9.c.15) or both (9.c.14).  
519 Contrary to  $difference_{humidity}$  that is a measurement transformation which is sufficient to detect  
520 window-opening as detailed in section 5.2,  $d_{humidity}$  seem to perform better combined with other  
521 measurements. For temperature transformations and associations a distinction has to be made between  
522 models. The sole  $d_{temperature}$  (9.a.13) seem to be consistent for LSTM and GRU models whereas it has to be  
523 combined with other transformations for LDA, SVM and RFC models (9.a.3, a.7, a.8, a.9 and a.14). For these  
524 last models, a large amount of temperature transformation association seems to be preferred in order to



525 get good results contrary to RNN models that appear to perform well with just one or no transformation  
 526 associations such as  $d_{temperature}$ ,  $difference_{temperature}$  or  $residue_{temperature}$  (9.a.13, a.15, a.18). Contrary to the  
 527 previous observations, a consensus doesn't seem to appear for CO<sub>2</sub> transformations associations.  
 528 Associations based on  $d_{CO_2}$  or a smoothed  $EMA_{CO_2}$  seems to be a bit more present and thus effective (9.b.2,  
 529 b.5, b.13 and b.16).

530 Overall and due to their absence or under representation, the use of sole measurements or sole  
 531 exponential moving average is not recommended regardless of the model. On the other hand, the sole use  
 532 of *derivation*, *STL-Residue* or *EMA-Difference* transformations appear to be enough to provide good results  
 533 on window opening detections for RNN models while LDA, SVM and RFC models favor the same  
 534 transformations but associated together.



535 **Figure 9** F1-score measurement transformation and associations for all tested combination best twenty F1-score

537 **5.5. Discussion and future work**

538 Measurement combination and transformation were performed on various machine learning models in  
 539 order to assess their efficiency and relevance on past window-status detection. However, although

540 performed on models that are not yet widely used on this domain, several limitations remains. Most  
541 transformations applied on measurements (*STL-Residue*, *derivate* or *EMA-Difference*) are built to reflect an  
542 impact created by a window opening between two environments with different characteristics (e.g., air  
543 temperature, CO<sub>2</sub> concentration). Thus, despite showing great results by improving and stabilizing models  
544 performances, they might not be suitable in other climates or seasons and need to be carefully evaluated  
545 beforehand. Therefore, a similar process will be followed on other seasons in order to highlight  
546 appropriate measurements combinations, transformations and associations.

547 Regarding measurements selection, LSTM and GRU models achieve satisfying results with the sole use  
548 of indoor temperature measurements although the addition of indoor CO<sub>2</sub> concentration appears to  
549 stabilize and slightly improve their results. For SVM, LDA and RFC models, the use of minimum both  
550 indoor temperature and CO<sub>2</sub> concentration tend to be recommended even if a small improvement in  
551 results can be observed by adding indoor humidity measurement.

552 It appears that, for untransformed data, results observed in other studies are consistent regarding the  
553 most important features that are indoor and outdoor air temperature [15]. However, the observed  
554 tendency tend to change when transformations are applied on indoor measurements and results  
555 deterioration can be observed by adding outdoor measurements. Furthermore, it is important to note that,  
556 as experienced by [40], air humidity appears to have a low impact on models.

557 Additional metrics introduced in this study provided a different perspective on models performances  
558 regarding window-status detection. These metrics offer a field perspective approach on models results  
559 that might allow selecting the model that best suits the needs for a project (e.g., by privileging the number  
560 of detected openings over their accuracy) or comparing results between relevant studies. However, unlike  
561 commonly used metrics adapted to unbalanced classes such as F1-score, their implementation is heavy  
562 and requires investigating simultaneously six different metrics.

563 A similar process is followed for real-time detection and future window-status prediction. The same  
564 work, conducted on real-time detection, shows identical results for RFC, LDA and SVM models as those  
565 presented in Table 7. However, an average drop of 0.01 to 0.10 on the average F1-score is observed for  
566 LSTM and GRU models. These differences are presented for both models in Table 8. It appears that LSTM  
567 performs significantly worse than the GRU for real-time detection despite being still better than other  
568 models. These differences can mainly be explained with additional metrics and are due to a drop in  
569 accuracy regarding window opening scores. In addition a predictive approach is currently in progress.

570 **Table 8** Average F1-score and standard deviation of the 20 best results for each measurements combination and models for past and  
 571 real time window-status detection

<b>F1-score:</b> <i>average ± standard deviation</i>		<b>GRU</b>	<b>LSTM</b>
T <sub>in</sub>	past	0.73 ± 0.07	0.68 ± 0.18
	real time	0.63 ± 0.17	0.59 ± 0.17
T <sub>in</sub> + T <sub>out</sub>	past	0.70 ± 0.03	0.70 ± 0.03
	real time	0.59 ± 0.10	0.56 ± 0.05
T <sub>in</sub> + H <sub>in</sub>	past	0.76 ± 0.01	0.75 ± 0.02
	real time	0.71 ± 0.02	0.67 ± 0.03
T <sub>in</sub> + C <sub>in</sub>	past	<b>0.78 ± 0.01</b>	<b>0.76 ± 0.01</b>
	real time	<b>0.76 ± 0.01</b>	<b>0.72 ± 0.02</b>
T <sub>in</sub> + H <sub>in</sub> + C <sub>in</sub>	past	<b>0.78 ± 0.01</b>	<b>0.76 ± 0.01</b>
	real time	<b>0.77 ± 0.01</b>	<b>0.73 ± 0.02</b>

572 **6. Conclusion**

573 This study presents a comparison of the performance of Gated Recurrent Unit (GRU), Long Short Term  
 574 Memory (LSTM), Linear Discriminant Analysis (LDA), Support Vector Machine (SVM) and Random Forest  
 575 Classifier (RFC) models in detecting window openings depending on several indoor and outdoor  
 576 measurements combinations, transformations and associations in the field of building energy during  
 577 heating season. The results showed that not only the choice of input data measurement was essential to  
 578 obtain satisfactory results but also that it was neither always optimum nor required to add more  
 579 information to the input of the models (e.g., outdoor measurements) and that a preliminary selection  
 580 might be necessary. Hence, if required, the sole use of a temperature sensor with adapted transformation  
 581 (e.g., *temperature derivate*, *temperature STL-Residue* or *temperature EMA-Difference*) might be sufficient to  
 582 provide satisfying results for window-openings detection. Adding other indoor measurements appears  
 583 recommended to obtain slightly more precise results for LSTM and GRU models and necessary for other  
 584 models. In this case, the combination of indoor temperature and CO<sub>2</sub> concentration measurement seems to  
 585 be the one to be privileged for all models.

586 This work also showed that a simple transformation of the data beforehand (e.g., *derivate*) or more  
 587 complex ones introduced in this paper (*STL-Residue* or *EMA-Difference*) could have a significant positive  
 588 impact on the quality of the window-openings detections by turning unusable results (e.g., temperature

589 sole or with other combinations) to satisfactory results. Depending on the model used, specific association  
590 of measurement transformation might be appropriate.

591 Furthermore, the additional metrics evaluations show that despite satisfying F1-scores results, the  
592 number of openings detected by all models may seem low (10 to 13 predicted out of 18 measured in total)  
593 but several openings have no or little impact on the indoor environment (a temperature decrease of 0.2°C  
594 for instance) and thus, does not offer enough information to the models to detect them. However, all  
595 models tend to detect an average of 18 to 22 hours of opening out of the 25.75 existing for an average of  
596 30 minutes to 2 hours of false opening. These results tend to show that the most impactful openings are  
597 detected over this one month test period. Thus, this may not be an issue depending of the application of  
598 these models, such as the estimation of the thermal losses of a building linked to window openings for  
599 example.

## 600 **Acknowledgements**

601 This research was conducted within the framework of the Intellib@t project within the joint research  
602 team GP2E (as in French, Guarantee of energy and environmental performances), between the I2M  
603 laboratory and Nobatek/INEF4, with funding from the Regional Committee of Nouvelle Aquitaine (CNRA)  
604 and the Investments for the Future Program (PIA).

## 7. Bibliography

- [1] European Commission. Directorate General for Energy., *Clean energy for all Europeans*. LU: Publications Office, 2019. Accessed: Sep. 28, 2022. [Online]. Available: <https://data.europa.eu/doi/10.2833/21366>
- [2] T. Ramesh, R. Prakash, and K. K. Shukla, 'Life cycle energy analysis of buildings: An overview', *Energy and Buildings*, vol. 42, no. 10, pp. 1592–1600, Oct. 2010, doi: 10.1016/j.enbuild.2010.05.007.
- [3] N. Aste, M. Manfren, and G. Marenzi, 'Building Automation and Control Systems and performance optimization: A framework for analysis', *Renewable and Sustainable Energy Reviews*, vol. 75, pp. 313–330, Aug. 2017, doi: 10.1016/j.rser.2016.10.072.
- [4] Imran, N. Iqbal, and D. H. Kim, 'IoT Task Management Mechanism Based on Predictive Optimization for Efficient Energy Consumption in Smart Residential Buildings', *Energy and Buildings*, vol. 257, p. 111762, Feb. 2022, doi: 10.1016/j.enbuild.2021.111762.
- [5] Z. Yang and B. Becerik-Gerber, 'The coupled effects of personalized occupancy profile based HVAC schedules and room reassignment on building energy use', *Energy and Buildings*, vol. 78, pp. 113–122, Aug. 2014, doi: 10.1016/j.enbuild.2014.04.002.
- [6] M. A. Hannan *et al.*, 'A Review of Internet of Energy Based Building Energy Management Systems: Issues and Recommendations', *IEEE Access*, vol. 6, pp. 38997–39014, 2018, doi: 10.1109/ACCESS.2018.2852811.
- [7] B. Brik, M. Esseghir, L. Merghem-Boulahia, and H. Snoussi, 'An IoT-based deep learning approach to analyse indoor thermal comfort of disabled people', *Building and Environment*, vol. 203, p. 108056, Oct. 2021, doi: 10.1016/j.buildenv.2021.108056.
- [8] A. Ioannou, L. Itard, and T. Agarwal, 'In-situ real time measurements of thermal comfort and comparison with the adaptive comfort theory in Dutch residential dwellings', *Energy and Buildings*, vol. 170, pp. 229–241, Jul. 2018, doi: 10.1016/j.enbuild.2018.04.006.
- [9] C. Fan, F. Xiao, and C. Yan, 'A framework for knowledge discovery in massive building automation data and its application in building diagnostics', *Automation in Construction*, vol. 50, pp. 81–90, Feb. 2015, doi: 10.1016/j.autcon.2014.12.006.

- [10] H. Esen, M. Inalli, A. Sengur, and M. Esen, 'Predicting performance of a ground-source heat pump system using fuzzy weighted pre-processing-based ANFIS', *Building and Environment*, vol. 43, no. 12, pp. 2178–2187, Dec. 2008, doi: 10.1016/j.buildenv.2008.01.002.
- [11] H. Esen, M. Inalli, A. Sengur, and M. Esen, 'Artificial neural networks and adaptive neuro-fuzzy assessments for ground-coupled heat pump system', *Energy and Buildings*, vol. 40, no. 6, pp. 1074–1083, Jan. 2008, doi: 10.1016/j.enbuild.2007.10.002.
- [12] E. Delzendeh, S. Wu, A. Lee, and Y. Zhou, 'The impact of occupants' behaviours on building energy analysis: A research review', *Renewable and Sustainable Energy Reviews*, vol. 80, pp. 1061–1071, Dec. 2017, doi: 10.1016/j.rser.2017.05.264.
- [13] M. Bonte, F. Thellier, and B. Lartigue, 'Impact of occupant's actions on energy building performance and thermal sensation', *Energy and Buildings*, vol. 76, pp. 219–227, Jun. 2014, doi: 10.1016/j.enbuild.2014.02.068.
- [14] S. Barlow and D. Fiala, 'Occupant comfort in UK offices—How adaptive comfort theories might influence future low energy office refurbishment strategies', *Energy and Buildings*, vol. 39, no. 7, pp. 837–846, Jul. 2007, doi: 10.1016/j.enbuild.2007.02.002.
- [15] X. Dai, J. Liu, and X. Zhang, 'A review of studies applying machine learning models to predict occupancy and window-opening behaviours in smart buildings', *Energy and Buildings*, vol. 223, p. 110159, Sep. 2020, doi: 10.1016/j.enbuild.2020.110159.
- [16] L. Wang and S. Greenberg, 'Window operation and impacts on building energy consumption', *Energy and Buildings*, vol. 92, pp. 313–321, Apr. 2015, doi: 10.1016/j.enbuild.2015.01.060.
- [17] L. Schnelle, G. Lichtenberg, and C. Warnecke, 'Using Low-rank Multilinear Parameter Identification for Anomaly Detection of Building Systems', *IFAC-PapersOnLine*, vol. 55, no. 6, pp. 470–475, 2022, doi: 10.1016/j.ifacol.2022.07.173.
- [18] L. Erhan *et al.*, 'Smart anomaly detection in sensor systems: A multi-perspective review', *Information Fusion*, vol. 67, pp. 64–79, Mar. 2021, doi: 10.1016/j.inffus.2020.10.001.
- [19] S. D'Oca and T. Hong, 'A data-mining approach to discover patterns of window opening and closing behavior in offices', *Building and Environment*, vol. 82, pp. 726–739, Dec. 2014, doi: 10.1016/j.buildenv.2014.10.021.
- [20] R. Markovic, 'Window opening model using deep learning methods', *Building and Environment*, p. 11, 2018.
- [21] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to data mining*, New internat. edition. Harlow: Pearson, 2014.
- [22] Y. Bengio, A. Courville, and P. Vincent, 'Representation Learning: A Review and New Perspectives'. arXiv, Apr. 23, 2014. Accessed: Aug. 03, 2022. [Online]. Available: <http://arxiv.org/abs/1206.5538>
- [23] M. A. R. Lopes, C. H. Antunes, A. Reis, and N. Martins, 'Estimating energy savings from behaviours using building performance simulations', *Building Research & Information*, vol. 45, no. 3, pp. 303–319, Apr. 2017, doi: 10.1080/09613218.2016.1140000.
- [24] European Environment Agency, 'Final energy consumption by sector and fuel'. 2015. [Online]. Available: <https://www.eea.europa.eu/data-and-maps/indicators/final-energy-consumption-by-sector-9/assessment-1>
- [25] J. T. Recht, J. Goffart, L. Mora, M. Woloszyn, and C. Buhé, 'Predicted and measured performances of near zero-energy houses: a comparison methodology', IBPSA Rome, Italy, 2019, p. 7
- [26] C. V. Gallagher, K. Leahy, P. O'Donovan, K. Bruton, and D. T. J. O'Sullivan, 'Development and application of a machine learning supported methodology for measurement and verification (M&V) 2.0', *Energy and Buildings*, vol. 167, pp. 8–22, May 2018, doi: 10.1016/j.enbuild.2018.02.023.
- [27] T. Hong, Z. Wang, X. Luo, and W. Zhang, 'State-of-the-art on research and applications of machine learning in the building life cycle', *Energy and Buildings*, vol. 212, p. 109831, Apr. 2020, doi: 10.1016/j.enbuild.2020.109831.
- [28] B. Grillone, S. Danov, A. Sumper, J. Cipriano, and G. Mor, 'A review of deterministic and data-driven methods to quantify energy efficiency savings and to predict retrofit scenarios in buildings', *Renewable and Sustainable Energy Reviews*, vol. 131, p. 110027, Oct. 2020, doi: 10.1016/j.rser.2020.110027.
- [29] S. Mahmoud, A. Lotfi, and C. Langensiepen, 'User Activities Outliers Detection; Integration of Statistical and Computational Intelligence Techniques: USER ACTIVITIES OUTLIERS DETECTION', *Computational Intelligence*, vol. 32, no. 1, pp. 49–71, Feb. 2016, doi: 10.1111/coin.12045.
- [30] D. Djenouri, R. Laidi, Y. Djenouri, and I. Balasingham, 'Machine Learning for Smart Building Applications: Review and Taxonomy', *ACM Comput. Surv.*, vol. 52, no. 2, pp. 1–36, Mar. 2020, doi: 10.1145/3311950.
- [31] O. Ardakanian, A. Bhattacharya, and D. Culler, 'Non-Intrusive Techniques for Establishing Occupancy Related Energy Savings in Commercial Buildings', in *Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments*, Palo Alto CA USA, Nov. 2016, pp. 21–30. doi: 10.1145/2993422.2993574.
- [32] M. Anastasiadou, V. Santos, and M. S. Dias, 'Machine Learning Techniques Focusing on the Energy Performance of Buildings: A Dimensions and Methods Analysis', *Buildings*, vol. 12, no. 1, p. 28, Dec. 2021, doi: 10.3390/buildings12010028.
- [33] F. Naspi, M. Arnesano, L. Zampetti, F. Stazi, G. M. Revel, and M. D'Orazio, 'Experimental study on occupants' interaction with windows and lights in Mediterranean offices during the non-heating season', *Building and Environment*, vol. 127, pp. 221–238, Jan. 2018, doi: 10.1016/j.buildenv.2017.11.009.
- [34] H. Kim, T. Hong, and J. Kim, 'Automatic ventilation control algorithm considering the indoor environmental quality factors and occupant ventilation behavior using a logistic regression model', *Building and Environment*, vol. 153, pp. 46–59, Apr. 2019, doi: 10.1016/j.buildenv.2019.02.032.
- [35] R. Markovic *et al.*, 'Comparison of Different Classification Algorithms for the Detection of User's Interaction with Windows in Office Buildings', *Energy Procedia*, vol. 122, pp. 337–342, Sep. 2017, doi: 10.1016/j.egypro.2017.07.333.
- [36] F. Stazi, F. Naspi, and M. D'Orazio, 'Modelling window status in school classrooms. Results from a case study in Italy', *Building and Environment*, vol. 111, pp. 24–32, Jan. 2017, doi: 10.1016/j.buildenv.2016.10.013.
- [37] H. B. Rijal, M. A. Humphreys, and J. F. Nicol, 'Development of a window opening algorithm based on adaptive thermal comfort to predict occupant behavior in Japanese dwellings', *Japan Architectural Review*, vol. 1, no. 3, pp. 310–321, Jul. 2018, doi: 10.1002/2475-8876.12043.
- [38] Z. Shi *et al.*, 'Seasonal variation of window opening behaviors in two naturally ventilated hospital wards', *Building and Environment*, vol. 130, pp. 85–93, Feb. 2018, doi: 10.1016/j.buildenv.2017.12.019.
- [39] F. Haldi and D. Robinson, 'Interactions with window openings by office occupants', *Building and Environment*, vol. 44, no. 12, pp. 2378–2395, Dec. 2009, doi: 10.1016/j.buildenv.2009.03.025.
- [40] R. Markovic, J. Frisch, and C. van Treeck, 'Learning short-term past as predictor of window opening-related human behavior in commercial buildings', *Energy and Buildings*, vol. 185, pp. 1–11, Feb. 2019, doi: 10.1016/j.enbuild.2018.12.012.
- [41] Y. Wei *et al.*, 'Comparison of different window behavior modeling approaches during transition season in Beijing, China', *Building and Environment*, vol. 157, pp. 1–15, Jun. 2019, doi: 10.1016/j.buildenv.2019.04.040.

- [42] R. Chalapathy and S. Chawla, 'Deep Learning for Anomaly Detection: A Survey'. arXiv, Jan. 23, 2019. Accessed: Sep. 29, 2022. [Online]. Available: <http://arxiv.org/abs/1901.03407>
- [43] B. Huchuk, S. Sanner, and W. O'Brien, 'Comparison of machine learning models for occupancy prediction in residential buildings using connected thermostat data', *Building and Environment*, vol. 160, p. 106177, Aug. 2019, doi: 10.1016/j.buildenv.2019.106177.
- [44] Y. Liu, Z. Pang, M. Karlsson, and S. Gong, 'Anomaly detection based on machine learning in IoT-based vertical plant wall for indoor climate control', *Building and Environment*, vol. 183, p. 107212, Oct. 2020, doi: 10.1016/j.buildenv.2020.107212.
- [45] Z. Chen, M. K. Masood, and Y. C. Soh, 'A fusion framework for occupancy estimation in office buildings based on environmental sensor data', *Energy and Buildings*, vol. 133, pp. 790–798, Dec. 2016, doi: 10.1016/j.enbuild.2016.10.030.
- [46] L. M. Candanedo and V. Feldheim, 'Accurate occupancy detection of an office room from light, temperature, humidity and CO 2 measurements using statistical learning models', *Energy and Buildings*, vol. 112, pp. 28–39, Jan. 2016, doi: 10.1016/j.enbuild.2015.11.071.
- [47] S. Hochreiter and J. Schmidhuber, 'Long Short-Term Memory', *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.
- [48] T. Fernando, H. Gammulle, S. Denman, S. Sridharan, and C. Fookes, 'Deep Learning for Medical Anomaly Detection – A Survey', *ACM Comput. Surv.*, vol. 54, no. 7, pp. 1–37, Sep. 2022, doi: 10.1145/3464423.
- [49] D. Skrobek *et al.*, 'Prediction of Sorption Processes Using the Deep Learning Methods (Long Short-Term Memory)', *Energies*, vol. 13, no. 24, p. 6601, Dec. 2020, doi: 10.3390/en13246601.
- [50] H. B. Rijal, P. Tuohy, M. A. Humphreys, J. F. Nicol, A. Samuel, and J. Clarke, 'Using results from field surveys to predict the effect of open windows on thermal comfort and energy use in buildings', *Energy and Buildings*, vol. 39, no. 7, pp. 823–836, Jul. 2007, doi: 10.1016/j.enbuild.2007.02.003.
- [51] F. Haldi and D. Robinson, 'On the behaviour and adaptation of office occupants', *Building and Environment*, vol. 43, no. 12, pp. 2163–2177, Dec. 2008, doi: 10.1016/j.buildenv.2008.01.003.
- [52] C. M. Bishop, *Pattern recognition and machine learning*. New York: Springer, 2006.
- [53] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*, 2. ed. New York Weinheim: Wiley, 2001.
- [54] T. Li, S. Zhu, and M. Ogihara, 'Using discriminant analysis for multi-class classification: an experimental investigation', *Knowl Inf Syst*, vol. 10, no. 4, pp. 453–472, Oct. 2006, doi: 10.1007/s10115-006-0013-y.
- [55] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. New York, NY: Springer New York, 2009. doi: 10.1007/978-0-387-84858-7.
- [56] L. Breiman, 'Random Forest', *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.
- [57] Y. Bengioy, P. Simardy, and P. Frasconiz, 'Learning Long-Term Dependencies with Gradient Descent is Difficult', p. 36.
- [58] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, 'Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling'. arXiv, Dec. 11, 2014. Accessed: Jul. 28, 2022. [Online]. Available: <http://arxiv.org/abs/1412.3555>
- [59] R. Jozefowicz, W. Zaremba, and I. Sutskever, 'An Empirical Exploration of Recurrent Network Architectures', p. 9.
- [60] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, 'LSTM: A Search Space Odyssey', *IEEE Trans. Neural Netw. Learning Syst.*, vol. 28, no. 10, pp. 2222–2232, Oct. 2017, doi: 10.1109/TNNLS.2016.2582924.
- [61] R. Razavi, A. Gharipour, M. Fleury, and I. J. Akpan, 'Occupancy detection of residential buildings using smart meter data: A large-scale study', *Energy and Buildings*, vol. 183, pp. 195–208, Jan. 2019, doi: 10.1016/j.enbuild.2018.11.025.
- [62] C. Rb, C. William S., M. Jean E., and T. Irma J., 'STL: A seasonal-trend decomposition procedure based on loess (with discussion)', Cleveland 1990, 1990.