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

Enguerrand de Rautlin de la Roy*[a], Thomas Recht[a], Akka Zemmari[b], Pierre Bourreau[c], Laurent Mora[a]

[a]*Univ. Bordeaux, CNRS, Bordeaux INP, I2M, UMR 5295, F-33400, Talence, France*

[b]*Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, F-33400, Talence, France*

[c]*Nobatek/INEF4, 9 rue Jean-Paul Alaux, F-33000 Bordeaux, France*

Abstract

The increasing use of monitoring systems such as Building Management System (BMS) or connected devices bring the opportunity to better evaluate, model or control both occupants' comfort and energy consumed by an operated building thanks to the consequent amount of data provided (e.g., air temperature, CO_2 concentration, electricity consumption). Occupants' behavior and more specifically window-openings affect both occupants' thermal comfort and building energy consumption and are therefore key components to consider. This paper presents a comparison of machine learning models applied on window-openings detection during the heating season such as: Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), Random Forest Classifier (RFC) and two Recurrent Neural Network (RNN), namely, Long Short Term Memory (LSTM) and Gated Recurrent Unit (GRU). While some 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

study proposes a detailed comparison of the main methods and focuses on the impact of feature engineering process considering four different data transformations based on field expertise and more than 800 different combinations built on six indoor and outdoor measurements. Results show that some of the proposed transformations and combinations positively impact all models performances. The best performances on window-openings detection are attained by using indoor temperature and CO_2 concentration on RNN models with an average F1-score of 0.78 while LDA, SVM and RFC models tend to provide satisfying but lower performance around 0.70-72. In addition, by using the right transformation, significant results can be achieved by detecting up to 84-88 % of window-opening times with the sole use of indoor air temperature measurements.

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

---

1. Introduction

The building sector accounts for approximately 40% of the final energy use in Europe [1]. In addition, life cycle analysis of buildings tends to show that most of the building life cycle energy consumption depends on its operation (80 to 90%) [2]. Thus, evaluating and defining operational loads of building are keys elements in order to reduce or comprehend buildings energy consumptions. The most common use of Building Management System (BMS) and connected devices in the building sector has led to a democratization of edifices called smart buildings. A *smart building* can be seen as an association of multiple systems, software and sensors [3] that aims to meet two main objectives: to reduce both operational and environmental costs by managing and optimizing the energy use [4, 5, 6] and to improve the comfort of the occupants [7, 8]. Hence, smart buildings generate a consequent amount of data from measurements (air temperature, CO_2 concentration, energy consumption, *etc.*) that can be studied in order to evaluate, model, correct or optimize specific operational loads during the building life cycle [9, 10, 11]. Among these loads, occupants' behavior can have a strong influence on the operational energy consumption of buildings as well as the thermal comfort [12, 13]. A common action is often identified:

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

Nevertheless, window-openings impacts on other measurements (such as air temperature, CO_2 concentration, *etc.*) can be observed through specific patterns that tend to deviate from other observations. These patterns can be recognized and classified by using machine learning techniques in order to determine the corresponding window-status (*open* or *close*).

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

The main contributions of this article are therefore:

- The comparison on window-status detections for several machine learning models classifier such as Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), Random Forest Classifier (RFC) and Recurrent Neural Network (RNN) such as Long Short Term Memory (LSTM) and Gated Recurrent Unit (GRU).

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

2. Related work

2.1. Occupants' behavior detection or prediction

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

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

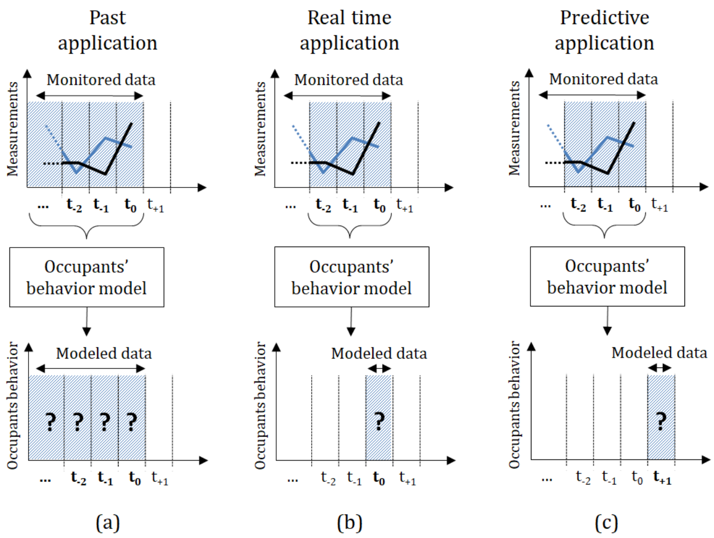

Figure 1 Occupants behavior detection based on monitored data: (a) past detection, (b) present detection and (c) future prediction

2.2. Window-opening behavior models

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

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

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

2.3. Feature selection and transformation for window opening models

Features used for window opening models can rather be divided into two groups [15] environmental and non-environmental. Environmental features are based on indoor or outdoor environment measurements such as air temperature, CO_2 concentration, wind speed, solar radiation, noise, *etc*. while non-environmental features are based on buildings, occupants or time characteristics such as room type, gender, age or time of the day. As analyzed in Xilei Dai review [15] and regardless of the models, the most used features come from environment measurements such as: outdoor and indoor air temperature, humidity, indoor CO_2 concentration and wind speed. Regarding logistic regression models, most of the studies tend to show that indoor and outdoor air temperature have the most impact [39, 50, 51]. However, concerning artificial neural network models, a detailed study highlighting, among others, measurements impact for window openings application, such as Romana Markovic [40], shows different results. Her study provides a relevant example based on an ANN model by analyzing neurons learned weights from more than twenty input features that highlight the importance of indoor environmental data and more specifically, CO_2 concentration. Regarding these results, a broad approach including different ambient indoor and outdoor measurements is privileged for this study. In addition, another main point should be specified regarding feature selection. Most of the features used in previous studies are measurements that are neither transformed (e.g., derivation, smoothing) nor combined (e.g., differences) despite positive impacts that might be obtained on models performances [22]. Hence, this study offers to test and compare various measurements transformations that are further presented in section 4.2.1 on several models in order to evaluate their contribution and relevance.

3. Experiment methodology

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

3.1. Evaluation process

3.1.1. F1-score

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

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

However, although this evaluation metric may provide a global overview on every models' performances, it alone might not be sufficient to choose which model is better especially in case of similar or identical results. As illustrated in Figure 2, a window opening state is defined, in this work, as one or successive open states (full-cells) bounded by one or successive close states (empty-cells). Both models evaluations (2.i) and (2.ii) provides the same F1-score values whereas both provide different results with only half of the openings perfectly detected for (2.i) and all openings detected but underestimated for (2.ii). Other metrics might be useful in order to compare the results to other studies, to deepen models outputs and guide or adapt the choice of models or measurements according to specific needs, with a focus on number of openings detected or on the total opening time for instance. Thus, more domain oriented metrics focusing on window opening classification and evaluation based on true and false opening detection are introduced and discussed in this paper.

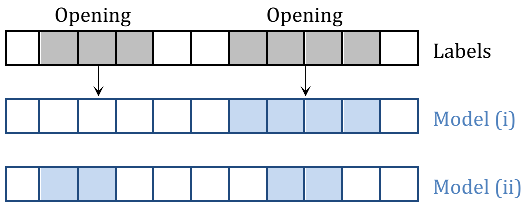

Figure 2 Window-openings detection for two different models

3.1.2. True and false openings

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

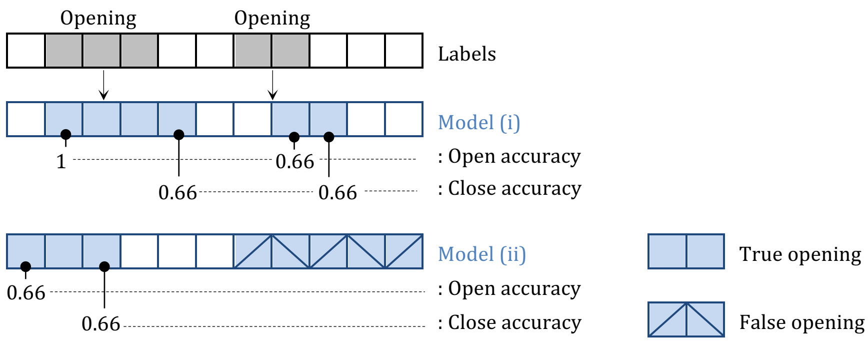

Figure 3 True and false opening examples

3.1.2.1. Total true and false openings number

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

3.1.2.2. Total true and false openings time

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

3.1.2.3. Average true opening accuracy score

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

3.2. Models

3.2.1. SVMs

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

3.2.2. LDA

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

3.2.3. Random Forest Classifier

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

3.2.4. RNNs

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

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

Table 1 List and range of Models' tuning parameters

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

4. Data description and data preprocessing

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

4.1. Data description

4.1.1. Data collection and preparation

The raw data used for this study is made of 1 minute time step measurements collected over two years (from July 2019 to February 2021) in a northwest bedroom of an apartment located in Bordeaux city-center (France). The raw data include indoor climate measurements (such as temperature, relative humidity and CO_2 concentration), outdoor climate measurements (such as temperature and relative humidity) and window-status measurements. Sensors positions for all studied measurements are illustrated in Figure 4.

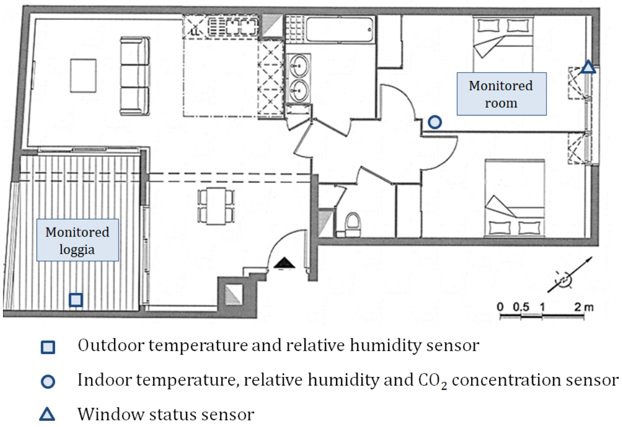

Figure 4 Studied sensors position

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

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_2 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	-	-

4.1.2. Data analysis

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

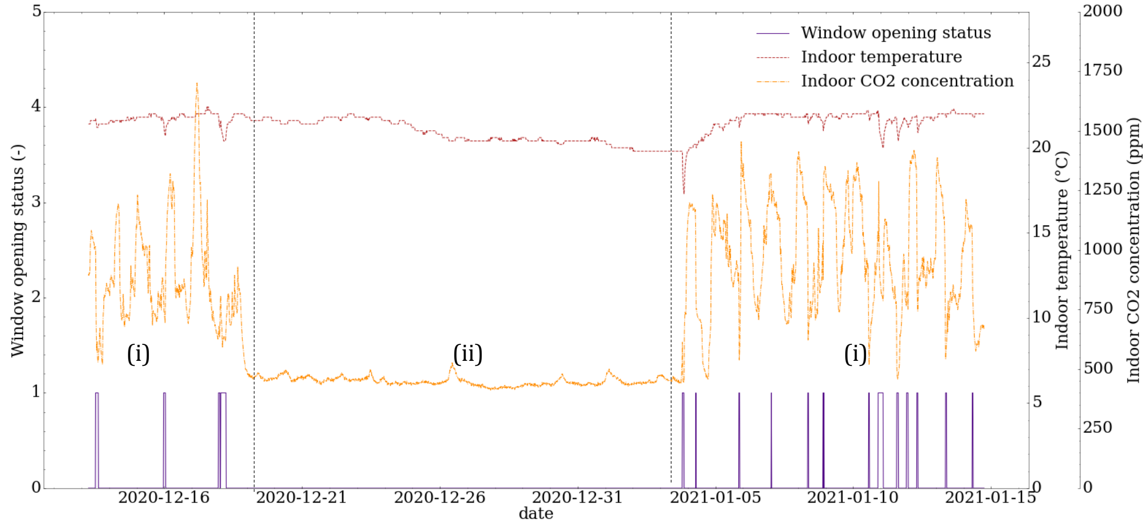

Figure 5 Test dataset of indoor temperature, indoor CO2 and window-status

The unoccupied period is characterized by a slow drift of approximately 2°C of indoor air temperature and a low stagnating CO_2 concentration. These differences are shown in Table 3. The occupied period is characterized by higher average indoor temperature and indoor CO_2 concentration. While period (5.i) presents data characteristics relatively close to the training set characteristics, those shown in period (5.ii) tend to differ. Such differences might impact the evaluation phase but may also provide useful information on models behavior and sensibility to uncommon data characteristics.

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

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_2 concentration (ppm)	1700.7	440.7	932.9	236.9
(2) Indoor CO_2 concentration (ppm)	526.2	411.5	449.2	17.8

4.2. Data preprocessing

4.2.1. Feature engineering

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

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

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

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

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

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

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

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

$$Diff_{M_t} = M_t - EMA_{M_t} \qquad \text{Eq. (5)}$$

These transformations are applied on indoor measurements such as temperature, relative and absolute humidity and CO_2 concentration. They can be associated in order to provide different information from the data to the model. Hence, for every indoor measurement, 20 different associations are performed by combining the measurement without modifications and the transformations presented above. As shown in Table 4 by comparing both periods presented previously on the test set to the training set, derivation, *STL-Residue* and *EMA-Difference* transformations appear to provide information without large variation such as global or local seasonality by centering the data. The same effect can be observed on three weeks extracted from the training set with Figure 6 where $d_{temperature}$, $residue_{temperature}$ and $difference_{temperature}$ transformations remain centered contrary to the temperature measurement. These transformations are intended to push models to be less sensitive to measurements values or global variations and more on local dynamics. However some drawbacks might be observed: a lag effect can be noticed on *EMA* based transformations such as *EMA-Difference* whereas derivation transformations can provide data with a low variance that might be, depending on the model, delicate to exploit.

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

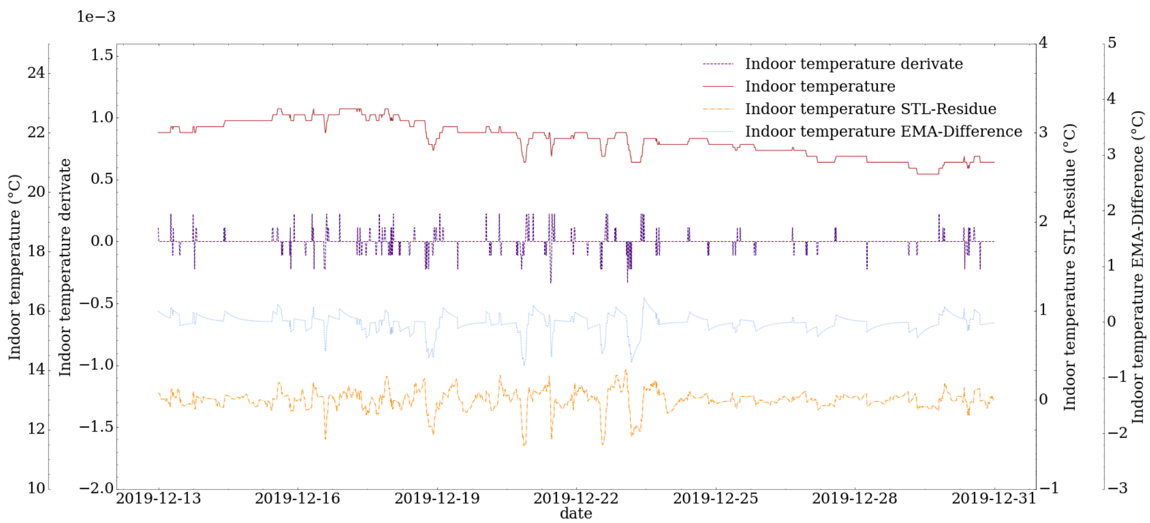

Figure 6 Applied data transformation and combinations representation on 3 training weeks

4.2.2. Data preparation and association

For this study, more than 800 different combinations of measurements and measurements transformations and associations were performed for every model. More specifically a base of 20 different associations, as presented in section 4.2.1, was processed for every indoor measurement (temperature, humidity and CO_2 concentration). Based on the previous combination results, 175 additional combinations were performed for dual indoor measurement combinations (temperature + humidity and temperature + CO_2). By following a similar process, around 100 different combinations for triple indoor measurements combination were created with around 300 more by adding outdoor measurements (temperature and humidity) to all previous combinations. Except for RFC models, features are normalized between 0 and 1 based on the training set characteristics.

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

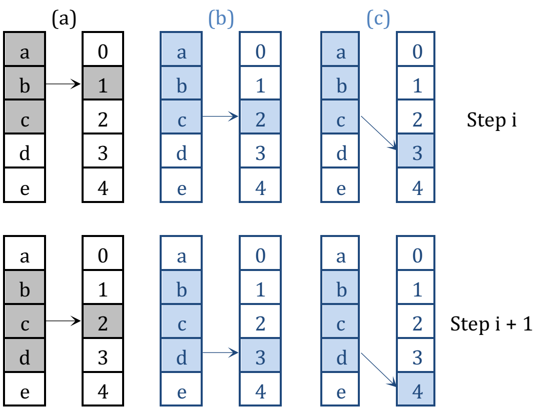

Figure 7 Training window for LSTM and GRU models for: (a) past window-openings, (b) real-time window-openings and (c) window-openings prediction

5. Results and discussion

Due to the specificities of LSTM, GRU and RFC models training, each of the 800 combinations has been performed ten times. Thus, only the model output with the best F1-score and accuracy results out of the ten was retained for evaluation. Since LDA and SVM models provide more stable outputs, each of the 800 combinations was only performed once. To not overload this study and since relative humidity systematically provides poorer results than absolute humidity, only absolute humidity based associations are presented. Hence, absolute humidity is further referred simply as humidity. All the following results are evaluated regarding their F1-score and alternative metrics presented in section 3.1 in order to discuss about measurements transformations, associations and combinations that might appear to be the most adequate to detect window openings. Models raw performances to recognize past window openings will be taken into consideration but are not the sole focus of this study. Thus, a comparison of the observed results between models and combination is preferred.

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

An overview of the performances of each model trained on the twenty bases transformations and associations (4.2.1) for every indoor measurement is presented in Table 5 Average F1-score and standard deviation of the 20 best results for each measurements combination and models. It appears that, regardless of the transformation or association used, indoor air humidity or CO_2 concentration sole base combination (referred as H_{in} and C_{in}, respectively) does not provide good results on window-status detection with an average F1-score that, at best, usually does not exceed 0.39. On the other hand, indoor air temperature sole base combinations (T_{in}) seem to provide better and exploitable results with a higher F1-score average for all models and especially for RNN based models with an average value close to 0.70.

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

F1-score: *mean ± standard deviation*	**GRU**	**LSTM**	**LDA**	**SVM**	**RFC**
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_2 (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**

However these results present a high dispersion that can be explained by looking at the F1-score results for the five base transformations without associations shown in Table 6. For air humidity measurements, it appears that *EMA-Difference* and *STL-Residue* based transformations increase models performances compared to other transformations or base measurement. Best results are observed with $difference_{humidity}$ for both LSTM and GRU models with an F1-score around 0.70. Regarding air temperature, the dispersion might be caused by the difficulty of all models to provide results when trained on untransformed temperature, $EMA_{temperature}$ or some other measurements transformations including them. These results might be explained by the differences between the training and testing set. Testing set measurements reach values and dynamics unseen during the training phase, allowing less contextual transformations such as the *derivate*, the *STL-Residue* or the *EMA-Difference* to perform better and increase the F1-score from 0.60 to 0.77 for LSTM and GRU models (representing an improvement of 25 %) and from around 0.00 to 0.6 for LDA, SVM and RFC models . Results including different measurements combination based on the same transformations and associations are presented in the following section.

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}	**0.77**	**0.75**	0.25	0.25	0.50
EMA-Difference .T_{in}	0.75	0.72	**0.65**	**0.62**	**0.60**
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}	**0.72**	**0.71**	0.13	0.17	**0.41**
STL-Residue.H_{in}	0.57	0.52	**0.41**	**0.40**	0.35

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

In order to have an overview of the best achievable performances of each model depending on the measurements combination used, the best twenty models are selected based on F1-score, for each performed combination. The average and standard deviation of those twenty best models outputs are presented in Table 7 with a total of 280 combinations out of the 800 originals for each model. Dual indoor measurements combinations (T_{in} + H_{in} and T_{in} + C_{in}) usually tend to provide higher opening window detection performances with a systematic increase of the maximum and average F1-score. For all models, indoor CO_2 concentration combined with indoor temperature seems to provide significate higher performances than humidity and temperature combinations. This performance enhancement is reflected by a consistent improvement on F1-scores averages for all models compared to temperature and humidity combinations. The combination of the three indoor measurements (T_{in} + H_{in} + C_{in}) seems to provide only slight to no improvement for all models on window opening detection compared to temperature and CO_2 combination. Based on F1-scores it appears that enough information are provided during training with this dual combination for LSTM, GRU, LDA and RFC models contrary to the SVM. Hence, LSTM and GRU models output on opening window detection seem to be capped around an average F1-score of 0.76 - 0.78 whereas LDA, SVM and RFC models appear to be lower limited around a 0.72 - 0.73 average score.

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

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

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

F1-score: *average ± standard deviation*	**GRU**	**LSTM**	**LDA**	**SVM**	**RFC**
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_2 (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}	**0.78 ± 0.01**	**0.76 ± 0.01**	**0.72 ± 0.01**	**0.70 ± 0.01**	**0.73 ± 0.02**
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}	**0.78 ± 0.01**	**0.76 ± 0.01**	0.71 ± 0.01	**0.72 ± 0.01**	**0.73 ± 0.01**
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	**0.73 ± 0.01**	0.69 ± 0.02	0.65 ± 0.02

To conclude, even if the combination of the three indoor measurements (T_{in} + H_{in} + C_{in}) seems to provide the best results on opening detection regardless of the model, two indoor measurements such as 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 very fluctuating with variations that are not only related to windows openings (occupancy, occupant position in the room, natural air movement) the indoor CO_2 concentration measurement seems to be preferable to indoor humidity. Furthermore, depending on the transformation used, the sole indoor temperature measurement proves to be consistent enough to provide opening window detection results on par with dual or triune combinations. Due to the observed tendency of outdoor measurements to decrease opening detection performance for the majority of the models, this study will further be focused on indoor measurements.

5.3. Measurements selection and combination impact based on additional metrics

The additional evaluations metrics introduced in section 0 are used in order to provide more in-depth explanations on the differences observed and described previously. The number of true and false openings, the total time (in hour) of true and false opening and the opening and closing score are recorded as a boxplot repartition in Figure 8. This figure is constructed by using the same best twenty models output, based on F1-scores, for each performed combination as in Table 7. However, a specific focus on one to three indoor measurements combinations is made with T_{in}, T_{in} + C_{in}, T_{in} + H_{in} and T_{in} + H_{in} + C_{in} that represent a total of 80 combinations out of the 470 originals for each model. Regarding the measurements combination, Figure 8 shows that for all models and for all additional metrics, indoor temperature and CO_2 combinations appear to perform slightly better than indoor temperature and humidity combinations. The difference seems to be mostly due to the fact that CO_2 based combinations tend to have a higher capacity to get better maximum results for true openings detections (8.a), opening and closing score (8.c and 8.c'). This observation might be explained by the propensity of the CO_2 concentration to fluctuate on higher levels than the humidity and thus, with an adequate transformation, to detect or to better define a few more openings. Furthermore, for all models, aside of GRU and LSTM, the use of a combination of minimum two indoor measurements provide a clear improvement on the results compared to the sole use of the indoor temperature even if the higher score tend to be close to all other combinations.

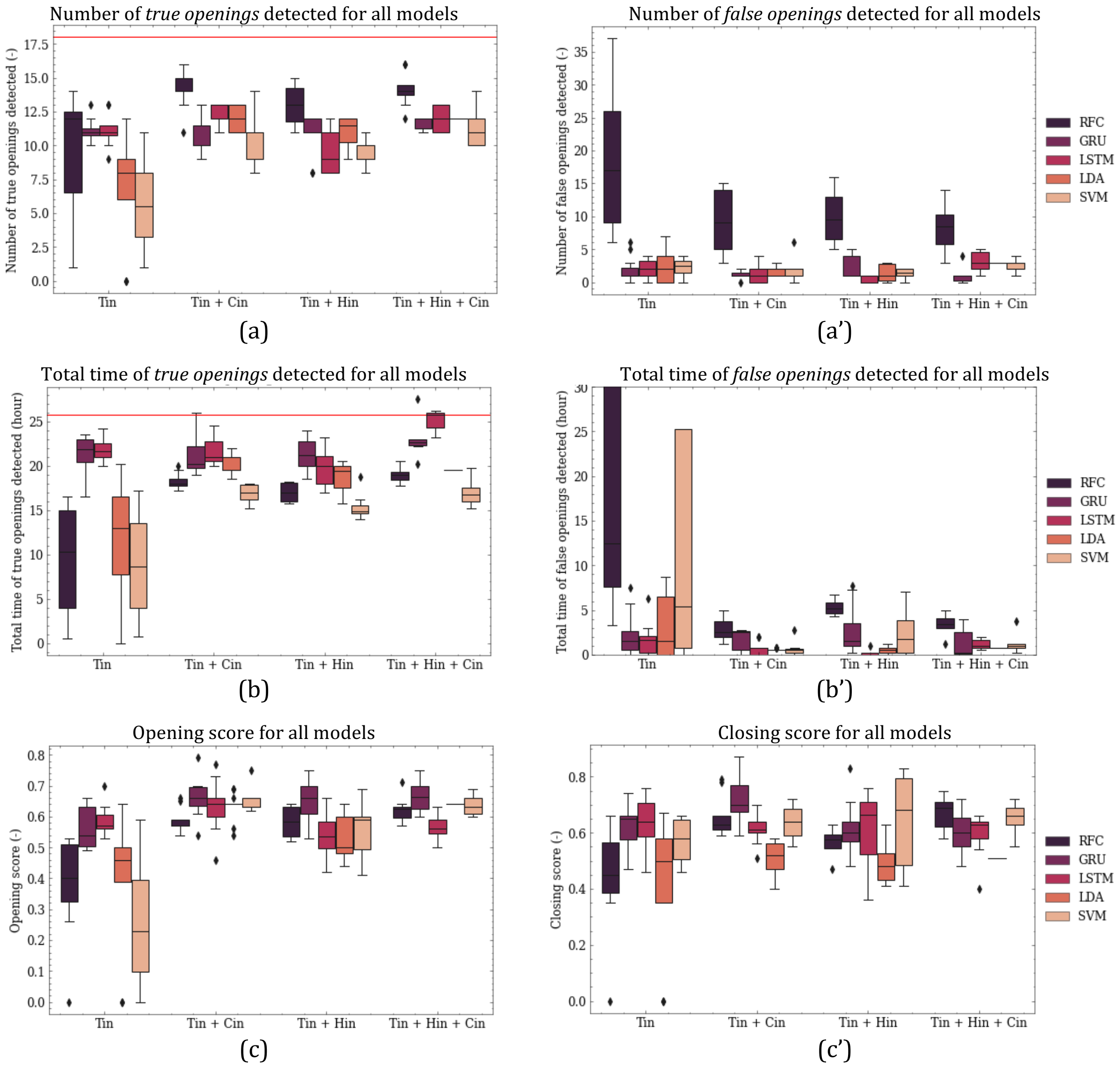

Figure 8 Additional evaluation metrics for every model and measurement combination

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.

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

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

In order to have an overview of the best performance of each model depending on the indoor measurements transformations and associations used, the best twenty models output, based on F1-score, 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 twenty best models outputs is presented in Figure 9 with a total of 60 combinations out of the 450 originals for each model. All measurements (referred as M) transformations and associations that are not part of the top twenty are not represented on this figure. Similarly, those under or low represented are displayed as a box with a small repartition or a small horizontal line.

Regarding humidity transformations and associations the three most recurrent and efficient for all models appear to be, by far, solely composed of $d_{humidity}$ (9.c.13), $difference_{humidity}$ (9.c.15) or both (9.c.14). Contrary to $difference_{humidity}$ that is a measurement transformation which is sufficient to detect window-opening as detailed in section 5.2, $d_{humidity}$ seem to perform better combined with other measurements. For temperature transformations and associations a distinction has to be made between models. The sole $d_{temperature}$ (9.a.13) seem to be consistent for LSTM and GRU models whereas it has to be combined with other transformations for LDA, SVM and RFC models (9.a.3, a.7, a.8, a.9 and a.14). For these last models, a large amount of temperature transformation association seems to be preferred in order to get good results contrary to RNN models that appear to perform well with just one or no transformation associations such as $d_{temperature}$, $difference_{temperature}$ or $residue_{temperature}$ (9.a.13, a.15, a.18). Contrary to the previous observations, a consensus doesn't seem to appear for CO_2 transformations associations. Associations based on d_{CO2} or a smoothed EMA_{CO2} seems to be a bit more present and thus effectives (9.b.2, b.5, b.13 and b.16).

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

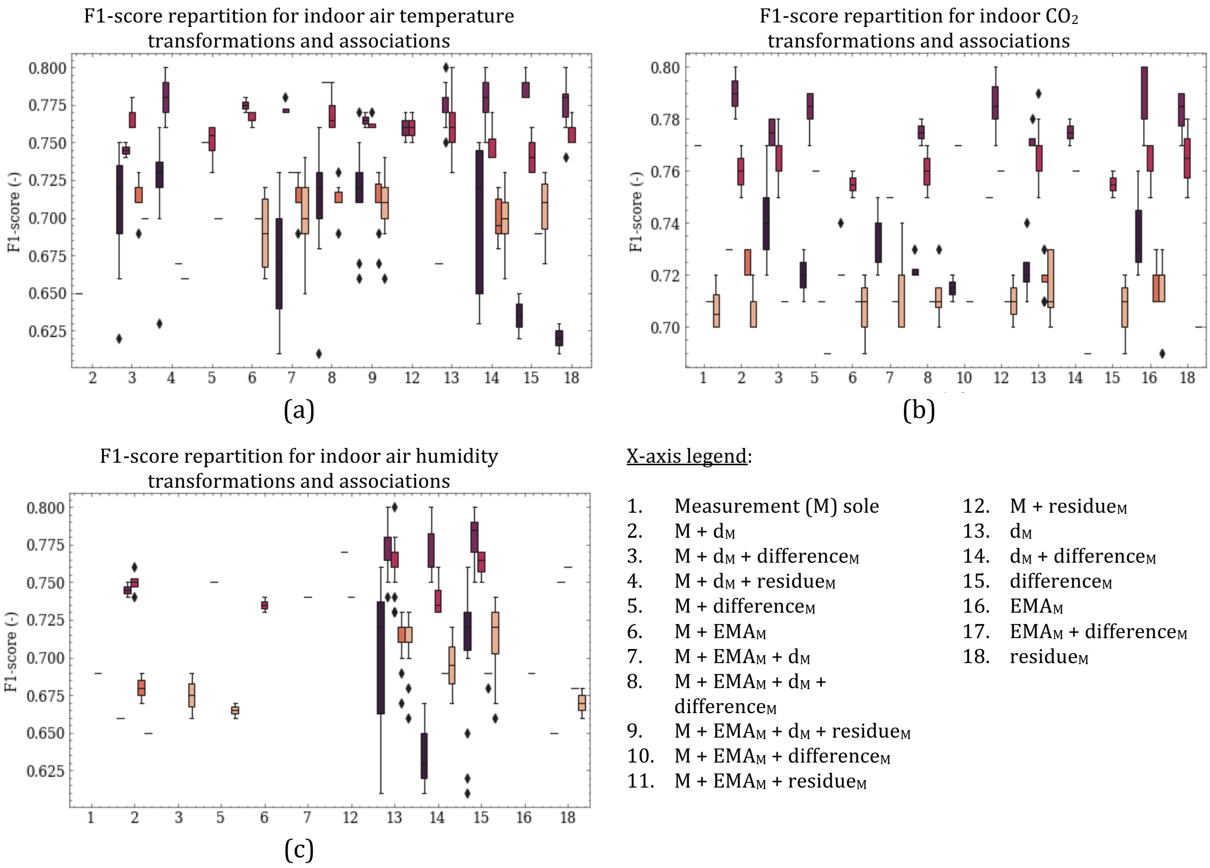

Figure 9 F1-score measurement transformation and associations for all tested combination best twenty F1-score

5.5. Discussion and future work

Measurement combination and transformation were performed on various machine learning models in order to assess their efficiency and relevance on past window-status detection. However, although performed on models that are not yet widely used on this domain, several limitations remains. Most transformations applied on measurements (*STL-Residue*, *derivate* or *EMA-Difference*) are built to reflect an impact created by a window opening between two environments with different characteristics (e.g., air temperature, CO_2 concentration). Thus, despite showing great results by improving and stabilizing models performances, they might not be suitable in other climates or seasons and need to be carefully evaluated beforehand. Therefore, a similar process will be followed on other seasons in order to highlight appropriate measurements combinations, transformations and associations.

Regarding measurements selection, LSTM and GRU models achieve satisfying results with the sole use of indoor temperature measurements although the addition of indoor CO_2 concentration appears to stabilize and slightly improve their results. For SVM, LDA and RFC models, the use of minimum both indoor temperature and CO_2 concentration tend to be recommended even if a small improvement in results can be observed by adding indoor humidity measurement.

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

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

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

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

F1-score: *average ± standard deviation*		**GRU**	**LSTM**
T_{in}	past	0.73 ± 0.07	0.68 ± 0.18
	real time	0.63 ± 0.17	0.59 ± 0.17
T_{in} + T_{out}	past	0.70 ± 0.03	0.70 ± 0.03
	real time	0.59 ± 0.10	0.56 ± 0.05
T_{in} + H_{in}	past	0.76 ± 0.01	0.75 ± 0.02
	real time	0.71 ± 0.02	0.67 ± 0.03
T_{in} + C_{in}	past	**0.78 ± 0.01**	**0.76 ± 0.01**
	real time	**0.76 ± 0.01**	**0.72 ± 0.02**
T_{in} + H_{in} + C_{in}	past	**0.78 ± 0.01**	**0.76 ± 0.01**
	real time	**0.77 ± 0.01**	**0.73 ± 0.02**

6. Conclusion

This study presents a comparison of the performance of Gated Recurrent Unit (GRU), Long Short Term Memory (LSTM), Linear Discriminant Analysis (LDA), Support Vector Machine (SVM) and Random Forest Classifier (RFC) models in detecting window openings depending on several indoor and outdoor measurements combinations, transformations and associations in the field of building energy during heating season. The results showed that not only the choice of input data measurement was essential to obtain satisfactory results but also that it was neither always optimum nor required to add more information to the input of the models (e.g., outdoor measurements) and that a preliminary selection might be necessary. Hence, if required, the sole use of a temperature sensor with adapted transformation (e.g., *temperature derivate*, *temperature STL-Residue* or *temperature EMA-Difference*) might be sufficient to provide satisfying results for window-openings detection. Adding other indoor measurements appears recommended to obtain slightly more precise results for LSTM and GRU models and necessary for other models. In this case, the combination of indoor temperature and CO_2 concentration measurement seems to be the one to be privileged for all models.

This work also showed that a simple transformation of the data beforehand (e.g., *derivate*) or more complex ones introduced in this paper (*STL-Residue* or *EMA-Difference*) could have a significant positive impact on the quality of the window-openings detections by turning unusable results (e.g., temperature sole or with other combinations) to satisfactory results. Depending on the model used, specific association of measurement transformation might be appropriate.

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

Acknowledgements

This research was conducted within the framework of the Intellib@t project within the joint research team GP2E (as in French, Guarantee of energy and environmental performances), between the I2M laboratory and Nobatek/INEF4, with funding from the Regional Committee of Nouvelle Aquitaine (CNRA) 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]]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 retrofitting 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)', Cleveland1990, 1990.