

# Machine learning and engineering feature approaches to detect events perturbing the indoor microclimate in Ringebu and Heddal stave churches (Norway)

Pietro Miglioranza, Andrea Scanu, Giuseppe Simionato and  
Nicholas Sinigaglia

*Department of Physics and Astronomy, University of Padua, Padua, Italy, and  
America Califano*

*Department of Physics and Astronomy, University of Padua, Padua, Italy and  
Department of Mechanical and Industrial Engineering,  
Norwegian University of Science and Technology, Trondheim, Norway*

## Abstract

**Purpose** – Climate-induced damage is a pressing problem for the preservation of cultural properties. Their physical deterioration is often the cumulative effect of different environmental hazards of variable intensity. Among these, fluctuations of temperature and relative humidity may cause nonrecoverable physical changes in building envelopes and artifacts made of hygroscopic materials, such as wood. Microclimatic fluctuations may be caused by several factors, including the presence of many visitors within the historical building. Within this framework, the current work is focused on detecting events taking place in two Norwegian stave churches, by identifying the fluctuations in temperature and relative humidity caused by the presence of people attending the public events.

**Design/methodology/approach** – The identification of such fluctuations and, so, of the presence of people within the churches has been carried out through three different methods. The first is an unsupervised clustering algorithm here termed “density peak,” the second is a supervised deep learning model based on a standard convolutional neural network (CNN) and the third is a novel *ad hoc* engineering feature approach “unexpected mixing ratio (UMR) peak.”

**Findings** – While the first two methods may have some instabilities (in terms of precision, recall and normal mutual information [NMI]), the last one shows a promising performance in the detection of microclimatic fluctuations induced by the presence of visitors.

**Originality/value** – The novelty of this work stands in using both well-established and in-house *ad hoc* machine learning algorithms in the field of heritage science, proving that these smart approaches could be of extreme usefulness and could lead to quick data analyses, if used properly.

**Keywords** Clustering, Wood, Microclimate, Heritage buildings, Machine learning,  
Convolutional neural networks

**Paper type** Research paper

© Pietro Miglioranza, Andrea Scanu, Giuseppe Simionato, Nicholas Sinigaglia and America Califano. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licenses/by/4.0/legalcode>

This work is part of the Symbol Research Project n.274749 and of the Spara Och Bevara Project n. 50049-1 and has been funded by the Norwegian Research Council and the Swedish Energy Agency respectively. This work has been possible thanks a collaboration between the University of Padua (Italy) and the Norwegian University of Science and Technology – NTNU (Norway). The ownership of microclimatic data collected and used in this work is of the Principal Investigator of the Symbol Project. Special thanks to Prof. Marco Baiesi and Prof. Chiara Bertolin for their precious guidance and support.



## 1. Introduction

During the medieval period, over 1,000 wooden stave churches were thought to have been built in Norway; however, currently only 28 of these churches survive. The indoor microclimate of these cultural sites has been directly governed by the external climate for centuries until the 19–20th century, when heating systems were installed in several sites to improve the indoor thermal comfort for the congregation. The main consequence of this climate control and of the presence of a great amount of people inside the stave churches is the resulting variation of the microclimate, which directly affects the churches' preservation as they are mainly made of wood (UNI EN 15757:2010 Standard - Conservation of Cultural Property - Specifications for Temperature and Relative Humidity to Limit Climate-Induced Mechanical Damage in Organic Hygroscopic Materials, 2010; Olstad *et al.*, 2020). The problem of detecting the events (i.e. occasions attended by people, such as celebrations, concerts, guided tour, etc.) taking place in the churches can be traced back to a time series classification (TSC) problem, which can be, nowadays, analyzed with both supervised and unsupervised machine learning techniques.

Another novel way to tackle this problem comes from the great success of deep learning-based models in supervised learning for TSC (Tang *et al.*, 2020), especially convolutional neural networks (CNNs) used for image recognition (Wu and Chen, 2016) and natural language processing (Young *et al.*, 2018).

In this work, the impact of visitors over the internal microclimate of Ringebu and Heddal stave churches (Section 2) located in Norway has been studied with the implementation of three different algorithms (Section 3); their performances have been compared (Sections 4) and main conclusions, in the view of future perspectives of optimal indoor climate management, have been deduced (Section 5).

In detail, the first applied method is a clustering algorithm (Mehta *et al.*, 2019) designed by Laio and Rodriguez and named “density peak clustering (DPC)” (Rodriguez and Laio, 2014). Its performances are evaluated with respect to the metadata (i.e. information about the use of the two churches) recorded by the churches' staff, which are considered ground truth.

The second applied method is a CNN inspired by the success of this kind of architecture in time series analyses (Ismail Fawaz *et al.*, 2020). Its performances are tested over the union of the Ringebu and Heddal datasets, thanks to a five-fold cross-validation (Refaeilzadeh *et al.*, 2009).

The third and last applied method is a novel engineering feature approach, named “unexpected mixing ratio (UMR) peak”, based on the assumption that days with events are characterized by at least one change in the internal climate, which is not dependent on the external climate. This model consists in a heavy data manipulation and a one parameter supervised linear classifier trained over Ringebu dataset and tested over Heddal dataset.

## 2. Materials

To carry out the proposed study, datasets comprising temperature ( $T$ ) and relative humidity (RH) time series measured in Ringebu and Heddal stave churches have been used. These time series were acquired from March 2019 to March 2020 within the framework of the Symbol (Sustainable Management of Heritage Building in a long term perspective) Project no. 274,749 undertaken by the Norwegian University of Science and Technology (NTNU), the Norwegian Institute for Cultural Heritage Research (NIKU), the Polish Academy of Sciences (PAS) and the Getty Conservation Institute (GCI); the measurements were collected by three dataloggers for each church: two of them located inside and measuring the indoor temperature ( $T_{in}$ ) and relative humidity ( $RH_{in}$ ) every 5 min, while the third one placed outside and measuring the outdoor temperature ( $T_{out}$ ) and relative humidity ( $RH_{out}$ ) every 15 min. In order to have equally long datasets and consistency between indoor and outdoor data, the outdoor data have been – in this work – resampled with a frequency of 5 min. Starting from the temperature and relative humidity time series, both the internal ( $MR_{in1}$ ,  $MR_{in2}$ ) and external ( $MR_{out}$ ) mixing ratio (MR) time series have been obtained, following the relation reported in

Camuffo (2018, 2019). The MR is a quantity that measures the number of water grams of moisture per kilogram of dry air, and it has a huge importance in the current application as it may be seen as a tracer for the presence of people inside the stave churches. As matter of fact, the MR provides a sort of threshold (as shown in the following sections), which quantifies and discriminates the presence of people inside a building. Alongside the microclimatic dataset, metadata with information about the scheduled events (namely, date and time of events, type of events and number of attendees) occurred in the two churches, and the usage of the heating systems was made available. To discriminate among days with people presence, two labels have been introduced and appended to the datasets: 0 for days without events and 1 for days with at least one event. It has to be strongly underlined that metadata could be, in some case, inaccurate as they are mainly recorded by hand by the churches' managers. This aspect has been further deepened in the following sections.

In both churches, the heating systems are used only during the cold season (mainly from September to April/May), but their managements are radically different. In Ringebu, the heating is switched on only on days with scheduled events. While in Heddal the heating is kept permanently on when the church is closed to visitors, i.e. in the cold season, time to time, in Heddal, because of sporadic events, the temperature is further risen for the comfort of churchgoers. Using data and metadata, three different approaches have been used to pursue the desired goal of tracing microclimate perturbations caused by the presence of people, and they are described in the following section.

### 3. Methods

In this section, the three used algorithms are described separately: the density peak clustering (DPC), the CNN and the novel proposed UMR approach.

#### 3.1 Density peak clustering (DPC)

There exists a large number of clustering algorithms, each one with different weak points and advantages. The first method has been based on the application of an unsupervised clustering algorithm. Assuming that in a given dataset there exists a certain probability distribution from which the data are taken, then the use of a clustering algorithm may be a good option. This concept is the main idea behind clustering (d'Errico *et al.*, 2021), and it allows to face the problem supposing that, in a certain space where the measurements may be represented, it is possible to recognize a pattern in the distribution. However, the main challenge of clusterization remains the choice of the parts in which it is meaningful to fragment the dataset. In this work, among the existing clustering algorithms (Lam and Wunsch, 2014), the DPC method (MacKay, 2005) has been chosen because it is simple to implement and has low computational costs. Moreover, given two parameters, it is capable to automatically find the number and centers of the clusters.

Considering  $T$ , RH and MR time series of each day as a data sample, then the data space dimension is extremely high. For this reason, in order to perform the clustering, few relevant features for each day need to be extracted. Due to the interest in the changes of the variables that describe the churches' indoor environment, the time series are mainly manipulated by computing their standard deviation (std) or their first discrete derivative, as follows:

- (1) *Derivative*: a discrete derivative has been computed, due to the interest in detecting the changes of the variables that are representative of the environment inside the churches;
- (2) *Standard deviation (std)*: the time series' *std* has been computed because it may be supposed that a day with an event is more affected by the variation of humidity, temperature and MR.

Starting from this idea, four features have been extracted for each day, reframing the problem to a clustering problem in a four-dimensional space. The features are as follows:

- (1) *Std* of  $MR_{in}$  time series;
- (2) *Std* of  $RH_{in}$  time series first derivative;
- (3) *Std* of  $(MR_{in}-MR_{out})$  time series first derivative and
- (4) *Std* of  $T_{in}$  time series first derivative.

As there is no clear evidence that the  $T$ ,  $RH$  and  $MR$  distributions can be adequately represented by a normal distribution, introducing the *std* is, mainly, a mean field approximation. It is a massive simplification, but it allows tackling the problem in object in a simple and quick way.

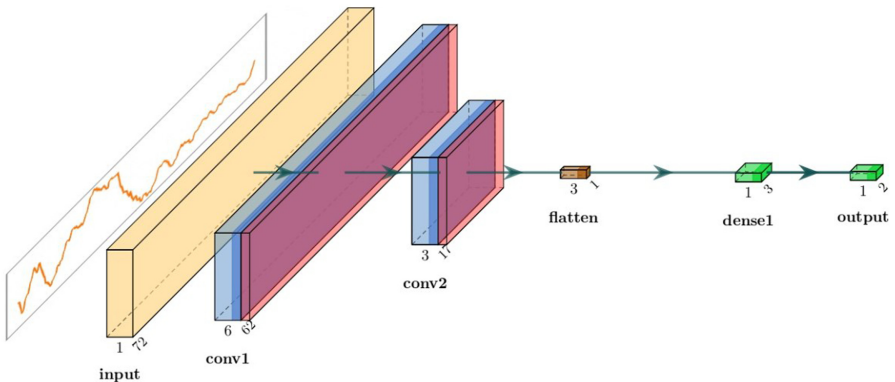
In addition, a normalization procedure has been operated in order to have the mean of each feature equal to 1. The reason behind this operation is that each feature should have the same weight in the algorithm. To evaluate if the algorithm looks for the correct clusters, metadata are fundamental; moreover precision, recall and the normalized mutual information (NMI) have been evaluated to have a quantitative measurement of the fragmentation goodness (MacKay, 2005). The main results are shown and discussed in Section 4.1.

### 3.2 Convolutional neural network (CNN)

Current approaches to TSC are based on CNNs because they are powerful tools able to find the underlying structure in datasets (Zhao et al., 2017). Starting from an input time series, a convolutional layer consists of sliding one-dimensional filters over the given sample, extracting non-linear features that are time-invariant and useful to classify the data. Thanks to the combination of multiple layers, the network can extract hierarchical features that could be useful for classification. The main reason is that the CNN can discover and extract the suitable internal structure to automatically generate deep features of the raw data, useful for classification.

In the current case, a simple one-dimensional CNN for TSC has been used. The input data consist in the  $MR_{in}$  time series, as the  $MR$ , for how it is defined, brings the maximum amount of useful information concerning the impact of people inside the churches. The architecture of this model is very simple, with few trainable parameters, while a more complex one will only lead to overfitting because of the reduced size of the dataset.

In Figure 1, a schematization of the used CNN is reported. The architecture is made by the input layer (yellow), the convolutional modules (light blue) followed by the rectified linear unit



**Figure 1.** The CNN architecture made by the input layer (yellow), the convolutional modules (light blue), the activation function (blue) and the maxpooling layer (red). In addition, the flatten (orange) and the dense (green) layers are reported as well

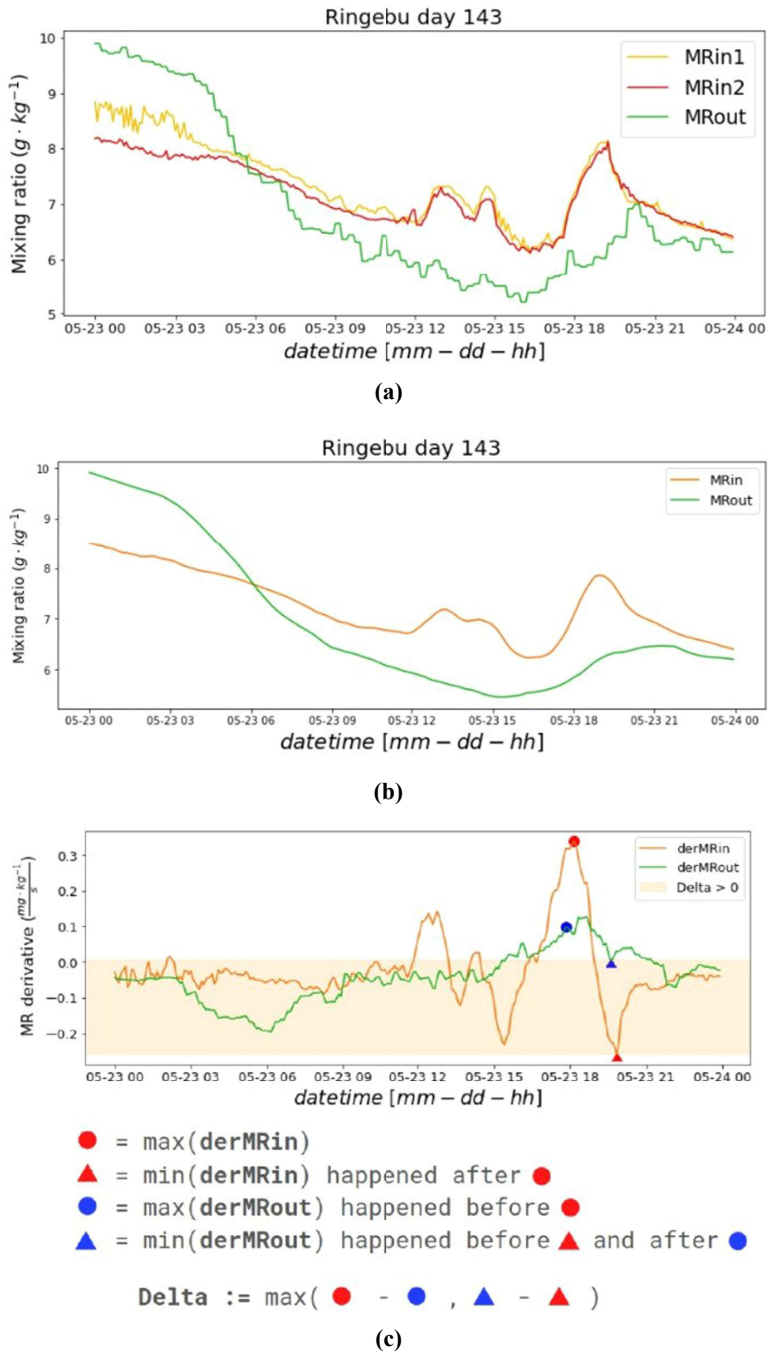
(RELU) activation function (blue), defined as the maximum value between 0 and  $x(f(x) = \max(0, x))$  and a max-pooling layer (red). The flatten and the dense layers are reported in orange and green, respectively; they provide the output value of the CNN: this last layer is made by two neurons as two is the total number of classes that compose this problem. This architecture has three convolutional layers with a few numbers of filters for each of them. The first convolution uses a bigger filter size than the one of the second convolution. The main idea is to previously look at the general trend of the series and then to focus on more detailed possible patterns. Therefore, before training this architecture, the data of the two churches are manipulated in three main steps: (1) the average between data collected by the two indoor sensors is computed to obtain a unique indoor series; (2) the curves are smoothed in order to limit the random fluctuations; (3) the series dimensionalities are reduced by taking one sample every four to decrease the computational cost without losing relevant information. However, due to the limited dataset, a data augmentation operation has also been mandatorily introduced.

When handling time series, there are a lot of techniques that may be used to enlarge a given dataset (Ismail Fawaz *et al.*, 2020). In this work, a mirror augmentation is performed. In this augmentation method, all the series are mirrored, reversing the order of their values but keeping the labels unchanged. It is important to underline that the augmentation is done only on the days in which there are scheduled events (identified with label 1). This is done because the difference between the number of samples for each label is too high; therefore, the dataset is defined *unbalanced* as the target variable (MR) has more observations in one specific class than the others. These data lead to a CNN architecture in which all the outputs are the most frequent targets. The main problem when dealing with models trained on unbalanced datasets is that they often provide poor results when generalizing. This means that the algorithm receives significantly more examples from one class, prompting it to be biased towards that particular class (Krawczyk, 2016). In this way, it is not able to learn what makes the other class “different” and fails in understanding the underlying patterns of class distinction. In particular, the challenge appears when the algorithms try to identify these rare cases in rather big datasets. Due to the disparity of classes, the algorithm tends to categorize the variables into the class with more instances, i.e. the *majority class*, while at the same time giving the false sense of a highly accurate model. Both the inability to predict rare events, i.e. variables in the *minority class*, and the misleading accuracy detract from the predictive models that have been built. For this reason, it is necessary to reduce the gap between the number of samples with different labels through data augmentation (Buda *et al.*, 2018). After the augmentation, because of the limited number of data, the performance has been further evaluated with the k-fold cross-validation to have a better estimation of the stability of the results looking at the scores. The cross-validation creates  $k$  different sets using the learning algorithm and tests it on  $k$  different test sets; hence, the results become more reliable. When a single evaluation is done on a test set, only one result is produced. By training on different splittings, the algorithm performance can be better understood.

### 3.3 Unexpected mixing ratio peak (UMR peak)

The third and last approach is a novel approach, herein proposed for the first time, called UMR peak. As the name suggests, this method is based on the heuristic assumption that days with scheduled events are characterized by at least one peak or variation in the indoor MR, which is not explained by the outdoor one.

For this reason, the starting point of UMR peak algorithm is a data manipulation operation which aims to extract one feature per each day, called *Delta* ( $\Delta$ ) and defined as the maximum value of the differences between two analogous peaks (as visible in Figure 2c and better explained later). This feature will ideally be accentuated when a variation in  $MR_{in}$  is not explained by  $MR_{out}$ , and it will be used to feed a supervised linear classifier capable to distinguish between days with events and days without events.



**Figure 2.** UMR peak algorithm data manipulation phases over the day 143 in Ringebu stave church: (a) original time series of  $MR_{in1}$  (yellow curve),  $MR_{in2}$  (red curve) and  $MR_{out}$  (green curve); (b) smoothed time series of  $MR_{in}$  (orange) and  $MR_{out}$  (green); (c) time series first derivatives:  $derMR_{in}$  (orange),  $derMR_{out}$  (green) and visual definition of feature *Delta*

In [Figure 2](#), the data manipulation operation is graphically shown and explained through all its steps. In [Figure 2a](#), the three time series of  $MR_{in1}$  (yellow curve),  $MR_{in2}$  (red curve) and  $MR_{out}$  (green curve) for Ringebu church over one day (day  $n$ . 143) of the monitored period are shown. They are computed starting from  $T$  and RH time series, which are directly measured by the three sensors, and they represent the input of the data manipulation operation.

Considering the dataset in its entirety (i.e. one-year long time series), the two indoor MR time series are averaged and smoothed. The smoothing phase is carried out by computing the moving average of the time series by choosing a time-window consisting of fifteen different measurements: the current measurement, the seven measurements before and the seven measurements after it for a total of 70 min time span. Since the two indoor sensors have a measurement rate which is three times bigger than the one of the outdoor sensor, it implies that the time window for outdoor time series must be three times larger than the time window for indoor time series. The smoothed curves of  $MR_{in}$  (orange) and  $MR_{out}$  (green) are plotted in [Figure 2b](#). As shown in [Figure 2c](#), the first derivatives of the time series,  $derMR_{in}$  in orange and  $derMR_{out}$  in green, are then computed. Starting from this point, in the data manipulation, it is necessary to divide these time series in 24-h-time windows, such that each day can be taken in consideration individually. The day chosen to visualize the final step of the procedure is still day  $n$  143 ([Figure 2](#)). The data manipulation goes on with looking for two peaks per each time series derivative that are defined following the heuristic assumption, which characterizes this method and an additional one: the presence of people inside the church causes an increase in the  $MR_{in}$  that is then followed by a decrease due to the people leaving the church. The first peak is defined as the maximum of  $derMR_{in}$  (red dot in [Figure 2c](#)). Assuming it is due to visitors entering the church, the second peak is defined as the minimum of  $derMR_{in}$  happened after it (red triangle in [Figure 2c](#)). This second peak should indicate MR decrease due to the people leaving the church. At this point, a comparison with external trend is needed: the third peak is defined as the maximum of  $derMR_{out}$  (blue dot in [Figure 2c](#)) happened before the first peak, and the fourth peak (blue triangle in [Figure 2c](#)) is defined in an analogous way. Once these four values are known, it is possible to define the feature  $\Delta$  as visually shown in [Figure 2c](#). The idea behind  $\Delta$  perfectly encodes the first heuristic assumption: the greater its value, the more likely it is that something unexpected occurred in the  $MR_{in}$  time series, with respect to the  $MR_{out}$  time series.

For this reason, once  $\Delta$  is computed for each day, the method consists in a one-parameter supervised classification. The only parameter is a threshold value for  $\Delta$ , called  $\Delta_{th}$ , such that the  $\Delta$  of the  $n$ -th day that exceeds this threshold is predicted to be a day with a scheduled event (day with label 1) as follows:

$$\Delta_n > \Delta_{th} \quad (1)$$

The parameter  $\Delta_{th}$  is optimized by maximizing the NMI of the classification over the Ringebu dataset, which is used as training set. Once the parameter is optimized, the method is tested over Heddal dataset.

## 4. Results

In this section, first the results of the different approaches are analyzed one by one and then some general considerations are deduced.

### 4.1 Density peak clustering results

Using the parameters of the DPC algorithm, the distribution has been forced to have only two clusters (days with events and days without events). In [Table 1](#), there is a synthesis of the results of the first approach tested within this algorithm. These results are related to the

analysis performed with the four features presented in Section 3.1. In particular in both churches, nonoptimal results are achieved in terms of precision, recall and NMI.

In addition, Figure 3 shows a projection of the four-dimensional clustering made with the *t*-SNE (*t*-distributed stochastic neighbor embedding) technique considering the Ringebu datasets (Van der Maaten and Hinton, 2008). On the axes, the two embedded dimensions (dim 1 and dim 2) of the clusters are reported in terms of an arbitrary unit (a.u.).

Every dot represents a day in the considered year that has been analyzed by the algorithm, which has tried to fragment this dataset into two components. In particular, the grey dots are those that the algorithm has recognized as the days with events, while the black dots as days without events.

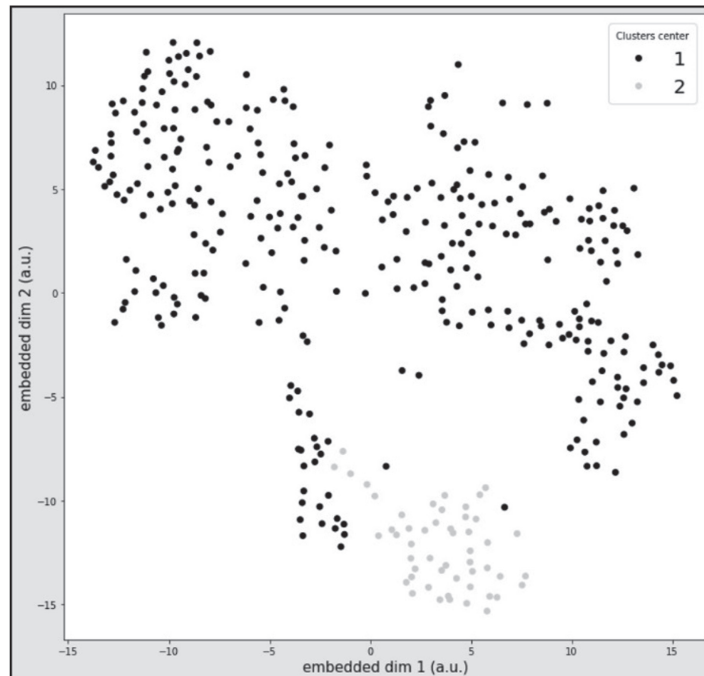
Since the results are not optimal, the analysis was run several times, and it was found out that the scores are even worse if the information about the temperature is not taken in consideration. This outcome is highlighted in Table 2.

As visible in Table 2, this analysis has worse outcomes compared to Table 1; the reason behind the difference in results is in the heating management of the two churches. In the case of Ringebu, there is a precise schedule that provides the moments when the church is heated. This

**Table 1.**

The three principal estimators for the goodness of the clusters (precision, recall and NMI) for the analyses carried out on Ringebu and Heddal church

	Precision	Recall	NMI
Ringebu	0.833	0.615	0.395
Heddal	0.164	0.962	0.111



**Figure 3.**

Projection of the four-dimensional clustering made with the *t*-SNE technique

**Note(s):** Every dot represents a day in the analysed year: days with events (grey) and days without events (black), as recognized by the algorithm

means that there is a recognizable pattern in the T time series. In particular, the heating is turned on in the days with celebrations in order to create a warm indoor climate for the visitors' comfort. Instead, the Heddal management for the heating is different: the internal climate is more stable since the heating is kept on continuously at moderate temperature during the cold period of the year. This means that there is no recognizable pattern for the algorithm. This leads to the conclusion that the clustering is able to just detect the days with the heating on in Ringebu and not really the days with events, thus not meeting the original goal. Actually, changing the subject in the computation of the NMI from days with events to days with heating on provides better results, as visible in Table 3.

As a matter of fact, the results that the algorithm gives in recognizing the days with the heating turned on are the best, even if this is not the initial purpose of the work.

In general, the choice of implementing an unsupervised method (DPC), albeit being weak in some cases, has been driven by the motivation that the recorded metadata could be affected by errors and inaccuracies that are mainly unknown to the users.

#### 4.2 Convolutional neural network results

The second approach is implemented using Keras-software® and Google-Colab© notebook. As already stated, the performances of this model are evaluated using a k-fold cross-validation (Berrar, 2018) with  $k = 5$ . The dataset is made by the union of the preprocessed time series of the indoor MR of the two churches: 730 samples are split in five chunks, with the training set being composed by 584 samples (80% of the whole dataset), and the validation set by 146 (20% of the whole dataset) one-dimensional time series. It can be noticed that the dimension of the training set is not large; therefore, according to literature (Luo *et al.*, 2019), it is not enough to train a CNN properly. The total number of trainable parameters is 155, and this number is voluntarily kept low to avoid a too complex architecture that, on a small dataset, may probably lead to overfitting. In Table 4, the results obtained for one run of the algorithm and the wide range of NMI for different splittings are reported. The other two scores reported in the table, i.e. loss and accuracy, are not relevant for the desired task because they are not giving the right feedback in reaching the desired goal, namely the detection of days with events. Indeed, if the accuracy is high, it is due to the large difference in the number of samples for the two labels. Hence, if the dataset is not enough balanced, the CNN cannot be properly trained for this task.

Herein (Table 4), the scores for the different folds of a single run are presented, while from the outcomes of more runs it emerges that the NMI has a fluctuating trend. Despite the cross-validation, it looks like the algorithm and/or the data is/are not consistent; it could be either that the algorithm is unable to learn or that the data are significantly complex. The results are

	Precision	Recall	NMI
Ringebu	0.475	0.292	0.072
Heddal	0.153	1.000	0.119

**Note(s):** The three principal estimators for the goodness of the clusters are reported, for the two analyzed churches

**Table 2.**  
Results for the second approach of the clustering method

	Precision	Recall	NMI
Ringebu	1.000	0.842	0.781

**Note(s):** The three principal estimator for the goodness of the clusters are reported for Ringebu Church

**Table 3.**  
Results for the last approach of the clustering method

strongly dependent on the splitting of the dataset and the cross-validation can only reduce the impact of the problem without avoiding it; this means that there is an inherent issue laying in the dataset. It is important that the data used to learn and those used to validate or test the model both follow a similar statistical distribution. One of the ways to achieve these criteria is to select the subsets randomly. Notwithstanding, the more unbalanced the dataset, the more difficult for the cross-validation to have sets following similar statistics. From a theoretical point of view, it can be stated that the CNN is a good solution for this kind of problem. Nevertheless, as seen from the results, the provided dataset is not optimal to train a CNN because it presents two critical points: first of all, the data are not enough for properly tuning the parameters; thus, the model is not able to learn the features to classify the time series. Then, as a consequence of the previous point, there is a strong class imbalance which increases the difficulty in training and evaluating the architecture. However, to face the issue of the limited datasets, data augmentation has been carried out in order to provide more observations to the model for a better generalization. This approach has been preferred respect to assigning different weights to the classes as the latter can strongly influence the classification during the training phase by setting a higher weight for the minority class (Wen *et al.*, 2020). As a matter of fact, the goal of the implemented approach is not about removing the class imbalance but about classifying the minority class even if this means losing in generalization capabilities.

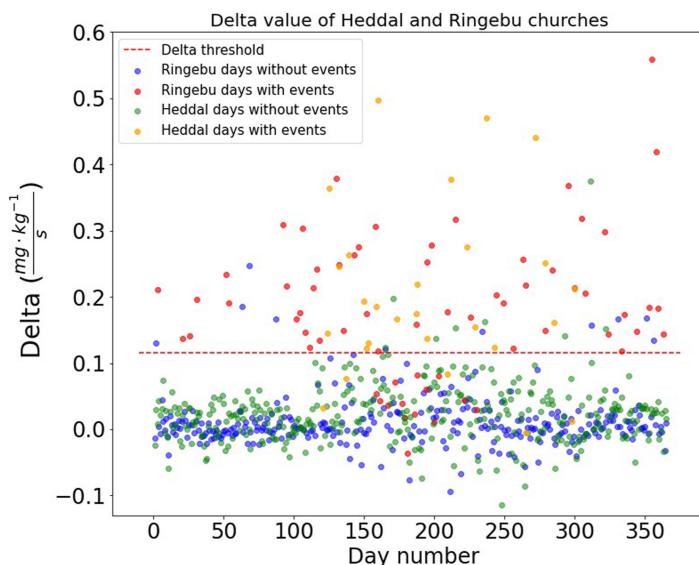
4.3 Unexpected mixing ratio peak results

The values of  $\Delta$  computed for each day in Ringebu and Heddal are reported in Figure 4. Here, the days with scheduled events (red dots for Ringebu and yellow dots for Heddal), the days without events (blue dots for Ringebu and green ones for Heddal) and  $\Delta_{th}$  (dashed line found by maximizing NMI over Ringebu dataset) are reported. Once known that  $\Delta_{th}$  is equal to  $0.115 \frac{mg}{kg s^3}$ , the performance over Heddal dataset has been computed and compared to the Ringebu's one. The performances in terms of precision, recall and NMI are reported in Table 5. From Figure 4, it is visible how for both churches there is a good separation between days with "zero-centered"  $\Delta$  and days with higher values of  $\Delta$ , in particular during the winter period (from day 270 on). It is interesting to notice that, during the same period, in Heddal church, there are no days (green or yellow dots) with a value of  $\Delta$  which is greater than  $\Delta_{th}$ . This makes sense since the church is closed to visitors in that period of the year, and the impact of the Heddal's heating system is not as intense as Ringebu's heating system. Moreover, it is also interesting to notice that most of misclassified days (yellow or red dots above the threshold) are summer days. This can be easily understood from the fact that in summer, due to weather and warm climate, the classification is more challenging as the church cannot be considered a closed system (as the doors and windows are kept opened more often). From performances reported in Table 5, it is clear how UMR peak outperforms previous methods. In particular, it seems a very solid approach, obtaining similar

	Loss	Accuracy (%)	NMI
Fold 1	0.18 ± 0.04	94 ± 2	0.5 ± 0.1
Fold 2	0.26 ± 0.04	89 ± 2	0.4 ± 0.1
Fold 3	0.43 ± 0.04	76 ± 2	0.1 ± 0.1
Fold 4	0.21 ± 0.04	93 ± 2	0.6 ± 0.1
Fold 5	0.41 ± 0.04	87 ± 2	0.3 ± 0.1
Average	0.30 ± 0.02	87.7 ± 0.9	0.37 ± 0.04

**Table 4.**  
Results of five-fold cross-validation

**Note(s):** Loss = categorical cross-entropy; Average = average between the scores of the different folds in percentage. Each error is computed as standard deviation



**Note(s):** The red dashed line represents the value of  $\Delta_{th} = 0.115 \frac{mg}{kg \cdot s}$ , estimated over Ringebu dataset

**Figure 4.** Value of  $\Delta$  per each day in Ringebu and Heddal datasets, classifying between days with events (red dots for Ringebu and yellow dots for Heddal) and days without events (blue dots for Ringebu and green ones for Heddal)

	Ringebu (training set)	Heddal (test set)
Precision	0.84	0.73
Recall	0.78	0.82
NMI	0.52	0.53

**Table 5.** Performances of UMR peak over the two churches

performances over both Heddal and Ringebu churches. Regarding this observation, it is visible from Figure 4 how the two  $\Delta$  distributions overlap; this may suggest that some aspects of  $\Delta$  distribution do not depend on the specific church, an aspect which may ask for further research in the future.

Concerning the meaning of  $\Delta_{th}$ , since  $\Delta$  is a difference of MR derivatives, it can be stated that  $\Delta$  encodes how much  $MR_{in}$  changes with respect to  $MR_{out}$  in a given time window. Then,  $\Delta_{th}$  characterizes how significant this change should be in order to distinguish between the two categories of days (i.e. with events and without events). Therefore, the UMR peak approach, although being a simple feature engineering procedure, provides an innovative way of manipulating time series and extracting useful information; moreover, it has shown to be more reliable and robust than the previous algorithms. Finally, the main results of the three proposed algorithms are summarized in Table 6.

	DPC	CNN	UMR
Ringebu	0.39	0.37	0.52
Heddal	0.11	0.37	0.53

**Table 6.** NMI for the three different methods for the two churches

## 5. Conclusions

In this work, starting from the analysis of indoor (and outdoor) temperature and relative humidity measurements collected in Ringebu and Heddal stave churches (Norway) during the 2019–2020 calendar year, there is an attempt in distinguishing between days with and without events to see the impact of visitors over the internal microclimate using three different methods of the modern machine learning technique.

It has been observed how the first method, the clustering approach, is strongly dependent on the church management: in particular, for Ringebu church, the DPC algorithm identifies the heated days better than the presence of people. This result underlines how the microclimate of this church is strongly conditioned by the heating system. At the opposite, with this method, it is more difficult to identify days with events in Heddal church because its heating management is less impacting. As a matter of fact, in Heddal church, a “conservation heating” strategy is adopted, based on keeping a stable indoor temperature level ( $\approx 5$  °C) for conservation rather than comfort purposes.

The second method, the CNN approach, is promising but, due to the small dataset, it cannot be properly trained and, consequently, shows unstable performances. To solve the problems related to the CNN, in future studies, it could be interesting to investigate a new architecture based on the combination of the CNN with a recursive neural network (RNN). In this way, it may be possible to find an underlying structure in the dataset, looking to what happens before a specific peak in the time series, similarly as it is done here with the UMR peak method.

The last method, the UMR peak approach, provides the best results; moreover, it seems solid over case studies with different indoor management strategies; in order to further prove this capability, more case studies may be taken in consideration for future research directions. However, the UMR peak method is based on an arbitrary smoothing procedure, which is specific for this problem. Future works may include the formalization of this method in the general task of comparing different time series and quantifying their mutual relations; in addition, it could be interesting to combine the extracted feature ( $\Delta$ ) with other variables and to use them as input to other machine learning algorithms.

Finally, although the information that was used as the ground truth for the above implementations is suitable to know the occurrence of the events taking place in the two churches, the practical usefulness of the proposed approaches lays in the possibility to catch patterns in incomplete, corrupted or inaccurate data. As a matter of fact, the metadata used as ground truth have been proven to be often inaccurate as they are mainly recorded by hand. Therefore, in principle, the algorithms can be used to track any kind of event, even unknown or unexpected. In addition, this aspect could be ideally applied to any field in which time series processing and manipulation is used as it would allow detecting anomalies or alterations.

In conclusion, it could be possible to extend the proposed approaches to other kinds of case studies in order to provide a methodology to identify microclimatic variations and/or unexpected events that may be risky for the conservation of historic buildings.

## References

- Berrar, D. (2018), “Cross-validation”, *Encyclopedia of Bioinformatics and Computational Biology*, Elsevier, Vol. 1, pp. 542-545.
- Buda, M., Atsuto, M. and Mazurowski, A. (2018), “A systematic study of the class imbalance problem in convolutional neural networks”, *Neural Networks*, Vol. 106, pp. 249-259.
- Camuffo, D. (2018), “The role of temperature and moisture”, *Basic Environmental Mechanisms Affecting Cultural Heritage*, Vol. 11, pp. 1-23.
- Camuffo, D. (2019), *Microclimate for Cultural Heritage*, Elsevier Science. doi: [10.1016/C2017-0-02191-2](https://doi.org/10.1016/C2017-0-02191-2).

- 
- d'Errico, M., Facco, E., Laio, A. and Rodriguez, A. (2021), "Automatic topography of high-dimensional data sets by non-parametric density peak clustering", *Information Sciences*, Vol. 560, pp. 476-492.
- Ismail Fawaz, H., Lucas, B., Forestier, G., Pelletier, C., Schmidt, D.F., Weber, J., Webb, G.I., Idoumghar, L., Muller, P.A. and Petitjean, F. (2020), "InceptionTime: finding AlexNet for time series classification", *Data Mining and Knowledge Discovery*, Vol. 34 No. 6, pp. 1936-1962.
- Krawczyk, B. (2016), "Learning from imbalanced data: open challenges and future directions", *Progress in Artificial Intelligence*, Vol. 5 No. 4, pp. 221-232.
- Lam, D. and Wunsch, D.C. (2014), "Clustering", *Academic Press Library in Signal Processing: Volume 1 - Signal Processing Theory and Machine Learning*, doi: [10.1016/b978-0-12-396502-8.00020-6](https://doi.org/10.1016/b978-0-12-396502-8.00020-6).
- Luo, C., Li, X., Wang, L., He, J., Li, D. and Zhou, J. (2019), "How does the data set affect CNN-based image classification performance?", *2018 5th International Conference on Systems and Informatics, ICSAI 2018*, IEEE, No. Icsai, pp. 361-366.
- MacKay, D.J.C. (2005), *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press.
- Mehta, P., Bukov, M., Wang, C.H., Day, A.G.R., Richardson, C., Fisher, C.K. and Schwab, D.J. (2019), "A high-bias, low-variance introduction to Machine Learning for physicists", *Physics Reports*, Vol. 810, pp. 1-124.
- Olstad, T.M., Ørnhøj, A.A., Jernæs, N.K., de Ferri, L., Freeman, A. and Bertolin, C. (2020), "Preservation of distemper painting: indoor monitoring tools for risk assessment and decision making in kvernes stave church", *Climate*, MDPI AG, Vol. 8 No. 2, doi: [10.3390/cli8020033](https://doi.org/10.3390/cli8020033).
- Refaeilzadeh, P., Tang, L. and Liu, H. (2009), "Cross-validation", in Liu, L. and Özsu, M.T. (Eds), *Encyclopedia of Database Systems*, Springer, doi: [10.1007/978-0-387-39940-9\\_565](https://doi.org/10.1007/978-0-387-39940-9_565).
- Rodriguez, A. and Laio, A. (2014), "Clustering by fast search and find of density peaks", *Science*, Vol. 344 No. 6191, pp. 1492-1496.
- Tang, W., Long, G., Liu, L., Zhou, T., Jiang, J. and Blumenstein, M. (2020), "Rethinking 1D-CNN for time series classification: a stronger baseline", available at: <http://arxiv.org/abs/2002.10061>.
- UNI EN 15757:2010 Standard - Conservation of Cultural Property - Specifications for Temperature and Relative Humidity to Limit Climate-Induced Mechanical Damage in Organic Hygroscopic Materials (2010).
- Van der Maaten, L. and Hinton, G. (2008), "Visualizing data using t-SNE", *Journal of Machine Learning Research*, Vol. 9 No. 86, pp. 2579-2605.
- Wen, Q., Sun, L., Yang, F., Song, X., Gao, J., Wang, X. and Xu, H. (2020), "Time series data augmentation for deep learning: a survey", *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI-21)*.
- Wu, M. and Chen, L. (2016), "Image recognition based on deep learning", *Proceedings - 2015 Chinese Automation Congress, CAC 2015*, pp. 542-546.
- Young, T., Hazarika, D., Poria, S. and Cambria, E. (2018), "Recent trends in deep learning based natural language processing", *IEEE Computational Intelligence Magazine*, Vol. 13 No. 3, pp. 55-75.
- Zhao, B., Lu, H., Chen, S., Liu, J. and Wu, D. (2017), "Convolutional neural networks for time series classification", *Journal of Systems Engineering and Electronics*, Vol. 28 No. 1, pp. 162-169.

### Corresponding author

America Califano can be contacted at: [america.califano@unipd.it](mailto:america.califano@unipd.it)

---

For instructions on how to order reprints of this article, please visit our website:

[www.emeraldgrouppublishing.com/licensing/reprints.htm](http://www.emeraldgrouppublishing.com/licensing/reprints.htm)

Or contact us for further details: [permissions@emeraldinsight.com](mailto:permissions@emeraldinsight.com)