buildings

MDPI

Article

Performance of Various Artificial Intelligence Models for Predicting Temperature in an Industrial Building—A Case Study

Johan Roussel [1,2,*,†], Zoubeir Lafhaj [1], Pascal Yim [1,†], Thomas Danel [2] and Laure Ducoulombier [1]

[1] Laboratoire de Mécanique, Multiphysique, Multiéchelle (LAMCube), UMR 9013, Centrale Lille, Université de Lille, 59000 Lille, France; zoubeir.lafhaj@centralelille.fr (Z.L.); pascal.yim@centralelille.fr (P.Y.); laure.ducoulombier@centralelille.fr (L.D.)

[2] DECIMA, ZI-Est Rue François Hennebique BP 51, 62052 Saint-Laurent-Blangy, Pas-de-Calais, France; tdanel@decima.fr

* Correspondence: johan.roussel@centralelille.fr; Tel.: +33-651-422-644

† These authors contributed equally to this work.

Abstract

This article presents a comparative analysis of the performance of various artificial intelligence models for predicting temperature in an industrial building. The main objective is to identify an optimal algorithm that enables efficient thermal management, which is essential for ensuring product quality, maintaining process safety, and optimizing energy consumption. Through an in-depth study, this research evaluates the accuracy, learning speed, and adaptability of the models in a dynamic industrial environment. Gradient Boosting algorithms, notably XGBoost and LightGBM, have demonstrated a capacity to give promising results, particularly for short- and mid-term predictions. The main results were obtained with XGBoost, achieving a mean absolute error of only 0.32 °C for a 5 min prediction horizon—below the 0.50 °C accuracy of the sensors used. The article also highlights the practical implications of these findings for Industry 4.0. By integrating high-performing predictive models into equipment control systems, it is possible to reduce energy costs, improve thermal comfort, and ensure efficient process stability. This research underscores the importance of an algorithmic approach tailored to both the thermal characteristics of the building and the computational capabilities of industrial infrastructure, with feature importance analysis confirming that external temperature and heating system states are the primary predictive factors, thereby helping to optimize industrial operations.

Keywords: temperature prediction; energy efficiency; artificial intelligence; industrial environment; Industry 4.0

check for updates

Academic Editor: Bo-Tao Huang

Received: 22 May 2025
Revised: 1 July 2025
Accepted: 4 July 2025
Published: 10 July 2025

Citation: Roussel, J.; Lafhaj, Z.; Yim, P.; Danel, T.; Ducoulombier, L. Performance of Various Artificial Intelligence Models for Predicting Temperature in an Industrial Building—A Case Study. *Buildings* **2025**, *15*, 2428. https://doi.org/10.3390/buildings15142428

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Optimizing energy use and managing thermal conditions in industrial buildings are critical challenges in the context of the energy transition and reducing greenhouse gas emissions. Efficient management of Heating, Ventilation and Air-Conditioning (HVAC) systems is essential to ensure occupant comfort, maintain optimal production conditions, and minimize energy consumption [1,2]. In this context, Artificial Intelligence (AI) is emerging as a solution for predicting and controlling indoor temperatures.

In response, recent French regulations promote infrastructure modernization to address energy efficiency goals. The French Law 'Évolution du Logement, de l'Aménagement et du Numérique (ELAN)' sets ambitious targets for the energy performance of buildings, notably through the integration of intelligent thermal management technologies [3]. The

tertiary sector decree requires buildings for tertiary use to progressively reduce their final energy, thus necessitating advanced control solutions and predictive systems for greater energy efficiency [4]. In addition, the Building Automation and Control Systems (BACS) decree requires the installation of building automation and control, reinforcing the interest in predictive models capable of optimizing thermal conditions in real time and anticipating temperature variations [5].

Thermal comfort is a central issue in the design and management of intelligent buildings, particularly in a context where energy sustainability and occupant wellbeing have become major priorities. Standards such as ISO 7730 [6] and ASHRAE 55 [7] play a fundamental role in providing methodological frameworks for assessing and guaranteeing a comfortable thermal environment. These standards are grounded in models that incorporate environmental and physiological parameters, such as Predicted Mean Vote (PMV) and Predicted Percentage of Dissatisfied (PPD) for ISO 7730, and adaptive approaches for ASHRAE 55 [8,9].

Historically, thermal prediction models have relied on physical and statistical approaches. Although robust, these methods have limitations when faced with the dynamic and complex environments of industrial buildings, notably due to their inability to capture the non-linear interactions and temporal dependencies of environmental parameters [10]. As traditional models reached their limits, new approaches emerged: Machine Learning (ML) and Deep Learning (DL) techniques have enhanced the accuracy and flexibility of predictive models, opening up new perspectives for energy optimization and intelligent thermal control [11].

This study proposes to analyze and compare seven AI-based approaches (including three ML approaches and four DL approaches) to predict temperature in an industrial building. This comparative study focuses on AI algorithm performance rather than comparison with physical models, whose limitations are well established in existing literature. Using a dataset from an industrial workshop equipped with environmental sensors, the performance of 7 predictive models is evaluated, with a focus on their accuracy and ability to adapt to dynamic environments.

The various artificial intelligence approaches utilized in this study can be categorized into distinct algorithmic families, as illustrated in Figure 1. This mind map provides a comprehensive overview of the machine learning and deep learning techniques relevant to temperature prediction tasks. The left branch represents traditional machine learning algorithms, including the Gradient Boosting methods (XGBoost, LightGBM, and CatBoost) employed in our comparative analysis. The right branch displays deep learning architectures, highlighting the neural network models (LSTM, GRU, CNN1D, and CNN2D) that we evaluate for their predictive capabilities in industrial environments.

This research evaluates and compares the performance of seven AI-based approaches for temperature prediction in an industrial workshop, including XGBoost, LightGBM, and CatBoost (ML), as well as LSTM, GRU, CNN1D, and CNN2D (DL). The objective is to identify the most suitable model by balancing predictive accuracy, computational efficiency, and robustness to varying industrial conditions, with the goal of maximizing energy efficiency and enhancing thermal comfort. This work contributes to the broader movement of technological innovation in the context of Industry 4.0, promoting intelligent management of industrial buildings through advances in AI and DL while adhering to current legal requirements.

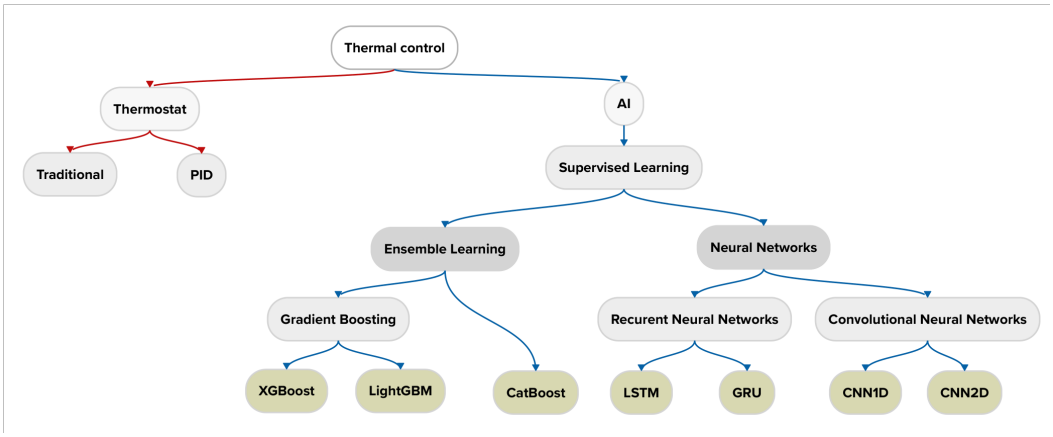

Figure 1. Mind map showing the different families of artificial intelligence algorithms used for prediction.

2. State of the Art: Temperature Modeling in Buildings

The evolution of modeling techniques for thermal prediction of industrial buildings reflects significant advances in AI and ML. This section presents a review of traditional methods, then explores modern approaches, focusing on the algorithms employed in this study.

2.1. Traditional Energy Optimization Methods: Mathematical and Physical Models

Before the emergence of AI-based optimization, energy management relied mainly on mathematical and physical models. These methods included Model Predictive Control (MPC), which minimized energy consumption by anticipating temperature variations and the thermal needs of buildings [12].

Among these approaches, three main categories have been widely studied:

- Linear predictive control models using differential equations to predict changes in indoor temperatures and adjust heating and ventilation systems accordingly [13,14].
- Thermal state models based on a simplified thermal representation of the building to optimize energy management according to heat loss and solar gain [15,16].
- Optimization by mathematical programming, taking into account physical and economic constraints in order to maximize energy efficiency while guaranteeing thermal comfort [17,18].

2.2. Limits of Traditional Models: Statistical and Physical Approaches

Early attempts at thermal prediction were based on physical models, such as PMV and thermodynamic equations [10]. These methods, while robust, suffered from a lack of flexibility in dynamic industrial environments. Their inability to capture the complex, non-linear interactions between different variables limited their accuracy [19].

Statistical approaches, such as linear regression and autoregressive models (AR, ARIMA), were then adopted. These models offered a better account of the relationships between variables, but remained insufficient to model the unpredictable fluctuations of temperature in an industrial environment [20].

2.3. The Contribution of Machine Learning Techniques: A New Era of Thermal Prediction

With the increase in computing power and the growing availability of data, DL has established itself as an effective alternative. Established machine learning models, including ensemble methods like Random Forest (RF) and kernel-based approaches such as Support Vector Machine (SVM), have improved prediction accuracy by exploiting multiple building characteristics and environmental conditions [10]. Given the substantial body of existing

research on these established methods, a detailed re-examination falls outside the scope of this current study.

2.4. Gradient Boosting Models: XGBoost, LightGBM, and CatBoost

Gradient Boosting algorithms, such as XGBoost, LightGBM, and CatBoost, have marked a major advance in thermal modeling. These methods iteratively combine several decision trees to minimize prediction error:

- XGBoost is widely used because of its robustness and computational optimization, notably through sparsity management and parallelism [1].
- LightGBM favors a leaf-by-leaf tree growth approach, which improves learning speed and reduces memory consumption [21].
- CatBoost is particularly well suited to categorical data and reduces overlearning thanks to specific permutation of categorical values [22].

These algorithms are particularly effective for thermal prediction, as they handle tabular data well and capture the complex relationships between sensors and environmental parameters [23].

2.5. Neural Network Models: Recurrent and Convolutional

Deep neural networks have also been explored for thermal prediction, including recurrent and convolutional models:

- Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) are recurrent neural networks capable of capturing long-term temporal dependencies in time series. They are suitable for modeling thermic variations, although their performance can be affected by high computational cost [24].
- CNN1D applies convolution filters along the time dimension extracting local patterns and improving short-term prediction [25].
- CNN2D processes the data in the form of spatio-temporal matrices, enabling identification of more complex correlations between different areas of the building [2].

2.6. Comparative Analysis and Research Prospects

Gradient Boosting models have shown high accuracy for thermal prediction, particularly for short to medium prediction horizons. Neural networks, while effective in capturing complex temporal dynamics, have a tendency to lose accuracy over the long term due to uncertainties in input data [26].

The integration of specialized libraries, such as Skforecast, could represent a future breakthrough in combining the power of Gradient Boosting models with advanced time series techniques. This approach would enable better adaptation to industrial environments and increased optimization of thermal management systems [27].

Evolution of predictive models towards hybrid approaches combining AI and energetic optimization opens up new prospects for the intelligent management of industrial buildings, in line with the requirements of Industry 4.0 [28].

2.7. Metrics for Assessing Model Performance

To assess the quality of predictions, four classic regression metrics were selected [23]:

1. Mean Absolute Error (MAE) allows the average error to be interpreted directly in degrees Celsius (or in units of measurement). This metric evaluates the average

deviation between predictions and observed values by maintaining a consistent unit (here in °C), thus making it easier to interpret the results. The MAE is calculated as

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{1}$$

where n is the number of observations, y_i is the actual value, and $\hat{y}_i$ is the predicted value. A good value of MAE is one that is close to zero, indicating that the predictions are, on average, very close to the actual values.

2. Root Mean Squared Error (RMSE) penalizes large errors to a greater extent. It gives greater weight to large deviations, which helps to identify models that are sensitive to extreme errors. The RMSE is computed as

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{2}$$

A good value of RMSE is also close to zero, and as it strongly penalizes large errors, it is particularly useful in contexts where large errors are more of a problem.

3. Mean Absolute Percentage Error (MAPE) allows you to assess the relative error, which is useful if the values can vary by different orders of magnitude. This metric is particularly useful when temperature amplitudes vary significantly, as it expresses the error as a percentage of the true value, providing a relative assessment of the accuracy of the model. The MAPE is defined as

$$\text{MAPE} = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{3}$$

A good value of MAPE is also close to zero, usually expressed as a percentage.

4. Coefficient of determination R^2 indicates the proportion of variance explained by the model. The coefficient of determination is calculated as

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{4}$$

where $\bar{y}$ is the mean of the observed values, SS_{res} is the sum of squares of residuals, and SS_{tot} is the total sum of squares. A good value of Coefficient of determination (R^2) is close to 1, indicating an excellent generalization capacity of the model. Conversely, a value close to 0 or negative indicates a poorly performing model.

3. Methodology: Case Studies and Data Collection

3.1. Industrial Site Overview: Characteristics and Technical Details

This study was conducted in a 6000 m^2 single-storey workshop located near Paris, characterized by 15 m-high ceilings and minimal insulation. The building is of concrete construction with a saw-tooth fibrocement roof and glass panes. To provide access for pedestrians and vehicles, it has 23 openings on the north and south facades. Pedestrian doors are equipped with a door closer, although those for vehicles are manually operated.

The heating system comprises 8 gas air heaters with a total output of 3.4 MW.

The overall architecture of the monitoring and control system is presented in Figure 2, showing a top view illustrating the relationship between sensors, control units, and the central data management system. This architecture provides the foundation for the data collection infrastructure described in the following sections. The computational infrastructure consists of standard industrial multi-service machines providing sufficient processing

capacity for real-time model inference without hardware constraints typical of edge computing environments.

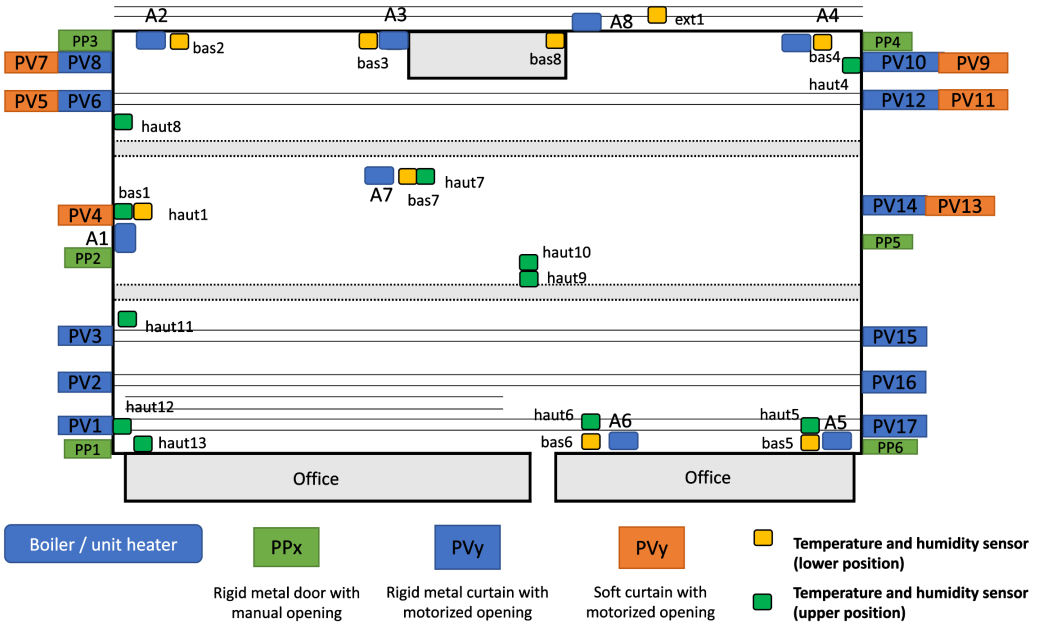

Figure 2. Diagram illustrating the layout of sensors and actuators in the industrial building.

3.2. Initial Diagnosis: Observations and Energy Issues

The observed average temperature is above 16 °C.

The peak consumption for the 2017/2018 heating period was 350,000 m^3 of gas, i.e., 650 kW h m^{-2} and an average of 322,000 m^3 over the following 6 years, i.e., 598 kW h m^{-2}.

These results were obtained using the formula below and the Lower Heating Value (*LHV*) of 11.15 data available on the official national gas supplier website [29].

$$E_{surf} = \frac{V_{gas} \times LHV}{A} \tag{5}$$

E_{surf}: Surface energy in kilowatt-hours per square meter (kWh/m^2)

V_{gas}: Volume of gas consumed in cubic meters (m^3)

LHV: Lower Heating Value of the gas, expressed in kilowatt-hours per cubic meter (kWh/m^3)

A: Area concerned in square meters (m^2)

3.3. Nomenclature of Installed Sensors and Actuators

The building was equipped as follows:

- The 6 pedestrian doors are equipped with a closing contact PP_x (x represents the door number).
- The 17 vehicle doors were also fitted with a closing contact noted PV_y (y represents the door number).
- The 5 double doors (flexible + metal) were fitted with closing contacts noted PV_y (y represents the door number).
- The 8 hot air generators were equipped as follows:
  - A temperature and humidity sensor (DHT22 with 0.50 °C accuracy) close to the original BAS_z (z represents the gas air heater number).
  - A 3-position multistate button allowing control by the original thermostat, by the installed solution, and by a forced shutdown noted $BTN_a_POS_b$ (a represents

gas air heater number, b represents position button 1, 2, or 3, respectively "auto", "stop", and "manual").
  - A gas air heater control $comRELAIS_c$ (c is the gas air heater's number).
  - A gas meter $impGAZ_d$ (d is the gas air heater's number).
- Temperature and humidity sensors were also installed at a height of 6 m.

4. Data Preparation and Processing

4.1. Database Data Collection and Analysis

Workshop data is stored in a MySQL/MariaDB relational database. The system's architecture enables the control units to manage several pieces of equipment simultaneously, each input/output being dedicated to the control of a specific element of the equipment. Figure 3 illustrates the structure of this database, revealing an organization centered on operational needs rather than AI modeling requirements. This architecture, designed for daily facility management, required significant transformation to be usable by our predictive algorithms.

An extraction of this database was performed on 17 April 2024 at 10:20, comprising 42,222,594 rows and 3 columns covering the period from 12 October 2022 at 17:09:35 to 17 April 2024 at 10:20:45. This considerable volume of raw data underwent comprehensive processing including cleaning, normalization, and extraction of relevant features for machine learning.

Following this preparation process, a structured dataset was constructed of 124,700 rows and 100 columns. This transformation allowed the isolation of essential predictive variables while preserving the integrity of temporal and causal relationships necessary for modeling the building's thermal behavior. Only a subset of these data proved relevant for training the predictive models, with the other variables primarily serving for contextualization and results validation. The feature set focuses on temporal lags and direct system states (relay positions, door states) rather than complex derived features, which is consistent with the building's low thermal inertia characteristics, where instantaneous relationships dominate over complex temporal patterns.



Figure 3. Schematic representation of the centralized building data.

4.2. Data Transformation Methodology for Modeling

The data transformation methodology is built around a set of essential steps [30]. The complete process, illustrated in Figure 4, is organized into three main phases:

- Preparation
- Processing
- Lag and normalization

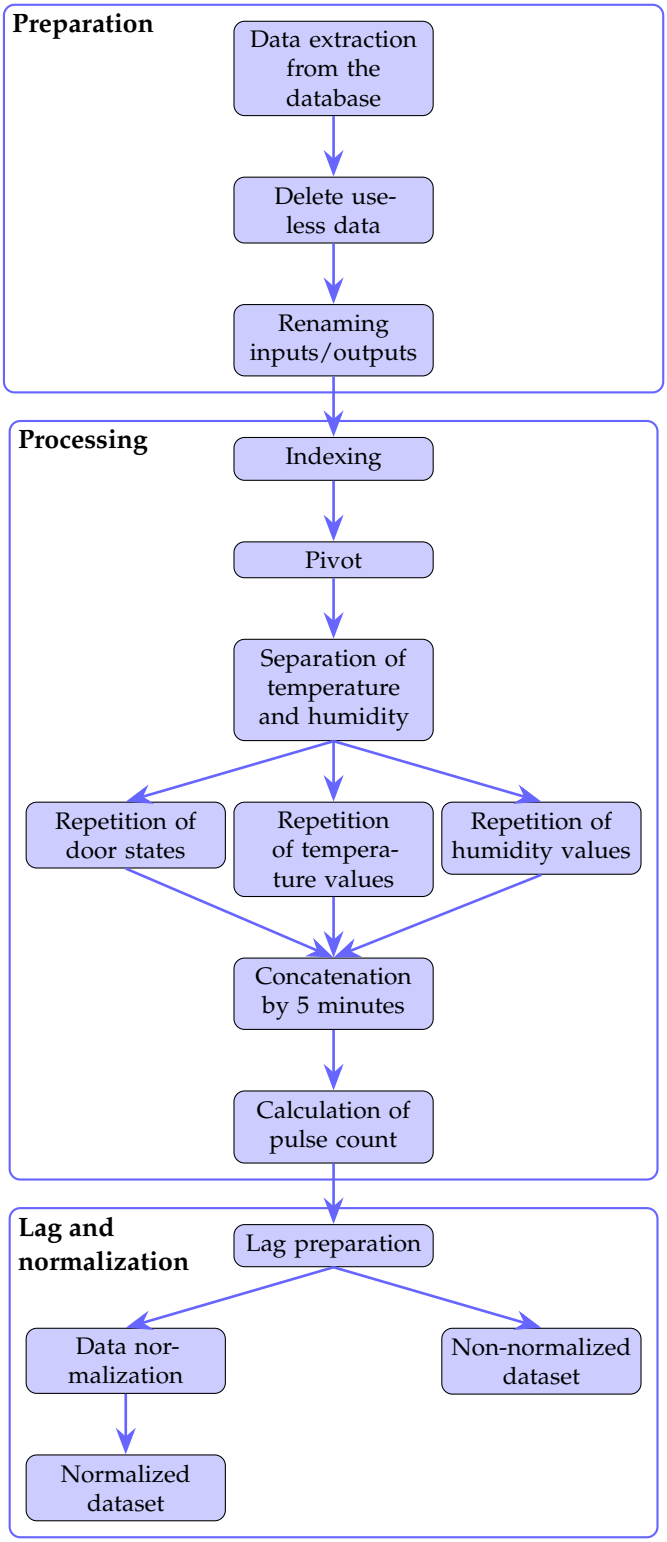

Figure 4. Flowchart of the steps involved in transforming and preparing raw data for modeling.

4.2.1. Preparation Phase

This phase includes the initial steps to structure and clarify the dataset:

- **Data extraction:** Raw data is extracted from the database in CSV format, retaining only relevant information. Tables 1 and 2 illustrates the structure of this dataset, moving from an initial format Table 1 to a transformed format Table 2.
- **Delete useless data:** Lighting control information (including "passageway lighting" and "outdoor lighting") has been removed from the dataset, as it is not relevant to the

analysis. System monitoring elements such as "heartbeat" and "Arduino Heartbeat" signals have also been excluded, since they do not provide meaningful information for this study.

- **I/O renaming:** Technical identifiers are replaced by explicit names to facilitate analysis (e.g., `tempEXT`, `HumExt`, `PP4`).

Table 1. Structure of extracted data before transformation for time series analysis and modeling.

date (timestamp) YYYY-MM-DD hh:mm:ss	**id in/out**	**Value**
2022-10-12 17:09:35	35	19.93°C 66.50% RH
2022-10-12 17:09:55	76	1
2022-10-12 17:10:14	80	1
2022-10-12 17:10:34	80	0
2022-10-12 17:10:54	80	1
2022-10-12 17:14:35	35	19.42 °C 66.12% RH
2022-10-12 17:14:55	37	1
2022-12-12 17:18:00	76	0
2022-10-12 17:19:55	37	0
2022-10-12 17:19:35	35	19.18 °C 66.03% RH
…	…	…

Table 2. Structure of extracted data after transformation for time series analysis and modeling.

Date (Timestamp)	**tempEXT (id 35)**	**HumExt (id 35)**	**PV4 (id 76)**	**comRELAIS2 (id 37)**	**impGAZ2 (id 80)**	**…**
2022-10-12 17:10:00	19.93	66.50	1	0	0	…
2022-10-12 17:15:00	19.42	66.12	1	1	3	…
2022-10-12 17:20:00	19.18	66.03	0	0	0	…
…	…	…	…	…	…	…

4.2.2. Processing Phase

This phase covers transformation and feature engineering operations required for modeling:

- **Indexing:** Each record is indexed by date and time to allow chronological sorting and time-based data analysis.
- **Pivot:** The dataset is reorganized with variables as columns and timestamps as rows, improving its suitability for analysis.
- **Temperature and humidity separation:** These variables are split into distinct series to allow independent analysis of their influence.
- **State repetition:** Values for door states, humidity, and temperature are repeated to create continuous time series. Temperature and humidity are interpolated, while door state data is filled using the forward fill (`ffill`) technique.

- **Concatenation by 5 min:** Data is grouped into 5 min intervals, forming regular time series as shown in Table 2. This interval balances the need for meaningful resolution and reduced sensitivity to noise, particularly given the high thermal inertia of the building.
- **Pulse count calculation:** The number of gas pulses is computed for each interval, providing a quantifiable measure of gas consumption.

4.2.3. Lag and Normalization Phase

This final phase prepares the dataset for input into modeling algorithms:

- **Lags preparation:** Time-shifted variables are created to represent the system's past states. This consists of associating variable values from previous intervals with the current time step.
- **Data normalization:** Features are normalized to reduce scale-related bias and enhance model convergence, particularly for algorithms sensitive to input magnitudes.

This process results in two versions of the dataset: a normalized dataset and a non-normalized one. The choice between them depends on the model used: neural networks typically require normalized inputs using MinMax scaling to facilitate convergence, whereas other models can operate directly on raw data.

4.3. Selection of Relevant Data for Analysis

The building under study is heated from October to April. Consequently, the analysis of temperature data focuses on this heating period, when thermal variations are significant and relevant to the study. During summer months, heating is unnecessary as external temperatures consistently exceed comfort thresholds, making this period irrelevant for evaluating heating system performance and control algorithms.

To guarantee the reliability of results, the selected period for data processing runs from January 25 to 3 April 2024. This period was chosen because it corresponds to the timeframe during which all sensors functioned perfectly, thus ensuring data collection without any gaps or major anomalies. This January–April period also encompasses significant seasonal variation, with external temperatures ranging from negative values in winter to approximately 20 °C in early spring, providing robust evaluation conditions for model generalization across changing climatic conditions. Earlier attempts at data collection were compromised by significant issues with the gas consumption sensors, which produced intermittent readings and occasionally reported physically impossible values. Additionally, the building management system experienced several software failures during the fall months, resulting in corrupted data logs and unreliable control signal records.

Figure 5 presents a comprehensive visualization of the prepared data spanning fifteen months, enabling assessment of their distribution and quality before implementation in predictive models. This visualization is structured in four distinct panels illustrating: internal (tempBAS1) and external (tempEXT) temperature variations, energy consumption readings in kWh, operational states of the heating system control relays, and building door activity. Seasonal cycles are clearly observable, with external temperatures fluctuating between negative values in winter and reaching approximately 30 °C in summer, while internal temperature is maintained around 15–20 °C during the heating season. The colored background sections indicate different operational periods, including off-heating seasons, data error instances, and the data segments effectively used for analysis.

Only the selected period from January to April provides suitable conditions for AI model testing, as it represents the only timeframe when the entire monitoring and control infrastructure was fully operational, with all sensors reporting accurate readings and the control system responding as designed.

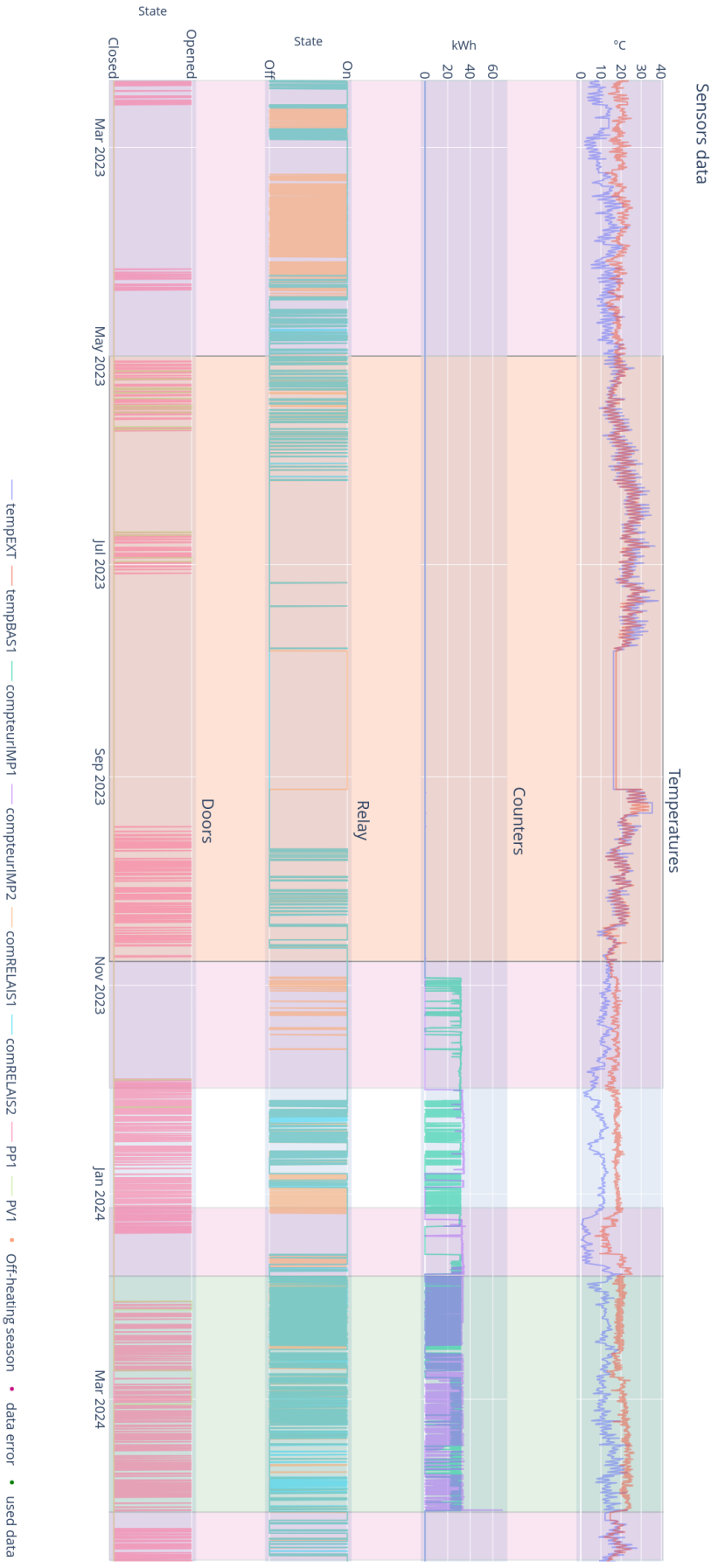

Figure 5. Visualization of data prepared to identify the optimum operating period.

5. Design and Evaluation of Prediction Models

5.1. Hypotheses and Experimental Design

Modeling focused on a specific temperature sensor: This strategy consists in developing a predictive model dedicated to the temperature of a single sensor, rather than seeking to simultaneously predict the temperatures of all the sensors on the shop floor. The temporal validation strategy employed a chronological split to evaluate model performance on unseen seasonal data, avoiding the data leakage issues inherent in traditional cross-validation approaches for time series. This decision is based on the principle of parsimony, aimed at building a robust and interpretable initial model, in line with the recommendations in the time series modeling literature [31].

Determining the optimum prediction horizon: The prediction horizon, defined as the time interval over which forecasts are made, is a crucial parameter for the practical application of the model. Too short a horizon limits the usefulness of forecasts for decision-making, while too long a horizon can lead to a degradation in accuracy. In this study, prediction horizons of 1, 12, and 24 time steps are tested, corresponding to 5 min, 1 h, and 2 h, respectively. A preliminary investigation into the thermal response characteristics of the building revealed that heating control equipment typically begins to produce measurable effects within 5 to 10 min after activation. This timescale is well suited to the computational capabilities of standard industrial hardware, where model inference times of milliseconds are easily achievable within 5 min control cycles. Reaches the target temperature in 1 h on average, with the slowest systems taking just under 2 h to fully respond. This finding indicates that a 2 h prediction horizon is sufficient for effective operational decision-making, as it encompasses the full response cycle of the heating systems while maintaining prediction accuracy within acceptable parameters.

Selecting the right number of lags: The number of lags, or past temperature values included as input variables, determines the model's ability to capture time dependencies. An insufficient number of lags can lead to under-modeling. on the opposite, an excessive number can introduce noise and increase the risk of overlearning (overfitting) [32]. During the preliminary investigation mentioned previously, a low significance of the lag effects was observed in this particular building's thermal behavior. The elbow method applied indicated an inflection point between 1 and 5 lags, depending on the algorithms used. Feature selection experiments confirmed that the selected variables effectively captured the essential predictive relationships without requiring complex derived features. However, for neural network models specifically, a lag parameter of 5 was determined to be sufficient to capture any relevant temporal dependencies while maintaining model efficiency and preventing overfitting issues. This conservative choice was made to ensure parameter count compatibility with the available dataset size (124,700 rows post-transformation).

5.2. AI Models Used for Prediction

For temperature prediction, several families of algorithms were compared. Figure 1 provides a mind map illustrating the different families of algorithms used in this study:

XGBoost and LightGBM: Tree-based Gradient Boosting methods. These approaches handle tabular data well and offer robust performance for many regression problems. Additionally, these models provide inherent interpretability through feature importance rankings, enabling identification of the most influential variables for industrial decision-making.

CatBoost: A Gradient Boosting variant adapted to categorical variables, direct use of relay positions, and other discrete sensors in nominal form.

LSTM and GRU: Two recurrent neural network architectures specifically designed for sequence modeling and time series analysis. LSTMs incorporate input, output, and forget gates to control information flow through the network and address the vanishing gradient

problem. GRUs use a simpler architecture with reset and update gates, requiring fewer parameters than LSTM while still effectively capturing long-term dependencies.

CNN1D and CNN2D: Convolutional network approaches. CNN1D considers the time vector as a one-dimensional sequence. The CNN2D, on the other hand, treats the lags and the various variables as an "image" (height = number of lags, width = number of sensors/features), enabling spatio-temporal motifs to be captured across all sensors. The neural network architectures were designed with controlled parameter counts to ensure compatibility with the dataset size. Specifically, the CNN2D model was limited to fewer than 50,000 trainable parameters through constrained filter numbers and fully connected layer sizes. All deep learning models incorporated regularization techniques: dropout layers (rates 0.2–0.5) for CNN architectures and L2 regularization for recurrent models to prevent overfitting.

For neural networks, a training protocol including early stopping on a validation set was used to limit overlearning. Gradient Boosting approaches (XGBoost, LightGBM, CatBoost) were trained on the training part, then validated on the test part.

5.3. Preparation of Lagged Variables

Data was collected and prepared to test the different algorithms. The data came from a workshop equipped with temperature, humidity and heating control sensors (knob positions, relays, gas meter). Measurements were collected chronologically, then assembled into a set of features, including outdoor temperature and humidity.

To enable models to exploit time dependency, *delayed features (lags)* were created. For each variable x_t (e.g., temperature), the corresponding lagged values $x_{t-1}, x_{t-2}, \ldots, x_{t-n}$ were generated, where n corresponds to the number of lags (past values). The objective is to predict, at time t, the value of the temperature at a horizon $t + H$ (where $H \in \{1, 12, 24\}$ in our experiments). All existing columns (temp, humidity, relay positions, etc.) were thus shifted in time, and a dataset with these lagged variables was constructed for training. The first few rows, lacking complete past values, were eliminated so as not to introduce any missing values. Preliminary tests with feature reduction techniques showed no significant performance improvement, indicating that the transformed dataset was appropriately dimensioned for the algorithms evaluated.

The dataset was then split into a *training part* (80% of observations) and a *test part* (20%), respecting the temporal order to ensure realistic evaluation of model performance on unseen future data, including seasonal transitions from winter to spring conditions. In some cases (XGBoost, LightGBM, LSTM, etc.), normalization or *scaling* (MinMax, for example) was applied to facilitate model convergence. For CatBoost, on the other hand, categorical columns (buttons, relays) were kept as is, as CatBoost is capable of handling categorical variables natively [21].

5.4. Hyperparameter Tuning Strategy

In this study, the primary objective consists of conducting a comparative analysis across a wide spectrum of machine learning and deep learning models under realistic deployment constraints. Therefore, a pragmatic and consistent tuning strategy was employed that balanced methodological rigor with computational feasibility across all seven models [33,34].

For the machine learning models (XGBoost, LightGBM, CatBoost), manual tuning informed by domain knowledge was employed, complemented by limited grid search over key hyperparameters. Following established best practices [35,36], the tuning focused on the following:

- Learning rate (0.01 to 0.1);

- Maximum tree depth (3 to 8);
- Number of estimators (100 to 1000);
- Subsampling ratios (0.7 to 1.0).

These parameter ranges were selected based on prior literature recommendations [35,37] and internal validation performance. The tuning protocol aimed to avoid overfitting while preserving computational efficiency, particularly given the recursive prediction scheme and multi-horizon evaluations.

For the deep learning models (LSTM, GRU, CNN1D, CNN2D), a constrained architecture search space was defined to maintain computational tractability [38,39]. Within this space, the following parameters were tuned:

- Number of recurrent units or filters (32 to 128);
- Dropout rate (0.2 to 0.5) [40];
- Number of layers (1 to 2);
- Batch size and learning rate (32 to 128 and 0.001 to 0.01, respectively).

Early stopping on validation loss was employed to determine optimal training epochs, with a patience parameter of 10 epochs and a minimum improvement threshold of 0.001 [39,41]. This approach prevented overfitting while ensuring adequate model convergence across all neural network architectures.

This approach prioritizes consistent methodology across model families rather than exhaustive optimization of individual models, reflecting the trade-offs industrial practitioners face when selecting and deploying predictive algorithms [37]. This strategy helps prevent selection bias that can arise when models receive unequal optimization effort [42], ensuring fair comparison across different algorithmic approaches.

While more sophisticated optimization methods such as Bayesian optimization could potentially enhance individual model performance [34], the chosen approach allowed for meaningful comparison across model families without introducing selection bias or overfitting to validation data, reflecting realistic constraints faced in industrial deployments.

5.5. Recursive Multi-Step Forecasting Strategy

To ensure data leakage prevention and reflect real-world deployment conditions, this study employs a recursive multi-step prediction approach, which is particularly well suited for real-time applications in industrial control. In this paradigm, the model is trained to predict the target variable (indoor temperature) at the next time step, denoted as y_{t+1}, based on a set of lagged input variables up to time t. To forecast further into the future—say, $y_{t+2}, y_{t+3}, \ldots, y_{t+H}$—the model recursively reuses its own previous predictions as inputs.

Formally, if the model f is trained to estimate $y_{t+1} = f(X_t)$, where X_t includes observed variables such as past temperatures, humidity levels, heating states, and external conditions, then for horizon $h > 1$, the prediction y_{t+h} is computed as:

$$y_{t+h} = f(X_{t+h-1}) \tag{6}$$

with some inputs in X_{t+h-1} being themselves predictions from earlier steps.

This recursive scheme introduces error accumulation, since the model's inputs become progressively contaminated with prediction noise. However, it faithfully reflects the operational deployment context, where future measurements are unavailable and decisions must rely on prior forecasts. This approach guarantees the absence of data leakage by construction, as no future information is ever used during prediction.

The observed performance degradation across increasing horizons (Section 6.3) serves as a proxy for each model's capacity to generalize under uncertainty. Notably, Gradient

Boosting models such as XGBoost and LightGBM demonstrate strong resilience, with limited loss in accuracy despite recursive error propagation.

5.6. Implementation and Specific Features of Gradient Boosting Models

The XGBoost, LightGBM, and CatBoost approaches are part of the Gradient Boosting framework, which consists of additively training a succession of decision trees to minimize a cost function. XGBoost focuses on advanced optimization in terms of computation and parallelism, thanks in particular to *sharding* and explicit *sparsity* management. LightGBM adopts a tree construction strategy based on a *leaf-wise* distribution with precise depth control, improving speed and memory requirements. CatBoost features native treatment of categorical variables, thanks to sequential encoding combined with permutation of data order, which limits the overlearning associated with nominal variables with high cardinality.

During training, the construction of the model is often formalized as the minimization of a loss $\mathcal{L}(F)$ by successively adding weak learning functions h_k , according to

$$F_{k+1}(x) = F_k(x) + \nu\, h_k(x), \tag{7}$$

where ν is a learning rate. The three variants (XG-Boost, LightGBM, and CatBoost) differ mainly in the way they construct each h_k and in their handling of regularization.

5.7. Architecture and Training of Recurrent Neural Networks

Recurrent neural networks incorporate internal memory mechanisms to handle possible longer temporal dependencies. The effectiveness of these models becomes apparent as soon as the data show marked sequential patterns, but they remain sensitive to signals essentially dictated by the punctual state of the system and to possible rapid breaks.

While both LSTM and GRU architectures were implemented in this study, it is worth noting their differences in complexity and efficiency. The more compact GRU architecture often proved easier to train with our dataset, requiring fewer computational resources while maintaining comparable prediction accuracy.

Convolutional networks of the CNN1D type treat each time series as a one-dimensional signal and apply filters (kernels) moving along the time axis. This method is ideal when the presence of local repeating patterns is suspected or the extraction of characteristic shapes over successive time windows is desired.

5.8. Convolutional Models: Two-Dimensional Data Processing

The CNN2D extension rearranges the input into a two-dimensional array: the first dimension (height) generally corresponds to the lags $\ell \in 1, \ldots, n$, and the second (width) to the various explanatory variables. The z_{ij} activations formed by the convolution operation are passed to a non-linear function (ReLU or other). Let $M \in \mathbb{R}^{n_\ell \times n_f}$ denote this input matrix, where n_ℓ is the number of delays and n_f the number of features. Each convolution is written as follows:

$$z_{ij} = \sum_{u=0}^{k_1-1} \sum_{v=0}^{k_2-1} ; M_{i+u,j+v}, W_{u,v} + b, \tag{8}$$

where $W \in \mathbb{R}^{k_1 \times k_2}$ is a two-dimensional filter, and b is a bias. The filters move in (i, j) space to extract joint features in the time axis and the feature axis. Once these feature maps have been calculated, a flattening is generally applied, followed by one or more perceptrons for the final regression. This spatial organization attempts to capture the interactions that depend on both the delay ℓ and the nature of the sensor. Although promising for cases with rich temporal patterns or for multivariate data linked by subtle correlations, this model

requires a large dataset for stable training and does not always take advantage of numerous lags if the essential part of the dynamic lies in the momentary state of the sensor [25].

6. Results and Analysis

This section analyzes the performance of the temperature prediction models based on the results of Table 3, which presents the evaluation metrics (MAE, RMSE, MAPE, and R^2) for each model and time horizon, including minimum, average (mean), and maximum values. The objective of the analysis is to assess the ability of the different algorithms to handle various prediction horizons and to compare their accuracy across increasing time intervals.

Table 3 presents a comprehensive performance matrix comparing seven distinct AI models across three critical prediction horizons (1, 12, and 24 h). The table structure allows for systematic comparison using four complementary error metrics, with each metric showing minimum, average, and maximum values across the eight temperature sensors (tempBAS1 to tempBAS8). This multi-dimensional evaluation framework ensures robust model assessment under varied conditions.

Several clear trends emerge from this performance matrix. First, Gradient Boosting methods (particularly LightGBM) consistently outperform neural network architectures across all horizons, with the performance gap widening as the prediction horizon increases. At the 1 h horizon, LightGBM achieves the best average MAE of 0.30 °C compared with the best neural network (GRU) at 0.53 °C. This gap becomes more pronounced at the 24 h horizon, where XGBoost's average MAE of 0.87 °C significantly outperforms the best neural network (CNN1D) at 1.38 °C.

Another notable pattern is the dramatic deterioration of CNN2D performance as prediction horizons extend, with R^2 values dropping from an already modest 0.687 at the 1 h horizon to negative values (−0.393) at 24 h, indicating performance worse than a simple mean prediction. The consistent superiority of tree-based methods suggests that the thermal behavior of the building may be governed by discrete conditional relationships rather than continuous sequential patterns.

These results reflect the building's relatively low thermal inertia, favoring models that excel at capturing short-term patterns and abrupt state changes. The manufacturing workshop's frequent door operations and intermittent heating cycles create thermal conditions that appear better modeled by the decision tree structures of Gradient Boosting algorithms than by the temporal memory mechanisms of recurrent neural networks. This finding contradicts conventional wisdom that often recommends recurrent architectures for time series forecasting, highlighting the importance of matching algorithms to the specific physical characteristics of the system under study.

Table 3. Temperature prediction results (min, moy, max) for tempBAS1 to tempBAS8. Bold values indicate best performance for each metric.

Horizon	Method	MAE °C			RMSE °C			MAPE %			R^2		
		Min	**Moy**	**Max**	**Min**	**Moy**	**Max**	**Min**	**Moy**	**Max**	**Min**	**Moy**	**Max**
1	XGBoost	**0.16**	0.32	0.66	**0.24**	0.50	1.00	**0.83**	1.84	4.10	0.855	0.931	**0.991**
	LightGBM	**0.16**	**0.30**	**0.54**	**0.24**	**0.48**	0.89	0.85	**1.70**	3.37	0.856	**0.937**	**0.991**
	CatBoost	0.19	0.33	0.55	0.28	0.50	**0.87**	1.20	1.81	**3.35**	**0.864**	0.936	0.978
	LSTM	0.31	0.55	0.80	0.43	0.73	1.04	1.78	2.90	4.77	0.771	0.870	0.948
	GRU	0.40	0.53	0.79	0.54	0.69	1.03	2.05	2.87	4.68	0.804	0.879	0.960
	CNN1D	0.44	0.60	0.88	0.57	0.82	1.29	2.22	3.28	4.73	0.534	0.807	0.958
	CNN2D	0.51	0.84	1.52	0.65	1.02	1.76	2.52	4.59	9.11	0.224	0.687	0.945
	XGBoost	0.41	0.66	1.12	0.56	0.90	1.46	2.53	3.38	6.83	0.708	0.819	0.909

Table 3. *Cont.*

Horizon	Method	MAE °C			RMSE °C			MAPE %			R^2		
		Min	**Moy**	**Max**	**Min**	**Moy**	**Max**	**Min**	**Moy**	**Max**	**Min**	**Moy**	**Max**
12	XGBoost	0.41	0.66	1.12	**0.56**	**0.90**	1.46	2.53	3.38	6.83	0.708	0.819	0.909
	LightGBM	**0.40**	**0.65**	1.07	**0.56**	**0.90**	1.41	**2.51**	**3.33**	6.45	**0.730**	**0.821**	**0.910**
	CatBoost	0.42	0.68	**1.01**	0.58	0.92	**1.35**	2.62	3.50	**6.13**	0.700	0.805	0.904
	LSTM	0.51	0.94	1.65	0.66	1.20	2.03	3.17	4.63	9.30	0.439	0.686	0.876
	GRU	0.63	0.93	1.61	0.82	1.16	1.95	3.50	4.75	9.13	0.479	0.681	0.852
	CNN1D	0.65	1.03	1.60	0.85	1.28	1.90	3.79	5.4	9.70	0.087	0.563	0.881
	CNN2D	0.98	1.34	1.91	1.19	1.61	2.25	5.15	6.72	10.69	−0.060	0.381	0.690
24	**XGBoost**	0.61	0.87	**1.26**	0.81	**1.15**	**1.62**	3.47	**4.43**	**7.60**	0.569	**0.702**	0.813
	LightGBM	0.61	**0.90**	1.34	0.85	1.17	1.70	**3.44**	4.50	7.94	**0.593**	0.695	0.800
	CatBoost	**0.60**	0.91	1.28	**0.80**	1.18	**1.62**	3.65	4.52	7.61	0.497	0.688	**0.816**
	LSTM	0.86	1.38	2.08	1.12	1.71	2.52	5.23	6.76	12.03	0.067	0.357	0.597
	GRU	1.02	1.54	2.12	1.32	1.90	2.63	6.19	7.52	12.22	−0.610	0.165	0.469
	CNN1D	0.81	1.38	1.87	1.03	1.73	2.32	5.00	6.74	9.40	0.013	0.326	0.694
	CNN2D	1.38	2.09	2.73	1.67	2.46	3.10	8.18	9.78	15.80	−1.249	−0.393	0.239

6.1. Objectives and Analysis Framework

This section presents an in-depth analysis of the performance of AI models applied to temperature prediction in an industrial environment. The results are evaluated over three prediction time horizons (1, 12, and 24 time steps) using four standard metrics: MAE, MAPE, RMSE, and R^2. The temporal validation approach ensures that models are tested on genuinely unseen future data, including seasonal transitions, reflecting real-world deployment conditions where models must perform reliably across varying climatic conditions. The main objective is to assess the predictive capability of the models under industrial operational conditions while identifying their strengths and limitations. The performances are interpreted taking into account the specificities of the building studied, in particular its thermal configuration, the characteristics of the sensors, and external climatic variations. Particular attention is paid to the degradation of performance as the time horizon increases and to the stability of predictions according to the different families of algorithms.

6.2. Comparison of Overall Model Performance

6.2.1. Gradient Boosting Models

XGBoost, LightGBM, and CatBoost algorithms demonstrate superior performance among all tested models. As indicated in Table 3, these Gradient Boosting models consistently achieve lower error metrics, particularly for short-term (1 time step) and medium-term (12 time steps) prediction horizons. Figure 6 visually confirms this superiority through a color-coded heatmap, where these models occupy the lightest regions (indicating better performance). Each cell presents results in the format: mean [minimum − maximum], providing a comprehensive view of performance distribution.

XGBoost: For short-term predictions (1 time step), XGBoost achieves excellent accuracy with MAE ranging from 0.16 °C to 0.66 °C (mean: 0.32 °C), RMSE from 0.24 °C to 1.00 °C (mean: 0.50 °C), MAPE from 0.83 % to 4.10 % (mean: 1.84 %), and R^2 from 0.855 to 0.991 (mean: 0.931). Performance remains robust at medium-term horizons (12 time steps) with MAE from 0.41 °C to 1.12 °C (mean: 0.66 °C), RMSE from 0.56 °C to 1.46 °C (mean: 0.90 °C), MAPE from 2.53 % to 6.83 % (mean: 3.38 %), and R^2 from 0.708 to 0.909 (mean: 0.819). For long-term predictions (24 time steps), XGBoost maintains acceptable performance with MAE from 0.61 °C to 1.26 °C (mean: 0.87 °C), RMSE from 0.81 °C to 1.62 °C (mean: 1.15 °C), MAPE from 3.47 % to 7.60 % (mean: 4.43 %), and R^2 from 0.569 to 0.813 (mean: 0.702).

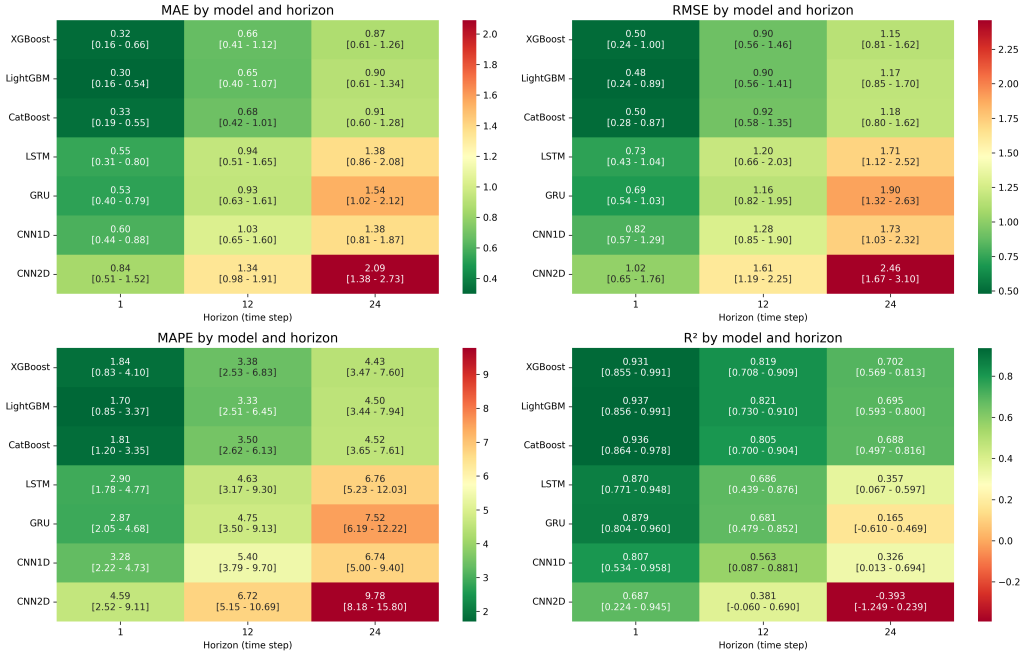

Figure 6. Schematic representation of the centralized building data.

LightGBM: This model performs comparably to XGBoost with short-term MAE from 0.16 °C to 0.54 °C (mean: 0.30 °C), RMSE from 0.24 °C to 0.89 °C (mean: 0.48 °C), MAPE from 0.85 % to 3.37 % (mean: 1.70 %), and R^2 from 0.856 to 0.991 (mean: 0.937). Medium-term metrics show slight deterioration with MAE from 0.40 °C to 1.07 °C (mean: 0.65 °C), RMSE from 0.56 °C to 1.41 °C (mean: 0.90 °C), MAPE from 2.51 % to 6.45 % (mean: 3.33 %), and R^2 from 0.730 to 0.910 (mean: 0.821). Long-term prediction maintains robustness with MAE from 0.61 °C to 1.34 °C (mean: 0.90 °C), RMSE from 0.85 °C to 1.70 °C (mean: 1.17 °C), MAPE from 3.44 % to 7.94 % (mean: 4.50 %), and R^2 from 0.593 to 0.800 (mean: 0.695).

CatBoost: Performance metrics are consistent with other Gradient Boosting models, showing short-term MAE from 0.19 °C to 0.55 °C (mean: 0.33 °C), RMSE from 0.28 °C to 0.87 °C (mean: 0.50 °C), MAPE from 1.20 % to 3.35 % (mean: 1.81 %), and R^2 from 0.864 to 0.978 (mean: 0.936). Medium-term predictions show expected degradation with MAE from 0.42 °C to 1.01 °C (mean: 0.68 °C), RMSE from 0.58 °C to 1.35 °C (mean: 0.92 °C), MAPE from 2.62 % to 6.13 % (mean: 3.50 %), and R^2 from 0.700 to 0.904 (mean: 0.805). Long-term metrics remain competitive with MAE from 0.60 °C to 1.28 °C (mean: 0.91 °C), RMSE from 0.80 °C to 1.62 °C (mean: 1.18 °C), MAPE from 3.65 % to 7.61 % (mean: 4.52 %), and R^2 from 0.497 to 0.816 (mean: 0.688).

6.2.2. Recurrent Neural Models

LSTM and GRU models demonstrate intermediate performance, positioned between Gradient Boosting algorithms and convolutional models. Figure 6 displays these models with middle-range shading, indicating their moderate effectiveness with notable performance degradation as prediction horizons increase.

LSTM: Short-term predictions show moderate accuracy with MAE from 0.31 °C to 0.80 °C (mean: 0.55 °C), RMSE from 0.43 °C to 1.04 °C (mean: 0.73 °C), MAPE from 1.78 % to 4.77 % (mean: 2.90 %), and R^2 from 0.771 to 0.948 (mean: 0.870). Medium-term performance decreases with MAE from 0.51 °C to 1.65 °C (mean: 0.94 °C), RMSE from 0.66 °C to 2.03 °C (mean: 1.20 °C), MAPE from 3.17 % to 9.30 % (mean: 4.63 %), and R^2 from 0.439 to 0.876 (mean: 0.686). Long-term prediction quality deteriorates significantly with MAE from 0.86 °C to 2.08 °C (mean: 1.38 °C), RMSE from 1.12 °C to 2.52 °C (mean: 1.71 °C),

MAPE from 5.23 % to 12.03 % (mean: 6.76 %), and R^2 declining to between 0.067 and 0.597 (mean: 0.357).

GRU: Performance follows similar patterns to LSTM with short-term MAE from 0.40 °C to 0.79 °C (mean: 0.53 °C), RMSE from 0.54 °C to 1.03 °C (mean: 0.69 °C), MAPE from 2.05 % to 4.68 % (mean: 2.87 %), and R^2 from 0.804 to 0.960 (mean: 0.879). Medium-term metrics worsen with MAE from 0.63 °C to 1.61 °C (mean: 0.93 °C), RMSE from 0.82 °C to 1.95 °C (mean: 1.16 °C), MAPE from 3.50 % to 9.13 % (mean: 4.75 %), and R^2 from 0.479 to 0.852 (mean: 0.681). Long-term performance drops substantially with MAE from 1.02 °C to 2.12 °C (mean: 1.54 °C), RMSE from 1.32 °C to 2.63 °C (mean: 1.90 °C), MAPE from 6.19 % to 12.22 % (mean: 7.52 %), and R^2 declining to between −0.610 and 0.469 (mean: 0.165).

6.2.3. Convolutional Models

CNN1D and CNN2D consistently demonstrate the lowest performance across all tested algorithms, particularly for longer prediction horizons. Figure 6 highlights this trend with darker shading in cells corresponding to these models, indicating higher error metrics.

CNN1D: Short-term predictions show MAE from 0.44 °C to 0.88 °C (mean: 0.60 °C), RMSE from 0.57 °C to 1.29 °C (mean: 0.82 °C), MAPE from 2.22 % to 4.73 % (mean: 3.28 %), and R^2 from 0.534 to 0.958 (mean: 0.807). Medium-term performance degrades with MAE from 0.65 °C to 1.60 °C (mean: 1.03 °C), RMSE from 0.85 °C to 1.90 °C (mean: 1.28 °C), MAPE from 3.79 % to 9.70 % (mean: 5.40 %), and R^2 from 0.087 to 0.881 (mean: 0.563). Long-term predictions show further deterioration with MAE from 0.81 °C to 1.87 °C (mean: 1.38 °C), RMSE from 1.03 °C to 2.32 °C (mean: 1.73 °C), MAPE from 5.00 % to 9.40 % (mean: 6.74 %), and R^2 from 0.013 to 0.694 (mean: 0.326).

CNN2D: This model exhibits the poorest overall performance with short-term MAE from 0.51 °C to 1.52 °C (mean: 0.84 °C), RMSE from 0.65 °C to 1.76 °C (mean: 1.02 °C), MAPE from 2.52 % to 9.11 % (mean: 4.59 %), and R^2 from 0.224 to 0.945 (mean: 0.687). Medium-term metrics show pronounced degradation with MAE from 0.98 °C to 1.91 °C (mean: 1.34 °C), RMSE from 1.19 °C to 2.25 °C (mean: 1.61 °C), MAPE from 5.15 % to 10.69 % (mean: 6.72 %), and R^2 from −0.060 to 0.690 (mean: 0.381). Long-term performance collapses further with MAE from 1.38 °C to 2.73 °C (mean: 2.09 °C), RMSE from 1.67 °C to 3.10 °C (mean: 2.46 °C), MAPE from 8.18 % to 15.80 % (mean: 9.78 %), and R^2 from −1.249 to 0.239 (mean: −0.393), indicating performance worse than a simple mean-based prediction model.

6.3. Analysis of Performance Degradation by Horizon

Figure 7 highlights a common phenomenon across all models: performance degradation with the increase in prediction horizon. However, this degradation is not uniform across different algorithm families.

Gradient Boosting models exhibit the lowest degradation, with an increase in mean MAE of approximately 0.57 °C between horizons 1 and 24 for XGBoost (from 0.32 °C to 0.87 °C). This robustness is reflected in the moderately sloped curves in Figure 7, indicating a better ability to maintain accurate predictions over extended horizons.

Recurrent models (LSTM and GRU) show intermediate degradation, with a rise in mean MAE of 0.83 °C for LSTM (from 0.55 °C to 1.38 °C) and 1.01 °C for GRU (from 0.53 °C to 1.54 °C) between horizons 1 and 24. Their curves in Figure 7 display a steeper slope, particularly between horizons 12 and 24.

Convolutional models experience the most significant degradation, with the mean MAE increase reaching 0.78 °C for CNN1D (from 0.60 °C to 1.38 °C) and 1.25 °C for CNN2D (from 0.84 °C to 2.09 °C) between horizons 1 and 24. This substantial decline is especially visible in Figure 7, where the corresponding curves have the steepest slope.

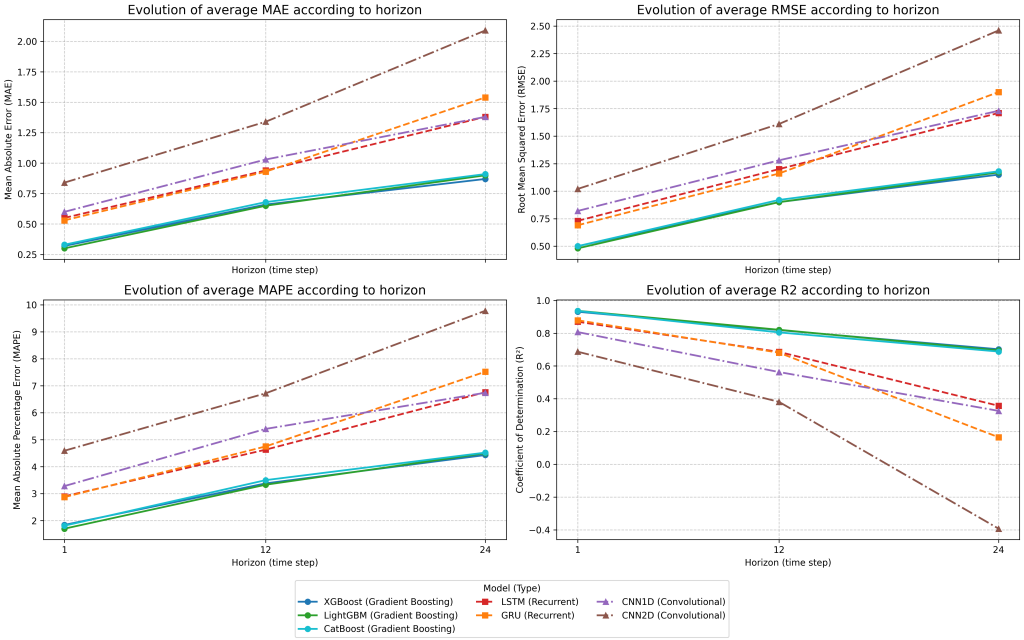

Figure 7. Comparison of metrics by horizon for all models

6.4. Impact of Sensor Location on Prediction Accuracy

An analysis of the prediction results highlights the significant influence of sensor placement within the studied industrial building on the performance of the temperature prediction models. Specifically, sensors tempBAS2 and tempBAS4 exhibit the poorest prediction accuracy. This degradation can be attributed to their proximity to vehicle doors, which are frequently opened. The frequent opening of these doors creates air currents that introduce additional thermal variability, disrupting the ability of the models to accurately predict temperature changes. Consequently, the metrics for these sensors show consistently higher errors compared with others. It is important to note that the temperature sensors used in this study have a precision of 0.50 °C. Therefore, any prediction error below this threshold is considered reliable, as it remains within the bounds of sensor accuracy. This consideration helps contextualize the observed model performance, especially in distinguishing between meaningful prediction discrepancies and those that fall within the expected measurement uncertainty. In contrast, sensors tempBAS3, tempBAS6, and tempBAS7 demonstrate excellent prediction accuracy. These sensors are located at the center of the building, far from the disruptive effects of the vehicle doors. Furthermore, their central positions are surrounded by other hot air generator units, which contribute to more stable and homogeneous thermal conditions. This configuration minimizes external disturbances and enhances the precision of the models. Lastly, sensor tempBAS8 also achieves above-average prediction accuracy, which is likely due to its placement within a small, enclosed room. The closed environment reduces the influence of external climatic variations and air currents, resulting in more predictable thermal behavior and improved model performance. This analysis underscores the importance of sensor location in industrial temperature prediction tasks. The proximity to external disturbances, such as open doors, can significantly degrade prediction performance, whereas central and enclosed locations provide favorable conditions for achieving higher accuracy.

Feature importance analysis conducted on the Gradient Boosting models confirms these observations, consistently ranking external temperature (tempEXT) as the most influential variable, followed by heating relay states (comRELAIS) and sensor location factors. This analysis validates that the models appropriately capture the physical relationships governing the building's thermal behavior.

6.5. Superiority of Traditional Models for Extreme Thermal Characteristics in Industrial Buildings

The building in question is characterized by two extreme thermal features: almost no thermal inertia and an overpowered heating system with forced-air diffusion. These factors create an atypical thermal behavior with rapid and large fluctuations, resulting in a highly responsive temperature profile.

An in-depth analysis highlights a counterintuitive phenomenon: the most recent and sophisticated learning models, particularly complex neural network architectures (LSTM, GRU, CNN1D, and CNN2D), perform significantly suboptimally compared with traditional Gradient Boosting algorithms. This observation warrants a detailed analysis in the context of the unique thermal characteristics of the studied industrial building.

While deep neural networks are theoretically designed to capture complex temporal dependencies and subtle spatial patterns, they paradoxically prove less suitable in this context. Several factors may explain this.

Overfitting to Slow Dynamics: LSTM and GRU networks, designed to model long-term dependencies, tend to overparameterize temporal relationships in an environment with nearly instantaneous changes. The mean MAE increases from 0.55 °C to 1.38 °C for LSTM and from 0.53 °C to 1.54 °C for GRU between horizons 1 and 24, a significantly steeper degradation compared with Gradient Boosting models (from 0.32 °C to 0.87 °C for XGBoost).

Inadequacy of Spatial Convolutions: CNN2D models, particularly ineffective with a catastrophic mean R^2 of -0.393 at the 24-time-step horizon, attempt to identify spatial thermal propagation patterns in an environment where forced ventilation rapidly homogenizes the temperature, making such patterns non-existent or irrelevant. The poor performance of CNN2D models cannot be attributed to inadequate training, as rigorous safeguards including early stopping, regularization, and controlled parameter counts were implemented. Instead, the fundamental mismatch between CNN spatial assumptions and the building's instantaneous thermal response characteristics explains this performance degradation.

Excessive Complexity for Simple Relationships: The building's thermal profile, dominated by heating power and the absence of inertia, presents almost linear relationships between input variables (heating power, external temperature) and the resulting temperature. Gradient Boosting models, using decision tree ensembles, efficiently capture these straightforward relationships without the unnecessary complexity introduced by deep neural networks.

An analysis of Figure 6 reveals that convolutional models, particularly CNN2D, occupy the darkest regions of the heatmap for all metrics, confirming their relative inadequacy for this industrial thermal prediction task. The analysis of the various metrics presented in Figure 6 provides a deeper understanding of model performance.

The MAE and RMSE show similar trends, with consistently higher values for RMSE due to its sensitivity to large errors. These two metrics confirm the superiority of Gradient Boosting models, followed by recurrent models and then convolutional models.

The MAPE offers a complementary perspective by normalizing errors relative to observed values. This metric reveals that even the most accurate models (XGBoost and LightGBM) exhibit relative errors reaching mean values of 4.43 % and 4.50 %, respectively, for the 24-time-step horizon. In contrast, the CNN2D reaches a mean MAPE of 9.78 %, confirming its low accuracy.

The R^2 assesses the models' ability to explain data variance. Gradient Boosting models maintain a mean R^2 above 0.688 even at the 24-time-step horizon, whereas recurrent models see their mean R^2 drop below 0.357 at this horizon. The CNN2D shows a negative mean R^2 at the long horizon, indicating its inability to capture long-term thermal trends.

6.6. Generalizability and Applicability Framework

The findings of this study are specifically tied to the thermal characteristics of the investigated industrial building, which exhibits two extreme features: minimal thermal inertia and an overpowered forced-air heating system. However, the methodological approach and key insights can be extended to other industrial facilities sharing similar thermal dynamics.

Applicability Criteria: The superiority of Gradient Boosting algorithms over deep learning approaches is expected to generalize to industrial buildings characterized by: low thermal inertia due to minimal insulation and high ceiling heights, forced-air heating systems with rapid thermal response, frequent thermal disruptions from operational activities (door openings, equipment cycling), and spatially distributed heating sources creating localized thermal zones rather than uniform temperature fields.

Conversely, buildings with high thermal mass, radiant heating systems, or strong spatial thermal gradients may benefit more from deep learning approaches, particularly CNN architectures designed to capture spatial relationships. Similarly, facilities with complex temporal thermal patterns or significant thermal lag effects might favor recurrent neural network architectures.

Validation Considerations: The proposed algorithmic selection framework requires validation across diverse industrial contexts. Key factors for assessment include: building envelope characteristics (insulation, thermal mass), HVAC system type (forced-air, radiant, mixed), operational patterns (continuous vs. intermittent), and external disturbance frequency. A preliminary assessment based on these criteria can guide initial algorithm selection, followed by empirical validation using the comparative methodology demonstrated in this study.

Sensor Placement Implications: The significant impact of sensor location on prediction accuracy (Section 6.4) suggests that the generalizability framework must also consider facility layout and sensor positioning strategies. Industrial buildings with similar thermal characteristics but different spatial configurations may require adapted sensor networks to achieve comparable prediction performance.

Ultimately, the criteria outlined above define the scope for the direct transfer of our results. However, it is essential to distinguish the limited generalizability of these specific findings from the broad applicability of our comparative methodology. The systematic protocol for data preparation, model training, and multi-horizon evaluation demonstrated in this study provides a robust and replicable framework. This framework can guide practitioners in selecting the optimal predictive algorithm for their own facilities, even those with thermal characteristics that differ significantly from our case study.

7. Conclusions

This comparative study rigorously evaluated the performance of various artificial intelligence algorithms for temperature prediction within an industrial building, with a clear objective: to optimize thermal management.

The key findings of this comparative analysis demonstrate the superior performance of Gradient Boosting algorithms, notably XGBoost and LightGBM, for temperature prediction in the studied industrial building. These models consistently exhibited the lowest prediction errors, as measured by the RMSE, the MAE, and the MAPE, when compared with the other evaluated algorithms, including LSTM and convolutional neural networks (CNNs). Furthermore, they achieved higher R^2, indicating a better ability to explain the variance in the temperature data. While the precise values of these metrics are detailed in the results sections, the overarching trend highlights the enhanced accuracy and reliability

of the predictions provided by XGBoost and LightGBM for short- and mid-term thermal management within this specific industrial context.

This study contributes several novel insights: systematic comparison of seven AI algorithms on real industrial data, empirical demonstration of Gradient Boosting superiority over deep learning in low thermal inertia environments, quantitative analysis of sensor placement impact on prediction accuracy, and methodological framework for algorithm selection based on building thermal characteristics.

The hyperparameter tuning strategy employed in this study demonstrates that pragmatic, consistent approaches can provide meaningful model comparisons while respecting computational constraints typical of industrial settings. By prioritizing methodological consistency over exhaustive optimization, the study avoided the selection bias that can occur when models receive unequal tuning effort, ensuring that our comparative results reflect genuine algorithmic differences rather than optimization artifacts.

These results underscore the crucial importance of carefully selecting models based on the specific thermal characteristics of the building. In this instance, Gradient Boosting approaches proved particularly effective at capturing the immediate causal relationships that define the industrial environment's thermal dynamics. This finding challenges the common assumption that deep learning architectures are systematically superior for predictive tasks, demonstrating instead the critical importance of architectural alignment with physical system properties.

The superiority of traditional models in this context can be attributed to several factors.

Simplicity of Thermal Relationships: The studied industrial building exhibits a thermal dynamic primarily driven by immediate factors such as heating power and external temperature, with minimal influence from thermal inertia. This characteristic favors models capable of efficiently capturing direct and straightforward relationships between input variables and internal temperature. The building's thermal characteristics favor direct variable relationships over complex derived features, as confirmed by preliminary feature selection tests that showed no performance improvement with reduced feature sets.

Robustness to Rapid Variations: Gradient Boosting models demonstrated superior adaptability to the rapid temperature fluctuations characteristic of this industrial environment, unlike neural network models, which tend to overfit or misinterpret these variations.

Computational Efficiency: In an industrial setting using standard multi-service infrastructure, traditional models offer significant advantages with inference times typically under 10 ms, well within the requirements for thermal control systems operating on 5 min cycles. This computational efficiency makes them particularly suitable for real-time industrial deployment.

Interpretability: Gradient Boosting models provide better interpretability of results through native feature importance analysis, which consistently identified external temperature, heating relay states, and sensor positioning as the primary predictive factors. This transparency is crucial in industrial environments where understanding the factors influencing predictions is essential for decision-making and process optimization.

Broader Implications and Future Research: While this study focuses on a specific industrial building with extreme thermal characteristics, the methodological framework provides a foundation for algorithm selection in broader industrial contexts. The key insight—that model architecture must align with physical system properties—extends beyond the specific case study to inform predictive modeling strategies across diverse industrial facilities.

Future research should investigate the proposed applicability framework across industrial buildings with varying thermal characteristics, HVAC configurations, and operational patterns. Particular attention should be given to developing automated assessment tools that can rapidly classify building thermal dynamics and recommend appropriate algorithmic approaches based on easily measurable facility characteristics.

Future research could also employ advanced interpretability techniques, such as SHAP (SHapley Additive exPlanations) analysis, to gain even finer-grained insights into individual prediction drivers.

The integration of these predictive models within broader Industry 4.0 ecosystems represents another avenue for investigation, particularly regarding real-time model adaptation and multi-building optimization strategies. Such developments could enable the scalable deployment of intelligent thermal management systems across large industrial portfolios.

Moreover, these results have significant implications for the integration of artificial intelligence within the context of Industry 4.0 and energy efficiency. The ability to accurately predict temperature enables proactive thermal management, potentially reducing energy consumption and ensuring optimal conditions for production and product quality. In an increasingly regulatory and societal landscape focused on reducing carbon footprint (as exemplified by the ELAN law in France), the adoption of such technologies represents a significant lever.

On a practical level, this study suggests that industrial companies should seriously consider the implementation of Gradient Boosting algorithms for temperature prediction within their infrastructures. The integration of these models with existing BACS could lead to a significant optimization of HVAC systems, yielding concrete economic and environmental benefits. The results obtained underscore the importance of a data-driven approach to improve energy efficiency in the industrial sector.

This research contributes to the understanding of the potential of AI for industrial thermal management, highlighting the effectiveness of Gradient Boosting algorithms and paving the way for practical applications for a more sustainable and efficient industry.

Author Contributions: Conceptualization, J.R. and T.D.; methodology, J.R., Z.L., and T.D.; software, P.Y.; validation, T.D.; investigation, J.R.; data curation, J.R. and P.Y.; writing—original draft preparation, J.R.; writing—review and editing, L.D., T.D., and Z.L.; visualization, J.R.; supervision, T.D. and L.D.; project administration, Z.L. and P.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable. This study did not involve humans or animals.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study is not publicly available due to privacy restrictions and industrial confidentiality agreements with the partner company DECIMA. However, data may be made available from the corresponding author upon reasonable request and appropriate confidentiality agreements.

Acknowledgments: Thank you to Centrale Lille and DECIMA for providing the experimental platform.

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

Abbreviations

ML	Machine Learning
DL	Deep Learning
MPC	Model Predictive Control
HVAC	Heating, Ventilation and Air-Conditioning
ELAN	Évolution du Logement, de l'Aménagement et du Numérique
MAPE	Mean Absolute Percentage Error
MAE	Mean Absolute Error
RMSE	Root Mean Squared Error
RF	Random Forest
BACS	Building Automation and Control Systems
AI	Artificial Intelligence
LSTM	Long Short-Term Memory
SVM	Support Vector Machine
LHV	Lower Heating Value
PPD	Predicted Percentage of Dissatisfied
CNN	convolutional neural network
R^2	Coefficient of determination
GRU	Gated Recurrent Unit
PMV	Predicted Mean Vote

References

1. Boutahri, Y.; Tilioua, A. Machine learning-based predictive model for thermal comfort and energy optimization in smart buildings. *Results Eng.* **2024**, *22*, 102148. [CrossRef]
2. Chen, X.; Wang, Q.; Srebric, J. Model predictive control for indoor thermal comfort and energy optimization using occupant feedback. *Energy Build.* **2015**, *102*, 357–369. [CrossRef]
3. Loi n° 2018-1021 du 23 novembre 2018 portant évolution du logement, de l'aménagement et du numérique. *Journal Officiel de la République Française* **2018**, 24 novembre. Available online: https://www.legifrance.gouv.fr/loda/id/JORFTEXT000037639478 (accessed on 17 November 2024).
4. Décret n° 2019-771 du 23 juillet 2019 relatif aux obligations d'actions de réduction de la consommation d'énergie finale dans des bâtiments à usage tertiaire. *Journal Officiel de la République Française* **2019**, 25 juillet, texte n° 53. Available online: https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000038812251 (accessed on 17 November 2024).
5. Décret n° 2020-887 du 20 juillet 2020 relatif au système d'automatisation et de contrôle des bâtiments non résidentiels et à la régulation automatique de la chaleur. *Journal Officiel de la République Française* **2020**, 21 juillet, texte n° 38. Available online: https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000042128488 (accessed on 17 November 2024).
6. International Organization for Standardization. *ISO 7730:2005*; Ergonomics of the thermal environment–Analytical determination and interpretation of thermal comfort using calculation of the PMV and PPD indices and local thermal comfort criteria. International Standard: Geneva, Switzerland, 2005. Available online: https://www.iso.org/standard/39155.html (accessed on 17 November 2024).
7. American Society of Heating, Refrigerating and Air-Conditioning Engineers. *ANSI/ASHRAE 55-2023*; Thermal Environmental Conditions for Human Occupancy. ANSI/ASHRAE Standard: Atlanta, GA, USA, 2023. Available online: https://www.ashrae.org/technical-resources/bookstore/standard-55-thermal-environmental-conditions-for-human-occupancy (accessed on 17 November 2024).
8. Bienvenido-Huertas, D.; Sánchez-García, D.; Tejedor, B.; Rubio-Bellido, C. Energy savings in buildings applying ASHRAE 55 and regional adaptive thermal comfort models. *Urban Clim.* **2024**, *55*, 101892. [CrossRef]
9. Moujalled, B.; Cantin, R.; Guarracino, G. Comparison of thermal comfort algorithms in naturally ventilated office buildings. *Energy Build.* **2008**, *40*, 2215–2223. [CrossRef]
10. Fard, Z.Q.; Zomorodian, Z.S.; Korsavi, S.S. Application of machine learning in thermal comfort studies: A review of methods, performance and challenges. *Energy Build.* **2022**, *256*, 111771. [CrossRef]
11. Ahmad, T.; Chen, H. A review on machine learning forecasting growth trends and their real-time applications in different energy systems. *Sustain. Cities Soc.* **2020**, *54*, 102010. [CrossRef]
12. Viot, H. Modélisation et Instrumentation d'un Bâtiment et de ses Systèmes Pour Optimiser sa Gestion Énergétique. Ph.D. Thesis, Université de Bordeaux, Bordeaux, France, 2016. Available online: https://theses.hal.science/tel-01503037 (accessed on 17 November 2024).

13. Moroşan, P.D.; Bourdais, R.; Dumur, D.; Buisson, J. Building temperature regulation using a distributed model predictive control. *Energy Build.* **2010**, *42*, 1445–1452. [CrossRef]
14. Freire, R.Z.; Oliveira, G.H.; Mendes, N. Predictive controllers for thermal comfort optimization and energy savings. *Energy Build.* **2008**, *40*, 1353–1365. [CrossRef]
15. Bacha, S.; Belhadji, L.; Missaoui, R.; Ploix, S. Validation of building energy management strategy: Application to home thermal zone. In Proceedings of the 4th International Conference on Power Engineering, Energy and Electrical Drives, Istanbul, Turkey, 13–17 May 2013; IEEE: Piscataway, NJ, USA, 2013; pp. 921–926. [CrossRef]
16. Fraisse, G.; Viardot, C.; Lafabrie, O.; Achard, G. Development of a simplified and accurate building model based on electrical analogy. *Energy Build.* **2002**, *34*, 1017–1031. [CrossRef]
17. Chen, T. Application of adaptive predictive control to a floor heating system with a large thermal lag. *Energy Build.* **2002**, *34*, 45–51. [CrossRef]
18. Prívara, S.; Široký, J.; Ferkl, L.; Cigler, J. Model predictive control of a building heating system: The first experience. *Energy Build.* **2011**, *43*, 564–572. [CrossRef]
19. Fanger, P.O.; Toftum, J. Extension of the PMV model to non-air-conditioned buildings in warm climates. *Energy Build.* **2002**, *34*, 533–536. [CrossRef]
20. Kapp, S.; Choi, J.K.; Hong, T. Predicting industrial building energy consumption with statistical and machine-learning models informed by physical system parameters. *Renew. Sustain. Energy Rev.* **2023**, *172*, 113045. [CrossRef]
21. Muraina, I.O. Ideal Dataset Splitting Ratios in Machine Learning Algorithms: General Concerns for Data Scientists and Data Analysts. In *Proceedings of the 7th International Mardin Artuklu Scientific Research Conference*; IKSAD Institute: Ankara, Turkey, 2022; pp. 496–515. Available online: https://www.researchgate.net/publication/358284895 (accessed on 17 November 2024).
22. Li, X.; Han, Z.; Zhao, T.; Zhang, J.; Xue, D. Modeling for indoor temperature prediction based on time-delay and Elman neural network in air conditioning system. *J. Build. Eng.* **2021**, *33*, 101854. [CrossRef]
23. Yin, H.; Wu, Z.; Wu, J.C.; Chen, Y.; Chen, M.; Luo, S.; Gao, L.; Hassan, S.G. A Multistep Interval Prediction Method Combining Environmental Variables and Attention Mechanism for Egg Production Rate. *Agriculture* **2023**, *13*, 1255. [CrossRef]
24. Li, D.; Chen, D.; Shi, L.; Jin, B.; Goh, J.; Ng, S.K. MAD-GAN: Multivariate Anomaly Detection for Time Series Data with Generative Adversarial Networks. In *Artificial Neural Networks and Machine Learning—ICANN 2019: Text and Time Series Munich, Germany, 17–19 September 2019*; Springer: Cham, Switzerland, 2019.
25. Sulzer, M.; Christen, A.; Matzarakis, A. Predicting indoor air temperature and thermal comfort in occupational settings using weather forecasts, indoor sensors, and artificial neural networks. *Build. Environ.* **2023**, *234*, 110077. [CrossRef]
26. Han, J.; Shu, K.; Wang, Z. Predicting energy use in construction using Extreme Gradient Boosting. *PeerJ Comput. Sci.* **2023**, *9*, e1500. [CrossRef]
27. Narocki, C. *Heatwaves as an Occupational Hazard*; The European Trade Union Institute: Brussels, Belgium, 2021; pp. 1–71.
28. Core Writing Team; Lee, H.; Romero, J. (Eds.) *IPCC, 2023: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*; Technical report, Intergovernmental Panel on Climate Change; IPCC: Geneva, Switzerland, 2023. [CrossRef]
29. GRDF (Gaz Réseau Distribution France). Coefficient de Conversion Gaz : Comment le Calculer? Available online: https://www.grdf.fr/particuliers/logement-gaz/premiers-pas/coefficient-conversion-commune (accessed on 17 November 2024).
30. Lavrač, N.; Škrlj, B.; Robnik-Šikonja, M. Propositionalization and embeddings: Two sides of the same coin. *Mach. Learn.* **2020**, *109*, 1465–1507. [CrossRef]
31. Wulff, S.S. Time Series Analysis: Forecasting and Control, 5th edition. *J. Qual. Technol.* **2017**, *49*, 418–419. [CrossRef]
32. Shumway, R.H.; Stoffer, D.S. *Time Series Analysis and Its Applications*; Springer Nature: Cham, Switzerland, 2017. [CrossRef]
33. Bergstra, J.; Bengio, Y. Random search for hyper-parameter optimization. *J. Mach. Learn. Res.* **2012**, *13*, 281–305.
34. Feurer, M.; Hutter, F. Hyperparameter optimization. In *Automated Machine Learning*; Springer: Cham, Switzerland, 2019; pp. 3–33.
35. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
36. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.Y. Lightgbm: A highly efficient gradient boosting decision tree. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Volume 30.
37. Probst, P.; Boulesteix, A.L.; Bischl, B. Tunability: Importance of hyperparameters of machine learning algorithms. *J. Mach. Learn. Res.* **2019**, *20*, 1934–1965.
38. Melis, G.; Dyer, C.; Blunsom, P. On the state of the art of evaluation in neural language models. *arXiv* **2018**, arXiv:1707.06779.
39. Reimers, N.; Gurevych, I. Optimal hyperparameters for deep LSTM-networks for sequence labeling tasks. *arXiv* **2017**, arXiv:1707.06799.
40. Gal, Y.; Ghahramani, Z. A theoretically grounded application of dropout in recurrent neural networks. In Proceedings of the Advances in Neural Information Processing Systems, Barcelona, Spain, 5–10 December 2016; pp. 1019–1027.

41. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
42. Cawley, G.C.; Talbot, N.L.C. On over-fitting in model selection and subsequent selection bias in performance evaluation. *J. Mach. Learn. Res.* **2010**, *11*, 2079–2107.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.