Politecnico di Torino

Master's degree Program in

Digital Skills for Sustainable Societal Transitions

Master's Degree Thesis

Data-driven Machine Learning and 3D Visualization for Forecasting Indoor Environmental Conditions

A Case Study on Campus Building (Aule R, Polito)

Supervisor:
Anna Osello

Co-Supervisors:
Enrico Macii
Matteo Del Giudice

Candidate:
Sajedeh Jalalnejad

A.Y. 2024/2025

Table of Contents

Acronyms

Acronym	Full Term
AI	Artificial Intelligence
ANN	Artificial Neural Network
ARIMA	AutoRegressive Integrated Moving Average
ASHRAE	American Society of Heating, Refrigerating, and Air-Conditioning Engineers
BIM	Building Information Modeling
BMS	Building Management System
CAD	Computer-Aided Design
CNN	Convolutional Neural Network
CO_2	Carbon Dioxide
CSV	Comma-Separated Values
DAX	Data Analysis Expressions
DL	Deep Learning
DST	Daylight Saving Time
EDA	Exploratory Data Analysis
EN	European Norm
ETL	Extract, Transform, Load
HVAC	Heating, Ventilation, and Air Conditioning
IAQ	Indoor Air Quality
IEQ	Indoor Environmental Quality
IoT	Internet of Things
KPI	Key Performance Indicator
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MC Dropout	Monte Carlo Dropout
ML	Machine Learning
MLR	Multiple Linear Regression
MSE	Mean Squared Error
PMI	Partial Mutual Information
RF	Random Forest
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
R^2	Coefficient of Determination
sMAPE	Symmetric Mean Absolute Percentage Error
SHAP	SHapley Additive exPlanations
UTC	Coordinated Universal Time
CET	Central European Time

CEST	Central European Summer Time
Speckle	Speckle Interoperability Cloud Platform
XAI	Explainable Artificial Intelligence

Acknowledgment

I would like to express my sincere gratitude to my supervisors, Professors. Anna Osello, Enrico Macii, and Matteo Del Giudice for their invaluable guidance and constructive feedback throughout this research. Their support and expertise were fundamental to the completion of this thesis.

My appreciation also goes to the faculty members of the Inter-University Department of Regional and Urban Studies and Planning (DIST) and the Department of Structural, Geotechnical and Building Engineering (DISEG) at Politecnico di Torino for providing an inspiring academic environment and interdisciplinary perspective.

I would like to thank the Polito Monitoring Team for granting access to the environmental sensor data of Aule R and for their technical assistance.

Finally, I express my heartfelt gratitude to my family for their unconditional love, support, and constant encouragement throughout my academic journey. This achievement would not have been possible without them. They are my greatest motivation.

Completing this work has been both challenging and rewarding, and I am humbled by the collective effort that has made it possible. I sincerely hope that this research contributes to promoting more sustainable, data-driven, and comfortable indoor environments for future generations.

Sajedeh Jalalnejad

Abstract

Indoor environmental conditions, particularly air temperature and CO_2 concentration, play a central role in shaping comfort, health, and energy performance in academic buildings. Despite the widespread deployment of IoT-based monitoring systems, most HVAC systems still operate reactively. This thesis addresses this gap by developing an explainable, data-driven forecasting framework that predicts short-term indoor environmental trends and integrates these predictions into a 3D BIM-based visualization environment for improved decision support. Using approximately 20 months of sensor data from the Aule R building at Politecnico di Torino, the study evaluates the performance of both machine-learning and deep-learning methods, including Random Forest, CNN, and LSTM for modeling temperature and CO_2 dynamics.

The methodology includes a pipeline of data cleaning, temporal alignment, feature engineering, multivariate time-series forecasting, and uncertainty estimation. The framework connects predictive outputs to an innovative dashboard, enabling spatial representation of forecasted indoor conditions within the building's 3D model.

The expected outcome is that machine-learning models can provide reliable short-horizon temperature forecasts by leveraging the temporal structure inherent in indoor sensor data. Furthermore, explainability tools help clarify the role of key variables, while the BIM-enabled data visualization aims to offer an intuitive platform for exploring environmental trends and supporting proactive HVAC-related decision-making.

Chapter 1 – Introduction

1.1 Problem Statement and Background

Indoor environmental conditions, such as air temperature, humidity, and CO_2 concentration, directly influence thermal comfort, health, and cognitive performance of building occupants. These aspects are particularly relevant in educational environments where classroom conditions change rapidly due to fluctuating occupancy, intermittent ventilation, and variable outdoor weather[1]. Although many university buildings are equipped with HVAC systems and dense sensor networks, the control logic of these systems is still predominantly reactive; that is, they intervene only after discomfort conditions or ventilation inadequacies have already occurred [2, 3]. A predictive approach requires short- to medium-term forecasts of indoor temperature and CO_2 levels so that building operations can be proactively tuned to mitigate discomfort and ventilation inadequacy.

For building data, which is inherently multivariate, nonlinear, and time-dependent, data-driven machine learning (ML) techniques, especially deep learning architectures such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) models, tend to outperform traditional linear approaches like ARIMA [4, 5, 6, 7]. For short forecasting horizons, CNNs tend to perform well because they capture quick, high-frequency changes in the data. LSTMs, on the other hand, are generally better at handling information that depends on patterns stretched over longer periods [6, 8, 9]. Furthermore, integrating exogenous variables such as outdoor weather conditions and calendar indicators has been shown to enhance predictive precision in academic research [5, 10]. Concurrently, progress in Explainable AI (XAI), particularly using SHAP (SHapley Additive exPlanations), provides a mechanism for transparently interpreting model outputs by quantifying the contribution of specific input features (e.g., lagged temperature/CO_2 values, outdoor air temperature, and time-of-day) to the final forecasts [11, 12].

Case Study: Aule R Building

This research is conducted as a real-world case study on Aule R, an academic building at Politecnico di Torino that is equipped with a long-term IoT monitoring system. The building includes eight instructional spaces (R1–R4 and R1B–R4B), where sensors continuously measure indoor air temperature, CO_2 concentration, HVAC setpoints, and outdoor temperature. The dataset used in this thesis comprises approximately 20 months of 15-minute environmental measurements, providing a high-resolution, multivariate time series suitable for machine-learning forecasting.

In addition to the sensor infrastructure, Aule R already has a detailed Building Information Model (BIM) and an interactive Power BI dashboard developed during a previous internship, where historical environmental data were linked to the 3D model through the Speckle platform. That system focused on

descriptive visualization of past measurements; in this thesis, it is extended by integrating short-term forecasts of indoor temperature and CO_2 into the BIM-Power BI environment.

Aule R provides an ideal case study for this research because the availability of long-term, high-frequency sensor data enables robust forecasting, and its BIM model allows spatial visualization of predicted environmental conditions, also the existing dashboard infrastructure makes it possible to evolve from descriptive monitoring toward a predictive decision-support tool.

By grounding the analytical framework in a real building with rich data and existing visualization pathways, this case study demonstrates how forecasting, explainability, and 3D representation can be integrated into a unified, operationally meaningful workflow.

1.2 Research Questions

This thesis aims to answer the following key research questions:

A. **Predictive Capability:**
   How accurately can short-term indoor variables, especially temperature and CO_2 concentration, be predicted for the Aule R building using historical indoor sensor data combined with outdoor weather information?

B. **Model Comparison:**
   To what extent do the performance characteristics of Random Forest, CNN, and LSTM models differ across short-term (1-hour) and medium-term (3-hour) forecasting horizons, and which algorithm offers the most consistent balance between accuracy, robustness, and computational cost?

C. **Interpretability and Feature Influence:**
   How can SHAP-based interpretability help identify the most influential temporal (lag/rolling) features contributing to the predictions produced by the best-performing models?

D. **Decision-Support Visualization:**
   How can forecast outputs be integrated within the BIM-linked Power BI dashboard to provide intuitive, spatially contextualized visual analytics for building operators and facility managers?

1.3 Objectives

The main objective of this thesis is to design and validate an explainable, data-driven forecasting framework capable of predicting indoor environmental conditions in Aule R and supporting HVAC-related decision-making. The study focuses not only on improving prediction accuracy but also on understanding model behavior and enhancing its practical applicability through visualization.

Specific objectives are:

- **Data Acquisition and Preprocessing:**

Collect and synchronize 20 months of 15-minute indoor sensor data (temperature and CO_2) with outdoor temperature and HVAC set point information, producing a reliable multivariate time-series dataset.

- **Feature Engineering and Selection:**
  Create a customized set of temporal (lag and rolling) and contextual (calendar) features, and use statistical feature-selection methods to eliminate redundancy while maintaining predictive power.
- **Predictive Modeling:**
  Implement and compare Random Forest, CNN, and LSTM models for 1-hour and 3-hour forecasting horizons, analyzing their accuracy, strengths, and limitations within the context of Aule R's operational characteristics.
- **Interpretability Analysis:**
  Use SHAP values to measure how individual features affect predictions, enabling a more precise explanation of the model's behavior.
- **Integration and Visualization:**
  Integrate the predictive results and uncertainty estimates into the existing BIM-Power BI dashboard to enable spatially contextual 3D visualization and enhance decision-support capabilities.

Chapter 2 - Literature Review

2.1 Scope and Structure

This chapter brings together the core ideas and studies that inform the approach taken in this thesis. It begins with a brief look at the indoor environmental quality (IEQ) standards that apply to educational buildings, explaining how thermal comfort and indoor air-quality requirements are typically defined in classrooms and why these criteria matter in practice. The chapter then turns to studies that use data-driven or machine-learning approaches to forecast indoor temperature and CO_2 levels, highlighting how such predictions can support comfort management.

In the following sections, the chapter reviews the deep-learning models that are most frequently used for multivariate time-series data—particularly CNNs and LSTMs—and discusses the kinds of indoor-environment patterns they are able to capture. Because the usefulness of predictive models depends not only on accuracy but also on how well their results can be understood, the chapter also introduces SHAP as a method for interpreting model behaviour and explaining the role of different input features.

The final part of the chapter shifts from modelling to visualization. Here, the focus is on how predictive results can be integrated into BIM-based dashboards, allowing forecasted conditions to be explored directly within a 3D visualization of the building. This combination of prediction, explanation, and spatial visualization forms the conceptual basis for the framework developed in the thesis.

2.2 IEQ Frameworks and Standards

Indoor environmental quality (IEQ) includes thermal, air quality, visual, and acoustic dimensions. When it comes to classroom comfort and indoor air quality, ASHRAE 55 and EN 16798-1 explain what counts as acceptable conditions. They set recommended temperature limits and use CO_2 levels as a simple indicator of whether a room is being ventilated properly [13, 14, 15]. In many studies, CO_2 concentrations near 1000 ppm are taken as a signal that a room is not being ventilated enough. It is worth noting, though, that CO_2 mainly reflects ventilation effectiveness and should not be interpreted as a direct health metric [16]. These standards contextualize the forecasting task: anticipating temperature and CO_2 helps maintain compliance and occupant comfort under variable occupancy.

Key insight: Educational buildings exhibit rapid fluctuations in IEQ; static HVAC schedules are typically inadequate, motivating the use of predictive analytics to anticipate and prevent excursions [2,17].

2.3 Data-Driven Forecasting of Indoor Temperature and CO_2

Models like ARIMA rely on linear relationships and usually handle only one variable at a time, so they tend to perform poorly when dealing with the multivariate and nonlinear dynamics that characterize indoor environmental data [7]. In contrast, Random Forest (RF) and other tree-based methods handle nonlinearities and mixed features (lags, rolling statistics, calendar/weather) effectively, often serving as

robust baselines with good interpretability [4, 18]. However, tree models require explicit temporal features to represent memory.

Deep learning learns temporal structure directly from sequences. Comparative studies of campus buildings show that CNNs often outperform LSTMs for short horizons by capturing localized temporal patterns, whereas LSTMs excel at longer horizons due to their gated memory [6, 8, 9]. For indoor temperature, Elmaz et al. (2021) reported CNN-LSTM and CNN dominance at 15-minute resolution [8]. Minassian et al. (2025) ran their tests on several rooms, trying out CNN, LSTM, and a combined CNN–LSTM model. What they noticed was pretty simple: the results changed a lot depending on how much past data they fed into the model and whether they added extra information, like the outdoor temperature. In other words, the choice of history length and external inputs played a big role in how well the models performed [6]. For CO_2, studies emphasize the value of occupancy-related patterns and weather inputs alongside recent CO_2 lags to improve forecasts [2].

Design implications for this thesis:

- Use multivariate inputs, including indoor temperature and CO_2 lags, outdoor temperature, to capture both thermal and occupancy-related dynamics.
- Consider history lengths tuned to forecast horizon (shorter for 1-hour, longer for 3-hour).
- Compare RF (tabular) and CNN/LSTM (sequence) models.
- Analyze the contribution of CO_2-related features to temperature forecasts to understand occupancy effects.
- Discuss implications for predictive and proactive thermal comfort management, enabling a transition from reactive to data-driven HVAC operation.

Beyond improving predictive accuracy, these models support more informed decision-making related to thermal comfort and energy-efficient building management.

2.4 Deep-Learning Architectures for Multivariate Time Series

CNN (1-D, causal). CNN models use temporal filters that scan across the input sequence to identify short-term patterns directly from the data. This makes them particularly suitable for high-frequency indoor environmental measurements, where local temporal variations play a significant role in short-horizon forecasting [6, 8, 19].

LSTM. LSTMs retain longer-range temporal dependencies via gating mechanisms and are widely used in building forecasting when multi-hour dynamics (thermal inertia, occupancy cycles) matter [5, 9, 20].

Hybrid CNN-LSTM architectures leverage both local pattern extraction (CNN) and long-term memory (LSTM), frequently yielding strong results in building datasets with 10-15-minute sampling [6, 8]. Since this thesis already compares CNN and LSTM explicitly, hybrids are discussed as context but not

implemented to keep the model zoo focused and reproducible. In the context of indoor environmental forecasting, CNNs and LSTMs not only offer superior accuracy but also serve as the computational core for data-driven comfort assessment and intelligent HVAC management.

2.5 Interpretability with SHAP

In building management, it is important for any predictive model to be understandable. SHAP helps with this by showing how each input feature influences a specific prediction, both at the individual level and across the whole dataset [11]. In building applications, SHAP has been used to reveal the relative contributions of outdoor temperature, recent indoor lags, and calendar effects, thereby improving trust in ML forecasts and informing operator decisions [12, 21]. In the case of Random Forest models, SHAP values can be computed exactly using TreeExplainer. For neural networks, however, the DeepExplainer version is used to estimate feature contributions, enabling comparison of how different models weigh their inputs.

2.6 BIM-Power BI Visualization and Predictive Reporting

BIM provides the spatial and semantic backbone for contextualizing sensor observations and model outputs within the physical layout of a building. Prior work has shown that coupling Revit models with data pipelines and analytics platforms, frequently via middleware such as Speckle, enables interactive exploration of environmental conditions across rooms and zones.

In this thesis, the same ideas are applied in a predictive way by sending the forecasted temperature and CO_2 values—along with their uncertainty ranges—to a Power BI dashboard connected to the Aule R BIM model. While this setup is not a real-time digital twin, it still provides a useful visual layer that makes predictions easier to understand and supports communication and decision-making about HVAC operation and comfort [22].

2.7 Synthesis and Research Gap

The review of existing studies shows a few clear tendencies. Deep-learning models such as CNNs and LSTMs typically outperform traditional methods when the goal is to forecast multiple indoor variables simultaneously. Random Forest, on the other hand, remains a strong and interpretable baseline that often delivers competitive results with less complexity. Another important finding is the value of SHAP, which helps make the behaviour of these models easier to understand by highlighting the influence of different features. Finally, several studies point out that BIM-based dashboards help place environmental data in its spatial context, making it easier to interpret room-level patterns and conditions.

However, existing studies rarely integrate these components into a unified, predictive workflow for forecasting indoor conditions. In particular, there is a lack of frameworks that combine ML/DL forecasting, explainability (SHAP), and uncertainty estimation.

This thesis addresses this gap by developing an explainable forecasting framework for Aule R that quantifies thermal and air-quality dynamics and supports analysis of HVAC operational efficiency through data-driven, spatially contextualized insights.

2.8 Summary and Conceptual Framework

The literature reviewed in this chapter shows that forecasting Indoor Environmental Quality (IEQ) in educational buildings requires analytical methods capable of capturing nonlinear behavior and strong temporal dependencies. Traditional linear time-series approaches, such as ARIMA, are generally insufficient for these complex conditions. In contrast, modern machine-learning and deep-learning models, including Random Forest (RF), Convolutional Neural Networks (CNNs), and Long Short-Term Memory (LSTM) networks, offer greater flexibility and consistently superior predictive performance for multivariate indoor-environmental data.

Furthermore, recent advances in explainable AI (XAI) tools such as SHAP make it possible to interpret the predictions of these otherwise non-transparent models. SHAP helps clarify how the model works by showing how much different inputs, such as recent values, weather conditions, or calendar patterns, contribute to each prediction. This makes the results easier to interpret and helps experts see what is driving the temperature and CO_2 forecasts.

From a systems perspective, BIM-linked dashboards offer a practical visual context for presenting sensor data and analytics outputs. However, most existing frameworks remain descriptive and lack predictive or decision-support capabilities relevant to HVAC operation.

Conceptual Framework

Building on the points discussed earlier, this thesis follows a framework comprising four interconnected components.

The first part focuses on preparing the data. The historical temperature and CO_2 readings collected in Aule R are aligned with outdoor weather information and calendar markers so that all variables share a common structure and time base. This step turns the raw sensor data into a clean and consistent multivariate time series that both tree-based models and deep-learning architectures can use.

The second element of the framework is the predictive modeling stage. Three algorithms (Random Forest, CNN, and LSTM) are trained to estimate indoor temperature and CO_2 levels one and three hours ahead. Their hyperparameters and input history lengths are adjusted based on how strongly the variables depend on past observations, allowing each model to learn the relevant temporal patterns in the data.

A third component involves understanding how the models make their predictions. Here, SHAP is used to examine the influence of different feature types, such as temporal variables, outdoor weather conditions, and calendar effects. This analysis provides explanations both for individual predictions and for the overall behavior of the models, helping building operators interpret the results with greater confidence.

The final part of the framework deals with visualization. The forecasted values are connected back to the BIM model through the Power BI dashboard developed during the earlier internship. This integration allows the predictions to be viewed directly within the 3D layout of the building, turning the dashboard from a purely descriptive tool into a simple, spatially contextualized decision-support interface.

Chapter 3 – Methodology

3.1 Overview

This chapter describes the methodological framework developed to forecast and visualize indoor environmental conditions in the Aule R academic building at Politecnico di Torino. The approach is organized as a sequence of practical steps, each of which contributes to building a coherent and reliable analytical pipeline.

The process begins with the collection of long-term environmental data from the building's sensor network, followed by a careful preprocessing phase in which the raw measurements are cleaned, aligned in time, and checked for consistency. Once a stable dataset is obtained, a set of temporal and contextual features is engineered to capture the short-term fluctuations, daily patterns, and seasonal trends that characterize indoor temperature and CO_2 dynamics.

These enriched datasets are then used to train three different machine-learning models (Random Forest, CNN, and LSTM) chosen because they reflect complementary ways of learning temporal behavior in buildings. Their performance is evaluated through standard forecasting metrics, and SHAP-based interpretability tools are applied to better understand which variables influence the predictions and why. Finally, the forecasting results are integrated into a BIM-linked Power BI dashboard, enabling the predictions to be explored directly within the 3D model of Aule R. In this way, the methodological framework connects data preparation, model development, and spatial visualization into a single workflow designed to support energy-efficient operation and improve comfort management in the building.

3.1.1 Overview of the Analytical Framework

The overall architecture of the proposed BIM-IoT Integrated Data Analytics Pipeline workflow is illustrated in Figure 3.1. The system integrates heterogeneous building information and sensor data into a unified pipeline for analysis, prediction, and visualization.

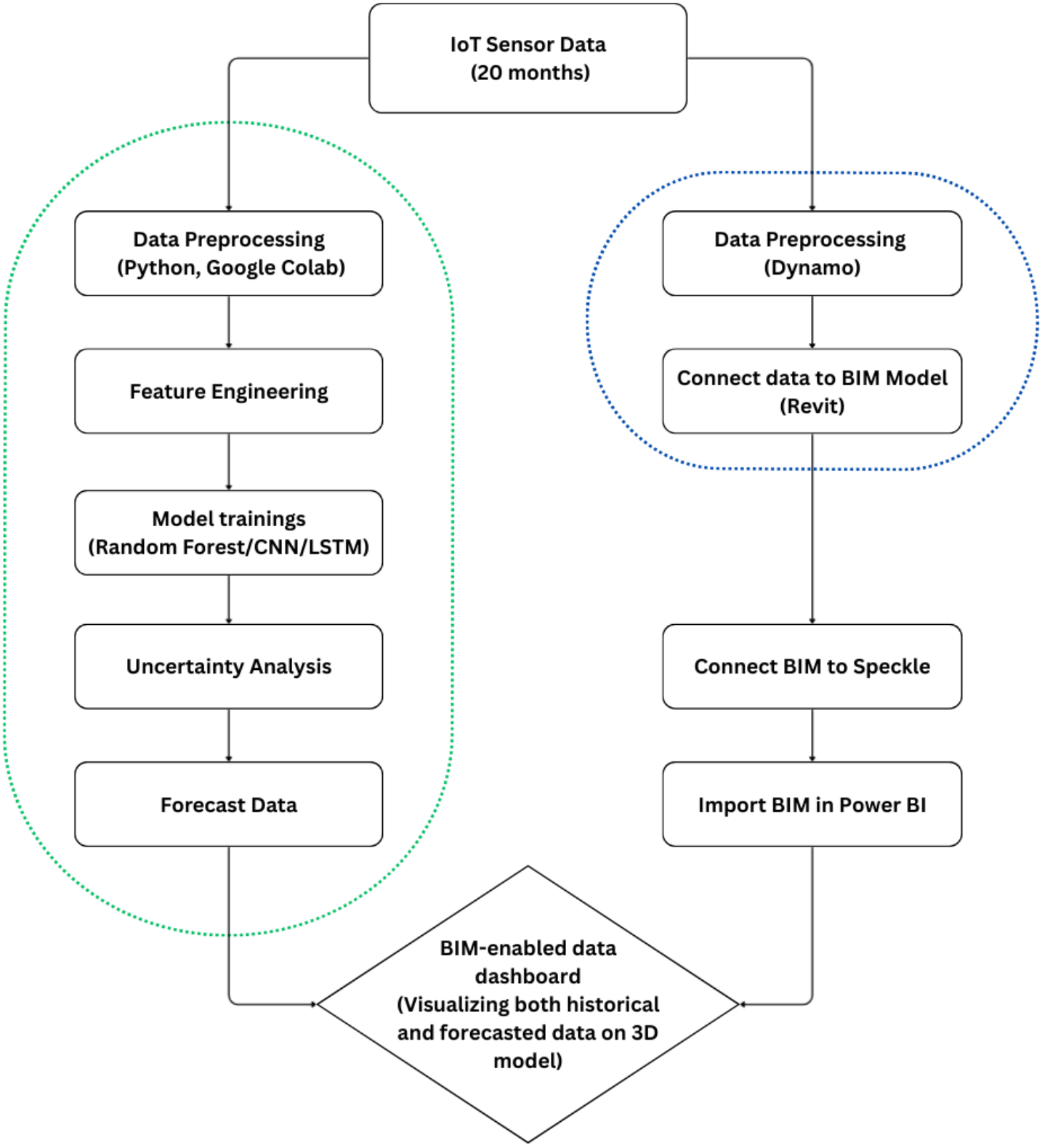

Figure 3.1 Overall architecture of the BIM-IoT Integrated Data Analytics Pipeline workflow

3.2 Data Collection and Preprocessing

3.2.1 Data sources and acquisition

The dataset covers approximately 20 months (5 March 2024 - 25 November 2025), collected every 15 minutes from the sensor network installed in Aule R, Politecnico di Torino. The environmental sensors measure:

Table 3.1 Parameters measured in Aule R

Parameter	Symbol	Unit	Description
Indoor air temperature	Tin	°C	Mean of upper (Z2) and lower (Z1) room sensors, or a single sensor in basement rooms
Outdoor air temperature	Tout	°C	External reference sensor on facade
Setpoint temperature	Tset	°C	Control signal from the HVAC system
CO_2 concentration	C	ppm	Indoor air quality/occupancy indicator

The extended monitoring period results in approximately 58,000-60,000 time-stamped samples per zone, enabling the analysis to capture both short-term dynamics (intra-hour and intra-day variations) and long-term seasonal effects (weekly and annual cycles). Such long-term, high-frequency data enhance the statistical reliability of the forecasting models and enable the learning of periodic behaviors, consistent with findings reported in recent studies on multi-season and annual indoor-environment forecasting [4, 5].

3.2.2 Data cleaning and alignment

All raw CSV sensor files were imported and standardized through a dedicated Python preprocessing module developed for this study. The procedure ensured data consistency, temporal integrity, and reliability across all monitored zones.

The main steps are summarized as follows:

- **File normalization and encoding:**
  Non-data header lines (e.g., sep=;) and hidden byte-order marks (BOM) were automatically detected and removed.
  Column names in both English and Italian were standardized to consistent labels across all datasets.

- **Timestamp parsing and localization:**
  All timestamps were localized to *Europe/Rome* (CET/CEST). To handle the DST fallback hour (02:00 duplicated), we used an ambiguity-masking strategy that identifies duplicate timestamps and consistently keeps the first occurrence (the DST instance), while discarding the second (standard-time duplicate). Nonexistent times during the spring transition were shifted forward. This yields a strictly increasing, unique datetime index before 15-minute resampling and canonical re-indexing.
- **Temporal resampling and alignment:**
  Records were resampled on a strict 15-minute grid covering the whole monitoring period (5 March 2024 - 25 November 2025).
  This ensured full temporal synchronization between temperature and CO_2 data for each zone.
- **Quality control and outlier handling:**
  Physical plausibility limits were applied to indoor temperature ($-10\ °C \leq T_{in} \leq 50\ °C$), with out-of-range values converted to NaN and short gaps (≤ 1 h) filled through time-based interpolation.
  Sensor inspection revealed that indoor CO_2 occasionally contained physically impossible low values (below the atmospheric background ≈ 400 ppm). To address this, a data-driven dynamic floor was defined as the maximum between a conservative 380 ppm baseline and a background estimate from the dataset (5th percentile − 30 ppm). All CO_2 values falling below this threshold were hard-clipped to restore physical plausibility, as these represent sensor errors rather than genuine missing data. Extremely high values (> 4000 ppm), if present, were converted to NaN, and short gaps (≤ 1 h) were interpolated. In contrast, longer gaps were removed during the final dropna() stage after lag/rolling feature generation. This correction improves the physical validity and stability of the time-series models [23].
- **Calendar feature mapping:**
  Each record was enriched with temporal attributes (hour of day, day of week, month, and season) to support the identification of diurnal and seasonal patterns and occupancy cycles.

Through these steps, the dataset was transformed into a uniform, timezone-aware, and quality-controlled dataframe comprising approximately 58,000-60,000 records per zone. This temporally consistent and seasonally complete dataset provides a robust foundation for feature engineering and subsequent machine-learning modelling stages [24].

3.3 Feature Engineering and Selection

To capture both short-term fluctuations and long-term thermal trends, a set of temporal and contextual features was engineered from the cleaned data.

3.3.1 Temporal lag features

To incorporate temporal memory, lag features of 15-minute data were generated for both indoor temperature (temp_mean) and CO_2 (co2) at multiple horizons:

L = {1, 2, 3, 4, 8, 12, 24, 48}

corresponding respectively to 15 min

These lag terms encode autocorrelation patterns essential for forecasting.

3.3.2 Rolling-window statistics

Rolling means smooth short-term fluctuations and reveal aggregated trends.

For each variable, rolling windows of 1 h (4 steps), 3 h (12 steps), 6 h (24 steps), and 24 h (96 steps) were computed, providing local trend information used by the learning algorithms. Such aggregated descriptors help capture mid-term thermal dynamics characteristic of massive building envelopes [18].

3.3.3 Calendar and contextual features

To capture the intrinsic periodicity of building operation and environmental dynamics, a set of calendar and contextual features was engineered.

Time-related variables, including hour of day, day of week, and weekend flag, were encoded using cyclical sine–cosine transformations to preserve the continuity between the end and start of each period [25].

Specifically:

$$\boldsymbol{hour\_sin} = \boldsymbol{sin}\left(\frac{\boldsymbol{2\pi \cdot hour}}{\boldsymbol{24}}\right), \qquad \text{hour\_cos} = \cos\left(\frac{2\pi \cdot \text{hour}}{24}\right)$$

And similarly for the weekly cycle:

$$\boldsymbol{week\_sin} = \boldsymbol{sin}\left(\frac{\boldsymbol{2\pi \cdot day\_of\_week}}{\boldsymbol{7}}\right), \qquad \text{week\_cos} = \cos\left(\frac{2\pi \cdot \text{day\_of\_week}}{7}\right)$$

These encodings allow the model to learn smooth transitions across 24-hour and 7-day boundaries, avoiding artificial discontinuities (e.g., between 23:00 and 00:00).

Since the dataset spans more than one calendar year (March 2024 – November 2025), annual seasonality was represented using cyclical month encodings to capture inter-year climatic and operational patterns:

$$\boldsymbol{month\_sin} = \boldsymbol{sin}\left(\frac{\boldsymbol{2\pi \cdot month}}{\boldsymbol{12}}\right), \qquad \text{month\_cos} = \cos\left(\frac{2\pi \cdot \text{month}}{12}\right)$$

In addition, an academic-term indicator (in_academic_term) was constructed to reflect the operational schedule of the Politecnico di Torino across multiple academic years covered by the dataset. Teaching activities and examinations typically occur from October through July, and in September, while August corresponds to a low-occupancy, non-academic period. The indicator was computed for both academic years (2023/24 and 2024/25) to ensure consistency across the extended monitoring period.

This combination of cyclic, seasonal, and academic-calendar features allows the forecasting models to distinguish long-term climatic trends from occupancy-driven behavioral patterns, thereby improving the characterization of indoor temperature and CO_2 variations. Outdoor temperature and HVAC setpoint were included as exogenous contextual drivers. At the same time, CO_2 concentration served both as an explanatory feature of occupancy patterns and as a target variable representing indoor air quality.

3.3.4 Feature selection

To avoid working with an unnecessarily large set of lagged, rolling, and time-encoding features, mutual information (MI) ranking was applied on the training set to determine which predictors contributed most to forecasting accuracy. MI measures how strongly each feature is related to the target variable, even when the relationship is nonlinear.

Based on these scores, the top K = 80 features were selected for all models. These typically included the most relevant temperature and CO_2 lags, short-term rolling averages, key calendar encodings (such as hour and day-of-week), and external drivers like setpoint and outdoor temperature.

Reducing the feature space in this way helped remove redundancy, sped up training, and improved model stability, while still preserving all essential information needed for accurate forecasting.

3.4 Model Design and Training Procedure

Three supervised-learning models were implemented to forecast indoor temperature and CO_2 concentration for horizons of 1 h (4 steps) and 3 h (12 steps) ahead. The selected models cover complementary learning paradigms:

Table 3.2 Forecasting models used in this study

Model	Type	Rationale
Random Forest (RF)	Ensemble of decision trees	Captures nonlinear relations in tabular features; interpretable via SHAP [26].
Convolutional Neural Network (CNN)	1-D sequence learner	Efficiently extracts short-term temporal patterns from recent time windows [27].

Long Short-Term Memory (LSTM)	Recurrent neural network	Learns long-term dependencies and seasonality in time-series data [28].

Normalization of Input Features for Deep Learning Models

Deep learning models such as CNNs and LSTMs are highly sensitive to the scale of input features. Since the indoor environmental dataset contains variables with different numerical ranges (for example, temperature values around 20–30 °C, CO_2 concentrations ranging from 400 to 2000 ppm, and cyclical encodings bounded between −1 and 1), unscaled inputs may lead to unstable training, slow convergence, or poor generalization.

To ensure stable and efficient optimization, all numerical input features used by the sequence models were standardized before training.

Standardization was performed using a StandardScaler, which transforms each feature as:

$$\frac{\mathbf{x} - \boldsymbol{\mu}}{\boldsymbol{\sigma}} = \boldsymbol{z}$$

where μ and σ are the mean and standard deviation of the feature computed only from the training portion of the dataset.

This prevents data leakage, ensuring that no information from the validation or test periods is used during preprocessing.

The fitted scaler was then applied consistently across training, validation, and testing sets.

Only the input features were normalized; the target variables (temperature and CO_2) were left in their physical units (°C and ppm) to preserve interpretability of the forecasting errors.

This normalization step improves optimization stability, balances the contribution of each feature, and allows both CNN and LSTM models to learn temporal patterns effectively.

3.4.1 Data splitting

To ensure a precise and leakage-free evaluation of forecasting performance, the dataset was partitioned using fixed chronological boundaries rather than ratio-based or randomized splitting. This approach reflects standard best practices in time-series forecasting, where the model must be trained exclusively on past data and evaluated strictly on future periods.

The full 20-month dataset (5 March 2024 – 25 November 2025) was therefore partitioned into three consecutive segments representing past, intermediate-future, and unseen-future conditions:

- **Training period:** 5 March 2024 → 31 May 2025

This segment spans roughly 450 days (~72%) and includes full seasonal variability. It provides the model with a broad range of climatic and operational scenarios necessary for learning robust temporal patterns.

- **Validation period:** 1 June 2025 → 31 August 2025
  A three-month summer interval (~14%) is used exclusively for hyperparameter tuning and early stopping. This period represents a distinct thermal regime, making it suitable for optimizing model behavior without affecting the final evaluation.
- **Testing period:** 1 September 2025 → 25 November 2025
  The final portion (~14%) corresponds to the beginning of the 2025/26 academic year and is held out strictly for unbiased performance assessment. Models never observe this period during training or tuning.

This splitting approach ensures that the forecasting models learn only from past observations and are evaluated solely on future, operationally relevant data. Compared to proportional splits (e.g., 70/15/15), the fixed-calendar strategy preserves seasonal, academic, and occupancy-related patterns. It provides a more realistic assessment of forecasting performance in real-world building operation scenarios.

Table 3.3 Chronological data-splitting strategy

Subset	**Date Range**	**Duration**	Percentage	Purpose
Training	5 March 2024 → 31 May 2025	~450 days	~72%	Model fitting: learning seasonal, daily, and CO_2/thermal patterns
Validation	1 June 2025 → 31 August 2025	~92 days (3 months)	~14%	Hyperparameter tuning, early stopping, and model selection
Testing	1 September 2025 → 25 November 2025	~86 days	~14%	Final unbiased performance evaluation on unseen future data

3.4.2 Model configuration

To capture the different temporal behaviors of indoor environmental data, three complementary machine-learning models were implemented:

A Random Forest (RF), a Convolutional Neural Network (CNN), and a Long Short-Term Memory network (LSTM).
Each model offers distinct advantages in terms of interpretability, the ability to model nonlinear dynamics, and the capacity to learn temporal dependencies.

Their configurations and design rationales are described below:

- **Random Forest (RF):**
  The Random Forest model serves as a strong non-parametric baseline that can model nonlinear feature interactions without requiring extensive hyperparameter tuning.
  Because RF does not rely on a sequential input structure, it is particularly effective when the target variable is primarily explained by engineered features such as lags, rolling statistics, and calendar embeddings.
  The RF model was trained separately for each zone, forecasting horizon, and target variable using a fixed configuration optimized through pilot experiments. Its main parameters are summarized in Table 3.4.

Table 3.4 Configuration of the Random Forest model

Parameter	Specification	Purpose
n_estimators	120	Balance of accuracy and speed
max_depth	15	Prevents overfitting
Min_samples_leaf	5	Smoother prediction surface
Max_features	"sqrt"	Standards for regression forests
n_jobs	–1	Full CPU parallelism
bootstrap	True	Improves generalization
Feature Selection	Top 80 features (Mutual Information)	Reduces redundancy and speeds training
SHAP Analysis	600 validation samples, check_additivity=False	Generates interpretable feature importance at low cost

SHAP analysis was computed on 600 samples of the validation set (check_additivity=False) to derive interpretable feature importance and reduce computational overhead, consistent with standard practice for tree explainers in large ensembles.

- **Convolutional Neural Network (CNN):**

  The CNN model is designed to learn short-term temporal dependencies directly from raw sequences. Unlike RF, which relies on engineered lags, the CNN automatically extracts local temporal patterns through convolutional filters.

  A causal architecture was employed to ensure that the model never uses information from the future. This is critical for forecasting tasks where the temporal order must be strictly preserved. The configuration is summarized in Table 3.5.

Table 3.5 Configuration of the CNN model

Parameter	Specification	Purpose
Input shape	(12 steps, n features) = 3 h window	Fixed history length per sequence.
Conv1D layers	2 × Conv1D (64 filters, kernel = 3, ReLU, causal padding)	Extracts local temporal patterns.
Pooling layer	Global Average Pooling 1D	Reduces dimensionality and prevents overfitting.
Dense layers	Dense 128 (ReLU) → Dense 1 (output)	Maps learned features to forecast value.
Optimizer	Adam (lr = 1×10^{-3})	Adaptive gradient optimization.
Loss function	Mean Absolute Error (MAE)	Robust to outliers, interpretable in °C/ppm.
Batch size	64	Efficient GPU utilization.
Epochs	≤ 120 with Early Stopping (patience = 15)	Stops training when no validation improvement.
Learning-rate schedule	ReduceLROnPlateau (factor = 0.5, patience = 5)	Refines convergence.

Training Strategy

CNN models were trained for up to 120 epochs, using the Adam optimizer and MAE loss function. Early stopping (patience=15) prevented overfitting, while a ReduceLROnPlateau schedule allowed the model to refine its learning rate during training.

- **Long Short-Term Memory (LSTM):**

  The LSTM model is responsible for capturing longer temporal dependencies that may span several hours.

  Whereas CNNs focus on local patterns, LSTMs maintain an internal memory that tracks trends and contextual information over extended periods.

  Although the input sequence length was kept consistent with the CNN (12 time steps = 3 hours) to allow fair comparison, the LSTM's memory cells will enable it to retain information far beyond the explicit input window.

  Its configuration is presented in Table 3.6.

Table 3.6 Configuration of the LSTM model

Parameter	Specification	Purpose
Input shape	(12 steps, n features) = 3 h window	Same sequence length as CNN for comparability.
LSTM layer	128 units, dropout = 0.2	Captures long-term temporal dependencies while preventing overfitting.
Dense layers	Dense 128 (ReLU) → Dense 1 (output)	Nonlinear mapping to forecast the target.
Optimizer	Adam (lr = 1×10^{-3})	Stable training of recurrent networks.
Loss function	Mean Absolute Error (MAE)	Same metric as CNN for fair comparison.
Batch size	64	Trade-off between stability and memory use.
Epochs	≤ 120 with Early Stopping (patience = 15)	Avoids overfitting and excessive runtime.
Learning-rate schedule	ReduceLROnPlateau (factor = 0.5, patience = 5)	Adaptive fine-tuning of learning rate.

Each model was trained and validated independently for every zone and forecast horizon. For the sequence models (CNN and LSTM), both architectures were first trained on the training subset and evaluated on the validation subset using mean absolute error (MAE) as the primary selection criterion. For each zone–target–horizon combination, the architecture achieving the lowest validation MAE was selected as the best-performing sequence model. The chosen model was then fine-tuned on the combined training and validation data for an additional 10 epochs before being evaluated on the held-out test set. This procedure ensured a fair comparison between architectures while exploiting all available data for final parameter refinement, and the overall multi-model setup enabled the capture of both short-term and longer-term temporal patterns of the indoor environment.

Hyperparameter Selection Rationale:

The chosen hyperparameters were selected through a small, literature-guided tuning process applied to a representative subset of the data. Full hyperparameter search was avoided due to the computational cost across all zones and horizons; instead, configurations that consistently provided a stable balance between accuracy and runtime were adopted. These settings follow standard practice in building forecasting studies and ensure reproducible performance without unnecessary model complexity.

3.4.3 Model Architectures

Table 3.7 Summary of the model architectures used for forecasting

Model	Key Layers	Strength
Random Forest	120 trees, depth ≤ 15	Interpretable via SHAP
CNN	Conv1D × 2 + Dense 128	Detects local temporal patterns
LSTM	LSTM 64 + Dense 128	Captures long-term dependencies

All models were implemented in Python 3.12, using scikit-learn, TensorFlow Keras, and the SHAP library in Google Colab.

3.5 Uncertainty Quantification in Forecasting Models

To evaluate the reliability of the predictive models, the uncertainty of forecasts was also quantified. This step indicates how confident each model is about its predictions and helps interpret the results beyond single deterministic values.

For the **Random Forest (RF)** model, the 95% confidence interval was computed from the variability of tree outputs as:

$$\sigma_{trees} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Where

- σ_{trees} represents the standard deviation across n trees
- $\bar{y}$ is the mean prediction across all trees

The 95% confidence interval was then computed as:

$$\bar{y} \pm 1.96 * \sigma_{trees}$$

This interval reflects epistemic uncertainty arising from variability in model structure and training data.

For the **LSTM** model, uncertainty was quantified using Monte Carlo (MC) Dropout, in which dropout layers remain active during inference. The model performs T=30 stochastic forward passes, yielding predictions $y^1, y^2, \ldots, y^T$.

The predictive mean is:

$$\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y^T$$

The corresponding uncertainty is obtained from the standard deviation of these outputs:

$$\sigma_{MC} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(y^T - \bar{y})^2}$$

The 95% confidence interval is defined as:

$$\bar{y} \pm 1.96 * \sigma_{MC}$$

This method captures model uncertainty due to stochasticity in neural network weights. These confidence bands highlight the model's reliability and variation over time, improving interpretability and supporting decision-making for the Aule R building. [29]

3.6 Evaluation Metrics and Validation Strategy

Model performance was assessed using three standard regression error metrics:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - y\hat{}_i|, \quad RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - y\hat{}_i)^2}, \quad sMAPE = \frac{200}{n}\sum_{i=1}^{n}\frac{|y_i - y\hat{}_i|}{|y_i| + |y\hat{}_i|}$$

MAE reflects the average magnitude of prediction errors in the physical units of the target (°C or ppm), RMSE penalizes larger deviations more heavily and therefore highlights instability or peak mispredictions, while sMAPE provides a unitless percentage error that facilitates comparison across zones and different environmental variables.

For evaluation, a fixed chronological split was adopted instead of ratio-based partitioning.

The full 20-month dataset (5 March 2024 – 25 November 2025) was divided into:

- **Training:** 5 March 2024 → 31 May 2025
- **Validation:** 1 June 2025 → 31 August 2025
- **Testing:** 1 September 2025 → 25 November 2025

This approach ensures that models are trained exclusively on past data and validated/tuned on an intermediate future period. At the same time, the final testing phase corresponds to a completely unseen segment reflecting the beginning of the 2025/26 academic year.

Unlike proportional splits (e.g., 70/15/15), the fixed-boundary strategy preserves seasonal, academic, and occupancy-related dynamics and avoids temporal leakage without requiring artificial buffer zones.

For sequence models (CNN and LSTM), the validation MAE was used for early stopping and model selection.

Final performance was continuously computed on the held-out test period using all three metrics (MAE, RMSE, sMAPE).

MAE evaluates the mean absolute deviation, RMSE penalizes larger errors more heavily, and sMAPE normalizes errors relative to magnitude, facilitating comparison across different zones [30].

3.7 Interpretability with SHAP Analysis

To ensure transparency, model interpretability was achieved using SHapley Additive exPlanations (SHAP) [26, 11]. It was employed to interpret the Random Forest predictions.

Each SHAP value quantifies the contribution of a specific feature to increasing or decreasing a given prediction relative to the model's baseline output.

Aggregating absolute SHAP values across all test samples provided global feature importance rankings, revealing the dominant factors driving the indoor temperature and CO_2 forecasts.

Short-term temperature lags (e.g., temp_mean_lag1–4), outdoor temperature, and time-of-day features consistently appeared among the most influential predictors, reflecting the physical intuition of thermal inertia and occupancy-driven dynamics.

For neural models (CNN and LSTM), the same interpretability concept can be extended using DeepExplainer to visualize temporal relevance patterns, although this was reserved for future work due to computational cost.

This interpretability approach aligns with best practices in explainable building-energy forecasting [12].

3.8 Visualization through BIM-enabled Power BI Dashboard

The final stage of the methodology focused on embedding the predictive results within the 3D environment of the Aule R building to create an innovative dashboard. In this phase, the Building Information Model (BIM) developed in Revit acted as a static geometric reference, a spatial container that provides room-level semantics but does not operate as a real-time digital twin. Its purpose was to serve as a structured spatial framework onto which analytical and predictive data could be mapped.

The forecast outputs, including predicted indoor temperature and CO_2 concentration together with their associated uncertainty ranges, were exported from the modelling pipeline as structured CSV tables. These tables followed a unified schema containing the zone identifier, timestamp, target variable, forecasting horizon, predictive mean, and upper–lower confidence bounds. Once exported, the prediction datasets were imported into Microsoft Power BI and linked to the corresponding Revit spaces using the Speckle connector.

This integration enabled the construction of a fully interactive 3D visualization of the Aule R building, where each classroom (R1-R4, R1B-R4B) displays its forecasted conditions directly within the spatial model. Through color-coded room shading and dynamic filtering controls, users can explore predicted values over different horizons, compare them against historical measurements, and visually inspect uncertainty propagation across zones and time periods. The dashboard also includes conventional time-series charts and slicers that complement the 3D perspective and allow users to navigate specific dates, variables, or modeling configurations.

Overall, the BIM-linked Power BI dashboard serves as a data-driven reporting layer that enhances the interpretability of the predictive analytics. Rather than functioning as a real-time digital twin, it provides a clear, intuitive environment for communicating model insights, supporting comparative analysis across rooms, and facilitating informed decision-making related to HVAC performance, comfort assessment, and operational planning within the Aule R building.

3.9 Summary

This chapter detailed the research design and methodological workflow adopted for forecasting indoor environmental conditions in Aule R. Starting from a 20-month dataset, the workflow encompassed data preprocessing, engineered temporal and contextual features, optimized Random Forest and deep-learning models, and interpretable SHAP analysis.

Finally, the predictive outputs were integrated into the BIM-enabled data visualization in an innovative Power BI dashboard, establishing a coherent workflow that enhances data-driven building-management decisions.

Chapter 4 – Results

4.1 Overview

This chapter presents the experimental results of the data-driven forecasting framework developed for predicting indoor environmental conditions in Aule R academic building. The analyses follow the methodological pipeline described in Chapter 3 and focus on evaluating the performance of Random Forest (RF), Convolutional Neural Network (CNN), and Long Short-Term Memory (LSTM) models for short-term forecasting of indoor temperature and CO_2 concentration across multiple zones of the building.

The chapter begins with an exploratory analysis of the cleaned 20-month dataset (March 2024 – November 2025), a high-resolution (15-minute) dataset used in this study. It then reports the forecasting accuracy of the three modeling approaches, compares their performance across prediction horizons, quantifies uncertainty, and provides model interpretability results using SHAP values. Finally, the chapter showcases the integration of model outputs into a BIM-Power BI dashboard, providing a concise visual decision-support layer for interpreting indoor environmental performance.

4.2 Exploratory Data Analysis (EDA)

4.2.1 Data Coverage and Completeness

The cleaned 20-month dataset of indoor environmental measurements (March 2024 – November 2025) contains 478,728 time-stamped samples at a 15-minute resolution. After applying the full preprocessing pipeline, comprising timestamp normalization, DST ambiguity correction, removal of corrupted values, physical-range filtering, and short-gap interpolation, the dataset achieves 100% data coverage for both indoor temperature (temp_mean) and CO_2 concentration.

No missing values remain in either variable (0 NaNs), meaning that the final model-ready dataset is fully complete and temporally continuous across the entire monitoring period. This complete coverage ensures that the subsequent feature-engineering steps (lags, rolling statistics, and temporal encodings) can be applied without introducing additional data loss, thereby providing a robust and consistent foundation for forecasting models.

4.2.2 Data Quality Checks

During the preprocessing stage, several quality-control procedures were applied to ensure the physical plausibility and temporal consistency of the 20-month dataset. This section summarizes the outcomes of these procedures.

CO_2 Dynamic Floor and Clipping

Inspection of the raw CO_2 measurements revealed the presence of physically implausible low values (often below 300 ppm), typically caused by sensor resets, power interruptions, or short communication faults.

To correct these anomalies, a dynamic CO_2 background floor was computed for each zone as:

- the conservative baseline (380 ppm), and
- the 5th percentile of observed values minus 30 ppm,

with the larger of the two selected as the minimum plausible CO_2 value.

All CO_2 measurements falling below this dynamic threshold were hard-clipped.

The number of corrected samples differs significantly across zones, reflecting differences in sensor stability and noise levels:

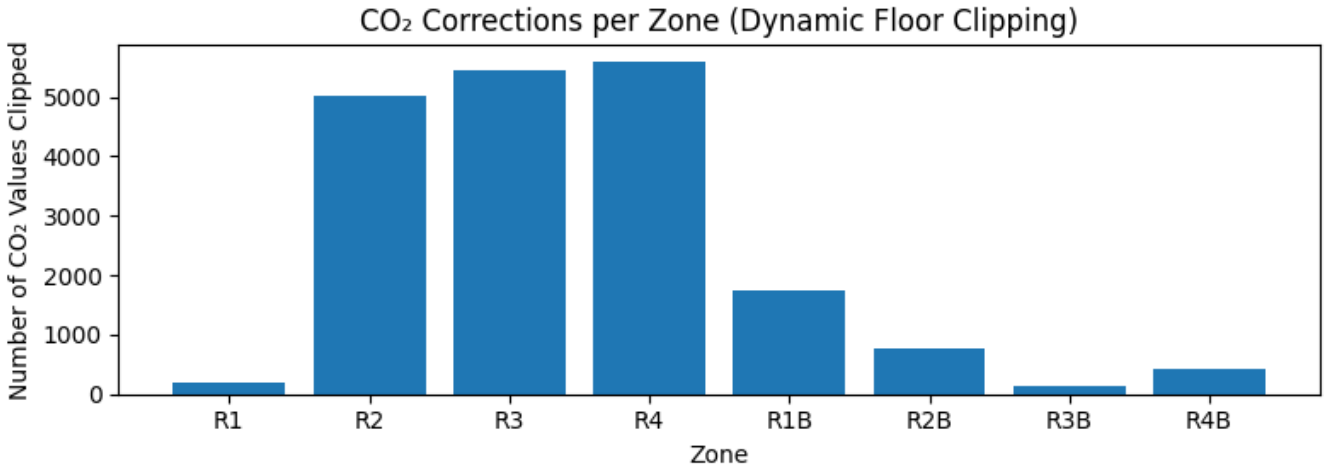

Figure 4.1 shows the distribution of CO_2 corrections across the eight monitored zones

These corrections ensure physical plausibility of the CO_2 time series while preserving short-term variability and the underlying occupancy-related patterns essential for forecasting.

Temperature Outlier Filtering

Indoor air temperature values outside the physical range −10 °C to 50 °C were identified as sensor faults and replaced with NaN.

Such anomalies were rare and typically originated from short communication interruptions or corrupted sensor packets.

Subsequently, time-based interpolation (≤ 1 hour) was applied to reconstruct short missing segments.

Longer gaps were eliminated automatically after generating lag and rolling features, ensuring that only physically plausible and temporally consistent data entered the forecasting models.

DST Ambiguity Resolution

The transition from daylight saving time (CEST → CET) on 27 October introduced duplicated timestamps at 02:00, creating ambiguity in chronological ordering.
These were automatically detected, and for each duplicated timestamp, the first occurrence (CEST instance) was retained, while the duplicate was removed.
This correction enforces a strictly monotonic timestamp index across all zones, preventing temporal leakage and ensuring consistent alignment during resampling and feature engineering.

4.2.3 Statistical Overview of Indoor Variables

Indoor temperature in Aule R ranges from 12.75°C to 33.80°C, with a mean of 23.06°C and a moderate variability (std = 2.61°C). This relatively narrow spread reflects the combined effect of HVAC control and seasonal changes over the 20-month monitoring period. The temperature distribution is approximately unimodal and close to symmetric around 23°C, indicating that most observations cluster within a few degrees of typical comfort conditions.

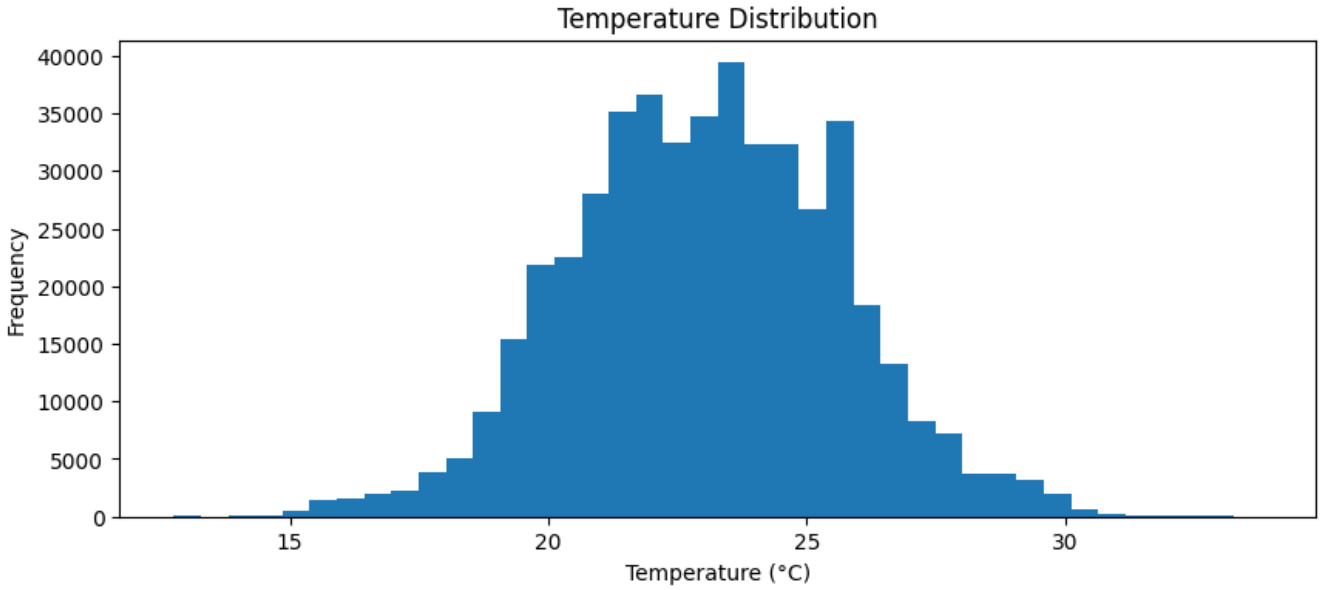

Figure 4.2 Histogram of indoor temperature values across all zones over the 20 months.

CO_2 concentration exhibits a much broader and more asymmetric distribution. After applying the dynamic background floor described in Section 4.2.2, values span from 380 ppm to 2013 ppm, with a mean of 590.65 ppm and a standard deviation of 211.09 ppm. The histogram is clearly right-skewed, with a large mass of observations around 400–600 ppm and a long tail extending toward higher concentrations. This pattern highlights the strong influence of occupancy and ventilation events: CO_2 remains near background levels during unoccupied periods but can rise above 1000 ppm during busy

class hours, confirming that CO_2 is considerably more variable and occupancy-driven than indoor temperature.

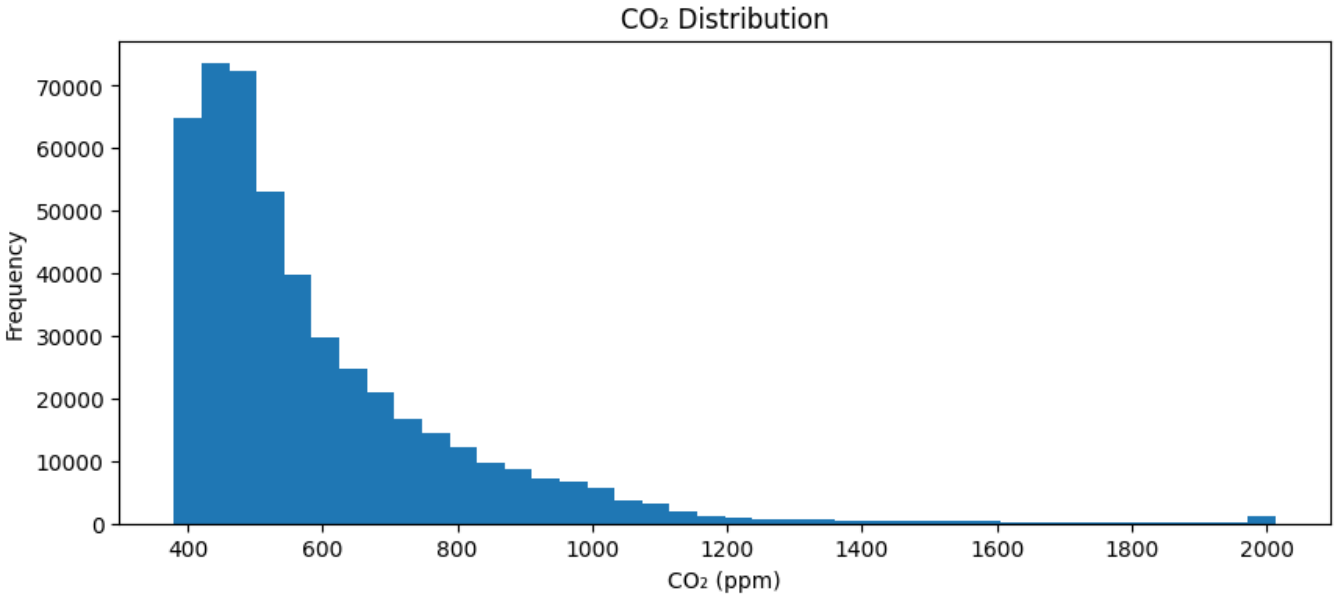

Figure 4.3 Histogram of indoor CO_2 concentrations across the dataset.

4.2.4 Temporal Patterns

The 20-month dataset exhibits clear and consistent temporal structures in both indoor temperature and CO_2 concentration, reflecting typical occupancy cycles and HVAC operation in Aule R.

Daily Patterns

Daily temperature dynamics follow a smooth and predictable cycle:

- Minimum temperatures occur around 06:00-07:00 (≈ 22.7°C), reflecting overnight cooling under reduced HVAC activity and no internal heat gains.
- From 08:00 onward, temperature gradually increases, reaching afternoon peaks of approximately 23.4-23.45°C between 15:00 and 17:00.
- After 18:00, the temperature slowly declines toward nighttime levels.

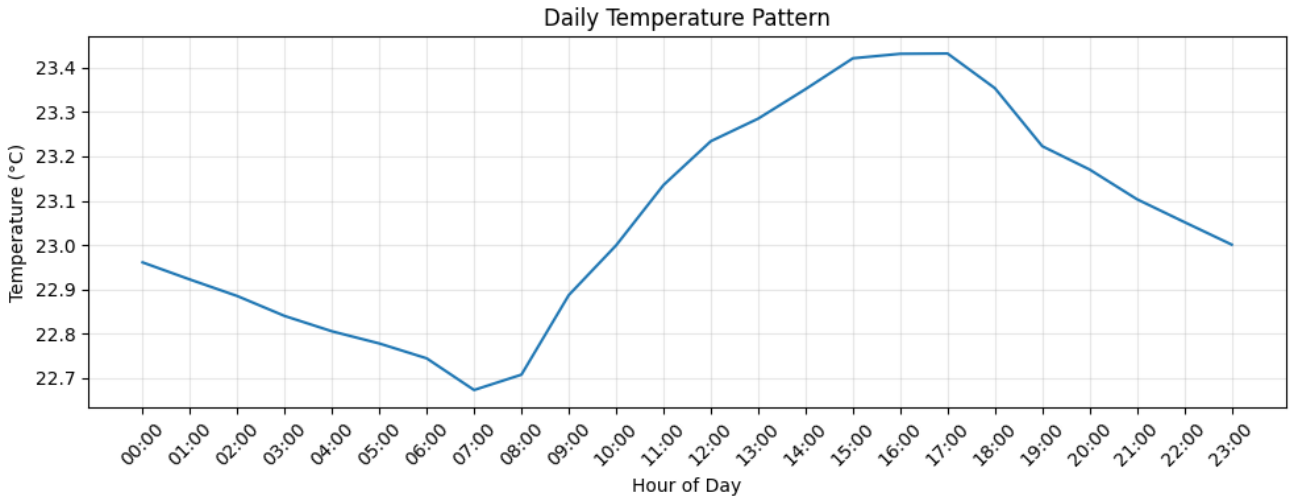

Figure 4.4 Average daily temperature cycle aggregated across all zones.

Daily CO_2 trends strongly reflect occupancy behavior:

- CO_2 is lowest in the early morning (≈ 480-500 ppm).
- A sharp rise begins between 07:00 and 09:00, aligned with the start of class activity.
- CO_2 reaches its maximum levels around 10:00-12:00, peaking between 680 and 700 ppm.
- Concentrations remain elevated during classroom hours and then drop rapidly after 18:00, returning toward background levels during the night.

These patterns clearly indicate that temperature dynamics are dominated by slow thermal inertia, whereas CO_2 responds rapidly to changes in occupancy.

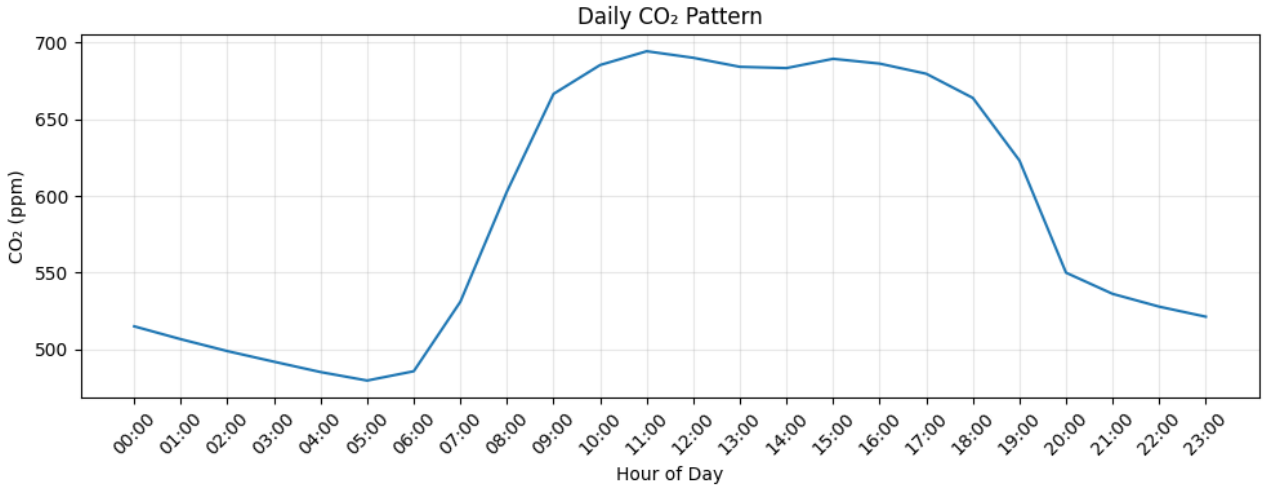

Figure 4.5 Average daily CO_2 cycle aggregated across all zones.

Weekly patterns

Weekly patterns provide another layer of insight into how indoor conditions evolve in relation to building usage. Unlike the more pronounced hour-to-hour variability observed in the daily cycles, changes across the week are subtler yet highly informative.

For **temperature**, the weekly trend shows a mild but very consistent difference between weekdays and weekends. During Monday–Friday, indoor temperatures tend to be slightly higher, with the weekly maximum occurring on Tuesday (around 23.18–23.19°C). This pattern can be attributed to two combined effects: regular occupancy throughout the workweek, which introduces internal heat gains, and the HVAC system operating in its standard weekday mode. Because heating, cooling, and ventilation schedules typically adjust to expected classroom activity, weekday temperatures remain marginally elevated.

As the week transitions into Saturday and Sunday, temperature values dip noticeably, reaching a minimum of approximately 22.85°C on Sunday. These lower values are consistent with reduced or absent occupancy and less intensive HVAC operation over the weekend, when the building is largely unused.

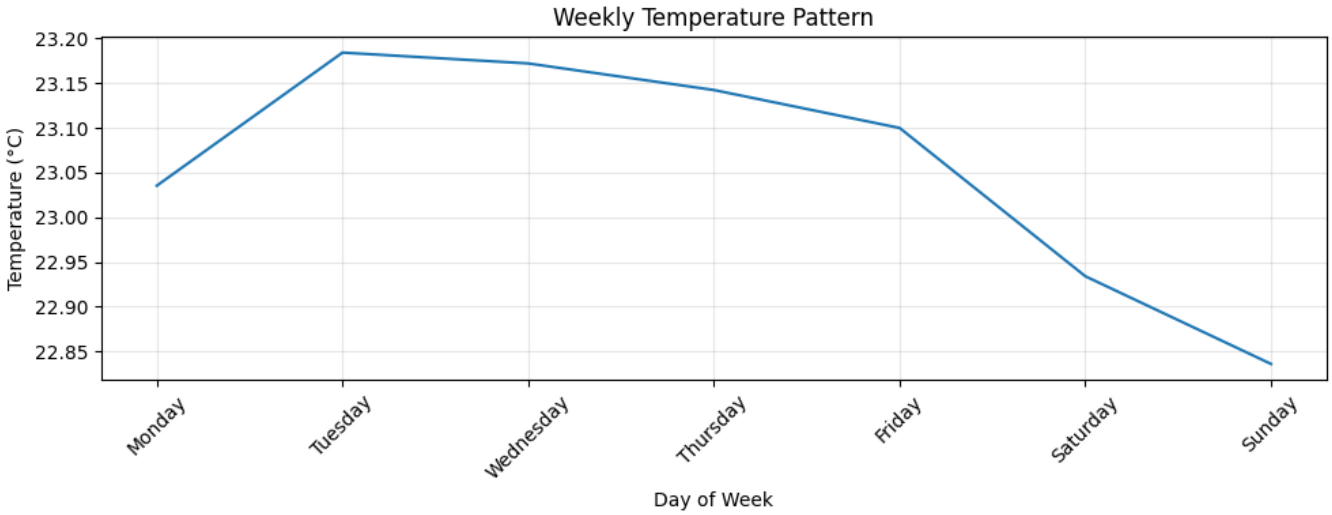

Figure 4.6 Average indoor temperature across days of the week.

CO_2 exhibits a much stronger weekly contrast. On weekdays, especially Monday and Tuesday, CO_2 concentrations peak above 650 ppm, reflecting sustained occupancy levels and continuous classroom turnover. Throughout the middle of the week, CO_2 remains elevated, mirroring the rhythm of teaching schedules.

In contrast, weekend CO_2 levels drop dramatically, often stabilizing below 480 ppm. These low and steady values align with background outdoor concentrations, clearly confirming that the building is either unoccupied or only minimally used. This distinct weekday–weekend separation highlights the direct sensitivity of CO_2 to human presence and ventilation control.

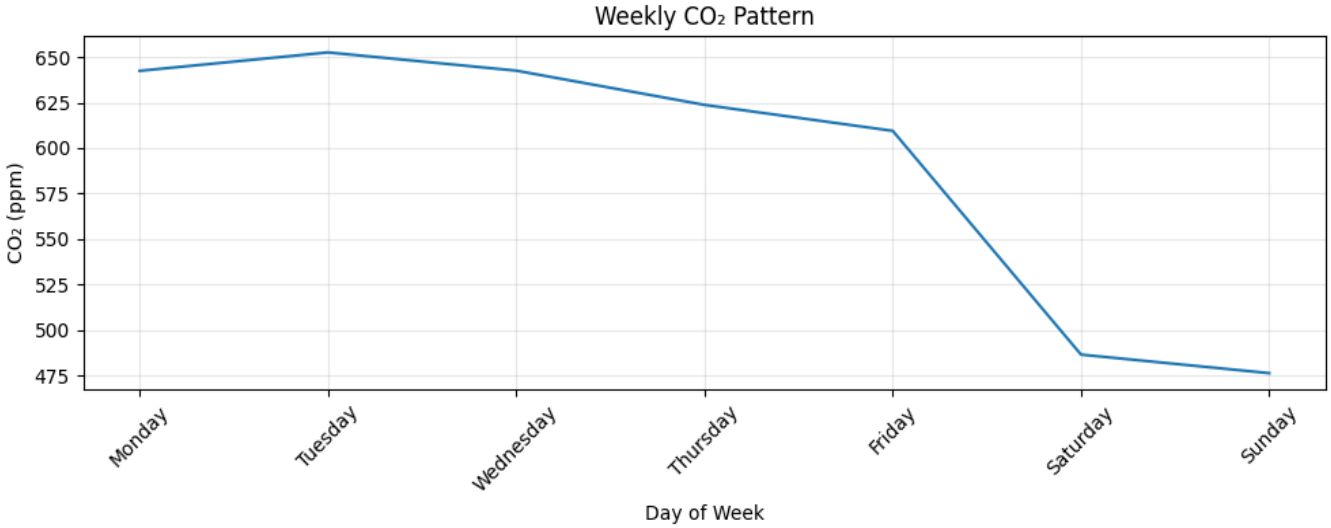

Figure 4.7 Weekly CO_2 pattern showing significant differences between weekdays

Together, the weekly and daily temporal structures, smooth thermal cycles, sharp CO_2 peaks aligned with occupancy, and apparent operational differences between weekdays and weekends, strongly justify incorporating hour-of-day, day-of-week, and cyclical sine–cosine temporal encodings into the feature-engineering pipeline used for forecasting.

4.2.5 Autocorrelation Analysis (ACF)

Autocorrelation analysis was performed to quantify the temporal persistence of indoor temperature and CO_2 concentration and to guide the selection of lagged features for forecasting. Figures 4.8 and 4.9 show the autocorrelation function (ACF) over the first 48 lags (equivalent to a 12-hour window at 15-minute resolution).

Temperature exhibits very high persistence across the entire 12-hour period.

ACF values remain above 0.95 for all 48 lags, indicating an extremely strong autoregressive structure and slow thermal dynamics typical of conditioned indoor environments. This high temporal memory confirms the suitability of including multiple long lags and longer rolling windows in temperature forecasting models.

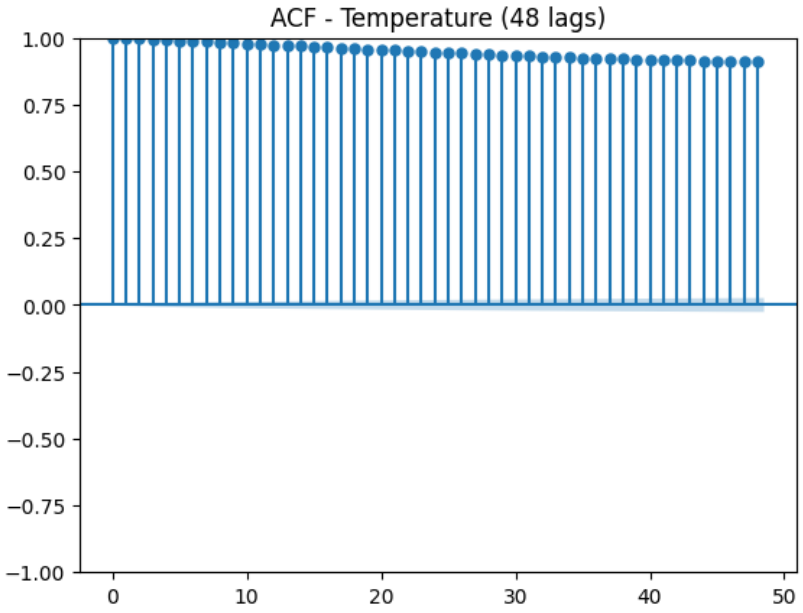

Figure 4.8 Autocorrelation function (ACF) of indoor temperature over 48 lags (12 hours). The very slow decay indicates high temporal persistence and strong autoregressive behavior.

In contrast, **CO_2 displays a much faster autocorrelation decay**.

The ACF decreases from near 1.0 at lag 1 to approximately 0.70 at lag 10, falls below 0.30 by lag 30, and approaches zero by lag 45. This rapid decay reflects the short-term and occupancy-driven nature of CO_2 dynamics, which respond quickly to classroom presence and ventilation events. These characteristics justify using short-horizon forecasting (1–3 hours) and emphasize the importance of recent CO_2 lags for predicting near-future concentrations.

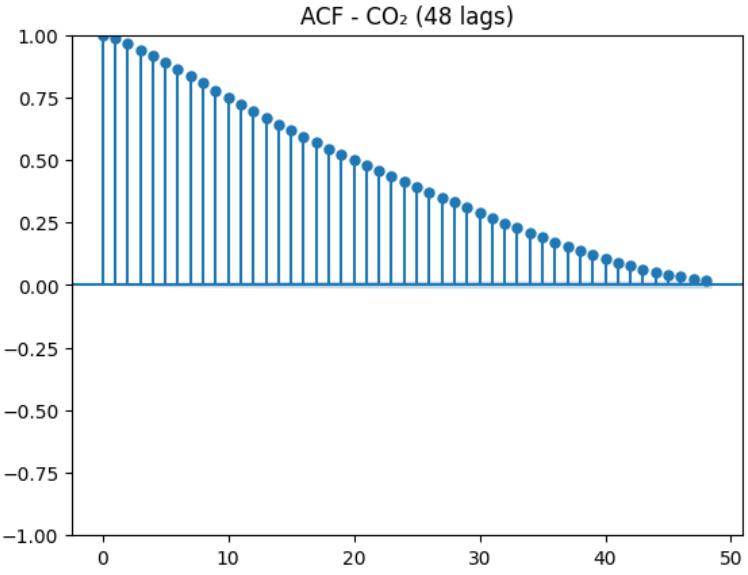

Figure 4.9 Autocorrelation function (ACF) of indoor CO_2 concentrations. The rapid decay reflects short-term occupancy-driven variability.

Together, the ACF results highlight the fundamentally different temporal behaviors of the two variables: temperature is highly persistent with long-term memory, whereas CO_2 is transient and dominated by short-term fluctuations.

4.3 Feature Engineering Verification

4.3.1 Most Informative Features (MI-based Selection)

To quantitatively assess the effectiveness of the engineered features, a mutual information (MI) analysis was carried out for the 1-hour-ahead temperature target. All numerical predictors derived in the feature-engineering pipeline (lags, rolling statistics, and contextual variables) were included, and the 20 features with the highest MI scores are reported in Table 4.1.

Table 4.1 Top 20 most informative features for the 1-hour temperature forecast, ranked by mutual information (MI)

Rank	Feature	MI
1	temp_mean_lag1	2.60
2	temp_mean_roll_1h_mean	2.59
3	temp_mean_lag2	2.48
4	temp_mean_lag3	2.37
5	temp_mean_lag4	2.27
6	temp_mean_roll_3h_mean	2.23
7	temp_mean_lag8	1.98
8	temp_mean_roll_6h_mean	1.88
9	temp_mean_lag12	1.78
10	temp_mean_roll_24h_mean	1.59
11	temp_mean_lag24	1.42
12	temp_mean_lag48	1.17
13	setpoint	0.85
14	temp_out	0.79
15	month	0.79
16	month_cos	0.46
17	month_sin	0.42
18	co2_roll_24h_mean	0.31
19	co2_lag1	0.22
20	co2_lag2	0.22

The results clearly show that short-term lagged and rolling statistics of indoor temperature dominate the ranking. The first positions are consistently occupied by temp_mean_lag1, temp_mean_lag2, temp_mean_lag3, temp_mean_lag4, together with temp_mean_roll_1h_mean and temp_mean_roll_3h_mean. Longer lags and aggregates (temp_mean_lag8, temp_mean_lag12, temp_mean_lag24, temp_mean_lag48, and the 6-h and 24-h rolling means) also exhibit substantial MI values, confirming that both very recent history and multi-hour thermal inertia contribute relevant predictive information. This pattern empirically supports the choice of lag set {1, 2, 3, 4, 8, 12, 24, 48} and multi-scale rolling windows (1 h, 3 h, 6 h, 24 h) adopted in Chapter 3.

Among the contextual variables, setpoint and outdoor temperature (temp_out) show non-negligible MI scores, indicating that HVAC control actions and external climatic conditions meaningfully influence the future indoor temperature. Seasonal calendar descriptors (month, month_sin, month_cos) also appear in the top-20 list, reflecting slower seasonal drifts in the building's thermal regime.

Finally, CO_2-based features (e.g., co2_roll_24h_mean, co2_lag1, co2_lag2) achieve lower but still positive MI values. This suggests that occupancy-related internal gains, as encoded by CO_2 dynamics, provide additional, but secondary, information for temperature forecasting compared to direct thermal history.

The MI analysis confirms that the engineered lag and rolling features, together with a compact set of contextual and calendar variables, form a highly informative predictor space and validates the feature-engineering design choices presented in the methodology chapter.

4.3.2 Cross-zone Variability in Selected Features

To investigate whether different rooms within Aule R exhibit distinct thermal behaviors, mutual information (MI) was computed separately for each zone. Table 4.2 summarizes the three most informative features for the 1-hour-ahead temperature forecast in all eight monitored zones.

Table 4.2 Top three most informative features for the 1-hour temperature forecast across all monitored zones, ranked by mutual information (MI).

zone	top1_feature	top1_MI	top2_feature	top2_MI	top3_feature	top3_MI
R1	temp_mean_lag1	2.492	temp_mean_roll_1h_mean	2.487	temp_mean_lag2	2.363
R2	temp_mean_lag1	2.545	temp_mean_roll_1h_mean	2.542	temp_mean_lag2	2.42
R3	temp_mean_roll_1h_mean	2.453	temp_mean_lag1	2.45	temp_mean_lag2	2.33
R4	temp_mean_lag1	2.477	temp_mean_roll_1h_mean	2.456	temp_mean_lag2	2.347
R1B	temp_mean_roll_1h_mean	2.756	temp_mean_lag1	2.706	temp_mean_lag2	2.604

R2B	temp_mean_roll_1h_mean	2.355	temp_mean_lag1	2.283	temp_mean_lag2	2.197
R3B	temp_mean_roll_1h_mean	2.557	temp_mean_lag1	2.506	temp_mean_lag2	2.395
R4B	temp_mean_roll_1h_mean	2.52	temp_mean_lag1	2.474	temp_mean_lag2	2.377

Across all zones, the results reveal a consistent and dominant reliance on short-term thermal history, with temp_mean_lag1, temp_mean_lag2, and the 1-hour rolling mean (temp_mean_roll_1h_mean) appearing systematically among the top-ranked predictors. This uniformity highlights that temperature evolution in Aule R is primarily governed by strong autoregressive dynamics and short-term thermal inertia, regardless of specific room configuration or HVAC node.

Some degree of cross-zone variability does however emerge. In R1, R2, and R4, the first-order lag (temp_mean_lag1) is the single most informative feature, suggesting a more immediate thermal responsiveness to short-term fluctuations. Conversely, in R1B, R2B, R3B, R4B, and especially R1B, the 1-hour rolling mean exhibits the highest MI score, indicating smoother temperature profiles with more substantial averaging effects, likely due to differences in HVAC exposure, airflow distribution, or thermal buffering within these rooms.

Notably, contextual variables (e.g., setpoint, outdoor temperature) and CO_2 features do not appear among the top predictors for any zone, reinforcing the conclusion that recent indoor temperature history overwhelmingly explains short-term temperature dynamics. At the same time, other signals provide secondary or indirect contributions.

The cross-zone comparison confirms that the feature-engineering strategy is robust and generalizable across the building. At the same time, the subtle inter-zone differences offer meaningful insights into room-level thermal characteristics and HVAC behavior.

4.4 Forecasting Performance of Machine Learning Models

This section evaluates the forecasting performance of the three implemented models—Random Forest (RF), Convolutional Neural Network (CNN), and Long Short-Term Memory (LSTM)—across all building zones and for both target variables: indoor temperature (temp_mean) and CO_2 concentration. The analysis considers two forecasting horizons: 1 hour (1h) and 3 hours (3h).

The forecasting results for indoor temperature and CO_2 concentration reveal clear, consistent patterns across models, horizons, and building zones. By examining both aggregated metrics and zone-level behavior (using MAE, RMSE, and a set of comparative plots), the distinctive characteristics of each model and variable become evident. This section provides a unified interpretation of the forecasting performance, drawing on all the supporting visualizations.

Table 4.3 reports the forecasting accuracy of the Random Forest (RF) and sequence models (CNN/LSTM) for indoor air temperature and CO_2 concentration at two predictive horizons (1 h and 3 h). Across all metrics, RF consistently outperforms the sequence models, achieving substantially lower MAE and RMSE values.

Table 4.3 Forecasting accuracy of Random Forest and sequence models (CNN/LSTM) for temperature and CO_2 across 1-hour and 3-hour horizons.

Model	Target	Horizon	MAE	RMSE	sMAPE
RF	temp_mean	1h	0.458	0.760	1.734
RF	temp_mean	3h	0.753	1.112	2.899
RF	CO_2	1h	30.60	57.91	5.48
RF	CO_2	3h	53.90	93.61	9.57
CNN/LSTM	temp_mean	1h	1.497	2.049	5.924
CNN/LSTM	temp_mean	3h	1.647	2.177	6.531
CNN/LSTM	CO_2	1h	43.25	92.06	7.58
CNN/LSTM	CO_2	3h	61.88	114.72	11.02

4.4.1 Temperature Forecasting Performance (Main Target)

1-hour ahead forecasting

Across all zones, the Random Forest (RF) model demonstrates the highest accuracy for short-term temperature prediction. According to the aggregated metrics table, RF achieves an MAE of 0.46 °C, which is considerably lower than the best-performing sequence model (CNN/LSTM, MAE ≈ 1.5 °C). This large performance gap is also reflected in the comparative bar charts for the 1-hour horizon Figure4.10, where the RF bar is visibly shorter than those of the deep learning models.

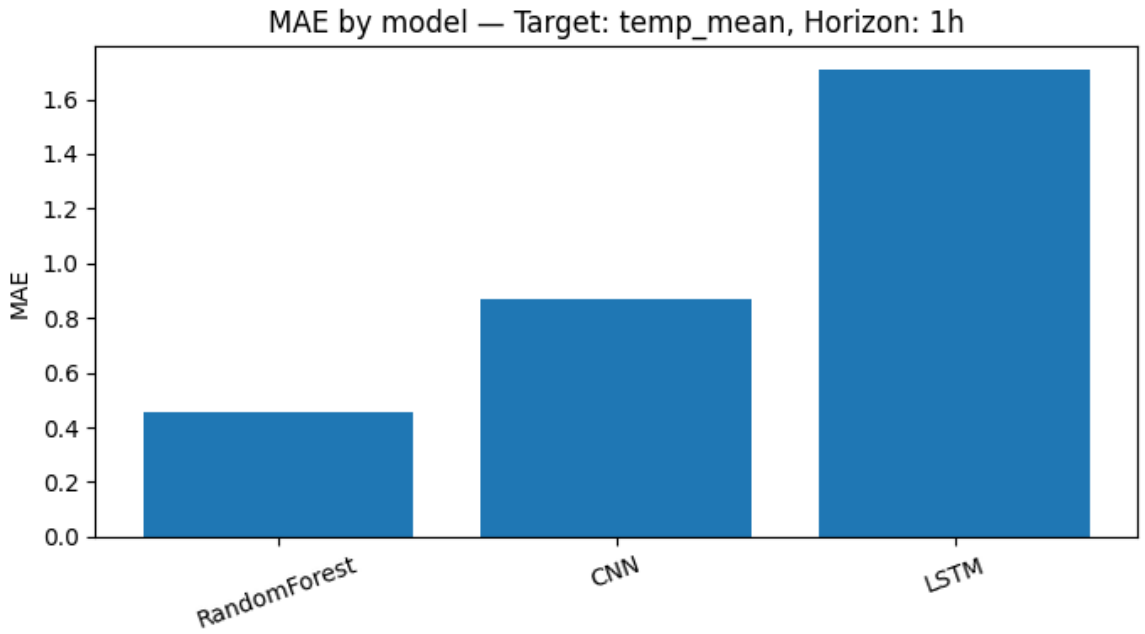

Figure 4.10 Comparison of 1-hour temperature forecasting errors (MAE) across Random Forest, CNN, and LSTM models.

This behavior is further clarified when looking at the zone-level breakdown in Figure 4.11. In every eight classrooms, RF consistently maintains a low MAE, typically between 0.35 °C and 0.60 °C. CNN and LSTM, on the other hand, show more fluctuation across rooms, and their errors are noticeably higher in all cases.

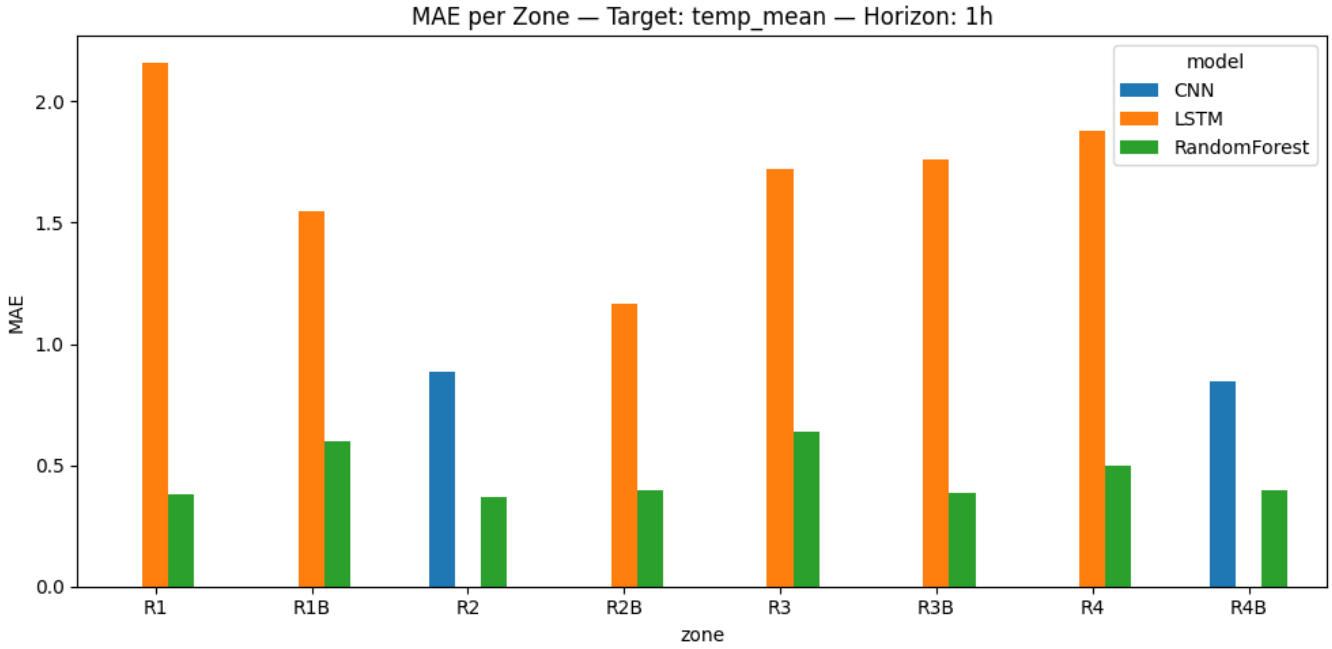

Figure 4.11 Zone-level 1-hour temperature forecasting errors (MAE) for Random Forest, CNN, and LSTM models across all monitored rooms.

Scatter plots also help visualize the structural difference between the models. The RF scatter plot, Figure 4.12, shows that predictions cluster tightly around the diagonal, indicating a very close match between

predicted and actual temperatures. The CNN/LSTM scatter plot in Figure 4.13 indicates much greater dispersion, especially at higher temperatures, which explains why their MAE values are consistently higher.

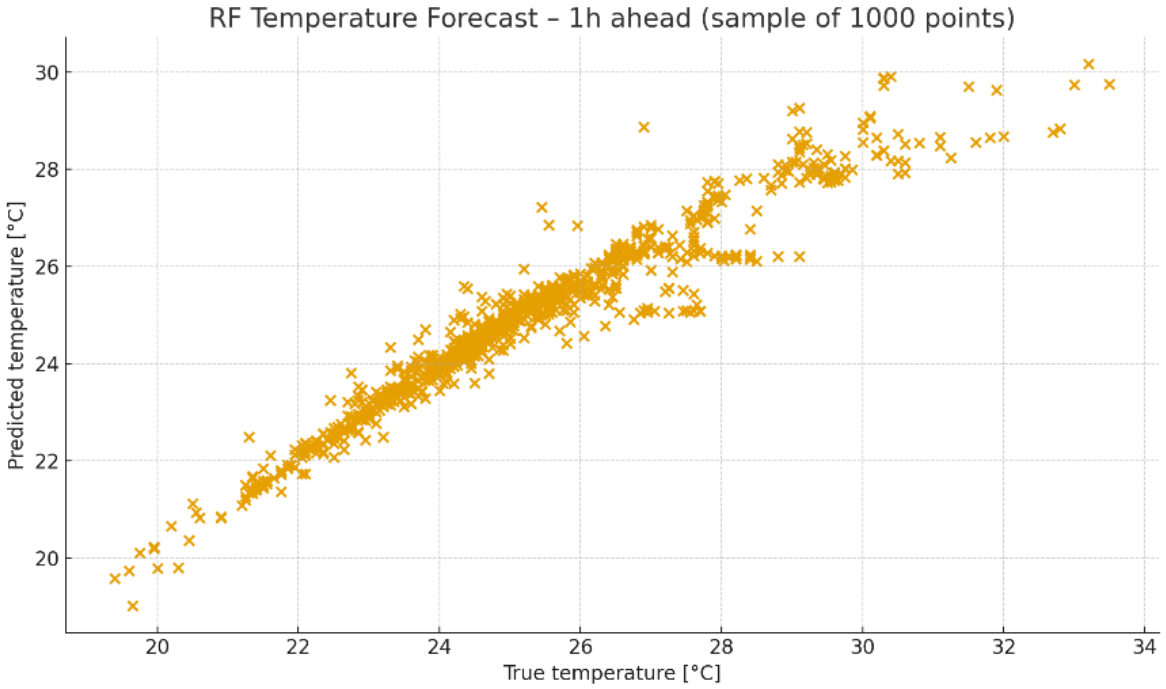

Figure 4.12 Scatter plot of 1 h ahead RF temperature forecasts vs measured values (sample of 1000 points)

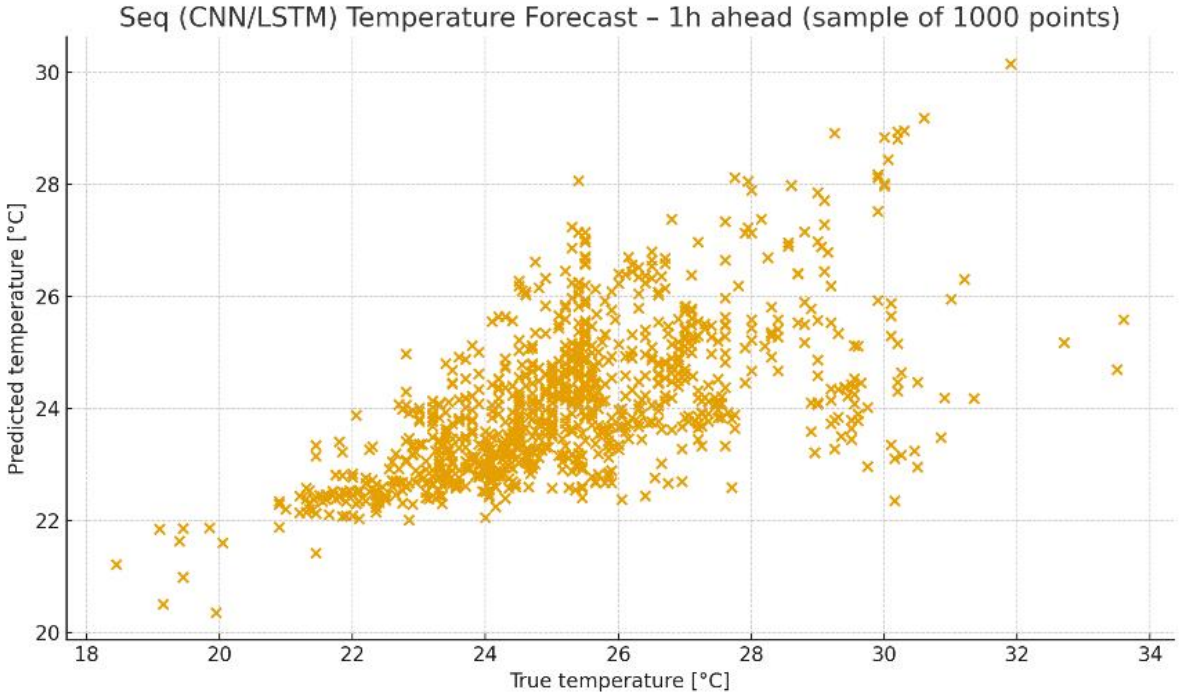

Figure 4.13 Scatter plot of 1 h ahead CNN/LSTM temperature forecasts vs measured values (sample of 1000 points)

The high performance of RF at the 1-hour horizon reflects the strong short-term persistence of indoor temperature, which was indicated by the autocorrelation analysis in Section 4.2.5. Since lagged features capture most of this temporal structure, RF can exploit them directly and effectively.

3-hour ahead forecasting

At the 3-hour horizon, errors naturally increase for all models, yet RF again remains the most stable and accurate method. The aggregated MAE for RF rises to 0.75 °C, whereas CNN and LSTM reach approximately 1.45–1.65 °C. This trend is evident in the comparison plot in Figure 4.14, where the error bars for CNN/LSTM remain considerably higher.

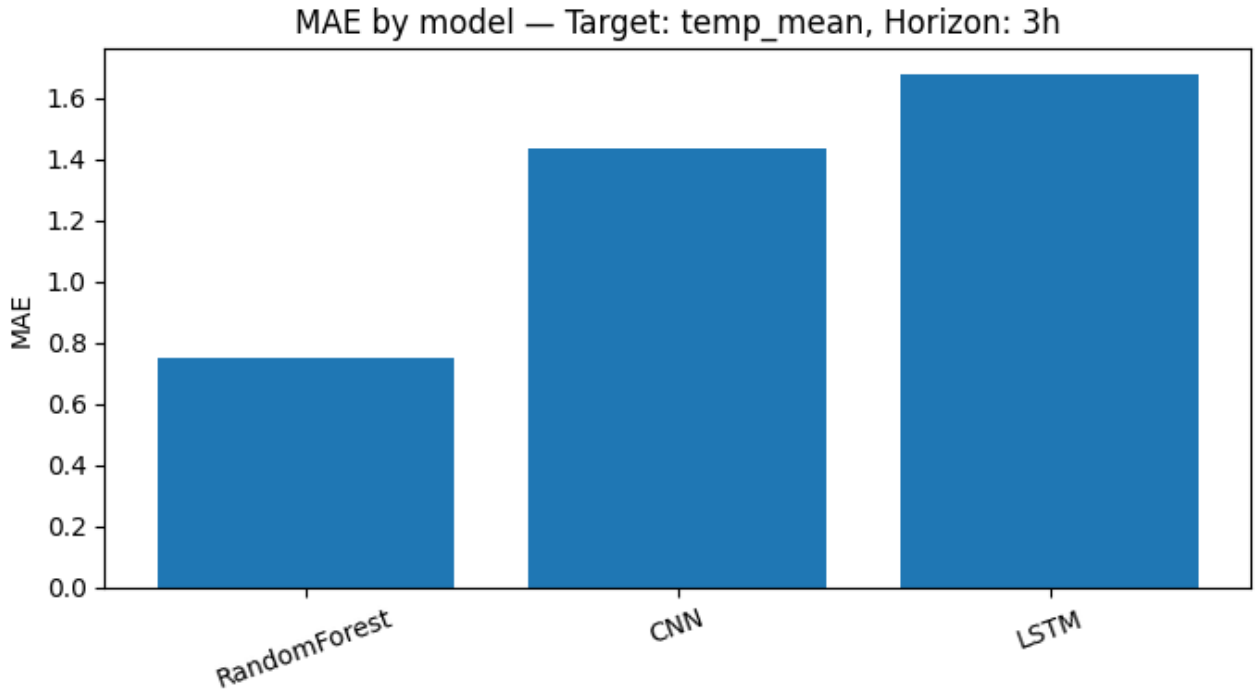

Figure 4.14 Comparison of 3-hour temperature forecasting errors (MAE) for Random Forest, CNN, and LSTM models.

Zone-level results display the exact pattern, Figure 4.15. Even in the more challenging longer horizon, RF maintains an MAE typically below 0.85 °C in most rooms, while LSTM often exceeds 2.0 °C. RMSE patterns reinforce this observation Figure 4.16, where RF shows the lowest variance and the smoothest error profile among all models.

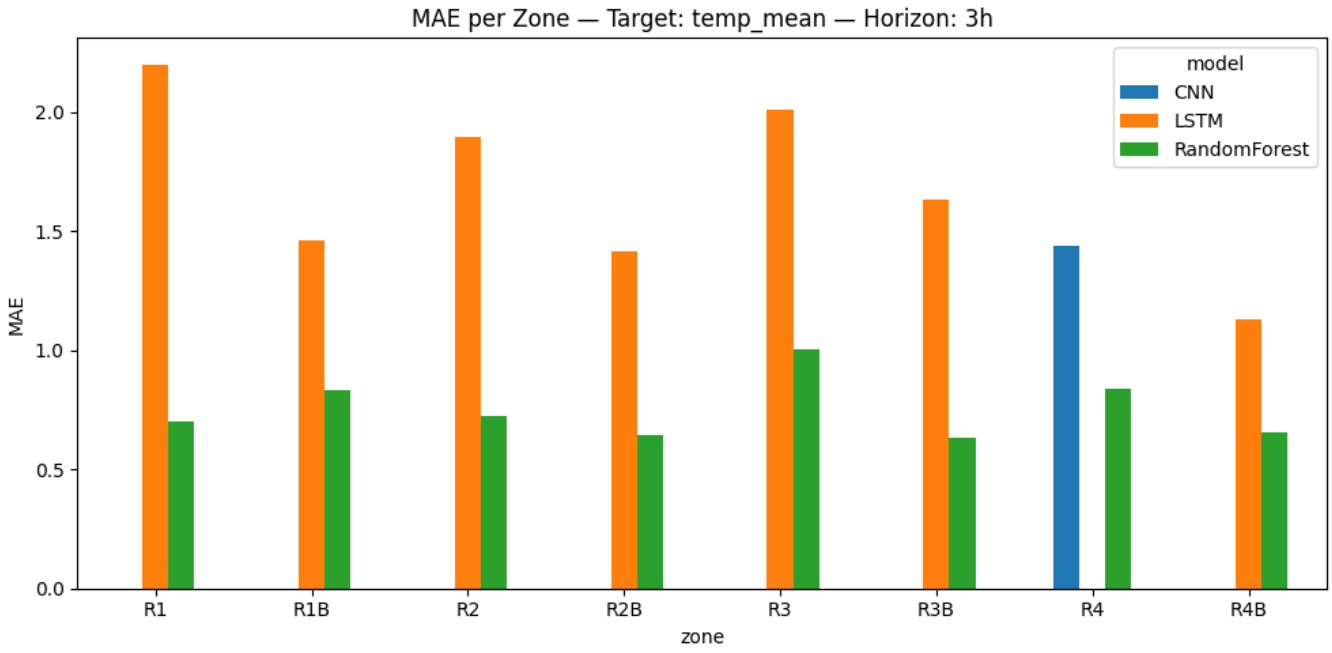

Figure 4.15 Zone-level 3-hour temperature forecasting errors (MAE) for Random Forest, CNN, and LSTM models across all monitored rooms.

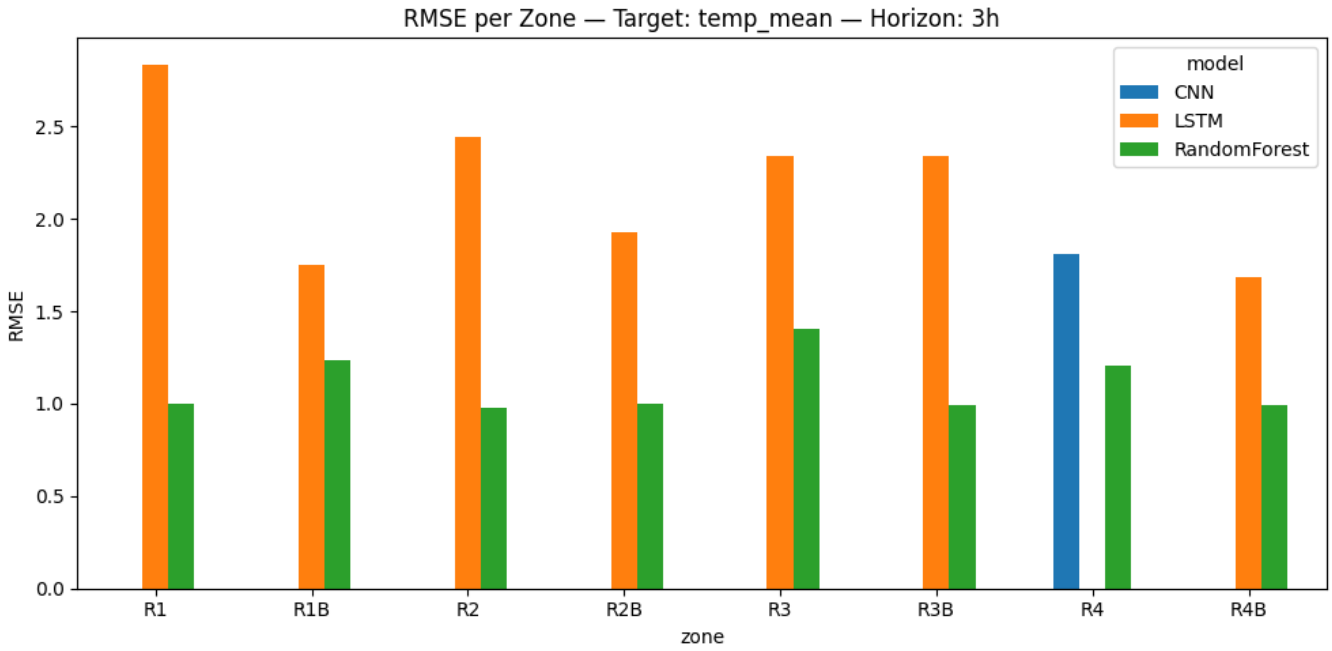

Figure 4.16 Zone-level 3-hour temperature forecasting errors (RMSE) for Random Forest, CNN, and LSTM models across all monitored rooms.

The “Forecast Error Vs Horizon” plot, Figure 4.17, illustrates how gracefully RF performance degrades compared to CNN/LSTM. Whereas the RF curve increases only mildly between 1h and 3h, the deep learning curve starts high at 1h. It shows only a marginal additional increase, indicating that the deep

models fail to capture the fine-grained short-term dynamics of the thermal environment and perform substantially worse overall.

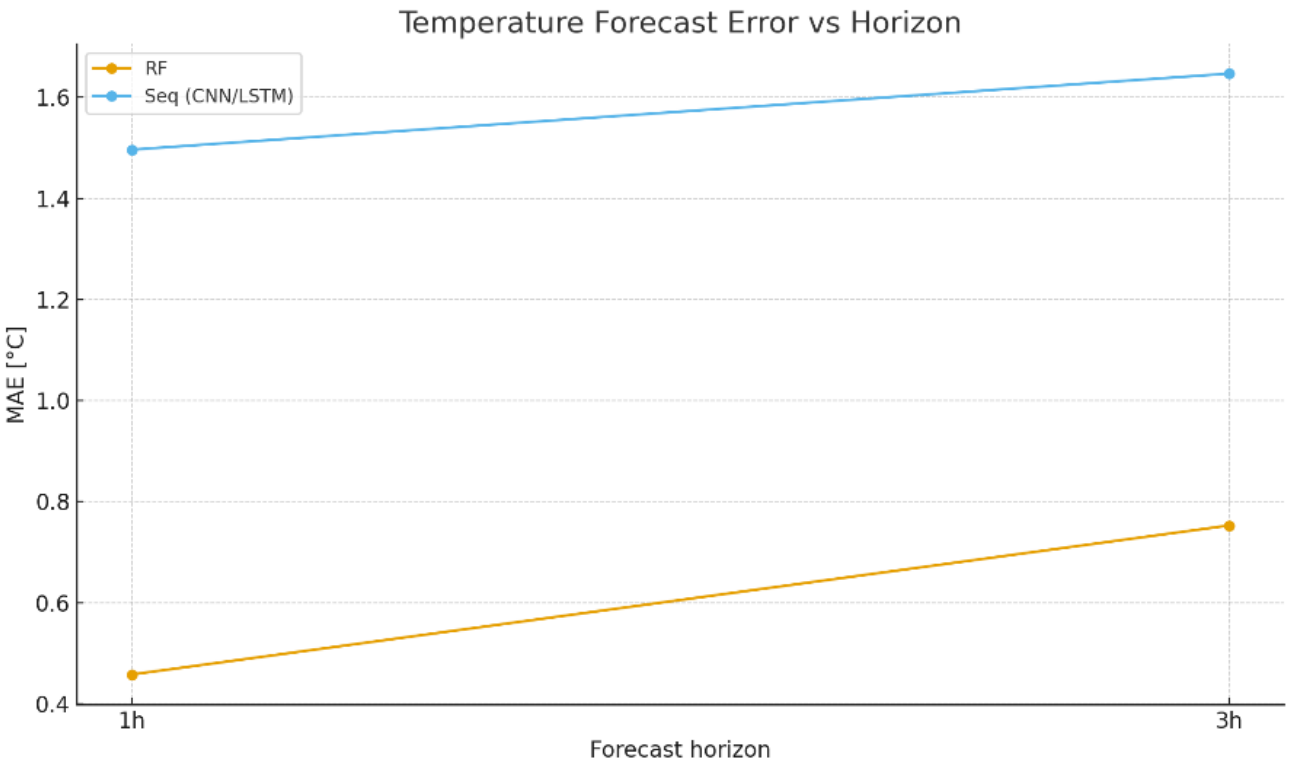

Figure 4.17 Evolution of MAE with forecast horizon (1 h vs 3 h) for RF and CNN/LSTM temperature models.

In summary, both the tabular and graphical evidence clearly show that RF provides highly reliable temperature forecasts in both 1h and 3h horizons, with errors consistently under 1 °C. This level of accuracy is notable given the complex thermal behavior arising from mixed HVAC influences, envelope inertia, and occupancy-related heat gains.

4.4.2 CO_2 Forecasting Performance (Secondary Target)

1-hour ahead forecasting

Forecasting CO_2 concentration is inherently more challenging because it depends on occupancy, which is not directly available in the dataset. Despite this limitation, RF again delivers the most accurate short-horizon predictions. The aggregated metrics report an MAE of 30.6 ppm for RF, compared to 43.3 ppm for CNN/LSTM. The corresponding bar chart, Figure 4.18, visually confirms this difference.

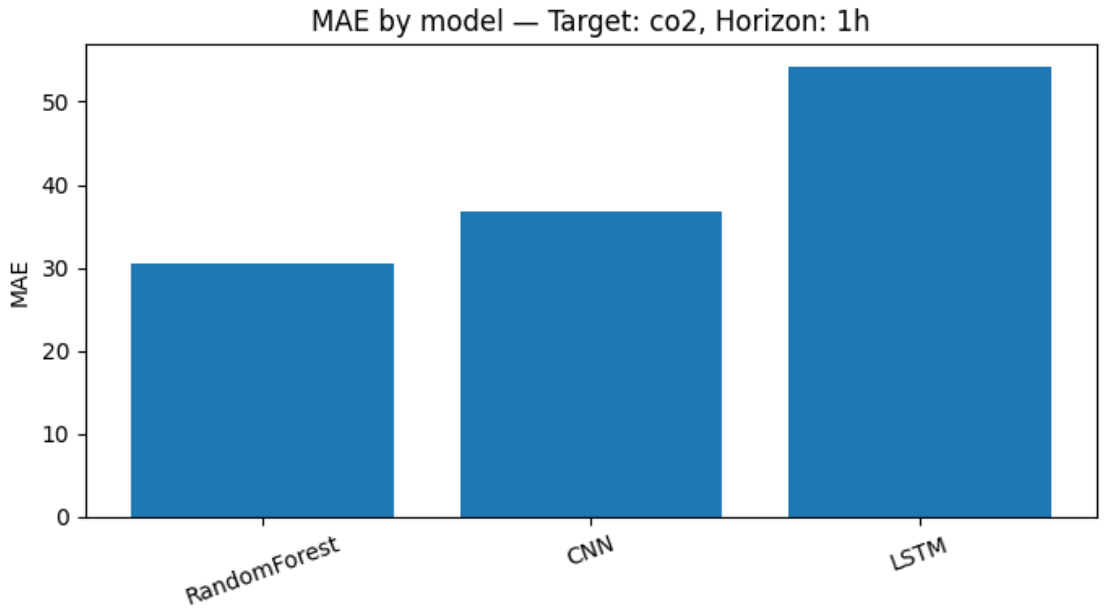

Figure 4.18 Comparison of 1-hour CO_2 forecasting errors (MAE) for Random Forest, CNN, and LSTM models.

Zone-level plots, Figure 4.19, highlight a wider variability in CO_2 forecasting error than in temperature. Certain rooms (notably R3 and R1B) experience higher spikes in error for all models. These zones are the largest classrooms and tend to exhibit more dynamic occupancy patterns, which explains their more volatile CO_2 profiles. Even in these difficult rooms, RF maintains the lowest MAE, whereas LSTM often exceeds 60 ppm.

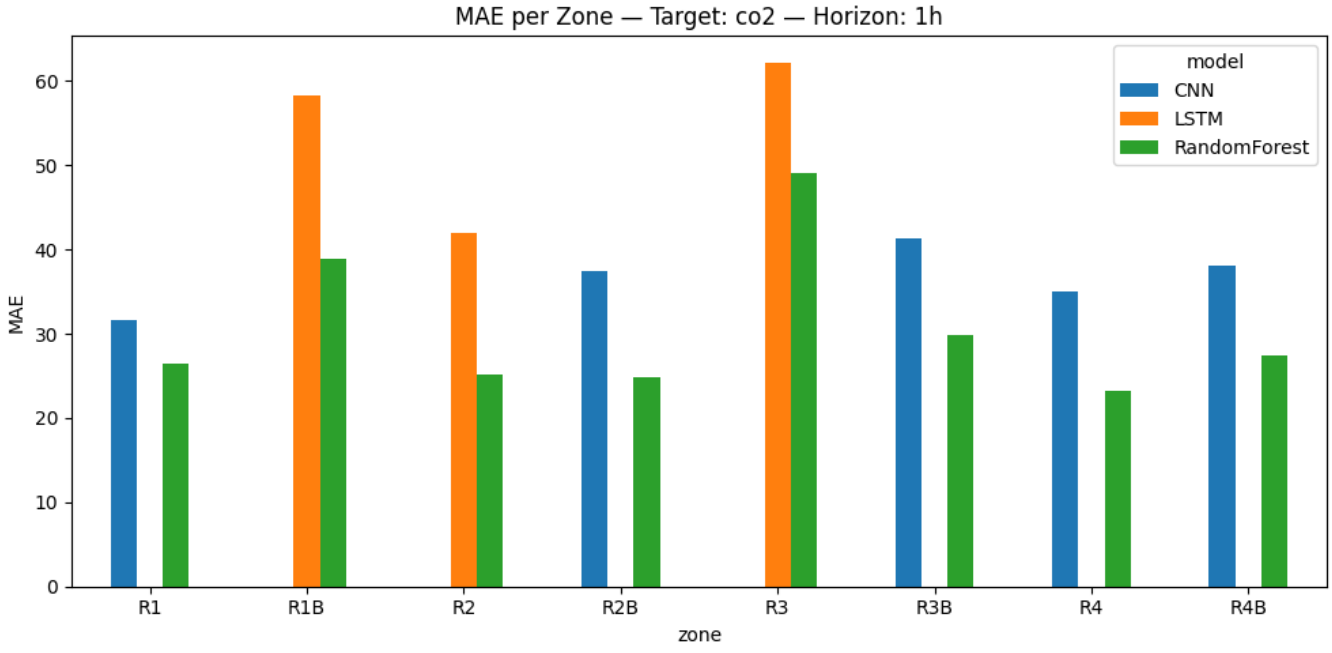

Figure 4.19 Zone-level 1-hour CO_2 forecasting errors (MAE) for Random Forest, CNN, and LSTM models across all monitored rooms.

3-hour ahead forecasting

The challenge of CO_2 prediction becomes even more apparent at the 3-hour horizon. All models experience a significant increase in error, but RF continues to perform best. The average MAE for RF increases to 53.9 ppm, while CNN/LSTM reach 61.9 ppm and often exceed this value in individual rooms. The bar chart Figure 4.20 clearly reflects this horizon-dependent degradation.

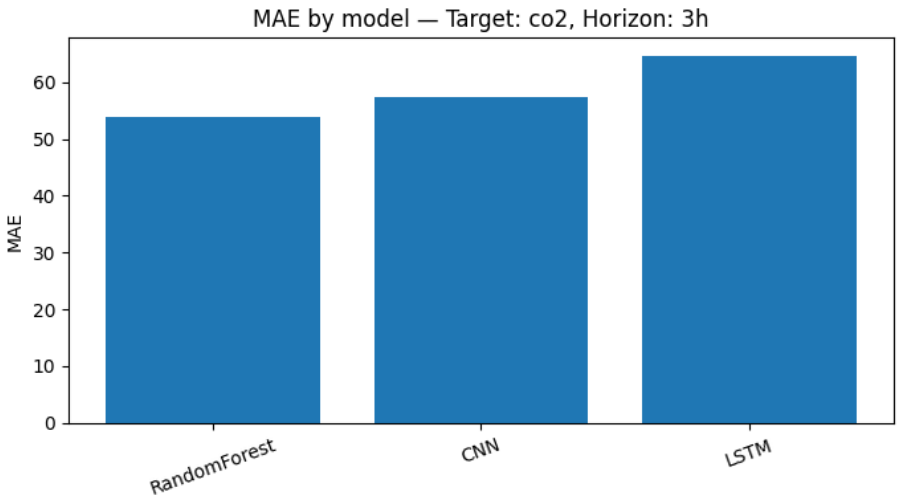

Figure 4.20 Comparison of 3-hour CO_2 forecasting errors (MAE) for Random Forest, CNN, and LSTM models.

The zone-level 3h chart Figure 4.21 indicates that error growth is not uniform across classrooms: zones with previously high volatility (e.g., R3, R1B) exhibit the greatest increases in error at longer horizons. Basement rooms tend to maintain slightly lower CO_2 variability, which is consistent with their lower occupancy density.

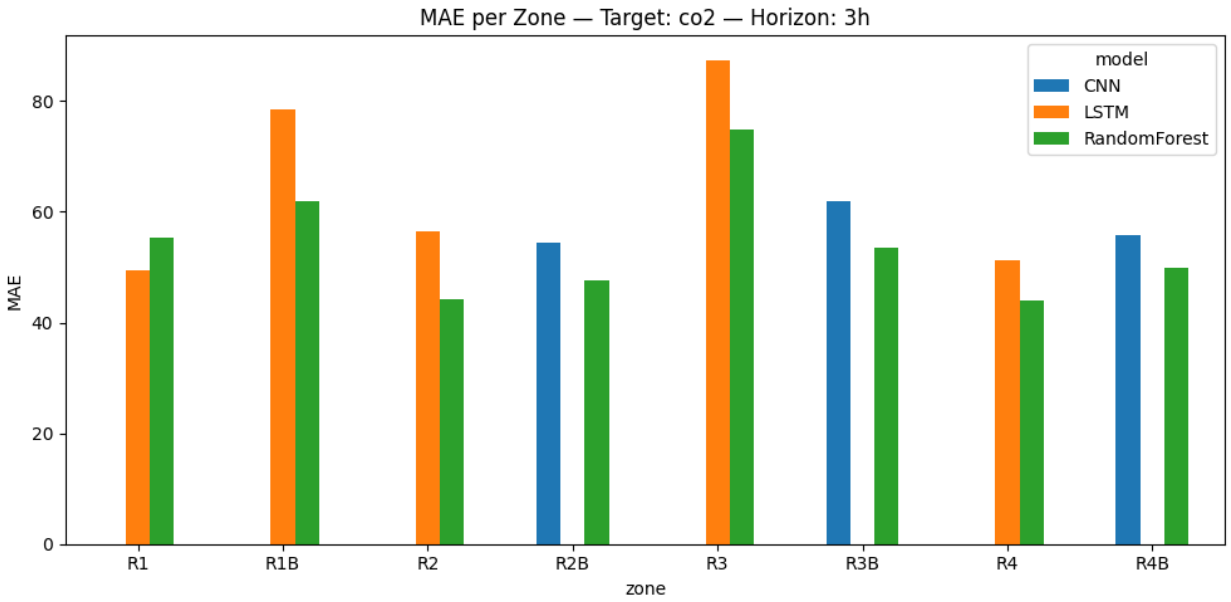

Figure 4.21 Zone-level 3-hour CO_2 forecasting errors (MAE) for Random Forest, CNN, and LSTM models across all monitored rooms.

These patterns also agree with the autocorrelation analysis for CO_2 in Section 4.2.5, which showed a rapid decay of temporal correlation within the first 30–45 minutes. This explains why the 3-hour forecasts deteriorate more sharply for CO_2 than for temperature.

The "CO_2 Forecast – MAE by Model and Horizon" plot Figure 4.22 summarizes how prediction accuracy changes between the 1h and 3h horizons. Similar to temperature forecasting, RF shows a smoother and more controlled increase in error, rising from about 31 ppm at 1h to 54 ppm at 3h. In contrast, the sequence models start from a higher error level and continue to diverge at the longer horizon, reaching approximately 62 ppm. This pattern highlights the difficulty of forecasting CO_2 (due to its rapid fluctuations and occupancy-driven behavior) and emphasizes the advantage of RF in short-term prediction.

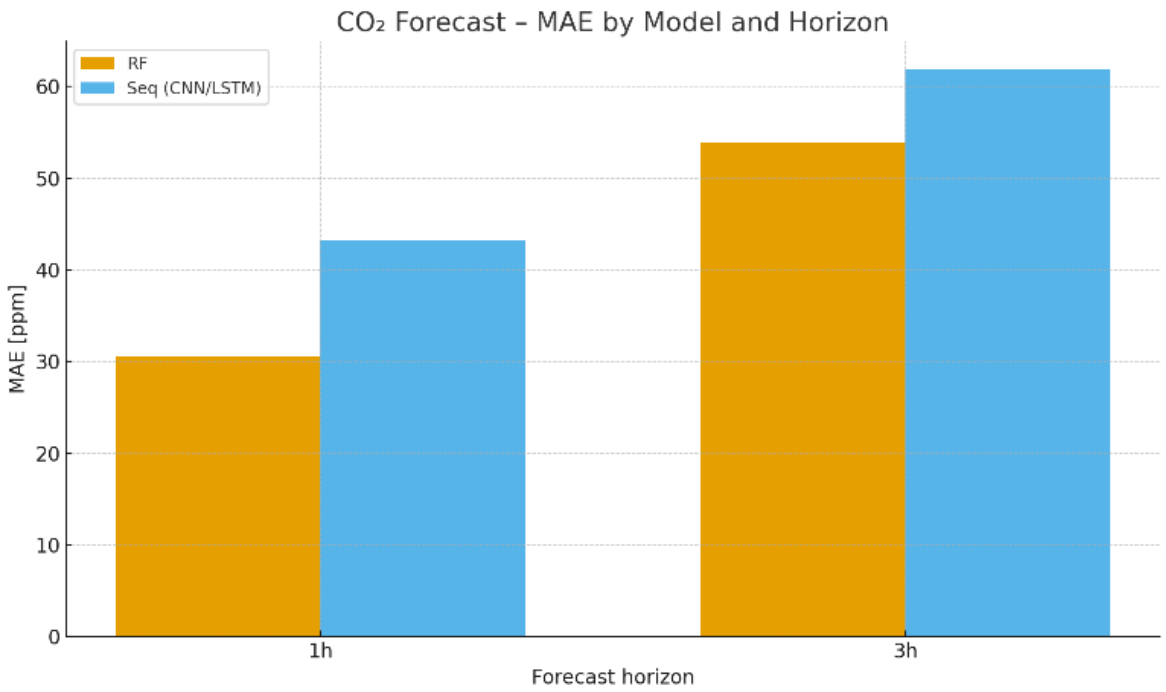

Figure 4.22 Mean absolute error (MAE) for 1 h and 3 h ahead CO_2 forecasts for RF and CNN/LSTM models.1

CO_2 forecasting remains more challenging than temperature prediction, but RF consistently provides the most reliable short-term performance, maintaining a clear error margin over the deep learning models across both horizons.

4.4.3 Why Random Forest Outperformed Sequence Models

The superior performance of the Random Forest model can be understood by looking at both the characteristics of the data and the way temporal information was represented in this study. Indoor temperature in Aule R changes gradually and has strong short-term persistence. Because of this, the lagged and rolling features created during feature engineering already capture most of the relevant temporal dynamics. Random Forest is particularly effective with such structured, tabular inputs, allowing it to turn these features into highly accurate forecasts.

In contrast, the CNN and LSTM models must learn temporal patterns directly from short raw sequences. Without longer histories or explicit occupancy information—particularly important for CO_2—they face a more difficult learning problem. Their predictions show greater variability and higher errors, especially as the forecasting horizon increases. Deep learning models are also more sensitive to hyperparameters and data scaling, making them less robust when the dataset is moderate in size or when specific drivers of the dynamics are excluded.

Taken together, the results across all zones and horizons indicate that the forecasting task in Aule R favors a feature-driven approach over sequence-based representation learning. Random Forest benefits directly from the strong autocorrelation of temperature and the carefully engineered inputs, whereas

CNN and LSTM cannot fully exploit their potential under these conditions. This explains why Random Forest consistently achieves lower errors than the sequence models for both temperature and CO_2.

4.5 SHAP Explainability for Random Forest

To better understand how the Random Forest model generates its temperature predictions, SHAP values were computed for all zones and then averaged to produce a global importance ranking. The resulting plot, Figure 4.23, offers a clear and intuitive picture of which signals the model relies on most strongly.

A dominant pattern emerges immediately:

the current indoor temperature (temp_mean) and its first lag (temp_mean_lag1) overwhelmingly drive the 1-hour-ahead forecast. This is fully consistent with the strong thermal persistence observed earlier in the autocorrelation analysis. In practice, this means that the building's temperature in the next hour is largely determined by where it stands right now and where it was just a short moment before.

Following these two core predictors, the 1-hour rolling mean (temp_mean_roll_1h_mean) also shows substantial influence. Its importance highlights the smoothing effect of recent thermal history and reinforces the idea that local thermal inertia plays a central role in shaping short-term indoor conditions.

The SHAP ranking also assigns meaningful weight to the individual sensor readings (temp_z1, temp_z2). Their presence among the top features indicates that small temperature variations within the room (caused perhaps by stratification, differences in airflow, or proximity to occupants) carry useful predictive content that the model can exploit.

Higher-order lags, and multi-scale rolling windows (e.g., lag2, lag3, lag4, lag8, rolling_3h, rolling_6h) contribute progressively less but still play a supporting role. These patterns reflect slower temperature dynamics and the influence of accumulated heat gains and HVAC effects across several hours.

Contextual drivers such as outdoor temperature (temp_out) and setpoint appear further down the list. While they do matter, their lower SHAP values indicate that short-term indoor temperature evolution is more dominated by internal conditions than by external ones. This agrees with the physical behavior of conditioned academic spaces, where HVAC systems tend to buffer indoor air from rapid outdoor fluctuations.

Finally, calendar-based features (e.g., month, hour_cos) show only minor effects. Their low SHAP values suggest that, at the 1-hour forecasting horizon, seasonal or time-of-day signals add relatively little beyond what is already encoded in the recent temperature history.

Overall, the SHAP analysis provides strong validation of the feature-engineering choices outlined in Chapter 3. The model's behavior aligns closely with physical intuition: indoor temperature forecasts depend heavily on very recent conditions, complemented by short-term averages and a modest contribution from longer-term thermal memory and contextual factors. The clarity and consistency of

these results help build confidence in the Random Forest model as both an accurate and interpretable forecasting tool.

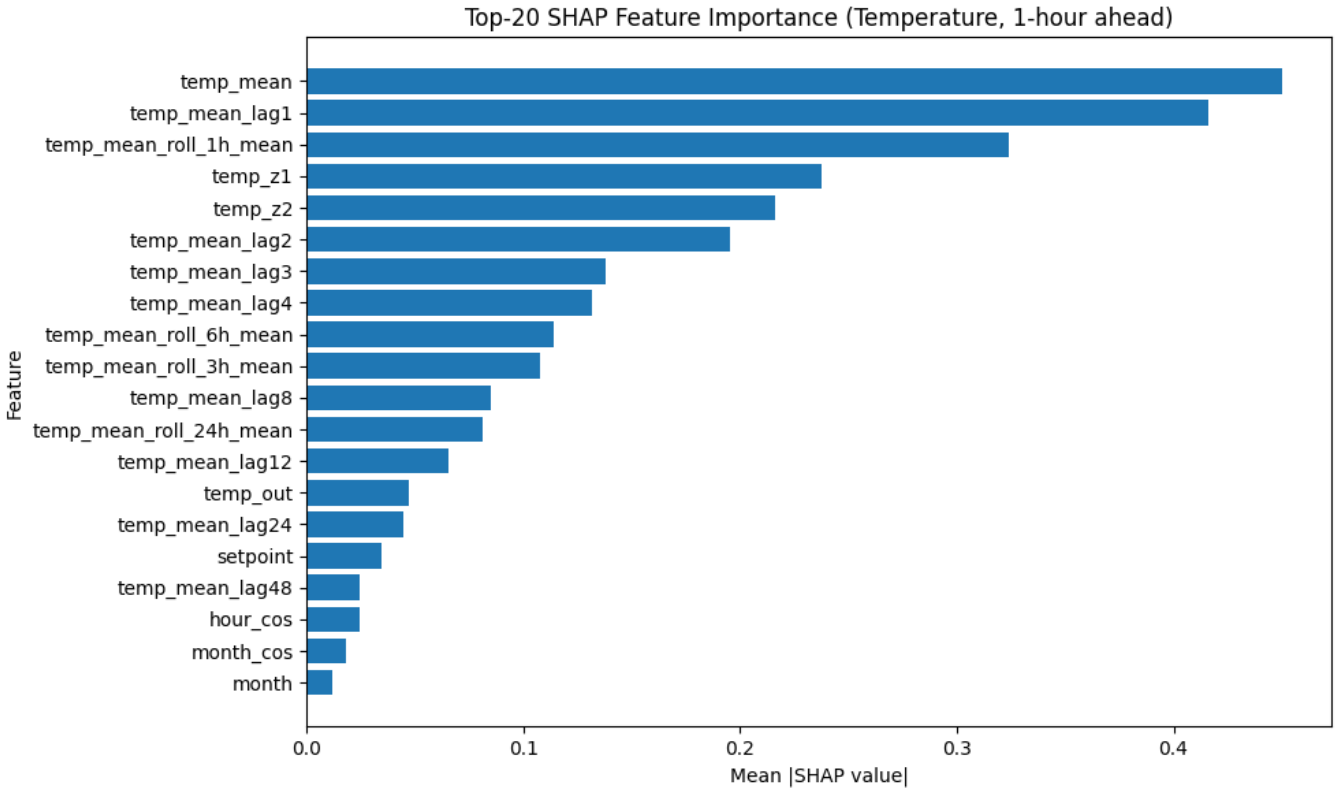

Figure 4.23 Top 20 most influential features for the 1-hour temperature forecast based on SHAP importance values from the Random Forest model.

4.6 Uncertainty Quantification

To better understand the reliability of the forecasting models, uncertainty intervals were estimated for both temperature and CO_2 predictions. These intervals represent the range within which the true value is likely to fall and therefore provide an additional layer of interpretability beyond point forecasts. Examining these bands helps clarify not only how accurate the models are, but also how confident they remain across different conditions, horizons, and variables.

Uncertainty analysis for temperature reveals a remarkably stable behavior. In the 1-hour-ahead temperature forecast for Zone R1, the Random Forest model produces consistently narrow prediction intervals, closely enveloping the true temperature curve throughout the day, Figure 4.24. This pattern reflects the strong persistence of indoor temperature: changes occur gradually, and the model largely relies on recent thermal history, which is well captured by the engineered lag and rolling features. Even

during moderate fluctuations in the early morning or midday periods, the confidence band remains tight, indicating that the model has a solid grasp of the underlying thermal dynamics. By contrast, the LSTM model exhibits much greater uncertainty and predicts a nearly flat temperature profile, Figure 4.25. The wide confidence band suggests that the model is unsure about the direction or magnitude of temperature changes, reinforcing the earlier conclusion that sequence models struggled to capture short-term temperature behavior in this building.

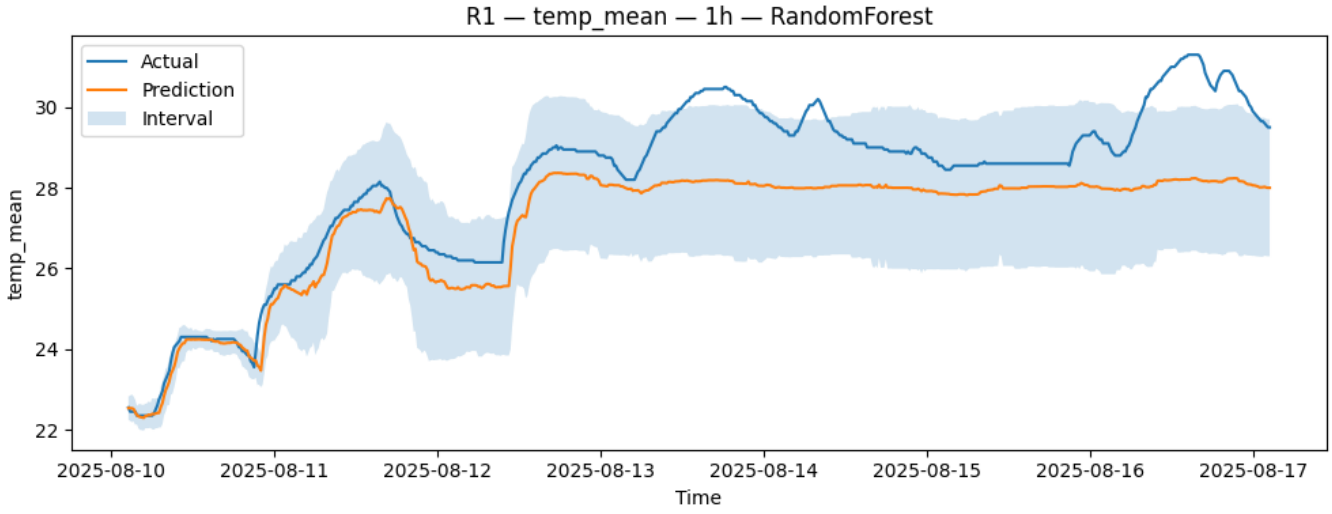

Figure 4.24 Random Forest 1-hour temperature forecast for Zone R1, showing predicted values with 95% uncertainty intervals compared to actual measurements.

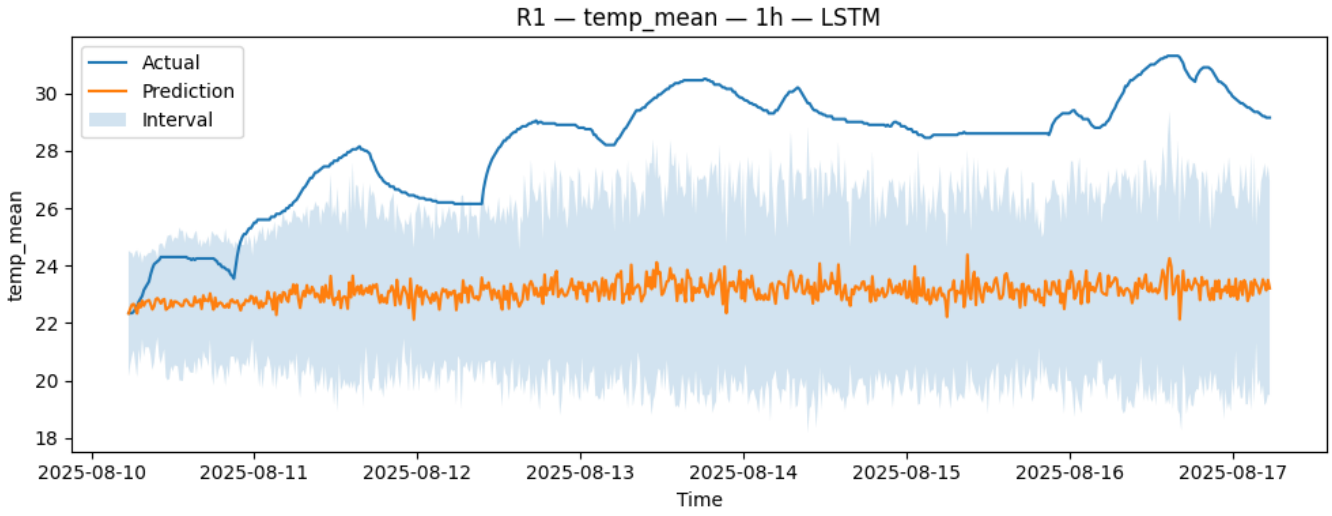

Figure 4.25 LSTM 1-hour temperature forecast for Zone R1, showing predicted values with uncertainty intervals compared to actual measurements.

A similar contrast can be observed when examining a different day with nearly flat cooling-driven temperature patterns, Figure 4.26. Even under these steady conditions, the Random Forest model retains narrow and coherent intervals, whereas the sequence model produces very wide uncertainty

bands, showing that it remains unsure even when the true signal varies only minimally. This example helps highlight that the limitations of the sequence model are not only tied to dynamic temperature changes but also appear when the signal is too stable for the model to interpret effectively.

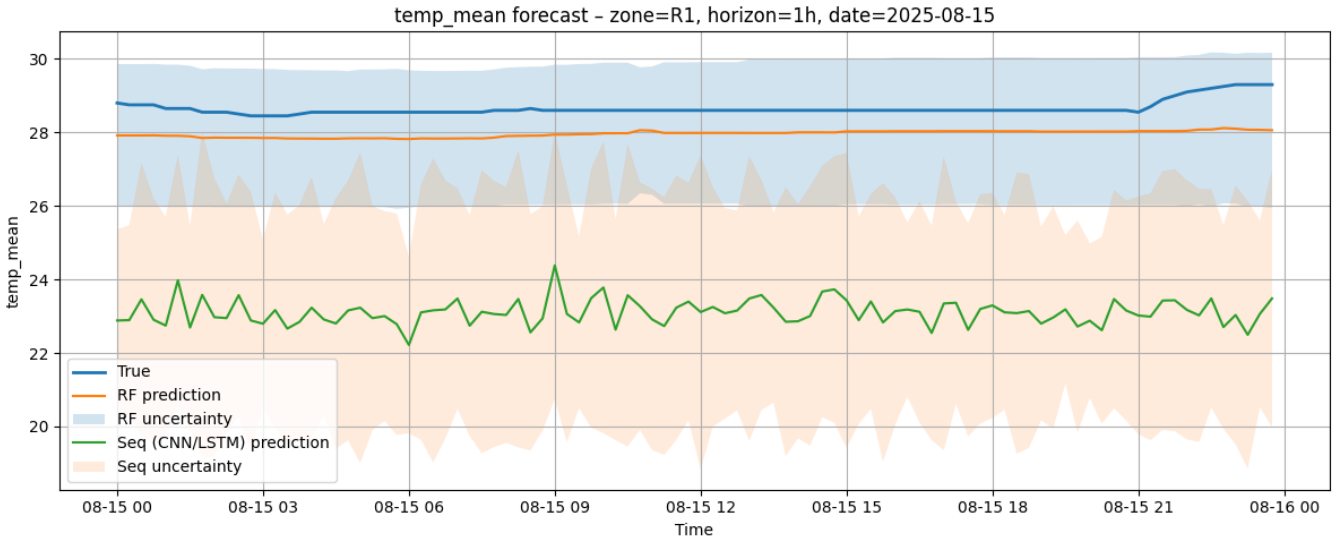

Figure 4.26 Comparison of 1-hour temperature forecasts and uncertainty intervals for Random Forest and sequence models (CNN/LSTM) in Zone R1 on 15 August 2025.

As the forecasting horizon increases, uncertainty naturally widens. In the 3-hour temperature forecast for the same zone, Figure 4.27, the Random Forest intervals expand slightly but remain coherent and informative. The model still follows the overall temperature pattern, but the greater spread reflects the reasonable increase in uncertainty when looking several hours ahead. This widening is expected and aligns with the physical nature of indoor environments, where the influence of recent conditions gradually diminishes with time. Despite this, the RF model maintains meaningful predictive structure, suggesting that it remains reliable even at longer horizons.

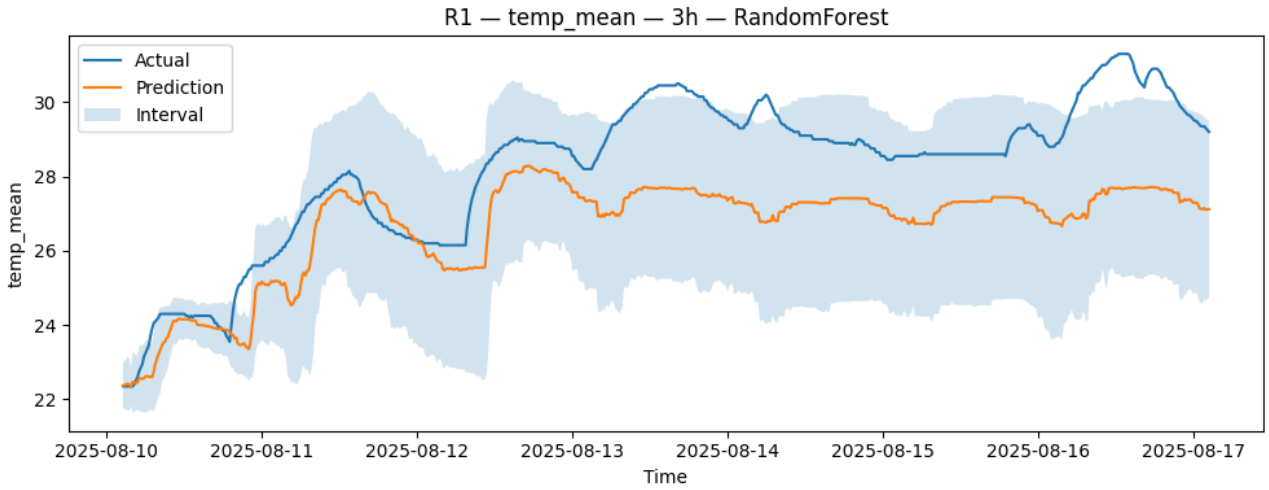

Figure 4.27 Random Forest 3-hour temperature forecast for Zone R1, showing predicted values with 95% uncertainty intervals compared to actual measurements.

Uncertainty patterns for CO_2 present a very different picture. CO_2 concentrations fluctuate far more suddenly than temperature because they are strongly affected by human occupancy and ventilation events. These rapid changes contribute to broader and more variable uncertainty intervals across all models. In the 1-hour-ahead forecast for Zone R1 on September 15, Figure 4.28, the Random Forest model still manages to capture the general shape of CO_2 variations, but its confidence band widens noticeably during periods of sudden increases or decreases. The sequence model shows even broader uncertainty intervals, particularly during sharp occupancy-driven transitions, highlighting its difficulty in anticipating CO_2 dynamics without explicit behavioral inputs.

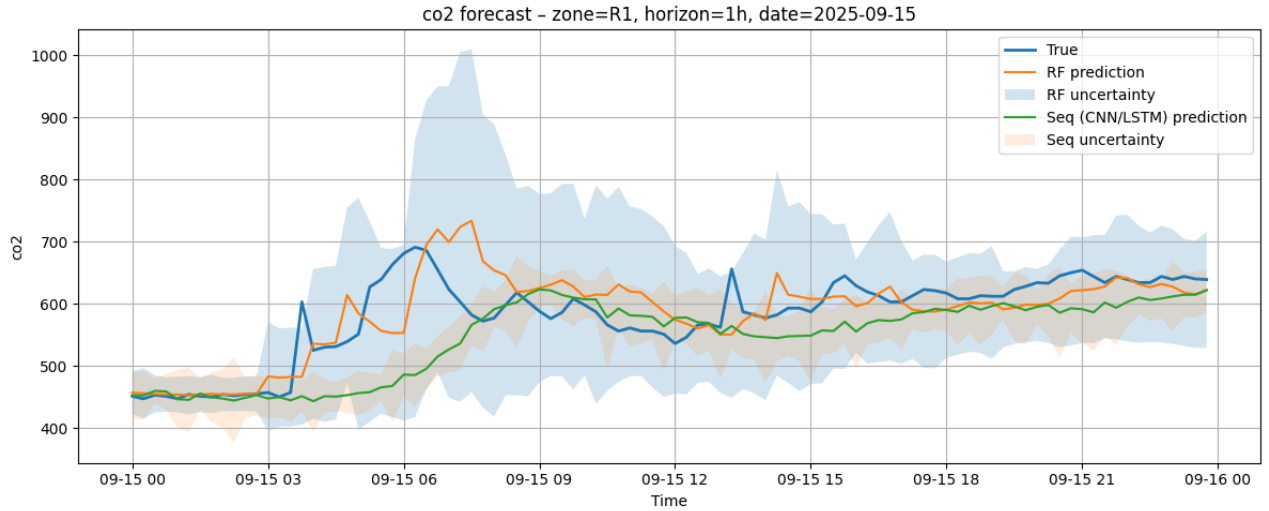

Figure 4.28 Comparison of 1-hour CO_2 forecasts and uncertainty intervals for Random Forest and sequence models (CNN/LSTM) in Zone R1 on 15 September 2025.

These observations become more pronounced at the 3-hour horizon. Both RF and LSTM show significantly wider uncertainty bands in multi-day CO_2 forecasts, reflecting the reduced predictability of this variable as the time window expands. While Random Forest maintains a recognizable structure that tracks the overall trend, the sequence model's intervals become extremely broad, suggesting low confidence across most of the forecast window. This behavior underscores the inherently stochastic nature of CO_2 dynamics and the limited ability of data-driven models to infer occupancy-driven peaks solely from past CO_2 values.

Together, these examples demonstrate the value of uncertainty quantification in contextualizing the forecasting results. Temperature forecasts exhibit high confidence and limited variability, especially over short horizons, which strengthens the case for using Random Forest as a dependable tool for short-term thermal management. CO_2 forecasts, on the other hand, inherently carry greater uncertainty due to their sensitivity to occupancy, and the corresponding intervals highlight the need for richer contextual or occupancy-related data if more reliable CO_2 predictions are desired. Across all scenarios, the Random Forest model expresses narrower and more stable uncertainty bounds than the sequence models, reinforcing its role as the most robust and trustworthy option for indoor environmental forecasting in Aule R.

4.7 BIM-enabled Data Visualization in Power BI

An important step in translating the forecasting models into a practical decision-support tool is integrating their outputs into the 3D BIM-enabled data visualization in the Power BI environment. During the internship phase, the dashboard was designed primarily to display historical sensor data, allowing users to explore temperature and CO_2 conditions across the building in a spatially intuitive way. Each room in the 3D model could be highlighted according to its measured parameter ranges, and users could switch between metrics or drill down into time-series plots for individual zones. This historical dashboard served as a proof of concept for a data-enhanced BIM model, demonstrating how sensor information can be visualized directly on top of a geometric representation of the building, figure 4.29 and figure 4.30.

Figure 4.29 BIM-Power BI dashboard developed during internship, visualizing historical data across the 3D model of Aule R.

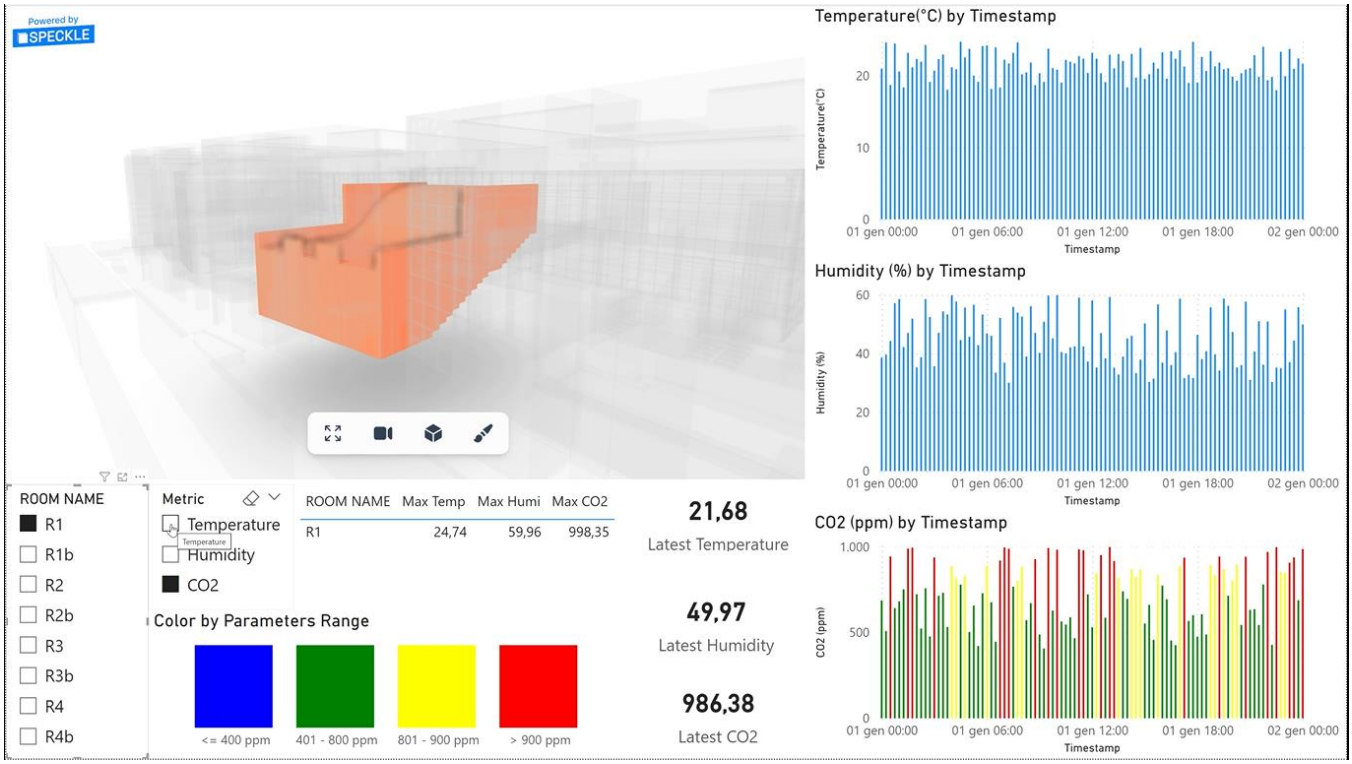

Figure 4.39 BIM-Power BI dashboard developed during internship, visualizing historical data across the 3D model of Aule R.

In the extended work for this thesis, the visualization framework is enriched by incorporating the forecasted environmental conditions generated by the machine-learning models developed in previous sections. Instead of limiting the dashboard to past measurements, it is now possible to display near-future predictions (e.g., 1-hour or 3-hour-ahead forecasts) alongside historical trends. These forecasts can be associated with the BIM geometry in the same way as the original data: each room in the 3D model can be dynamically colored based on its predicted temperature or CO_2 concentration, while the time-series plots update to show both the recent observations and the upcoming forecast interval.

This integration fundamentally shifts the purpose of the dashboard, from a retrospective monitoring tool to a forward-looking decision-support system. Building managers, HVAC operators, or researchers can visually assess not only what has happened in the past 24–48 hours, but also how indoor conditions are expected to evolve shortly. For example, zones predicted to reach higher CO_2 levels can be highlighted in advance, enabling proactive ventilation adjustments. Similarly, predicted temperature trends can inform thermal comfort management or HVAC scheduling.

The ability to visualize future conditions within the spatial context of the 3D model also strengthens the interpretability of the forecasting pipeline. Forecast anomalies, uncertainty intervals, or rapid transitions become easier to identify when viewed in combination with the building layout and room-specific characteristics.

Overall, extending the BIM-enabled data visualization in Power BI dashboard to include forecasted values represents a natural evolution of the initial internship work. By merging predictive analytics with data-driven 3D visualization, the system provides a more comprehensive understanding of the indoor environment, one that supports both monitoring and predictive insights. This enhancement demonstrates how machine learning, sensing data, and BIM visualization can be combined into a unified platform capable of supporting smarter, more informed building-management decisions.

4.8 Summary of Findings

The results of this chapter show a consistent pattern across all analyses. Indoor temperature in Aule R exhibits smooth and highly predictable behavior, which allowed the Random Forest model to achieve very strong forecasting performance, with errors remaining low even for longer horizons. CO_2, in contrast, fluctuates sharply due to occupancy and is therefore harder to predict; yet Random Forest still performed more reliably than the sequence models. SHAP analysis confirmed that short-term indoor temperature history is the dominant driver of accurate predictions, while contextual factors such as outdoor conditions play a smaller role. Uncertainty quantification further highlighted the differences between variables and models: temperature forecasts showed tight and stable confidence bands, while

CO_2 uncertainty was naturally larger. Integrating these forecasts into the BIM–Power BI dashboard transformed the visualization into a forward-looking tool, enabling users to see not only what has happened in the building, but also what is likely to occur next. Altogether, the findings demonstrate that a feature-driven machine-learning approach, supported by explainability and visualization, can effectively anticipate indoor environmental conditions and provide meaningful insights for building operations.

Chapter 5 – Conclusion

5.1 Summary of Research and Key Contributions

This thesis presented an integrated, data-driven framework for forecasting indoor environmental conditions in the Aule R building at Politecnico di Torino. Using 20 months of high-resolution (15-minute) sensor measurements of indoor air temperature and CO_2, combined with outdoor temperature and HVAC setpoint data, the study developed an analytical pipeline that spans data preprocessing, feature engineering, machine-learning modelling, uncertainty quantification, interpretability through SHAP, and 3D visualization in a BIM-linked Power BI environment. This work extends a previous internship project that focused on BIM modeling and historical data visualization in Power BI by introducing a predictive layer and connecting it to the BIM-based dashboard.

Three major methodological pillars define the contributions of this thesis:

1) **Multivariate data preparation and feature engineering**
   Twenty months of IoT sensor data were cleaned, temporally aligned, corrected for DST shifts, and enriched with multi-scale lag features, rolling-window statistics, and contextual calendar descriptors. The temporal persistence patterns, very strong for temperature and moderate for CO_2, justified the feature design and were validated through autocorrelation analysis. Feature-selection results confirmed that short-term indoor temperature history dominated predictive information, while setpoint and outdoor temperature played secondary but meaningful roles.
2) **Comparative modelling using Random Forest, CNN, and LSTM**
   Across all zones and all horizons, the three modelling approaches; Random Forest, CNN, and LSTM were systematically trained and evaluated across multiple zones and forecasting horizons. Overall, the Random Forest model emerged as the most robust and stable approach for this specific building context. Its advantage stems from the highly autoregressive nature of indoor temperature and the strong effectiveness of engineered lag features, which provide a rich representation of temporal patterns. The deep-learning sequence models, although common in time-series forecasting, were less suited to this specific building context due to limited occupancy observability, strong short-term persistence, and the relatively moderate dataset size.
3) **Explainability and BIM-linked visualization**
   SHAP analysis revealed that recent indoor temperature values, short-horizon rolling averages, and secondarily outdoor temperature were the principal drivers of RF predictions. These insights provide interpretability and strengthen trust in the model's behavior. Finally, the predictive outputs, together with uncertainty ranges, were linked to the Aule R BIM model and integrated within Power BI, transforming the existing dashboard from a purely descriptive historical viewer

into a spatially contextualized predictive decision-support tool. This expands the earlier internship work by creating a unified pipeline that merges forecasting, explainability, and 3D visualization.

The combined results demonstrate that explainable and computationally efficient machine-learning models can reliably forecast short-term indoor environmental conditions in educational buildings, laying a strong foundation for data-driven HVAC assessment.

5.2 Implications and Practical Insights

The findings of this research have several practical implications for building operation, HVAC performance assessment, and occupant comfort. The stability of short-term forecasts indicates that HVAC operation can gradually shift from reactive control toward proactive, forecast-informed strategies, particularly for regulating indoor temperature. Although CO_2 forecasting proved more challenging due to the absence of explicit occupancy data, short-term predictions remain sufficiently reliable to support ventilation-on-demand approaches, helping to anticipate periods of insufficient air exchange. Moreover, the use of SHAP interpretability provides facility managers with clear insight into how and why environmental conditions evolve, making model outputs more transparent and easier to adopt in practice. Finally, embedding predictive results into the BIM-Power BI 3D visualization enhances communication and operational awareness by allowing building operators to spatially visualize upcoming thermal and air-quality patterns at room level.

5.3 Future Development

Base on the outcomes of this work, future research can advance the predictive framework in several directions:

- **Integrating real-time occupancy data**
  Using Wi-Fi counts, camera-based anonymized occupancy, or course timetables data would significantly improve the accuracy of both temperature and CO_2 forecasts and enable predictive ventilation strategies.
- **Expanding the forecasting horizons and model diversity**
  Exploring a broader range of modelling approaches, including hybrid CNN–LSTM architectures, and transformer-based time-series models, would enable more accurate medium and long-term forecasting.
- **Developing a real-time BIM-IoT monitoring platform**
  Transitioning from descriptive visualization to a dynamic digital twin capable of continuous prediction and control optimization.
- **Deployment in additional buildings for generalizability**

Applying the workflow to other Polito buildings or similar academic environments would test its generalizability across diverse spatial layouts, thermal characteristics, and usage patterns.

This thesis demonstrated that combining IoT sensor data, explainable machine-learning models, and BIM-based 3D visualization can substantially enhance understanding and forecasting of indoor environmental conditions in educational buildings. The proposed framework provides actionable insights into HVAC behavior, enables proactive comfort management, and moves toward more intelligent, data-driven building operation. While the framework does not yet constitute a fully real-time digital twin, it establishes a solid foundation for future predictive building-management platforms at Politecnico di Torino and represents a meaningful step toward data-driven and occupant-centric environmental control.

Chapter 6 – Bibliography

[1] ▲ Al Horr, Y., et al. (2016). Occupant productivity and office indoor environment quality: A review. Building and Environment, 105, 369–389. https://doi.org/10.1016/j.buildenv.2016.06.001

[2] ▲ Ouf, M., Stathopoulos, T., & Bahadori-Jahromi, A. (2022). Dynamic classroom ventilation based on CO_2 forecasting. Energy and Buildings, 264, 112054. https://doi.org/10.1016/j.enbuild.2022.112054

[3] ▲ Lu, X., et al. (2019). IoT-enabled smart buildings: An overview. Energy and Buildings, 199, 187–199. https://doi.org/10.1016/j.enbuild.2019.06.013

[4] ▲ Amasyali, K., & El-Gohary, N. M. (2018). A review of data-driven building energy prediction studies. Renewable and Sustainable Energy Reviews, 81, 1192–1205. https://doi.org/10.1016/j.rser.2017.04.095

[5] ▲ Li, C., Tang, Y., & Peng, L. (2021). Deep learning-based temperature forecasting for buildings. Energy and Buildings, 231, 110621. https://doi.org/10.1016/j.enbuild.2020.110621

[6] ▲ Minassian, R., Mihăiţă, A.-S., & Shirazi, A. (2025). Optimizing indoor environmental prediction in smart buildings: A comparative analysis of deep learning models. Energy and Buildings, 327, 115086. https://doi.org/10.1016/j.enbuild.2024.115086

[7] ▲ Wen, L., Zhou, K., & Yang, S. (2019). A comparison of ARIMA and LSTM models for short-term building energy consumption prediction. Energy and Buildings, 194, 116–125. https://doi.org/10.1016/j.enbuild.2019.04.015

[8] ▲ Elmaz, F., Eyckerman, R., Casteels, W., Latré, S., & Hellinckx, P. (2021). CNN-LSTM architecture for predictive indoor temperature modeling. Building and Environment, 206, 108327. https://doi.org/10.1016/j.buildenv.2021.108327

[9] ▲ Li, Z., et al. (2020). Short-term building load and temperature prediction via LSTM networks. Applied Energy, 272, 115200. https://doi.org/10.1016/j.apenergy.2020.115200

[10] ▲ Chou, J.-S., & Tran, D. T. (2020). Feature engineering for building energy prediction using temporal data. Applied Energy, 261, 114386. https://doi.org/10.1016/j.apenergy.2019.114386

[11] ▲ Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. NeurIPS 30. https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf

[12] ▲ Zhao, Z., et al. (2023). Explainable machine learning for building energy prediction: A review. Energy and AI, 13, 100212. https://doi.org/10.1016/j.egyai.2023.100212

[13] ASHRAE. (2023). ANSI/ASHRAE Standard 55—Thermal Environmental Conditions for Human Occupancy. https://www.ashrae.org/technical-resources/bookstore/standard-55

[14] CEN. (2019). EN 16798-1: Energy performance of buildings—Indoor environmental input parameters for design and assessment. https://epb.center/document/en-16798-1/

[15] Frontczak, M., & Wargocki, P. (2011). Influence of indoor environmental factors on human comfort. Building and Environment, 46(4), 922–937. https://doi.org/10.1016/j.buildenv.2010.10.021

[16] ASHRAE. (2025). Position Document on Indoor Carbon Dioxide. https://www.ashrae.org/File%20Library/About/Position%20Documents/pd-on-indoor-carbon-dioxide-english.pdf

[17] Wang, S., & Yan, C. (2020). Data-driven building energy and environment modeling. Applied Energy, 263, 114683. https://doi.org/10.1016/j.apenergy.2020.114683

[18] Ahmad, M. W., Mourshed, M., & Rezgui, Y. (2017). Trees vs. neurons: Comparison between Random Forest and ANN for building prediction. Energy and Buildings, 147, 77–89. https://doi.org/10.1016/j.enbuild.2017.04.038

[19] Bai, Y., et al. (2020). Convolutional neural networks for time-series modeling in buildings: A review. Energy and Buildings, 215, 109917. https://doi.org/10.1016/j.enbuild.2020.109917

[20] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

[21] Ribeiro, M. H. D. M., et al. (2020). Explainable ML and SHAP for energy/IEQ prediction. Applied Energy, 262, 114313. https://doi.org/10.1016/j.apenergy.2019.114313

[22] Hu, X., & Assaad, R. H. (2024). A BIM-enabled digital twin framework for real-time indoor environment monitoring and visualization by integrating autonomous robotics, LiDAR-based 3D mobile mapping, IoT sensing, and indoor positioning technologies. Journal of Building Engineering, 86, 108901. https://doi.org/10.1016/j.jobe.2024.108901

[23] Chou, J.-S., et al. (2019). Outlier detection in building sensor data. Automation in Construction, 104, 102 – 112. https://doi.org/10.1016/j.autcon.2019.03.009

[24] Park, S., & Kim, H. (2020). Data preprocessing for IoT-based building energy analysis. Energies, 13(18), 4721. https://doi.org/10.3390/en13184721

[25] ▲ Guo, L., Zhang, H., & Li, H. (2020). Encoding time features for deep energy forecasting. Applied Energy, 275, 115318. https://doi.org/10.1016/j.apenergy.2020.115318

[26] ▲ Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

[27] ▲ LeCun, Y., et al. (2015). Deep learning. Nature, 521, 436–444. https://doi.org/10.1038/nature14539

[28] ▲ Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

[29] ▲ Gal, Y., & Ghahramani, Z. (2016). Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. Proceedings of the 33rd International Conference on Machine Learning (ICML), 1050–1059. https://proceedings.mlr.press/v48/gal16.html

[30] ▲ Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? Geoscientific Model Development, 7(3), 1247–1250. https://doi.org/10.5194/gmd-7-1247-2014