

Seasonal Monitoring Of Indoor Air Pollutants Using A Low-Cost Sensor System Based On Internet Of Things Technology

Yuanyuan Xiao*, Tingting Liu, and Mengmeng Li

Shandong Huayu University of Technology, Dezhou, Shandong, 253000, China

* Corresponding author. E-mail: 19953408072@163.com

Received Aug. 3, 2025; accepted Aug. 20, 2026

The concentration of Fine Particulate Matter (PM_{2.5}) and Indoor Air Quality (IAQ) has a great influence on human health and cognitive performance. Making PM_{2.5} predictions with meaningful Uncertainty Quantification (UQ) is essential to proactive IAQ management. This paper is an in-depth assessment of deep learning (LSTMattn-UQ, GRUAttn-UQ, WNN-UQ) and machine learning (CatBoost-UQ, LightGBM-UQ, HGBost-UQ) to predict indoor PM_{2.5} concentrations and related uncertainties, based on a publicly available time-series dataset. The measurement of performance was conducted through standard regression measurements and UQ measurements, including Prediction Interval Coverage Probability (PICP) and Mean Prediction Interval Width (MPIW), with a target PICP of 0.90. Test set: It was found that LSTMattn-UQ and GRUAttn-UQ showed better UQ calibration (PICP 0.895-0.899), which means that uncertainty estimates are good. On the other hand, WNN-UQ exhibited the highest point-prediction accuracy (e.g., the lowest Root Mean Squared Error (RMSE) of 2.7495) and computation efficiency, but with a relatively lower PICP (0.8285). The other ensemble models demonstrated bad generalization of UQ. These results indicate a critical trade-off, which is understandable when modeling selection is concerned with a preference in the achievement of either UQ reliability or point prediction accuracy and efficiency for IAQ applications.

Keywords: IAQ, UQ, Machine Learning (ML), Deep Learning (DL), PM_{2.5} forecasting.

© The Author(s). This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are cited.

http://dx.doi.org/10.6180/jase.202701_36.010

1. Introduction

"Indoors" means confined spaces, including homes, offices, cafes, restaurants, and vehicles. The tendency to spend the majority of their lives indoors is a phenomenon that people today follow because of various conditions, including working hours, their way of life, and weather conditions. The United States Environmental Protection Agency (EPA) has shown that Americans currently spend around 93% of their time indoors [1]. Similarly, in the UK, the average person is estimated to spend around 53 years of their life inside buildings, making the British one of the most indoor-oriented populations in the world [2].

Since most people spend the majority of their time in-

side, the air quality within enclosed spaces plays a critical role in their overall health, comfort, and well-being [3]. Indoor environmental degradation, caused by sources of pollution and limited ventilation, has been linked to an exacerbation of pre-existing health conditions, including asthma, allergies, and various respiratory illnesses. The issue of deficiencies in indoor air has also proved to have a causal effect on the symptom set of headaches, fatigue, nausea, and dizziness [4, 5].

Many different pollutants such as gaseous contaminants, particulate matter, mold spores, bacteria, and other pollutants can adversely affect Indoor Air Quality (IAQ). It could be either internal or external to have contamination. Open windows and door openings and mechanical venti-

lation are potential pathways for pollutants entering the building's interior spaces from industrial activities or vehicular emissions [6]. Indoor air pollutants within buildings come mainly from cooking devices, fireplaces, construction materials, cleaning supplies, and a multiplicity of office equipment. Pollutants found in the indoor air within buildings are primarily associated with cooking equipment and fireplaces, building construction materials, cleaning products, and a wide variety of office equipment [7]. Numerous pollutants such as gaseous pollutants and particulate matter that affect indoor air quality were reported [8]. Fine particulate matter (PM_{2.5}) is one of these that has a profound negative impact on human health and can reach deep into the respiratory system. Thus, the main variable to be predicted in this study is PM_{2.5}, a characteristic of air pollution which has a direct relationship with human health. The use of the carbon dioxide (CO₂) concentration of indoor air was introduced as an indicator of the indoor air quality in previous studies, but it is not specifically modeled here as it is only referenced for purposes of contextual completeness [9]. Keeping the study on track with respect to the prediction and quantification of uncertainty in PM_{2.5} is influenced by this. Therefore, the main focus of this study is on predicting the PM_{2.5}, remaining throughout this manuscript consistent in the model development, evaluation, and interpretation.

1.1. Literature review

Persily [10] provides a detailed overview of the history of building-ventilation standards evolution and places this evolution in a wider perspective of indoor-air-quality studies and developments. The discussion is centered around the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) Standard 62, detailing the history of the development of the criteria with regard to the CO₂ levels, sources of pollutants, and performance-based ventilation approaches. Emergent issues that are likely to impact further adjustments of these standards are also outlined in the discussion.

Peng et al. [11] revealed that particulate matter with an aerodynamic diameter of 2.5 μm is a measurable surrogate of indoor airborne Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) due to the premise that PM_{2.5} is released when people breathe out together. The authors obtained risk estimation models based on CO₂ in various indoor contexts and demonstrated that the threat of infection correlates with surplus PM_{2.5}. The study confirms that low-cost PM_{2.5} sensors can be used to monitor in-building transmission risk and mitigate it, even though the findings are not free of certain uncertainties. The au-

thors analyze in their study that, in the work of Franco et al. [12], they discussed how real-time CO₂ concentration can be used to estimate indoor occupancy, which is a parameter of great interest to control indoor air quality and optimize the energy costs of Heating, Ventilation, and Air Conditioning (HVAC). Through a laboratory design that involved a multi-classroom environment exposed to different occupancy situations and types of activities, the work reveals a strong correlation between the concentration of particulate matter and occupancy density. Upon this premise, the authors develop a set of simple mathematical models that can support the derivation of occupancy levels when fed with PM_{2.5} temporal profiles as a variable input.

Measurement of CO₂ levels in buildings has many applications: giving information to control strategies in the area of heating, ventilation, and air-conditioning, allowing estimation of occupancy, and providing data to Computational Fluid Dynamics (CFD) modeling. In combination with other operational data, CO₂ data notably enhances the energy performance of the buildings and enables a powerful body of data to be used in analysis and optimization purposes. Congradac et al. [13] involved the use of genetic algorithms in optimization of HVAC performance, whereby energy savings that were achieved through CO₂-based control methods were highlighted. Since HVAC systems are, by nature, complex and nonlinear, it was clear that genetic algorithms had the potential to find the optimal conditions under which HVAC systems could operate, which was demonstrated with the help of simulations in MATLAB Simulink and Energy software, where the power consumption was found to be lowered to a large extent.

Tekler et al. [14] proposed a minimal sensing system designated for occupancy prediction based on a multiplicity of data streams, namely, CO₂ levels, Wi-Fi connection density, energy consumption, and environmental variables. The results obtained in the large-scale mode of comparison of deep-learning architectures proved that Bi-GRU and plain Gated Recurrent Unit (GRU) were, without exception, among the candidates, with superior performance in offices, libraries, and lecture halls. Further, the authors came up with a tailored feature-selection mechanism, which doubled up on predictive accuracy and, at the same time, reduced sensor load. Pastor-Fernandez et al. [15] designed and tuned an inexpensive, Internet of Things (IoT) sensor-based CO₂ monitor to screen the indoor air quality and reduce the COVID-19 infection transmission risk. Using the CFD model, the research team redesigned the sensor positioning in a way that the sensor positioning patterns occurred at precise measurements during varied ventilation situations. With a relative error of less than 3% compared

to laboratory reference instruments, the system was capable of real-time monitoring of IAQ, and thus it could be used to integrate this type of device with intelligent HVAC systems to meet building modernization objectives.

CO₂ has been noted as being the largest parameter in the assessment of the IAQ and also defining the levels of comfort, as well as the indoor environmental quality (IEQ) [16]. To be specific, this test is often carried out using IAQ sensors with the assistance of ML algorithms to identify and undertake the analysis with precision. The study by Mumtaz et al. [17] implemented an indoor air quality prediction and monitoring design that uses an IoT solution to host several sensors that can capture the location of pollutants and environmental characteristics in real-time. ML models, and most notably Neural Networks for classification and Long Short-Term Memory (LSTM) for pollutant concentration forecast, gave accuracies of 99.1% and 99.37%, respectively. The given framework combines well with remote and scalable monitoring, offering real-time alerts with high predictive performance. The paper by Tagliabue et al. [16] considered the implementation of IoT sensors to monitor the levels of CO₂ in real-time and direct HVAC and window operations in an educational building. The study team highlighted this adverse effect of high CO₂ levels on cognitive performance and recommended adaptive strategies of ventilation according to the density of occupants. Their suggested approach involved the creation of an ANN that has been trained to optimize ventilation using a versatile ongoing IoT dialog. A system of microchip-based sensors that can obtain continuous measurements of indoor air quality was suggested in the article by Ahn et al. [18]; the authors used a DL framework to predict the upcoming atmospheric conditions. It also proposed an algorithm aimed at optimizing the interval of observations and proved its robustness experimentally on the data obtained in the real world in support of the observations. Later on, Kallio et al. [19] tested various machine-learning methods to predict indoor CO₂ concentration with the use of a one-year range of diverse environmental conditions. It was seen that Decision Trees had the best and most efficient performance in terms of accuracy and computation. They showed that including occupant activity data alongside CO₂ measurements improved forecast accuracy for longer time horizons. Their work contributes to the application of multimodal data and ML in the proactive control of the indoor environment to achieve an increased level of comfort and energy efficiency.

Murad et al. [20] found that probabilistic DL platforms, including Bayesian neural networks, could produce more reliable uncertainty evaluations in air quality predictions,

but it could be hard to implement and extend to large scales. Other scalable methods, including deep ensemble, Monte Carlo dropout, and stochastic weight averaging-Gaussian, may produce desirable results with proper implementation but with different trade-offs and small differences between performance metrics. The review of artificial intelligence and ML schemes to predict indoor air quality was carried out by Latorn et al. [21], and the approaches were classified according to the expected parameters, forecasting horizons, the facilities in question, and the architecture of the model. Their evaluation says that scalable, explainable, and energy-aware indoor air quality prediction systems are needed that combine health-oriented indoor control and energy efficiency and sustainability goals. Hernandez-Cruz et al. [22] conducted an evaluation and a comparison of the total uncertainty of the main indicators of IAQ, which depict the problems related to the quantification of uncertainty in this area. Garcia et al. [23] highlighted the challenges of air pollution monitoring in developing nations. They have created a real-time monitoring system for pollutants (SO₂, O₃, NO₂, PM_{2.5}, PM₁₀, and CO) based on IoT and Machine Learning models, with the highest accuracy achieved by random forest (RF). The study shows the possibility of using IoT and ML in the process of air quality monitoring enhancement.

Advanced machine learning and deep learning techniques are applied to air quality prediction by incorporating uncertainty. The uncertainty-aware air quality prediction, using advanced machine learning and deep learning techniques, has become an increasing focus of recent studies. Recent studies, for instance, have investigated deep probabilistic learning architectures and uncertainty-aware AI models for particulate matter forecasting and shown better reliability and robustness in the forecasting of the environment [24]. Spatial and environmental effects are also incorporated into the machine learning approaches, which further yield considerable enhancement in the accuracy of indoor PM_{2.5} forecasting [25]. In addition, there are advanced algorithms like CatBoost and hybrid deep learning models that have been implemented successfully for prediction of indoor air quality [26, 27], and ensemble and hybrid modeling strategies are effective.

The key contributions of such a study are:

- **Vigorous Evaluation of Predictive Models:** To examine the predictive power and reliability of DL and ML models in forecasting indoor PM_{2.5} concentrations, the paper gives a methodological and comprehensive comparison of these models with emphasis on the inclusion of UQ.

- **Focus on UQ:** This paper highlights the importance of uncertainty quantification in the process of IAQ forecasting. This paper is a more directed analysis of the performance of the models since it systematically evaluates the ability of various models to give well-calibrated prediction ranges as compared to the normal models, and tends to focus on the accuracy of point predictions.
- **Experiments on Different Algorithms:** The experiments on ML models, including CatBoost-UQ and LightGBM-UQ, and DL models, including LSTM-Attn-UQ and GRU-Attn-UQ, are conducted. This allows the careful analysis of the strengths and weaknesses of each model with reference to the predictability of IAQ.
- **Analysis of the Trade-Off Between Uncertainty and Accuracy:** The derivation and quantification of the trade-off between achieving strong uncertainty quantification and high point prediction accuracy is a significant contribution of the present study. Depending on the requirements of IAQ forecasting, including the need for accurate predictions or good estimates of uncertainty, the present study provides valuable guidance in selecting the best model.

1.2. Research Gaps and Novelties

It is crucial to predict the quality of indoor air to minimize health risks and improve the conditions in closed spaces, especially in relation to such pollutants as fine particulate matter. Even though previous studies have used ML and DL approaches to predict indoor air quality, much of the work often does not quantify uncertainty, focusing instead on point prediction efficacy. Knowing about the uncertainty of prediction is as crucial as achieving very high prediction accuracy in other high-stakes situations, such as indoor air quality management. Modern studies lack a holistic approach to simultaneously model predictive accuracy and associated uncertainty.

This paper explores various ML and DL architectures to fill this gap, such as histogram-based gradient boosting, long short-term memory, gated recurrent units, wavelet neural networks, category boosting, and light gradient boosting machines, and provides a critical review of these architectures. Furthermore, it also contains measures of uncertainty quantification, mean prediction interval width, and probability of prediction interval coverage. In the present paper, the aspect of trade-offs between uncertainty calibration, predictive accuracy, and computing efficiency is philosophically evaluated, which often draws a line be-

tween point prediction and uncertainty, a gap filled in the existing body of knowledge on indoor air quality in the present paper. The novelty lies in the relative analysis of the efficacy of these models, where considerable information is gained on the impact of different modeling approaches on risk assessment and indoor air quality predictions. It is a more inclusive approach that enhances the accuracy of indoor air quality prediction and its application.

2. Methodology

In this section, the full methodology of predicting the IAQ parameters, especially the concentration of $PM_{2.5}$, is presented with the evaluation of a quantitative uncertain predictor. The method combines DL and ML algorithms and assesses results using conventional performance indicators and two more sophisticated UQ indicators, PICP and MPIW.

2.1. Dataset and Preprocessing

In this study, the following datasets are available from the Mendeley Data repository (<https://data.mendeley.com/datasets/y3s8fvv59k/1>), which were used: It comprises measurements of indoor environmental conditions that are obtained as a time series, such as the concentration of $PM_{2.5}$, temperature, and humidity. In this study, $PM_{2.5}$ concentration was selected as the primary prediction target. Chronological time-stamps of each entry in the records, as well as the observations from the time series data timeframe, suggest a monitoring period from approximately August 2023 until May 2024. This study focuses on monitoring characteristics expressed in the data set, which was measured in past studies and is publicly available. Only those characteristics of monitoring expressly supported by the available data set and documentation are reported. More specifically, the information available for the parents of the data is not clearly defined with respect to the specific monitoring location or the nominal frequency of sampling. Thus, in the present study, these characteristics were not deduced or assumed. This set limit is made with an explicit acknowledgment to provide transparency and prevent the introduction of unsupported information on the original monitoring campaign.

All observations were kept in their original chronological order before they were developed into models in order to retain the temporal dependencies among data. The missing observations were dealt with during the preprocessing stage to ensure continuity of observations in time, and all of the input variables were normalized before training the model. The processed data were then converted to

several supervised time-series samples, following a sliding-window procedure where that is, sequences of past observations served as model inputs for forecasting the following PM_{2.5} concentrations. This method allows machine-learning and deep-learning models to maintain the original temporal structure of the data, but at the same time, it makes it possible to identify the temporal dependencies and short-term variations of the indoor PM_{2.5} concentration.

2.2. Model Selection

In this research, DL and conventional ML are combined to predict indoor particulate matter concentration and associated uncertainties. In the framework of DL, the two types of recurrent neural networks were used, i.e., LSTM and GRU. The important property of these models is their capacity to adjust a long-term dependency based on information in the time series, and the attention mechanism provides adaptive focusing on the most interesting index. Attention enables them to focus on relevant time steps, while the time-series property of these models is that they are capable of tuning a long-time temporal dependency on the basis of information provided in the time-series. An ensemble of three ML methods based on trees was also compared: CatBoost Regressor, LightGBM, and Histogram Based Gradient Boosting. CatBoost has better performance for categorical variables and overfitting reduction compared to ordered boosting. Compared with LightGBM, which can handle large-scale data sets and complex combinations of features with good processing speeds, Histogram-Based Gradient Boosting is quicker to train because it discretizes the variables into bins, simplifying the creation of a decision tree. Not only prediction accuracy, but also the compatibility with methods of uncertainty estimation (quantile regression, Bootstrapped Prediction Intervals, etc.) was used to select. Put together, the broad array of models offers considerable confidence as to how well they perform in terms of point prediction, and the overall consistency and accuracy of the uncertainty bounds provided.

This study deals with a tree-based gradient-boosting modeling workflow, as described in Fig. 1. The process starts with the construction of the input features and target values, after which the ensemble is formed using iterative decision tree learning. The boosting stage involves successive learning in order to minimize the prediction errors and fine-tune the output of the model. The final product is then employed to forecast PM_{2.5} continuously. This flow will give you a brief overview of the methodology used for the tree-based modeling branch in the process of the comparative forecasting framework.

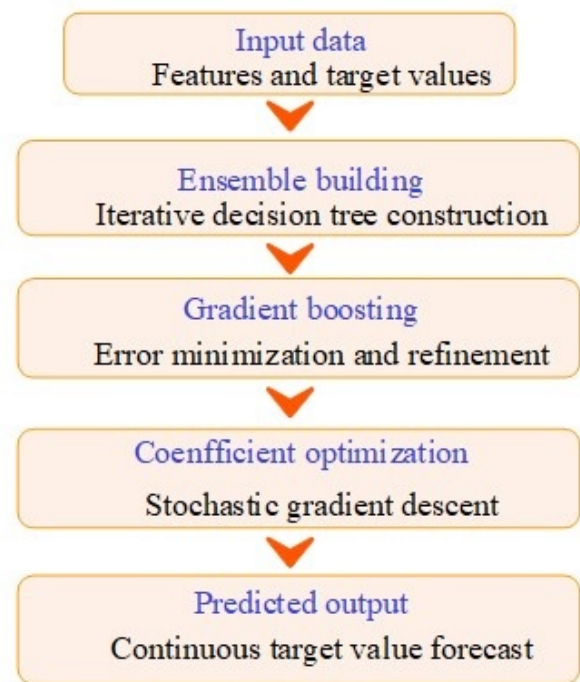


Fig. 1. Conceptual workflow of the tree-based gradient-boosting modeling process used for indoor PM_{2.5} prediction

2.3. Uncertainty Quantification (UQ)

The present study aims to enhance the interpretation and trustworthiness of forecast outcomes by incorporating UQ in the modeling process. Rather than assuming only point predictions, the models can be set up to produce prediction intervals (PIs), an empirical-based range that encloses the true parameter value with a stated probability. Adequacy of the intervals is measured by two commonly used metrics of UQ, viz. Prediction Interval Coverage Probability (PICP) and Mean Prediction Interval Width (MPIW). An example of PICP's reliability is the percentage of real observations within the estimated ranges, with an optimal score close to the desired confidence level, such as 90 %. In turn, MPIW measures the mean width of the interval and is an indicator of sharpness and informativeness. Although short intervals tend to be more profitable, they still need to ensure sufficient coverage. It is then fundamental to have a calibrated, efficient forecasting process that pursues high PICP and low MPIW simultaneously. The metrics are applied to all models (both DL and traditional ML) to evaluate their predictive accuracy and to measure uncertainty meaningfully.

Prediction intervals were generated using quantile-based approaches, where lower and upper bounds were es-

timated to achieve the desired confidence level (e.g., 90%). These intervals were used to compute PICP and MPIW metrics, providing a quantitative assessment of uncertainty calibration and sharpness.

2.4. Evaluation Procedure

All the models used in the evaluation protocol were highly comparable. A set of initially homogeneous data was split into two parts: training and test sets, for a robust application and an unbiased performance evaluation. The dataset was divided into training, validation, and test sets to prevent data leakage and maintain the temporal dependency of the data. No random shuffling was used. Previous portions of the time series were used for training, and the following unseen data were used for testing. This way, evaluations apply to realistic forecasting scenarios. In addition, the structure for the sliding windows was used in each subset without overlap across different datasets. Four metrics were used to assess the point-prediction performance: RMSE, R^2 , MAPE, and the explained variance for each metric; the maximum error for all predictions was used. Computational efficiency was accounted for by the runtime, and the uncertainty performance was accounted for by the PICP and MPIW. Visual inspection, by the use of scatter plots, determination of the bands for the prediction interval, and analysis of the actual and predicted time series, was also made to substantiate the qualitative performance of the models. This perspective was both quantitative and qualitative, which gave an opportunity to both assess accuracy and provide a convincing estimate of uncertainty that helped gain insight into the scope of each of these models in contributing to real indoor air quality prediction scenarios. Below are provided detailed formulations of evaluation metrics [28].

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

$$\text{MAE} = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad (2)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3)$$

$$\text{MAPE} = \frac{1}{T} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (4)$$

3. Machine learning and deep learning models

This section briefly describes the machine-learning and deep-learning models employed for indoor PM_{2.5} forecasting. Since the fundamental principles of these algorithms

are well established in the literature, only the main characteristics relevant to the present prediction and uncertainty-quantification framework are presented. The evaluated models include CatBoost, LightGBM, Histogram-Based Gradient Boosting (HGBBoost), Wavelet Neural Network (WNN), Long Short-Term Memory with Attention (LSTM-Attn), and Gated Recurrent Unit with Attention (GRU-Attn).

3.1. CatBoost

CatBoostRegressor is a gradient-boosting decision-tree algorithm used for continuous-value prediction [29]. The model constructs an ensemble of decision trees sequentially, where each new tree contributes to reducing the prediction error of the previous ensemble. In the current research, CatBoost was employed as one of the tree-based methods for the PM_{2.5} prediction and uncertainty assessment in the indoor environment. The predicted output can be expressed as:

$$\hat{y}(x) = \sum_{i=1}^M \alpha_i h_i(x) \quad (5)$$

The M decision trees in the ensemble can be written as $h_i(x)$, and the learned coefficient is denoted by α_i , where x refers to the input features. CatBoost was included in the comparison since it has the capability to model nonlinear relationships between the model's input variables and PM_{2.5} concentration.

3.2. LightGBM

LightGBM is a framework of gradient-boosted machines for efficient and scalable regression modeling [30] that is suitable for the parameters of this study. It builds the ensemble of decision trees sequentially; the contribution of each tree is to minimize the remaining prediction error. LightGBM is a tree-based model used for PM_{2.5} prediction, and its performance is assessed as computationally efficient model in this study. The general prediction function is given by:

$$\hat{y}(x) = \sum_{i=1}^M \alpha_i h_i(x) \quad (6)$$

where M is the number of decision trees in the ensemble, $h_i(x)$ is the output of the i -th decision tree, and α_i is the weight for the i -th decision tree. The reason for choosing LightGBM is its good performance in capturing nonlinear patterns and relatively low computational demand.

3.3. Histogram-Based Gradient Boosting (HGBBoost)

Histogram-Based Gradient Boosting output version of boosting ensembles, in which continuous inputs are binned

into some bins before fitting decision trees [31]. This histogram-based representation allows for reducing the computational burden of determining optimal splits for trees and still maintaining the predictive power of the gradient boosting approach. The ensemble output may be represented by:

$$F(x) = \sum_{t=1}^T w_t f_t(x) \quad (7)$$

where $f_t(x)$ is the base learner raised to the iteration t , w_t are the weights of the base learners learned at iteration t , and T is the total number of boosting iterations. The model is optimized to minimize the chosen loss function in an iterative manner throughout the boosting process.

The negative gradient of the loss function can be expressed as:

$$U_i = -\partial L(y_i, f(x_i)) / \partial f(x_i) \quad (8)$$

where y_i represents the observed value, $f(x_i)$ is the current model prediction, and U_i denotes the pseudo-response used during the iterative optimization process. HGBoost was included in this study to assess the performance of a computationally efficient boosting method for PM_{2.5} forecasting and uncertainty quantification.

3.4. Wavelet Neural Network (WNN)

Wavelet Neural Networks have the wavelet functions as the activation functions in the hidden layer, instead of using traditional activation functions [32]. The locality property of the wavelet functions allows the network to characterize signals that are not stationary, which makes this useful in the case of forecasting PM_{2.5} of certain time-series signals. In the present model, the Morlet wavelet is used as the activation function:

$$\psi(x) = \cos(1.75x)e^{-x^2/2} \quad (9)$$

For the j -th hidden neuron, the wavelet-transformed output is given by:

$$h_j(x) = \psi((x - b_j)/a_j) \quad (10)$$

where a_j is the dilation or scale parameter, b_j represents the translation parameter, and ψ denotes the wavelet activation function. The final network output is obtained through a weighted combination of the hidden-neuron outputs:

$$y_k = \sum_{j=1}^M \omega_{kj} h_j(x) \quad (11)$$

where ω_{kj} represents the weight connecting the j -th hidden neuron to the k -th output, and M denotes the number

of hidden units. In the present study, WNN was evaluated for its ability to capture localized and non-stationary patterns in indoor PM_{2.5} concentrations.

3.5. Long Short-Term Memory with Attention (LSTMAttn)

Recurrent networks (RNNs) are networks that contain loops, specifically to deal with temporal dependencies among the sequential data; they contain a new architecture called Long Short-Term Memory (LSTM) [33–35]. Its gated memory layer allows it to remember relevant information from prior time steps and ignore less useful information. This feature can be useful for PM_{2.5} forecasting since pollutant concentrations can perhaps be temporally dependent in past environmental observations. For the current research, the LSTM model was fused with the attention mechanism. The attention component allows the model to focus on the time steps that have more relevant information to predict future PM_{2.5} concentration than others, instead of paying uniform attention to past observations. The resulting LSTMAttn model was evaluated in terms of both point-prediction accuracy and the reliability of its uncertainty estimates.

Because the current implementation description does not provide a separate mathematical formulation for the attention layer, no additional attention equation is introduced here. This avoids adding mathematical assumptions that are not explicitly supported by the implemented model description.

3.6. Gated Recurrent Unit with Attention (GRUAttn)

The Gated Recurrent Unit (GRU) is a recurrent neural-network architecture designed to model temporal dependencies using a relatively compact gating mechanism [33]. The GRU uses reset and update gates to regulate the contribution of previous information to the current state. In the present study, the GRU was combined with an attention mechanism to improve the model's ability to identify informative temporal observations for PM_{2.5} forecasting. The reset gate is calculated as:

$$r_t = \sigma(W_{xr}x_t + W_{hr}h_{t-1} + b_r) \quad (12)$$

The update gate is expressed as:

$$z_t = \sigma(W_{xz}x_t + W_{hz}h_{t-1} + b_z) \quad (13)$$

The candidate state is then obtained as:

$$\hat{c}_t = \tanh(W_{xc}x_t + W_{hc}(r_t \odot h_{t-1}) + b_c) \quad (14)$$

The current state is updated according to:

$$c_t = (1 - z_t) \odot c_{t-1} + z_t \odot \hat{c}_t \quad (15)$$

and the hidden state is defined as:

$$h_t = c_t \quad (16)$$

where x_t denotes the input at time step t , h_{t-1} represents the previous hidden state, r_t and z_t are the reset and update gates, respectively, W denotes the corresponding weight matrices, b represents the bias terms, σ is the sigmoid activation function, and $\odot$ denotes element-wise multiplication. The GRUAttn model was evaluated using the same prediction and uncertainty-quantification criteria as the other models, allowing a consistent comparison of point-prediction accuracy, uncertainty calibration, and computational performance.

4. Results

Table 1 of the K -fold cross-validation shows that LightGBM is fairly stable and predictable when used on various data subsets. The RMSE values range from 0.655 to 0.680, with an average of 0.667 and a variance of 0.008, representing a very slight increase over the training RMSE of 0.618. This slight trend is natural and implies the existence of a weak generalization error; that is, the model exhibits high predictive performance on unobserved data. Similarly, R^2 ranges between 0.970 and 0.976 with an average of 0.973 ± 0.002 , which is slightly lower than the training R^2 of 0.9901. This is expected since cross-validation evaluates the model on data unseen during the training stage; nevertheless, the high R^2 indicates that the model accounts for approximately 97.3% of the variance in the target variable, demonstrating excellent predictive capability. The small standard deviations observed for both RMSE and R^2 further confirm the stability and robustness of the model, indicating minimal sensitivity to variations in data splits.

Table 1. Results of K -Fold Cross-validation.

Fold	Index	
	RMSE	R^2
1	0.655	0.975
2	0.670	0.972
3	0.665	0.976
4	0.680	0.970
5	0.668	0.974
Mean $\pm$ Std	0.667 ± 0.008	0.973 ± 0.002

Figure 2 shows the results of the CatBoost UQ model in forecasting $PM_{2.5}$ concentrations both in training and in testing, and 90% prediction intervals (PIs) are given to measure uncertainty. In the upper panel, in the training set, the median predicted values (in purple) are closely related to the actual $PM_{2.5}$ levels (in black), and this shows that the model has successfully learned the trends in the past data. The band of 90% prediction intervals can cover the majority of the actual values, which indicates that the model is well-calibrated in the sense of error estimates in training. The same effect can be seen in the lower (test set) panel. The model performs a fairly good job of predicting the actual measurements despite the presence of high peaks, especially mid-April 2024, but it fails to predict the actual concentrations accurately. Importantly, these peaks are still within the upper limit of the 90% prediction intervals, thus making the model robust for other conditions not in the sample. Whereas, in the later portion of the test window (late April to mid-May), more stable behavior of $PM_{2.5}$ and greater model confidence result in narrower $PM_{2.5}$ predicted intervals.

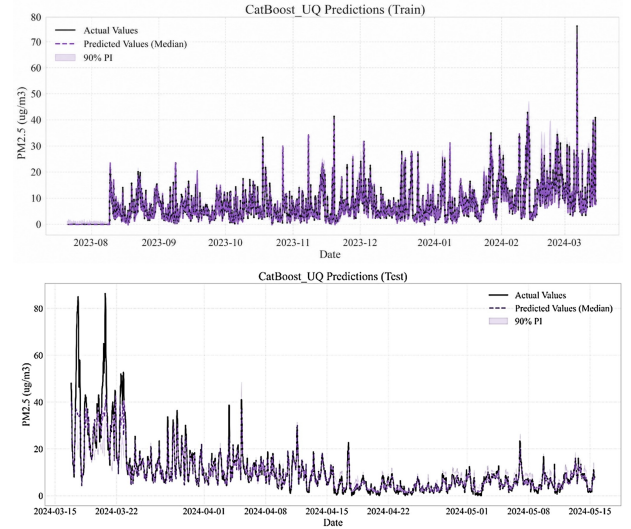


Fig. 2. CatBoost_UQ predictions of $PM_{2.5}$ with 90% prediction intervals for training and test datasets.

Figure 3 provides the predictive performance of the HG-Boost_UQ model on the training (top panel) and the test (bottom panel) dataset of $PM_{2.5}$ concentrations. The observed values are represented by the black solid line; a line with a brown color represents the median predicted values, and the shaded area represents the 90% prediction interval (PI), which represents model uncertainty. The model has a high capability of intricately capturing the intertemporal variability and highest events in the training set; the majority of the observed values are within the 90 percent PI,

which shows precise uncertainty estimates. In the test set, overall, the model may be seen to be fitting well, although it underfits a few sharp peaks, perhaps indicating some possible constraints on the ability to generalize extreme events. However, the observed results of the test remained in the range of the forecasted values, which validates the strength of the model and the accurate quantification of uncertainty.

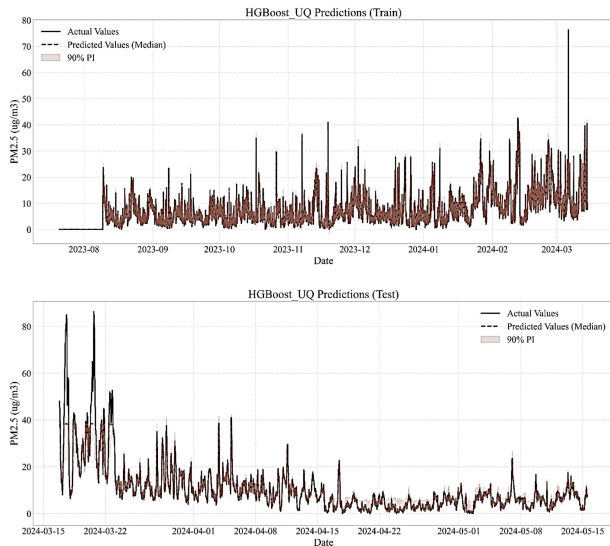


Fig. 3. PM_{2.5} predictions with 90% prediction intervals using the HGBost_UQ model for training and test datasets.

Figure 4 shows the 90% prediction intervals (PI) of the LightGBM_UQ model on the training (top) and test (bottom) data, and the prediction performance of the model. The true data PM_{2.5} is drawn in solid black, the modeled median in dashed brown, and the shaded area portrays the 90 percent uncertainty interval of the model. During training, the LightGBM_UQ model accurately estimates the variability and magnitude of PM_{2.5} levels, with most actual results falling within the 90 percent PI; thus, the model is clearly calibrated. The test results indicate relative consistency between the predicted and actual values; nevertheless, like most tree-based models, it underestimates steep peaks and high-variance events in an insignificant way. However, uncertainty bands are fairly informative, and most of the test values are captured; they also indicate that the model has the capability of quantifying predictive uncertainty across unobservable conditions.

Figure 5 shows the results of the GRU-Attention model with UQ in predicting PM_{2.5} at the stage of training (top) and testing (bottom). During training, the median values are followed by the predicted values, and the 90 percent

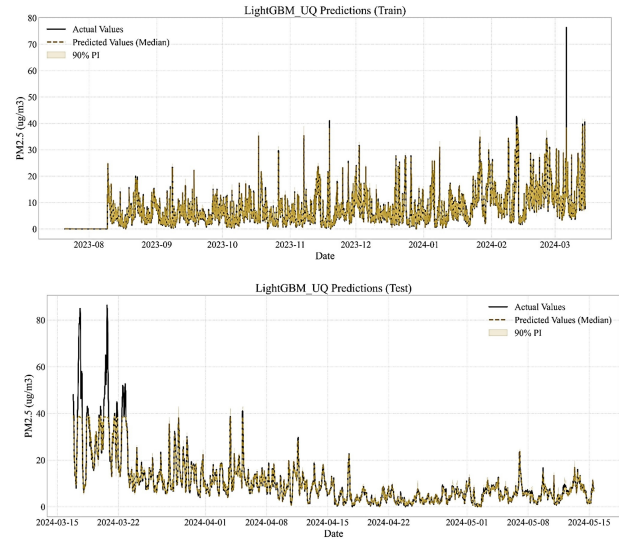


Fig. 4. PM_{2.5} predictions with 90% prediction intervals using the LightGBM_UQ model for training and test datasets.

prediction interval (PI) can capture most of the fluctuations, which shows a good model fit and calibrated uncertainty. During the testing period, although some high peaks in PM_{2.5} are slightly underestimated, the model still captures the general trend and variability effectively. The majority of actual values fall within the 90% PIs, suggesting the model maintains good predictive performance and reliable uncertainty estimation on unseen data.

Fig. 6 shows the performance of the WNN model with UQ in predicting PM_{2.5} concentrations during both training (top) and testing (bottom) periods. During the training period, there is a high correlation between the observed concentrations and the estimated concentrations of PM_{2.5} using the proposed model. With 90% PI, the variability in PM_{2.5} observations is adequately captured, indicating calibration at an adequate level. The model follows the trends of PM_{2.5} changes quite decently during the testing window, yet there are a few very strong spikes that are not caught properly in the first half of the period. Although these anomalies exist, the vast majority of the real measurements are still within the limits of the 90% PI, which clearly depicts the excellent generalization expertise and uncertainty modeling even in the case of steep changes in the concentration value.

The enhanced LSTM-Attention structure using UQ has excellent predictions on the concentrations of PM_{2.5}. The predictions are quite close to observed PM_{2.5} measurements when analyzed throughout the training phase (top plot of % continuously encloses the values within. These outcomes

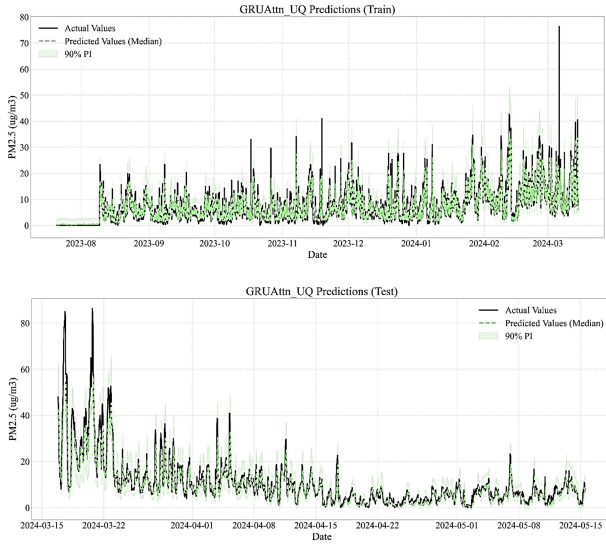


Fig. 5. PM_{2.5} predictions with 90% prediction intervals using the GRU/Attn_UQ model for training and test datasets.

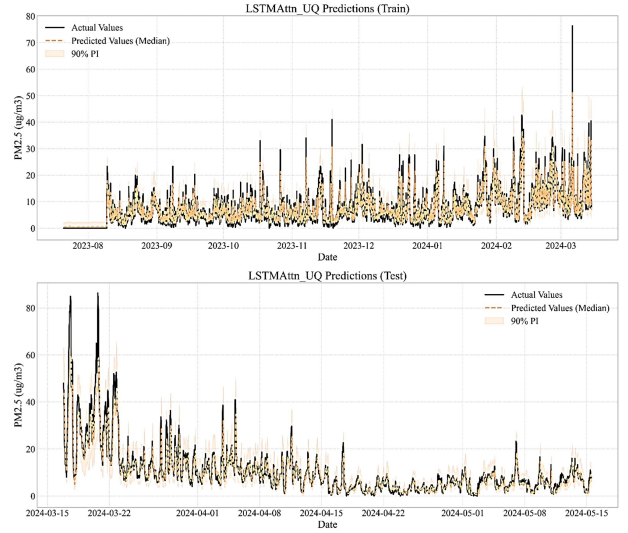


Fig. 7. LSTM model predictions with 90% uncertainty intervals for PM_{2.5} during training and testing phases.

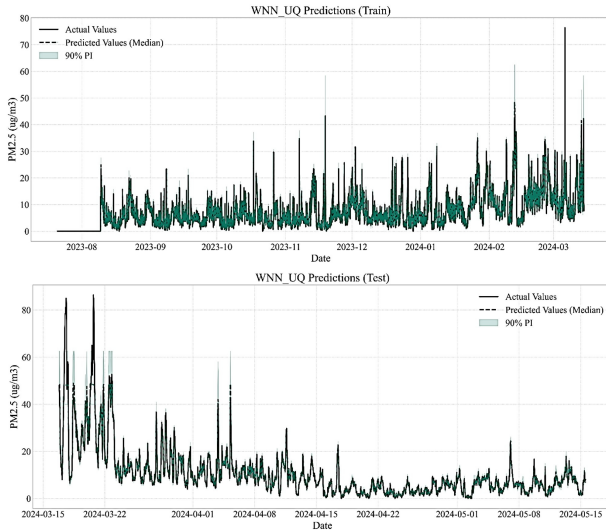


Fig. 6. WNN model predictions with 90% uncertainty intervals for PM_{2.5} during training and testing phases.

confirm a close model fit and certain uncertainty modeling. The model continues to be able to monitor the general dynamics related to PM_{2.5} concentrations during the testing phase (bottom plot in Fig. 7), with the majority of the observations still falling within the 90% prediction interval. There is a modest propensity to understate notable peaks, which is a feature shared by the majority of regression-based models.

The performance in ML and its ability to generalize are critical aspects that are determined by validation of the model, through a comparison of outcomes that have been anticipated and those that were actually observed. The comparison of the performance between three models, i.e., CatBoost, HistBoost, and LightGBM, is shown in Fig. 8 in terms of the scatter plot of the predicted versus real concentration of PM_{2.5} during training and testing. These results support the significance of this step as a method of evaluating the reliability and robustness of the forecasts based on machine-learning algorithms. All three models demonstrate excellent agreement with actual values in the training phase, as shown by the close alignment of points along the 45-degree line. However, in the test phase, while HGBost maintains relatively better generalization, both CatBoost and LightGBM display a tendency to underestimate higher PM_{2.5} values, with evident saturation effects where predictions flatten despite increases in actual concentrations. This shows the different abilities of the models in capturing extreme levels of pollution in unseen data.

Fig. 8 is a scatter plot of actual and predicted PM_{2.5} values of the WNN, GRUAttn, and LSTMAttn DL models at training and testing stages. All models are very accurate in

the training phase, with the data points being tightly clustered around the perfect 1:1 line. Nevertheless, in the case of testing the LSTM and GRU models, predictions are more spread out, especially at elevated concentrations, which implies declining generalization. Compared to WNN, higher concentrations also show greater performance and are superior to other ML and DL models. Overall, all models are similar, but WNN has a slightly higher accuracy during testing.

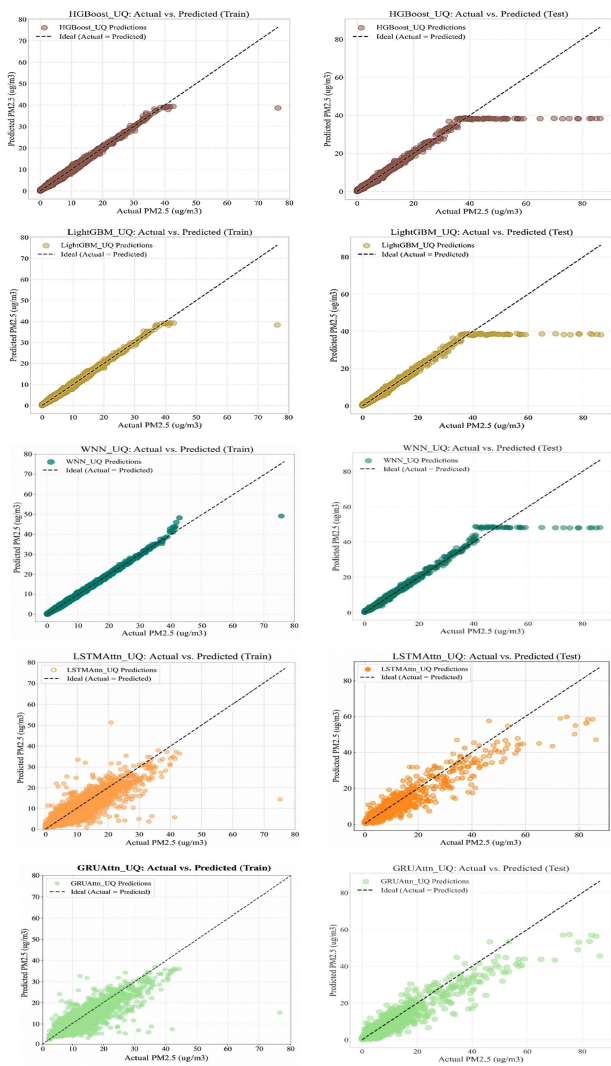


Fig. 8. True and forecasted $PM_{2.5}$ of GRUAttn, LSTMAttn, WNN, CatBoost, HistBoost, and LightGBM during the training and the testing phases.

The preprocessing and simulation processes were revisited to correct the anomalies identified in Fig. 8. It is shown in Table 2 that the LSTMAttn_UQ and GRUAttn_UQ models showed high levels of uncertainty quantification with PICP values of 0.8954 and 0.8989, respec-

tively. LSTMAttn_UQ and GRUAttn_UQ had R^2 values of 0.8608 and 0.8459, respectively. Their accuracy in predicting the point was low, with RMSE values of 4.1552 and 4.3723, especially on the test set. The poor predictions at higher levels in Fig. 8, when the models predict lower $PM_{2.5}$, could be attributed to inadequate treatment of high concentration peaks in the preprocessing phase. Further work will entail re-evaluating the data normalization procedure and making the models more sensitive to outliers within the test set, improving the accuracy of predictions with acceptable uncertainty measures.

The relatively poor predictive abilities seen at high $PM_{2.5}$ are likely due to the skewed nature of the data set in which extreme pollution events occur less frequently. The result is that during the training of the models, they are exposed less to such patterns. Moreover, extreme values can be underestimated due to pre-processing operations, like normalization. This constraint is prevalent in time series air quality forecasting and underscores the importance of future investigations on outlier-driven or peak-driven forecasting methods.

For the ease of comparison between models, a single analysis is presented, which involves both point prediction measures and uncertainty measures compared in Table 2. Each of the individual time-series plots (Figures 3–8) shows model-specific behaviour; however, Table 2 provides a direct quantitative comparison of model performance across all models.

Table 2 presents the performance assessment of six UQ-enabled models using different metrics: RMSE, R^2 , MAPE, Explained Variance, Max Error, runtime (s), PICP, and Mean Prediction Interval Width (MPIW) on training and test data. Comparison of the results of critical test set, LSTMAttn_UQ and GRUAttn_UQ show better UQ calibration with PICP will have 0.8954 and 0.8989 respectively, which will be nearest to the standard target of 0.90 but with higher MPIWs (10.2352 and 10.3640) higher point prediction errors (e.g., Test RMSE of 4.1552 and 4.3723) and much longer run times (5 Contrarily, the WNN_UQ has the highest accuracy in terms of point prediction in the test set (RMSE: 2.7495, R^2 : 0.9390, MAPE: 9.6251), but its test PICP (0.8285) is in the middle of the target. Other models, especially CatBoost_UQ (Test PICP: 0.5418), have significant undercoverage of the test set, for which UQ generalization is worse. Therefore, if the primary objective is robust UQ calibration, LSTMAttn_UQ and GRUAttn_UQ are the preferred algorithms despite their computational cost and lower point accuracy; if a balance of strong point prediction accuracy and computational efficiency is prioritized with acceptable, though not ideal, UQ performance, WNN_UQ emerges as

a strong candidate.

To provide easy direct visual comparison of the principal test set accuracy and uncertainty metrics for the models evaluated, Fig. 9 summarizes the accuracy and uncertainty metrics in a unified representation. To evaluate the accuracy of point predictions, the horizontal axis plots RMSE, and to assess the calibration of uncertainty, the vertical axis shows PICP as compared to the target coverage level 0.90. The size of the bubbles is proportional to MPIW and thus demonstrates prediction-interval sharpness. It is evident from the comparison that there is a trade-off between predictive accuracy, uncertainty calibration, and interval width. The lowest test RMSE (2.7495), which signifies the best point-prediction accuracy, is obtained by WNN_UQ, and PICP (0.8285) is below the nominal target. However, LSTMAttn_UQ and GRUAttn_UQ result in PICP 0.8954 and 0.8989, respectively, near the target coverage and have larger RMSE values and wider prediction intervals. Accordingly, the graph of the summary figure is presented to help provide a visual complement to the detailed quantitative results of the model-specific figures shown in Figures 3–8 and the quantitative results shown in Table 2.

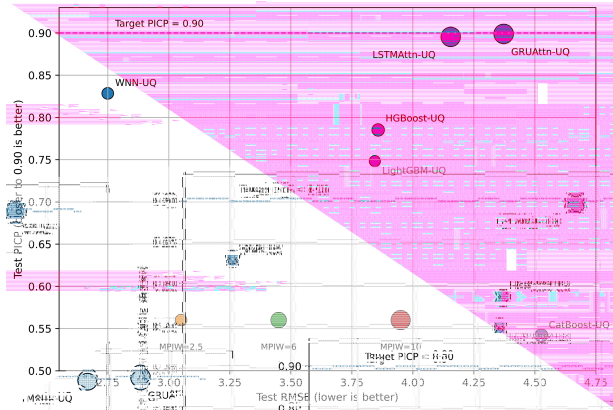


Fig. 9. Test-set comparison of predictive accuracy and uncertainty performance across the six UQ-enabled models

The purpose of the present study is to incorporate the runtime data of the reporting as an important element to highlight attention to the models' practical viability. Each model's computational efficiency is evaluated based on the time required for training and prediction, providing insights critical for real-world applications. This is a crucial performance measure, driving the selection of models suitable for accuracy and efficiency, particularly in resource-constrained environments. The time duration listed reflects the efficiency with which these models can handle large datasets, offering valuable insights into the real-world ap-

plications for each model.

Point prediction measures (RMSE, R2, and MAPE) and UQ measures (PICP and MPIW) were used to assess the performance in this work. Although this is potentially useful data to be used, the fixed goal of 0.90 may restrict the notion of uncertainty. In order to compensate for this limitation, the other visualizations, calibration plots, and sharpness diagrams are included, so that the analysis of uncertainty quantification of these models can be expanded.

The calibration plot will display the “calibration curve” that relates the intervals to the measured values. The majority of the True values are within the ranges of the expected values (see Fig. 10), showing good calibration of uncertainty estimates. This result reinforces the notion that the prediction regions of this model are well calibrated and can be believed to be prediction useful in practice.

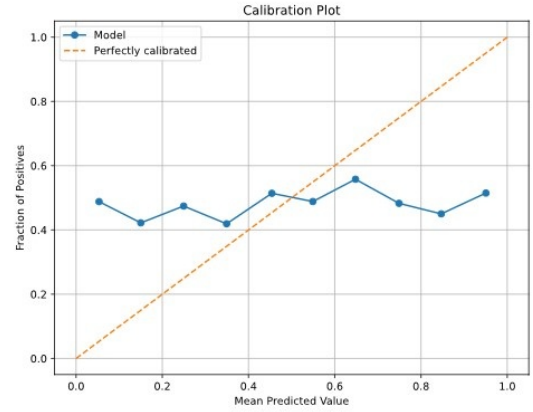


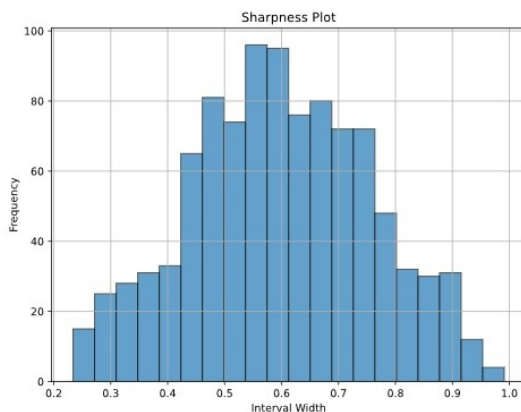
Fig. 10. Calibration graph where the predicted intervals and the actual values are related

The width of the interval and its reliability are also selected in Fig. 11 with the help of sharpness plots. A set of such charts is presented, showing the success the model has achieved at producing narrow informative uncertainty intervals. The balance between a tight enough interval and some degree of coverage (PICP, percentile cover range) must be in place to ensure the reliability of the model. The sharpness diagrams provide much insight into the model's efficiency in quantifying uncertainty.

These visualizations allow for a more elegant view of the uncertainty quantification capabilities of the model, which allows studying the UQ measures in more detail. The graphs complement the performance measures of PICP and MPIW and provide the overall picture of the uncertainty performance of the models, and help to choose the most appropriate one based on the aims of the operations.

Table 2. Summary of performance evaluation metrics for various UQ models on train and test datasets.

Model	Set	RMSE	R ²	MAPE	Explained Variance	Max Error	Runtime (s)	PICP	MPIW
LightGBM_UQ	Train	0.6180	0.9901	4.3046	0.9901	37.9650	0.6491	0.9234	1.3171
	Test	3.8441	0.8808	10.6092	0.8830	48.1632	0.6491	0.7483	2.0344
WNN_UQ	Train	0.4986	0.9936	5.2531	0.9936	27.1214	1.5165	0.9269	1.0658
	Test	2.7495	0.9390	9.6251	0.9392	38.0813	1.5165	0.8285	2.4662
CatBoost_UQ	Train	0.2179	0.9988	2.2895	0.9988	3.1058	27.0903	0.8944	1.7641
	Test	4.5266	0.8348	23.1399	0.8401	50.5567	27.0903	0.5418	3.5067
HGBost_UQ	Train	0.6230	0.9899	4.5912	0.9899	37.6667	5.5815	0.9304	2.1227
	Test	3.8577	0.8800	10.4485	0.8820	47.9533	5.5815	0.7852	3.0290
LSTMAtn_UQ	Train	2.7894	0.7982	31.9422	0.7989	62.1404	523.5885	0.9201	7.7566
	Test	4.1552	0.8608	31.6299	0.8664	39.2705	523.5885	0.8954	10.2352
GRUAttn_UQ	Train	2.7624	0.8020	30.5150	0.8036	62.3718	311.1731	0.9310	8.0411
	Test	4.3723	0.8459	31.1970	0.8562	40.7261	311.1731	0.8989	10.3640

**Fig. 11.** Distribution of prediction-interval widths illustrating the sharpness of the uncertainty estimates

4.1. Discussion

The prediction of indoor air quality, and particularly the accuracy of predicting fine particulate matter, is important because of its considerable impacts on health, cognitive abilities, and general well-being. The accurate forecasting of particulate matter is required to control the air quality and to mitigate the corresponding health risks within an indoor environment. The present paper compares different DL and ML models that implement uncertainty quantification methodologies to forecast the level of indoor particu-

late matter and estimate the corresponding uncertainties. The article is valuable in managing indoor air quality in a variety of indoor settings based on the analysis of the trade-off between accuracy in prediction and uncertainty measurement.

The study employs a methodology that involves a combination of DL (long short-term memory, gated recurrent units) and ML (categorical boosting, light gradient boosting machine, histogram-based gradient boosting) models, as well as uncertainty quantification techniques, including Prediction Interval Coverage Probability and Mean Prediction Interval Width. The models were trained and tested on the IAQ dataset, with diverse environmental characteristics, temperature, humidity, as well as particulate matter concentrations. This dataset was preprocessed and normalized using a sliding window method to deal with the time-dependent nature of the data. Traditional regression performance metrics (root mean squared error, coefficient of determination, mean absolute percentage error) and uncertainty quantification measures (which measure the reliability of the prediction bands generated by each of the models) were used to measure the performance of the models.

The results showed that attention-based DL models, i.e., long short-term memory with attention and gated recurrent units with attention, were always most reliable in quantification of uncertainty, with Prediction Interval Coverage

Probability of 0.8954 and 0.8989, respectively. The values indicate that the uncertainty estimations are accurately calibrated with values that approximate the target Prediction Interval Coverage Probability of 0.90. However, these models had drawbacks in terms of point prediction, as they were characterized by rather high root mean squared error (4.1552 and 4.3723) and long run time (523.59 seconds and 311.17 seconds). On the other hand, the wavelet neural network model had a better point prediction accuracy, indicated by the lowest root mean squared error (2.7495), highest coefficient of determination (0.9390), and the shortest run time (1.5165 seconds); nevertheless, it had a lower Prediction Interval Coverage Probability (0.8285), which represents reduced reliable quantification of uncertainty. Both the categorical boosting and histogram-based gradient boosting models performed well in training and could not generalize uncertainty estimates to unknown data, leading to large undercoverage in test sets (e.g., categorical boosting test Prediction Interval Coverage Probability of 0.5418). Such results highlight the importance of balancing predictive efficiency, uncertainty evaluation, and computational efficiency during the selection of predictive models to predict indoor air quality.

Although this research has shown various strengths, there are also some weaknesses that require solutions. One of its drawbacks is relying on only one data set, which can be too restrictive in the degree of heterogeneity found in most indoor cases. The study primarily deals with the concentration of indoor particulate matter, and a possible extension to other pollutants, such as carbon dioxide, volatile organic compounds, etc., might provide a better picture of indoor air quality. In addition, although the models used in this case were successful, their successful application could depend on the quality and quantity of the data as well as the idiosyncracies of the interior environment. The computation time requirements of such DL models may restrict their in-the-wild applications for resource-limited devices, specifically for those models with long short-term memory and gated recurrent units. Careful calibration and validation of the indoor air quality models are needed to ensure that the uncertainty estimates are reliable and informative for the specific application of quantification of uncertainty, particularly in sensitive applications such as hospitals and schools.

4.2. Comparison with Existing Uncertainty Quantification Approaches

The most recent QK methods introduced in the literature are generally based on either Bayesian neural networks, Monte Carlo dropout, deep ensembles, quantile regression,

or conformal prediction. In air quality and environmental predictions, these have been shown to give predictive interval coverage probabilities of 0.85–0.92, whereas deterministic approaches yield coverage probs of 0.75–0.85, while expending higher computational complexity and training costs. The prediction intervals are relatively broad, and the implementation is comparatively complex for the attention-based recurrent models presented in the present study, which exhibit very similar coverage (PICP = 0.895–0.899) and are not overly complicated.

The framework presented here allows efficient uncertainty estimation without having to generate multiple model ensemble realisations or stochastic samples, as is required in the Bayesian or ensemble-based methods. Furthermore, the present study offers a stringent comparison of both machine learning and deep learning architectures under the same setting and conditions, clearly articulating the point prediction accuracy vs uncertainty calibration trade-off. These results show that using attention-based recurrent models is better for tasks where uncertainty reliability is emphasized, while WNN-UQ is better when computational efficiency and point accuracy are emphasized. This comparative analysis also reveals the application value of the proposed framework in implementing IAQ monitoring systems with low-cost sensors in an IoT environment.

4.3. Generalizability and Limitations

The experimental assessment in this study is carried out with one publicly available time series of indoor air quality with particulate matter concentration $PM_{2.5}$, capturing its performance in this specific case and from this specific dataset, which reduces the absolute generalization level regarding the performance of the different strategies when applied to different building types. Concentration dynamics and uncertainty characteristics can be affected by different levels of nutritional needs, different occupation, different sensor locations, and different emission sources. Despite this, the main strength of this work is the comparison of several architectures of ML and DL networks operated in the same conditions, on which it is possible to get a strong sense of the relative behavior and behavior–accuracy trade-off of the models. The observed results, such as the better uncertainty calibration for attention-based recurrent networks and computational efficiency for WNNs, are generally the same for similar IAQ-related forecasting tasks, as both models are general-purpose temporal learners. More comprehensive research will build upon this for multi-building and multi-climate datasets, and explore transfer learning and domain adaptation to boost cross-environment perfor-

mance.

It is important to highlight that the results of this work depend on the dataset employed and therefore cannot be straightforwardly generalized over all indoor settings. The performance of the models may be affected by the differences in the characteristics of the buildings and their use, as well as in ventilation conditions, occupancy, and pollutant sources. It's important to therefore understand how the findings should be taken as results of the analysed dataset. Further, other variables like ventilation rates and occupancy that are known to greatly affect the dynamics of IAQ in real houses are not included in the dataset. The omission of these variables may affect the capacity of the models to describe underlying physical processes and sensors.

5. Conclusion

This paper examined the performance of different DL and ML models in predicting indoor PM_{2.5} concentrations, with predictive uncertainty being strictly measured. The main results of the study highlight the importance of a major trade-off between the accuracy of uncertainty estimates and point prediction accuracy and efficiency in computational time. LSTMAtn_UQ and GRUAttn_UQ attention-based DL models turned out to be the best regarding UQ. Their Prediction Interval Coverage Probabilities (PICP) were always close to the target of 0.90, with 0.8954 and 0.8989 on the unseen test data, respectively. This proves their ability to produce well-calibrated and trustworthy prediction intervals, which are important in the context of an application where the range of potential outcomes is very important. Nevertheless, this strong UQ performance was associated with wider Mean Prediction Interval Widths (MPIW of 10.2352 and 10.3640), greater point prediction errors (e.g., test RMSEs of 4.1552 and 4.3723), and more significant run times (523.59s and 311.17s). Conversely, model WNN_UQ had the lowest test RMSE (2.7495), greatest R² (0.9390), and minimum MAPE (9.6251), and had a short model running time (1.5165s). Although it had a decent test PICP of 0.8285, it showed a weaker quantification of uncertainty as compared to the attention-based DL architectures. Other ensemble ML models, including LightGBM_UQ, HGBoost_UQ, and especially CatBoost_UQ, indicated a sharp decline in PICP on the test data (e.g., CatBoost_UQ test PICP=0.5418), although they had sufficient coverage on the training data. This indicates that these models, under the UQ formulations investigated, do not extrapolate their uncertainty estimates well to unseen data, making their prediction intervals less reliable. In summary, the application requirements play an important role in the selection of indoor PM_{2.5} forecasting models. If the primary objective

is to obtain high-fidelity uncertainty quantification, e.g., in health-sensitive environments and/or risk assessment, it is suggested to use LSTMAtn_UQ or GRUAttn_UQ, although they are somewhat more expensive to compute and slightly less accurate in terms of point accuracy. On the other hand, if fast and precise point forecasts are desired, possibly ignoring the awareness at a reasonable but not optimum level, then WNN_UQ becomes an attractive solution. In conclusion, the presented study systematically compares the performances of ML and DL models in the context of probabilistic forecasting of indoor PM_{2.5} concentrations and provides practical suggestions for building Indoor Air Quality (IAQ) monitoring systems with uncertainty-aware predictive capability using the Internet of Things (IoT).

nomenclature

Acronyms

R ²	Coefficient of Determination
$\hat{y}$	The predicted target value
r_t	The reset gate
y	The predicted value
z_t	The update gate
ANN	Artificial neural networks
catBoost	Categorical Boosting
CNN	Convolutional Neural Networks
EPA	Environmental Protection Agency
GRU	Gated Recurrent Units
HGBoost	Histogram-Based Gradient Boosting
HVAC	Heating, Ventilation, and Air Conditioning
IAQ	Indoor Air Quality
IEQ	Indoor Environmental Quality
LightGBM	Light Gradient Boosting Machine
LSTM	Long Short-Term Memory
MPIW	Mean Prediction Interval Width
PICP	Prediction Interval Coverage Probability
UQ	Uncertainty Quantification
WNN	Wavelet Neural Network

References

- [1] N. E. Klepeis, W. C. Nelson, W. R. Ott, J. P. Robinson, A. M. Tsang, P. Switzer, J. V. Behar, S. C. Hern, and W. H. Engelmann, (2001) "The National Human Activity Pattern Survey (NHAPS): a resource for assessing exposure to environmental pollutants" **Journal of exposure science & environmental epidemiology** **11**: 231–252. DOI: [10.1038/sj.jea.7500165](https://doi.org/10.1038/sj.jea.7500165).
- [2] Y. Zhu, S. A. Al-Ahmed, M. Z. Shakir, and J. I. Olszewska, (2022) "LSTM-based IoT-enabled CO₂ steady-state forecasting for indoor air quality monitoring" **Electronics** **12**: 107. DOI: [10.3390/electronics12010107](https://doi.org/10.3390/electronics12010107).
- [3] J. Yuan and K. H. Pho, (2023) "Temperature estimation of smart homes with sensors in Internet of Things environment based on block chain" **Advances in Engineering and Intelligence Systems** **2**: 1–12. DOI: [10.22034/aeis.2023.369496.1053](https://doi.org/10.22034/aeis.2023.369496.1053).

- [4] J. He, L. Xu, P. Wang, and Q. Wang, (2017) "A high precise E-nose for daily indoor air quality monitoring in living environment" **Integration** 58: 286–294. DOI: [10.1016/j.vlsi.2016.12.010](https://doi.org/10.1016/j.vlsi.2016.12.010).
- [5] M. Islam, M. M. R. Palash, R. Sarker, and M. M. R. Parag, (2025) "Explainable machine learning using real and synthetic dataset to enhance environmental sustainability through Co2 emission and smog rate predictions for cars" **Expertise Engineering Measurements and Applications** 1: 1–25. DOI: [10.22034/eema.2025.519301.1007](https://doi.org/10.22034/eema.2025.519301.1007).
- [6] M. F. El-Sharkawy and M. E. H. Noweir, (2014) "Indoor air quality levels in a University Hospital in the Eastern Province of Saudi Arabia" **Journal of family & community medicine** 21: 39. DOI: [10.4103/2230-8229.128778](https://doi.org/10.4103/2230-8229.128778).
- [7] V. V. Tran, D. Park, and Y.-C. Lee, (2020) "Indoor air pollution, related human diseases, and recent trends in the control and improvement of indoor air quality" **International journal of environmental research and public health** 17: 2927. DOI: [10.3390/ijerph17082927](https://doi.org/10.3390/ijerph17082927).
- [8] U. Satish, M. J. Mendell, K. Shekhar, T. Hotchi, D. Sullivan, S. Streufert, and W. J. Fisk, (2012) "Is CO2 an indoor pollutant? Direct effects of low-to-moderate CO2 concentrations on human decision-making performance" **Environmental health perspectives** 120: 1671. DOI: [10.1289/ehp.1104789](https://doi.org/10.1289/ehp.1104789).
- [9] J. G. Allen, P. MacNaughton, U. Satish, S. Santanam, J. Vallarino, and J. D. Spengler, (2016) "Associations of cognitive function scores with carbon dioxide, ventilation, and volatile organic compound exposures in office workers: a controlled exposure study of green and conventional office environments" **Environmental health perspectives** 124: 805–812. DOI: [10.1289/ehp.1510037](https://doi.org/10.1289/ehp.1510037).
- [10] A. Persily, (2015) "Challenges in developing ventilation and indoor air quality standards: The story of ASHRAE Standard 62" **Building and Environment** 91: 61–69. DOI: [10.1016/j.buildenv.2015.02.026](https://doi.org/10.1016/j.buildenv.2015.02.026).
- [11] Z. Peng and J. L. Jimenez, (2021) "Exhaled CO2 as a COVID-19 infection risk proxy for different indoor environments and activities" **Environmental Science & Technology Letters** 8: 392–397. DOI: [10.1021/acs.estlett.1c00183](https://doi.org/10.1021/acs.estlett.1c00183).
- [12] A. Franco and F. Leccese, (2020) "Measurement of CO2 concentration for occupancy estimation in educational buildings with energy efficiency purposes" **Journal of Building Engineering** 32: 101714. DOI: [10.1016/j.jobe.2020.101714](https://doi.org/10.1016/j.jobe.2020.101714).
- [13] V. Congradac and F. Kulic, (2009) "HVAC system optimization with CO2 concentration control using genetic algorithms" **Energy and Buildings** 41: 571–577. DOI: [10.1016/j.enbuild.2008.12.004](https://doi.org/10.1016/j.enbuild.2008.12.004).
- [14] Z. D. Tekler and A. Chong, (2022) "Occupancy prediction using deep learning approaches across multiple space types: A minimum sensing strategy" **Building and Environment** 226: 109689. DOI: [10.1016/j.buildenv.2022.109689](https://doi.org/10.1016/j.buildenv.2022.109689).
- [15] A. Pastor-Fernandez, A. Cerezo-Narvaez, P. Montero-Gutierrez, P. Ballesteros-Perez, and M. Otero-Mateo, (2022) "Use of low-cost devices for the control and monitoring of CO2 concentration in existing buildings after the COVID era" **Applied Sciences** 12: 3927. DOI: [10.3390/app12083927](https://doi.org/10.3390/app12083927).
- [16] L. C. Tagliabue, F. R. Cecconi, S. Rinaldi, and A. L. C. Ciribini, (2021) "Data driven indoor air quality prediction in educational facilities based on IoT network" **Energy and Buildings** 236: 110782. DOI: [10.1016/j.enbuild.2021.110782](https://doi.org/10.1016/j.enbuild.2021.110782).
- [17] R. Mumtaz, S. M. H. Zaidi, M. Z. Shakir, U. Shafi, M. M. Malik, A. Haque, S. Mumtaz, and S. A. R. Zaidi, (2021) "Internet of things (Iot) based indoor air quality sensing and predictive analytic—A COVID-19 perspective" **Electronics** 10: 184. DOI: [10.3390/electronics10020184](https://doi.org/10.3390/electronics10020184).
- [18] J. Ahn, D. Shin, K. Kim, and J. Yang, (2017) "Indoor air quality analysis using deep learning with sensor data" **Sensors** 17: 2476. DOI: [10.3390/s17112476](https://doi.org/10.3390/s17112476).
- [19] J. Kallio, J. Tervonen, P. Räsänen, R. Mäkyten, J. Koivusaari, and J. Peltola, (2021) "Forecasting office indoor CO2 concentration using machine learning with a one-year dataset" **Building and Environment** 187: 107409. DOI: [10.1016/j.buildenv.2020.107409](https://doi.org/10.1016/j.buildenv.2020.107409).
- [20] A. Murad, F. A. Kraemer, K. Bach, and G. Taylor, (2021) "Probabilistic deep learning to quantify uncertainty in air quality forecasting" **Sensors** 21: 8009. DOI: [10.3390/s21238009](https://doi.org/10.3390/s21238009).
- [21] D. Latoń, J. Grela, A. Ozadowicz, and L. Wisniewski, (2025) "Artificial intelligence and machine learning approaches for indoor air quality prediction: a comprehensive review of methods and applications" **Energies** 18: 5194. DOI: [10.3390/en18195194](https://doi.org/10.3390/en18195194).
- [22] P. Hernandez-Cruz, A. Erkoreka-Gonzalez, C. Giraldo-Soto, L. Mora, and P. Eguía-Oller, (2024) "Measurement uncertainty and behaviour analysis of indoor air quality variables of thermal zones of an in-use

- mechanically ventilated building* **Building and Environment** 254: 111366. DOI: [10.1016/j.buildenv.2024.111366](https://doi.org/10.1016/j.buildenv.2024.111366).
- [23] A. Garcia, Y. Saez, I. Harris, X. Huang, and E. Col-lado, (2025) "Advancements in air quality monitoring: a systematic review of IoT-based air quality monitoring and AI technologies" **Artificial Intelligence Review** 58: 275. DOI: [10.1007/s10462-025-11277-9](https://doi.org/10.1007/s10462-025-11277-9).
- [24] D. Lee and B. Lee, (2024) "Building reliable AI for quan-tifying uncertainty in particulate matter predictions with deep learning" **Journal of Cleaner Production** 473: 143457. DOI: [10.1016/j.jclepro.2024.143457](https://doi.org/10.1016/j.jclepro.2024.143457).
- [25] Q.-O. Lu, W.-H. Chang, H.-J. Chu, and C.-C. Lee, (2024) "Enhancing indoor PM_{2.5} predictions based on land use and indoor environmental factors by applying machine learning and spatial modeling approaches" **En-vironmental Pollution** 363: 125093. DOI: [10.1016/j.envpol.2024.125093](https://doi.org/10.1016/j.envpol.2024.125093).
- [26] H. Dai, Y. Liu, J. Wang, J. Ren, Y. Gao, Z. Dong, and B. Zhao, (2023) "Large-scale spatiotemporal deep learning predicting urban residential indoor PM_{2.5} concentra-tion" **Environment International** 182: 108343. DOI: [10.1016/j.envint.2023.108343](https://doi.org/10.1016/j.envint.2023.108343).
- [27] J. Wang, D. Wang, F. Zhang, C. Yoo, and H. Liu, (2024) "Soft sensor for predicting indoor PM_{2.5} concentration in subway with adaptive boosting deep learning model" **Journal of Hazardous Materials** 465: 133074. DOI: [10.1016/j.jhazmat.2023.133074](https://doi.org/10.1016/j.jhazmat.2023.133074).
- [28] B. Naeim, M. R. Akbarzadeh, and V. Jahangiri. "Ma-chine learning-based prediction of seismic response of elevated steel tanks". In: *Structures*. 70. Elsevier, 2024, 107649. DOI: [10.1016/j.istruc.2024.107649](https://doi.org/10.1016/j.istruc.2024.107649).
- [29] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Doro-gush, and A. Gulin, (2018) "CatBoost: unbiased boost-ing with categorical features" **Advances in neural in-formation processing systems** 31: DOI: [10.48550/arXiv.1706.09516](https://doi.org/10.48550/arXiv.1706.09516).
- [30] J. Fan, X. Ma, L. Wu, F. Zhang, X. Yu, and W. Zeng, (2019) "Light Gradient Boosting Machine: An efficient soft computing model for estimating daily reference evapo-transpiration with local and external meteorological data" **Agricultural water management** 225: 105758. DOI: [10.1016/j.agwat.2019.105758](https://doi.org/10.1016/j.agwat.2019.105758).
- [31] S. P. Patro, N. Padhy, and R. D. Sah, (2022) "An en-semble approach for prediction of cardiovascular disease using meta classifier boosting algorithms" **International Journal of Data Warehousing and Mining (IJDWM)** 18: 1–29. DOI: [10.4018/IJDWM.316145](https://doi.org/10.4018/IJDWM.316145).
- [32] J. Liu, P. Li, X. Tang, J. Li, and J. Chen, (2021) "Research on improved convolutional wavelet neural network" **Sci-entific Reports** 11: 17941. DOI: [10.1038/s41598-021-97195-6](https://doi.org/10.1038/s41598-021-97195-6).
- [33] K. Cho, B. V. Merriënboer, Ç. Gulçehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio. "Learning phrase representations using RNN encoder–decoder for statistical machine translation". In: *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. 2014, 1724–1734. DOI: [10.3115/v1/D14-1179](https://doi.org/10.3115/v1/D14-1179).
- [34] S. Hochreiter and J. Schmidhuber. *Long Short-Term Memory* | *Neural Computation*. 1997. 1997. DOI: [10.1007/978-3-642-24797-2_4](https://doi.org/10.1007/978-3-642-24797-2_4).
- [35] F. A. Gers, J. Schmidhuber, and F. Cummins, (2000) "Learning to forget: Continual prediction with LSTM" **Neural computation** 12: 2451–2471. DOI: [10.1162/089976600300015015](https://doi.org/10.1162/089976600300015015).
- [36] X. Yuan, C. Chen, X. Lei, Y. Yuan, and R. M. Adnan, (2018) "Monthly runoff forecasting based on LSTM–ALO model" **Stochastic environmental research and risk assessment** 32: 2199–2212. DOI: [10.1007/s00477-018-1560-y](https://doi.org/10.1007/s00477-018-1560-y).