Contents lists available at ScienceDirect

Building and Environment

journal homepage: www.elsevier.com/locate/buildenv

Learning-based CO2 concentration prediction: Application to indoor air quality control using demand-controlled ventilation

Saman Taheri, Ali Razban *

Department of Mechanical and Energy Engineering, Indiana University-Purdue University Indianapolis, Indianapolis, IN 46202, United States of America

ARTICLE INFO

Keywords:
Air quality index
CO2 prediction
HVAC energy consumption
Occupant-centric control
Machine learning

ABSTRACT

There have been increasing concerns over the air quality inside buildings as high levels of bio-effluents can cause nausea, dizziness, headaches, and fatigue to the people working in those spaces. First published in 2004 as Standard 62.1, ASHRAE Standard 62.2-2019 requires highly occupied spaces to implement heating, ventilation, and air conditioning (HVAC) that can dilute contaminants produced by occupants. In this regard, occupant-centric ventilation control has been regarded as an effective practice to maintain a satisfactory indoor air quality (IAQ) when dealing with highly variable occupancy environments. However, few established models in current literature and practice consider dynamic occupancy behavior and adaptive IAQ control. To address this gap, a dynamic indoor CO2 model is constructed using machine learning algorithms to forecast CO2 concentrations across a range of forecasting horizons. Herein, we tuned and compared six state-of-the-art learning algorithms—including Support Vector Machine, AdaBoost, Random Forest, Gradient Boosting, Logistic Regression, and Multilayer Perceptron. The algorithms' performances are validated using CO2 and historical meteorological data collected from a campus classroom with a variable occupancy rate. Simulation results showed that Multilayer Perceptron can strongly predict the volatile CO2 behavior and also outperforms other algorithms in terms of accuracy. Furthermore, a control strategy capable of modeling and detecting dynamic patterns of CO2 level is utilized to modulate the ventilation rate in real-time and also reduce the energy consumption. The proposed controller reduced the HVAC fan's energy consumption by 51.4% and provided ventilation as needed per the ASHRAE standards.

1. Introduction

Per the US Energy Information Administration (EIA), the building sector accounts for over 33% of final energy consumption in residential, commercial, and industrial settings [1]. In particular, heating, ventilation, and air conditioning (HVAC) systems are responsible for the largest category of end-use energy consumption in buildings, more than lighting, refrigeration, and water heating [2]. As such, HVAC systems are a source of unexploited, avoidable energy loss, and their improper operation leads to excessive greenhouse gas emissions, an inordinate waste of energy, and occupant thermal discomfort. To operate HVAC with increased energy efficiency without compromising a satisfactory indoor environment, it is crucial to monitor its performance and optimize the operation. The HVACs' efficient operation is mostly determined by its control and optimization parameters, as discussed in [3]. In this regard, authors in [4] have indicated that improving the HVAC's control algorithm is far more reliable and profitable than replacing HVAC equipment to achieve higher efficiency. A recent study has revealed that proper control of HVAC can deliver energy-savings of 30% while preserving comfort [5]. Many researchers have given importance to advanced control algorithms whenever improving HVAC energy efficiency is desirable [6,7].

A good number of control methods for HVAC systems has been addressed in previous studies. These control strategies can be classified into three major categories: classical control methods, intelligent control approaches, and model predictive control (MPC). Classical methods include on–off, proportional, proportional–integral (PI), and proportional–integral derivative (PID) controllers. They are utilized for indoor temperature control [8], dynamic control of supply air pressure [9], cooling coil unit control [10], management of supply air temperature [11], evaporator supply heat control [12], and control of variable air volume unit temperature [13]. Despite being intuitive and easy to install, on/off and PID methods cannot deal with temporal-dependent processes with time delays, leading to an inconsistent performance with such systems. To account for this problem, intelligence and predictive-based techniques are adapted [14,15].

* Corresponding author.
E-mail address: arazban@iupui.edu (A. Razban).
URL: https://et.iupui.edu/people/arazban (A. Razban).

https://doi.org/10.1016/j.buildenv.2021.108164
Received 7 May 2021; Received in revised form 13 July 2021; Accepted 16 July 2021

Intelligent control models are usually based on artificial intelligence (AI), Fuzzy Logic (FL), and Genetic algorithms. FL is commonly employed to tune PID controller gain and to optimize its response on a global scale. The FL controller can overcome contradictions between local and global controller goals by prioritizing pre-determined individual controllers over others to minimize energy usage and preserve thermal comfort. In this context, a three-level supervisory FL architecture is incorporated [16] to control the setpoints of the lower-level controllers of water and air subsystems. An FL-based controller is designed in [17] to control humidity, air velocity, and temperature in an air handling unit (AHU). Moreover, artificial neural networks (ANNs) are usually trained on the performance data of HVAC systems to learn nonlinear and time-varying dynamics associated with the system. ANN is considered a black-box approach that does not need an understanding of the process's underlying physics. ANN is commonly used in feedforward control, and it can be trained to explore a relationship between input(s) and output(s) using mathematical techniques. For instance, a predicted mean vote thermal comfort controller is designed in [18] to predict occupancy behavior and conduct a multi-zone temperature control. In another study, ANN is applied along with a genetic algorithm to optimize an HVAC system with several chillers in a residential building [19]. A detailed review of the existing literature associated with HVAC control techniques and their associated performance can be found in [20–23].

Although the aforementioned control methodologies have emphasized the overall energy efficiency in HVAC operation, dynamic occupancy trends have been largely ignored. This is because energy-saving targets often interfere with occupant comfort and indoor air quality standards, creating complex optimization issues. To account for this problem, multi-objective optimization can be incorporated to consider both energy saving and thermal comfort goals simultaneously [24]. However, to achieve a balance between both goals, extensive knowledge regarding occupant activities is crucial. Information on occupation states might be used to regulate setback temperatures to conserve energy throughout unoccupied hours or to maintain an adequate degree of comfort for occupants upon arrival. To monitor occupancy behavior, information from a wide range of tools like Wi-Fi and Bluetooth systems, infrared sensors, power meters, and CO2 sensors can be used. Among all these options, CO2 sensor data has become a focus of recent attention as it has a high correlation with the presence/absence of occupants, and it also preserves people's privacy [25].

Moreover, there have been increasing concerns over the air quality inside buildings as traditional HVAC control strategies might not comply with the new ventilation requirements of ASHRAE Standard 62.1 2019 [26]. For example, a recent study shows that many educational environments in the US might not have sufficient ventilation to cope with the CO2 levels when the classrooms are full of students [27]. The CO2 concentration level above 1000 ppm is deemed to be high and associated with discomfort or health-related issues like nausea, dizziness, headaches, and fatigue [28].

To model CO2 level in buildings, authors in [29] proposed a machine learning (ML) algorithm, namely decision trees, trained with indoor and outdoor CO2 concentration data. A Markov Chain model is implemented afterward, resulting in 90% accuracy for occupancy state estimation. Using historical CO2 concentration info, another ML algorithm called long-short term memory is trained in [30] to predict CO2 concentration level in a short-term forecasting horizon. By examining the association between CO2 predictions and occupant numbers, they calculated the number of inhabitants as a function of CO2 concentration with more than 70% accuracy. In Ref. [31], an occupancy behavior recognition model is developed based on CO2 concentration, motion detectors, and lighting sensors. Based on this information, a Markov Chain is employed for occupancy. Although using ML algorithms for CO2 prediction has proven beneficial, there is a lack of a comprehensive study to compare the state-of-the-art algorithms on such a task. Also, it is necessary to show whether environmental features have some effect on model prediction results. In [32], a vision-based machine learning method is presented that allows the detection and identification of occupant activity inside building areas. Through the development of occupancy heat emission profiles, the data may be fed into building energy management systems, assisting in the reduction of excessive HVAC energy loads and the efficient control of interior conditions.

In this context, Ref. [33] proposes a combined machine learning and ventilation model for enhancing indoor environmental quality. A gray box model based on CO2 concentration prediction is proposed in [34] to enhance the ability of predictive models in model predictive control environment. [35] used a deep Q-network to provide model-free optimum control balancing across several HVAC systems. The optimization objective was to reduce the building's energy consumption while keeping the interior Co2 levels below a certain threshold. [36] used machine learning algorithms to investigate CO2 prediction. There are, however, numerous significant variations between the Ref. [36] and our work. The primary contribution of their study is to determine which feature sets are appropriate for a CO2 prediction study utilizing machine learning algorithms. These feature sets could include the following: (I) historical CO2 values; (II) historical CO2 and passive infrared (PIR) sensor values; (III) historical CO2, PIR, temperature, and humidity values; and (IV) historical PIR, temperature, and humidity data. Our work, on the other hand, tries to tune/optimize several machine learning algorithms using a fixed feature set (past PIR, temperature, Dewpoint, and humidity). Second, whereas the Ref. [36] employs MAE as an accuracy statistic, we employ MAE, MAPE, R2, and RMSE. We believe that utilizing many performance metrics is useful because, for example, the MAE metric is scale- and size-dependent. Furthermore, while the machine learning techniques utilized in Ref. [36] are tuned for predicting horizons of 5, 10, and 15 min, we forecast CO2 in advance of 1, 6, and 24 h. When it comes to practical applications, we believe that longer forecasting horizons with high accuracy metrics are the most desirable.

To control the HVAC system based on CO2 prediction, an internal model control in conjunction with PID-controller was used and resulted in desired ventilation air quality (80% of the time) at lower costs [37]. The work validated that the IMC-PI controller has a faster response to the changes in CO2 level when compared to the conventional PI CO2 control. This study lacked model validation and did not compare their proposed controllers with the traditional ON–OFF controllers. Multi-objective optimization MPC and PI controllers were used by other investigators to maintain the air quality of a full-fledged HVAC model with ventilation [38]. This implementation is case-specific and cannot be generalized for other models due to the integrated nature of the model and control. Discrete On/Off and Fuzzy Logic controller techniques were simulated using Matlab & Simulink, and the results were shown based on occupancy reflected by the collected rooms' CO2 data for energy reduction and system performance. It indicated 62.8% reduction in fan energy consumption using a Fuzzy Logic controller [39]. The investigators [40] have used DCV (demand-controlled ventilation) to change the fan speed according to the change in the room's CO2. Most of these approaches show significant energy savings; however, they either lack full model validation, fail to compare results with the conventional ON–OFF control, or propose complex control systems that demand heavy computation power and are neither versatile nor general in nature. Most of these models have a severe shortcoming of slow response time compared to the existing system or are highly complex, thereby increasing the computation time [41]. Thus, a simple model is needed to capture the CO2 dynamics in the space and to be used as a reference for the control system design. This paper addresses those gaps by starting with the indoor CO2 model, validating it, using a simple yet robust control strategy, and comparing it to the more common ON–OFF control strategy. The development of an adaptive control algorithm allows the ventilation rate to be predicted and tuned ahead of changes in CO2 concentration and occupancy rates. Based on the above discussion, the major contributions of this study can be summarized as follows:

- It is necessary to adjust and compare a number of learning-based models of CO2 concentration prediction. We aim to rank a large number of machine learning algorithms with varying capacities on such a job in order to extract a customized model from a potentially endless number of alternatives (Section 2).
- It is suggested to use an occupant-centric control method to produce ventilation that not only satisfies the new ASHRAE standard in terms of CO2 concentration level but also consumes the least amount of energy. The effectiveness of the suggested control method is shown by embedding it into an existing HVAC system and comparing its performance to that of the conventional control technique (Section 3).

The paper continues as follows:Section 2 details the representative CO2 data with features and characteristics. Then, six machine learning methods are tweaked and compared using multiple performance criteria spanning different predicting horizons. Section 3 goes into detail on the occupancy and demand control models. Finally, Section 4 draws the conclusion and suggests several potential research directions.

2. General structure of the proposed algorithm

This section describes the step-by-step framework used for demand-controlled ventilation (DCV) based on CO2 concentration prediction. The machine learning algorithms used for CO2 prediction are then described along with the performance metrics used to evaluate the accuracy thereof.

2.1. Occupancy-based demand controlled ventilation

As already mentioned, DCV is capable of maintaining the needed indoor air quality while reducing energy usage. Indoor CO2 levels are employed as a measure of indoor air quality in the environment, and people are assumed to be the primary producer of CO2 indoors, resulting in an increase in indoor concentrations relative to outside levels. The air conditioning system can modify ventilation rates in response to variations in indoor CO2 production, i.e., the rate of ventilation is regulated over time in response to signals from the indoor CO2 concentration. Using this method, we can reduce energy waste by catering to the ventilation needs based on the occupancy levels/number. In some ventilated areas like classrooms, banquet halls, and auditoriums, the number of occupants and their actions can be monitored to validate the CO2 prediction models. In this study, one of the most consistently occupied classrooms was chosen to conduct the CO2 data collection and model validation experiment. The occupancy number was monitored based on the class schedules from Monday through Friday. It was assumed that more than 50% of the students would arrive within 10 min of the class start time and that all students would leave within 3 min of the class end time.

To ensure sufficient IAQ, the ventilation device should optimize its decision before the CO2 level exceeds its steady-state value. Therefore, the control algorithm must predict the CO2 concentration level in advance and must control the fan ahead of steady-state conditions. The current system considers a fixed number of inhabitants while the proposed system is based on the introduced real-time indoor occupancy estimation algorithm. It is assumed that the number of occupants is constant over short-term periods, such as 10 min, and it changes over lengthy periods, such as one hour. Accordingly, we use the actual CO2 generation rate to predict the number of occupants and total CO2 generation rate in the zone for a future time, say the next 10 min. Since the data resolution for the MLP process is 1 min, we average the following ten predictions at each step. This will also help to reduce the noise associated with the MLP predictions. Other input features of the MLP algorithm are also calculated using persistence forecasting [42] which relies on past observations to predict the following values.

Based on MLP CO2 prediction, the ventilation system is controlled to maintain the required indoor CO2 level. The ventilation rate of

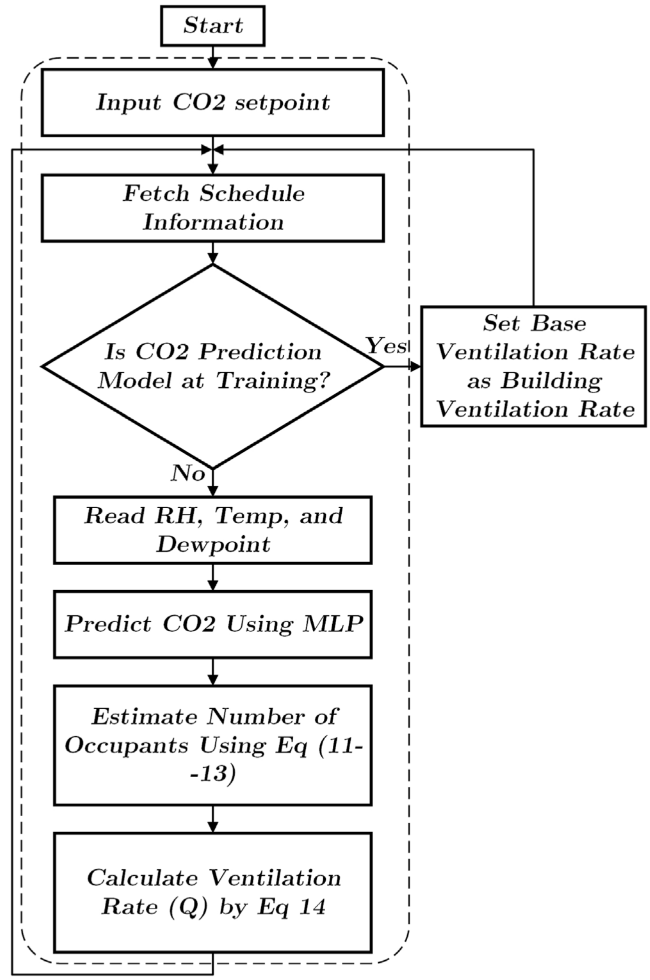

Fig. 1. The proposed framework for indoor air quality control using demand-controlled ventilation.

the system is adjusted according to the occupancy prediction and ventilation rates at different occupancy levels where E is fan energy consumption in kWh, d_p is the total pressure of the fan (obtained from the fan's nameplate), and H is the hours of operation. Fig. 1 represents the flow chart of the proposed method for the indoor air quality control of the system.

2.2. Representative machine learning models

Based on the degree and amount of human supervision they undergo during the training phase, machine learning (ML) models may be classified as supervised, semi-supervised, unsupervised, or reinforcement learning. The most popular ML method is supervised learning in which the desired labels are fed to the algorithm with the goal of estimating a numeric value. Classification and regression are two popular supervised learning activities. Our case is a supervised problem since we need to predict the CO2 concentration level ahead of time and have the requisite historical information.

To address supervised problems, many ML algorithms have been developed, each with its own set of implications. Here, Support Vector Machine (SVM), AdaBoost (AdB), Random Forest (RF), Gradient Boosting (GB), Logistic Regression (LR), and Multilayer Perceptron (MLP) are adopted as representative ML models. We selected these six algorithms for various reasons: ability to learn complex and nonlinear interactions, adaptability to different types of problems, fast results, and simple interpretation [43]. Also, each of these algorithms has specific and unique capabilities, making them attractive for a comparison study. Notice that the mathematical background of these algorithms is extensive and out of the scope of this paper. As such, we only reflect on the important concepts or formulas that determine unique characteristics of each algorithm. For a detailed discussion regarding the unique characteristics of those algorithms, interested readers are referred to [44].

Support Vector Machine: Pioneered by Vapnik [45], SVM has become a subject of intense research because of its effectiveness in regression tasks. The basic principle behind SVM is to use a regressor that minimizes the error between the predictions and ground data while providing enough margin to provide higher generalizability than simple regression. Given a set of training data D as $\{(\mathbf{X}_1, y_1), (\mathbf{X}_2, y_2), \ldots, (\mathbf{X}_i, y_i), \ldots, (\mathbf{X}_N, y_N)\} \in D$, $\mathbf{X}$ is the vector of feature space $\mathbf{X}$:$\{x_1 = \text{Temperature}, x_2 = \text{Humidity}, x_3 = \text{Dewpoint}\}$, y is the CO2 concentration level, $i \in \{1, 2, \ldots, N\}$ is the index for instances included in the dataset, and N is the number of samples. The goal is to find a function that can accurately approximate all the data. The estimation function for SVM can be written as Eq. (1).

$$f(\mathbf{X}) = \mathbf{w} \bullet \mathbf{X} + \mathbf{b} \tag{1}$$

Here, $(\bullet)$ is the inner product, and $\mathbf{w}$ and $\mathbf{b}$ are the weighting and bias vectors. These vectors must be computed in such a way minimize the associated regression error. The error function is incorporated as follows:

$$e_{reg}(f) = S \sum_{i=1}^{N} \tau\left(f(\mathbf{X}_i) - y_i\right) + \frac{1}{2}(\mathbf{w} \bullet \mathbf{w}) \tag{2}$$

where S is a predetermined value and $\tau(.)$ is the loss function associated with empirical risks, as represented in (3):

$$\tau\left(f(\mathbf{X}_i) - y_i\right) = \begin{cases} 0 & \left|f(\mathbf{X}_i) - y_i\right| \le \varepsilon \\ \left|f(\mathbf{X}_i) - y_i\right| - \varepsilon & \text{Otherwise} \end{cases} \tag{3}$$

As it is formulated, the cost function is sensitive to the value of ε. When predictions have less than a $\pm\varepsilon$ difference with the ground truth data, the weighting and bias vectors are not updated; otherwise, the configuration is adjusted in a way that minimizes the overall error. Since this hyperparameter controls SVM's accuracy, it is set as a fraction of the standard deviation associated with the CO2 concentration. As such, the error threshold is set to 27.8 during our experiments.

Random Forest:A typical way to enhance a machine learning algorithm's effectiveness is to average the forecasts of a set of several predictors. This leads to obtaining superior performance compared to a single model. This approach is called ensemble learning. RF is an ensemble-based ML algorithm that trains each predictor on a different data proportion. A detailed mathematical formulation and explanation for RF can be found in [46].

Logistic Regression: As opposed to linear regression, a logistic model computes a weighted total of the input features; however, instead of outputting the raw data like regression, it outputs a logarithm of the logistic value between zero and one as in (4).

$$p = \sigma(\mathbf{w} \bullet \mathbf{X}) \tag{4}$$

To squeeze the output between zero and one, LR uses a function called sigmoid. For a given scalar s, the sigmoid function is expressed as (5):

$$\sigma(s) = \frac{1}{1 + e^{-s}} \tag{5}$$

AdaBoost:Another way to enhance the performance of a learning model is to adjust the configuration based on the instances that contribute more to the error. One way to do so is first to train a function on the dataset and then to obtain the results, increase the weight of instances, and tune a new model based on the updated weights. This process is repeated sequentially until the final model yields a reasonable performance. This is the central idea behind sequential-based ML techniques like ADB. Interested readers should turn to [47] for more information on the mathematical context of the ADB algorithm.

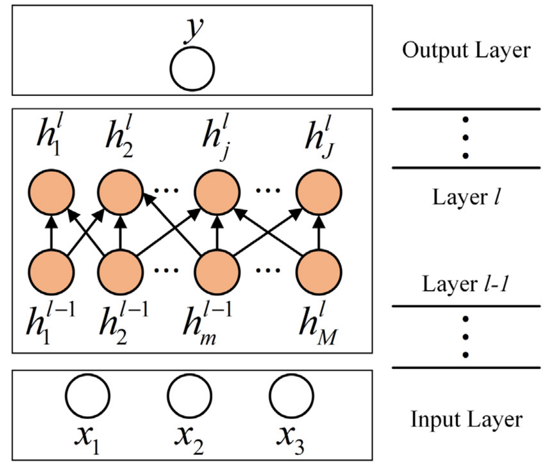

Fig. 2. Multilayer perceptron schematic.

Multilayer perceptron: An MLP is a standard neural network in which the output is a weighted sum of inputs plus a bias squashed by an activation function like a sigmoid function. A schematic of MLP is shown in Fig. 2. The relationship between the input(s) and output is determined by (6)

$$h_j^l = \sigma\left(\sum_{m=1}^{M} w_{j,m}^l h_m^{l-1}\right); \forall j \in \{1, 2, \ldots, J\}, \forall l \in \Omega \tag{6}$$

where Ω is the set of MLP hidden layers, $w_{j,m}^l$ is the connection weight of node j in layer l coming from node m in layer $l-1$, M is the number of nodes in hidden layer $l-1$, and J is the number of hidden layer nodes in the hidden layer l. The parameters of the network are calculated based on an approach called stochastic gradient descent [48].

Gradient Boosting: GB is another sequential ML model that builds an ensemble of regression models, each of which is trained sequentially and is dependent on the previous predictor. When all of the models are tuned, a highly accurate generalization model is obtained on the task. The hallmark of GB is its ability to strike the optimal balance between model sophistication and generalization performance.

2.3. Performance metrics

Various predictive measures can be used to evaluate the efficiency of forecasting models. Bias, variance, complexity of calibration, refinement, variability, precision, and resolution are all factors that influence the consistency of a forecasting model. The bulk of related literature has been validated using precision metrics such as root mean squared error (RMSE), as characterized by (7).

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2} \tag{7}$$

where y_i and $\hat{y}_i$ are the predicted and ground truth data, respectively. Another way to assess accuracy is to average the absolute difference between the predicted and actual values. This is called mean absolute error (MAE) as represented in (8).

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} \left|y_i - \hat{y}_i\right| \tag{8}$$

Forecasts that have lower MAE and RMSE values are more reliable. However, a weakness of the MAE and RMSE metrics is that they are not normalized in regard to the scale of the labels. To provide a better understanding of the error scale, a performance metric based on the percentages of errors is incorporated as in (9). This is called mean absolute percentage error (MAPE).

$$\mathrm{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{\hat{y}_i}\right| \times 100 \tag{9}$$

MAPE values of less than 10% represent an extremely precise forecast, for 11% to 20% MAPE indicates a decent forecast, 21% to 50% MAPER specifies a fair forecast, and more than 50% indicates a highly incapable model [49]. A problem with MAPE is that it does not consider prediction variance. To cope with this, the coefficient of determination, denoted by R^2 in (10), is commonly used to reflect the variance of predictions with regard to the mean of observations ($\bar{y}$).

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{10}$$

R^2 is an indicator of how well the projections match the observations, with values varying from 0 to 1. In other terms, it provides valuable information regarding the forecasting model's capacity to match, with a value near 1 indicating greater prediction accuracy.

3. Simulation results and discussion

In this part, we will discuss the findings of a real-world case study that we conducted based on the proposed algorithm.

3.1. Data characteristics

The data were collected from a university campus classroom over three months during the fall session that included 13,003 weather-related values and a corresponding CO2 concentration with a resolution of 1-min. Weather-related variables constitute ambient temperature, relative humidity (RH), and dew points. Feature-output correlation is significant for two reasons: It is critical to perform dimension reduction on large datasets. Reduction in the number of dimensions reduces the number of characteristics that must be tracked. Two characteristics that have a correlation of more than 0.7 or less than −0.7 must be reduced to one feature [50]. Heat map representation of the dataset correlation matrix is shown in Fig. 3. Our characteristics do not have features with correlations less than −0.7 and higher than 0.7. Also, learning algorithms that utilize characteristics most associated with the label are further developed based on those attributes. For example, stratified sampling is performed for the training phase depending on the values of the most correlated characteristic. In this research, temperature was shown to be strongly predictive of the result (CO2). Second, knowing which characteristics are most associated with the label is important because these features are utilized when constructing the learning algorithm. In this research, temperature was shown to be strongly predictive of the result (CO2).

The correlations between the CO2 concentration and the corresponding features are illustrated in Fig. 4. The features constitute dew points, relative humidity, and temperature. All of the features positively correlate with CO2 concentration, with temperature as the most-correlated predictor with a correlation of 0.45, followed by dew point and RH with correlations of 0.37 and 0.11, respectively. Although positively correlated, it can be seen that the system has complex behavior. To better understand the data intricacy, the probability distribution functions of each feature as well as the CO2 are shown in Fig. 5. The Gaussian or normal distribution is the most often used model for quantifying the variance in original data. In this context, data are summarized using the arithmetic mean and standard deviation. As a result, the normal distribution has become the de facto standard for describing data and its variance. Additionally, any movement away from the symmetric normal distribution toward an asymmetry perspective significantly increases data distribution identification and interpretation quality. As depicted, the CO2 average level is 545 ppm; however, we have a CO2

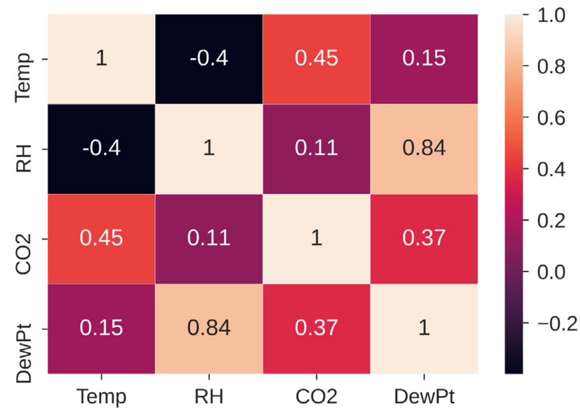

Fig. 3. Heat map representation of the correlation matrix for the representative dataset.

level of more than 1000 ppm for 1132 samples (8.7 percentage of the data). The ASHRAE standard 62.2 recommends 1000 ppm of CO2 and above as a threshold of concentration whenever ventilation correction is needed. Noncompliance results not only in detrimental effects for employees and students but also in increased visits from administrative agency officials, which could result in mandates or penalties.

3.2. Learning-based CO2 prediction

Accurate CO2 prediction is a key task in achieving demand control ventilation (DCV) since ASHRAE 62.2 requires the outdoor air flowrate to be calculated based on ventilation rates needed to dilute the contaminants produced by occupants and building materials. However, the number of occupants and their activities are uncertain to a great extent, leading to the poor performance of physics-based and rule-based CO2 prediction models [36]. Machine learning algorithms have proven beneficial in this regard, as they are capable of learning high uncertainty and variability associated with CO2 data. This section first describes the representative CO2 data along with the associated features and their characteristics. Then, six machine-learning algorithms that can learn those characteristics are tuned and compared based on several performance metrics over different forecasting horizons. It is necessary to consider various time horizons for CO2 prediction as uncertainty introduced by occupancy behavior propagates when the forecasting horizon extends to longer-term periods. In this study, short-term (a few minutes in advance), mid-term (a few hours in advance), and long-term (a day in advance) are considered as representative horizons for CO2 predictions.

The data collected from a campus classroom at IUPUI (Indiana University-Purdue University, Indianapolis) is used to evaluate the performance of ML models. The dataset is split into two sections by proportions of P% as the training set (TS) and $(1 - P)$% as the cross-validation (CV). Random subsampling is used to divide the dataset into training and test sets. Random samples are chosen during the training phase. However, when evaluating the models' performance, the cross-validation set is retained in its original structure. As such, we may argue that the cross-validation set reflects the model's predictive capability (the sequential order of the samples is maintained). Data points are assumed to be drawn from the same likelihood distribution. We then select P% of these samples at random for the training set and the remaining $(1 - P)$% for the evaluation set. We use different P values to assess the generalizability of the ML models over different forecasting time horizons. For example, $P = 89$ means that 11% of the data (1430 samples out of 13003) is associated with CV; thus, the corresponding forecasting horizon is approximately $1430/60 = 24$ h. Herein, 1-h ($P = 99.5$), 12-h ($P = 97.3$), and 24-h ($P = 89$) forecasting horizons have been selected for the experiments. Notably, it is essential to ensure

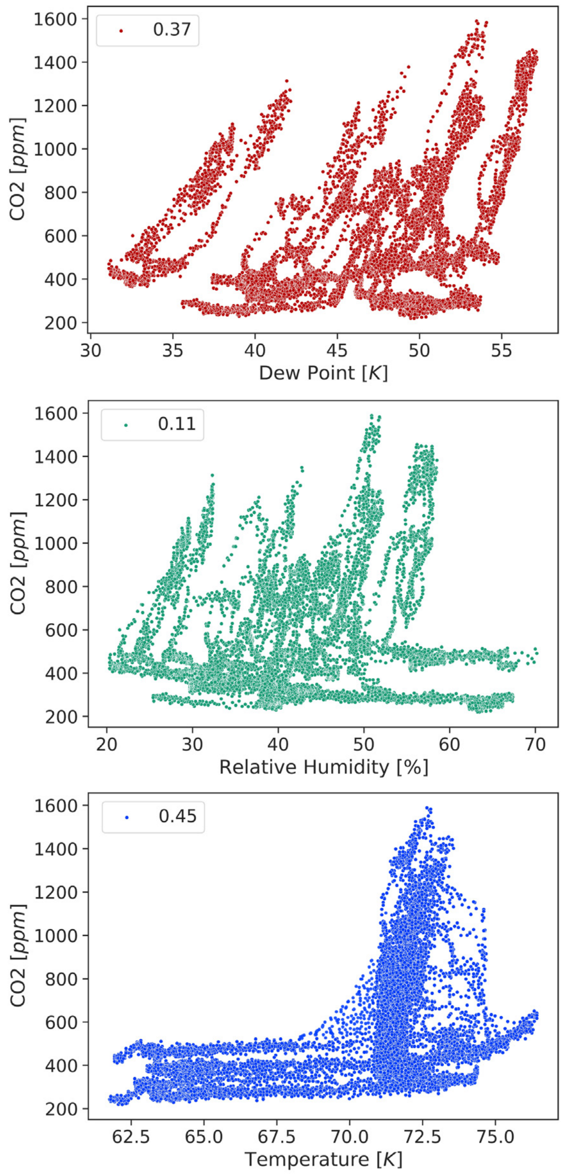

Fig. 4. Depiction of Correlation between CO2 concentration and other variables.

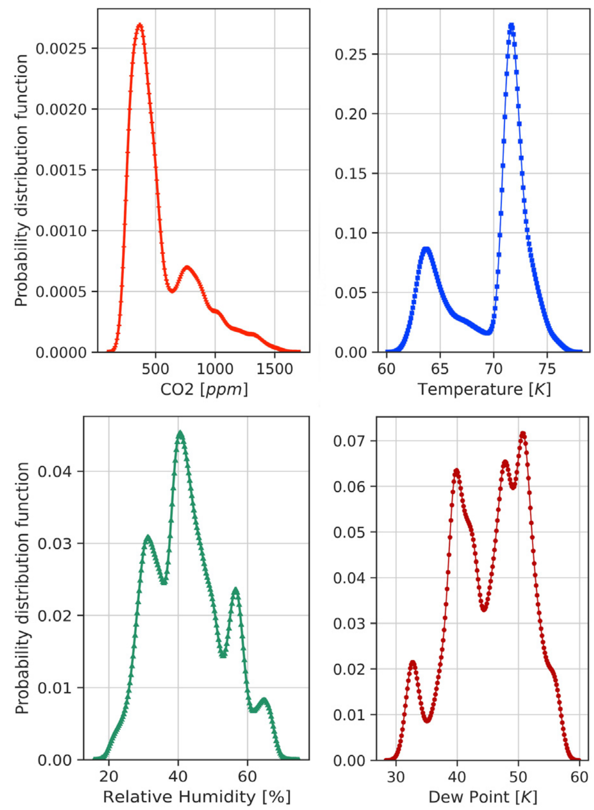

Fig. 5. Probability distribution function of different features and CO2 concentration.

Table 1
CO2 forecasting performance of the ML models—1h-ahead horizon.

Phase	Metric	SVM	RF	LR	AdB	MLP	GB
Training	RMSE	34.73	38.86	36.44	35.99	**33.29**	38.11
	MAE	30.18	33.22	30.22	37.37	**29.14**	30.16
	MAPE	17.67	12.54	18.32	15.55	14.68	**11.28**
	R^2	0.931	**0.944**	0.887	0.906	0.952	0.924
Cross validation	RMSE	35.13	39.98	37.65	36.62	**34.32**	38.55
	MAE	**30.13**	30.32	30.29	30.47	30.16	30.23
	MAPE	19.66	15.59	21.56	18.87	16.09	**13.72**
	R^2	0.895	0.902	0.812	0.883	**0.912**	0.865

Table 2
CO2 forecasting performance of the ML models—6h-ahead horizon.

Phase	Metric	SVM	RF	LR	AdB	MLP	GB
Training	RMSE	34.81	41.75	37.58	36.86	**33.78**	39.76
	MAE	38.19	41.28	39.28	33.43	42.20	**36.19**
	MAPE	19.67	15.45	21.54	18.81	**18.42**	19.64
	R^2	0.859	**0.901**	0.850	0.894	0.873	0.842
Cross validation	RMSE	46.33	43.37	49.19	48.01	**44.70**	46.75
	MAE	**0.19**	0.45	0.43	0.69	0.24	0.34
	MAPE	23.92	18.45	27.93	23.63	20.25	**17.32**
	R^2	0.782	0.788	0.691	0.755	**0.812**	0.760

that the split process does not exhibit discriminating behavior toward certain types of samples. Python's Stratified split function is used to divide the data into train and test sets to reflect the frequency of different values. This enables the creation of homogeneous training and testing sets and the elimination of sampling bias. Additionally, missing values are filled using the Simple-Imputer function, which substitutes the median for missing values. Then, all the characteristics are scaled by removing the mean value from each sample and dividing it by the standard deviation, resulting in a distribution with unit variance. This is referred to as "Standardization". Table 1 shows the performance of the proposed forecasting models for the TS and CV when a 1-h in advance CO2 prediction is needed.

As represented, different algorithms show different characteristics and performances as expected. Some algorithms like RF tend to perform better during the training phase but can be/are outperformed by other models in the validation phase. MLP, however, has shown a solid performance for both TS and CV. The forecasting period is extended to 6 h and 24 h to assess the generalizability. The ML algorithms' generalizability can be regarded as the ability of the model to predict the output accurately when the unseen feature values are fed to the algorithm. As such, the results for 6-h ahead and 24-h ahead predictions are shown in Tables 2 and 3 respectively.

As can be seen, both MLP and SVM can reasonably tackle the non-stationary problem associated with CO2 forecasting, with the former outperforming the latter for different forecasting horizons on average according to simulation performance. The better prediction accuracy of MLP compared to SVM is also depicted in Fig. 6, where MLP follows the pattern of CO2 changes more quickly and with higher accuracy. As

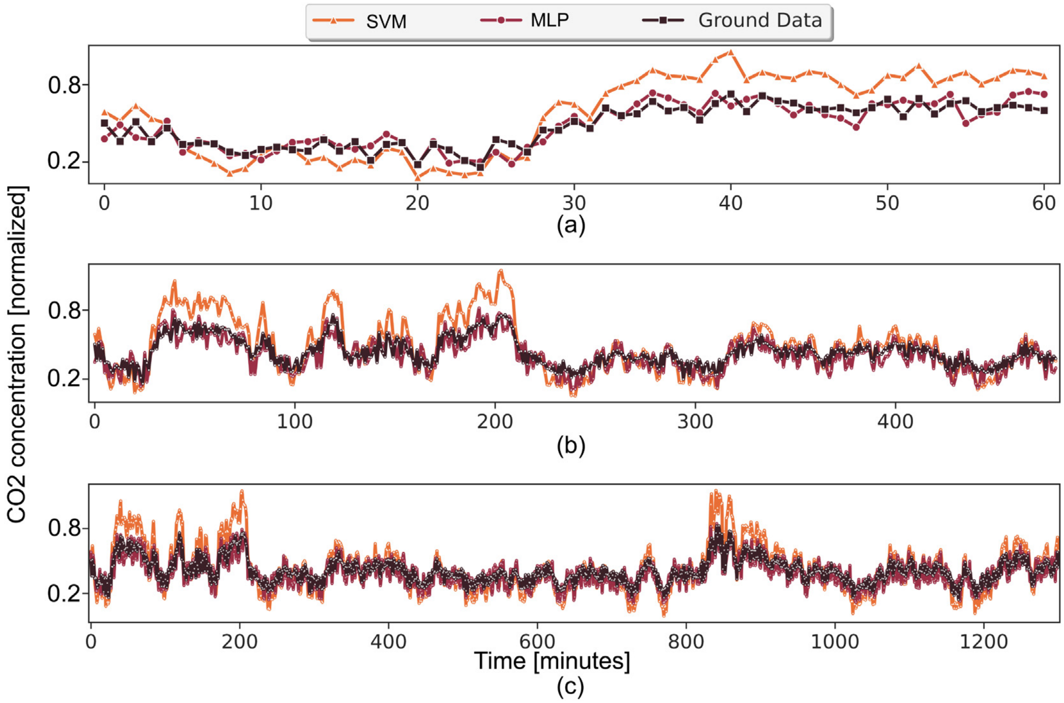

Fig. 6. CO2 concentration forecasting of MLP and SVM across different time horizons (a) one-hour in advance, (b) 6-h in advance, (c) 24-h in advance.

Table 3
CO2 forecasting performance of the ML models—24h-ahead horizon.

Phase	Metric	SVM	RF	LR	AdB	MLP	GB
Training	RMSE	45.03	46.48	46.99	46.48	**44.39**	49.07
	MAE	44.21	47.24	49.26	42.42	**40.16**	46.17
	MAPE	21.64	14.93	22.42	19.06	19.45	**13.78**
	R^2	0.780	0.835	0.792	**0.843**	0.819	0.779
Cross validation	RMSE	55.66	56.84	58.62	57.23	**54.78**	59.67
	MAE	**52.16**	57.39	53.33	51.59	55.19	57.28
	MAPE	25.98	20.39	26.71	22.99	22.27	**18.13**
	R^2	0.719	0.769	0.713	0.772	**0.805**	0.721

Table 4
Hyperparameters associated with the proposed MLP algorithm for CO2 prediction.

Activation function	Sigmoid
Number of hidden layers	4
Hidden layer nodes	130
Solver	AdaDelta
Learning rate	0.01
Validation fraction	0.2
Random state	None

such, MLP is used in further experiments. All the configuration files are developed and run using Python and Tensorflow2 as backend.

Developing an efficient machine learning model is a complicated and time-consuming process that requires selecting the right procedure and adjusting the model's hyperparameters to achieve the optimum network. Hyperparameters are used to either setup a machine learning model (for example, the error threshold e in SVM or the learning rate for training the network) or to specify the algorithm used to minimize the loss function (e.g., the activation function and optimizer types of the networks). Hyperparameters, in general, define both the layout and training of a neural network. The number of network layers, nodes in each layer, activation mechanism, and other ML architectures features are all hyperparameters. Hyperparameter tuning is the process of designing the optimum model architecture with the optimal hyperparameter setting. Traditional methods for tuning hyperparameters include manual testing, grid search, and decision-theoretic optimization. We utilized a decision-theoretic optimization method in this research, tweaking the hyperparameters using a Python program named "HyperOpt". Additional information about this function and its implementation is available in [48]. The hyperparameters associated with the ultimate MLP model are listed in Table 4.

3.3. Model validation

The number of occupants and their activities is uncertain to a great extent. As such, accurate CO2 prediction is a critical task in achieving DCV due to the CO2 high correlation with the volume of activities and number of people in a particular space. ASHRAE [51] suggests the following formula as the governing equation for the indoor CO2 models to calculate the relationship between CO2 concentration and number of occupants:

$$V\frac{dC}{dt} = QC_0 - QC + G.P \tag{11}$$

where V is the room volume [m^3], C is the CO2 concentration in the room [ppm], C_0 is the outdoor CO2 concentration [ppm], Q is the fan flow rate [m^3/s], G is the CO2 generation per person [ppm m^3/s], and P is the room occupancy [number of people in the room]. Assuming constant occupancy/generation and ventilation rates in time, the governing relationship during the transient phase is as follows:

$$C(t) = C_0 + \frac{G.P}{Q}\left(1 - e^{\frac{-Qt}{V}}\right) \tag{12}$$

where Q/V is referred to as the building's exterior air change volume, and its inverse, V/Q, is referred to as the system's time constant. If the class starts the day with the outside CO2 concentration and is then filled, the indoor CO2 concentration would continue to increase at a rate determined by the ratio of fan ventilation rate to the room volume. The room would ultimately have a steady concentration, as time goes to infinity, provided by the expression in (13).

$$C_u = C_0 + \frac{G.P}{Q} \tag{13}$$

Fig. 7 represents the equilibrium CO2 concentration vs the fan ventilation rate for different occupancy levels. As more people occupy the

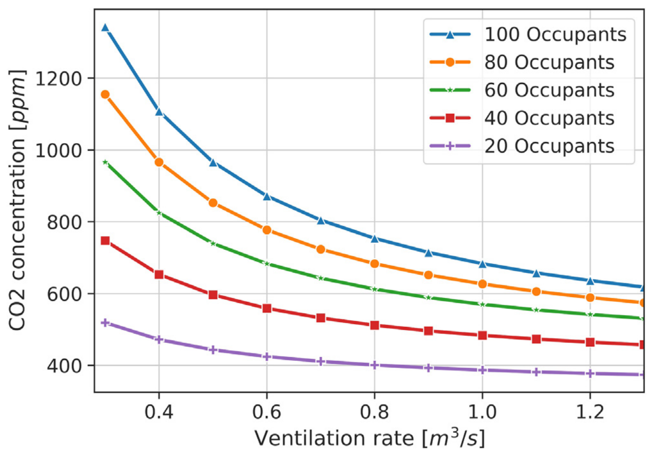

Fig. 7. Equilibrium CO2 concentration and ventilation rates at different occupancy levels.

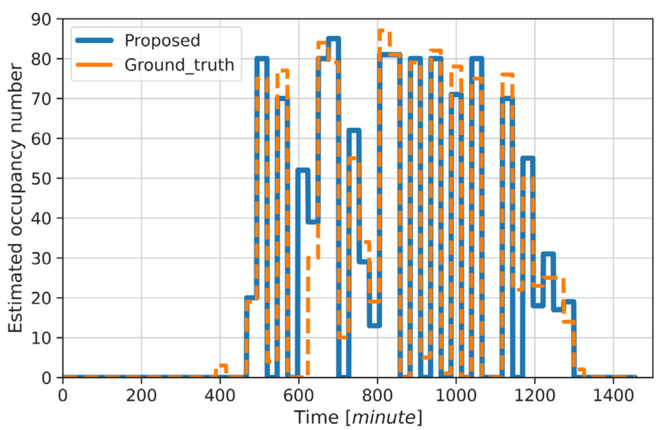

Fig. 8. Occupancy estimates of the subject campus classroom for a chosen day (Nov, 08).

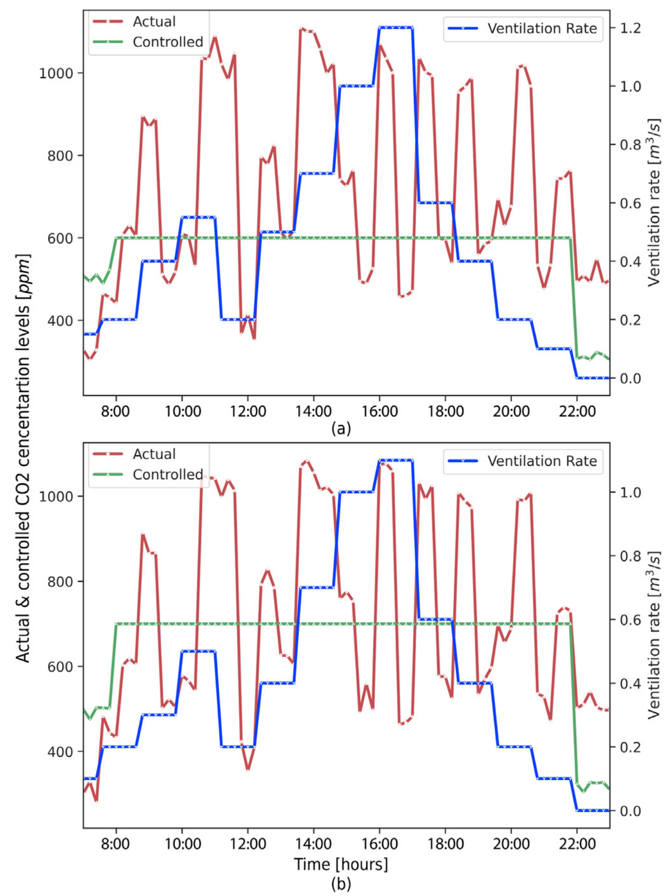

Fig. 9. Controlled ventilation to maintain the necessary amount of CO2 indoors, (a) 600 ppm setpoint, (b) 700 ppm setpoint.

space, ventilation must increase to retain the same CO2 concentration. Eqs. (11)–(13) are used to validate the occupancy estimation results for one of the classrooms where the size of the hall, number of occupants, and occupant activities are accessible. The Onset CO2 data logger was used to measure the CO2 in the room with the sampling rate of 1 sample/min. The occupancy number responsible for the rate of change of CO2 is calculated for the class timings throughout the week. The classroom volume is 500 m^3 and the maximum flowrate of the HVAC system is 0.9 m^3/s. The occupancy state's period and the pace at which people entered and left the zone were compared to actual data to determine the occupancy model's performance. Fig. 8 illustrates the occupancy estimate for a day that indicates that the occupancy rate assumption is reasonable. From Figs. 2 and 5, it is evident that CO2 concentration exceeded 1000 ppm as the occupancy increased, making the classroom under-ventilated. As represented, the simulation results followed the trend of actual data. As such, the proposed model can fairly mimic the response characteristics of the existing system. Based on this prediction, a monitoring mechanism that ensures that the CO2 concentrations never go beyond a predetermined threshold can be put in place.

3.4. Demand controlled ventilation

For CO2 thresholds of 600 ppm and 700 ppm, the controlled ventilation and indoor CO2 are shown in Fig. 9. As depicted, the controller tries to track these setpoints by lowering the ventilation rates whenever the occupancy drops and increasing the ventilation rate when the number of occupants rises. The ventilation rates for the 600 ppm threshold are usually higher than the 700 ppm threshold as the fan has to work more to compensate for the excessive CO2 generation. This means that the CO2 setpoint level has a massive impact on the energy consumption level of the system. For example, for setpoints 600 ppm and 700 ppm, the overall indoor ventilation (fresh air) for the chosen day (11/08/2018) are 22 438.8 m^3 and 18 860.6 m^3, respectively. The total indoor ventilation and the fan energy consumption for different setpoints, ranging from 500 to 1000 ppm, are shown in Table 5. This table also displays the total energy consumption difference between the current system (with an average fan flowrate of 0.9 m^3/s) and the proposed DCV system. As can be seen, the total fresh air and energy consumption rates are reduced in higher CO2 thresholds. Increased energy savings can also be obtained by increasing the CO2 setpoint values, since the fan needs to work less in this case. However, significant energy savings are anticipated as compared to the conventional ON/OFF control scheme. This is due to the fan's flexibility to decrease the flowrate whenever occupancy is dropping and also the capacity of increasing the flowrate when CO2 is highly generated. To further assess the fan energy consumption based on the ventilation rate, Eq. (14) is used.

$$E = d_p \times Q \times H \tag{14}$$

4. Conclusion

This research provided a step-by-step methodology for controlling the HVAC system of a campus classroom based on the level of CO2

Table 5
Average fan flowrate, total fresh air, and energy consumption comparison for different DCV setpoints.

System	Average fan flowrate [m^3/s]	Operation hours	Total fresh air [m^3]	Energy consumption [kWh]	Energy saving [%]
Current	0.9000	10	32 400.0	36.0	–
DCV-500 ppm	0.7435	10	26 766.0	29.7	17.5
DCV-600 ppm	0.6233	10	22 438.8	24.9	30.8
DCV-700 ppm	0.5238	10	18 860.6	21.0	41.6
DCV-800 ppm	0.4873	10	17 542.8	19.5	45.8
DCV-900 ppm	0.4513	10	16 246.8	18.1	49.7
DCV-1000 ppm	0.4366	10	15 717.6	17.5	51.4

concentration in the environment. In a detailed comparison study, six of the most advanced machine-learning algorithms in the field were selected and fine-tuned for the purpose of CO2 prediction. As a result of its great ability to learn nonlinearities connected with the CO2 data, the multilayered perceptron network outperforms other representative algorithms, according to our findings. We created an occupant-centric control strategy for monitoring the levels of indoor air quality based on the ultimate machine learning model. Model-based control schemes beat conventional ON/OFF controllers in terms of control precision and energy savings, as demonstrated by our research. The proposed model and control technique proved to be effective in reducing total energy usage by up to 51.4 percent, according to the results. Using the dynamic behavior of the occupancy patterns, as well as the uncertainties and disturbances in the system, the control strategy was able to address the shortcomings of the current control system. Unlike other controllers such as MPC, this control system may be constructed quickly and at a minimal cost, and it does not necessitate the use of additional computer power due to its simplicity of design using Hardware-in-the-Loop (HIL) testing for validation, this control system can be connected to the HVAC fan for full-fledged testing and validation in the future as part of a research project.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Zhihong Pang, Yan Chen, Jian Zhang, Zheng O'Neill, Hwakong Cheng, Bing Dong, How much hvac energy could be saved from the occupant-centric smart home thermostat: A nationwide simulation study, Appl. Energy 283 (2021) 116251, https://www.sciencedirect.com/science/article/pii/S0306261920316421.

[2] Sokratis Papadopoulos, Constantine E. Kontokosta, Alex Vlachokostas, Elie Azar, Rethinking hvac temperature setpoints in commercial buildings: The potential for zero-cost energy savings and comfort improvement in different climates, Build. Environ. 155 (2019) 350–359, https://www.sciencedirect.com/science/article/pii/S036013231930232X.

[3] Mingya Zhu, Yiqun Pan, Zejun Wu, Jiantong Xie, Zhizhong Huang, Risto Kosonen, An occupant-centric air-conditioning system for occupant thermal preference recognition control in personal micro-environment, Build. Environ. 196 (2021) 107749, https://www.sciencedirect.com/science/article/pii/S0360132321001578.

[4] Farrokh Jazizadeh, Vedant Joshi, Francine Battaglia, Adaptive and distributed operation of hvac systems: Energy and comfort implications of active diffusers as new adaptation capacities, Build. Environ. 186 (2020) 107089, https://www.sciencedirect.com/science/article/pii/S0360132320304650.

[5] Eva Schito, Paolo Conti, Luca Urbanucci, Daniele Testi, Multi-objective optimization of hvac control in museum environment for artwork preservation, visitors thermal comfort and energy efficiency, Build. Environ. 180 (2020) 107018, https://www.sciencedirect.com/science/article/pii/S036013232030398X.

[6] K.E. Mary Reena, Abraham T. Mathew, Lillykutty Jacob, A flexible control strategy for energy and comfort aware hvac in large buildings, Build. Environ. 145 (2018) 330–342, https://www.sciencedirect.com/science/article/pii/S036013231830564X.

[7] S. Taheri, R. Ghoraani, A. Pasban, M. Moeini-Aghtaie, A. Safdarian, Stochastic framework for planning studies of energy systems: a case of EHs, IET Renew. Power Gener. 14 (2020) 435–444, http://dx.doi.org/10.1049/iet-rpg.2019.0642.

[8] Shiyu Yang, Man Pun Wan, Wanyu Chen, Bing Feng Ng, Swapnil Dubey, Model predictive control with adaptive machine-learning-based model for building energy efficiency and comfort optimization, Appl. Energy 271 (2020) 115147, https://www.sciencedirect.com/science/article/pii/S0306261920306590.

[9] Tianyi Zhao, Ying Zhou, Jili Zhang, Xiuming Li, Online differential pressure reset method with adaptive adjustment algorithm for variable chilled water flow control in central air-conditioning systems, in: Building Simulation, Springer, 2021, pp. 1–16.

[10] Rodney Hurt, Gang Wang, Song Li, Experimental validation of cooling coil control valve performance with cascade control, ASHRAE Trans. 126 (2) (2020) 233, https://www.ulib.iupui.edu/cgi-bin/proxy.pl?url=https://search.ebscohost.com/login.aspx?direct=true&db=vsh&AN=146858050&site=eds-live.

[11] M.D. Shamim Ahamed, Radu Zmeureanu, Nunzio Cortrufo, Jose Candanedo, Gray-box virtual sensor of the supply air temperature of air handling units, Sci. Technol. Built Environ. 26 (8) (2020) 1151–1162, URL https://doi.org/10.1080/23744731.2020.1785812.

[12] Chen Wu, Zhou Xingxi, Deng Shiming, Development of control method and dynamic model for multi-evaporator air conditioners (meac), Energy Convers. Manage. 46 (3) (2005) 451–465, https://www.sciencedirect.com/science/article/pii/S0196890404000834.

[13] Xiuming Li, Tianyi Zhao, Jili Zhang, Tingting Chen, Predication control for indoor temperature time-delay using elman neural network in variable air volume system, Energy Build. 154 (2017) 545–552, https://www.sciencedirect.com/science/article/pii/S0378778817319035.

[14] Iman Fakhari, Amirmohammad Behzadi, Ehsan Gholamian, Pouria Ahmadi, Ahmad Arabkoohsar, Design and tri-objective optimization of a hybrid efficient energy system for tri-generation, based on PEM fuel cell and MED using syngas as a fuel, J. Cleaner Prod. 290 (2021) 125205, https://www.sciencedirect.com/science/article/pii/S0959652620352495.

[15] Iman Fakhari, Amirmohammad Behzadi, Ehsan Gholamian, Pouria Ahmadi, Ahmad Arabkoohsar, Comparative double and integer optimization of low-grade heat recovery from PEM fuel cells employing an organic rankine cycle with zeotropic mixtures, Energy Convers. Manage. 228 (2021) 113695, https://www.sciencedirect.com/science/article/pii/S019689042031219X.

[16] Raad Z. Homod, Assessment regarding energy saving and decoupling for different ahu (air handling unit) and control strategies in the hot-humid climatic region of iraq, Energy 74 (2014) 762–774.

[17] A. Berouine, E. Akssas, Y. Naitmalek, F. Lachhab, M. Bakhouya, R. Ouladsine, M. Essaaidi, A fuzzy logic-based approach for hvac systems control, in: 2019 6th International Conference on Control, Decision and Information Technologies (CoDIT), 2019, pp. 1510–1515.

[18] Zhipeng Deng, Qingyan Chen, Reinforcement learning of occupant behavior model for cross-building transfer learning to various hvac control systems, Energy Build. 238 (2021) 110860, https://www.sciencedirect.com/science/article/pii/S0378778821001444.

[19] Sholahudin Nasruddin, Pujo Satrio, Teuku Meurah Indra Mahlia, Niccolo Giannetti, Kiyoshi Saito, Optimization of hvac system energy consumption in a building using artificial neural network and multi-objective genetic algorithm, Sustain. Energy Technol. Assess. 35 (2019) 48–57, https://www.sciencedirect.com/science/article/pii/S2213138818305629.

[20] Maryam Gholamzadehmir, Claudio Del Pero, Simone Buffa, Roberto Fedrizzi, Niccolo' Aste, Adaptive-predictive control strategy for hvac systems in smart buildings– a review, Sustainable Cities Soc. 63 (2020) 102480, https://www.sciencedirect.com/science/article/pii/S2210670720307009.

[21] Mohammad Esrafilian-Najafabadi, Fariborz Haghighat, Occupancy-based hvac control systems in buildings: A state-of-the-art review, Build. Environ. (2021) 107810, https://www.sciencedirect.com/science/article/pii/S0360132321002171.

[22] A. Ahmadian, A. Shafiee, M. Alidoost, A. Akbari, Flexible paper-based li-ion batteries: A review, World J. Eng. Technol. 9 (2021) 285–299, http://dx.doi.org/10.4236/wjet.2021.92020.

[23] D. Mariano-Hernandez, L. Hernandez-Callejo, A. Zorita-Lamadrid, O. Duque-Perez, F. Santos Garcia, A review of strategies for building energy management system: Model predictive control, demand side management, optimization, and fault detect & diagnosis, J. Build. Eng. 33 (2021) 101692, https://www.sciencedirect.com/science/article/pii/S2352710220310627.

[24] Shide Salimi, Amin Hammad, Optimizing energy consumption and occupants comfort in open-plan offices using local control based on occupancy dynamic data, Build. Environ. 176 (2020) 106818, https://www.sciencedirect.com/science/article/pii/S0360132320301761.

[25] Wooyoung Jung, Farrokh Jazizadeh, Human-in-the-loop hvac operations: A quantitative review on occupancy, comfort, and energy-efficiency dimensions, Appl. Energy 239 (2019) 1471–1508, https://www.sciencedirect.com/science/article/pii/S030626191930073X.

[26] Xing Lu, Tao Yang, Zheng O'Neill, Xiaohui Zhou, Zhihong Pang, Energy and ventilation performance analysis for co2-based demand-controlled ventilation in multiple-zone vav systems with fan-powered terminal units (ASHRAE RP-1819), Sci. Technol. Built Environ. 27 (2) (2021) 139–157, URL https://doi.org/10.1080/23744731.2020.1831318.

[27] Wanyu R. Chan, Xiwang Li, Brett C. Singer, Theresa Pistochini, David Vernon, Sarah Outcault, Angela Sanguinetti, Mark Modera, Ventilation rates in california classrooms: Why many recent hvac retrofits are not delivering sufficient ventilation, Build. Environ. 167 (2020) 106426, https://www.sciencedirect.com/science/article/pii/S0360132319306365.

[28] Andrew Persily, Challenges in developing ventilation and indoor air quality standards: The story of ashrae standard 62, Build. Environ. 91 (2015) 61–69, https://www.sciencedirect.com/science/article/pii/S0360132315000839. Fifty Year Anniversary for Building and Environment.

[29] Seung Ho Ryu, Hyeun Jun Moon, Development of an occupancy prediction model using indoor environmental data based on machine learning techniques, Build. Environ. 107 (2016) 1–9, https://www.sciencedirect.com/science/article/pii/S0360132316302463.

[30] H. Elkhoukhi, M. Bakhouya, M. Hanifi, D. El Ouadghiri, On the use of deep learning approaches for occupancy prediction in energy efficient buildings, in: 2019 7th International Renewable and Sustainable Energy Conference (IRSEC), 2019, pp. 1–6.

[31] Bing Dong, Burton Andrews, Sensor-based occupancy behavioral pattern recognition for energy and comfort management in intelligent buildings, in: Proceedings of Building Simulation, 2009, pp. 1444–1451.

[32] Paige Wenbin Tien, Shuangyu Wei, John Kaiser Calautit, Jo Darkwa, Christopher Wood, A vision-based deep learning approach for the detection and prediction of occupancy heat emissions for demand-driven control solutions, Energy Build. 226 (1) (2020) 110386, https://www.sciencedirect.com/science/article/pii/S0378778820316820.

[33] Hao-Cheng Zhu, Chen Ren, Shi-Jie Cao, Fast prediction for multi-parameters (concentration, temperature and humidity) of indoor environment towards the online control of HVAC system, Build. Simul. 14 (1) (2021) 649—665, https://rdcu.be/cnq08.

[34] Quinten Carton, Bart Merema, Hilde Breesch, Finding the optimal set points of a thermal and ventilation control system under changing outdoor weather conditions, E3S Web Conf. 246 (1) (2021) 11006, http://dx.doi.org/10.1051/e3sconf/202124611006.

[35] Ki Uhn Ahn, Cheol Soo Park, Application of deep Q-networks for model-free optimal control balancing between different HVAC systems, Sci. Technol. Built Environ. 26.1 (1) (2020) 61–74, http://dx.doi.org/10.1080/23744731.2019.1680234.

[36] Johanna Kallio, Jaakko Tervonen, Pauli Rasanen, Riku Makynen, Jani Koivusaari, Johannes Peltola, Forecasting office indoor co2 concentration using machine learning with a one-year dataset, Build. Environ. 187 (2021) 107409, https://www.sciencedirect.com/science/article/pii/S0360132320307770.

[37] Igor Skrjanc, Barbara Subic, Control of indoor co2 concentration based on a process model, Autom. Constr. 42 (2014) 122–126, https://www.sciencedirect.com/science/article/pii/S0926580514000375.

[38] Hongbin Liu, SeungChul Lee, MinJeong Kim, Honglan Shi, Jeong Tai Kim, ChangKyoo Yoo, Finding the optimal set points of a thermal and ventilation control system under changing outdoor weather conditions, Indoor Built Environ. 23 (1) (2014) 118–132, URL https://doi.org/10.1177/1420326X14522669.

[39] Ali Razban, Arash Edalatnoor, David Goodman, Jie Chen, Energy optimization of air handling unit using co2 data and coil performance, in: ASME 2016 International Mechanical Engineering Congress and Exposition, American Society of Mechanical Engineers Digital Collection, 2016.

[40] Nabil Nassif, A robust co2-based demand-controlled ventilation control strategy for multi-zone hvac systems, Energy Build. 45 (2012) 72–81, https://www.sciencedirect.com/science/article/pii/S0378778811004622.

[41] H. Wang, J. Zhao, J. Duan, M. Wang, Z. Dong, Greenhouse co2 control based on improved genetic algorithm and fuzzy neural network, in: 2018 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), 2018, pp. 1537–1540, http://dx.doi.org/10.1109/IMCEC.2018.8469546.

[42] Dimitris Lazos, Alistair B. Sproul, Merlinde Kay, Development of hybrid numerical and statistical short term horizon weather prediction models for building energy management optimisation, Build. Environ. 90 (2015) 82–95, https://www.sciencedirect.com/science/article/pii/S0360132315001341.

[43] Hansaem Park, Dong Yoon Park, Comparative analysis on predictability of natural ventilation rate based on machine learning algorithms, Build. Environ. 195 (2021) 107744, https://www.sciencedirect.com/science/article/pii/S0360132321001529.

[44] Tanveer Ahmad, Huanxin Chen, A review on machine learning forecasting growth trends and their real-time applications in different energy systems, Sustainable Cities Soc. 54 (2020) 102010, https://www.sciencedirect.com/science/article/pii/S2210670719335516.

[45] Corinna Cortes, Vladimir Vapnik, Support-vector networks, Mach. Learn. 20 (3) (1995) 273–297, URL https://doi.org/10.1007/BF00994018.

[46] Haitao Wang, Daoguang Feng, Kai Liu, Fault detection and diagnosis for multiple faults of vav terminals using self-adaptive model and layered random forest, Build. Environ. 193 (2021) 107667, https://www.sciencedirect.com/science/article/pii/S0360132321000780.

[47] Xinyue Li, Shuqin Chen, Hongliang Li, Yunxiao Lou, Jiahe Li, Multi-dimensional analysis of air-conditioning energy use for energy-saving management in university teaching buildings, Build. Environ. 185 (2020) 107246, https://www.sciencedirect.com/science/article/pii/S036013232030617X.

[48] Saman Taheri, Mohammad Jooshaki, Moein Moeini-Aghtaie, Long-term planning of integrated local energy systems using deep learning algorithms, Int. J. Electr. Power Energy Syst. 129 (2021) 106855, https://www.sciencedirect.com/science/article/pii/S0142061521000958.

[49] A. Ahmadi, M. Nabipour, B. Mohammadi-Ivatloo, V. Vahidinasab, Ensemble learning-based dynamic line rating forecasting under cyberattacks, IEEE Trans. Power Deliv. (2021) 1, http://dx.doi.org/10.1109/TPWRD.2021.3056055.

[50] A. Géron, Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems, O'Reilly Media, Sebastopol CA, USA, 2019.

[51] Jiaming Li, Josh Wall, Glenn Platt, Indoor air quality control of hvac system, in: Proceedings of the 2010 International Conference on Modelling, Identification and Control, 2010, pp. 756–761.