Journal of Building Engineering 97 (2024) 110760

Contents lists available at ScienceDirect

Journal of Building Engineering

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

Assessment of airborne transmitted infection risk in classrooms using computational fluid dynamics and machine learning-based surrogate modeling

Hyeonjun Lee, Donghyun Rim *

Architectural Engineering Department, Pennsylvania State University, 222 Engineering Unit A, University Park, PA 16801, USA

ARTICLE INFO

Keywords:
Computational fluid dynamics
Data-driven machine learning
Surrogate modeling
Airborne transmission
Infection risk

ABSTRACT

In recent years, the assessment of airborne transmitted infection risk has been extensively performed using Computational Fluid Dynamics (CFD) simulations, especially in response to the COVID-19 pandemic. Nevertheless, the high computational demands and time-intensive nature of CFD simulations highlight the need for fast or real-time infection risk predictions. This capability is crucial for swift decision-making in dynamic environments where timely health interventions are critical. This paper presents a thorough analysis of airborne infection risks in classroom environments based on CFD simulations to understand key factors such as ventilation strategies, air change rates, occupant arrangements, source locations, and particle sizes. This study also employs data-driven supervised learning methods—specifically, Long Short-Term Memory (LSTM) and Artificial Neural Networks (ANN)—to generate surrogate models for predicting airborne infection risk. Key findings reveal that different ventilation strategies significantly affect airborne infection risk, reducing it by 49 %–77 %. Moreover, the conventional Wells-Riley model was identified as lacking in its ability to accurately predict local infection risks. The study further challenges the assumption that higher air change rates are universally beneficial, considering that occupants seated in the back rows of a classroom experienced up to a 166 % increased risk, despite elevating air change rates from 1.1 h^{-1} to 11 h^{-1}. These results suggest that physical distancing alone may be insufficient and highlight the importance of considering other factors such as occupant arrangements. Regarding the model performance, the ANN-based surrogate model demonstrated varying prediction accuracy. For inhalable particle concentration predictions for susceptible occupants, R^2 values ranged from 0.31 to 0.65 with CVRMSE values between 100 % and 180 %. In contrast, the model achieved an R^2 of 0.79 and a CVRMSE of 34 % for infectors. The insights and methodologies from this study can inform HVAC system design and operation strategies to better mitigate infectious disease transmission in densely occupied indoor environments.

1. Introduction

The COVID-19 pandemic has left enduring imprint on the annals of human history, inflicting a profound and far-reaching toll on public health, daily life, and worldwide economy. The repercussions for public health have been manifold, including overwhelming healthcare facilities [1], instigating global mental health crises [2], and widening the chasm of healthcare disparities [3]. While the

* Corresponding author.
E-mail addresses: hzl5603@psu.edu (H. Lee), dxr51@psu.edu (D. Rim).

https://doi.org/10.1016/j.jobe.2024.110760
Received 7 July 2024; Received in revised form 28 August 2024; Accepted 13 September 2024
Available online 17 September 2024

SARS-CoV-2 can be transmitted by both direct and indirect contact, it has been notably established that the primary mode of infection is through airborne transmission [4].

One of the most frequently used approaches is to solve the transport equations and numerically model fluid flow and particle transport, thus enabling quantitative risk assessment. Due to the computational complexity of these equations, especially in turbulent flows, researchers often employ Reynolds-Averaged Navier-Stokes (RANS) equation to compromise the calculations. This approach has made Computational Fluid Dynamics (CFD) simulation an increasingly popular tool for studying the risk of infection, as it offers detailed insights into the spatial dispersion and transmission of infectious aerosols. Thanks to its inherent advantages, CFD has empowered researchers to investigate various indoor operating conditions and assess their impacts on infection risk, encompassing factors such as ventilation strategies [5,6], air change rate [7], source location [8,9], social distancing [10,11], air distribution [12,13], and personal protective equipment [14–16]. Nonetheless, given its substantial computational demands and time requirements, CFD does have limitations in examining general infection risk, especially for fast or real-time predictions [17,18].

In response to these challenges, machine learning-based surrogate models have been gaining significant popularity for their potential to reduce computational costs and improve simulation efficiency. These surrogate models serve as approximation methods to emulate the behavior of complex simulations in a computationally efficient manner. By leveraging detailed CFD simulation results, they capture the high-dimensional, nonlinear relationships between boundary conditions and their outcomes. Although generating data for these models can be computationally expensive and repetitive, leading to data scarcity, their high accuracy has made them a widely studied area [19]. For example, machine learning-based surrogate models have been developed using CFD data to predict velocity fields around different objects in real time [20–22], airborne pollutant dispersion on an urban scale [23,24], and wind velocity by wind turbine blades [25]. These studies demonstrate that surrogate models enable real-time or much faster predictions while significantly reducing computational demands.

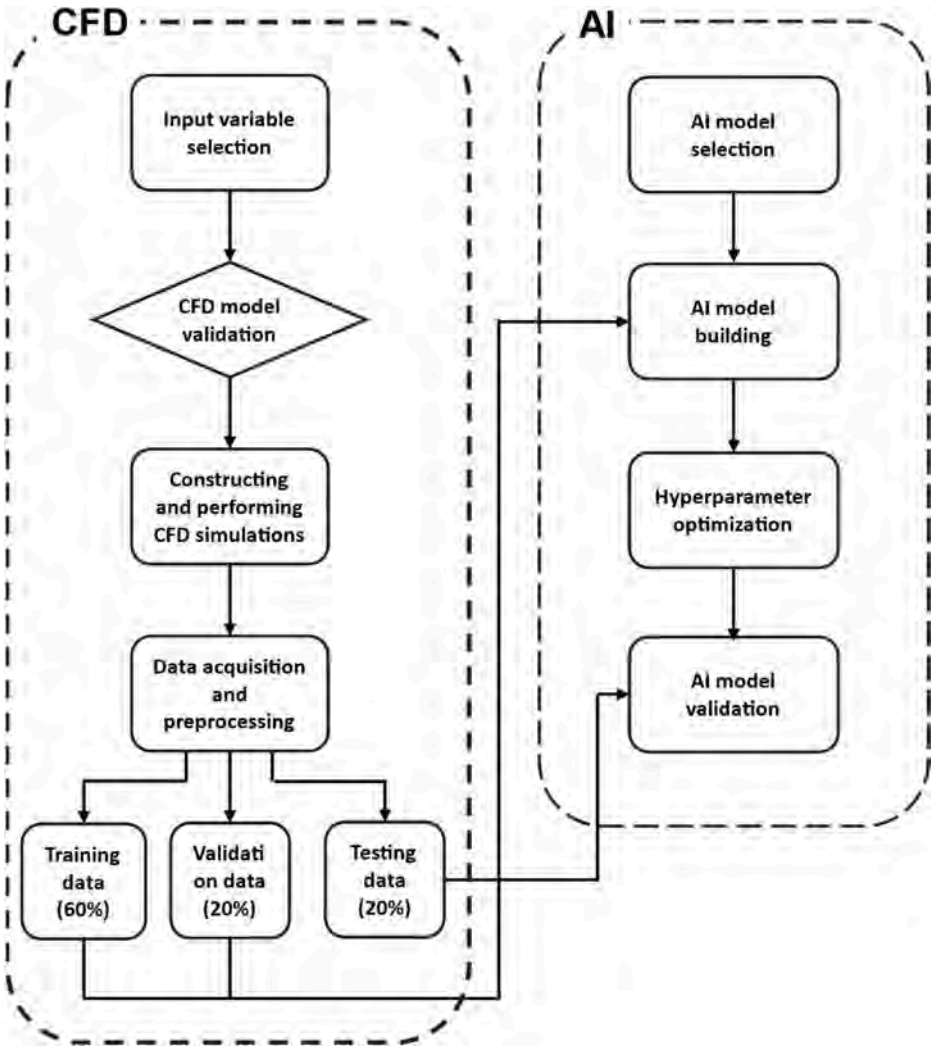

Fig. 1. Flow chart of the study.

In the context of indoor environments, predictions of airflow patterns and temperature distribution are commonly performed. For example, Quang et al. [26] trained a Deep Neural Network on indoor airflow distribution in multi-story buildings using CFD simulations, reducing computational time by 80 % compared to traditional CFD simulations while maintaining accuracy. Similarly, Zhou and Ooka [27,28] and Faulkner et al. [29] developed machine-learned surrogate models to predict indoor airflow patterns and temperature based on CFD data. Real-time predictions of thermal comfort have also been investigated in various settings, such as stadiums [30], vehicles [31], and air cabins [32]. For airborne particle transportation, Mesgarpour et al. [33] proposed a computational framework for a Multi-Input Multi-Output (MIMO) AI model based on CFD results to rapidly predict droplet spread from a sneezing passenger in a bus. However, no attempt has yet been made to surrogate airborne infection risk that provides real-time infection risk prediction.

Building on this background, our study has two primary objectives. First, we aim to perform a comprehensive analysis of infection risk in a classroom setting using CFD simulations. This involves investigating the impact of various parameters—including ventilation mode, occupant arrangement, air change rate, and distance from the infector—on each occupant's infection risk, with the goal of assessing their suitability as inputs for a surrogate model. Second, based on the simulation results, we will train a machine learning model to develop a surrogate model capable of quickly predicting the inhalable particle concentration for each occupant.

2. Method

This study is organized into two sections: 1) Computational Fluid Dynamics (CFD) and 2) Artificial Intelligence (AI). Infection risks estimated by CFD simulations under various scenarios were used to train and test models in the AI section. Fig. 1 presents the overall flow of the two tasks. The flowchart detailed a methodology for integrating CFD simulations with AI to predict inhalable particle concentration for each occupant. It started with selecting input variables and validating the CFD model. Following the model validation, CFD simulations with varied parameters were conducted, and the resulting data were preprocessed through cleaning, normalization, and feature extraction. The data were then split into training, validation, and testing sets. An AI model was selected and optimized, with performance checks ensured using metrics such as R^2, RMSE, and CVRMSE. Finally, the predictions by the surrogate model were compared with CFD testing data that were kept separate from the training data.

2.1. Description of geometry

For the CFD model, the Department of Energy (DOE) reference building specifications for a secondary classroom, a space measuring

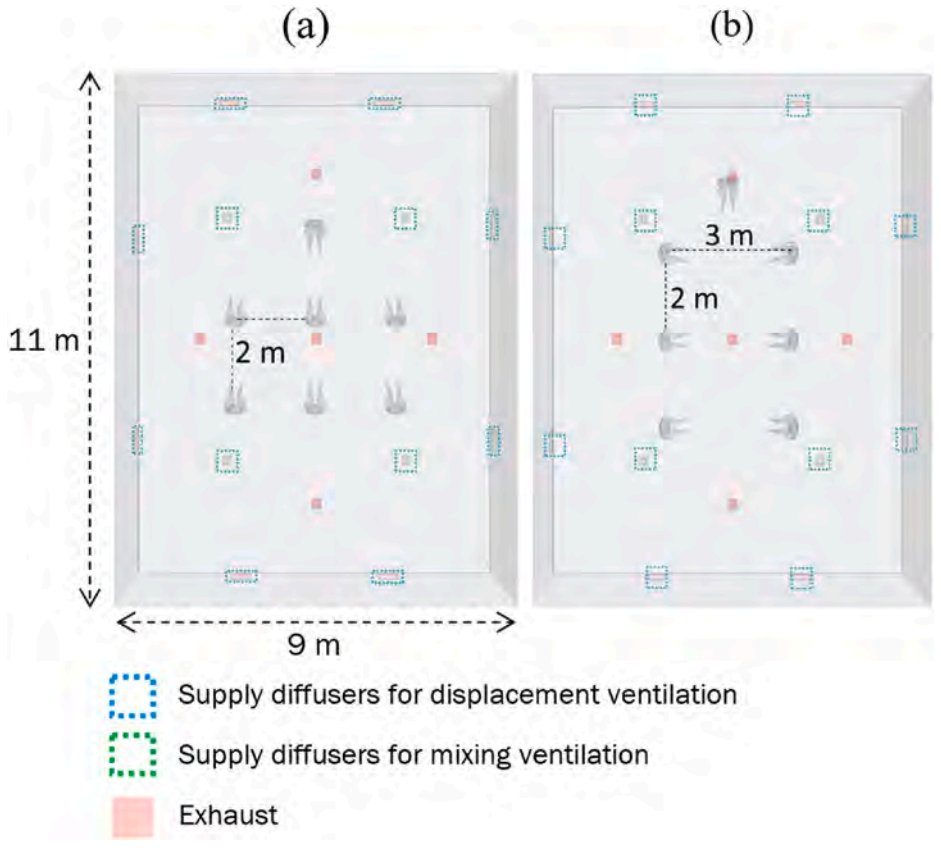

Fig. 2. Top view of two different arrangements a) normal classroom arrangement b) debate-style arrangement.

9 m × 11 m x 3 m (length × width × height), were configured to replicate a standard small classroom setting accommodating seven sedentary occupants. It is important to note that the selected occupancy level of seven individuals was specifically chosen for this research to achieve two objectives: 1) to systematically investigate the effects of occupant arrangement on particle distribution, and 2) to simplify the problem complexity for the AI model. Simplifying the problem complexity is crucial for enhancing computational efficiency and ensuring that the AI model can effectively capture and analyze non-linear interactions within the classroom environment. Fig. 2 illustrates two different configurations: one with mixing supply diffusers and the other with displacement supply diffusers, both using shared exhausts. In the mixing configuration, four 0.2 m × 0.2 m diffusers were located on the ceiling, each blocked at the center. These diffusers supply air at a 30-degree angle from the ceiling in four directions (east, west, north, and south), promoting uniform air distribution throughout space. In contrast, the displacement ventilation configuration features eight 0.5 m × 0.5 m diffusers situated on the lower side of each wall. This setup creates vertical temperature stratification [6]. Fig. 2a shows the first occupant arrangement. An instructor was located at the front of the room with six students facing the instructor, each maintaining 2-m distance from each adjacent occupant to delineate typical classroom setting. Fig. 2b illustrates the second arrangement with six occupants facing each other and one occupant at the top to emulate debate-style setting. The two arrangements were considered to capture two common set up of classroom as well as increasing diversity of occupant arrangements.

2.2. Sensitivity analysis: grid and time step

This study primarily aims to investigate the release of particulate matter from an individual identified as an infector and assess the subsequent exposure experienced by individuals classified as susceptible. In the simulation, the choice of mesh quality was critical to ensure an accurate representation of the physical phenomena being studied. In order to represent the complex human model, a combination of surface remeshing, polyhedral meshing, and prism layer meshing were employed. To accurately capture airflow and mass transfer phenomena near the occupant surfaces, grid sizes of y + value < 7 were generated. This approach adheres to the recommendations by Pei and Rim. [34], which advocate for y + values below 10 to obtain reliable estimations of airflow and mass transfer rate adjacent to the human surface.

To ensure the reliability of the generated computational mesh, the Grid Convergence Index (GCI) method was utilized to assess discretization errors [35]. Steady-state velocity measurements were obtained at a vertical distance of 0.2 m above the occupant's head. These measurements were conducted for three separate mesh resolutions, designated as Grid 1 (623,416 cells), Grid 2 (417,770 cells), and Grid 3 (200,967 cells). Subsequently, the GCI was calculated in accordance with Eq. (1). A comprehensive list of the mesh resolutions and their respective GCI values is presented in Table 1. Comparative analyses of the steady-state velocities across the three mesh resolutions are depicted in Fig. S 1 in Supplementary materials. In consideration of both computational efficiency and grid convergence metrics, Grid 2 was examined for use in the remaining portions of the study.

$$GCI_{fine}^{21} = \frac{1.25 \times e_a^{21}}{r_{21}^{p} - 1} \tag{1}$$

Where e_a^{21} is approximate relative error of parameter interested, r is grid refinement factor, and p is apparent order of convergence.

In CFD simulations, the choice of an optimal time step size holds great significance, particularly in situations characterized by unsteady conditions. The present section seeks to identify a time step size that not only ensures computational efficiency but also yields a converged solution. To this end, three disparate time step sizes—0.5 s, 0.1 s, and 0.01 s—were systematically evaluated. These simulations were conducted in the context of a conventional classroom setting equipped with mixing ventilation operating at an air change rate of 11.1 h^{-1}. Table 2 presents a comparison of the integrated volume fraction of particles over 30 min for these three-time step sizes. When comparing the time step sizes of 0.5 s and 0.1 s, the average percentage difference amounted to 67 %. Between 0.1 s and 0.01 s, this difference was reduced to 11 %. Taking into account computational time and the resources required, a time step size of 0.1 s was chosen for subsequent simulations.

2.3. Validation of CFD model

For the validation of the CFD model, numerical results pertaining to flow and particle distribution were compared with experimental results from Chen et al. [36]. A small chamber, measuring 0.8 m × 0.4 m x 0.4 m (length x width x height), was modeled using the Eulerian-Eulerian approach. In this framework, both the flow field and particle distribution were treated as separate continua. Employing the realizable k-epsilon turbulence model, the x-direction air velocity and normalized particle concentration were measured at various locations and compared with the numerical outcomes.

Measurements of the x-direction velocity and normalized particle concentrations were taken at the center of the y-plane at three different x-positions: x = 0.2 m, x = 0.4 m, and x = 0.6 m. These specific points were chosen to provide a representative cross-sectional profile of the flow and particle distribution within the chamber. The experimental results at these points were juxtaposed with the

Table 1
Grid convergence index for normal classroom arrangement.

	Grid 1 (fine)	Grid 2 (medium, examined)	Grid 3 (coarse)
Number of cells	623,416	417,770	200,967
Air velocity at 0.2 m above occupant's head (m/s)	0.225	0.217	0.215
GCI (%)	0.12	1.59	–

Table 2
Integrated volume fraction of particle during 30 min for three different time step size comparison.

	0.01 s	0.1 s	0.5 s
Occupant 1	9.12×10^{-10}	7.77×10^{-10}	8.09×10^{-10}
Occupant 2	4.28×10^{-12}	4.90×10^{-12}	1.13×10^{-12}
Occupant 3	6.36×10^{-12}	5.81×10^{-12}	1.80×10^{-12}
Occupant 4	5.50×10^{-12}	4.60×10^{-12}	9.56×10^{-12}
Occupant 5	2.93×10^{-12}	2.74×10^{-12}	5.00×10^{-12}
Occupant 6	2.10×10^{-12}	1.83×10^{-12}	2.42×10^{-13}
Occupant 7	4.04×10^{-12}	3.50×10^{-12}	6.13×10^{-13}
Average percent difference (%)	–	11	67

corresponding numerical results obtained from the CFD model.

The comparative analysis indicated that the numerical results closely matched the experimental data across the different locations. For instance, at x = 0.2 m, the numerical and experimental x-direction velocities showed an close alignment, as illustrated in Fig. 3a, with a Mean Squared Error (MSE) of just 0.00012, as shown in Table 3, indicating a high degree of accuracy. Similarly, at x = 0.4 m and x = 0.6 m (Fig. 3b and c), the MSE values were slightly higher at 0.00013 and 0.0001, respectively, but still demonstrated a strong correlation between the numerical and experimental results, as summarized in Table 3.

For normalized particle concentrations, the comparison also revealed a strong agreement between the CFD model and the experimental data. As shown in Fig. 4 and detailed in Table 3, the MSE values for the normalized concentration at x = 0.2 m, x = 0.4 m, and x = 0.6 m were 0.0032, 0.0036, and 0.0013, respectively. Consequently, it is concluded that the Eulerian-Eulerian approach, combined with the realizable k-epsilon turbulence model, offers reasonable accuracy for particle distribution, and can be utilized for the remaining portions of the study.

2.4. Building operation scenarios

To comprehensively evaluate the risk of infection due to airborne transmission in enclosed spaces, various parameters were considered. These include ventilation strategy, air change rate, occupant arrangement, source location, and particle size, as delineated in Table 4. Numerous studies reported the significance of ventilation strategies, which encompass aspects such as airflow and distribution [37,38]. A systematic review by Tsang et al. [39] indicated that most of the existing research focuses on the effects of different ventilation strategies, particularly within hospital settings. In a similar vein, Zhao et al. [40] identified social distancing as one of three dominant factors influencing airborne transmission, along with ventilation systems and environmental conditions such as humidity and temperature. The present study set the minimum air change rate at 1.1 h^{-1} according to ASHRAE Standard 62.1 and varied it up to 10 times the minimum rate [41].

In addition to these factors, particle size is also crucial. A study by Miller et al. [42] highlighted its importance in different modes of COVID-19 transmission, including direct contact, large droplets, and aerosols. Larger droplets tend to fall quickly and travel less than six feet, whereas smaller, aerosolized particles can remain in the air for extended periods, as noted by the EPA. Our study specifically examined two particle sizes: 1 μm and 10 μm, representing particulate matter 2.5 and 10, respectively. The 1 μm particle size was chosen because particles smaller than 1 μm is a significant component of $PM_{2.5}$ in a classroom, contributing to the overall particulate

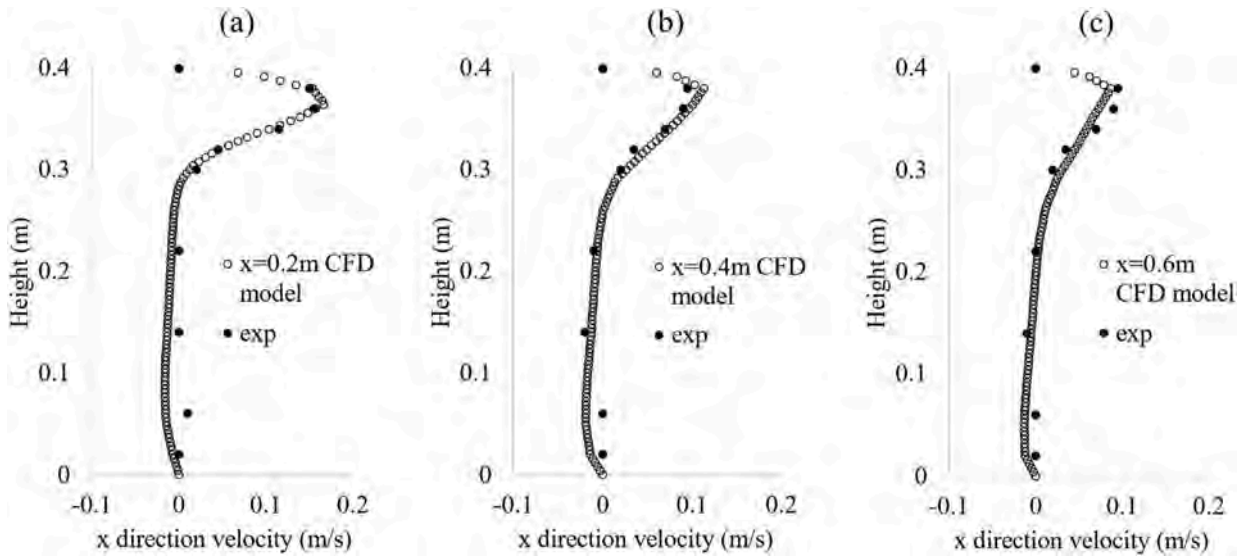

Fig. 3. x-direction velocity validation results at different locations, a) at x = 0.2 m, b) at x = 0.4 m, and c) at x = 0.6 m.

Table 3
Mean Squared Errors at different locations for each parameter.

Parameters\Locations	x = 0.2 m	x = 0.4 m	x = 0.6 m
X-direction air velocity	0.00012	0.00013	0.0001
Normalized concentration	0.0032	0.0036	0.0013

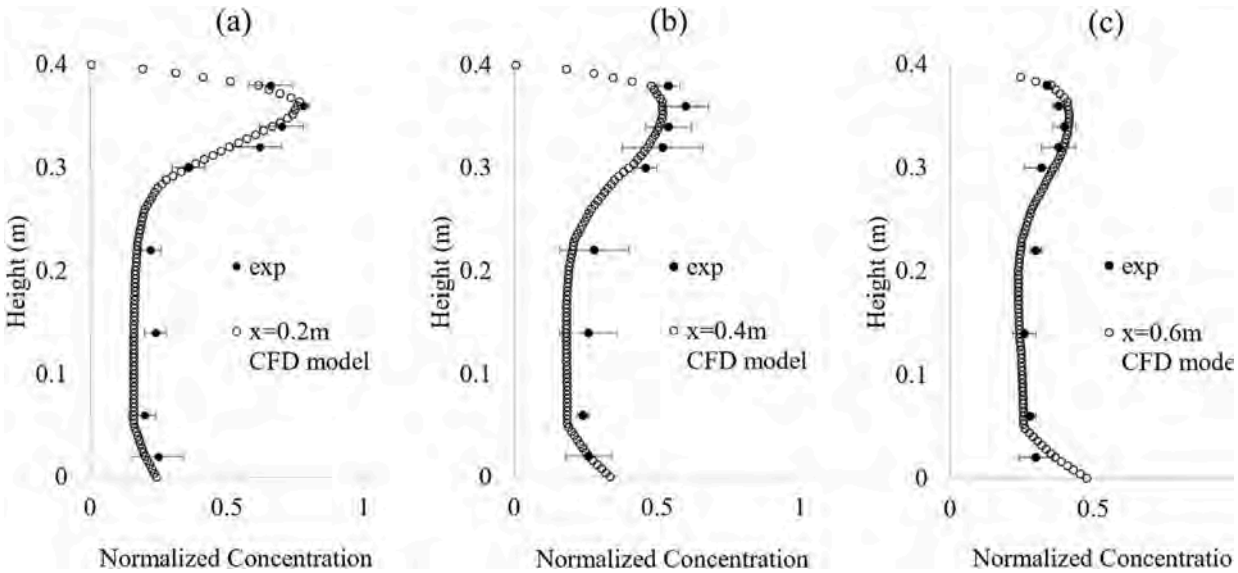

Fig. 4. Normalized concentration validation results at different locations, a) at x = 0.2 m, b) at x = 0.4 m, and c) at x = 0.6 m.

Table 4
Varied parameters.

Input variables	Variations			
Ventilation strategy	Mixing Ventilation		Displacement Ventilation	
Air change rate (h^{-1})	1.1	4.6	8.0	11.1
Occupant arrangement*	Normal classroom		Debate-style	
Source location	Occupant 1 to 7			
Particle size (μm)	1		10	

concentration [43]. Furthermore, particles emitted during typical speech often average around 1 μm in diameter, making this size particularly important for understanding airborne transmission dynamics [44–46].

To comprehensively evaluate the impact of multiple factors on the risk of airborne transmission, we designed a detailed and systematic study that involved varying several critical parameters. Specifically, we examined two distinct ventilation strategies, each with unique implications for air circulation and contaminant dispersion in occupied spaces. We also considered four different air change rates, ranging from minimal to highly efficient ventilation, to capture the effects of varying levels of air supply on the dilution and removal of airborne pathogens. Additionally, we explored two different occupant arrangements, simulating scenarios where individuals are either clustered together or more evenly dispersed throughout the space. This allowed us to understand how proximity and spatial distribution of occupants might affect the transmission dynamics.

Furthermore, we investigated seven different source locations, strategically placing the emission points of airborne pathogens at various spots within the environment to observe how the position of an infector influences the spread of contaminants. To add another layer of complexity, we included two particle sizes in our analysis, recognizing that different-sized particles have distinct behaviors in terms of settling, dispersion, and inhalation risk.

By combining these variations, our study encompassed a total of 224 CFD simulation cases. This extensive dataset provided a foundation for thoroughly exploring the intricate interplay of ventilation, occupant placement, source location, and particle size in shaping the risk of airborne transmission in indoor environments. Through this comprehensive approach, we aimed to uncover nuanced insights that could inform more effective strategies for mitigating the spread of airborne pathogens.

2.5. Particle emitting dynamics and inhalation analysis

2.5.1. Particle transport modeling

In the present study, particles emitted from an infector are assumed to be droplet nuclei. These are small particles with diameters ranging from 1 to 10 μm that represent the dried or evaporated residue of droplets [47]. Netz. [48] found that 10-μm diameter droplets evaporate completely within 0.1 s at a relative humidity of 50 %, with smaller particles evaporating even more quickly. The study also indicated that the effects of droplet evaporation on flow buoyancy and mass density change can be neglected. Yang et al. [49] further noted that the amount of virus in the droplet is unaffected by the evaporation process. Therefore, this study assumes that the particles emitted from an infector are fully evaporated droplet nuclei.

2.5.2. Particle volume fraction emission

The volume fraction of the particle emitted is another factor to be considered, as it varies with different particle sizes that are emitted from the infector. Yang et al. [49] indicated that during typical speech and loud speech, the volume fraction of droplets at the mouth ranged from 6×10^{-9} to 1×10^{-8} respectively. For this study, the volume fraction of 1×10^{-8} was used for particles with diameter of 10 μm. In addition, volume fraction of 1 μm particles was used as 1×10^{-11} to equalize the number of particles emitted for both sizes, ensuring a precise understanding of how particle size influences the dispersion of infectious particles within an indoor space. This approach allowed for a direct comparison of the dispersion patterns of differently sized particles under identical emission conditions. Additionally, because the different particle sizes were emitted in separate scenarios, this method did not complicate the calculation of infection risk using the dilution-based method, which relies on concentrations at two sampling volumes: the source of infection (the infector) and the point of potential exposure (the susceptible).

2.5.3. Sampling volume

To fully replicate the inhalation of particles by individual occupants, a 1-L cubic volume was specifically positioned in close proximity to the nasal region (see Fig. S 2 in Supplementary materials). This cube served as a representation of the sampling volume, which is designed to measure the volume fraction of particles that are emitted by an infector, reflecting the typical tidal volume of a normally active individual [50].

2.5.4. Source emission and inhalability

In the framework of the Eulerian approach, where both air and particles are treated as a continuum phase, this study focused on the volume fraction of infectious particles emitted. In each simulation case with duration of 30 min, a single occupant, designated as the "infector," emitted particles while in the talking mode. This mode was characterized by a continuous airflow rate of 13.5 L per minute, as defined by Gupta et al. [51]. The remaining six occupants, vulnerable to these particles, are referred to as "susceptible individuals."

To enhance the accuracy of assessing inhalability for these susceptible individuals, breathing mode was also taken into account, as characterized by the same study by Gupta et al. [51]. This approach aimed to realistically emulate a scenario where one occupant is talking while the rest are listening. However, it's worth noting that some research, such as the study by Rim and Novoselac. [52], suggested that the breathing mode has a negligible impact on particle inhalability. To address this, a straightforward comparative test simulation was carried out, comparing the 20-min cumulative volume fraction of particles with and without the breathing simulation at air change rate of 11.1 h^{-1}. While the breathing mode created fluctuations within sampling volumes, there was no significant difference in 20-min integrated exposure as shown in Table 5 with 4.93 % difference at maximum. The same comparative test at an air change rate of 1.1 h^{-1} was also conducted, resulting in a difference of less than 2.49 % (see Fig. S 3 and Table S 1 in Supplementary Materials). Therefore, the breathing mode for susceptible individuals was not considered in the current study.

2.6. Dilution-based evaluation of airborne infection risk

The evaluation of infection risk in this study was conducted using the dilution-based method proposed by Zhang et al. [53]. This approach leverages the concept of dilution of airborne contaminants and extends the Wells-Riley equation to accommodate a spatially non-uniform distribution of aerosols. The time-dependent infection risk based on the Wells-Riley equation is defined as Eq. (2) and the dilution-based method is defined as Eq. (3)

$$P_I = \frac{C}{S} = 1 - exp\left(-\frac{Iqpt}{Q}\right) \tag{2}$$

Table 5
20 min - Integrated volume fraction of particle under normal classroom arrangement with mixing ventilation at air change rate of 11.1 h^{-1} comparing with and without breathing mode.

	Without breathing mode (−)	With breathing mode (−)	Percent difference (%)
Occupant 1	2.84E-10	2.70E-10	4.93
Occupant 2	9.52E-13	9.39E-13	1.37
Occupant 3	1.19E-12	1.14E-12	4.20
Occupant 4	6.48E-13	6.40E-13	1.23
Occupant 5	4.98E-13	5.16E-13	−3.61
Occupant 6	1.82E-13	1.85E-13	−1.65
Occupant 7	3.46E-13	3.46E-13	0.00

Where P_I is the probability of infection, C is the number of infection cases, S is the number of susceptible, I is the number of infector individuals in the space, q is the quanta generation rate in $\frac{\#}{hr}$, p is the breathing rate of individuals in $\frac{m3}{hr}$, t is the exposure time in hour, and Q is the air change rate in m^3/hr.

$$P_D = 1 - e^{-N_{quantum}} \tag{3}$$

Where

$$N_{quantum} = \int_0^T p_{susceptible} \times C_{quantum}(t)dt \tag{4}$$

$$C_{quantum} = \frac{\varnothing \times q}{p_{infector} \times D} \tag{5}$$

$$D = \frac{C_{source}}{C_{target}} \tag{6}$$

P_D is the airborne infection risk at the target position during time t exposure, $N_{quantum}$ is the inhaled quanta by a susceptible individual, $C_{quantum}$ is the quantum concentration in $\frac{quanta}{m^3}$, $\varnothing$ is the penetration ratio of personal protective equipment such as mask, q is the quantum generation rate in $\frac{quanta}{s}$, $p_{infector}$ is the breathing rate of the infector, D is dilution ratio, C_{source} is airborne contaminant concentration at the source, and C_{target} is the airborne contaminant concentration at the target.

Within the framework of this study, the "source" refers to the sampling volume of the infecting individual, whereas the "target" is construed as the sampling volumes for susceptible individuals. A recent study revealed that the quantum generation rates for three SARS-CoV-2 variants—Alpha, Delta, and Omicron—are back-calculated to be 89-165 h^{-1}, 312–935 h^{-1}, and 725-2345 h^{-1}, respectively [54]. For the purposes of the present study, the constant quantum generation rate for the Alpha variant was assumed to be at the lower bound of 89 h^{-1}.

2.7. Machine learning based surrogate model

The surrogate model is fundamentally designed to replicate the behavior of the CFD simulations, enabling faster predictions of inhalable particle concentrations at a lower computational cost. The model architecture consisted of an input layer representing the various physical parameters (e.g., air change rates, room geometry, and occupant locations), multiple hidden layers where nonlinear relationships are captured through interconnected neurons, and an output layer that provides the predicted particle concentrations. Key parameters of the model include the learning rate, which controls the step size during optimization and influences convergence efficiency; the batch size, which determines the number of training samples processed simultaneously and affects both computational efficiency and update stability; and the number of neurons and layers, which dictate the model's capacity to capture complex patterns but require careful tuning to avoid overfitting. A normalization technique was applied to the input data to standardize the range of values, improving the training stability and convergence rate by ensuring that all features contribute equally to the learning process. The following sections will provide a detailed explanation of the supervised learning approach used, the data preprocessing steps, and the hyperparameter optimization process.

To predict the 30-min period of inhalable particle number in the 1-L sampling volume located in front of each occupant's nose, this study employed supervised learning. The supervised learning is widely used approach when the inputs and outputs are both available [55]. By letting both input and output know to be trained, the primary objective is to discern a relationship between the inputs and their corresponding outputs, enabling predictions for new input data.

The CFD data, specifically the volume fraction of the particle phase for each occupant over a 30-min interval, were exported into. xlsx files. Each of the 224 simulation cases represented a different combination of input parameters, such as ventilation strategies and air change rates. The CFD output data for each simulation was combined with its corresponding input parameters into a single file, resulting in a total of 224 files.

Using Python, these.xlsx files were loaded into Pandas dataframes, which organized the data into a tabular format. This structured data was then ready for preprocessing, including handling missing values, normalizing features, and ensuring consistency across all datasets. After preprocessing, the data was split into training, validation, and testing sets. This prepared the data for subsequent training of the AI model within the TensorFlow and scikit-learn environments, where the focus was on predicting the inhalable particle number concentration based on the input parameters from the CFD simulations.

In the current study, two separate models were created and trained: the susceptible model and the infector model. Each model underwent a creation phase where its architecture was defined, followed by a training phase where the model learned from the training data. After training, each model's performance was first evaluated using the validation set to fine-tune the model parameters and avoid overfitting. During training, early stopping was used to monitor the validation loss and stop training if no improvement was observed, thus fine-tuning the model effectively.

Once fine-tuning was complete, the models were evaluated using the test set to ensure their accuracy and reliability. Following this evaluation, predictions were made using the test data. These predictions were then reshaped and underwent inverse transformations to revert them back to their original scale, making the results interpretable and comparable to the original data.

2.7.1. Input data structure

This study introduced a range of initial parameters, as detailed in Table 4, to predict the 30-min inhalable particle concentration for each occupant. These parameters served as the input, while the particle concentration was the output. To enhance the model's generality, the arrangement parameter was replaced with the x and y coordinates and the facing direction of each occupant.

Each case in this study comprised seven distinct time series datasets, representing the number of inhalable particles over a 30-min period for each of the seven occupants. Due to the nature of particle emission by an infector, the particle count in front of the infector was significantly higher, often by orders of magnitude, compared to that of a susceptible individual. This disparity posed a challenge for the model, which was trained using mean absolute error, as it struggled to learn effectively from the smaller numbers associated with susceptible occupants. To address this, the study distinctively isolates the infector data from the susceptible data. This isolated susceptible data was then zero-padded, ensuring that both datasets were subsequently trained separately.

2.7.2. Model selection

In this study, we explored the performance of two different models, Artificial Neural Networks (ANN) and Long Short-Term Memory (LSTM) networks, for the given research problem. ANNs, which are composed of interconnected artificial neurons, have the advantage of identifying nonlinear relationships [56]. On the other hand, LSTMs, a type of Recurrent Neural Networks (RNNs), are designed to handle contextual information with dynamic inputs. One of the key features of LSTMs is their ability to address the vanishing gradient problem, a common issue in traditional RNNs that leads to the loss of earlier information over time. This makes LSTMs particularly suitable for time series prediction tasks [57,58].

2.7.3. Hyperparameter optimization

To identify the optimal hyperparameters for both the infector and susceptible models, a comprehensive evaluation was conducted using the determination coefficient (R^2), root mean square error (RMSE), and coefficient of variation of the root mean square error (CVRMSE) metrics. Various combinations of hyperparameters were tested for both ANN and LSTM models. The hyperparameters and their variations used for optimization are summarized in Table 6. To prevent overfitting, the early stopping technique was employed during training.

Table 6 outlines the varied hyperparameters and their corresponding variations, including the number of hidden layers (1, 2, 3, and 4), number of neurons (ranging from 2 to 129 in intervals of 2), batch size (4, 8, 16, and 32), learning rate (0.01, 0.001, and 0.0001), training set ratio (0.6 and 0.7), and normalization techniques (StandardScaler and MinMaxScaler).

Both models utilized ANN architectures and StandardScaler normalization technique with the infector model employing 2 hidden layers, 38 neurons, a batch size of 32, a learning rate of 0.001, and a training set ratio of 0.6. Similarly, the susceptible model also used 2 hidden layers but with 76 neurons, a batch size of 8, a learning rate of 0.0001, and a training set ratio of 0.6.

This rigorous hyperparameter tuning process ensured the models were optimized for accurate predictions while mitigating the risk of overfitting, thereby enhancing their reliability and performance.

3. Results

3.1. Effects of ventilation mode

Fig. 5 illustrates the varying infection risks associated with different air change rates under both mixing and displacement ventilation systems. Notably, the risk of infection for susceptible individuals was found to be between 49 % and 77 % lower under displacement ventilation as compared to mixing ventilation. These findings are consistent with previous research, which has also demonstrated the performance of displacement ventilation in mitigating the risk of airborne transmission with both numerical and experiment results [59–61].

The Wells-Riley equation was also applied to estimate the infection risk under both mixing and displacement ventilation systems at various air change rates. While the equation yielded reasonable approximations under specific conditions—namely, at low air change rates for mixing ventilation and at high air change rates for displacement ventilation—its limitations were evident. In particular, the upper quartile of infection risk was found to be up to 6.45 times higher than that predicted by the traditional Wells-Riley model with significant range. This highlights the model's limitations in capturing the complexities of spatial distribution in infection risk.

3.2. Effects of airflow pattern

As illustrated in the previous section, occupants under displacement ventilation experienced lower infection risk compared to those under mixing ventilation across different air change rates. To further examine the effects of airflow patterns under different ventilation

Table 6
Varied hyperparameters with their variations for optimization.

Hyperparameters	Variations			
Number of hidden layers	1	2	3	4
Number of neurons	2 -129 (Interval of 2)			
Batch size	4	8	16	32
Learning rate	0.01	0.001	0.001	
Training set ratio	0.6		0.7	
Normalization technique	StandardScaler		MinMaxScaler	

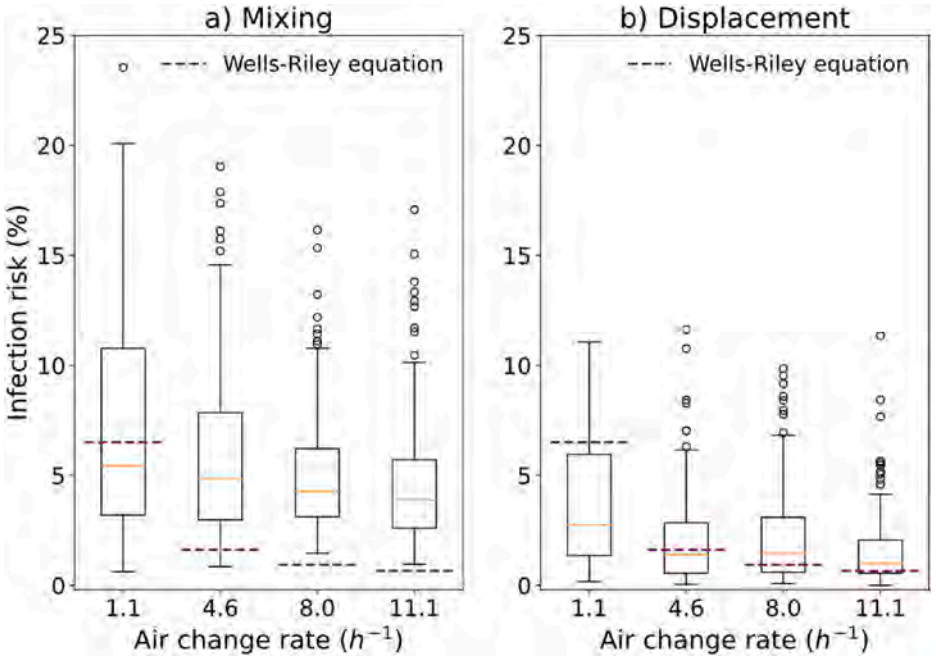

Fig. 5. Comparison of infection risk (%) under different air change rates (h^{-1}) for both Mixing Ventilation and Displacement Ventilation. The dotted blue lines represent the Wells-Riley equation results. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

strategies and air change rates, contour and streamline plots of particle volume fraction at 30 min are presented in Figs. 6 and 7. At an air change rate of 1.1 h^{-1}, particle dispersion is primarily driven by buoyancy forces, causing particles to rise to the upper level. However, under mixing ventilation, as shown in Fig. 7a, the elevated particles are redistributed by airflow from the ceiling diffuser, resulting in a higher infection risk compared to displacement ventilation, where dispersed particles remain at the upper level and are effectively removed, as illustrated in the streamline. At a higher air change rate of 11.1 h^{-1}, the particle dispersion pattern changes significantly. Under mixing ventilation, emitted particles are pushed to relatively lower levels, resulting in widespread particle dispersion throughout the space. In contrast, under displacement ventilation, even at the higher air change rate, buoyancy forces remain dominant, maintaining vertical stratification and effectively removing particles.

3.3. Effects of air change rate

As shown in Fig. 5, the infection risk decreases as the air change rate increases. However, many studies indicate that the air change rate is neither proportional to the reduction in infection risk nor does it necessarily guarantee a lower infection risk, as expressed in the Wells-Riley model [62–64]. Fig. 8a presents the individual infection risk at two air change rates of 1.1 h^{-1} and 11.1 h^{-1} for a normal classroom arrangement under mixing ventilation. The occupant indices are indicated in Fig. S 4 in Supplementary materials. With an air change rate of 1.1 h^{-1}, occupants 2, 3, and 4, who are in the front row, have a higher infection risk compared to occupants 5, 6, and 7 in the back row. However, as the air change rate increases to 11.1 h^{-1}, while infection risks for occupants in the front row decrease significantly, the infection risk for occupants in the back row increases. Specifically, the infection risk increased by 88 % and 166 % for

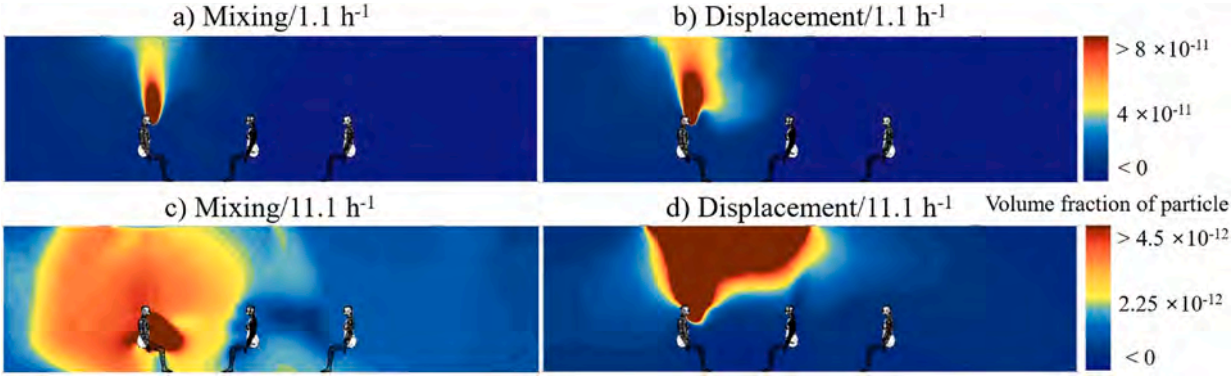

Fig. 6. Contour plots of the volume fraction of particles under mixing and displacement ventilation at air change rates of 1.1 h^{-1} and 11.1 h^{-1}.

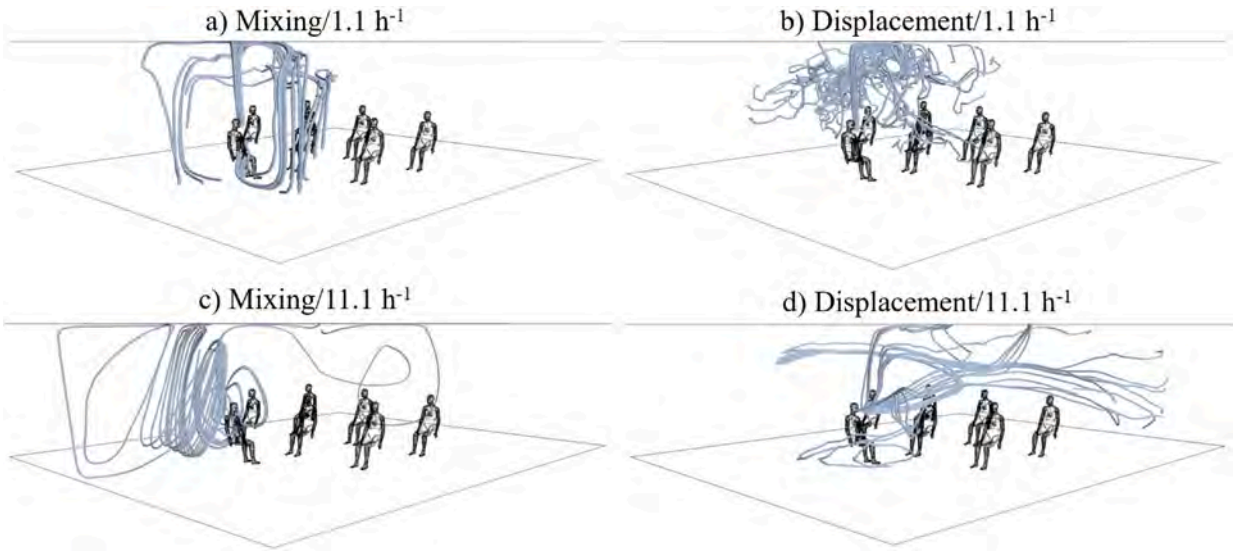

Fig. 7. Particle trajectories emitted by an infector (occupant 1 in this case) under mixing and displacement ventilation at air change rates of 1.1 h^{-1} and 11.1 h^{-1}.

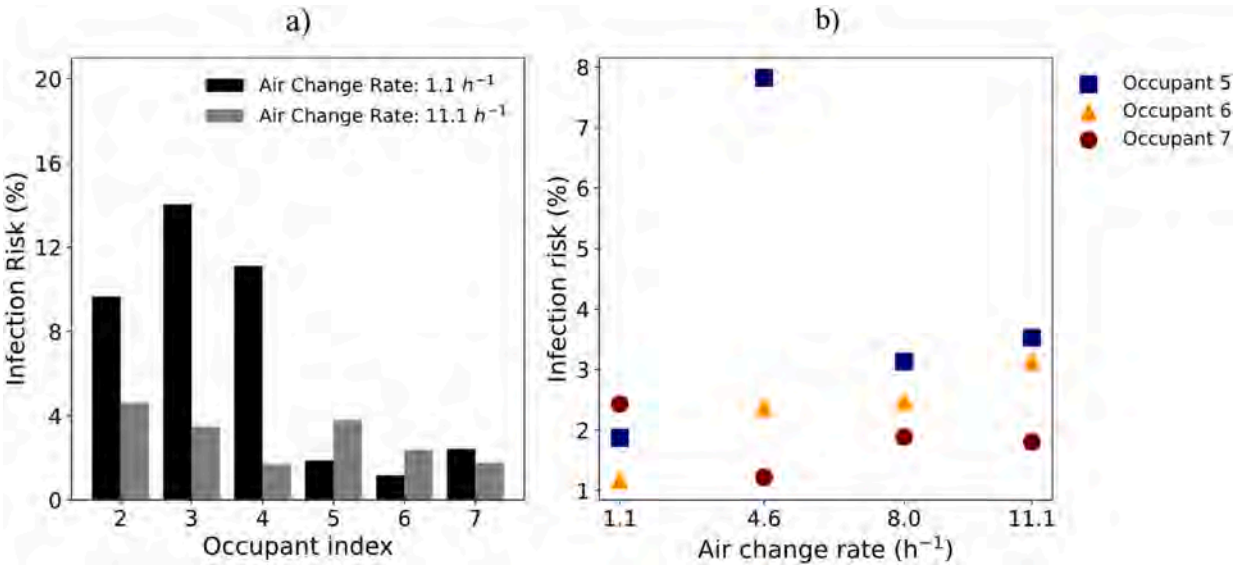

Fig. 8. Infection risk with occupant 1 as the infector under mixing ventilation for a normal classroom arrangement. (a) Individual infection risk at air change rates of 1.1 h-1 and 11.1 h-1. (b) Infection risk of occupants 5, 6, and 7 at different air change rates.

occupants 5 and 6, respectively. Fig. 8b illustrates the infection risk for occupants in the back rows across four different air change rates. As the air change rate increases, occupants could be at greater risk of exposure to infectious particles, possibly due to the farther travel distance of the particles caused by increased airflow. This finding contradicts the widely accepted view that higher ventilation rates consistently lead to reduced exposure. However, it aligns with specific studies. Notably, Bolashikov et al. [65,66] reported that while elevated air change rate (12 h^{-1}) generally reduced average exposure, under certain conditions, it may increase exposure compared to that of lower ventilation rates (3 h^{-1} and 6 h^{-1}). Moreover, Pantelic and Tham [67] reported that an increase in supply airflow rate could lead to an increase in exposure in some cases. This paradoxical effect was attributed to higher airflow rates causing complex airflow patterns and further dispersion of infectious particles.

3.4. Effects of occupant arrangement

Fig. 9 presents the percentage difference in average infection risk between the two arrangements, with the normal classroom arrangement serving as the reference. For mixing ventilation, the percentage difference decreases as the air change rate increases, from 20 % greater infection risk at the ASHRAE's minimum ventilation rate to 1 % at 11.1 air change rate compared to that in a normal classroom arrangement. On the other hand, under displacement ventilation, the percentage difference is about 20 % lower at 4.6 h^{-1}

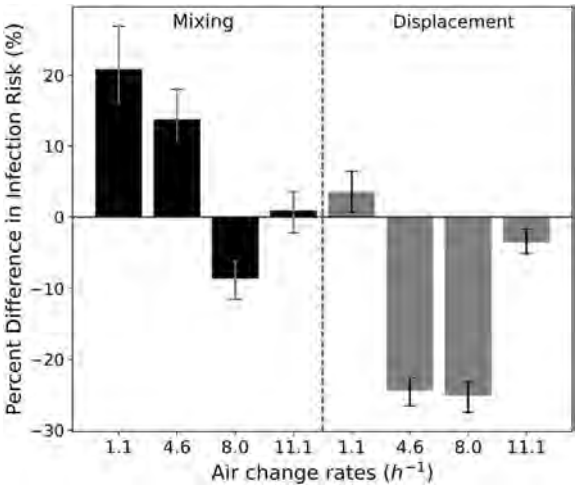

Fig. 9. Percent difference in average infection risk between the two arrangements, with normal classroom arrangement as the reference, for various ventilation strategies and air change rates with error bars representing one standard deviation.

and 8 h^{-1} compared to that in a normal classroom arrangement with no specific trend with varying air change rate. This suggests that with a higher ventilation rate under mixing ventilation, the effect of occupant arrangement becomes smaller. This seems reasonable, as mixing at higher air change rates would tend to provide a more uniform condition. In contrast, susceptible occupants under the debate-style arrangement tend to experience similar or lower infection risk compared to those under the normal classroom arrangement when using displacement ventilation. Although further research is necessary to pinpoint the reason, it's shown that the influence of occupant arrangement on infection risk is significant, particularly with displacement ventilation.

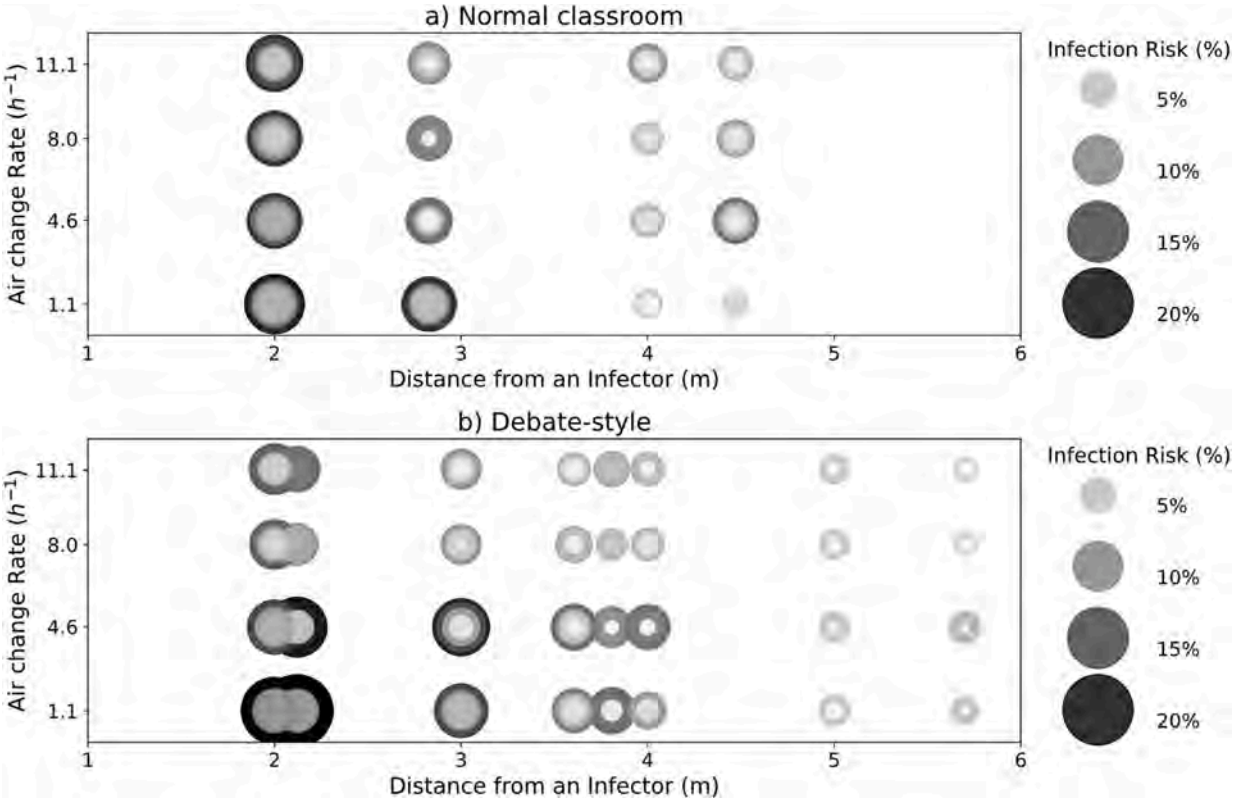

Fig. 10. Infection risk by distance from an infector at different arrangements and air change rates per hour for a) normal classroom and b) debate-style arrangements.

3.5. Effects of the distance from an infector

Fig. 10 illustrates infection risk as a function of distance from an infector, across varying air change rates. The plot employs bubble size and color intensity to represent infection risk, with larger bubbles and darker colors indicating higher risks. The data generally show a trend where infection risk decreases with increasing distance from the infector, applicable to both normal classroom and debate-style arrangements. However, this trend isn't always consistent; in some cases, those situated farther from the infector experience higher infection risks compared to those closer. For example, Fig. 11 illustrates the infection risk when occupant 6 is the infector under mixing ventilation in a debate-style arrangement at an air change rate of 4.6 h^{-1}. In this scenario, the infection risk for occupant 4 is approximately 280 % and 87 % higher than that for occupants 2 and 3, respectively, despite occupant 4 being at the same distance from the infector as occupant 2 and farther away than occupant 3. Additionally, the infection risks for occupants 1 and 5 are similar, even though occupant 1 is about 2.5 m farther from the infector than occupant 5. It is important to note that although the diffusers are symmetrical, the particle dispersion within the space, as shown in Fig. 11c, is not. While the configuration could impact particle dispersion, the symmetry of the supply diffusers and exhausts does not necessarily result in symmetric particle dispersion. This asymmetry suggests that particle dispersion is more influenced by other factors affecting local airflow patterns, such as heat sources, air stagnation, and source location. The result regarding the impact of airflow patterns on particle dispersion aligns with a previous study by Yan et al. [68] which investigated a conference room with an occupant arrangement akin to the current study's debate-style arrangement, under various ventilation strategies. They reported that, while it differs by supply and exhaust diffuser locations, mixing ventilation could lead to extensive droplet dispersion across the room, resulting in a higher infection risk for occupants farther away, compared to those closer. Therefore, relying solely on distance as a preventive measure may not be sufficient in certain situations, and other factors, such as airflow patterns, should also be considered.

3.6. Effects of particle size

Fig. 12 presents boxplots of infection risk distribution for two distinct particle sizes, showing that larger diameter particles lead to a lower infection risk. The trajectory and duration of airborne particles in enclosed spaces are significantly influenced by their size. Particles with a diameter of 10 μm settle much faster due to gravitational forces compared to those that are 1 μm in size. This accelerated descent effectively reduces their residence time in the ambient environment, thereby decreasing the likelihood of inhalation and the associated infection risk. At the air change rate of 1.1 h^{-1}, the percent difference between the medium of the two particle sizes is 18 %, but as the air change rate gets larger to 11.1 h^{-1}, then the percent difference significantly increases to 51 %, indicating an amplified effects of smaller particles remaining suspended while larger particles settle. This underscores the importance of particle settling due to gravity which affects the airborne infection risk.

3.7. Machine learning based surrogate model

3.7.1. Performance of the surrogate model on the test dataset

Fig. 13 compares the inhalable particle number over a 30-min period between the test dataset generated by CFD simulation, which was not used for training or validation, and its prediction by the surrogate model for both the infector and each occupant as susceptible individuals. Each data point represents the inhalable particle number concentration at a specific time step, and each line composed of dots, which are colored differently for each case, represents a case for 30-min, with the diagonal line indicating a perfect prediction.

Table 7 presents the evaluation metrics, including the Coefficient of Variation of the Root Mean Square Error (CVRMSE) values for the infector and susceptible models. The surrogate model for the infector exhibited relatively better performance, achieving R^2 and CVRMSE values of 0.79 and 34 %, respectively. This improved performance can be attributed to the infector's constant speaking and particle emission, which might have simplified the model's learning and prediction tasks. In contrast, the susceptible models demonstrated the ability to capture some nonlinear relationships, with R^2 values ranging from 0.31 to 0.65 and CVRMSE values from

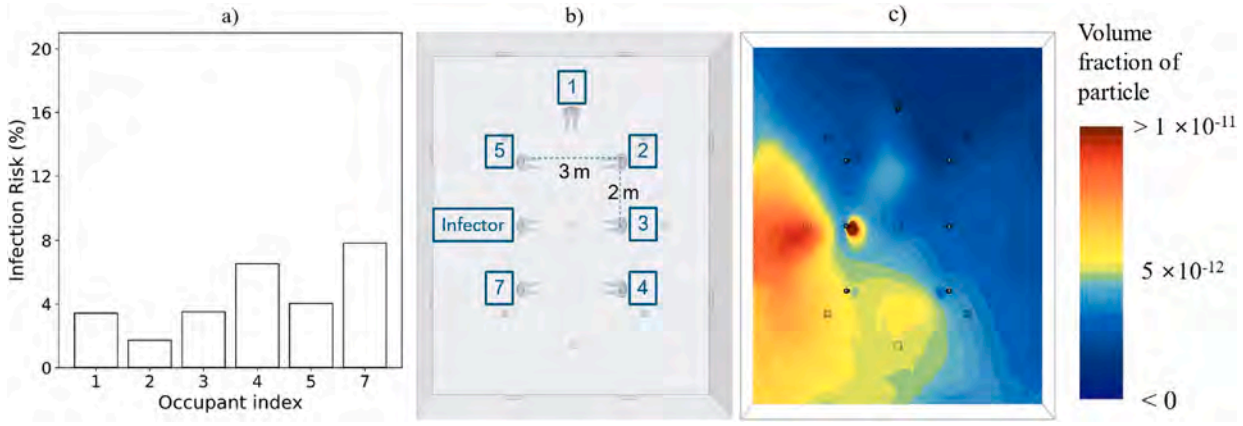

Fig. 11. a) Individual infection risk when occupant 6 is the infector under mixing ventilation in a debate style-arrangement at an air change rate of 4.6 h^{-1}, b) debate-style arrangement with the infector and occupant index annotated, and c) horizontal volume fraction of particle contour plot at mouth level (z = 1.2 m) at 30 min.

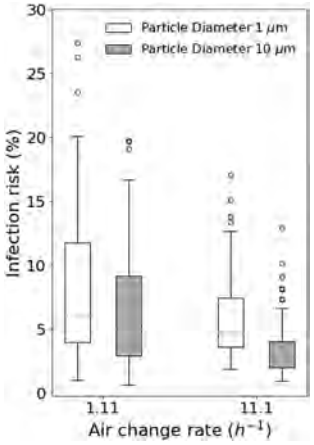

Fig. 12. Infection risk box plot for 1.11 h^{-1} and 11.1 h^{-1} between the two particle diameters.

100 % to 181 %.

While R^2 values above 0.8 are often sought for strong predictive accuracy, values below this threshold can still offer valuable insights, particularly in complex environmental modeling. For instance, in a study on suspended sediment load prediction, models with R^2 values between 0.35 and 0.74 effectively captured crucial trends such as the influence of rainfall, discharge, and stage on sediment transport [69]. Similarly, in PM2.5 concentration predictions, R^2 values as low as 0.22 to 0.49 provided meaningful insights into spatial variability and key predictors like aerosol optical depth and elevation [70]. These findings highlight that even with moderate R^2 values, the models can still uncover important patterns, providing valuable guidance for further refinement and practical application.

The R^2 values observed for the surrogate model, particularly for susceptible individuals, could be influenced by several key factors. One significant factor is the inherent complexity of the problem, specifically the mechanisms of particle dispersion within the environment. While the current dataset is extensive, the intricate and dynamic nature of particle behavior suggests that more data may be necessary to adequately capture all relevant scenarios. This potential limitation might contribute to the lower R^2 values (0.31–0.65) observed for susceptible models, as the model may struggle to generalize effectively. To address this, increasing the volume of data through techniques such as data augmentation could be beneficial. By expanding the dataset with additional simulated scenarios, the model's ability to generalize and predict more accurately could be enhanced.

Another factor that could contribute to the observed R^2 values is the limitation of using primarily static input parameters in training the AI model. While the CFD simulations provide detailed data on airborne particle dispersion, including the dynamic nature of inhalable concentration as an output, the current model does not incorporate dynamic input factors that might be beneficial in predicting infection risk. For instance, CO_2 concentration distribution over time could serve as an indirect indicator of infection risk, as CO_2 is exhaled along with aerosols that may carry SARS-CoV-2 from COVID-19-infected individuals, making it a potential proxy for indoor SARS-CoV-2 concentrations [71–73]. Incorporating such time-dependent variables into the AI model's training process could improve its ability to capture non-linear relationships and temporal variations in particle behavior, potentially enhancing predictive accuracy.

3.7.2. AI model prediction pattern

Fig. 14 displays a scatter plot of the AI model's predictions regarding susceptible occupants for each case. For each case, average and median R^2 values of the six susceptible occupants are computed, illustrating the performance trend. The figure indicates that for the normal classroom arrangement, the susceptible AI model predictions' median and average R^2 values clustered within the −20 range, showing a consistent performance without notable outliers. This clustering suggests that the model is adept at generalizing for such a normal classroom, though the performance is not exemplary. In contrast, the debate-style layout poses additional complexities, especially when considering displacement ventilation with average R^2 drops to −120. This indicates that the AI model struggles to predict outcomes accurately due to the complex airflow patterns and varying exposure levels in this setup. Specifically, as shown in Fig. 14c and d, the model failed to effectively learn the airflow patterns under displacement ventilation, leading to lower prediction accuracy and more dispersed R^2 values. Nonetheless, the model's predictive capabilities for both arrangements improve with mixing ventilation at increased air change rates, hinting at a better grasp of scenarios with more homogeneous air distribution. While these observed patterns could be influenced by various factors, including the composition of the training set or adjustments in model hyperparameters, it's clear that occupant arrangement emerges as a significant variable. The differing levels of prediction accuracy between the two arrangements underscore the need to consider occupant configuration in future enhancements of the model's design and training.

3.7.3. Time-varying inhalable particle number concentration comparison: Best and worst cases

Fig. 15 shows the time-varying comparison of inhalable particle number concentration over 30 min for each occupant in a specific case. This case is the best-predicted among the test data, as indicated in the far upper right corner of Fig. 14a. It was conducted under mixing ventilation at 8 air change rate in a normal classroom arrangement. The model's predictions closely aligned with the CFD

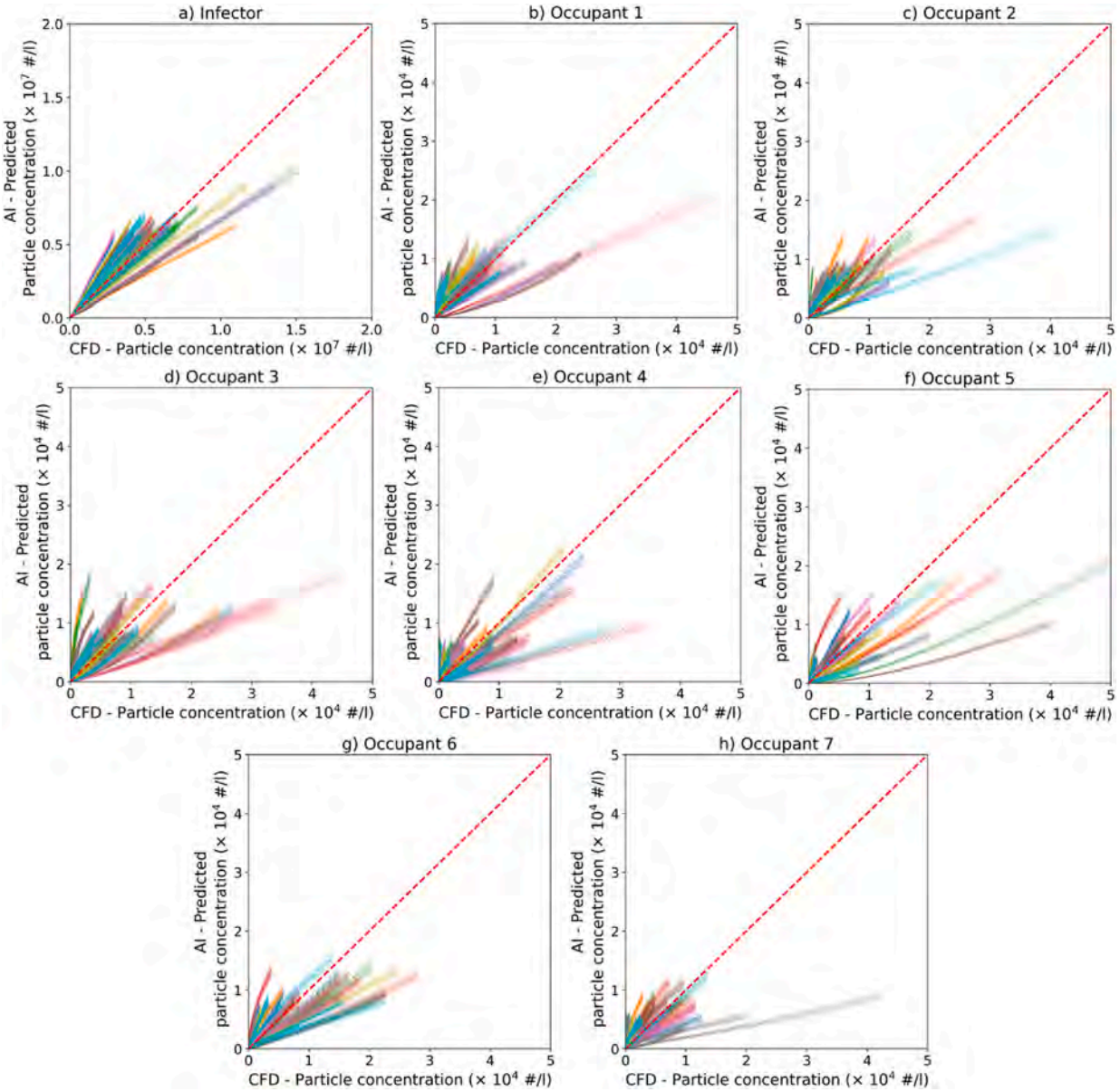

Fig. 13. Scatter plots of inhalable particle number comparing CFD data with corresponding AI model predictions for a) the infector model, and b) to h) the susceptible model for occupant 1 to 7.

Table 7
Performance evaluation for the susceptible and infector model.

Occupant Index	R^2	RMSE (Particle #)	CVRMSE (%)
Susceptible Occupant 1	0.37	3821	155
Susceptible Occupant 2	0.31	3441	162
Susceptible Occupant 3	0.44	4330	127
Susceptible Occupant 4	0.65	2746	108
Susceptible Occupant 5	0.62	3788	123
Susceptible Occupant 6	0.51	3319	100
Susceptible Occupant 7	0.37	3005	181
Average	0.47	3493	134
Infector	0.79	1038927	34

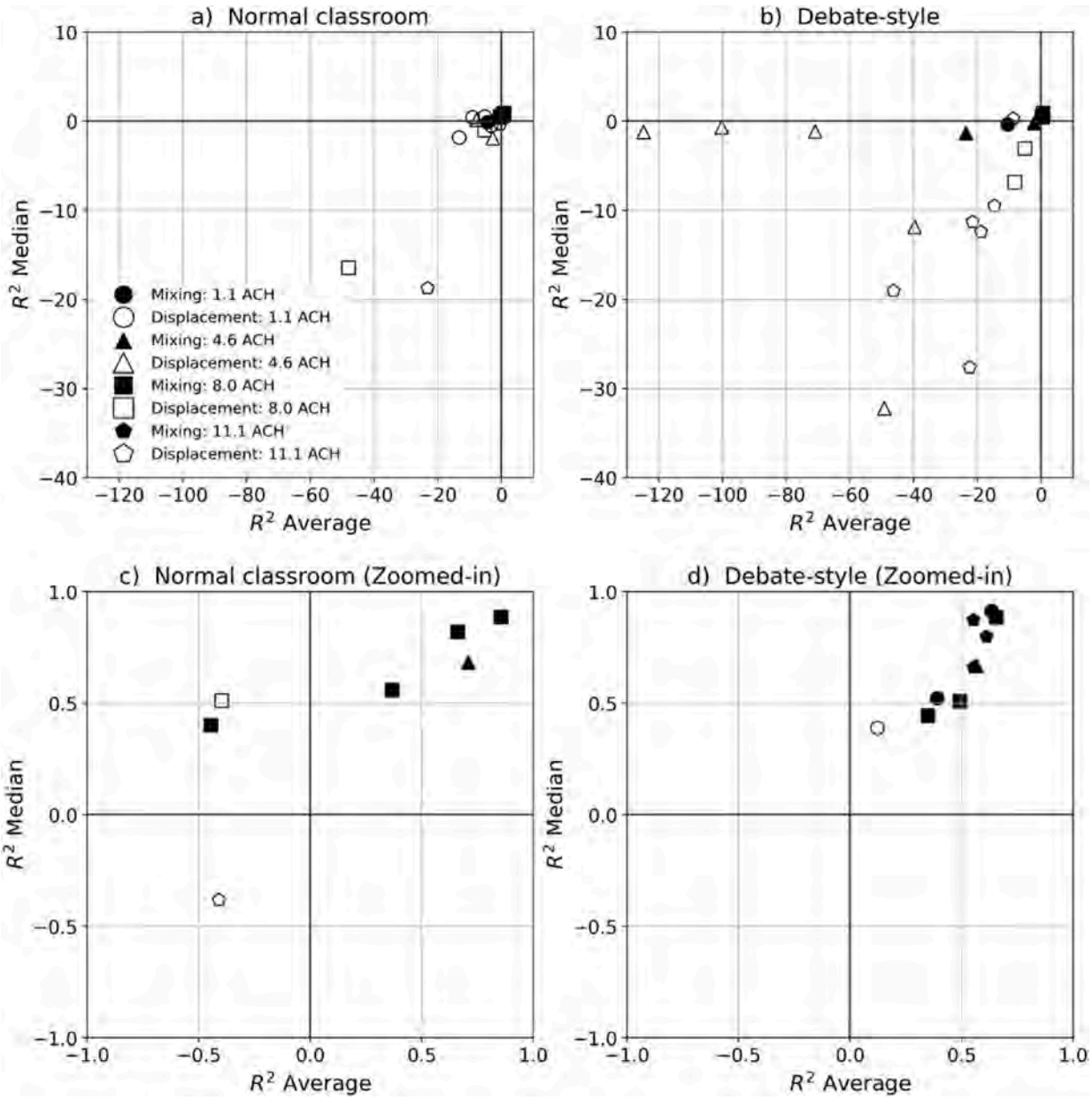

Fig. 14. Scatter plots demonstrating the relationship between average and median R^2 values from the AI model's predictions across testing data cases in different room arrangements. (a) Normal classroom arrangement, (b) Debate-style setup; (c) and (d) are the respective zoomed-in views.

simulation values, resulting in a maximum percent difference of 29 % in the Area Under the Curve (AUC). This low percent difference in the AUC provides valuable insight into the model's performance and its ability to estimate infection risk accurately. This suggests that the model effectively captures the underlying patterns of emitted particle dispersion in this configuration.

Fig. 16, on the other hand, depicts the worst prediction case, shown in the far lower portion of Fig. 14b his case involved displacement ventilation at a higher rate. While the surrogate model accurately predicted the inhalable particle number concentration for occupant 2, who was right next to the infector, occupant 3, the AI surrogate model's predictions for other occupants deviated significantly from the true CFD values with percent difference in AUC up to 1147 %. This deviation is reflected in the higher percent difference in the AUC for each occupant, indicating increased variability in the model's predictions.

Overall, the comparative analysis of the best and worst prediction cases highlights the varying accuracy of the ANN-based surrogate model under different ventilation conditions and occupant arrangements. While the model demonstrated high accuracy for certain scenarios with mixing ventilation at higher rates, its performance significantly deteriorated in other configurations such as displacement ventilation at lower rates. These findings underscore the model's current limitations and the necessity for further refinement, particularly in handling different ventilation strategies and air flow patterns. Future work should focus on expanding the

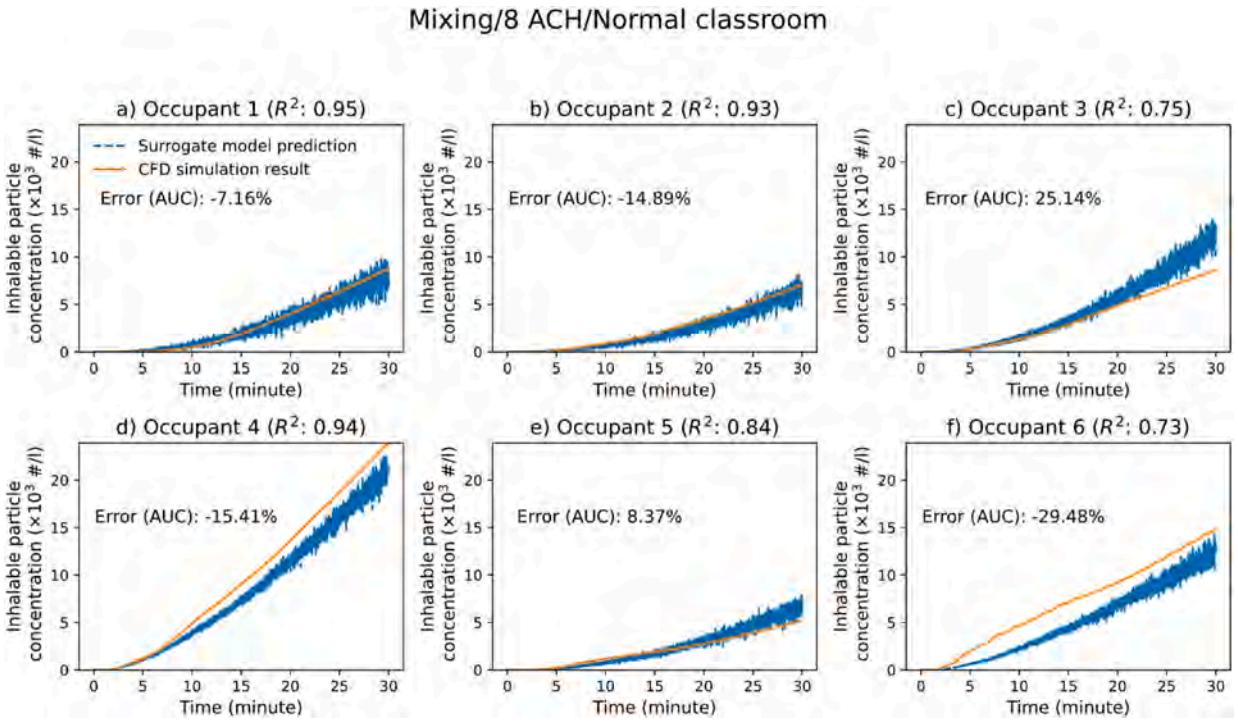

Fig. 15. Comparison of time-varying inhalable particle number concentrations over 30 min between CFD simulation and surrogate model predictions for each occupant in the normal classroom arrangement with mixing ventilation at 8 air change rate, with occupant 7 as the infector. The percent difference in Area Under the Curve (AUC) is indicated for each occupant.

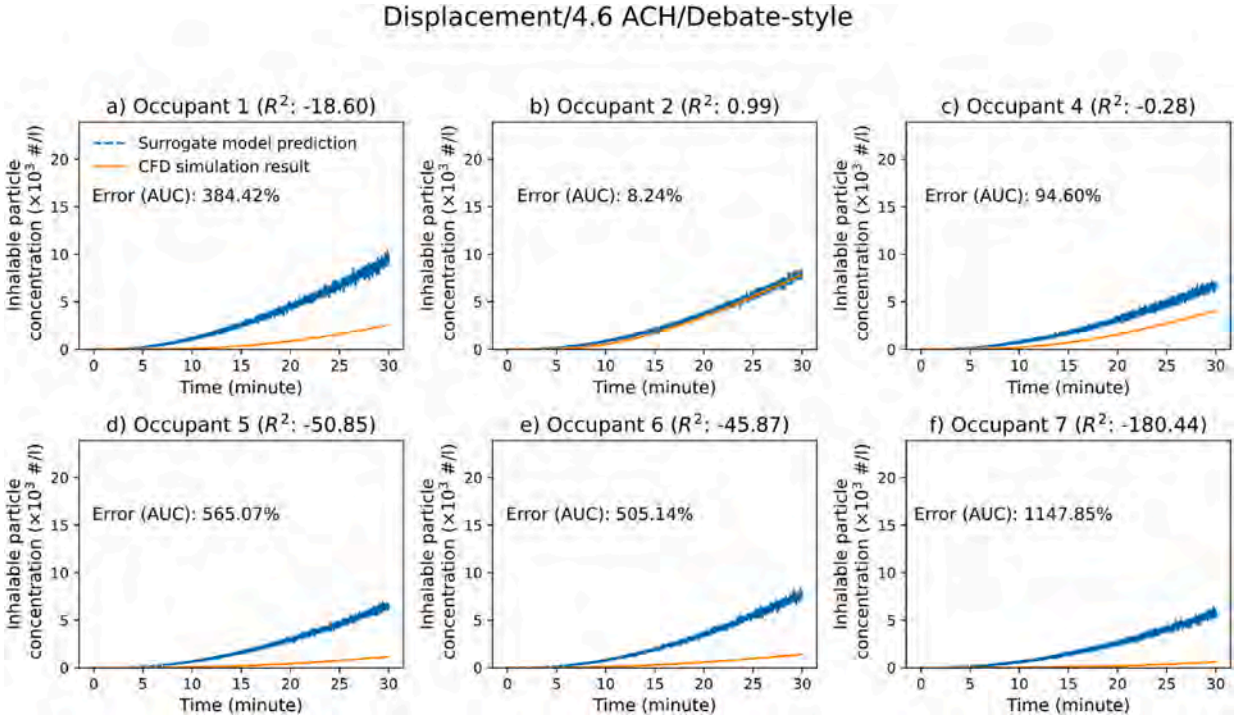

Fig. 16. Comparison of time-varying inhalable particle number concentrations over 30 min between CFD simulation and surrogate model predictions for each occupant in the debate-style arrangement with displacement ventilation at 4.6 air change rate, with occupant 3 as the infector. The percent difference in Area Under the Curve (AUC) is indicated for each occupant.

training dataset with more varied scenarios (e.g., using data augmentation techniques) and incorporating additional features to enhance the robustness and generalizability of the surrogate model. This will ultimately contribute to more reliable infection risk predictions in diverse indoor environments.

3.7.4. Inference performance by the AI based surrogate model

An out-of-range prediction using the trained model is referred to as inference. To evaluate the model's performance on inference, two additional Computational Fluid Dynamics (CFD) cases were conducted. These cases had an air change rate of 13.8 h^{-1} under both mixing and displacement ventilation conditions, with occupant 1 designated as the infector in a normal classroom arrangement.

As shown in Table 8, the model's performance under mixing ventilation was notably accurate (see Fig. S 5 and S 6 in Supplementary Materials). This level of accuracy is likely due to the enhanced air mixing, which results in a more uniform distribution of infectious particles throughout the space. The homogeneous nature of the well-mixed environment likely facilitated easier training and improved prediction accuracy.

In contrast, the model's performance under displacement ventilation is significantly poorer. This discrepancy underscores the complex nature of infectious particle dispersion in displacement ventilation systems. The stratification created by displacement ventilation introduces additional variables and complexities that the current model may not adequately capture.

These findings, once again, emphasize the critical role of ventilation strategies in developing and applying predictive models for infectious particle dispersion in indoor environments. While the model shows promise in scenarios involving mixing ventilation, its limitations under displacement ventilation highlight the ongoing challenges in accurately predicting airborne disease transmission risks across different ventilation conditions.

4. Conclusions

The present study offers a thorough analysis of airborne transmission in classroom settings, focusing on key factors that significantly influence infection risk, including ventilation strategies, air change rates, occupant arrangements, and source locations. We meticulously generated a comprehensive set of 224 CFD simulation cases to explore these variables. Furthermore, we trained data-driven supervised learning models—specifically Long Short-Term Memory (LSTM) networks and Artificial Neural Networks (ANNs)—to enable rapid prediction of airborne infection risks. The major findings are as follows:

1) The upper quartile of infection risk was found to be up to 6.5 times greater than the predictions made by the traditional Wells-Riley model, highlighting a significant discrepancy. This discrepancy underscores the model's limitations in accurately capturing the complex spatial distribution of infection risk.
2) Increasing air change rates were found to reduce infection risks for occupants in the front rows of a classroom. However, unexpectedly, these higher air change rates were associated with a significant increase in risk—up to 166 %—for occupants seated in the back rows. This paradox challenges the conventional belief that higher air change rates uniformly enhance safety.
3) In the absence of control measures, displacement ventilation leads to a significant reduction in infection risk, ranging from 49 % to 77 %, compared to mixing ventilation.
4) Higher ventilation rates under mixing ventilation reduce the impact of occupant arrangement, while under displacement ventilation, susceptible occupants in a debate-style arrangement experience similar or up to 30 % lower risks compared to those in a normal classroom. This highlights the substantial effect of occupant arrangement on infection risk, particularly with displacement ventilation.

Some limitations should be noted. Although Artificial Neural Networks (ANNs) effectively modeled the relationship between particle distribution and infection risks under mixing ventilation, the surrogate model had limitations in fully capturing the intricate relationships, particularly with displacement ventilation. Future research should explore advanced data augmentation techniques and the integration of additional variables to enhance predictive accuracy and robustness in this context. Moreover, simulating a relatively small number of occupants in a classroom may not fully capture the complexity and variability of real-world environments, potentially affecting the generalizability of our findings. Future studies should investigate the behavior and interactions of a larger population in occupied spaces. Additionally, validating the model with full-scale field measurements could provide a more robust basis for assessing its accuracy and applicability. Incorporating dynamic inputs such as CO_2 concentrations and other environmental variables would further help capture nuanced interactions and improve predictive accuracy.

Table 8
Performance evaluation on inference, comparing two different ventilation strategies.

	Mixing		Displacement	
	R^2	CVRMSE (%)	R^2	CVRMSE (%)
Occupant 2	0.75	37	−37	542
Occupant 3	0.69	39	−53	532
Occupant 4	0.03	68	0.94	17
Occupant 5	0.89	25	−89	1020
Occupant 6	0.12	86	−10	302
Occupant 7	0.98	12	0.01	75

Funding sources

This research was supported by Centers for Disease Control and Prevention (CDC, Contract #75D301-23-P-17789) and the U.S. National Science Foundation (NSF Grant 1944325).

CRediT authorship contribution statement

Hyeonjun Lee: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis. **Donghyun Rim:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Funding acquisition, Conceptualization.

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.

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jobe.2024.110760.

References

[1] A. Shechter, et al., Psychological distress, coping behaviors, and preferences for support among New York healthcare workers during the COVID-19 pandemic, Gen. Hosp. Psychiatr. 66 (Sep. 2020) 1–8, https://doi.org/10.1016/j.genhosppsych.2020.06.007.
[2] K. Usher, J. Durkin, N. Bhullar, The COVID-19 Pandemic and Mental Health Impacts, Blackwell Publishing, Jun. 01, 2020, https://doi.org/10.1111/inm.12726.
[3] D.B.G. Tai, A. Shah, C.A. Doubeni, I.G. Sia, M.L. Wieland, The Disproportionate Impact of COVID-19 on Racial and Ethnic Minorities in the United States, Feb. 15, 2021, https://doi.org/10.1093/cid/ciaa815. Oxford University Press.
[4] A. A. Rabaan et al., "Airborne Transmission of SARS-CoV-2 Is the Dominant Route of Transmission: Droplets and Aerosols King Fahad Medical City.".
[5] H. Motamedi, M. Shirzadi, Y. Tominaga, P.A. Mirzaei, CFD modeling of airborne pathogen transmission of COVID-19 in confined spaces under different ventilation strategies, Sustain. Cities Soc. 76 (Jan) (2022), https://doi.org/10.1016/j.scs.2021.103397.
[6] S. Park, R. Mistrick, D. Rim, Performance of upper-room ultraviolet germicidal irradiation (UVGI) system in learning environments: Effects of ventilation rate, UV fluence rate, and UV radiating volume, Sustain. Cities Soc. 85 (2022) 104048.
[7] H. Dai, B. Zhao, Association of the infection probability of COVID-19 with ventilation rates in confined spaces, Build. Simulat. 13 (6) (Dec. 2020) 1321–1327, https://doi.org/10.1007/s12273-020-0703-5.
[8] M.F. King, C.J. Noakes, P.A. Sleigh, Modeling environmental contamination in hospital single- and four-bed rooms, Indoor Air 25 (6) (Dec. 2015) 694–707, https://doi.org/10.1111/ina.12186.
[9] H.C. Yu, K.W. Mui, L.T. Wong, H.S. Chu, Ventilation of general hospital wards for mitigating infection risks of three kinds of viruses including Middle East respiratory syndrome coronavirus, Indoor Built Environ. 26 (4) (Apr. 2017) 514–527, https://doi.org/10.1177/1420326X16631596.
[10] Y. Feng, T. Marchal, T. Sperry, H. Yi, Influence of wind and relative humidity on the social distancing effectiveness to prevent COVID-19 airborne transmission: a numerical study, J. Aerosol Sci. 147 (Sep) (2020), https://doi.org/10.1016/j.jaerosci.2020.105585.
[11] Y. Shang, J. Dong, L. Tian, F. He, J. Tu, An improved numerical model for epidemic transmission and infection risks assessment in indoor environment, J. Aerosol Sci. 162 (May 2022), https://doi.org/10.1016/j.jaerosci.2021.105943.
[12] W. Su, et al., Infection probability under different air distribution patterns, Build. Environ. 207 (Jan) (2022), https://doi.org/10.1016/j.buildenv.2021.108555.
[13] S. Zhang, D. Niu, Y. Lu, Z. Lin, Contaminant removal and contaminant dispersion of air distribution for overall and local airborne infection risk controls, Sci. Total Environ. 833 (Aug) (2022), https://doi.org/10.1016/j.scitotenv.2022.155173.
[14] A. Khosronejad, et al., Fluid dynamics simulations show that facial masks can suppress the spread of COVID-19 in indoor environments, AIP Adv. 10 (12) (Dec. 2020), https://doi.org/10.1063/5.0035414.
[15] H. Alenezi, M.E. Cam, M. Edirisinghe, A novel reusable anti-COVID-19 transparent face respirator with optimized airflow, Biodes Manuf 4 (1) (Mar. 2021) 1–9, https://doi.org/10.1007/s42242-020-00097-1.
[16] S. Leonard, et al., Reducing aerosol dispersion by high flow therapy in COVID-19: high resolution computational fluid dynamics simulations of particle behavior during high velocity nasal insufflation with a simple surgical mask, JACEP Open 1 (4) (Aug. 2020) 578–591, https://doi.org/10.1002/emp2.12158.
[17] S. Tan, Z. Zhang, K. Maki, K.J. Fidkowski, J. Capecelatro, Beyond well-mixed: a simple probabilistic model of airborne disease transmission in indoor spaces, Indoor Air 32 (3) (Mar. 2022), https://doi.org/10.1111/ina.13015.
[18] G. Calzolari, W. Liu, Deep Learning to Replace, Improve, or Aid CFD Analysis in Built Environment Applications: A Review, Elsevier Ltd., Dec. 01, 2021, https://doi.org/10.1016/j.buildenv.2021.108315.
[19] D. Panchigar, K. Kar, S. Shukla, R.M. Mathew, U. Chadha, S.K. Selvaraj, Machine Learning-Based CFD Simulations: a Review, Models, Open Threats, and Future Tactics, Springer Science and Business Media Deutschland GmbH, Dec. 01, 2022, https://doi.org/10.1007/s00521-022-07838-6.
[20] X. Jin, P. Cheng, W.L. Chen, H. Li, Prediction model of velocity field around circular cylinder over various Reynolds numbers by fusion convolutional neural networks based on pressure on the cylinder, Phys. Fluids 30 (4) (Apr. 2018), https://doi.org/10.1063/1.5024595.
[21] X. Guo, W. Li, F. Iorio, Convolutional neural networks for steady flow approximation, in: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery, Aug. 2016, pp. 481–490, https://doi.org/10.1145/2939672.2939738.
[22] J. Zhang, X. Zhao, Machine-learning-based surrogate modeling of aerodynamic flow around distributed structures, AIAA J. 59 (3) (2021) 868–879, https://doi.org/10.2514/1.J059877.
[23] X. Jurado, N. Reiminger, M. Benmoussa, J. Vazquez, C. Wemmert, Deep learning methods evaluation to predict air quality based on Computational Fluid Dynamics, Expert Syst. Appl. 203 (Oct. 2022), https://doi.org/10.1016/j.eswa.2022.117294.
[24] P. Kastner, T. Dogan, A GAN-based surrogate model for instantaneous urban wind flow prediction, Build. Environ. 242 (Aug) (2023), https://doi.org/10.1016/j.buildenv.2023.110384.

[25] B. Wilson, S. Wakes, M. Mayo, Surrogate modeling a computational fluid dynamics-based wind turbine wake simulation using machine learning, in: 2017 IEEE Symposium Series on Computational Intelligence (SSCI), Institute of Electrical and Electronics Engineers, Honolulu, HI, USA, Feb. 2017.
[26] T. Van Quang, D.T. Doan, N.L. Phuong, G.Y. Yun, Data-driven prediction of indoor airflow distribution in naturally ventilated residential buildings using combined CFD simulation and machine learning (ML) approach, J. Build. Phys. 47 (4) (Jan. 2024) 439–471, https://doi.org/10.1177/17442591231219025.
[27] Q. Zhou, R. Ooka, Neural network for indoor airflow prediction with CFD database, in: Journal of Physics: Conference Series, Institute of Physics, Dec. 2021, https://doi.org/10.1088/1742-6596/2069/1/012154.
[28] Q. Zhou, R. Ooka, Influence of data preprocessing on neural network performance for reproducing CFD simulations of non-isothermal indoor airflow distribution, Energy Build. 230 (Jan) (2021), https://doi.org/10.1016/j.enbuild.2020.110525.
[29] C.A. Faulkner, et al., Fast prediction of indoor airflow distribution inspired by synthetic image generation artificial intelligence, Build. Simulat. 16 (7) (Jul. 2023) 1219–1238, https://doi.org/10.1007/s12273-023-0989-1.
[30] R. Zhang, D. Liu, L. Shi, Thermal-comfort optimization design method for semi-outdoor stadium using machine learning, Build. Environ. 215 (May 2022), https://doi.org/10.1016/j.buildenv.2022.108890.
[31] A. Warey, S. Kaushik, B. Khalighi, M. Cruse, G. Venkatesan, Data-driven prediction of vehicle cabin thermal comfort: using machine learning and high-fidelity simulation results, Int. J. Heat Mass Tran. 148 (Feb) (2020), https://doi.org/10.1016/j.ijheatmasstransfer.2019.119083.
[32] T. hu Zhang, X. yi You, A simulation-based inverse design of preset aircraft cabin environment, Build. Environ. 82 (2014) 20–26, https://doi.org/10.1016/j.buildenv.2014.08.002.
[33] M. Mesgarpour, et al., Prediction of the spread of Corona-virus carrying droplets in a bus - a computational based artificial intelligence approach, J. Hazard Mater. 413 (Jul) (2021), https://doi.org/10.1016/j.jhazmat.2021.125358.
[34] G. Pei, D. Rim, Quality control of computational fluid dynamics (CFD) model of ozone reaction with human surface: effects of mesh size and turbulence model, Build. Environ. 189 (Feb) (2021), https://doi.org/10.1016/j.buildenv.2020.107513.
[35] I.B. Celik, U. Ghia, P.J. Roache, C.J. Freitas, H. Coleman, P.E. Raad, Procedure for estimation and reporting of uncertainty due to discretization in CFD applications, Journal of Fluids Engineering, Transactions of the ASME 130 (7) (Jul. 2008) 780011–780014, https://doi.org/10.1115/1.2960953.
[36] F. Chen, S.C.M. Yu, A.C.K. Lai, Modeling particle distribution and deposition in indoor environments with a new drift-flux model, Atmos. Environ. 40 (2) (Jan. 2006) 357–367, https://doi.org/10.1016/j.atmosenv.2005.09.044.
[37] X. Zhao, S. Liu, Y. Yin, T. Zhang, Q. Chen, Airborne Transmission of COVID-19 Virus in Enclosed Spaces: an Overview of Research Methods, John Wiley and Sons Inc, Jun. 01, 2022, https://doi.org/10.1111/ina.13056.
[38] N. Izadyar, W. Miller, Ventilation Strategies and Design Impacts on Indoor Airborne Transmission: A Review, Elsevier Ltd., Jun. 15, 2022, https://doi.org/10.1016/j.buildenv.2022.109158.
[39] T.W. Tsang, L.T. Wong, K.W. Mui, Experimental Studies on Airborne Transmission in Hospitals: A Systematic Review, SAGE Publications Ltd, Apr. 01, 2024, https://doi.org/10.1177/1420326X231205527.
[40] X. Zhao, S. Liu, Y. Yin, T. Zhang, Q. Chen, Airborne Transmission of COVID-19 Virus in Enclosed Spaces: an Overview of Research Methods, John Wiley and Sons Inc, Jun. 01, 2022, https://doi.org/10.1111/ina.13056.
[41] ASHRAE and ANSI, Ventilation and acceptable indoor air quality, Atlanta, GA, www.ashrae.org/technology, 2022.
[42] S.L. Miller, et al., Transmission of SARS-CoV-2 by inhalation of respiratory aerosol in the Skagit Valley Chorale superspreading event, Indoor Air 31 (2) (Mar. 2021) 314–323, https://doi.org/10.1111/ina.12751.
[43] M. Braniš, P. Řezáčová, M. Domasová, The effect of outdoor air and indoor human activity on mass concentrations of PM10, PM2.5, and PM1 in a classroom, Environ. Res. 99 (2) (Oct. 2005) 143–149, https://doi.org/10.1016/j.envres.2004.12.001.
[44] D.A. Edwards, et al., Inhaling to mitigate exhaled bioaerosols, Proc. Natl. Acad. Sci. USA 101 (50) (2004) 17383–17388 [Online]. Available: www.pnas.orgcgidoi10.1073pnas.0408159101.
[45] P. Fabian, et al., Influenza virus in human exhaled breath: an observational study, PLoS One 3 (7) (Jul. 2008), https://doi.org/10.1371/journal.pone.0002691.
[46] S.J Lakey, Pascale, Youngbo Won, David Shaw, Freja F. Østerstrøm, James Mattila, Emily Reidy, Brandon Bottorff, et al., Spatial and temporal scales of variability for indoor air constituents, Commun. Chem. 4 (1) (2021) 110.
[47] M. Porta, A dictionary of epidemiology, Sixth (2014).
[48] R.R. Netz, Mechanisms of airborne infection via evaporating and sedimenting droplets produced by speaking, Am. Chem. Soc. (2020), https://doi.org/10.1021/acs.jpcb.0c05229. Aug. 20.
[49] F. Yang, A.A. Pahlavan, S. Mendez, M. Abkarian, H.A. Stone, Towards improved social distancing guidelines: space and time dependence of virus transmission from speech-driven aerosol transport between two individuals, Phys Rev Fluids 5 (12) (Dec. 2020), https://doi.org/10.1103/PhysRevFluids.5.122501.
[50] J.D. Pleil, M. Ariel Geer Wallace, M.D. Davis, C.M. Matty, The Physics of Human Breathing: Flow, Timing, Volume, and Pressure Parameters for Normal, On-Demand, and Ventilator Respiration, IOP Publishing Ltd, Oct. 01, 2021, https://doi.org/10.1088/1752-7163/ac2589.
[51] J.K. Gupta, C.H. Lin, Q. Chen, Characterizing exhaled airflow from breathing and talking, Indoor Air 20 (1) (2010) 31–39, https://doi.org/10.1111/j.1600-0668.2009.00623.x.
[52] D. Rim, A. Novoselac, Transport of particulate and gaseous pollutants in the vicinity of a human body, Build. Environ. 44 (9) (Sep. 2009) 1840–1849, https://doi.org/10.1016/j.buildenv.2008.12.009.
[53] S. Zhang, Z. Lin, Dilution-based evaluation of airborne infection risk - thorough expansion of Wells-Riley model, Build. Environ. 194 (May 2021), https://doi.org/10.1016/j.buildenv.2021.107674.
[54] H. Dai, B. Zhao, Association between the infection probability of COVID-19 and ventilation rates: an update for SARS-CoV-2 variants, Build. Simulat. 16 (1) (Jan. 2023) 3–12, https://doi.org/10.1007/s12273-022-0952-6.
[55] T. Hastie, R. Tibshirani, J. Friedman, in: Overview of Supervised Learning, 2009, pp. 1–33, https://doi.org/10.1007/b94608_2.
[56] I.A. Basheer, M. Hajmeer, Artificial neural networks: fundamentals, computing, design, and application [Online]. Available: www.elsevier.com/locate/jmicmeth, 2000.
[57] Y. Yu, X. Si, C. Hu, J. Zhang, A Review of Recurrent Neural Networks: Lstm Cells and Network Architectures, MIT Press Journals, Jul. 01, 2019, https://doi.org/10.1162/neco_a_01199.
[58] R.C. Staudemeyer, E.R. Morris, Understanding LSTM – a tutorial into Long short-term memory recurrent neural networks [Online]. Available: http://arxiv.org/abs/1909.09586, Sep. 2019.
[59] J.M. Villafruela, I. Olmedo, F.A. Berlanga, M. Ruiz de Adana, Assessment of displacement ventilation systems in airborne infection risk in hospital rooms, PLoS One 14 (1) (Jan. 2019), https://doi.org/10.1371/journal.pone.0211390.
[60] R.K. Bhagat, P.F. Linden, Displacement ventilation: a viable ventilation strategy for makeshift hospitals and public buildings to contain COVID-19 and other airborne diseases: ventilation strategy for COVID-19, R. Soc. Open Sci. 7 (9) (Sep. 2020), https://doi.org/10.1098/rsos.200680.
[61] Y. Yin, et al., Experimental study on displacement and mixing ventilation systems for a patient ward, HVAC R Res. 15 (6) (2009) 1175–1191, https://doi.org/10.1080/10789669.2009.10390885.
[62] J. Pantelic, K.W. Tham, Adequacy of air change rate as the sole indicator of an air distribution system's effectiveness to mitigate airborne infectious disease transmission caused by a cough release in the room with overhead mixing ventilation: a case study, HVAC R Res. 19 (8) (Nov. 2013) 947–961, https://doi.org/10.1080/10789669.2013.842447.
[63] E.S. Mousavi, K.R. Grosskopf, Ventilation rates and airflow pathways in patient rooms: a case study of bioaerosol containment and removal, Ann. Occup. Hyg. 59 (9) (Dec. 2014) 1190–1199, https://doi.org/10.1093/annhyg/mev048.
[64] C. Wang, S. Holmberg, S. Sadrizadeh, Numerical study of temperature-controlled airflow in comparison with turbulent mixing and laminar airflow for operating room ventilation, Build. Environ. 144 (Oct. 2018) 45–56, https://doi.org/10.1016/j.buildenv.2018.08.010.
[65] Z.D. Bolashikov, A.K. Melikov, W. Kierat, Z. Popioek, M. Brand, Exposure of health care workers and occupants to coughed airborne pathogens in a double-bed hospital patient room with overhead mixing ventilation, in: HVAC and R Research, Aug. 2012, pp. 602–615, https://doi.org/10.1080/10789669.2012.682692.

[66] Z.D. Bolashikov, A.K. Melikov, M. Barova, Exposure to exhaled air from a sick occupant in a two-bed hospital room with mixing ventilation: effect of posture of doctor and air change rate, in: In Proceedings of 11th REHVA World Congress and the 8th International Conference on Indoor Air Quality, Ventilation and Energy Conservation in Buildings, 2013.
[67] J. Pantelic, K.W. Tham, Adequacy of air change rate as the sole indicator of an air distribution system's effectiveness to mitigate airborne infectious disease transmission caused by a cough release in the room with overhead mixing ventilation: a case study, HVAC R Res. 19 (8) (Nov. 2013) 947–961, https://doi.org/10.1080/10789669.2013.842447.
[68] Y. Yan, X. Li, X. Fang, Y. Tao, J. Tu, A spatiotemporal assessment of occupants' infection risks in a multi-occupants space using modified Wells–Riley model, Build. Environ. 230 (2023), https://doi.org/10.1016/j.buildenv.2023.110007. Feb.
[69] B. Choubin, H. Darabi, O. Rahmati, F. Sajedi-Hosseini, B. Kløve, River suspended sediment modelling using the CART model: a comparative study of machine learning techniques, Sci. Total Environ. 615 (Feb. 2018) 272–281, https://doi.org/10.1016/j.scitotenv.2017.09.293.
[70] Y. Xu, et al., Evaluation of machine learning techniques with multiple remote sensing datasets in estimating monthly concentrations of ground-level PM2.5, Environ. Pollut. 242 (Nov. 2018) 1417–1426, https://doi.org/10.1016/j.envpol.2018.08.029.
[71] S. Park, D. Song, CO2 concentration as an indicator of indoor ventilation performance to control airborne transmission of SARS-CoV-2, J Infect Public Health 16 (7) (Jul. 2023) 1037–1044, https://doi.org/10.1016/j.jiph.2023.05.011.
[72] S.N. Rudnick, D.K. Milton, Risk of indoor airborne infection transmission estimated from carbon dioxide concentration, Indoor Air 13 (3) (2003) 237–245 [Online]. Available: www.blackwellpublishing.com/ina.
[73] Z. Peng, J.L. Jimenez, Exhaled CO2as a COVID-19 infection risk proxy for different indoor environments and activities, Environ. Sci. Technol. Lett. 8 (5) (May 2021) 392–397, https://doi.org/10.1021/acs.estlett.1c00183.