Indian Journal of Engineering & Materials Sciences
Vol. 31, October 2024, pp. 650-666
DOI: 10.56042/ijems.v31i5.9862

Event specific transmission prognosis gleaned from machine learning for covid-19 prophylaxis in air-conditioned office

Nishant Raj Kapoor[a], Denise-Penelope N. Kontoni[b,c*], Ashok Kumar[a], Anuj Kumar[a], Aman Kumar[a,d] & Harish Chandra Arora[a,d]

[a] Academy of Scientific and Innovative Research (AcSIR), Ghaziabad 201 002, India

[b]Department of Civil Engineering, School of Engineering, University of the Peloponnese, GR 26334 Patras, Greece

[c]School of Science and Technology, Hellenic Open University, GR 26335 Patras, Greece

[d]Structural Engineering Department, CSIR - Central Building Research Institute, Roorkee 247 667, India

Received: 24 March 2024; accepted: 29 July 2024

The unprecedented rate of emerging and re-emerging cases of COVID-19, along with novel variants and the recent recognition of long COVID-19, has made it essential to explore the transmission of SARS-CoV-2. An air-conditioned office room has been investigated to understand the event-specific possibility of viral transmission of the coronavirus in composite climatic conditions. This work has introduced newly developed machine learning models to forecast the R-Event using artificial neural networks (ANN), Gaussian Process Regression (GPR), Support Vector Machine (SVM), and Curve-Fitting (CF). Real-time monitoring of the air-conditioned office room has been carried out during the months of April and May of 2022 to collect data. The proposed models have been evaluated for their performances. To assess the merit of the four models, statistical parameters—namely the correlation coefficient, mean absolute error, root mean square error, mean absolute percentage error, Nash-Sutcliffe efficiency index, and a20-index have been considered. Eight input features have been used to predict the target R-Event in this study. The main objective of this study has been to develop a model that links CO_2 concentration with the R-Event value for forecasting event-specific infections in an air-conditioned office environment. Results have indicated that the developed ANN prediction model is the best among the four models for forecasting the R-Event.

Keywords: Machine learning (ML), Artificial neural network (ANN), Indoor air quality (IAQ), Viral transmission, COVID-19, Air-conditioning, Indoor environmental quality (IEQ), Carbondioxide (CO_2), Air-conditioned office, Building.

1 Introduction

SARS-CoV-2 is a virus that is related to the enormous coronavirus family. Both people and certain animals can be infected by the variants of these viruses. SARS-CoV-2 is an extremely infectious human pathogen. The virus is transmitted from person to person by several ways predominantly by droplets expelled when coughing, sneezing, or talking by an infected person. A less prevalent method of transmission is through contacting one's mouth, nose, or eyes after touching a surface that has the virus on it[1]. As of February 21, 2022, there have been more than 757 million confirmed cases of coronavirus infected humans worldwide, resulting in approximately 6.9 million human fatalities [2]. The global cases of several strains of the coronavirus disease are surging again in several parts of the globe as the virus in its fourth year since its first reported case in December 2019. The growth of this airborne virus is exponential and the infectivity rate is also higher. This pandemic causing virus has affected more than 200 nations, regions, and territories. With a 7.8 billion human population, COVID-19 has infected roughly 1 in every 12 people living today and 1 in every 1200 people were already dead throughout the globe with several variants of this virus.SARS-CoV-2 has been referred to throughout the study as SC-2.

Many countries already faced three waves of this pandemic and some countries even faced a fourth wave. There may be cases of silent waves as well throughout many regions of the world. It is well-accepted that SC-2 is an airborne virus, which is quite a big concern to reduce its transmission. Another concern is "long-COVID", a novel term for the long-term negative influence of these viruses on human health [3,4]. To tackle this virus transmission several methods have been tried by several governments, regulatory bodies, and organizations. These methods include social distancing, compulsion of mask,

* Corresponding author (Email: kontoni@uop.gr ; kontoni.denise@ac.eap.gr)

carrying sanitizers, isolations, limited occupancy, hygiene maintenance, and several other prevention and control measures[5]. In buildings it may be tempting to enact social isolation or occupancy limitations that are "one-size-fits-all", but the reality is that indoor air quality (IAQ) varies greatly depending on the environment, kind of building, and occupants' behavioral pattern [6]. As a result, it is difficult to say that one type of ventilation is ubiquitous.

Widely occurred "super-spreading" instances, in which numerous people are confirmed as infected, took place in enclosed built settings like workplaces, schools, restaurants, hospitals, houses, apartment buildings, assembly halls, etc. [7-12].The bulk of people's time is consumed in built spaces, and the condition of their interior environment directly affects their health [13–15]. Pathogen transmission inside is influenced by indoor environmental quality (IEQ). According to research, SC-2 may spread through the air, which explains why COVID-19 cases have dramatically increased in confined settings with poor ventilation [16]. Several countries throughout the world have issued COVID-19-fighting suggestions while considering energy efficiency for sustainability concerns [17,18]. Almost all of the standards recommend natural ventilation, distinct zonation in buildings, CO_2 as an indicator, window/door opening, and other factors throughout these recommendations for efficient SC-2 prevention in buildings. However, air conditioning is widely used nowadays and research on the transmission of pathogens in AC rooms needs to be conducted.

The process of air-conditioning is used to establish and maintain certain relative humidity levels, temperature, and air quality levels in enclosed environments. Usually, this procedure is used to sustain the occupant's level of comfort. However, due to the enclosed environment, the ventilation is affected generally. In a poorly ventilated space, the aerosol suspension time rose tenfold [19]. Interior ventilation and indoor CO_2 levels have a significant relation in a steady-state situation [20].If an appropriate technique to identify the internal ventilation rate is not available, the indoor CO_2 levels may be utilized as an indication. Traditional CO_2 measuring techniques are unable to detect changes in CO_2 concentration across time or space frequently. There are now possibilities to monitor CO_2 and precisely anticipate the concentration levels within buildings owing to recent technology advancements in CO_2 monitoring and prediction. As a result, a richer understanding of CO_2 exposure in people can be attained with less effort [20].

Recently, several researchers tried different ways to link indoor CO_2 levels with viral transmissions in built environments of different types [21-28]. Rudnick and Milton[21] estimate the risk of viral transmission through indoor air based on the CO_2 level. They develop a model that calculates how much of the air being inhaled has already been expelled by someone within the building by using the CO_2 indoors as a measure for breathing out infected air exposure. Outside air supply rate was not taken into account, nor was it assumed that it would remain constant over time, when the researchers created a CO_2-based risk model. They also refrained from presuming that an infectious agent's concentration had achieved a steady state. A similar idea is given by Peng and Jimenez[22] to assess CO_2 indoors as a SC-2 infection risk proxy. They established analytical methods for CO_2-based risk proxies. They also mention a drawback that there are considerable indecisions in the infection risk approximations, which are mostly based on viral exhalation rates. However, in order to improve public health and safety, they advocate implementing inexpensive CO_2 based indoor infection risk monitoring devices. Bazant and Bush[23] suggested limiting the duration spent in a mutual environment with a sick occupant to reduce virus aerial transmission. Afterwards, Bazant *et al.*[24] take into account inhabitation duration, as well as the mean concentration of CO_2 exhaled inside, showcasing the utilization of CO_2 sensors in the risk evaluation of aerial viral transmission. Using CO_2 levels, those authors created an analytical model to predict the likelihood of aerial transmission. While CO_2 levels are thought of as a direct indication of air mixing as well as ventilation, the investigators came to the definite conclusion that CO_2 levels can be used to assess the likelihood of SC-2 transmission via air.

Wang *et al.*[25] published a review study on the aerial communication of viruses via respirational activities. Authors suggest that CO_2 sensors can help in monitoring as well as improving ventilation by acting as an indicator of the accumulated breath-out air. The recommended range for carbon dioxide concentrations is 700 to 800 ppm, however, ventilation techniques should also be considered. Martinez *et al.*[26] created a new agent-based simulator called “Arch ABM.” A few attempts to forecast airborne virus transmission in a naturally ventilated

office space using AI were recently reported by Kapoor *et al*. [1,27,28]. The authors of the study[1] used ANN and curve-fitting techniques to relate IEQ parameters (CO_2 levels, humidity, and temperature) to occupant behavior of window-door opening, occupancy, outside pollution, ceiling fan usage, and R-Event likelihood. The aforementioned factors were linked using several additional machine learning (ML) techniques in a different research by Kapoor *et al*. [27,28]. By combining data on indoor CO_2 concentration, occupancy, occupant behavior, and other outdoor-indoor environmental parameters that are obtained in a real-time air-conditioned office setting, this research extends prior studies by using AI to anticipate the airborne transmission.

Other significant indoor environmental elements that affect virus survival are temperature and humidity. Both temperature and humidity significantly influence indoor human comfort and directly affect how individuals feel. After reviewing the available data on the effects of temperature as well as humidity on SC-2transmission, Mecenas *et al*.[29] came to the conclusion that impending studies should take into account these variables since they could affect the virus's spread. A number of international recommendations have listed temperature as well as humidity as crucial aspects. Most recommendations offered boundary parameters for interior temperature and relative humidity that considered both indoor occupant comfort and the SC-2 aerial transmission. Kapoor *et al*.[30] proposed several strategies for COVID-19 prophylaxis in buildings of different typologies. As the indoor temperature and humidity are essential parameters for indoor human comfort, these were considered as important parameters in the proposed strategies. It was recommended to update the National Building Code according to the current state of research and the growing requirements to prevent viral transmissions in the future as well[30]. The boundary parameters for inside humidity as well as temperature recommended by various standards for minimizing SC-2 transmission while preserving comfort are given in the literature[31-41]. According to all of the aforementioned recommendations, a safe comfortable indoor relative humidity and temperature range is between 40 to 60 percent and 24 to 26 degrees Celsius, respectively.

Occupant behavior, outside environmental conditions, and interior occupant density are also important variables in limiting SC-2 transmission in a confined environment[42,43]. Several governments as well as organizations worldwide have released guidelines and recommendations on COVID-19 appropriate behavior in workplace situations [44,45]. While complying with these recommendations and guidelines the most important factors in air-conditioned environments are energy needs. The energy consumption rises as a result of the opening requirements in air-conditioned spaces. For the total population, it is quite difficult to strike an optimum balance between comfort, safety, and energy in any air-conditioned environment.

PM2.5 and PM10, which represent the quantity of pollutants in the air, are the most important part of the AQI. Several researchers attempted to establish a link between particulate matter and SC-2 transmission[46-51]. Because the SC-2 is so tiny, it may readily settle on particulate matter as well as dust and then be spread by the re-suspension in built environments caused by human activity as well as natural factors. In confined locations with high occupancy, the risks of breathing free-floating virus load in the air are substantially higher. In addition, microscopic microbes can enter deeply into our bodies and cause severe disorders.

Tupper *et al*.[52] proposed the concept of “event R” in relation to COVID-19. According to Tupper *et al*., “Event-R is the estimated number of new infections attributable to the presence of a single infected individual at an event”. The authors established a fundamental relationship between “event R” as well as other variables namely transmission strength, mixing degree, individual closeness, and exposure time. REHVA included R-Event in their improved version of the REHVA calculator for aerial SC-2 prediction tool in addition to infection likelihood[53]. Scientific study on the antecedent spreading of different Coronavirus strains in indoor environments across the world is presently determining the precise infection likelihood. The “REHVA calculator” is used to calculate the R-Event in AC settings to establish a link between R-Event and indoor CO_2 concentration in an office environment. The goal of this work is to anticipate the R-Event using interior CO_2 concentration in an office setting. Aside from CO_2, exterior as well as interior environmental factors, and occupancy are also important features.

Many researchers developed numerical models and simulations to anticipate infection likelihood utilizing objective as well as subjective data. Only a few

studies have attempted to find the relation between indoor CO_2 levels and coronavirus transmission [21-28]. However, much as CO_2 may be used scientifically as a proxy for indoor SC-2 content, indoor CO_2 concentration can also be used to forecast likely infected patients. Artificial Intelligence (AI) has not yet been utilized in this research and has great potential to nudge the related works. The research aims are: monitoring and collection of real-time spatio-temporal data of objective (environmental) as well as subjective (occupancy-related) variables in an air-conditioned (AC) office to apply ML techniques for linkingmonitored parameters; comparing newly developed models (for an air-conditioned indoor environment) for prediction of R-Event for COVID-19 prophylaxis.

The R-Event value was predicted in this work for forecasting the likelihood of coronavirus transmission using Machine Learning (ML) models and Curve Fitting (CF). Section 2 discusses the materials and procedures employed in this study. This part also covers data collecting, standardization, and CF as well as ML methodology in depth. Section 3 presents the outcomes of the work and discussion on the same with highlighting the limitations. At last, section 4 concludes the work.

2 Materials and Methods

CO_2 levels in the office were continually monitored using an "EXTECH Indoor Air Quality Meter/Data-logger Model EA80." Temperature and humidity readings were taken using the "EXTECH Heat Stress WBGT Meter Model HT200". The details of all the materials and methods are presented in the natural ventilation case and mixed-mode case by Kapoor *et al.*[1,51].

The model was developed using 695 data sets. This study includes indoor temperature (T_{In}), indoor relative humidity (RH_{In}), occupancy (O), area per person (A_P), volume per person (V_P), CO_2 concentration (CO_2), air quality index (AQI), fan air speed (F_S), and one target variable, R-Event. The chance of at least two people being present in the workplace had a role in the selection of the data sets. Solitary occupancy readings were not taken into consideration in this investigation. Additionally, readings without air conditioning device turned on were not taken into account.The circumstances within the AC office building are also influenced by its surroundings. It is common knowledge that office buildings near busy roads, markets and densely populated areas are impacted by environmental pollution. As a result, AQI, an external environmental variable, is also considered in this study.

Since people are the main source of emissions in any indoor space, occupancy highly affects CO_2 concentration in occupied building spaces. The R-Event is shown in equation 1[52]. However, the coronavirus new cases probability calculator of REHVA Version 2.1 is used in this work to scientifically determine the R-Event. Equation 2 represents the R-Event mathematical form as per the REHVA calculator[53]. Where IP is infection probability, NS is number of susceptible person, and NI is number of infectitious person.

$$R - Event = \frac{kT}{t}(1 - e^{-\beta t}) \quad \text{... (1)}$$

where "T" signifies the total event duration, "β" signifies the transmission probability (constant regarding time), "t" signifies the duration that a single susceptible person made contact with a sick person, "$1\text{-}e^{-\beta t}$" signifies the chances that any susceptible person will become sick, and "k" signifies the total person who made contact with the sick person.

$$R - Event = \frac{IP \times NS}{NI} \quad \text{... (2)}$$

Figure 1 shows the conceptual framework applied for this investigation. In order to anticipate SC-2 transmission in an AC office setting utilizing indoor CO_2 as a proxy indicator, a unique ANN-based model must first be constructed. Data collection, normalization, and splitting for training/ testing are the main processes in the database preparation process.

2.1 Data Collection

Using a CO_2 analyzer and a heat stress meter, the data for this investigation was gathered in Roorkee, India. The readings were taken in April and May in the year 2022. The coordinates for the test site are EL 77°54'10" and NL 29°51'54". The details of the office room were presented in previous research [1,27,28,51]. The main entrance door was also a static component in this investigation and remained closed. The second entrance, a side door (single door), was located on the east side wall. One ceiling fan is positioned in the center of the room. A single window air conditioner was installed at a height of 0.3 meters above the ground level, underneath the fixed aluminium-glass window on the south side wall.

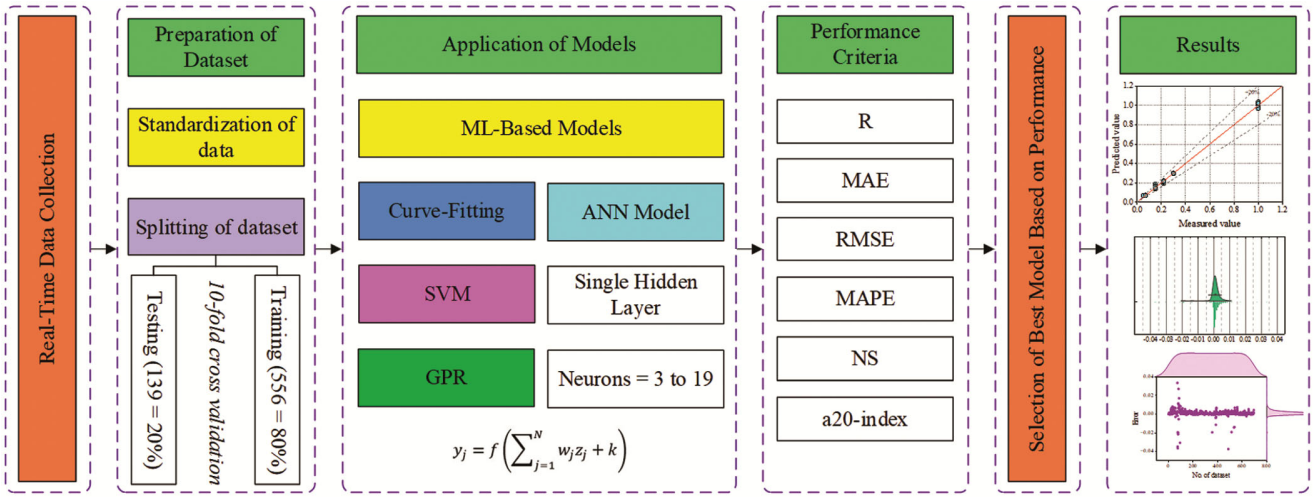

Fig. 1 — Conceptual framework.

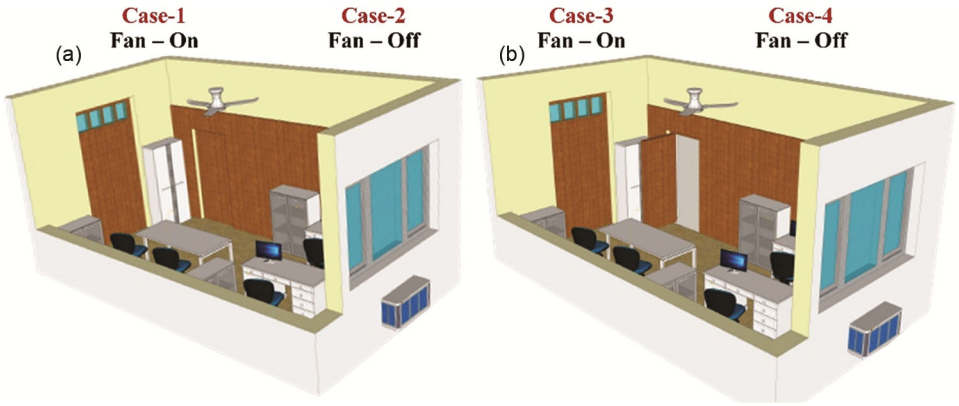

Fig. 2 — AC office room operational cases 3-D illustration.

Table 1 — Details of office room components which are altered for AC case.

Component	Front Door	Side Door	Side Window-1	Side Window-2	Fan (Ceiling)	Air-Conditioner
Area/Volume	3.0 m^2	1.8 m^2	0.6 m^2	0.6 m^2	-	-
Fixed/Variable	Fixed (Closed)	Variable	Fixed (Closed)	Fixed (Closed)	Variable	Fixed (On)

Table 1 provides details of the office room components which are changed for AC case[1,51].

The study goal was kept under wraps from office staff during monitoring since the Hawthorne effect[54] would negatively affect the participants' (subjects) usual behavior. It is possible that individuals' behavior altered when they realized they were being observed. They also changed their movements, breathing patterns, and behaviors if they were aware of the study's goals during CO_2 monitoring.

All of the individuals in this study were informed about the study's final goal after the data was collected, and their signed consent was acquired for further utilizing the data obtained for analysis and publication purposes. All of the static participants were seated longitudinally in a "L" configuration, about one meter apart. In Kapoor *et al.*[1], the specifics of the research setting and subjects are detailed. The monitored office space is shown in three dimensions in Figure 2. On the functioning of the ceiling fan, and the operability of side doors are predicated. Data was gathered at intervals of ten minutes forty-four times daily from Monday through Friday.

Figure 3 and 4 show the selected dataset heat plot matrix and distribution of input and output parameters respectively.

The data for eight input parameters were collected as presented in Table 2. The office room CO_2 concentration

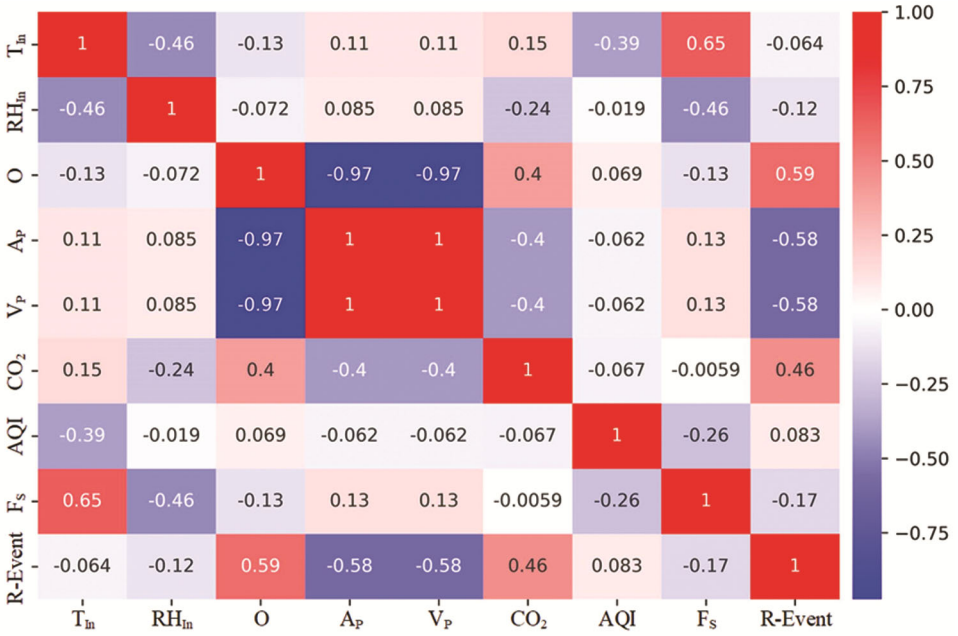

Fig. 3 — Heat plot matrix.

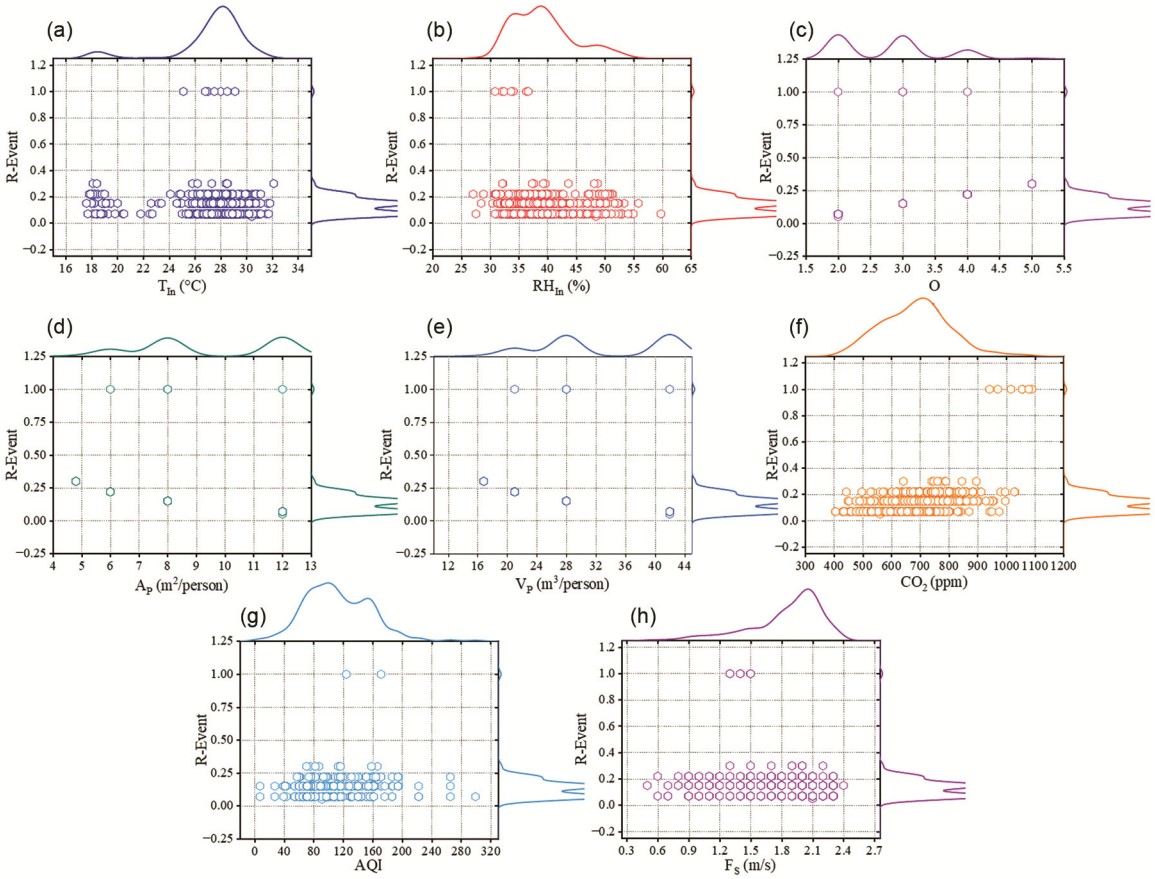

Fig. 4 — Distribution of input and output parameters.

Table 2 — Parameters considered in developing model for AC office and their statistical analysis.

Parameters	Symbol	Unit	Min.	Mean	Max.	Std.	Skewness	Type
Indoor Temp.	T_{In}	°C	17.60	27.34	32.10	2.85	-2.05	Input
Indoor RH	RH_{In}	%	27	39.14	59.70	5.46	0.76	
Occupants	O	Nos.	2	2.76	5	0.76	0.61	
Area per person	A_P	m^2/person	4.80	9.35	12	2.40	-0.01	
Volume per person	V_P	m^3/person	16.80	32.71	42	8.39	-0.01	
CO_2 Level Inside	CO_2	ppm	404	682.89	1089	118.12	0.27	
AQI	AQI	-	7	112.99	299	44.73	0.61	
Fan Speed	F_S	m/s	0.50	1.85	2.40	0.36	-1.27	
R-Event	*R-Event*	-	0.05	0.14	1	0.10	5.72	Output

levels range from 404 to 1089 ppm, and there will be 2 to 5 occupants during the research, according to the input data. The 0.05 to 1.00 range covers the R-Event output parameter. All of the feature parameters as well as the target withtheir lowest, maximum and mean values are presented in the table. It also includes information on skewness, kurtosis, and standard deviation. The indoor RH mean was 39.14 percent, and the average interior temperature was 27.34 degrees Celsius. The mean occupancy was 2.76 for the AC office room, 9.35 m^2 and 32.71 m^3 were the average area and volume per person, respectively. The average CO_2 concentration in the office room during the measuring period was 682.89 ppm. The mean AQI was 112.99. The average fan speeds were 1.85 m/s. The mean R-Event is 0.137.

2.2 Standardization of Selected Data

The chosen database for this investigation was standardized following the dataset collection. Standardization is a step in defining the measurements into specific ranges, like 0 to 1, -1 to +1, 0 to +0.9, or 0 to +0.8 [1, 27, 28, 55]. Values for the selected input parameters might vary within a predetermined range. Standardization makes the computations faster. The datasets were standardized in the range of 0 to +0.8 using Equation 3.

$$N_{standardized} = 0.8 \times \frac{N - N_{min}}{N_{max} - N_{min}} \quad \text{... (3)}$$

where N is the value from the database of interest, N_{min} is the parameter's smallest value, N_{max} is its highest value, and $N_{standardized}$ is the value that has to be standardized.

2.3 Forecasting R-Event

Four methods were applied to develop the correlation and prognosis of the R-Event. The first approach is the CF method. The following ML techniques applied were SVM and GPR. The last approach uses ANN, a method of AI. The R-Event as an output for the AC office room is predicted using ANN with eight inputs.

2.3.1 Curve-Fitting

Curve fitting is the process of finding a mathematical model that best describes a set of data points. The goal of curve fitting is to create an equation that can be used to approximate the values of a dependent variable based on the values of one or more independent variables. The process of curve fitting involves selecting a mathematical function that closely matches the data points, and then adjusting the parameters of the function to minimize the difference between the predicted values and the actual data points. This is typically done using a least squares approach, where the sum of the squared differences between the predicted and actual values is minimized. Curve fitting is commonly used in a variety of fields, including science, engineering, finance, and economics. It is often used to make predictions or extrapolate data beyond the range of the available data points.

First-order degree curve fitting, also known as linear regression, is a method of fitting a straight line to a set of data points. In this method, the relationship between the dependent variable and one independent variable is modeled using a linear function of the form:

$$y = mx + c \quad \text{... (4)}$$

where y is the dependent variable, x is the independent variable, c is the y-intercept of the line (the value of y when x=0), and m is the slope of the line (the change in y for a one-unit change in x).

The goal of first-order degree curve fitting is to estimate the values of the parameters m and c that best fit the data. Once the parameters are estimated, the fitted line can be used to make predictions or estimate the value of the dependent variable for a given value of the independent variable. Additionally, the goodness of

fit of the model can be assessed using measures such as the coefficient of determination (R-squared) or the residual standard error and other performance metrics. Based on the curve-fitting methodology, the obtained expression is shown in Equation 5.

$$\begin{aligned} \mathrm{R-Event} \;=\; & -0.0357\,T_{In} - 0.0989\,RH_{In} \\ & + 0.2487\,O \;-\; 0.0028\,A_P \\ & - 0.00003\,V_P \;+\; 0.1286\,CO_2 \\ & + 0.021\,AQI \;-\; 0.0675\,F_S \\ & + 0.0637 \end{aligned} \quad \ldots (5)$$

Let $N = 0.0637$

Now, for all the parameters the correlation coefficient (C_c) can be stated in computational form as:

$$\begin{aligned} C_c \;=\; & -0.0357\,T_{In} - 0.0989\,RH_{In} \;+\; 0.2487\,O \\ & - 0.0028\,A_P \;-\; 0.00003\,V_P \\ & + 0.1286\,CO_2 \;+\; 0.021\,AQI \\ & - 0.0675\,F_S \end{aligned} \quad \ldots (6)$$

Eventually, the equation becomes:

$$\mathrm{R-Event} \;=\; C_c \;+\; N \quad \ldots (7)$$

2.3.2 Artificial Neural Networks

The first ANNwas developed by psychologist Frank Rosenblatt in 1958. During the last four decades, ANN has made important contributions to the development of a number of sophisticated information-processing techniques[1]. The ANN modeling approach uses computers to simulate several fundamental properties of the human brain, like the ability to solve difficulties in new situations by employing knowledge from earlier experiences. The effectiveness of the model learning process is significantly influenced by the hidden layer neurons[55,56]. By comparing measured outputs with projected outcomes, a trained ANN can detect relationships among the data to produce the best results feasible.

The neural networks are able to develop models from complex datasets with multiple features. It improves the usability and accuracy of models [57]. The main goal of training is to reduce mistakes while raising the correlation coefficient's value (R = 1). The "MSE" as well as "R" shows the ANN performance. The equations of both the performance indices are presented below in equations 8 and 9.

$$R = \frac{\sum(x_i - \bar{x}) \times (y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \times \sum(y_i - \bar{y})^2}} \quad \ldots (8)$$

where, x_i denotes measured values, $\bar{x}$denotes measured values mean, y_i denotes predicted values, $\bar{y}$denotes predicted values mean.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2 \quad \ldots (9)$$

The practice of describing a phenomenon in a set of computer assertions is known as modeling. It is vital to determine the optimum network design, which provides high accuracy. The “R” as well as “MSE” were determined for each neuron. Figure 5 presents

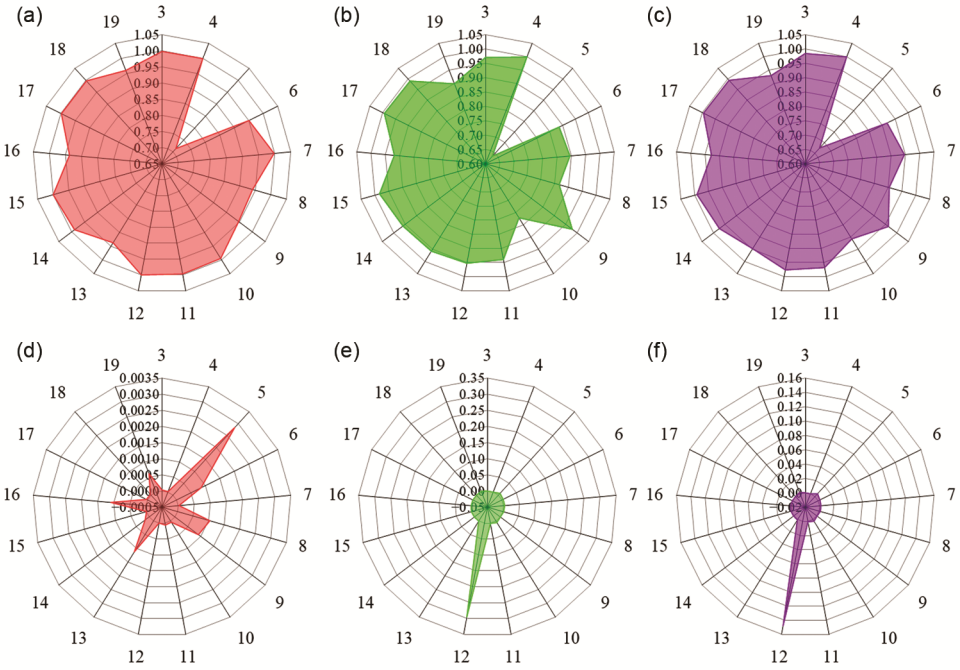

Fig. 5 — Selection of best neuron based on R and MSE value (a) R--training, (b) R-testing, (c) R-all, (d) MSE-training, (e) MSE-testing, and (f) MSE-all.

the performance of the network depending on the MSE. 3 to 19 neurons were employed to train the ANN model. The training correlation coefficient peaked at the fourth neuron with the value of 0.99925. The "MSE" and "R" of each neuron were taken into account while deciding its ranking. The neurons with greater "R" were ranked from higher value to lower value. However, the neurons with lower "MSE" were given a higher ranking. The best ranked neurons were determined by adding together the ranks of "MSE" as well as "R" of every neuron for training, and testing. Out of all the trials, the fourth neuron trial gets the lowest overall rating score, making it the best ranked neuron. One hidden layer with four neurons makes up the network that predicts each neuron's performance the best. The LM training method's superior convergence, high accuracy, and rapid network training was the reason for opting for the same in this work. This approach is speedier but frequently calls for more memory.

$$y = tangsig(x) = \frac{2}{(1+e^{-2x})} - 1 \quad \text{... (10)}$$

$$y = purlin(x) = x \quad \text{... (11)}$$

The "MSE" values at various neurons for training, as well as testing, are shown in Figure 5. For each trained neuron, the training, as well as testing "MSE" were recorded along with the "R". The highest "MSE" for training as well as is at the 5thneuron with the values of 0.002834 and 0.009308 serially. Whereas the highest testing "MSE" is at the 12th neuron with the value of 0.295579.

2.3.3 Support vector machines

Vladimir Vapnik devised the Support Vector Machine (SVM) paradigm in the 1990s, which is predicated on the foundations of Statistical Learning Theory. SVMs delineate a decision boundary by identifying the extreme boundaries and constructing the corresponding hyperplanes that effectively separate two distinct classes. Anomalies in decision boundaries can lead to erroneous classification of new datasets. The key principle of SVMs is the identification of support vectors, which correspond to extreme data points and assist in identifying the constraints of the decision boundary.

SVMs have gained widespread utility in classification tasks. The conventional approach to regression analysis involves estimating the function f(x) that minimizes the difference between empirical responses and predictions for all training datasets. However, SVMs prioritize achieving the shortest generic error bound over the minimal observed training error to obtain optimal performance. The regularization term, which constrains the complexity of the function space, is combined with the training error to yield the generalization error bound.

2.3.4 Gaussian process regression

Gaussian Process Regression (GPR) is a nonparametric Bayesian machine-learning technique employed for regression analysis. Carl Edward Rasmussen and Christopher K. I. Williams developed this method in the mid-1990s. GPR is based on the concept of representing a function as a Gaussian process, which is a set of random variables having a joint Gaussian distribution for any finite number of them. This distinctive feature permits the modeling of uncertainty in predictions, in contrast to conventional linear regression which only accounts for the mean value.

The formulation of GPR involves defining "a prior over functions" that is then modified based on observed data to generate a posterior distribution over functions. This posterior distribution is subsequently utilized to create predictions by determining the expected value of the function at all new input points and the prediction's uncertainty. The forecast is represented as a Gaussian distribution with a mean and covariance that captures the level of uncertainty in the prediction.

In summary, GPR is a robust technique for approximating functions and quantifying uncertainty, with applications in various fields such as geostatistics, civil engineering, time series analysis, and robotics.

2.3.5 Performance criteria

For assessing the precision of ML as well as CF models, performance metrics including R, MAE, MAPE, MSE, RMSE, NS, and a20-index are taken into consideration to establish the models' predictability[58-61]. The equations for the same are given in the formulae equations 12 to 16 below. To gauge a model's correctness, higher values (near to 1) of the NS index as well as R are required. Equations 8 and 9, which stand for the "R" and "MSE" equations, were previously stated in section 2.3.2.

$$MAE = \frac{1}{N}\sum_{I=1}^{N}|x_i - y_i| \quad \text{... (12)}$$

$$MAPE = \frac{1}{N}\sum_{I=1}^{N}\left|\frac{x_i - y_i}{x_i}\right| \times 100 \quad \text{... (13)}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(x_i - y_i)^2}{N}} \quad \text{... (14)}$$

$$NS = 1 - \frac{\sum_{i=1}^{N}(x_i - y_i)^2}{\sum_{i=1}^{N}(x_i - \overline{y_i})^2} \quad \text{... (15)}$$

$$a20 - index = \frac{m20}{N} \quad \text{... (16)}$$

where N is the total experimental dataset number, $x_i$and$y_i$are the measured andanticipated values at i[th] level, $\overline{y_i}$ is the mean value of the anticipated results.

3 Results and Discussion

In this section, an analysis has been conducted on the outcomes of all the models that have been developed. The performance indices, pertaining to each model, have been tabulated and presented in Table 3. The R-value of CF, SVM, GPR and ANN model for the complete dataset is 0.6462, 0.8182, 0.9936 and 0.9992, respectively.

In terms of predicting R-Event, it has been observed that the ANN model exhibits superior performance as compared to the CF model. The correlation coefficient of the ANN model is 35.33% higher than that of the CF model. Additionally, the performance indices of the CF model, including MAE, MAPE, RMSE, NS, and a20-index, have respective values of 0.0311, 22.1321, 0.0802, 0.4084, and 0.6738. Conversely, the performance indices of the ANN model are 0.0017, 1.0575, 0.0042, 0.9984, and 0.9986, respectively.

Furthermore, it has been observed that the MAE, MAPE, and RMSE values of the CF model are 94.53%, 95.22%, and 94.76% higher than those of the ANN model, respectively. In terms of the a20-index and NS performance factors, the ANN model has demonstrated superior performance, exhibiting values that are 32.53% and 59.09% greater than those of the CF model, respectively. Hence, based on the performance indices it can be concluded that the ANN shows stronger precision as compared to other CF and ML models.

A scatter plot has been generated to illustrate the relationship between the measured R-Event and predicted R-Event values. As depicted in Figure 6, it can be observed that only 36.83% of the dataset is situated on the redline, which represents the model fit. Additionally, a marginal box-plot has been constructed to visualize the number of datasets and their corresponding error values. The marginal plot confirms that only eight data points lie outside the error range values of 0.0743. The range of error is confined between -0.09141 to 0.83663.

Figure 7 shows the scatter and marginal plot of the training, testing and complete dataset. In Figure 7a, the first graph shows the scatter plot in which the R-Event values of measurement and prediction have been presented. 53.60% of the dataset is directly covering the fitting line of the model as shown in red color in Figure 7a (leftside). The error range is

Table 3 — Performance comparison of CF, GPR, ANN and SVM models.

Performance Indicator	Data Type	ANN	GPR	SVM	CF
R	Training	0.9992	1	0.8304	0.6462
	Testing	0.9994	0.9752	0.7855	
	All	0.9992	0.9936	0.8182	
MAE	Training	0.0017	0.00004	0.0102	0.0311
	Testing	0.0017	0.0066	0.0143	
	All	0.0017	0.0014	0.0109	
MAPE	Training	1.0695	0.0272	4.7271	22.1320
	Testing	1.0094	4.2581	5.9515	
	All	1.0575	0.8733	4.9719	
MSE	Training	0.00002	0	0.0035	0.0064
	Testing	0.00002	0.0007	0.0062	
	All	0.00002	0.0001	0.0041	
RMSE	Training	0.0045	0	0.0592	0.0802
	Testing	0.0045	0.0265	0.0787	
	All	0.0045	0.0118	0.0639	
NS	Training	0.9983	0.9999	0.6473	0.4084
	Testing	0.9986	0.9508	0.5599	
	All	0.9984	0.9871	0.6246	
a-20 Index	Training	0.9982	1	0.9802	0.6738
	Testing	1	0.9568	0.9496	
	All	0.9986	0.9914	0.9741	

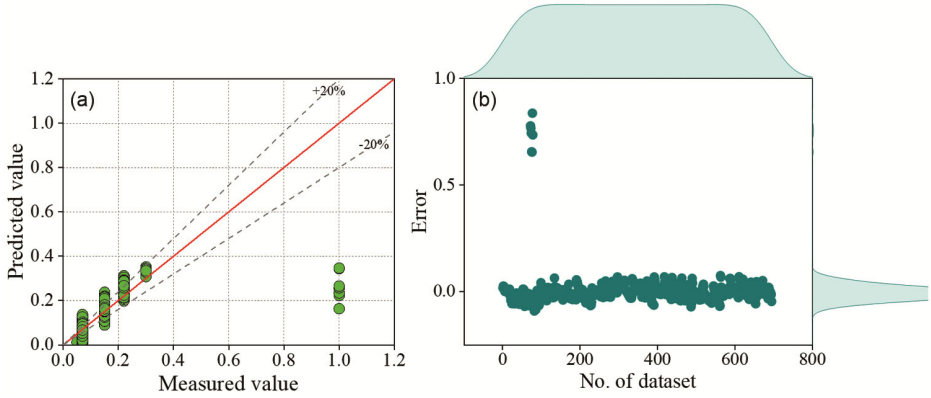

Fig. 6 — Results of curve fitting model (a) scatter plot, (b) marginal plot.

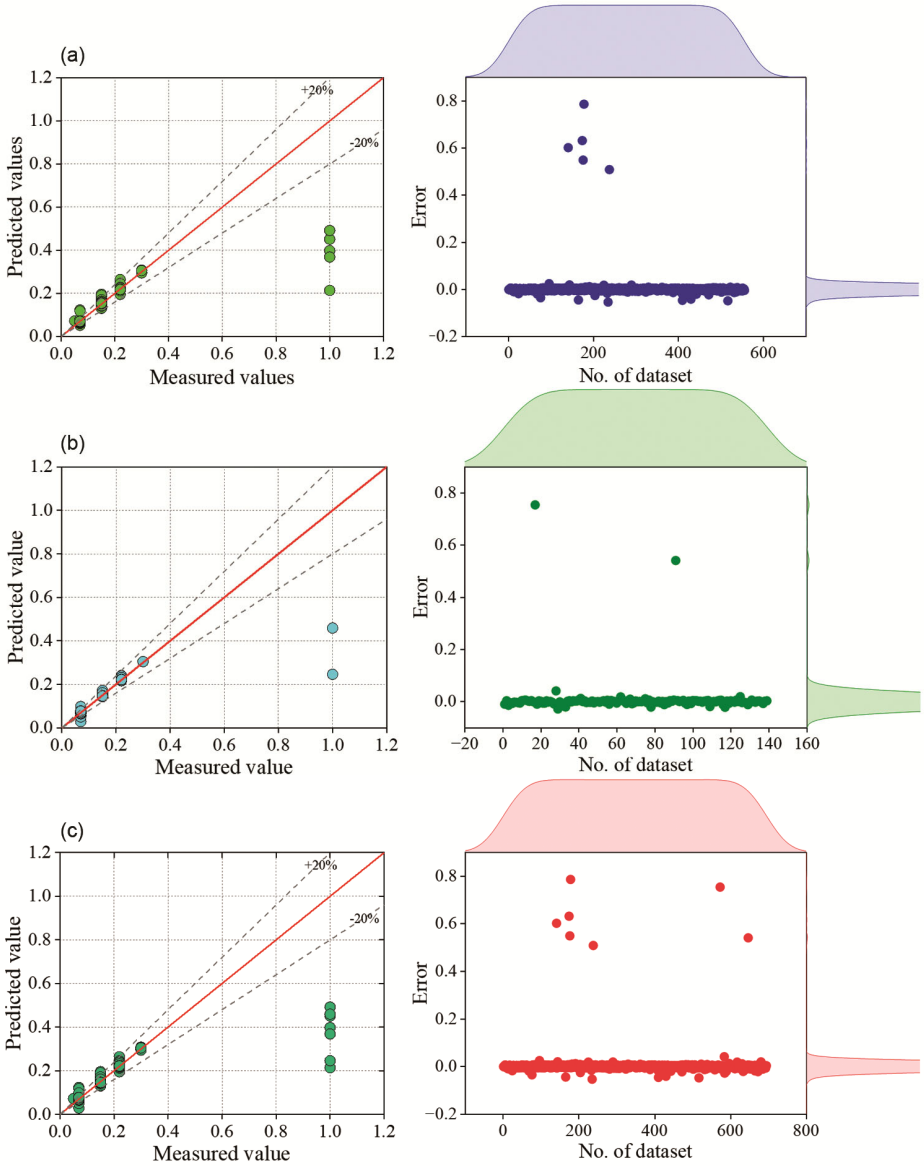

Fig. 7 — Results of SVM model (a) training dataset, (b) testing dataset, and (c) complete dataset.

between -0.0533 to 0.7859 as depicted by the marginal plot in Fig. 7a (right-side), and only 5 values exceed the error range value of 0.0249.

Similarly, Fig. 7b is drawn for testing the dataset for the SVM model. The scatter plot shown in Fig. 7b (left-side) presents that 53.24% of the dataset is lying over the redline showing the fitting of the model. The error range is between -0.02816 to 0.75381 as presented in Fig. 7b (right-side) and the marginal plot shows that only 2 values are above the error range value of 0.04114.

Fig. 7c shows the fitting of the SVM model for the complete dataset. In Fig. 7c, the first part is a scatter plot, which shows that 53.53% of the dataset is directly covering the fitting line of the model. The range of the error is lying between -0.0533 to 0.7859 as shown in Fig. 7c (right-side). It is confirmed by the marginal plot that only seven values exceed the error range value of 0.0411.

The performance of GPR and ANN models are shown in Figs 8 and 9, respectively. Figure 8 shows the fitting of the GPR model for the training dataset. In Fig. 8, the first plot shows the scatter plot in which the R-Event value of measurement and predictions has been presented and 50% of the data is covering the fitting line of the model as shown in Fig. 8a (left-side). The error is between -0.0001 to 0.0009 as presented in the second plot of Fig. 8a (right-side).

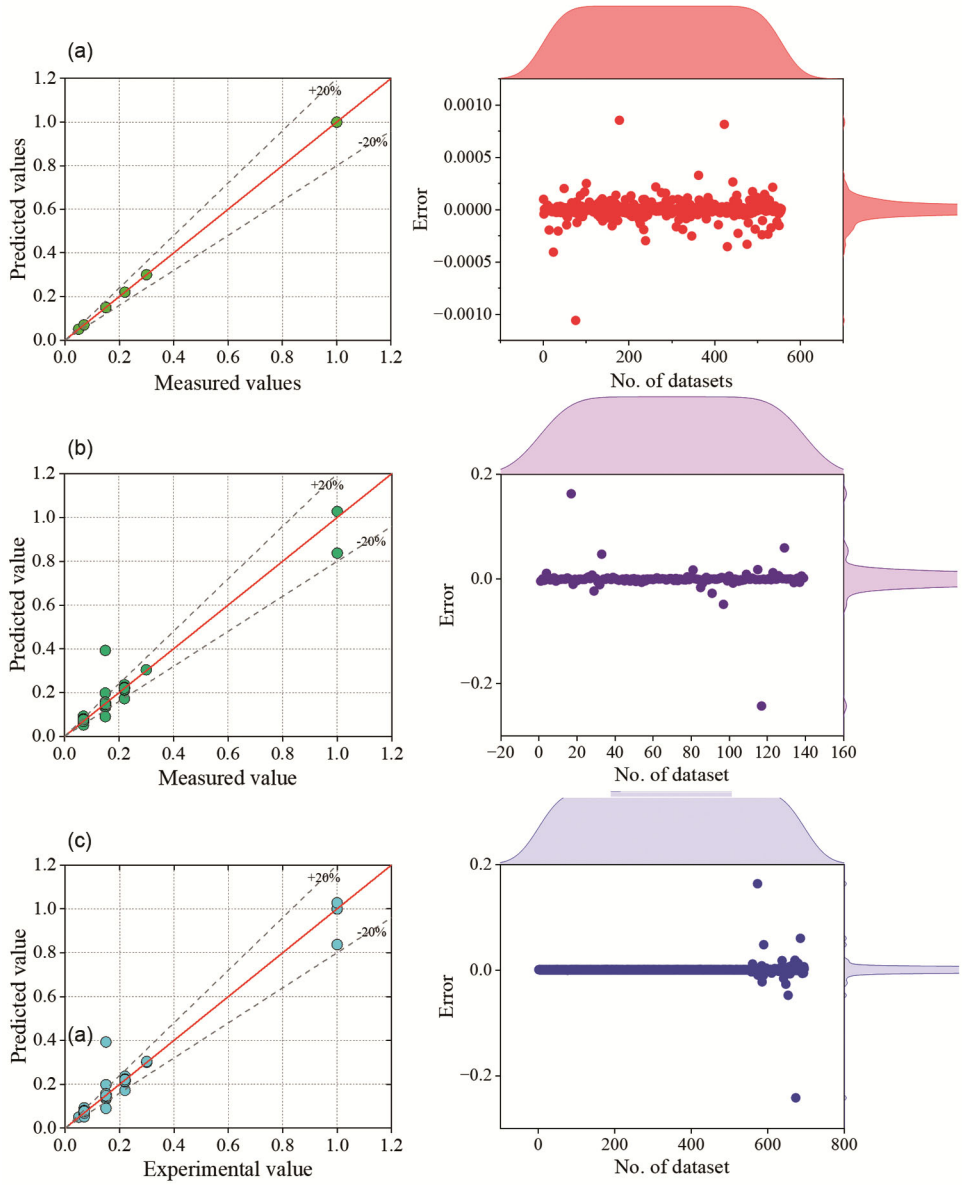

Fig. 8 — Results of GPR model (a) training dataset, (b) testing dataset, and (c) complete dataset.

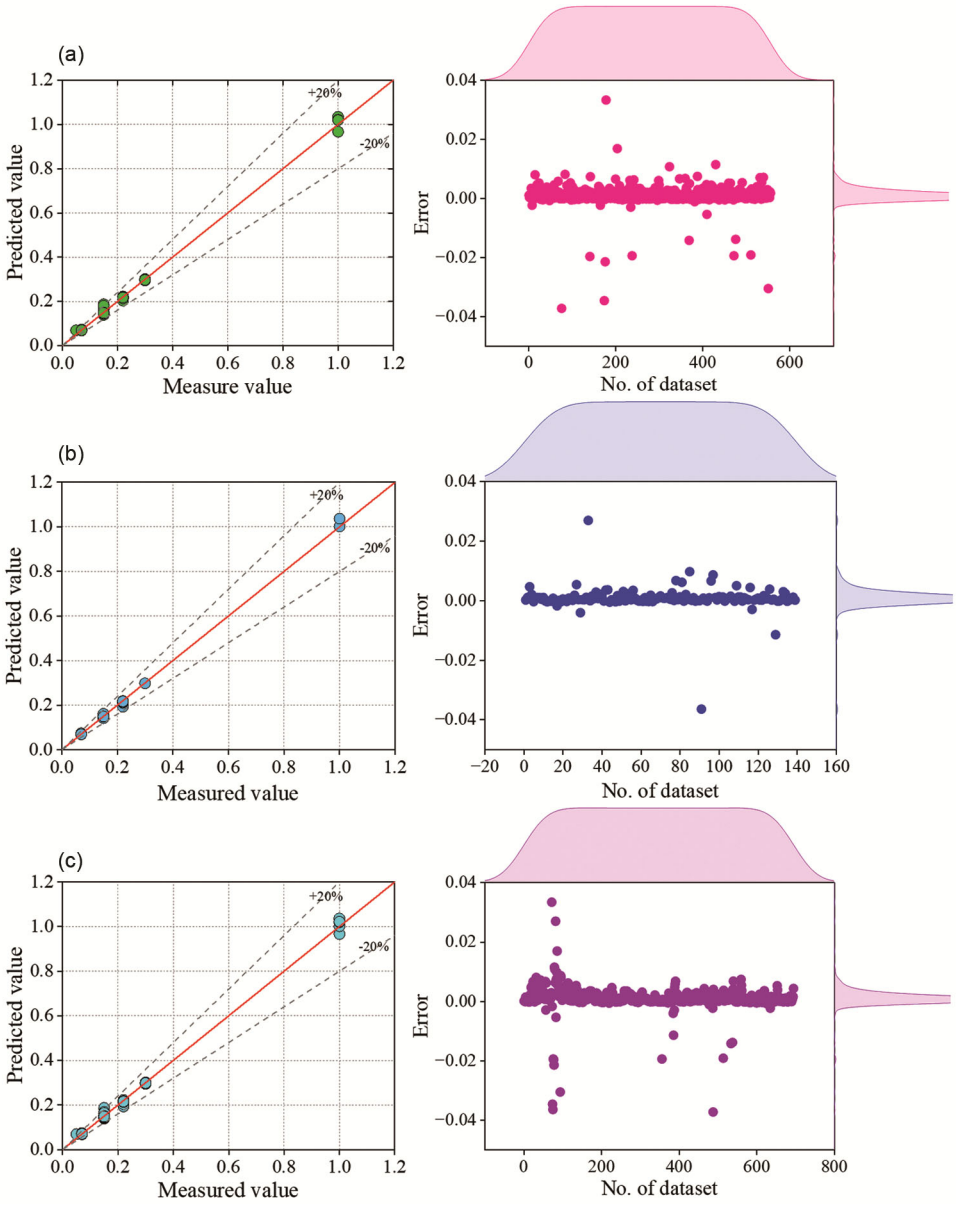

Fig. 9 — Results of ANN model (a) training dataset, (b) testing dataset, and (c) complete dataset.

The third plot shows that no value exceeds the error range value of 0.0009.

The scatter plot in Fig. 8b (left-side) presents that only 43.17% data of the GPR testing test is nearly aligned to the fitting line of the model. The error range plot shows the limit between -0.2427 to 0.1630 for the GPR testing dataset. Only one value exceeds the error range of 0.0596 as shown in Figure 8b (right-side).

Figure 8c shows the scatter plot, error range plot, and marginal plot of the complete dataset used for the GPR model. The 48.63% of data is aligned with the fitting line drawn between the measured R-Event and predicted R-Event as shown in Fig. 8c (left-side). Only one value is exceeding the error range of 0.0596 in the marginal plot and the error range is between -0.2427 to 0.1360 as shown in Figure 8c (right-side).

In Figure 9, there are three sections in each figure, where the first section of all the ANN figures contains 76.08%, 74.82% and 75.83% dataset on the fitting line for each training, testing and complete dataset, sequentially. The error range of the training, testing and complete dataset are -0.0373 to 0.0333, -0.0365 to 0.0269, and -0.0373 to 0.0333, respectively. For all three ANN datasets, no value exceeds the maximum error range.

The Taylor plot, also known as the Taylor diagram or Taylor chart, is a graphical tool that is commonly used to compare the performance of multiple climate models against a reference dataset or observation. It was first introduced by Karl Taylor in 2001[62,63].The plot consists of a polar coordinate system, where the distance from the origin represents the normalized standard deviation of the model compared to the reference dataset, and the angle represents the correlation coefficient between the model and the reference dataset. The reference dataset is usually represented as a point at the origin. The models are plotted as points on the diagram, with their distance and angle relative to the reference dataset indicating their performance compared to the reference. Models that have high correlation coefficients and low normalized standard deviations are considered to be better performers than those that have lower correlation coefficients and higher normalized standard deviations. In summary, the Taylor plot is a useful tool for visually comparing the performance of multiple models against a reference dataset or observation, allowing researchers to identify which models perform best and worst and to gain insights into the strengths and weaknesses of each model. The fitting of the developed has been shown in Fig. 10. In Fig. 10, the GPR and ANN models directly lie over the dotted blue line that represents the standard deviation. However, the CF and SVM models are far away from the standard deviation line of the measured R-Event. The Taylor plot also confirms the accuracy of the ANN model.

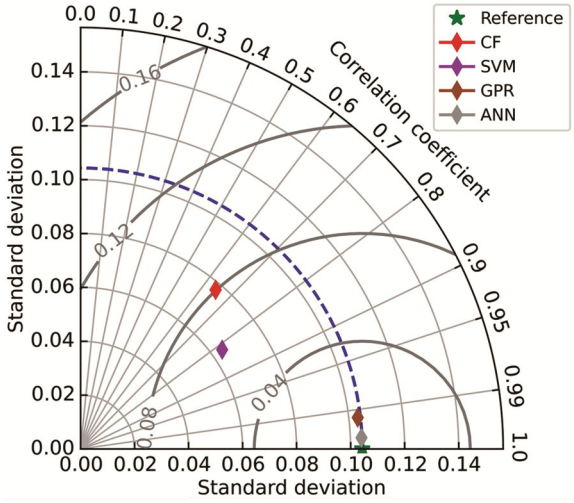

Fig. 10 — Taylor plot of the developed ML models.

3.1 Proposed ANN Formulation to estimate the R-Event

ANN model demonstrated excellent accuracy for prediction. The ANN model is presented to workout R-Event. The final equation for the prediction purpose is Equation 17:

$$R - Event = purlin(W_{HO}Y_i + B_{HO}) = W_{HO}Y_i + B_{HO} \quad \text{... (17)}$$

Equation 17 shows the general form for input to hidden layer Yi.

$$Y_i = W_{IH}N_{i,normalized} + B_{IH} \quad \text{... (18)}$$

where W_{IH} and B_{IH} denote matrix weight and biases between the input-to-hidden layer, $W_{HO}$and $B_{HO}$are the weights and biases of the hidden-to-output layer, $N_{i,normalized}$ denotes normalized inputs.

The proposed model comprises eight input parameters. Equation 20 may be used to determine $Y_i$value. Equation 19 expresses the final R-Event equation.

$$R - Event = -0.748020\,P_1 - 1.045879\,P_2 + 0.329071\,P_3 + 0.082082\,P_4 - 0.306008 \quad \text{... (19)}$$

Some other general observations were obtained over the course of the research. Increased indoor activity raises the risk of viral spread and has an immediate impact on CO_2 levels. The volume and size of the workplace are key components for a safe working environment. The primary variable that has a strong link with R-Event is occupancy. As a result, occupancy levels must be reduced to a safe level for a secure workplace. In order to maintain low CO_2 levels inside buildings and prevent the spread of SC-2,

$$\begin{bmatrix}P_1\\P_2\\P_3\\P_4\end{bmatrix} = tansig \begin{bmatrix}-0.202944 & -0.660461 & 0.019195 & -0.143245 & -0.094812 & 1.801014 & -0.572412 & -1.110980\\ -2.196323 & 2.533653 & 7.018275 & 1.570779 & 2.164265 & -12.649054 & 7.401999 & 8.477463\\ -0.014979 & -0.014443 & -0.053712 & -0.164204 & -0.622859 & 0.022504 & -0.017725 & -0.004253\\ 0.142843 & 0.121390 & 1.368264 & -0.951628 & -0.729442 & 0.237410 & -0.066099 & -0.060606\end{bmatrix} \times \begin{bmatrix}T_{In}\\RH_{In}\\O\\A_P\\V_P\\CO_2\\AQI\\F_S\end{bmatrix} + \begin{bmatrix}-2.956664\\13.370908\\-0.430535\\-2.601925\end{bmatrix} \quad \text{... (20)}$$

ventilation is essential. Significant correlations between CO_2 levels and indoor airborne transmissions show the necessity for enhanced ventilation in air-conditioned situations. It is suggested to keep up with all ventilation systems for better operational management. The performance of full-scale operation is improved by effective maintenance. In the workplace, continuous talking must be avoided since it increases the danger of an infection spreading when it takes place without a mask. By using sensors, maintaining hygiene, properly utilizing personal safety equipment, providing instruction, and enforcing the law in buildings, the threat of viral transmission can be significantly reduced.

3.2 Limitations and Scope of Future Research

This study has a number of constraints. The focus of this study is only on the interior spaces of a single-zone air-conditioned office room. A solitary infection source is taken into consideration with pathogen proliferation occurring entirely through airborne transmission channels in an office setting functioning under the composite climatic environment. This work is constrained by a variety of financial, social, technical, and other factors. The method is uniform, though, and may be applied to a wide range of structures in different climatic zones with other environmental circumstances and parameters. The direction as well as air flow patterns were not considered in this experiment. They could, however, significantly affect how airborne viruses travel, which limits the study's potential application. Only diurnal environmental changes are considered since office hours hinder the monitoring of night oscillations. Aged between 21 and 30, the participants were healthy individuals without any breathing problems. The information obtained is case-based as well as dependent on a number of additional environmental conditions that are not covered in this study, such as pressure, several other pollutants present indoors and outdoors, and wind direction. This study takes into account subjects using the workplace space, not wearing a face mask, and is therefore susceptible to SC-2. For purposes of prediction, this study takes into account the SC-2 DELTA strain, which is twice as contagious as the initial viral strain of SC-2. Extremely inconsistent behavior is seen by both recent and historic SC-2 versions. This study does not address the relationship between viruses and people. Every individual has a distinctive behavior depending on their brain-prints as well as experiences, the broad variation in human behavior under various conditions is in and of itself a constraint. This technique has additional restrictions because the model can only forecast outcomes within the bounds of the dataset. Other respiratory actions like coughing, sneezing, etc. were just not deliberated in this study either. Although the mobility of individuals outside of offices, close to doors and windows, is another crucial feature, this research does not take it into account because of the very dynamic surroundings. Other case-based adjustments are limits of the created model, which estimates the average rate of event reproduction for the office space with diurnal fluctuations. Close touch and the spread of fomite were not taken into account. Non-uniform concentration is another constraint of the created model. The created model is a single-zone model. Many of the parameters, which were obtained from the literature or extrapolated using the best information currently available, are imprecise. Future research will address some of these shortcomings. Although more complicated as well as realistic models can be created, the parametric uncertainty may still predominate in the overall uncertainty. To include fresh knowledge, futuristic models might take parameters depending on the results of recent studies. For more “features” i.e. inputs and more “feature sets”, the “Target”, i.e. output (here R-Event) can be more accurate with lower cost-function value. This will result in better predictions of the transmission prognosis of SC-2 in any particular built environment the model is developed for. Parametric (i.e. feature) uncertainty and alterations in viral behavior will affect the real-time overall uncertainty.

4 Conclusion

The AI-based modeling process is typically a difficult method. The four models presented in this paper are an analytical CF model, an SVM model, a GPR model, and an ANN model. These prediction models were developed and put to the test to predict the R-Event in an AC office environment. In an AC office space that was located in a composite climate of India, real-time monitoring of the office environment was done in the months of April as well as May of 2022. Eight input parameters were used for developing the models. For particular occurrences, R-Event is the target showing the chances that SARS-CoV-2 would spread indoors. The event specific transmission (R-Event) prognosis will help in COVID-19 prophylaxis in air-conditioned office rooms and will definitely nudge public safety

and health as an outcome. The data were statistically examined before being subjected to CF, SVM, and GPR and then ANN training, and testing, The ANN model outperformed all the developed models in terms of accuracy in forecasting the R-Event. With observed findings, ANN displayed a greater correlation as well as lower error (R=0.9992, RMSE=0.0042, MSE=0.0000, MAPE=1.0575, MAE=0.0017, NS=0.9984, and a20-index=0.9986). The models created in this work are helpful in forecasting the R-Event in enclosed air-conditioned spaces. As a long-term outcome, this study and related types of research may be able to save lives as well as time, money, and efforts.

References

1 Kapoor N R, Kumar A, Kumar A, Kumar A, & Kumar K, *IEEE Access*, 10 (2022) 121204.

2 WHO Coronavirus (COVID-19) Dashboard. Available online: https://covid19.who.int/

3 Crook H, Raza S, Nowell J, Young M, & Edison P, *BMJ*, 374 (2021) 1648.

4 Raveendran A V, Jayadevan R, & Sashidharan S, *Clin Res Rev*, 15(3) (2021) 869.

5 Agarwal N, Meena C S, Raj B P, Saini L, Kumar A, Gopalakrishnan N, Kumar A, Balam N B, Alam T, Kapoor N R, & Aggarwal V, *Sustain Cities Soc*, 70 (2021) 102942. https://doi.org/10.1016/j.scs.2021.102942

6 Kumar A, Kapoor NR, Kumar A, Deep A, Kulkarni K S, & Arora H C, *Proceeding ofi-Converge 2022: Changing Dimensions of the Built Environment,* (2022).

7 Frieden T R & Lee C T, *EmergInfect Dis,* 26(6) (2020) 1059.

8 Althouse BM, Wenger EA, Miller JC, Scarpino SV, Allard A, Hébert-Dufresne L, & Hu H, *PLoS Biol,* 18(11) (2020) e3000897.

9 Lewis D, *Nature*, 590(7847) (2021) 544.

10 Kapoor NR, Kumar A, Alam T, Kumar A, Kulkarni KS, & Blecich, P, *Sustainability,*13 (2021) 11855.

11 Kapoor N R, Kumar A, Meena C S, Kumar A, Alam T, Balam N B, & Ghosh, A, *Adv Civ Eng*, 2021 (2021) 8851685.

12 Ismail SA, Saliba V, Bernal J L, Ramsay ME, & Ladhani SN, *Lancet Infect Dis,* 21(3) (2021) 344.

13 Kapoor NR, &Tegar JP, *Int Res J Eng Technol,* 5(7) (2018) 1744.

14 Raj N, Kumar A, Kumar A, Goyal S, *Proceedings of the International Conference on Building Energy Demand Reduction in Global South (BUILDER'19),* (2019) 1–9. Available online: https://nzeb.in/event/builder19/

15 Kumar A, Kapoor NR, Arora HC, & Kumar A, *CRC Press: Boca Raton, Florida, USA,*9781003287186 (2022).

16 Greenhalgh T, Jimenez JL, Prather KA, Tufekci Z, Fisman D, & Schooley R, *The Lancet*, 397(10285) (2021) 1603.

17 Guo M, Xu P, Xiao T, He R, Dai M, & Miller SL, *Build Environ*, 187 (2021) 107368.

18 Zheng W, Hu J, Wang Z, Li J, Fu Z, Li H, Jurasz J, Chou SK, & Yan J, *Adv Appl Energy*, 3 (2021) 100040.

19 Somsen GA, van Rijn C, Kooij S, Bem RA, & Bonn D, *Lancet Respir Med*, 8(7) (2020) 658. https://doi.org/10.1016/S2213-2600(20)30245-9

20 Kapoor NR, Kumar A, Kumar A, Kumar A, Mohammed MA, Kumar K, Kadry S, & Lim S, *Wirel Commun Mob Comput*, 2022 (2022) 9404807.

21 Rudnick SN & Milton DK, *Indoor Air*, 13(3) (2003) 237.

22 Peng Z & Jimenez JL, *Environ Sci Technol Lett,* 8(5) (2021) 392.

23 Bazant MZ & Bush JWM, *ProcNatl Acad Sci USA,*118(17) (2021) e2018995118.

24 Bazant MZ, Kodio O, Cohen AE, Khan K, Gu Z, & Bush JW, *Flow*, (2021) 1:E10.

25 Wang CC, Prather KA, Sznitman J, Jimenez JL, Lakdawala SS, Tufekci Z, & Marr LC, *Science*, 373 (2021) 6558.

26 Martinez I, Bruse JL, Florez-Tapia AM, Viles E, & Olaizola IZ, *Build Environ*, 207 (2022) 108495.

27 Kapoor NR, Kumar A, & Kumar A, *Proceedings ofi-Converge 2022: Changing Dimensions of the Built Environment,* (2022).

28 Kapoor NR, Kumar A, Kumar A, *Proceedings of International Conference on Information Systems and Management Science (ISMS 2022),* (2022).

29 Mecenas P, Bastos RTDRM, Vallinoto ACR, & Normando D, *PLoS one*, 15(9) (2020) e0238339.

30 Kapoor NR, Kumar A, Kumar A, Arora HC, Kumar A, & Gaur S, *Sustainability*, 16(2) (2024) 516.

31 CSIR Guidelines on Ventilation of Residential and Office Buildings for SARS-Cov-2 Virus, (Version 2.0) (2022). Available online:https://www.csir.res.in/sites/default/files/2023-07/CSIR%20Ventilation%20Guidelines%20Version%202.0%20%28Dr.%20Ashok%20Kumar%2C%20CSIR-CBRI%29%20-FINAL.pdf

32 PHO, COVID-19: Heating, Ventilation and Air Conditioning (HVAC) Systems in Buildings, (2020).

33 REHVA, REHVA COVID-19 guidance document, (2020).

34 CCIAQ, Addressing COVID-19 in Buildings,(2020).

35 NHC, Hygienic Specifications for Operation and Management of Airconditioning Ventilation Systems in Office Buildings and Public Places during COVID-19 Epidemic, (2020).

36 ECDC, Heating, ventilation and air-conditioning systems in the context of COVID-19, (2020).

37 ASHRAE, ASHRAE Position Document on Infectious Aerosols, (2020).

38 ASHRAE, ASHRAE Issues Statements on Relationship Between COVID-19 and HVAC in Buildings,(2020).

39 ISHRAE, Start up and Operation of Air conditioning and Ventilation systems during Pandemic in Commercial and Industrial Workspaces,(2020).

40 SHASE, Q and A on Ventilation in the Control of SARS-CoV-2 Infection,(2020).

41 ASC, Guidelines for office buildings to deal with "new coronavirus" operational management emergency measures, (2020).

42 Franceschini PB, & Neves LO, *Energy Build,*258(2022) 111831.

43 Fan M, Fu Z, Wang J, Wang Z, Suo H, Kong X, & Li H, *Build Environ*, 212 (2022) 108831.

44 COVID-19 Employer Information for Office Buildings. Available online: https://stacks.cdc.gov/view/cdc/88411

45 Employers and workers, World Health Organization. Available online: https://www.who.int/teams/risk-communication/employers-and-workers?gclid=CjwKCAjwiuuRBhBvEiwAFXKaNLFawGyMm3ocZoKOlTK9OIgXKDv3CNd4ThLIYMYZnpEBorvy0GdIrRoCS1YQAvD_BwE

46 Zhu C, Maharajan K, Liu K, & Zhang Y, *Environ Res*, 198 (2021) 111281.

47 Santurtún A, Colom ML, Fdez-Arroyabe P, Del Real A, Fernández-Olmo I, & Zarrabeitia MT, *Environ Res,* 206 (2022) 112261.

48 Li Z, Tao B, Hu Z, Yi Y, & Wang J, *J Infect*, 84(5) (2022) 684.

49 Rzymski P, Poniedziałek B, Rosińska J, Rogalska M, Zarębska-Michaluk D, Rorat M, Moniuszko-Malinowska A, Lorenc B, Kozielewicz D, Piekarska A, &Flisiak R, *Environ Pollut*, 306 (2022) 119469.

50 Piscitelli P, Miani A, Setti L, De Gennaro G, Rodo X, Artinano B, Vara E, Rancan L, Arias J, Passarini F, & Domingo JL, *Environ Res,* 211 (2022) 113038.

51 Kapoor NR, Kumar A, Kumar A, Zebari DA, Kumar K, Mohammad MA, Al-Waisy AS, Albahar MA, *Int J Environ Res Public Health*, 19(24) (2022) 16862.

52 Tupper P, Boury H, Yerlanov M, & Colijn C, *Proc Natl Acad Sci USA*, 117(50) (2020) 32038.

53 REHVA Calculator. Available online: https://www.rehva.eu/covid19-ventilation-calculator

54 McCambridge J, Witton J, & Elbourne DR, *J Clin Epidemiol,* 67(3) (2014) 267.

55 Suri RS, Dubey V, Kapoor NR, Kumar A, & Bhushan M, *Proceedings of the International Conference on Information Systems and Management Science (ISMS 2021)*, (2023).

56 Lin W & Chen G, *IEEE Trans Neural Netw*, 20 (2009) 1340.

57 Abiodun OI, Jantan A, Omolara AE, Dada KV, Mohamed NA, & Arshad H, *Heliyon*, 4 (2018) e00938.

58 Kumar A, Arora HC, Kapoor NR, & Kumar K, *Struct Concr*, 24(3) (2023) 3990.

59 Singh R, Arora H C, Bahrami A, Kumar A, Kapoor N R, Kumar K, & Rai H S, *Materials,* 15(23) (2022) 8295.

60 Kumar A, Arora H C, Kapoor N R, Mohammed M A, Kumar K, Majumdar A, & Thinnukool O, *Sustainability*, 14(4) (2022) 2404.

61 Arora HC, Kumar S, Kontoni D-PN, Kumar A, Sharma M, Kapoor N R, & Kumar K, *Buildings*, 12(12) (2022) 2137.

62 Zhang Y & Ye A, *Heliyon*, 8(3) (2022) e09153.

63 Taylor K E, *J Geophys Res*, 10 6(D7) (2001) 7183-7192.