Transmission Probability of SARS-CoV-2 in Office Environment Using Artificial Neural Network

NISHANT RAJ KAPOOR\textsuperscript{1,2}, ASHOK KUMAR\textsuperscript{1,2}, ANUJ KUMAR\textsuperscript{1,3}, (Senior Member, IEEE), ANIL KUMAR\textsuperscript{4}, (Senior Member, IEEE), AND KRISHNA KUMAR\textsuperscript{5}, (Graduate Student Member, IEEE)

\textsuperscript{1}Academy of Scientific and Innovative Research (AcSIR), Ghaziabad 201002, India
\textsuperscript{2}Department of Architecture and Planning, Central Building Research Institute (CSIR), Roorkee 247667, India
\textsuperscript{3}Building Energy Efficiency Division, Central Building Research Institute (CSIR), Roorkee 247667, India
\textsuperscript{4}School of Computing, DIT University, Dehradun 248009, India
\textsuperscript{5}Department of Hydro and Renewable Energy, Indian Institute of Technology Roorkee, Roorkee 247667, India

Corresponding author: Anil Kumar (dahiyaanil@yahoo.com)

\section*{ABSTRACT}
In this paper, curve-fitting and an artificial neural network (ANN) model were developed to predict R-Event. Expected number of new infections that arise in any event occurring over a total time in any space is termed as R-Event. Real-time data for the office environment was gathered in the spring of 2022 in a naturally ventilated office room in Roorkee, India, under composite climatic conditions. To ascertain the merit of the proposed ANN and curve-fitting models, the performances of the ANN approach were compared against the curve fitting model regarding conventional statistical indicators, i.e., correlation coefficient, root mean square error, mean absolute error, Nash-Sutcliffe efficiency index, mean absolute percentage error, and a20-index. Eleven input parameters namely indoor temperature ($T_{\text{In}}$), indoor relative humidity ($\text{RH}_{\text{In}}$), area of opening ($A_{\text{O}}$), number of occupants ($O$), area per person ($A_{\text{P}}$), volume per person ($V_{\text{P}}$), $CO_2$ concentration ($CO_2$), air quality index (AQI), outer wind speed ($W_S$), outdoor temperature ($T_{\text{Out}}$), outdoor humidity ($\text{RH}_{\text{Out}}$) were used in this study to predict the R-Event value as an output. The primary goal of this research is to establish the link between $CO_2$ concentration and R-Event value; eventually providing a model for prediction purposes. In this case study, the correlation coefficient of the ANN model and curve-fitting model were 0.9992 and 0.9557, respectively. It shows the ANN model’s higher accuracy than the curve-fitting model in R-Event prediction. Results indicate the proposed ANN prediction performance ($R = 0.9992$, RMSE = 0.0018708, MAE = 0.0006675, MAPE = 0.8643816, NS = 0.9984365, and a20-index = 0.9984300) is reliable and highly accurate to predict the R-event for offices.

\section*{INDEX TERMS}
ANN, carbon dioxide concentration, COVID-19, indoor air quality, natural ventilation, real-time monitoring, office building, curve fitting, public health.

I. INTRODUCTION
Governments of several nations are predicting the occurrence of fourth wave of Coronavirus Disease (COVID-19) in their countries, while many countries are already facing the critical situation. The Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) instigated the COVID-19 and it has been spreading rapidly across the world since the beginning of 2020. On 30\textsuperscript{th} January, 2020, the World Health Organization (WHO) declared the outbreak a “Public Health Emergency of International Concern”, and it was designated a “pandemic” on 11\textsuperscript{th} March, 2020. Many countries immediately implemented preventive methods like social (physical) distancing and lockdowns due to the COVID-19 pandemic. This results in negligible or limited outdoor activities during the COVID-19 pandemic, consequently, people spend more time indoors. SARS-CoV-2 has been abbreviated as ‘SC-2’ during this study.

This outbreak of SC-2 causes immense omnidirectional losses (human lives, economic loss, physical and...
psychological health loss, educational and cultural loss, job loss) to every nation [1], [2], [3], [4], [5], [6]. Doubts regarding the pathways of transmission SC-2 have been raised because of the prompt rise of COVID-19 symptoms as well as mortalities throughout the world. WHO COVID-19 Dashboard [7] shows that the death toll will soon cross the 6.5 million mark along with more than 5 billion total cases of COVID-19 as reported to the WHO throughout the globe.

Contact transmission and droplet transmission are widely recognized by the public [8]. Albeit, in addition to the more commonly established transmission routes of direct contact with sick individuals or contaminated surfaces and larger respiratory droplets, inhaling tiny airborne droplets (aerosols) is a likely third route of infection. According to several recent studies, different sizes of droplets are generated when infected person breathe, speak, sneeze, and cough, the larger droplets swiftly settle in the 1-2 metres range [9], [10]. As illustrated in Figure 1, droplets with smaller diameters can travel a longer distance, depositing greater than 2 metres away from the virus-emitting source [11]. Smaller particles carry more viral concentration and due to negligible influence of gravitational force and the presence of air flow they can travel up to 6 metre distance. Figure 2(b) illustrates the relationship between the droplet and airborne transmission with the distance between infected and susceptible persons. The susceptible person ‘1’ is at high risk due to less distance from Infected person ‘I’, whereas susceptible person ‘2’ is at moderate risk and susceptible person ‘3’ is at low risk due to distance of more than 6 meters from the person ‘I’. Additionally, Figure 2(a) presents that smaller particles penetrate deeply into the human respiratory system and affect more.

Various studies around the globe found that airborne transmission is the prime cause of the rapid infection spread [12], [13], [14], [15], [16], [17], [18], [22],[27], [28], [29], [30], [31]. To keep the economy growing, viral transmission concerns have been given increased attention in building facilities such as offices, schools, hotels, commercial buildings, hospitals, and other buildings. It is well documented that most of the ‘super-spreading’ cases where multiple people are found infected; occurred in closed spaces like offices, restaurants, schools, homes, building apartments, hospitals, assembly halls, etc. References [12], [13], [14], [15], [16], [17], and [18]. People spend the majority of their time indoors, and the quality of their indoor environment has a direct impact on their health [19], [20], [21]. Indoor Environmental Quality (IEQ) is responsible for indoor transmissions of pathogens. Transmission of SC-2 via air has been proven, explaining the rapid increase in COVID-19 cases in closed structures without proper ventilation [22]. Moreover, anthropogenic respiratory activities also affect the transmission rate as there is variation among respirational expelled particle’s count, velocity, size, settling distance, as well as activity occurrence duration and frequency. Breathing, talking, coughing, and sneezing are the most common respirational activities in an office environment [23], [24], [25], [26], the aforementioned activity-associated variations are tabulated in Table 1.

Several cases of suspected aerial transmission of various infectious illnesses (including COVID-19) in indoor spaces have been recorded. The first instance comprised thirty-seven confirmed cases out of sixty-eight people who shared a bus in China for a hundred minutes’ journey [27]. The Diamond Princess cruise ship is the second notable occurrence, with 712 confirmed cases recorded out of 3711 people traveling on the ship [28]. The third instance had 9 confirmed cases; an index patient (who had earlier visited Wuhan, the epicentre of the pandemic) who lived in apartment infected 4 other
neighbouring residents in a housing complex in China [29]. In an experimental study conducted by Van Doremalen et al. (2020) [30], SC-2 remained alive in aerosol form and suspended in the air for the whole 3-hour experiment. Shrestha et al. [31] studied SC-2 aerosol dispersion in a virtual office building and concluded that unventilated spaces are more prone to transmissions. They found that when a building is under-ventilated, significant aerosol build-up can develop even with low occupancy.

Poor ventilation in the building’s interior areas has been linked to an increase in the spread of other respiratory diseases apart from SC-2. However, according to the current priority researchers pointed out that improper ventilation is an important factor in the spread of SC-2 inside different types of buildings. After recognising the significance of ventilation systems in preventing the transmission of SC-2, various researchers, policymakers, building-related research institutes, and Heating, Ventilation and Air-Conditioning (HVAC) associations and societies of different countries have published revised guidelines on the management as well as operation of ventilation systems during pandemic. Several nations across the globe, including the United States, India, China, Japan, Canada, and the European Union, have released recommendations to combat COVID-19 [32], [33]. Approximately all the guidelines suggested that it is quite beneficial to increase external air ventilation throughout the day in a building. Table 2 presents a comparable overview of several important guidelines and recommended temperature and relative humidity limits. These guidelines also considered natural ventilation, specific spaces, \( \text{CO}_2 \) as a proxy indicator, door/window opening, etc. for effective prevention of the SC-2 spread in buildings.

In buildings, to decrease the use of energy, Natural Ventilation (NV) is one of the most basic technique. NV permits...
substantial air changes per hour for minimal operating costs with effective heat removal. Airflow through the building with NV is caused by both buoyancy forces and wind effect, a wind pressure component is generally not considered as design principles. NV is dependent on buoyant forces created by temperature variations within the structure and wind from the outside. Buildings can be built to take advantage of one or both of these driving forces [46], [48], [47].

According to a research study [48], in a poorly ventilated room, the aerosol suspension duration increased tenfold. In a steady-state condition, the interior ventilation and the indoor CO2 concentration have a specific relationship [49]. As a result, if a suitable device to detect interior ventilation is not available, the indoor CO2 concentration might be utilised as a proxy indication. Indoor CO2 levels reflect a building’s ventilation performance in relation to the air changes, air flow pattern (depending upon indoor and outdoor environment, anthropogenic activities indoors), occupant density, opening area, and spatial volume of occupied space. Conventional CO2 measurement strategies fail to capture spatial or temporal variations in the CO2 concentration level. Recent technological developments in CO2 monitoring and prediction provide opportunities to monitor CO2 and accurately predict the concentration levels inside buildings. This develops deeper insights into individuals’ CO2 exposure. Additionally, exact air changes per hour (ACH) is hard to find in a real-time office working environment because conditions are highly dynamic in naturally ventilated building spaces. For regular indoor environments, the CO2 content should be kept at 1000 ppm or below. However, according to the SAGE-EMG guidelines [50] for COVID-19 like situations, one should keep CO2 levels below 800 parts per million (ppm) in indoor environments where a lot of aerosols are produced. If the concentration reaches 1500 parts per million or more, it should be prioritised for improvement. CO2 could be used as a proxy to detect the chances of infection spread due to SC-2 as both COVID-19 spreading and CO2 concentration depend upon occupant density inside any confined space.

Rudnick and Milton [51] evaluate the risk of indoor airborne viral transmission based on carbon dioxide concentration. They develop a model employing CO2 concentration as a marker for exhaled-breath exposure and computing the proportion of inhaled air that has already been discharged by someone in the building. Without assessing the outside air supply rate or assuming that it stays constant over time, authors created a CO2-based risk model without presuming that the concentration of an infectious agent has achieved steady-state. Similarly, in 2021; Peng and Jimenez [52] nudge the idea to reconsider the CO2 concentration as a COVID-19 infection risk proxy for different indoor environments and activities. Authors create analytical formulations for CO2-based risk proxies and applied them to a variety of common indoor settings. Authors limits the scope of study by stating a limitation that there are large uncertainties associated with infection risk estimates mainly from virus exhalation rates but they recommended to install low-cost CO2 based indoor infection risk monitoring systems to promote public health and safety. Bazant and Bush [53] recommended limited time spent in a shared place with an infected individual for minimising COVID-19 indoor airborne transmission. Later ensuing the same, Bazant et al. [54] reformulated this safety recommendation in terms of occupancy duration and mean exhaled carbon dioxide content in an indoor environment, allowing the use of CO2 sensors in the risk assessment of airborne transmission of pathogenic viruses. Authors developed a mathematical model to forecast the probability of airborne transmission based on real-time CO2 observations. Authors strongly concluded that as CO2 levels are considered as direct indicator for ventilation and mixing of air it can also be used to assess the risk of airborne transmission of viruses like SC-2. Wang et al. [55] presented a review summary on the airborne transmission of respiratory viruses. Carbon dioxide sensors can be used to monitor and improve ventilation by acting as indications of the accumulation of exhaled air. Although the ventilation type should also be considered, it has been recommended to keep carbon dioxide levels below 700 to 800 ppm. Martinez et al. [56] developed a novel agent-based simulator named “ArchABM”. ArchABM is intended to aid in the design of new buildings or the adaptation of existing ones by calculating sufficient room sizes, ventilation characteristics, and assessing the effect of policies while accounting for IAQ as a result of complex human-building interaction patterns. In this work, a previously published aerosol model was modified to compute time-dependent carbon dioxide (CO2) and viral quanta concentrations in each room, as well as inhaled CO2 and virus quanta concentrations for each occupant over the course of a day as a measure of physiological response. Recently, Tang et al. [57] consider spatio-temporal airborne viral infection risk assessment in real time using CO2 concentration field monitoring and occupancy in the built environment. Authors suggested a technique that can assess real-time airborne disease infection risks via aerosol transmission. The method was applied in a university office and the cumulative infection risk was evaluated for three residents with three-day work. The current work nudges these research by introducing AI to predict the airborne transmission by using real-time collected data of CO2 and other environmental parameters.

Humidity and temperature are important parameters of an indoor environment and affect the virus survival rate. Both humidity and temperature play an important role in indoor human comfort and directly affect the occupant’s perception. Mecenas et al. [58] reviewed the literature on the effects of humidity and temperature on COVID-19 transmission and explored that it is important to include these parameters in future studies as these can affect the spreading of the disease. Several guidelines of different countries also considered relative humidity and temperature as important parameters. Many of these guidelines (as mentioned in Table 2) recommended boundary conditions for indoor relative humidity and temperature; considering both human comfort indoors as well as prevailing SC-2 spreading. Additionally, outdoor
TABLE 2. Parameters considered by different guidelines for preventing SC-2 spread in built environments.

| Reference | Organisation | Country (Year) | Specific Spaces | RH (%) | Temp. (°C) | CO\textsubscript{2} as Indicator | Window Open | Outer Air Supply |
|-----------|--------------|----------------|-----------------|--------|------------|-------------------------------|-------------|-----------------|
| [34], [35] | CSIR         | India (4.2021) | Yes             | 40 – 70| 24 – 30    | -                             | Yes         | Yes             |
| [36]      | PHI          | Canada (8.2020) | Yes             | 40 – 60| -          | -                             | Yes         | Yes             |
| [37]      | REHVA        | EU (8.2020)    | Yes             | 40 – 60| -          | -                             | Yes         | Yes             |
| [38]      | CCHQ         | Canada (8.2020) | Yes             | 40 – 60| -          | -                             | Yes         | Yes             |
| [39]      | NIC          | China (7.2020) | Yes             | -      | -          | -                             | Yes         | Yes             |
| [40]      | ECDC         | EU (6.2020)    | Yes             | -      | -          | -                             | Yes         | Yes             |
| [41, 42]  | ASHRAE       | USA (4.2020)   | Yes             | 40 – 60| 18 – 26    | -                             | Yes         | Yes             |
| [43]      | ISHRAE       | India (4.2020) | Yes             | 40 – 70| 17 – 28    | -                             | Yes         | Yes             |
| [44]      | SHASE        | Japan (3.2020) | Yes             | 40 – 70| 24 – 30    | -                             | Yes         | Yes             |
| [45]      | ASC          | China (2.2020) | Yes             | -      | -          | -                             | Yes         | Yes             |

environmental conditions, occupant behaviour, and indoor occupant density are among other important parameters to reducing SC-2 spread in an indoor environment [59], [60]. Several organizations and governments of different countries around the world had issued guidelines on COVID-19 appropriate behaviour in office environments [35], [39], [43], [45], [50], [61], [62]. Reduced occupancy, personal hygiene, good ventilation, proper personal safety, timely training, authentic information circulation, and worker’s health monitoring were some of the important recommendations to employers for a safe working environment.

A suitable door and window opening strategy according to spatiotemporal variation and occupant activity behavioural pattern helps in efficiently reducing indoor CO\textsubscript{2} concentration and thus, affecting the probability of SC-2 transmission in an office environment. Table 2 shows that approximately all the guidelines recommended opening doors/windows for enhancing natural ventilation and diluting exhaled stale air indoors. However, in some cases, opening doors/windows are not suitable. Especially in case of like extreme exterior noise, visual privacy concerns, high level of exterior pollution (depending on AQI) infiltrating indoors, unwanted excessive light, uncomfortable air movements/drafts, invasion of flies and other insects, etc. which can reduce occupant concentration and increasing errors resulting in productivity loss. AQI is also an important parameter as it represents pollutants level, PM\textsubscript{2.5} and PM\textsubscript{10} are the most important component of AQI.

Tupper et al. [63] introduced a concept named “event R” in relation with SC-2. According to Tupper et al. [63], “event-R is the expected number of new infections due to the presence of a single infectious individual at an event”. Using reports of small outbreaks authors obtained a fundamental correlation between “event R” and four other parameters namely, proximity of individuals, exposure duration, degree of mixing, and transmission intensity. The event-R is also termed as R-Event. REHVA also used R-Event alongside infection probability in their updated version (Version 2.1) of the airborne COVID-19 prediction calculator [64].

Currently, throughout the world, scientific investigation of the foregoing transmission of different variants of SC-2 virus in indoor settings is trying to find out the accurate infection probability. Worldwide scientific community look at two types of infection probability in bounded spaces: (i) based on air change rate, and (ii) based on exhaled CO\textsubscript{2} by the occupants. In this study, the above-mentioned second type; based on exhaled CO\textsubscript{2} by the occupants is linked with the real-time dynamic spatiotemporal indoor and outdoor environmental parameters to predict the R-Event value. R-Event value is an enhanced and more informative substitute for infection probability. The Version 2.1 calculator developed by the REHVA (mentioned above) is used to find out the R-Event value in the existing conditions and recorded simultaneously with other real-time parameters to develop a link between exhaled CO\textsubscript{2} and R-Event in an office environment. Unknown single infected occupant (among four static occupants) in the office environment is considered as an infection source to find the possibility of airborne transmission (through aerosols and droplets only). This study aims to help readers to accurately predict the R-Event value using indoor CO\textsubscript{2} levels in an office room. Along with CO\textsubscript{2}, indoor as well as outdoor environmental parameters, occupancy, and occupant information is also required for attaining prediction results.

In this work, the curve fitting and supervised machine learning-based ANN model have been used to forecast the R-Event value as a proxy to the probability of SC-2 spreading. The objectives and novelty of the proposed study is explained in Section 2. The details of the input parameters used to construct these models is in Section 3. Section 4 presents the description of curve fitting and ANN models. The findings of the curve fitting model and the results of the ANN models are described in Section-5. The last section of this study contains the conclusion, limitations, and future scope of the work.

II. NOVELTY AND OBJECTIVES

If an infected occupant is present inside a building with susceptible occupants and all are exhaling and inhaling at the same time, lower ventilation rates and increased occupant density lead to a build-up of carbon dioxide (CO\textsubscript{2}) with free-floating SC-2 virus responsible for the airborne spread of COVID-19. Various simulation studies and numerical models have been developed to predict infection probability using subjective and objective data. However, just a few researchers have attempted to establish a link between occupant CO\textsubscript{2} emissions and the transmission of SC-2. No real-time study has yet published a model based on advanced computing techniques such as artificial intelligence (AI).
There is very limited literature on indoor CO₂ concentration relation with viral transmissions [48], [51], [52], [54], [56], [57]. However, the indoor CO₂ concentration can be used to forecast the number of infected people as scientifically CO₂ can be used as a proxy for SC-2 concentration indoors. CO₂ concentration is not yet linked with the amount of infectious particles present in the air in real-time dynamic conditions like an uninterrupted office environment or classroom environment. The objectives of this research study are threefold:

(i) Collecting real-time spatial and temporal data of both environmental and personal parameters to establish a relationship among parameters using machine learning (ML)

(ii) Developing a novel link between CO₂ with infection probability-based R-Event value

(iii) Comparing novel ANN and curve-fitting model for R-Event prediction.

The other advantages of this study were to develop the CO₂ concentration database based on high-resolution real-time office-environment data and to investigate individuals’ exposure to the human exhaled CO₂ by mapping occupant’s activities inside the office and linking it with R-Event.

III. MATERIALS AND METHODS

CO₂ was measured continuously during office time using EXTECH “Indoor Air Quality Meter/Datalogger” Model EA80. The EA80 checks for CO₂ concentrations using a maintenance-free CO₂ sensor with a range of 0 to 6,000 ppm. The device was kept distant from the static occupants as directly exhaled CO₂ over the device will degrade the accuracy of the measurement. Temperature and humidity were measured by EXTECH “Heat Stress WBGT Meter” Model HT200. The HT200 enables accurate measurements for Air Temperature (TA), and Relative Humidity (RH). Heat Stress Index measures how hot it feels when humidity is combined with radiant heat, air movement, as well as temperature. Both the instruments were set at a height of 0.8m above the floor level. The technical specification of both the devices is presented in Table 3 and Table 4. Figure 3. (a) and (b) respectively shows the CO₂ measurement device EXTECH Model EA80 and heat stress meter EXTECH Model HT200 for temperature and humidity measurement. The instruments were produced by FLIR Commercial Systems Inc. under the EXTECH Instruments brand. As per the stringent guidelines of FLIR Commercial System the instruments hold ISO-9001 certification for Quality Management System; this is the prime reason for selecting these models for this study. Both the instruments were calibrated and inspected before use. Some better models are also available for the same purpose with higher accuracy but due to economic constraints, these models were used during this study. The calculation process was carried out in a MATLAB R2021a environment on a laptop computer (Dell Inspiron 15R, Intel(R) Core(TM) i5-4200U CPU @160GHz 2.30GHz, 6GB RAM).

A total of 638 data sets out of 950 data sets were used to build the prediction model. In literature, it is found that the amount of data required to develop any ML based model is the number of input parameters multiplied by ten. This study includes eleven input variables such as indoor temperature ($T_{inh}$), indoor relative humidity ($RH_{inh}$), area of opening ($A_{O}$), number of occupants ($O$), area per person ($A_{p}$), volume per person ($V_{p}$), CO₂ concentration ($CO_{2}$), air quality index ($AQI$), outer wind speed ($W_{S}$), outdoor temperature ($T_{out}$), outdoor humidity ($RH_{out}$), and one output variable that is Event-R (R-Event). The selection of the dataset was based on the chances of having at least two subjects in the office room. Single occupancy data sets were not considered for this study so 312 data sets were excluded from the whole datasets. The CO₂ concentration inside the office is mainly affected by the occupancy inside the office room as workers themselves are the emitting source. The output R-Event is the expected number of new infections that arise in any event occurring over a total time ‘T’ in any space. The mathematical representation of R-Event is presented in equation 1 [63]. However, for this study, the R-Event is numerically worked out by using REHVA Version 2.1 calculator for SC-2 infection probability prediction. The mathematical representation of the REHVA calculator-based R-Event is presented in equation 2 [64]. The conditions inside the NV office building also depend upon the external environment around the building. Office block near dense areas, marketplaces, and busy roads are generally seen affected by external conditions. Hence, outdoor environmental parameters are also considered in this study.

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

where, ‘T’ is the total event duration, ‘t’ is the contact time between a single susceptible person and an infected person, ‘$\beta$’ is constant per unit time probability of transmission, $1 - e^{-\beta t}$ is the probability for any susceptible person being infected, and ‘k’ is the total number of susceptible persons contacting the infectious person. Simply, it can be written
Table 3. Extitechmodel Ea80 Specifications.

| Specification          | CO₂     | Air Temperature | Relative Humidity (%RH) |
|------------------------|---------|-----------------|-------------------------|
| Range                  | 0 to 6000 ppm | -20°C to +60°C | 10% to 95%              |
| Resolution             | 1 ppm   | 0.1°C           | 0.1% RH                 |
| Accuracy               | ±3% of reading or ±50ppm, whichever is greater @ 101.4 kPa @ 25°C | ±0.5°C | ±3%RH @ 25°C, 30~95%RH, ±5%RH @ at 23°C, 10~30%RH |
| Sensor type            | Dual-wavelength detector with non-dispersive infrared (NDIR) sensor | Thermistor | Precision capacitance sensor |
| Response time          | <10 minutes in still air | 10°C/2 sec | 45% → 95% ≤ 1 min, 95% → 45% ≤ 3 min |
| Warm up time           | 10 seconds | -               | -                       |

Table 4. Extitechmodel Ht200 Specifications.

| Specification          | Air Temperature | Relative Humidity (%RH) |
|------------------------|-----------------|-------------------------|
| Range                  | 0 to 50.0°C     | 1% to 99%               |
| Resolution             | 0.1°C           | 0.1% RH                 |
| Accuracy               | ±0.8 (±15 to 40°C) | ±3.0%RH (20 to 80%), ±5.0%RH (<20% or >80%) |

as equation 2.

\[ R - \text{Event} = \frac{\text{IP} \times \text{Number of susceptible person}}{\text{Number of infectious person}} \]  (2)

where, IP is infection probability.

Figure 4 depicts the methodology adopted for this study. An artificial neural network (ANN) based novel approach for predicting SC-2 transmission linked with exhaled CO₂ as a proxy indicator in an office environment is commenced by database preparation. The database preparation includes three main steps (i) collection of data, (ii) normalization of data, and (iii) splitting of data for training, testing, and validation.

A. Collection of Data

The data used in this work was generated using a CO₂ analyzer and heat stress meter in the office room of Central Building Research Institute, Roorkee; a national laboratory of CSIR, India. The measurements were taken during the month of February and March in the year 2022. The coordinates of the tested location are NL 29°51'54'' and EL 77°54'10''. The floor area of the office room was 24 m² and the volume of the room was 84 m³. Approximately, 40% of the floor area was covered by the furniture (storage cabinet, table, chair, etc.) and approximately 20% of the volume is taken by the static furniture primarily covering more volume near the floor at the height of one meter. The office is on the ground floor and contains one big glass and aluminum frame window at the height of 1.2m on the south wall having a height and width of 1.5m and 2.5m, respectively. The window is divided into three parts; the middle portion is fixed while parts at both the ends of the window were operable and considered as a variable components of the room. The office room contains two wooden doors front door and a side door. The first door (front door) was on the north wall having the dimensions of 1.2m × 2.4m and it opened in the corridor. The front door was having attached ventilation above it having a dimension of 1.2m × 0.5m. During the study period, the ventilation is closed all the time and considered as a fixed component of the room while the front door was a variable component in this study. The second door was the side door, situated on the east side wall of the office room. The type of front door was a double door and the side door was a single door. The details of room components are presented in Table 5.

Considering the Hawthorne effect [65], both static and dynamic occupants (subjects) were not told about the purpose of the study as it would impact their normal behavior (talking/working/activity level/free movement, etc.). Occupants would change their behaviour after knowing that they were being monitored by someone; their breathing patterns, movements, and other activities also got changed if they know the purpose of the study. However, only one occupant (the first author) knows the purpose of this study as he is observing the details precisely for this study. The details of the subjects are tabulated in Table 6. All the static subjects were sitting in the longitudinal direction (‘L’ shape) at a distance of approximately one meter as shown in Figure 5.

The 3-dimensional representation of the 12 various cases during the monitoring in the office room with variable scenarios of two doors and two operable windows along with furniture arrangement is shown in Figure 6. The green numbered operational cases (1, 2, 4, 5, and 10) were observed during this study. The data were collected forty-four times a day at every 10-minute interval from Monday to Friday. The data collection schedule is shown in Figure 7.

Figure 8 represents the plot matrix of the dataset used in this study. The ‘R’ and ‘P’ are the two important factors of the scatter plot diagram. The value of correlation coefficient ‘R’ is maximum in between input parameters area per person (AP) and volume per person (VP) having the value of 1 (dark red circle at the center of Figure 8). Whereas, the value of the correlation coefficient is minimum at two points; between area per person (AP) and the number of occupants (O), as well
as volume per person ($V_p$), and the number of occupants ($O$). The minimum value of $R$ is $-1$ at the above-mentioned points and is shown in the dark blue circles at the center of Figure 8. The maximum value of $P$ is 0.93 and the minimum value of $P$ is 0.

The collected eleven input parameters are indoor temperature ($T_{in}$), indoor humidity ($RH_{in}$), area of opening ($A_o$), number of occupants ($O$), area per person ($A_p$), volume per person ($V_p$), $CO_2$ concentration ($CO_2$), air quality index ($AQI$), outer wind speed ($W_s$), outdoor temperature ($T_{out}$), and outdoor humidity ($RH_{out}$). As per the input data, the range of $CO_2$ concentration levels inside the office room is 327-964 ppm and the number of occupants during this study is 2-5 persons. The output parameter R-Event ranges between 0.04-0.24. The various statistical parameters of the dataset taken are tabulated in Table 7.

### B. STANDARDIZATION OF SELECTED DATA
The data was normalized once the database had been finalized for this study after the data collection. Normalization is the most important step to define the data into certain ranges such as 0 to +0.8, 0 to +0.9, 0 to 1, or maybe $-1$ to $+1$ [66]. The values of the selected input parameters may fluctuate. Accordingly, data standardization was employed to make the computations easier. In this research work, the data was normalized between the range of 0 to +0.8 by using equation 3.

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

where, $N_{normalized}$ is the value to be normalized, $N$ is the random value in the selected database, $N_{min}$ is the minimum value of the selected dataset and $N_{max}$ is the maximum value of the selected dataset.

### TABLE 5. Details of office room components.

| Component                  | Area/Volume | Fixed/Variable | Material                      |
|----------------------------|-------------|----------------|-------------------------------|
| Floor Area                 | 24 m²       | Fixed          | Tile Flooring                |
| Area Covered by Furniture  | 10 m²       | Fixed          | Wooden, Steel, Iron          |
| Room Volume                | 84 m³       | Fixed          | -                             |
| Front Door                 | 3 m²        | Variable       | Wooden                       |
| Side Door                  | 1.8 m²      | Variable       | Wooden                       |
| Ventilation Area           | 0.6 m²      | Fixed          | Wooden, Glass                |
| Fixed Window               | 1 m²        | Fixed          | Aluminium, Glass             |
| Side Window-1              | 0.6 m²      | Variable       | Aluminium, Glass             |
| Side Window-2              | 0.6 m²      | Variable       | Aluminium, Glass             |

### TABLE 6. Occupant details occupying the observed office environment.

| Worker | Gender | Age (years) | Weight (kg) | Height (cm) | Condition | Education | COVID History | Vaccination | Health Status |
|--------|--------|-------------|-------------|-------------|-----------|-----------|---------------|-------------|---------------|
| 1      | Female | 22          | 69          | 170.0       | Static    | UG        | None          | Yes         | Healthy       |
| 2      | Female | 25          | 55          | 160.0       | Static    | Graduation| None          | Yes         | Healthy       |
| 3      | Female | 29          | 57          | 152.4       | Static    | Graduation| None          | Yes         | Healthy       |
| 4      | Male   | 28          | 65          | 169.0       | Static    | PG        | None          | Yes         | Healthy       |
| 5      | Male   | 28          | 56          | 170.0       | Dynamic   | Graduation| None          | No          | Healthy       |
C. FILTRATION OF DATA
Studies done by Kumar and Saini [67] considered 70% of data for training, 15% for validation, and 15% data for testing to implement the multi-layer feed-forward back-propagation learning algorithm. A total of 4808 randomly selected datasets were chosen by the computer itself for training, the rest of the 2060 datasets were further split into two equal parts of 1030 datasets in each group; used for testing and validation. Similarly, in another study, Kumar et al. [68] divided the data into three parts. 70% of the data is used for training, 15% for testing, and the remaining 15% is used for validation in ANN and ABC-ANN. After having a quick review of the existing literature on ANN, the data was split into three parts having 70%, 15%, and 15% of the total dataset for training, testing, and validation purpose respectively. After splitting the 638 datasets, 446 datasets were used for training purposes, 96 datasets were used for testing and 96 datasets were used for validation in this study.
IV. PREDICTION OF R-EVENT
In this research article two methods were employed to develop the correlation and predict the values of R-Event, a proxy for infection probability. The first method is based on the curve fitting method by simply applying the first-degree order equation to the input and output parameters. The second method is based on artificial intelligence known as ANN and is part of supervised machine learning.
A. CURVE FITTING METHOD

In the curve fitting procedure, a constant value that depends on the remaining input parameters was generated after first defining the relationship between the first input parameter and the output parameter. The functional connections between the parameters were then established, and constants were obtained for each input parameter throughout the whole process, using the same integrated approach for the second input parameter and so on until the last input parameter. In detail, a first-order linear equation was used to determine the functional correlation between the coefficient of $R$-Event and the indoor temperature ($T_{in}$). A scatter diagram between $R$-Event as a function of $T_{in}$ is presented in Figure 9(a), and equation 4 shows the connection between $R$-Event and $T_{in}$.

\[
R\text{-Event} = -0.0586 \times T_{in} + 0.2875 \tag{4}
\]

Let, $N_1 = 0.2875$, as the constant value and furthermore the functions of the rest of inputs i.e. $RH_{in}$, $AO$, $Ap$, $VP$, $CO_2$, $AQI$, $W_S$, $T_{out}$, and $RH_{out}$. Therefore, the updated equation 4 can be expressed as:

\[
R\text{-Event} = -0.0586 \times T_{in} + N_1 \tag{5}
\]

For obtaining the correlation between the coefficient of $T_{in}$ and indoor relative humidity ($RH_{in}$), the relationship between $N_1$ and $RH_{in}$ is drawn, as shown in Figure 9(b). The obtained equation between $N_1$ and $RH_{in}$ is expressed as in equation 6.

\[
N_1 = 0.1641RH_{in} + 0.2193 \tag{6}
\]

Putting the value of $N_1$ in equation 5, the updated equation expressed as:

\[
R\text{-Event} = -0.0586 \times T_{in} + 0.1641RH_{in} + 0.2193 \tag{7}
\]
Let, $N_2 = 0.2193$, as the constant value and furthermore the functions of the rest of inputs i.e. $A_O$, $O$, $P$, $V_P$, $CO_2$, $AQI$, $W_S$, $T_{Out}$, and $RH_{Out}$. Therefore, equation 7 can be expressed as:

$$R-\text{Event} = -0.0586 \, T_{In} + 0.1641 \, RH_{In} + N_2$$

(8)

For obtaining the correlation between the coefficient of indoor temperature, indoor relative humidity, and area of opening ($A_O$) on R-Event the relationship between $N_2$ and $A_O$ is drawn, as shown in Figure 9(c). The obtained equation between $N_2$ and $A_O$ is expressed as in equation 9.

$$N_2 = -0.2857 \, A_O + 0.3590$$

(9)

Putting value of $N_2$ in equation 8, the equation 8 become:

$$R-\text{Event} = -0.0586 \, T_{In} + 0.1641 \, RH_{In} - 0.2857 \, A_O + 0.3590$$

(10)

Let, $N_3 = 0.3590$, as the constant value and furthermore the functions of the rest of inputs i.e. $O$, $A_P$, $V_P$, $CO_2$, $AQI$, $W_S$, $T_{Out}$, and $RH_{Out}$. Therefore, the equation 10 can be expressed as:

$$R-\text{Event} = -0.0586 \, T_{In} + 0.1641 \, RH_{In} - 0.2857 \, A_O + N_3$$

(11)

For obtaining the correlation between the coefficient of indoor temperature, indoor relative humidity, area of opening, number of occupants, area per person, volume per person ($V_P$) on R-Event, the relationship between $N_3$ and $V_P$ is drawn, as shown in Figure 9(f). The obtained equation between $N_3$ and $V_P$ is expressed as in equation 12.

$$N_3 = 0.9980 \, O - 0.0417$$

(12)

Putting value of $N_3$ in equation 11, the equation 11 become:

$$R-\text{Event} = -0.0586 \, T_{In} + 0.1641 \, RH_{In} - 0.2857 \, A_O + 0.9980 \, O - 0.0417$$

(13)

Let, $N_4 = 0.0417$, as the constant value and furthermore the functions of the rest of inputs i.e. $A_P$, $V_P$, $CO_2$, $AQI$, $W_S$, $T_{Out}$, and $RH_{Out}$. Therefore, the equation 13 can be expressed as:

$$R-\text{Event} = -0.0586 \, T_{In} + 0.1641 \, RH_{In} - 0.2857 \, A_O + 0.9980 \, O - N_4$$

(14)

For obtaining the correlation between the coefficient of indoor temperature, indoor relative humidity, area of opening, number of occupants, and area per person ($A_P$) on R-Event, the relationship between $N_4$ and $A_P$ is drawn, as shown in Figure 9(e). The obtained equation between $N_4$ and $A_P$ is expressed as in equation 15.

$$N_4 = 0.0017 \, A_P + 0.0414$$

(15)

Putting value of $N_4$ in equation 14 and equation 14 become:

$$R-\text{Event} = -0.0586 \, T_{In} + 0.1641 \, RH_{In} - 0.2857 \, A_O + 0.9980 \, O - 0.0017 \, A_P - 0.0414$$

(16)

Let, $N_5 = 0.0414$, as the constant value and furthermore the functions of the rest of inputs i.e. $V_P$, $CO_2$, $AQI$, $W_S$, $T_{Out}$, and $RH_{Out}$. Therefore, the equation 16 can be expressed as:

$$R-\text{Event} = -0.0586 \, T_{In} + 0.1641 \, RH_{In} - 0.2857 \, A_O + 0.9980 \, O - 0.0017 \, A_P - N_5$$

(17)

For obtaining the correlation between the coefficient of indoor temperature, indoor relative humidity, area of opening, number of occupants, area per person, and volume per person ($V_P$) on R-Event, the relationship between $N_5$ and $V_P$ is drawn, as shown in Figure 9(f). The obtained equation between $N_5$ and $V_P$ is expressed as in equation 18.

$$N_5 = 0.00002 \, V_P + 0.0414$$

(18)

Putting value of $N_5$ in equation 17, the equation 17 become:

$$R-\text{Event} = -0.0586 \, T_{In} + 0.1641 \, RH_{In} - 0.2857 \, A_O + 0.9980 \, O - 0.0017 \, A_P - 0.00002 \, V_P - 0.0414$$

(19)

Let, $N_6 = 0.0414$, as the constant value and furthermore the functions of the rest of inputs i.e. $CO_2$, $AQI$, $W_S$, $T_{Out}$, and $RH_{Out}$. Therefore, the equation 19 can be expressed as:

$$R-\text{Event} = -0.0586 \, T_{In} + 0.1641 \, RH_{In} - 0.2857 \, A_O + 0.9980 \, O - 0.0017 \, A_P - 0.00002 \, V_P - N_6$$

(20)

For obtaining the correlation between the coefficient of indoor temperature, indoor relative humidity, area of opening, number of occupants, area per person, volume per person, and human exhaled carbon dioxide ($CO_2$) on R-Event, the relationship between $N_6$ and $CO_2$ is drawn, as shown in Figure 9(g). The obtained equation between $N_6$ and $CO_2$ is expressed as in equation 21.

$$N_6 = -0.0791 \, CO_2 + 0.0601$$

(21)

Put value of $N_6$ in equation 20 and updated equation 20 is expressed as:

$$R-\text{Event} = -0.0586 \, T_{In} + 0.1641 \, RH_{In} - 0.2857 \, A_O + 0.9980 \, O - 0.0017 \, A_P - 0.00002 \, V_P + 0.0791 \, CO_2 - 0.0601$$

(22)

Let, $N_7 = 0.0601$, as the constant value and furthermore the functions of the rest of inputs i.e. $AQI$, $W_S$, $T_{Out}$, and $RH_{Out}$. Therefore, the equation 22 can be expressed as:

$$R-\text{Event} = -0.0586 \, T_{In} + 0.1641 \, RH_{In} - 0.2857 \, A_O + 0.9980 \, O - 0.0017 \, A_P - 0.00002 \, V_P + 0.0791 \, CO_2 - N_7$$

(23)
FIGURE 9. Curve fitting models.

For obtaining the correlation between the coefficient of indoor temperature, indoor relative humidity, area of opening, number of occupants, area per person, volume per person, human exhaled carbon dioxide, and air quality index (AQI) on R-Event, the relationship between $N_7$ and $AQI$ is drawn, as shown in Figure 9(h). The obtained equation between $N_7$
Let, \( N_9 = 0.0587 \), as the constant value and also the functions of the rest of inputs i.e. \( T_{Out} \), and \( RH_{Out} \). The equation 28 can be expressed as:

\[
R-\text{Event} = -0.0586 T_{In} + 0.1641 RH_{In} - 0.2857 A_0 + 0.0998 O - 0.0017 A_P - 0.00002 V_P + 0.0791 CO_2 - 0.0333 AQI + 0.0381 W_S - N_9
\]  
(29)

For obtaining the correlation between the coefficient of indoor temperature, indoor relative humidity, area of opening, number of occupants, area per person, volume per person, human exhaled carbon dioxide, air quality index, wind speed, and outdoor temperature (\( T_{Out} \)) on R-Event, the relationship between \( N_9 \) and R-Event is drawn, as shown in Figure 9(j). The obtained equation between \( N_9 \) and \( T_{Out} \) is expressed as in equation 30.

\[
N_9 = -0.0433 T_{Out} + 0.0793
\]  
(30)

By putting value of \( N_9 \) in equation 29, the equation 29 become:

\[
R-\text{Event} = -0.0586 T_{In} + 0.1641 RH_{In} - 0.2857 A_0 + 0.0998 O - 0.0017 A_P - 0.00002 V_P + 0.0791 CO_2 - 0.0333 AQI + 0.0381 W_S + 0.0433 T_{Out} - 0.0793
\]  
(31)

Let, \( N_{10} = 0.0793 \), as the constant value and furthermore the function input parameter i.e. \( RH_{Out} \). Therefore, the equation 31 can be expressed as:

\[
R-\text{Event} = -0.0586 T_{In} + 0.1641 RH_{In} - 0.2857 A_0 + 0.0998 O - 0.0017 A_P - 0.00002 V_P + 0.0791 CO_2 - 0.0333 AQI + 0.0381 W_S + 0.0433 T_{Out} - N_{10}
\]  
(32)

For obtaining the correlation between the coefficient of indoor temperature, indoor relative humidity, area of opening, number of occupants, area per person, volume per person, human exhaled carbon dioxide, air quality index, wind speed, outdoor temperature, and outdoor relative humidity (\( RH_{Out} \)) on R-Event, the relationship between \( N_{10} \) and \( RH_{Out} \) is drawn, as shown in Figure 9(k). The obtained equation between \( N_{10} \) and \( RH_{Out} \) is expressed as in equation 33.

\[
N_{10} = 0.0087 RH_{Out} + 0.0761
\]  
(33)

By putting value of \( N_{10} \) in equation 32, the equation 32 become:

\[
R-\text{Event} = -0.0586 T_{In} + 0.1641 RH_{In} - 0.2857 A_0 + 0.0998 O - 0.0017 A_P - 0.00002 V_P + 0.0791 CO_2 - 0.0333 AQI + 0.0381 W_S + 0.0433 T_{Out} - 0.0087 RH_{Out} - 0.0761
\]  
(34)
Let $N_{11} = 0.0761$ The equation become:

$$R\text{-Event} = -0.0586 \, T_{\text{In}} + 0.1641 \, R_{H_{\text{In}}} - 0.2857 \, A_{O} + 0.9980 \, O - 0.0017 \, A_{P} - 0.00002 \, V_{P} + 0.0791 \, C_{O} - 0.0333 \, A_{QI} + 0.0381 \, W_{S} + 0.0433 \, T_{\text{Out}} - 0.0087 \, R_{H_{\text{Out}}} - N_{11} \quad (35)$$

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

$$C_c = -0.0586 \, T_{\text{In}} + 0.1641 \, R_{H_{\text{In}}} - 0.2857 \, A_{O} + 0.9980 \, O - 0.0017 \, A_{P} - 0.00002 \, V_{P} + 0.0791 \, C_{O} - 0.0333 \, A_{QI} + 0.0381 \, W_{S} + 0.0433 \, T_{\text{Out}} - 0.0087 \, R_{H_{\text{Out}}} \quad (36)$$

Hence, the final equation can be written as:

$$R\text{-Event} = C_c - N_{11} \quad (37)$$

**B. ARTIFICIAL NEURAL NETWORK (ANN)**

The first ANN, known as a perceptron, was created in 1958 by psychologist Frank Rosenblatt with the goal of modelling how the human brain processes visual input and learns to recognize things. Over the last four decades, artificial neural networks have made significant contributions to implementing many types of intelligent information processing, understanding the core roles of actual neurons and brain activity, and also in many engineering applications [69]. The ANN modelling technique is a computer-based methodology for simulating several key aspects of the human nervous system. The capacity to solve issues by applying knowledge gained from previous experiences to new challenges or case scenarios.

Learning models like ANN build a network of core linkages between chosen outputs and available inputs that are linked together by connections known as neurons that are hidden behind layers. Every ANN neuron has a weight parameter and linkages to neurons from the preceding layer (hidden or input), with each connection having its own assigned weight. The number of neurons in the ANN’s hidden layer has a significant impact on model learning performance. The network learning function will be unable to converge to an ideal value if the number of neurons in the hidden layer is too low, resulting in oscillatory behavior of the error function, and hence will be unable to learn the connection between the input-output designs. Overtraining is the primary cause of overfitting [70]. A trained ANN assembly can identify causal relationships between inputs and outputs by comparing measured outputs with predicted outputs to get the best possible result [71]. Figure 10 shows the usual structure of a neural network, as well as a single node overview. Advantage of using neural networks is that it may create models from complicated natural systems with massive inputs easier to use and more accurate [72]. As a result, an artificial structure may be built in the manner of natural networks, as well as the relationship between its components can be determined by altering the values of each connection. Applying a certain input after training the neural network results in a definite output.
The most crucial aspect of training is to reduce the errors maximize the value correlation coefficient \( R = 1 \). This is accomplished by adjusting the weights throughout the learning phase until the error function is attained. The ‘\( R \)’ and ‘\( \text{MSE} \)’ metrics are used to assess the network’s performance in ANN. The equations of ‘\( R \)’ and ‘\( \text{MSE} \)’ are expressed in equations (38) and (39) respectively.

\[
R = \frac{\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N} (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{N} (y_i - \bar{y})^2}}
\]  

\[
\text{MSE} = \frac{\sum_{i=1}^{N} (x_i - y_i)^2}{N}
\]  

R represents the Pearson correlation coefficient, \( x_i \) is the measured values (x-variable) in a sample, \( \bar{x} \) is the mean value of the measured values, \( y_i \) is the predicted values (y-variable) in a sample and \( \bar{y} \) is the mean value of the predicted values.

The ‘gradient-descent’ strategy is an excellent way to quickly identify the least sets of errors. As the name indicating gradient descent employs the error gradient to slope down the error [73]. Backpropagation determines the error, which is proportional to the network’s output and is defined by the weighted output of the hidden neurons as well as the weights. Werbos [74] proposed backpropagation as a mechanism for learning and afterward it was suggested by Rumelhart [75]. Backpropagation is a kind of gradient descent technique in which the network’s weights move in the reverse direction of the performance function slope. This study proposed an ANN model to predict the R-Event for an office environment in the observed boundary conditions of the eleven input variables.

1) MODELING OF PROPOSED ANN MODEL

Procedure of representing a real-world object or phenomenon as a sequence of computational statements is ‘Modeling’. It is essential to find the network’s optimal design, which simultaneously delivers both high accuracy and a well-set. Figure 11 represents the ANN architecture for this study with eleven input parameters and R-Event as an output parameter. The number of ideal neurons and hidden layers was determined by trial and error because there is no formula for determining the exact number of neurons and hidden
The optimum structure was found after testing various architectures with various numbers of neurons in each layer. The MSE and R of the network were calculated for every neuron in the hidden layer, whereas the network’s other parameters remained unchanged. Figure 12 shows the network performance based on MSE-value. In this work, the ANN model was trained from 3 to 17 neurons. At the 3rd neuron, the correlation coefficient of training was 0.9905 and at the 17th neuron, the correlation coefficient of training reached 0.9999. At least 20 iterations were performed on each of the ANN models. Figure 13 illustrates the optimal design of the proposed ANN model. The hit and trial method was executed for finding the MSE and R of the data. The rank for different neurons was decided on the basis of their respective MSE and R-value performance. The neurons with higher R values ranked higher but on the contrary, the neurons with lower MSE were ranked higher as presented in Table 8. For finding the overall best neurons the ranks of all neurons for R and MSE were totaled for training, testing, and validation. For the 8th neuron trial, the total ranking is the least among all the trials. The network with a single hidden layer and 8 neurons in the hidden layer predicts the
best performance among other neurons. The database characteristics were normalized linearly in between 0 to 0.8 to increase the speed of the learning operation and lead to faster convergence. The computations were done with the ANN toolbox in MATLAB (R2021a) environment. The available training algorithms in MATLAB are Bayesian Regularization, Scaled Conjugate Gradient, and Levenberg–Marquardt. Out of these algorithms, the Levenberg–Marquardt (LM) training algorithm was employed in this work to train the network because of its appropriate convergence, high accuracy, and low time consumption [76], [77], [78]. This approach usually necessitates more memory, but it takes less time. When generalisation stops improving, as shown by a rise in the MSE of the validation samples, training automatically terminates. This method divided the data into three portions at random: 15% for testing, 15% for validation, and 70% for training; the same procedure is used in this study. TANSIG (equation (40)) and PURELIN (equation (41)) were chosen as the activation functions in the hidden and output layer, respectively.

As mentioned above and presented in Table 8, the neurons were trained from 3 to 17. Figure 12 represents the MSE values at different neurons for training, testing, and validation. The training, testing, and validation MSE were recorded with ‘R’ value for every neuron trained. The minimum MSE for training, testing, and validation are at the 16th, 10th and 8th neuron respectively with having the values of 0.00000103, 0.00002730, and 0.00002546 sequentially. The maximum MSE for training, testing, and validation are at the 3rd neuron with having the values of 0.00066024, 0.00105268, and 0.00062585 respectively.

In Figure 13(a) the errors lie in the range between 0.01918 to 0.03197 (normalized values). At the 92 epoch, the minimum value of MSE was achieved having the value of 0.00027214. In Figure 13(b) the green line shows the validation MSE, the blue coloured line shows the training MSE and the red coloured line shows the testing MSE. Figure 13(c) shows the learning process of the ANN.

$$y = \text{tansig}(x) = \frac{2}{1 + e^{-2x}} - 1$$ (40)

$$y = \text{purelin}(x) = x$$ (41)

### 2) PERFORMANCE INDICES

To evaluates the performance of curve fitting and ANN models, basic performance indices such as R, MAE, RMSE, MAPE, NS, a20-index are considered to find the accuracy of the model [79], [80]. The equations for the above-mentioned performance indices are presented in equation 42 to equation 47 below. R is the Pearson correlation coefficient, MAE is the mean absolute error, RMSE is the root mean square error, MAPE is the mean absolute percentage error, and NS is the Nash-Sutcliffe efficiency index. The accuracy of a model is defined by higher values of R and NS approaches to 1. The MAE, RMSE, and MAPE, have the lowest values, closer to zero defining the best accurate model. These performance functions were used to evaluate the model’s ability to predict R-Event accurately. The equations of R and MSE have been already mentioned in subsection 4.2, with equation no. 38 and 39 respectively.

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |x_i - y_i|$$ (42)

$$\text{MAPE} = \frac{100\%}{N} \sum_{i=1}^{N} \left| \frac{x_i - y_i}{y_i} \right|$$ (43)

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} (x_i - y_i)^2$$ (44)

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - y_i)^2}$$ (45)

$$\text{NS} = 1 - \frac{\sum_{i=1}^{N} (x_i - y_i)^2}{\sum_{i=1}^{N} (x_i - \bar{y})^2}$$ (46)

$$\text{a20-index} = \frac{m20}{N}$$ (47)

where, N is the total number of values in the experimental dataset, $x_i$ is the measured value at $i^{th}$ level, $\bar{x}$ is the mean of the experimetal values, $\bar{y}$ is the predicted value at $i^{th}$ level, $\bar{y}$ is the mean of the predicted results.

### V. RESULTS AND DISCUSSION

Curve fitting and ANN models were used in this study to develop the correlation among input parameters and R-Event. MAE and RMSE quantify the inaccuracy in parameter estimation, whereas R and NS quantify the similarity and proximity between measured and predicted values of the same parameters. The R of the ANN model for training, testing, and validation are at the 8th neuron in Figure 14. The training, testing, and validation R-value is 0.99978, 0.99632, and 0.99965 respectively. The overall R at the 8th neuron is 0.99922 as shown in Figure 14.

In this study, based on the various performance indices it is observed that ANN outperformed the curve fitting model. As shown in Figure 15, for the ANN model scatter plot between measured and predicted values all the dataset approximately lies in the range of 10% to -20% reference lines. However, for the curve-fitting model, the dataset exceeds the range of 30% to -30% reference lines. The comparison of different performance indices of both models is presented in Table 9. The ANN model’s performance is better than the curvefitting model and all the error indices are lower.
in the ANN model compared to the curve-fitting model. The correlation coefficient of the ANN model is 4.55% higher than the curve-fitting model. The MAE, MAPE, MSE, and RMSE of the ANN model is respectively 93.35%, 91.82%, 98.16%, and 86.55% lower than the curve fitting model. The Nash-Sutcliffe efficiency index (NS) and a-20 index value of the ANN model is 9.36% and 13.75% higher than the curve-fitting model.

Figure 16 shows the results obtained by the curve fitting model. The maximum error in the dataset lies in the range between -0.05 to 0.08. The frequency-error histogram (Figure 16(b)) shows that maximum error lies in between
−0.04 to 0.04 and the error frequency is distributed evenly with the frequency peak at zero.

In this study, the ANN model surpassed the curve-fitting model with a correlation coefficient (R) of 0.9992. Figure 17 shows the results obtained by the ANN model. The maximum error in the dataset lies in the range between −0.04 to 0.01. All the data sets are showing fewer errors in the ANN model compared to the curve-fitting model. The frequency-error histogram (Figure 17(b)) shows that maximum error lies in between −0.005 to 0.005 and the error frequency peak at zero. The comparison between Figure 16 and Figure 17 shows that ANN is more effective in predicting R-Event for office-built environments. The error trend between the ANN model and curve-fitting is visibly different as ANN error graphs show less deviation than the curve-fitting graph.

A. ANN FORMULATION

As stated earlier, the proposed ANN model, demonstrated high accuracy and interpretability. The explicit formula for R-Event may be directly obtained from the ANN model, which is a function of weights, biases, and associated activation functions. The weight and biases for input and output found in this study are presented in Table 10. The final equation to predict the R-Event is expressed as equation 48:

\[ R - Event = \text{purlin} \left( W_{HO}Y_i + B_{HO} \right) = W_{HO}Y_i + B_{HO} \]  (48)

The generalized formulation for input to hidden layer \( Y_i \) is represented in equation 49 and is calculated using equation 50, as shown at the top of page 22.

\[ Y_i = \left( W_{IH}N_i,\text{normalized} + B_{IH} \right) \]  (49)

where, \( N_i,\text{normalized} \) are the normalized inputs, \( W_{IH} \) is the weight of the matrix in between the input layer and hidden layer, \( B_{IH} \) is the biases between the input layer and hidden layer, \( W_{HO} \) and \( B_{HO} \) are the weights and bias of the hidden layer to the output layer.

The developed model can predict the R-Event for normalized projected data of inputs for the defined ranges. The proposed model includes eleven parameters to predict the indoor R-Event values for office buildings. After putting the values from Table 10 in equations 48 and 49 the final ANN model is presented in equation 51, as shown at the top of page 22.

VI. LIMITATIONS OF THE STUDY

This study has certain limitations. The scope of this study is restricted to office building indoor environments.

---

FIGURE 16. Results of curve fitting model (a) Variation in the predicted results (b) Frequency distribution of errors and (c) Error plot with respect to number of datasets.
Single source of infection is assumed with only airborne transmission route consideration in an office environment operating under natural ventilation mode in composite climatic condition. This research is limited by a number of technological, economic, social, and temporal restrictions. The procedure, however, stays the same and may be used in a variety of structures in a variety of climatic areas with varying environmental circumstances. Mechanical devices such as fans, air conditioning systems, exhaust systems, and so on were not used in this study and so are not taken into account. They can, however, have a significant impact on the spread of airborne infections, which restricts the scope of the study. Only diurnal fluctuations in the environment are taken into account since office time limits the observation of nocturnal variations. Healthy people between the ages of 21 and 30 were selected, with a normal breathing rate of 12-20 breaths per minute and no respiratory disorders. The data gathered is case-based and is dependent on a variety of other environmental parameters such as pressure, wind direction, and so on, which are not taken into account in this study. The behaviour of several old and new SARS-CoV-2 variants is extremely unpredictable. As a result, the human-virus interaction is outside the scope of this investigation. Human behaviour varies greatly depending on a variety of conditions, which is a restriction itself because each individual has a unique behavioural pattern based on their brain-prints.
and experiences. Most of these limitations will be addressed in the future research by the authors.

VII. CONCLUSION AND FUTURE WORK

In this study, predictive models such as ANN and curve fitting were developed and tested to predict the R-Event inside an office room. Real-time data for the office environment was gathered in the spring of 2022 in a naturally ventilated (NV) office room situated in composite climate. The main conclusions are as follows:

- Eleven parameters such as indoor temperature ($T_{in}$), indoor relative humidity ($RH_{in}$), area of opening ($A_O$), number of occupants ($O$), area per person ($A_p$), volume per person ($V_P$), $CO_2$ concentration ($CO_2$), air quality index ($AQI$), outer wind speed ($W_S$), outdoor temperature ($T_{out}$), outdoor humidity ($RH_{out}$) were used as input for the models. R-Event is the output that is used as the proxy to SARS-CoV-2 infection spread probability.
- The data were analyzed first and then curve fitting was applied, afterwards, the analyzed data were trained, tested, and validated with ANN.
- While comparing the curve-fitting and the ANN model, both the curve fitting and ANN model performed well in predicting the R-Event. However, ANN showed a stronger correlation and lesser error ($R = 0.9992$, $MSE = 0.0000035$, $RMSE = 0.00018708$, $MAE = 0.0006675$, $MAPE = 0.8643816$, $NS = 0.9984365$, and $a20$-index $= 0.9984300$) with measured results.
- The ANN model can be used to predict R-Event accurately when the data of necessary input parameters are available. This will help in saving time and effort.
- This study presents a novel computational model based on real-time collected data to predict the R-Event value as a proxy indicator for SARS-CoV-2 spreading in indoor office spaces.
- There are certain limitations of this work as the model can only predict the results according to the dataset limits. Additionally, sneezing, coughing like other respiratory activities were not taken into account in this study.
- Outer movement (near office doors/windows) of people is also an important parameter, however, due to highly dynamic conditions, this factor is not considered in this study.
- This model is created for naturally ventilated indoor settings (office rooms) where no mechanical equipment is used for ventilation; only door and window operations were used for ventilation purposes.
- The models developed in this study are useful to predict the R-Event as a proxy to SARS-CoV-2 infection probability using $CO_2$ level in similar other office environments and built spaces.

Some other general observations were also made during the study; these are as follows:
- Human activities affect the carbon dioxide concentration directly and a higher activity level increases the chances of infection spread.
- The area and volume of the office are very important for a safe working environment. However, occupancy is the prime factor that is strongly correlated with R-Event. Hence, for safe working environment occupancy must be reduced to the safe level.
- $CO_2$ levels are the second most correlated parameter to R-Event after occupancy, showing that it is very important to
ventilate indoor spaces for maintaining low CO\textsubscript{2} levels for preventing SARS-CoV-2 spread.

- People moving outside also affect the CO\textsubscript{2} levels but are negligible until they stand out in front of the door for long talks. This is a dangerous condition as this suddenly rises the CO\textsubscript{2} levels as incoming air is having high CO\textsubscript{2} levels itself.
- Maintaining all doors and windows is recommended for better operational control. Proper maintenance improves the efficiency of full-scale door and window functioning. Additionally, nets are also recommended to use for preventing the entry of insects inside the office.
- Frequent talking inside the office environment is not promoted and that to without a mask, it enhances the chances of infection spreading.
- It is recommended to install CO\textsubscript{2} sensors in offices to monitor the CO\textsubscript{2} levels.
- Cross-ventilation strategies are comparatively better than other single side ventilation scenarios to mitigate the viral transmission inside office spaces and other indoor spaces.
- Training and enforcement of guidelines significantly reduces the risk of viral transmission.

**FUTURE WORK**

The authors will extend this work to establish the relationship between CO\textsubscript{2} and viral spreading probability in different seasons and in different types of ventilation scenarios. The authors are currently gathering data for mixed-mode ventilated and air-conditioned office rooms.

**CONFLICTS OF INTEREST**

The authors declare that there are no conflicts of interest regarding the publication of this paper.

**ACKNOWLEDGMENT**

The work described in this research article is part of the first author’s AcSIR Ph.D. research work being carried out at the CSIR-CBRI, Roorkee, India. The authors are grateful to the Director at the Central Building Research Institute (CSIR), Roorkee, for his permission to publish the work. The authors would like to thank Dr. Kishor and S. Kulkarni for providing the scientific instruments used during this research work.

**REFERENCES**

[1] M. A. Rahman, N. Zaman, A. T. Asyhari, F. Al-Turjman, M. Z. A. Bhuiyan, and M. F. Zolkipli, "Data-driven dynamic clustering framework for mitigating the adverse economic impact of COVID-19 lockdown practices," *Sustain. Cities Soc.*, vol. 62, Nov. 2020, Art. no. 102372, doi: 10.1016/j.scs.2020.102372.

[2] J. Lee, "Mental health effects of school closures during COVID-19," *Lancet Child Adolesc. Health*, vol. 4, no. 6, p. 421, 2020, doi: 10.1016/S2352-4642(20)30109-7.

[3] N. Agarwal, C. S. Meena, B. P. Raj, L. Saini, A. Kumar, N. Gopalakrishnan, A. Kumar, N. B. Balam, T. Alam, N. R. Kapoor, and V. Aggarwal, "Indoor air quality improvement in COVID-19 pandemic: Review," *Sustain. Cities Soc.*, vol. 70, Jul. 2021, Art. no. 102942, doi: 10.1016/j.scs.2021.102942.

[4] D. L. Blustein, R. Duffy, J. A. Ferreira, V. Cohen-Scali, R. G. Cimamon, and B. A. Allan, "Unemployment in the time of COVID-19: A research agenda," *J. Vocational Behav.*, vol. 119, Jun. 2020, Art. no. 103436, doi: 10.1016/j.jvb.2020.103436.

[5] M. Lawson, M. H. Piel, and M. Simon, "Child maltreatment during the COVID-19 pandemic: Consequences of parental job loss on psychological and physical abuse towards children," *Child Abuse Neglect*, vol. 110, Dec. 2020, Art. no. 104709, doi: 10.1016/j.chiabu.2020.104709.

[6] A. Kumar, N. R. Kapoor, A. Kumar, A. Deep, K. S. Kulkarni, and H. C. Arora, “Public perception on SARS-CoV-2 transmission and air disinfection systems: A study," in *Proc. 2nd Int. Conf. I-Converge*. Dehradun, India: DIT Univ., Sep. 2022, pp. 1–8.

[7] World Health Organisation (WHO). Coronavirus (COVID-19) Dashboard. Accessed: Jun. 13, 2022. [Online]. Available: https://covid19.who.int/
[8] World Health Organisation (WHO). (2020). Transmission of SARS-CoV-2: Implications for Infection Prevention Precautions. Accessed: Aug. 23, 2021. [Online]. Available: https://www.who.int/newsroom

[9] L. Liu, Y. Li, P. V. Nielsen, J. Wei, and R. L. Jensen, “Short-range airborne transmission of expiratory droplets between two people,” Indoor Air, vol. 27, no. 2, pp. 362–367, Mar. 2017, doi:10.1111/ina.12314.

[10] Z. T. Ai and A. K. Melikov, “Airborne spread of expiratory droplet nuclei between the occupants of indoor environments: A review,” Indoor Air, vol. 28, no. 4, pp. 500–524, Jul. 2018, doi:10.1111/ina.12465.

[11] S. K. Das, J.-E. Alam, S. Plummeri, and V. Greco, “Transmission of airborne virus through sneezed and coughed droplets,” Phys. Fluids, vol. 32, no. 9, Sep. 2020, Art. no. 091702, doi:10.1063/5.0022859.

[12] L. Hamner, P. Dubbel, I. Capron, A. Ross, A. Jordan, J. Lee, J. Lynn, A. Ball, S. Narwal, S. Russell, D. Patrick, and H. Leibrand, “High SARS-CoV-2 attack rate following exposure at a choir practice—Skagit county, Washington, March 2020,” MMWR. Morbidity Mortality Weekly Rep., vol. 69, no. 19, pp. 606–610, May 2020, doi:10.15585/mmwr.mmrn6916e.

[13] N. R. Kapoor, A. Kumar, C. C. Meena, A. Kumar, T. Alam, N. B. Balam, and A. Ghosh, “A systematic review on indoor environmental quality in naturally ventilated school classrooms: A way forward,” Adv. Civil Eng., vol. 2021, pp. 1–19, Feb. 2021, doi:10.1155/2021/851685.

[14] B. M. Althouse, E. A. Wenger, J. C. Miller, S. V. Scarpino, A. Allard, L. Hébert-Dufresne, and H. Hu, “Superspreading events in the transmission dynamics of SARS-CoV-2: Opportunities for interventions and control,” PLOS Biol., vol. 18, no. 11, Nov. 2020, Art. no. e3000987, doi:10.1371/journal.pbio.3000987.

[15] T. R. Frieden and C. T. Lee, “Identifying and interrupting superspreading events—Implications for control of severe acute respiratory syndrome coronavirus 2,” Emerg. Infectious Diseases, vol. 26, no. 6, pp. 1059–1066, Jun. 2020, doi:10.2302/icid2606.200495.

[16] N. R. Kapoor, A. Kumar, T. Alam, A. Kumar, K. S. Kulkarni, and P. Blechich, “A review on indoor environment quality of Indian school classrooms,” Sustainability, vol. 13, no. 21, p. 11855, Oct. 2021, doi:10.3390/su13111855.

[17] D. Lewis, “Superspreading drives the COVID pandemic—and could help to tame it,” Nature, vol. 590, no. 7847, pp. 544–546, Feb. 2021, doi:10.1038/d41586-021-00460-x.

[18] S. A. Ismail, V. Saliba, J. L. Bernal, M. E. Ramsay, and S. N. Ladhani, “Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1,” New England J. Med., vol. 382, no. 16, pp. 1564–1567, 2020, doi:10.1056/NEJMra2004973.

[19] P. Shrestha, J. Wu, D. Ge, M. Zhagdajav, and X. Liu, “Multizonal modeling of SARS-CoV-2 aerosol dispersion in a virtual office building,” Building Environ., vol. 206, Dec. 2021, Art. no. 108347, doi:10.1016/j.buildenv.2021.108347.

[20] M. Guo, P. Xu, T. Xiao, R. He, M. Dai, and S. L. Miller, “Review and comparison of HVAC operation guidelines in different countries during the COVID-19 pandemic,” Building Environ., vol. 187, Jan. 2021, doi:10.1016/j.buildenv.2021.107368.

[21] W. Zheng, J. Hu, Z. Wang, J. Li, Z. Fu, H. Li, J.Jaras, S. K. Chou, and J. Yan, “COVID-19 impact on operation and energy consumption of heating, ventilation and air-conditioning (HVAC) systems,” Adv. Appl. Energy, vol. 3, Aug. 2021, Art. no. 100040, doi:10.1016/j.adapen.2021.100040.

[22] CSIR. (2021). CSIR Guidelines on Ventilation of Residential and Office Buildings for SARS-CoV-2 Virus. [Online]. Available: https://www.csit.res.in

[23] CSIR. (2022). CSIR Guidelines on Ventilation of Residential and Office Buildings for SARS-CoV-2 Virus. [Online]. Available: https://www.csit.res.in/readbook/bid=MTQ5ODcx&sub=submit=view

[24] COVID-19: Heating, Ventilation and Air Conditioning (HVAC) Systems in Buildings, Public Health Ontario, ON, Canada, 2020.

[25] REHVA COVID-19 Guidance Document, Representatives Eur. Heating Ventilating Assoc., Brussels, Belgium, 2020.

[26] Addressing COVID-19 in Buildings, Can. Committee Indoor Air Qual., Canada, 2020.

[27] HVAC specifications for operation and management of air-conditioning ventilation systems in office buildings and public places during COVID-19 epidemic. Nat. Health Commission, China, 2020.

[28] Heating, Ventilation and Air-conditioning Systems in the Context of COVID-19, Eur. Centre Disease Prevention Control, Eur. Union, 2020.

[29] AIRAE Position Document on Infection Aerosols, Amer. Soc. Heating, Refrig. Air-conditioning Eng., GA, USA, 2020.

[30] ASHRAE Issues Statements on Relationship Between COVID-19 and HVAC in Buildings, Amer. Soc. Heating, Refrig. Air-conditioning Eng., GA, USA, 2020.

[31] Start up and Operation of Air Conditioning and Ventilation Systems During Pandemic in Commercial and Industrial Workspaces, Indian Soc. Heating, Refrig. Air-conditioning Eng., New Delhi, India, 2020.

[32] Q and A on Ventilation in the Control of SARS-CoV-2 Infection, Soc. Heating, Air-conditioning Sanitary Eng., Tokyo, Japan, 2020.

[33] Guidelines for Office Buildings to Deal With ‘New Coronavirus’ Operational Management Emergency Measures, Archit. Soc. China, China, 2020.

[34] C. Ghiaus, F. Allard, M. Santamouris, C. Georgakis, and F. Nicol, “Urban environment influence on natural ventilation potential,” Building Environ., vol. 41, no. 4, pp. 395–406, Apr. 2006, doi:10.1016/j.buildenv.2005.02.003.

[35] P. O. Veld, “Introduction to EC RESHYVENT-EU cluster project on demand controlled hybrid ventilation for residential buildings,” Building Environ., vol. 43, no. 8, pp. 1342–1349, Aug. 2008, doi:10.1016/j.buildenv.2007.01.043.

[36] G. A. Somsen, C. van Rijn, S. Kooij, R. A. Bem, and D. Bom, “Small droplet aerosols in poorly ventilated spaces and SARS-CoV-2 transmission,” Lancet Infectious Diseases, vol. 8, no. 7, pp. 658–659, 2020, doi:10.1016/S1473-3099(20)30245-9.
ASHOK KUMAR is currently an Outstanding Scientist (Scientist ‘H’), a Senior Professor at AcSIR, Inventor, and an Architect-Planner. He has experience of more than 33 years. He has executed international and national projects related to IEQ and he is a Senior Member of the CSIR Task Group “WAYU” created to find indigenous solutions for ventilation problems and IAQ in COVID-19 like situations. He also invented several market-ready designs and technologies and their execution/applications to deal with IEQ problems of buildings. He has authored/coauthored over 110 research publications in international and national journals, conferences, reports, and book chapters. He is currently guiding five Ph.D. scholars and handled more than 85 research and development projects. The total costing of last five main projects are more than 30 crores. He has visited Germany, Japan, and Australia in recent years. He is a Principal/Alternative Member as an expert in more than a dozen committees of BIS, National Building Code 2016, Energy Conservation Building Code (ECBC 2017 and 2018), and ISO/WD 19467:2018. He was a Senior Member of the team that provided guidelines for reducing the spread of SARS-CoV-2 in Indian office and residential buildings. He has several awards to his credit from the various prestigious societies and organizations. He also received many “Best Paper Awards” for his brilliant publications. On the occasion of the Foundation Day of the CSIR-CBRI, he also earned the prestigious Diamond Jubilee Best Technology Award for the technology “HVAC Ducting System for Integration of COVID-19 Disinfection Solutions,” in 2021. He is also serving as the Governing Council Member for the Bureau of Energy Efficiency (BEE), GOI.

ANIL KUMAR (Senior Member, IEEE) received the M.Tech. degree from the Delhi College of Engineering, New Delhi, and the Ph.D. degree from Manipal University. He is currently working as a Professor of computer science and engineering, an Accreditation Coordinator, and the Head at the Data Science Research Group, DIT University, India. He has more than 25 years of teaching and industrial experience. He has served various reputed origination, such as Manipal University, Bharti Vidyapeeth, Mody University Science and Technology, and DRDO. He has guided over ten research scholars. He has published more than 200 research articles and patents. His research interests include image processing algorithm, cryptography, artificial intelligence, signals and systems, neural networks, genetic algorithm, and machine learning. He has done various Indian government projects as a Principal Investigator. He is a Senior Member of ACM and CSI; also worked as an Executive Committee Member of the IEEE Computer Society India Council, in 2015 and 2016, and the IEEE Rajasthan Sub-Section, in 2018. He has also been a Consultant to various industries. He is a Reviewer of many international journals of IEEE, Elsevier, Springer, and ACM.

KRISHNA KUMAR (Graduate Student Member, IEEE) received the B.E. degree in electronics and communication engineering from the Govind Balabhad Pant Engineering College, Pauri Garhwal, and the M.Tech. degree in digital systems from the Motilal Nehru NIT Allahabad. He is currently pursuing the Ph.D. degree with the Indian Institute of Technology Roorkee. He worked as an Assistant Professor at BTKIT, Dwarahat. He is currently working as a Research and Development Engineer at UJVN Ltd. He has more than 11 years of experience and has published numerous research papers in international journals, such as IEEE, Elsevier, Taylor and Francis, Springer, and Wiley. His research interests include renewable energy and artificial intelligence.