

Context-aware IoT-enabled framework to analyse and predict indoor air quality

Krati Rastogi, Divya Lohani*

Department of Computer Science and Engineering, Shiv Nadar University, Delhi NCR, India



ARTICLE INFO

Keywords:

Context aware
Indoor air quality
Internet of things
Extended Kalman filter
Discrete-time Markov chains
State of indoor air

ABSTRACT

For a productive and healthy life, air quality plays an important role. This paper addresses the requirements to develop a system capable of providing real-time information, predictions, and alerts about the indoor environment using context-awareness. The proposed IoT system serves for data collection, pre-processing, defining rules, and forecasting the predicting states of the indoor environment by giving information to the end-user about the alerts and recommendations. A novel approach based on the indoor pollutants T, RH, CO₂, PM_{2.5}, PM₁₀, and CO for the determination of the status of the environment is proposed. The pre-processing is used for filtering data using and extended Kalman filter. Further, the system uses an adaptive neuro-fuzzy inference system (ANFIS) and discrete-time Markov chains (DTMC) to predict the state of the indoor environment with the help of daily air pollution concentrations and environmental parameters. The ANFIS model predictor considers the value of indoor pollutants to form a new index: State of indoor air (SIA). For analysis and forecasting of the new index SIA, the DTMC model is used. The collected and measured data is stored in the IoT cloud using the sensing setup, and sensed information is used to develop the SIA transfer matrix, generating return durations corresponding to each SIA and providing alerts based on the data to the end-user. The models are assessed using the expected and actual return durations. The most frequent interior ventilation states, according to the predictions, are poor and moderate. Only 0.08 percent of the time does the IAQ remain in a good state. Two-thirds of the time (66%), the indoor ventilation is severe (poor, very poor, or hazardous); 19% of the time it is very bad, and 15% of the time it is hazardous, suggesting and warning that there is a very high probability of unhealthy AQI in educational institutions in the Delhi-NCR region. Performance is measured by the comparison between actual and forecasted return periods, and the forecast error for our system is low, with an absolute forecast error of 3.47% on an average.

1. Introduction

The presence of air pollutants in the surrounding atmosphere of a house, such as volatile organic compounds (VOCs), particulate matter (PM), physical and chemical elements, and inorganic compounds, is termed indoor air pollution. It is rapidly becoming a global public health issue, affecting millions of people around the world. Indoor air quality (IAQ) is one of the top five environmental and health threats, according to the Environmental Protection Agency (EPA) and its Science Advisory Board. Cities in Asia have seen rapid degradation in air quality, which has a significant impact on indoor air quality. Pollutants reach the indoor environment through windows, doors, holes in the walls, and sealants on doors and windows. The presence of harmful contaminants in indoor air is a significant danger to the occupants' health, and indoor

air quality (IAQ) is preserved to save the inhabitants from the effects of contaminants. According to the World Health Organization (WHO), people spend about 90% of their time in enclosed buildings with heating, ventilation, and air conditioning (HVAC) equipment. As a consequence, it is important to keep track of IAQ to safeguard the health and productivity of the people who live there.

Indoor air pollution defines the presence of air pollutants, consisting of particulate matter, volatile organic compounds, inorganic compounds, physical and chemical factors, etc., within and around the enclosed area. These pollutants pose a serious threat to the human body and to protect and prevent individuals from the impact of pollutants, 'Indoor Air Quality' is maintained. According to reports by the World Health Organization (WHO), people in urban areas spend about 90% of their time in indoor spaces or buildings, which are mostly centralised or

* Corresponding author.

E-mail address: divya.lohani@ieee.org (D. Lohani).

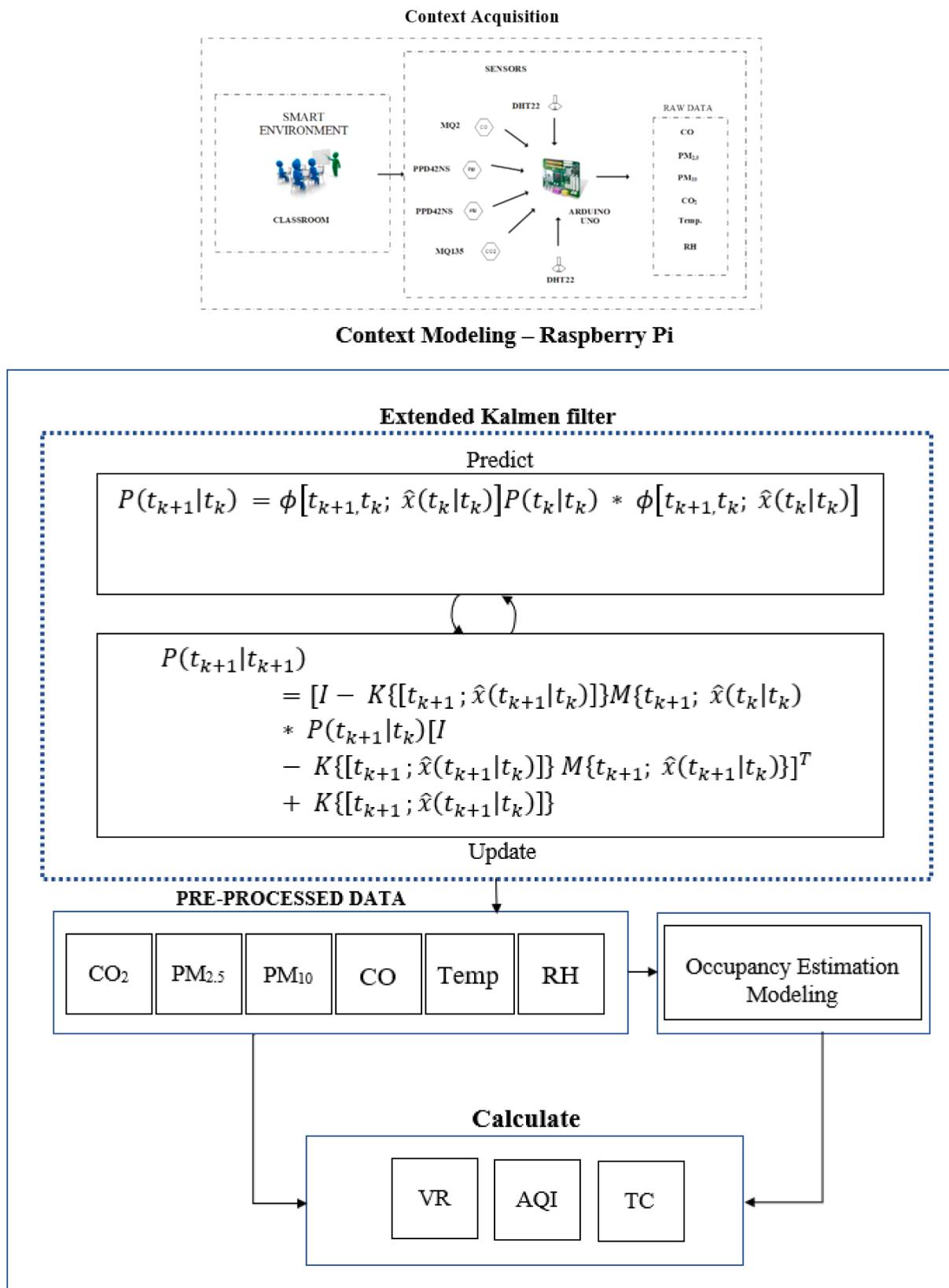


Fig. 1. Context-aware IoT system. (a) Context Acquisition IoT system (b) Context modelling IoT system (c) Context reasoning IoT system (d) Context Prediction IoT system (e) Context Recommendations / Alerts IoT system.

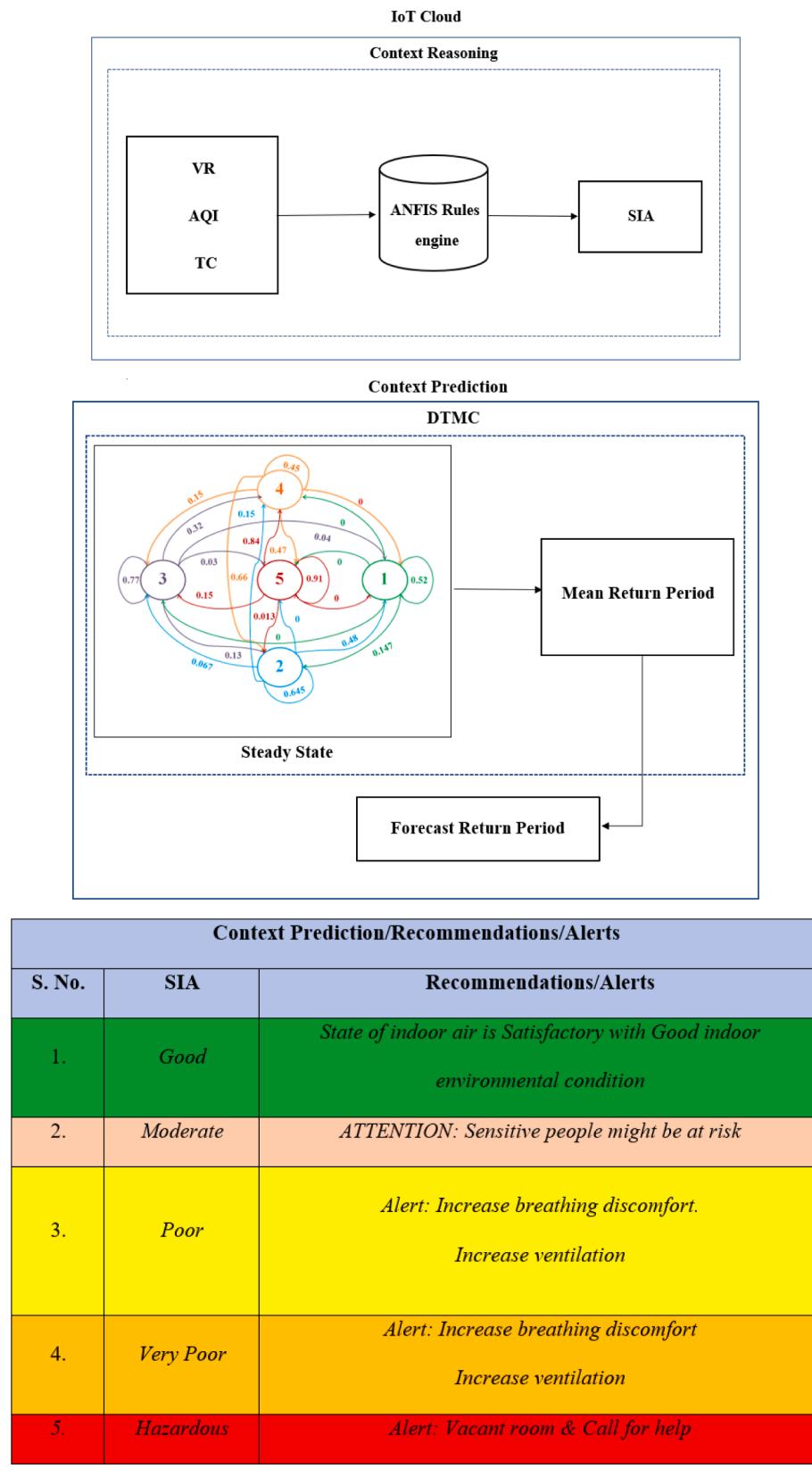


Fig. 1. (continued).

regulated using heating, ventilation, and air conditioning devices. According to data of the Environmental Protection Agency (EPA) (2016), the level of indoor pollutants is 2–5 times more polluted than that of outdoor pollutants.

Indoor air quality has become a vital parameter that needs to be emphasized in non-residential areas, especially educational institutions. Almost all students spend most of their time indoors, in school buildings, classrooms, and hostels during their educational life. In these

Context Prediction/Recommendations/Alerts		
S. No.	SIA	Recommendations/Alerts
1.	Good	<i>State of indoor air is Satisfactory with Good indoor environmental condition</i>
2.	Moderate	<i>ATTENTION: Sensitive people might be at risk</i>
3.	Poor	<i>Alert: Increase breathing discomfort. Increase ventilation</i>
4.	Very Poor	<i>Alert: Increase breathing discomfort Increase ventilation</i>
5.	Hazardous	<i>Alert: Vacant room & Call for help</i>

Fig. 1. (continued).

institutions, pollutants from several sources affect the health, comfort, and performance of the students and employees considerably, especially, due to the negative and harmful impact on the memory and concentration of students, productivity, decision-making ability, and lifestyle (Wyon et al., 2004). To build a healthy atmosphere for the students, constant monitoring of indoor air quality in classrooms is a must. Many studies have investigated the effects of good indoor air quality on human cognitive abilities and the improved performance of students and teachers (Satish et al., 2012).

Poor air quality may have adverse effects on people living mostly in an enclosed environment, especially the younger generation, which spends most of its time in an air-conditioned domain, with little scope for circulation of fresh air. Indoor air pollutants cause hazardous health issues such as Sick Building Syndrome (SBS) and Building Related Illness (BRI). Sickness and discomfort experienced by the occupants due to the Indoor Air Quality (IAQ) are related to SBS and BRI. For the health, productivity, and comfort of human beings it is important to maintain the state of air quality, it helps to make decisions when there is fresh and clean air.

Context-awareness is a core feature of ubiquitous and pervasive computing systems. A Context-aware system collects data able to understand and can be used to characterize the situation by providing a strategy regarding decision-making by collecting information as input and reaction as output. The essential task is to predict the expectations of the users and helps them improve at the right time by providing appropriate information. In the recent past, context-based information and communication amongst the data have been used to provide end-user information and alerts for indoor air quality. Most researchers have confined their work to improving air quality by using the concentration of a single pollutant (Zhou et al., 2020). Also, most systems that collect and monitor real-time information are limited to provide a system that combines environmental pollutants to form an index for the calculation of indoor air quality. Another drawback of the current air quality systems is their inability to predict and provide the end-user with information from the past values to control and take measures.

A context-aware system is presented in this work, that can adapt to the end user's requirements according to the contextual environmental parameters. The contextual information is the input, and the system's reaction following the inputs is the output. The system collects contextual data that can be used to characterize the situation by providing a strategy for decision-making. The essential task is to modify the indoor environment according to the expectation of the users and help in

improving the end user's experience by providing appropriate information at the right time. The proposed context-aware system works in 4 stages. In the first step, i.e., context acquisition, a sensing system is used to collect suitable contextual information from the indoor environment. The outputs of temperature, humidity, carbon monoxide, carbon dioxide, and particulate matter are used as contexts. The sensor output may have inconsistencies such as missing data points, errors, and outliers. The second stage of the proposed system works to eliminate such discrepancies from sensed data. An extended Kalman filter is employed for this purpose. The error-free contextual inputs are used in calculating the three indices: ventilation rate, air quality index, and thermal comfort. These indices are further combined into a single index, SIA which can comprehensively describe the state of the indoor environment. In the context prediction stage, Discrete-time Markov chains (DTMC) are used to forecast the future state of the indoor environment. These predictions are useful in estimating future energy demands, as well as in timely control of deterioration in an indoor environment.

Hypothesis

The hypotheses examined in this study are listed below:

H_0 : is there any such method that determines the real-time collection and analysis of data, including major indoor pollutants that contributes to index formulation in the invented works, which helps to decide the current status of indoor pollution with parameter classification based on the environment?

H_1 : how to predict and forecast the values of the state occurrence with the impact on human health, using an IoT-based system that monitors real-time environmental parameters, providing context-aware information?

In this research paper, the above questions are emphasized and answered in the preceding discussion.

Novelty: To maintain healthy air quality, optimum ventilation, and comfortable indoor environment, there is a need for an end-to-end system that provides context-aware information about the current ventilation, thermal comfort, and the state and indoor air. This work introduces a novel approach to providing context-aware IoT systems for modelling different situations and conditions of an enclosed environment, the status of indoor air using multiple indoor pollutants as well as environmental parameters. IAQ, thermal comfort, and ventilation issues are perceived as distinct domains, and solutions for each of these issues are provided individually. The quality of the indoor environment, however, is assessed by considering all these factors simultaneously, and not distinctively. The ANFIS model predictor considers the value of

indoor pollutants to form a new index: State of indoor air (SIA). The proposed index takes all possible factors of human health and comfort into consideration in classifying the state of the indoor environment into various categories such as good, describing the state of air as good with the least amount of pollutants present; moderate known for fair amount of pollutants present, poor and very poor states for a large number of pollutants and with a vast amount of concentration, and hazardous, describing the state with a higher concentration of pollutants. A novel method has been developed to categorize and describe the real-time indoor ventilation with the aid of contaminants that exist in the indoor atmosphere, ventilation states are established. The technique's unique feature is its ability to forecast the state of an enclosed area and help to reduce the number of pollutants present by improving the quality of air for the end-user using a context-based model.

Contribution of the Paper: The aim of this work is to create an IoT-based framework for data acquisition and context modelling based on the retrieval of a large volume of data that is not present in the IoT sensing system. As a result, these incomplete data points must be estimated if the adequate analysis is to be carried out. Extended Kalman Filter (EKF) method is used in this article based on a series of measurements over time from the setup, is used to deal with inaccuracies, missing data, and errors. The indoor pollutant concentrations obtained from EKF are used in context reasoning for the development of a new index and categorization into states. Fig. 1 shows that the context-aware system predicts the ventilation state and the human discomfort caused by it in the last phase. Context refers to the immediate environment of the individual. The features used in this work to estimate the immediate environment of an indoor occupant are the indoor pollutants, ventilation and indoor environment parameters. The parameters considered are temperature (T), relative humidity (RH), carbon dioxide (CO_2), carbon monoxide (CO), and particulate matter ($\text{PM}_{2.5}$ and PM_{10}). In this paper, six indoor factors are employed as inputs, which are required to compute thermal comfort, ventilation rate, and air quality index (AQI). Thermal comfort is represented by the percent of dissatisfied people (PPD), ventilation rate (VR) is estimated using air change rate (ACH) from CO_2 concentration, and AQI is obtained from CO, $\text{PM}_{2.5}$, and PM_{10} concentrations. The PPD, VR, and AQI values are provided to an adaptive neuro-fuzzy inference system (ANFIS) to determine the current state of IAQ and the proportion of time for which the air quality in the classroom is not healthy. The prediction of the indoor air is done with the help of the DTMC model. Real-time air pollution data is used to measure the performance (context prediction) of the system. The indoor state of air is thus validated using the confusion matrix, recall, F score, and precision. The predicted return periods are validated against the actual return periods of SIA for the testing period.

Indoor air quality prediction models use statistical/ stochastic time series modelling or machine learning techniques to provide air quality forecasts in the form of pollutant concentrations over the days ahead. However, forecasting of indoor air quality using discrete time Markov chains (DTMC) has not been widely employed. The duration in which an SIA level is surpassed twice is known as the return period. The series of potential SIA levels is described by the stochastic DTMC model, where the likelihood of each SIA level occurring depends only on the state of the preceding SIA.

To create the most accurate prediction of the future IAQ state, the present stochastic process state must be known. As very few parameters are needed to analyse such a SIA process, DTMC has been employed in this work to simulate SIA states. The SIA levels are Good, Moderate, Poor, Very Poor, and Hazardous. The predicted SIA return periods are compared to the actual return periods for the testing period to ensure accuracy.

Fig. 1 shows the different processes that are taking place in the present system. In the first phase, displayed in Fig. 1(a), context acquisition is performed. It involves the collection of indoor parameters and pollutants levels for the purpose of analysis and prediction. Context modelling is performed using an extended Kalman filter in the second

phase (Fig. 1(b)). In the third phase (Fig. 1(c)), the context reasoning phase converts raw VR, AQI, and TC values into SIA states. In the fourth phase (Fig. 1(d)), context prediction is performed using the DTMC model. In the final (Fig. 1(e)) phase, alerts and recommendations are provided to the indoor occupants.

Fig. 1 shows a context-based IoT system in 4 phases - context acquisition, context modelling, context reasoning, and context prediction. Sensor-based IoT architecture for data collection from the indoor environment pollutants parameters considered are temperature (T), relative humidity (RH), carbon dioxide (CO_2), carbon monoxide (CO), and particulate matter ($\text{PM}_{2.5}$ and PM_{10}) provide raw data for context acquisition and context modelling based on the processing of the non-negligible amount of data that is absent from the IoT sensor system. Thus, there is a need to estimate these missing data points if proper modelling is to be carried out. In this paper, an Extended Kalman Filter (EKF) algorithm, based on a series of measurements over time from the setup, is used to deal with inaccuracies, missing data, and errors. Further, the error-free values of thermal comfort, ventilation rate, and air quality index are computed using six indoor characteristics that are utilized as inputs in this study (AQI). The percentage of people who are unsatisfied with thermal comfort is measured (PPD), ventilation rate (VR) is estimated using air change rate (ACH) from CO_2 concentration, and AQI is obtained from CO, $\text{PM}_{2.5}$, and PM_{10} concentrations. The indoor pollutant concentrations changed to form parameters, PPD for thermal comfort, TC, VR, and AQI are used in context reasoning for the development of the rule of a new index and categorization into states. The PPD, VR, and AQI values are provided to an adaptive neuro-fuzzy inference system (ANFIS) to determine the current state of IAQ and the proportion of time for which the air quality in the classroom is not healthy. The context-aware system predicts the state of indoor air in the last phase. The DTMC technique is used to anticipate the status of the air within a building. The suggested system's context prediction performance also delivers suggestions and alerts to end users via the application or web servers depending on the category described by the rules created for a better, healthier, and more productive life.

The following sections describe how the proposed architecture operates. The remainder of the sections are organised as follows: A current literature analysis on indoor air quality, ventilation rate, thermal comfort and context-based air quality analysis of buildings is presented in Section II. Section III explains the IoT sensing architecture, while Section IV describes the input parameters used for the calculation of SIA. Sections V and VI explain the generic SIA calculations using ANFIS rules and the construction of the prediction model DTMC for the SIA index. Performance prediction is presented in Section VII, which includes predicting consistency and comparing forecast values to real results. The paper comes to a close with Section VIII.

2. Related work

Papers that are related to the monitoring and analysis of IAQ with the help of wireless and IoT-based systems are discussed first. Authors have reported using IoT and wireless sensor methods to monitor and analyse IAQ. Kim et al. (2014) created a system for real-time monitoring and alerting of IAQ. The complexities of designing and implementing an integrated sensing network for real-time indoor air quality monitoring, facilities, and information processing are described. The work aims to detect the concentrations of indoor pollutants, and notification of air quality is provided to the inhabitants. Saad et al. (2013) propose a web-based framework for monitoring environmental parameters and indoor gaseous contaminants. The paper discusses how data is transferred from sensors to the sink and the web-based cloud infrastructure. Spachos & Hatzinakos (2016) use particulate matter and environmental criteria to estimate the health of an enclosed room. It provides inputs on the air quality and environmental parameters. This data is used by the application to keep track of the HVAC conditions in a room. Tsang-Chu et al. (2013) have developed an air quality control device that decreased

the error rate from 15 to 7 percent. [Turanjanin et al. \(2014\)](#) described systems for calculating the ventilation thresholds using indoor pollutants and carbon dioxide concentrations.

Sick building syndrome (SBS) is often attributed to a lack of airflow and the ingestion of toxic pollutants. Association between SBS and IAQ is investigated by a few authors. [Dorizas et al. \(2015\)](#) and Norhidayah et. al. (2013) investigated the connection between IAQ and SBS. The important predictors of sick building syndromes are ventilation and accumulation of possible contaminants within the indoor environment, [Norhidayah et al. \(2013\)](#). Some authors have studied the impact of temperature and humidity on indoor air quality and thermal comfort. A device was designed by [Fang et al. \(1998\)](#) with different temperature and humidity combinations in the 18–28 °C and 30–70 percent RH ranges, respectively. A specially designed test system was built, and a series of experiments were planned to independently observe the effect of temperature and humidity on air quality/odour intensity perception and pollutant emission from the products. Different combinations of ranges of humidity and temperature were found to be linearly correlated. The air was perceived as less acceptable with increasing temperatures and humidity. [Stazi et al. \(2017\)](#) created an automated mechanism that opens a window based on IAQ and thermal comfort correlations.

[Liu et al. \(2020\)](#) have proposed a context-aware system that detects a user-specific situation and offers customized services that meet the user's requirements. Mobile operators and platforms help in providing context-awareness services smoothly and quickly. [Marco et al. \(2020\)](#) have presented a context-aware model to estimate the thermal dynamics of buildings and to predict indoor temperature evolution. The model uses an unscented Kalman filter to analyse a realistic data-set generated by Energy+. For example, CO₂ is used as a tracer gas in [Nowak-Dzieszko & Kisilewicz et al. \(2020\)](#) for computation and evaluation of the ventilation system efficiency. Simultaneous measurements of PM levels inside and outside the building help determine the infiltration factor. The authors conclude that while the ventilation is poor, the low air change rate is a protection against the infiltration of PM from the outdoor environment. [Ha et al. \(2020\)](#) have proposed a new index by combining the indoor air quality index (IAQI) and humidex called enhanced indoor air quality index (EIAQI). Waspmove sensors are used to measure indoor pollutants and temperature- humidity levels and fuse them together using an extended fractional-order Kalman filter (EKF). The obtained EIAQI is used to provide timely air quality alerts. The authors in [Ruiqiao et al. \(2020\)](#) present a new thermal comfort index based on the argument that as the predicted mean vote (PMV) does not consider the effect of solar radiation on indoor environments, it is not capable of estimating thermal comfort in solar radiation conditions. [Becerra et al. \(2020\)](#) investigates the impact of outdoor and indoor pollutants on schools. The effect of outdoor air pollution sources in the vicinity of the school, like vehicular traffic or industrial units on the indoor air quality is demonstrated.

[Chanjuan Sun et al. \(2021\)](#) conducted experiments in residential buildings using logistic regression to corroborate that di(2-ethylhexyl) phthalate (DEHP) had a significant correlation with respiratory diseases in children. [Rachel et al. \(2022\)](#) explored the feasibility of monitoring air quality using low-cost sensors. 30 low cost sensors were deployed indoors and outdoors. The authors concluded that low-cost sensors can be used effectively to monitor indoor AQI levels. [Li et al. \(2022\)](#) have proposed the use of a computational fluid dynamics (CFD) based back propagation neural network along with a particle swarm optimizer (BPNN-PSO) to predict real time indoor air quality.

[Sikora et al. \(2022\)](#) showed that when the BLDC motor is powered by an additional DC/DC converter and the control signals are derived from an external magnetic disc provides the lowest energy losses. According to [Wozniak et al.](#), heuristic optimization increases efficiency and decreases energy usage. Simulations of applied heuristic models show that such models were able to quickly optimise the system settings. The outcomes demonstrate that optimised models are more effective and

that the suggested method is to minimise voltage pulsation. [Wozniak et al.](#) developed a system based on 6 G network communication standards. The solution is designed for the next IoT level. The proposed control paradigm is effective at managing water flow, controlling windshields, addressing security concerns, and limiting carbon dioxide through adaptive ventilation.

[Nwiabu et al. \(2011\)](#) have combined context awareness, situation awareness, general domain knowledge, and case-based analysis to build a decision support system. The proposed architecture can solve specific domain problems by accepting uncertain knowledge and processing it to predict the state of the environment. The trial implementation presented in the paper is used to predict hydrate formation in under-water oil and gas pipelines with high accuracy. [Feng et al. \(2009\)](#) have presented a context-aware decision support (CaDS) system comprising of a shared situation awareness model and an entity agent for each individual user. The entity agents provide services such as event classification, proactive decision-making, and action recommendation with the help of a rule-based inference engine. A simulated command and control application is used to demonstrate the working and performance of the proposed system.

[Schürholz et al. \(2019\)](#) developed a context-aware indoor air quality monitoring and prediction system which uses contextual information such as health conditions, occupant's preferences, sensitivities, and air pollution data to provide notifications and recommendations to individual occupants. The validation of the performance of the air quality prediction algorithm is presented. [Metia et al. \(2013\)](#) have developed an algorithm based on the Extended Kalman Filter (EKF) and Kriging method to estimate air pollutant concentrations in the Sydney basin. EKF makes up for discrepancies such as noise, missing, and incorrect data, while the Kriging method serves to spatially interpolate the contour of ozone concentrations. The validation of the EKF model shows a significant improvement over station data.

The present work proposed a system that can monitor indoor parameters in real-time and notify building inhabitants to potentially dangerous circumstances. Solutions offered by authors and researchers in the recent times include models that are focussed on prediction of either IAQ, TC or VR. [Saini et al. \(2022\)](#)). [Li, L. et al. \(2022\)](#) study the indoor pollutants using an IoT based system. The authors designed CFD and ANN based generic algorithm for prediction. [Tian et al. \(2022\)](#) selected artificial neural network (ANN) models with back-propagation (BP) to predict the energy performance, air quality and thermal comfort. Thermal comfort is represented PMV and draft rate (DR), and IAQ by air age and air change efficiency (ACE), using the air supply and exhaust temperatures. But the prediction performance of these models is low. [Nan, Ma et al. \(2021\)](#) focused on models for prediction and analysis of data using different approaches. This paper reviews the analytical models and identifies the corresponding input variables, discussing their application in models based on artificial neural networks (ANNs) and reinforcement learning (RL). ANN and RL models have accurately described non-linear systems with uncertain dynamics and provided predictive and adaptive control strategies.

[Dong et al. \(2019\)](#) have provided information in managing optimal energy saving, thermal comfort, visual comfort, and indoor air quality in the built environment. The study discusses the occupants' health and comfort by measuring data in terms of energy saving, thermal comfort, visual comfort and indoor air quality. [Putra et al. \(2018\)](#) predicted indoor air quality using artificial neural networks. Past historic data from the sensing setup is analysed and artificial neural network is used for prediction. Levenberg-Method shows strong correlation between input and target values.

[Wong et al. \(2022\)](#) used different classification models, support vector machine (SVM) is used with various kernel functions. K-Nearest Neighbors (kNN), logistic regression, decision tree (DT), random forest (RF) and multilayer perceptron artificial neural network (MLP-ANN) are also explored. [Liu et al. \(2021\)](#) used various models such as Principal Component Regression (PCR), Support Vector Regression (SVR)

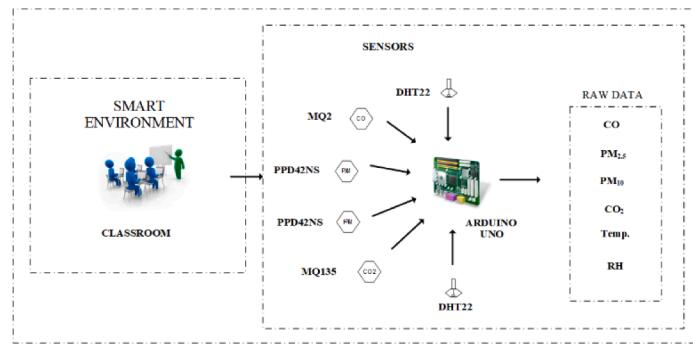


Fig. 2. Sensing setup of our IoT system in classroom.

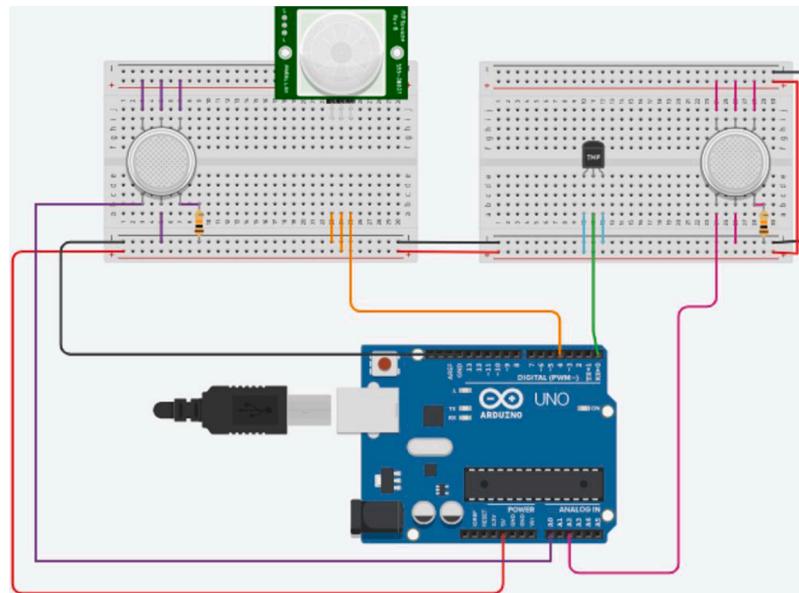


Fig. 3. Circuit Diagram of our IoT Sensing System.

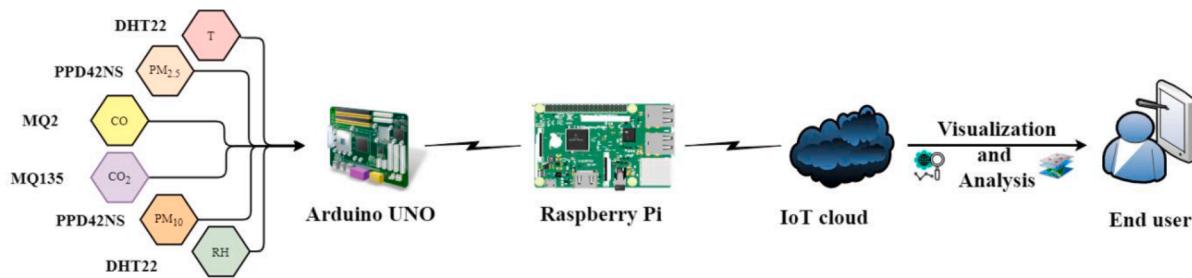


Fig. 4. Our IoT sensing system.

machine and Autoregressive Moving Average (ARMA). First, data extraction is done by principal component analysis, and then predicted values of six types of pollutants are obtained. Secondly, regressed values are generated using the SVR. ARMA model with the residual modelling corrects the residual items, and the predicted values of six types of pollutants are obtained. Hybrid model showed better performance accuracy in this case instead of using a single model.

In all these works, the authors have either used separate or combined prediction models which have been low in accuracy. The neuro fuzzy ANN model used in the present work has been able to predict the parameters with higher accuracy. Hence, its use is justified.

3. Sensing Setup and data collection

3.1. IoT Sensing setup

An IoT sensing architecture consists of sensors for continuous monitoring, collection, storage, analysis, and visualization of real time data is shown in Fig. 2. A large amount of data based on real-time generation from pollutants present in the indoor atmosphere, for real-time data storage, visualization and analysing pollutants concentrations, IoT cloud service is used. This sensing system uses a microcontroller i.e., Arduino Uno to sense and transmit indoor air pollutant concentrations to an end user by providing alerts and recommendations.

Arduino Uno is an 8-bit microcontroller with twenty input-output pins, fourteen being digital and six being analogue, a USB jack, a power connector, an In-Circuit Serial Programming header, a sixteen mega-hertz quartz crystal and a reset button. The sensors that are used to measure environmental data are connected to an Arduino Uno board. The sensing setup is fixed for the classroom and is for enclosed areas, so instead of batteries, the regular required power supply for the system. The sensing system consists of various sensors attached for measuring pollutants to microcontroller are as:

- (i) The sensor for collecting data about temperature and humidity is DHT-22 sensor having VCC pin connected to Arduino by 5 V pin, ground pins is connected to the GND of Arduino, the output is displayed by using Digital Pin 1 of Arduino connected to O/P pin of the sensor. The unit of measurement for the ambient temperature is degrees Celsius and relative humidity is percentage.
- (ii) For measuring the level of CO₂ concentration inside the room, the gas sensor used is MQ135. The pin configuration is as follows where the voltage is applied through the VCC pin of the sensor having a connection to the 5 V pin of Arduino UNO; the ground pin of sensor is connected to the ground pin of Arduino; the output is taken on pin 0 for the Arduino which is connected to the O/P of the sensor. The unit for measuring CO₂ concentration is in ppm (parts per million).
- (iii) For measuring the CO level in the room, the gas sensor used is the MQ2 and, which has a fast response time. The pin configuration is as follows: the MQ2 sensor's VCC pin connects to the 5 V pin of Arduino Uno, the ground pins of the MQ2 sensor is connected to the ground pin of the Arduino, and the O/P pin is connected to analogue Pin 2 of the Arduino UNO.
- (iv) *PM sensor*: An optical sensor PPD42NS is used for measuring the PM concentration. It records the low pulse occupancy duration, which is based on the light scattering theorem. It can detect particles with a radius up to 0.5 μm. This sensor's pin configuration is: Digital pin of PPD42NS connects to Arduino pin 4, the VCC pin is connected to the Arduino pin with 5 V, and the sensor ground is connected to Arduino ground. It has been used to collect PM_{2.5} and PM₁₀ concentrations. The measurement of PM_{2.5} and PM₁₀ is in $\mu\text{g}/\text{m}^3$.

For standardizing the pollutant concentration in the IoT sensing system, an off-the-shelf sensor for CO₂ and PM concentration is used. The CO₂ concentration is standardized using Sensordrone, and the indoor PM_{2.5}, PM₁₀ levels are standardized using the Air Visual Air Quality Monitor. These two sensors can send structured data directly to an Android app via Bluetooth and WiFi.

The Wi-Fi module ESP 8266 is utilized by Arduino Uno in this system to communicate pollution measures to the internet of things cloud where they are stored for display, analysis, and prediction. Additionally, as the whole campus is WiFi-enabled, the, ESP8266 is connected to Arduino board.

In our framework, Amazon Web Services (AWS) is a cloud computing architecture that is available on demand for large-scale data collection, retrieval, analysis, and visualization. AWS IoT is a cloud-managed platform that lets connected devices communicate with cloud applications and other devices easily and securely. AWS IoT can support trillions of devices and trillions of messages, and can reliably and securely process and route those messages to AWS endpoints and other devices.

With its broad and deep platform, Amazon Web Services provides users with the highest level of security they need. To access the data, the user will get the account ID and password. By offering certificates, public and private keys, the AWS helps connect the board to the account through Wi-Fi.

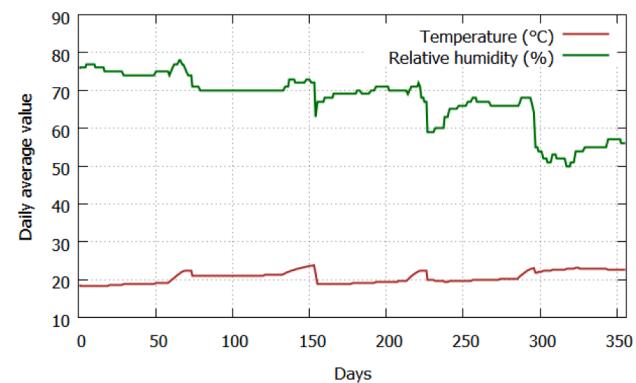


Fig. 5. Temperature and relative humidity change with time.

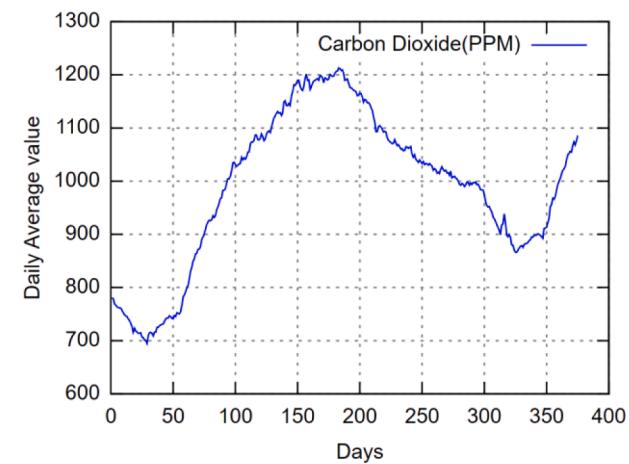


Fig. 6. CO₂ build-up with time.

3.2. Test bed and data collection

For this study, the classrooms (average size: 49.51×39.50 feet floor plan and 9.58 feet high ceiling) of the Engineering block of the University were chosen. An integrated heating, air conditioning, and ventilation (HVAC) system was installed in each classroom. Throughout the experiment, the windows and doors of the class-rooms were completely shut. The sensor equipment is positioned in the classroom such that it is neither in direct sunlight or heat, nor is it exposed to the human respiratory process or the HVAC setup. Despite the fact that the windows\ doors are shut, the setup was placed 6 feet away from the windows\ doors and at 2 feet over the level of ground.

Data was collected from 9 AM to 5 PM (Monday to Saturday), for one year and two months, from August 2018 to October 2019. The sampling rate must be kept small to capture the subtle variation, so it was kept at 1 sample per minute for indoor air pollutant concentrations. Fig. 5 shows the daily variation in indoor temperature and relative humidity for the period of observation. The air pollutant data was recorded daily, from 9:00 A.M. to 5:00 P.M., at a sampling rate of 1 per minute. The relative humidity and temperature values have been used for the calculation of PPD in determining the thermal comfort. Fig. 6 shows the daily indoor CO₂ concentrations, with the highest, minimum, and mean values of 3256 ppm, 721 ppm, and 1478.77 ppm, respectively. The daily average CO₂ concentration in this work is used in estimating the rate of ventilation (VR).

The concentration of the indoor pollutant CO did not vary much over the course of the study, with a high of 3.89 ppm, a minimum of 1.01, and a mean of 2.06 ppm. However, there is a wide range of PM 2.5 and PM₁₀ concentrations, with the highest and minimum values of 301, 365, and

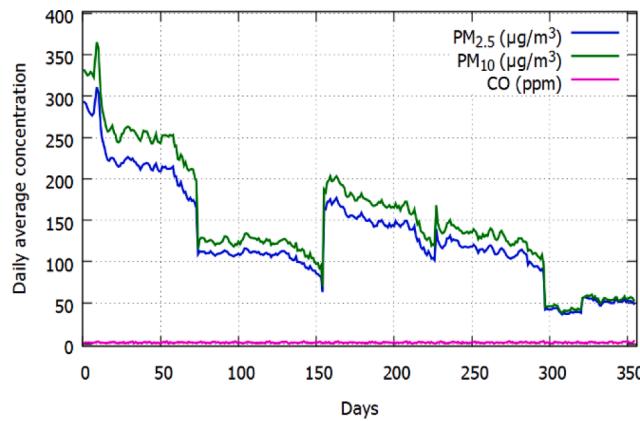


Fig. 7. Graph depicting the variation of indoor CO, PM_{2.5}, and PM₁₀.

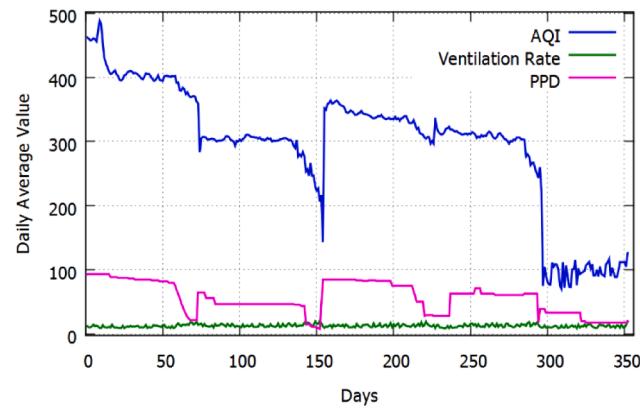


Fig. 8. Variation of daily average indoor PPD, VR, and AQI.

32, 35, respectively. Also, PM_{2.5} and PM₁₀ concentrations exhibit a strong and positive correlation, with the Karl Pearson correlation being 0.93. These values have been used for the calculation of indoor AQI.

Correlation coefficient using Karl Pearson between PM (PM_{2.5} and PM₁₀)

Karl Pearson Correlation coefficient determines the relationship amongst the particulate matter concentrations (PM_{2.5} and PM₁₀), and is calculated using the formula:

$$\rho = \frac{n * \sum a_i b_i - \sum a_i \sum b_i}{\sqrt{n \sum a_i^2 - (\sum a_i)^2} \cdot \sqrt{n \sum b_i^2 - (\sum b_i)^2}} \quad (1)$$

Where (a_i, b_i) is the ith pair observation value of PM_{2.5} (a) and PM₁₀ (b) respectively, i = 1,2,3, . . . ,n.

From August 2018 to October 2019, Figs. 6–8 depicts the relative variance in daily average VR (calculated using daily CO₂ concentration), PPD (calculated using temperature and relative humidity), and AQI (calculated using PM₁₀, PM_{2.5} and CO concentration).

3.3. Threats

3.3.1. Threats to internal validity

(i) Failure to Randomize: The input dataset consists of data points that represent the indoor air pollutants and the environmental parameter values. Stratified random sampling has been used to assign the data points randomly to the prediction model so as to get rid of the correlation between adjacent values. It involves separation of the population, before sampling, into uniform subgroups. The population should be divided up according to the

strata. Each component of the population is assigned to a single stratum in order for it to be mutually exclusive and collectively exhaustive. Then each stratum is subjected to basic random sampling. Reducing sampling error is the goal in order to increase sample precision.

- (ii) Failure to Follow the Treatment Protocol: In order to reduce sampling error, observation samples are randomly assigned to either training or testing sets in the experimental dataset. Stratified random sampling is used for this purpose.
- (iii) Attrition: The relationship between the independent and dependent variables of the prediction model may be impacted by the attrition bias. It can either make correlations between variables appear to exist when they do not, or vice versa. As the samples for training and testing are randomly selected, there is no threat to validity due to attrition bias.
- (iv) Experimental Effects: To avoid any bias in readings while collecting the data, the sensing node was kept in the centre of the lecture room, at a height of few feet, and also away from doors, windows or air conditioner vents. The way the experiment was conducted, i.e., data collection, did not affect the measurement of sensor data in any way.
- (v) Small Sample Sizes: The dataset consists of indoor parameter values that have been recorded for a period of one year and two months, from August 2018 to October 2019. It was collected at a sampling rate of 1 reading per minute. The same dataset has been used for creating training and test datasets. Hence, the size of the dataset is sufficiently large and the problem of small sample size does not arise.

3.3.2. Threats to external validity

Non-representative Sample: As the samples from the same dataset are randomly used for training and testing, the issue of non-representative samples does not arise.

4. Context modelling

Context modelling reduces the complexity by enabling and providing relatively accurate estimates. Moreover, context modelling is necessary to know about the system and its uses, high performance, and efficiency; it helps to handle the imperfect data which are inaccurate, repeated, or missing data. Also, context modelling helps to deal with the heterogeneity of the context data. One of the biggest challenges of IoT is being able to integrate and provide an accurate noise-free vast amount of heterogeneous data. To address this heterogeneity, it is necessary to use accurate modelling techniques, such as the Extended Kalman filter, to provide a formal way for the data. EKF provides simplicity (Ljung, 1979).

The parameters used in the above sections are highly uncertain and nonlinear. The use of Extended Kalman Filter (EKF) can be a better option in that case. Noise in sensed data indicates that the results are unreliable (Zhao & Wang, 2011). The current mean and covariance are linearized using a Kalman filter. However, a significant quantity of data may be missing from those obtained IoT sensor sets, necessitating the requirement to compute this missing data in order to execute accurate modelling. The Extended Kalman Filter is used to estimate the pollutant concentrations of time-varying systems, which are indirectly observed through noisy measurements or missing data, in order to improve the performance of the indoor pollutant concentration estimation model. The state of the system refers to the concentrations of dynamic factors such as CO₂, CO, PM_{2.5}, PM₁₀, temperature, and humidity. The pollutant data collected from sensors is portrayed as a time series, where concentration uncertainly changes over time, thus following a stochastic process. The sequence of pollutant values is randomly distributed over time and hence can be modelled as a stochastic differential equation. Assuming that the suggested model's stochastic differential equation is subjected to a white noise process N(t), we get:

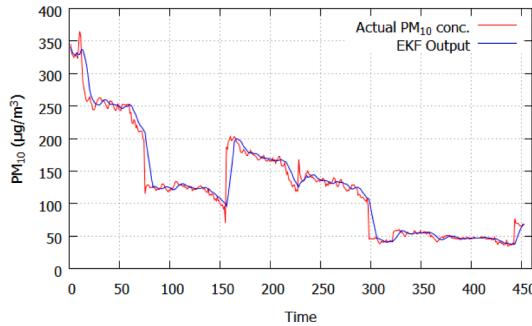


Fig. 9. Actual & Extended Kalman Filter PM₁₀ Concentration change with time.

$$\frac{dX_t}{dt} = f(X_t, t) \quad X_{t_0} \sim N(\hat{X}_{t_0}, P_{t_0}) \quad (2)$$

where, $X_{t_0} \sim N(\hat{X}_{t_0}, P_{t_0})$ represents the Gaussian random variables with mean \hat{X}_{t_0} , initial state vector X_{t_0} , and the error covariance P_{t_0} . The observation vector y_{t_k} may be expressed as:

$$y_{t_k} = h(X_{t_k}, t_k) + V_{t_k} \quad (3)$$

where, V_{t_k} is a white noise Gaussian processes vector with zero mean and $E[V_{t_k}V_{t_1}^T] = R(k)\delta_{kl}$; and δ_{kl} = the Kronecker delta. Eqs. (1) and (2) are the basis of the extended Kalman filter (EKF) algorithm. The observation data from the sensor, y_{t_k} is processed with the error covariance matrix $P(t_0|t_0)$ and the initial state vector, $X(t_0|t_0)$, to estimate the state vector, $X(t_{j+1}|t;t)$, and the error covariance, $P(t_k|t_k)$, from the Eqs. (3)–(7), as follows:

$$\hat{x}(t_{k+1}|t_k) = \hat{x}(t_k|t_k) + \int_{t_k}^{t_{k+1}} f[\hat{x}(t|t_k), t] dt \quad (4)$$

$$P(t_{k+1}|t_k) = \phi[t_{k+1}, t_k; \hat{x}(t_k|t_k)] P(t_k|t_k) * \phi[t_{k+1}, t_k; \hat{x}(t_k|t_k)] \quad (5)$$

$$\hat{x}(t_{k+1}|t_{k+1}) = \hat{x}(t_{k+1}|t_k) + K[t_{k+1}; \hat{x}(t_{k+1}|t_k)] * \{y_{t_{k+1}} - h[\hat{x}(t_{k+1}|t_k), t_{k+1}]\} \quad (6)$$

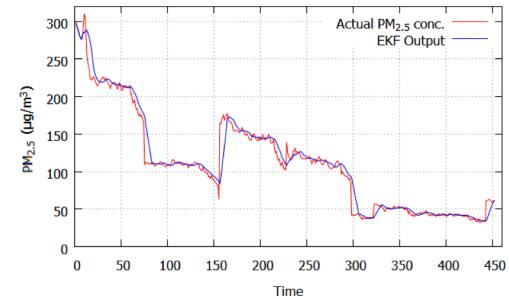


Fig. 10. Actual & Extended Kalman Filter PM_{2.5} Concentration change with time.

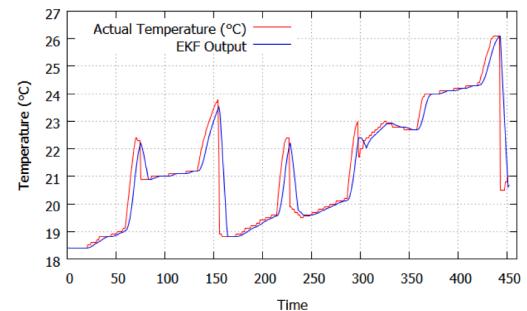


Fig. 11. Actual & Extended Kalman Filter Temperature Concentration change with time.

$$M[t_k; \hat{x}(t_k|t_k)] = \left[\frac{\partial h_i(X_{t_k}, t_k)}{\partial x_j} \right] \text{ at } X_{t_k} = \hat{x}(t_k|t_k) \quad (9)$$

here, h_i = the z'th component of $h(X_{t_k}, t_k)$, I = a unit matrix and X_j = the i'th component of the vector, X_{t_k} . Taylor's expansion of first order is used in the ϕ matrix in this algorithm as follows:

$$\begin{aligned} F[t_k; \hat{x}(t_k|t_k)] &= \left[\frac{\partial f_i(X_{t_k}, t_k)}{\partial x_j} \right] \text{ at } X_{t_k} = \hat{x}(t_k|t_k) \phi[t_{k+1}, t_k; \hat{x}(t_k|t_k)] \\ &= I + AF[t_k; \hat{x}(t_k|t_k)] \end{aligned} \quad (10)$$

$$\begin{aligned} P(t_{k+1}|t_{k+1}) &= \\ &[I - K\{[t_{k+1}; \hat{x}(t_{k+1}|t_k)]\}] M\{t_{k+1}; \hat{x}(t_k|t_k)\} * P(t_{k+1}|t_k) [I - K\{[t_{k+1}; \hat{x}(t_{k+1}|t_k)]\}] M\{t_{k+1}; \hat{x}(t_{k+1}|t_k)\}^T \\ &+ K\{[t_{k+1}; \hat{x}(t_{k+1}|t_k)]\} \end{aligned} \quad (7)$$

where A = sampling interval of observation states. The parameters to be estimated are incorporated as additional state variables in the state

$$K\{[t_{k+1}; \hat{x}(t_{k+1}|t_k)]\} = P(t_{k+1}|t_{k+1}) M^T[t_{k+1}; \hat{x}(t_{k+1}|t_k)] * M\{t_{k+1}; \hat{x}(t_{k+1}|t_k)\} P(t_{k+1}|t_k) * M\{t_{k+1}; \hat{x}(t_{k+1}|t_k)\} + R(k+1)^{-1} \quad (8)$$

where, $P(t_k|t_k)$ = covariance matrix of error in $\hat{x}(t_k|t_k)$, $\hat{x}(t_k|t_k)$ = state estimate at t_k given Y_{t_k} ; $\phi[t_{k+1}, t_k; \hat{x}(t_k|t_k)]$ = state transfer matrix from t_k to t_{k+1} ; $K[t_{k+1}; \hat{x}(t_{k+1}|t_k)]$ = Kalman gain matrix at t_{k+1} ; y_{t_k} { $y_{t_0}, y_{t_1}, \dots, y_{t_k}$ } = observation at t_k ; and

vector in the identification Eqs. (2)–(8). The EKF algorithm for state estimation can be directly used for identification.

Figs. 9–14 show the daily variation in indoor pollutant concentrations and corresponding Extended Kalman filter (EKF) outputs for the period of observation. As shown in Figure, the extended Kalman filter concentrations are low as compared actual concentration of data

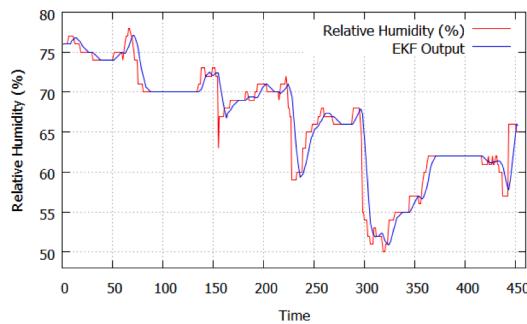


Fig. 12. Actual & Extended Kalman Filter Relative Humidity change with time.

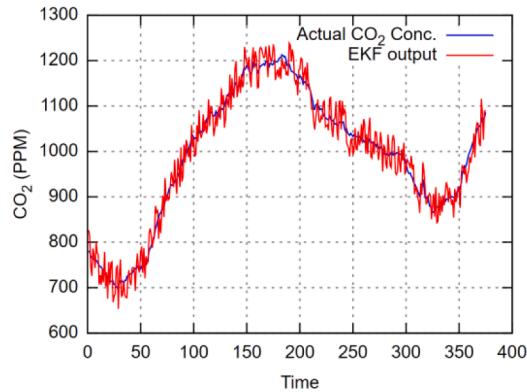


Fig. 13. Actual & Extended Kalman Filter CO₂ Concentration change with time.

implying that EKF reduced the data uncertainty from the collected pollutants' data. As is evident from the images, the EKF provides the best estimate from the noisy sensor data by filtering out the noise. It is noticed from Figs. 13 and 14 that the peak values of EKF outputs are lower than that of actual sensor concentrations because the EKF performs real-time outlier detection and correction.

5. Context reasoning

Context reasoning is a process of deriving the user's current situation from the incoming data from the context model, pollutant concentration, to generate rules on the data. It consists of ANFIS based rules generation which helps to maps the data for the determination of the state of indoor air quality.

5.1. Determination of SIA index

Data after removal of noise, smoothed data is used to categorize for calculation of ventilation rate, thermal comfort, and air quality index. Three prominent air pollutants PM_{2.5}, PM₁₀, and CO are used to

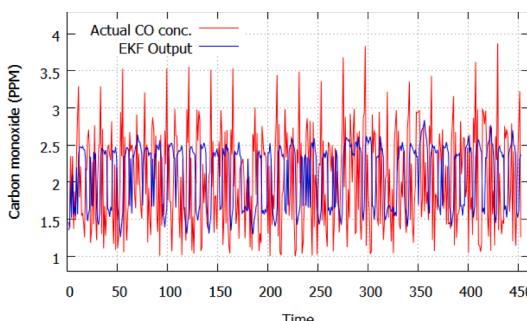


Fig. 14. Actual & Extended Kalman Filter CO Concentration change with time.

Table 1
ASHRAE Thermal Comfort Index [ANSI/ASHRAE Standard 55-2010 (2010)].

PPD	Category
100 - 75	Satisfactory
75 - 50	Moderate
50 - 25	Poor
25 - 10	Very Poor
10 - 0	Severe

Table 2
ASHRAE VR category (ANSI/ASHRAE Standard 62-2001, 2003).

Range	VR Category
>15	Satisfactory
13-15	Moderate
11-13	Poor
9-11	Very Poor
<9	Severe

determine the AQI of indoor space (Rastogi et al., 2021). The ventilation rate is determined using the indoor CO₂ concentration. The temperature and relative humidity values in the room are used to evaluate the thermal comfort of occupants with the help of PPD.

5.1.1. Thermal comfort

The basic indices of thermal comfort are predicted mean vote (PMV) and predicted percentage dissatisfied (PPD). The main factors of discomfort are temperature and relative humidity. PMV is defined by the six most important thermal variables: human activity level, clothing insulation, mean radiant temperature, humidity, temperature, and indoor air velocity. PMV is used as a thermal comfort index that indicates the mean thermal sensation value on a standard scale for a large group of persons. ASHRAE defines the PMV scale with levels for thermal comfort: +3 (very hot) +2 (warm) +1 (slightly warm), 0 (neutral), -1 (cool); -2 (cold) -3 (very cold). On this scale, the PMV index is between -3 and +3. PMV is computed using the equation [Al Horr et al., 2016]:

$$\begin{aligned} PMV = & (0,303 e^{-0.036M} + 0,028) * \{ (Met - W) - 3,05 * 10^{-3} \\ & * [5733 - 6,99 (Met - W) - p_a] - 0,42 * [(Met - W) - 58,15] \\ & - 1,7 * 10^{-5} * Met(34 - t_a) - 3,96 * 10^{-8} f_{cl} * h_c (t_{cl} - t_a) \end{aligned} \quad (11)$$

where,

Met- Metabolic rate is measured in met/m^2 .

W- Work done externally measured in W/m^2 , zero for sedentary work like receptionist or driver.

f_{cl} - the ratio of the body covered to the clothed to that of the nude body area.

t_a - ambient temperature measured in °C.

p_a - water vapour or humidity measured in Pa.

h_c - convectional heat transfer coefficient measured in (W/m^2K) .

t_{cl} - temperature while clothing on surface measured in °C.

Assuming that all students are in light cotton formal clothing and the classroom is closed with students seated, W is considered as 0, metabolic rate as $58 W/m^2$, the classroom ambient temperature is p_a , f_{cl} is considered 10, h_c is taken as 10.45, and t_{cl} is equal as t_a (Schiavon et. al. 2017; Djamilia et. al., 2017). Thermal dissatisfaction is represented as a percentage using PMV, to obtain the PPD index, which indirectly indicates the thermal comfort satisfaction. PPD is calculated using the equation below:

$$PPD = 100 - 95 * EXP(-0.03353 * PMV^4 - 0.2179 * PMV^2) \quad (12)$$

where PPD Predicted Percentage Dissatisfied and PMV predicted mean

Table 3

AQI Categories[Central Pollution Control Board of India, "Air Quality Standards," 2014].

State	Health Concern	AQI
State 1	Satisfactory	≤ 100
State 2	Moderate	101 – 200
State 3	Poor	201 – 300
State 4	Very Poor	301 – 400
State 5	Severe	> 400

vote. The comfort level is determined by combining PMV and PPD, according to ASHRAE 55 and ISO 7730. It is a seven-point scale (from -3 to +3 PMV values) used to estimate the mean reaction of a considerable assembly of occupants to perceive the thermal comfort level (Rastogi et al., 2020a). The relationship between PPD and thermal comfort is described in Table 1. The ASHRAE thermal sensation criterion for acceptability of global thermal comfort as:

5.1.2. Ventilation rate

The CO₂ concentration increases if the optimal amount of air from the outside air is not mixed with the air present inside the building. It also indicates an occupied indoor space, indicating if the building's air exchange balance is appropriate with the exchange of fresh air. The production of CO₂ concentration occurs throughout the respiratory processes of human breathing. An adult's exhaled breath creates 35,000–50,000 parts per million of CO₂ (WSUEP13-005, 2013). The number of inhabitants in a higher indoor pollution concentration is proportional to the level of outside air and ventilation rate. The CO₂ concentration in a confined environment may be tracked over time using

a mass balance equation. The mass balance equation is used to represent the temporal evolution of CO₂ concentration in an enclosed environment (Cheng et al., 2012):

$$dC_i(t)/dt = -(C_i - C_a) * \lambda + E/V_R \quad (13)$$

Where C_i is the concentration of tracer gas in indoor space, E is the quantity of tracer (CO₂) gas produced per unit time, V_R is the classroom volume, ACR is the change in air rate, and t is the period.

ACR stands for air change rate which is defined as the ratio of the volume of indoor air added or removed from a classroom in one unit of time to the air-volume of the lecture room (Rastogi & Lohani, 2019a). The following equation describes how CO₂ accumulates in the atmosphere:

$$C(t) = C_0 \exp(\lambda t) \quad (14)$$

The linear connection between the level of tracer gas CO₂ C_i(t) and the time elapsed 't' is determined by taking the logarithm on both sides (Rastogi & Lohani, 2020):

$$\ln C_i(t) = \ln C_0 + \lambda * (t) \quad (15)$$

The mentioned equations were calculated using the linear regression technique. The Eq. (15) is thus reduced to:

$$\ln(C(t) - C_a) = \ln(C_o - C_a) + \lambda * t \quad (16)$$

Where carbon dioxide level in ppm, at time 't', measured in hours, is C(t), Background and starting carbon dioxide levels are C_a and C_o, respectively.

The value of λ is used to determine VR calculated in Eq. (17). Cubic feet per minute (CFM) can be expressed to calculate VR as (Cheng et al.,

Table 4
Rules for calculation of proposed sia index using PPD, VR, AND AQI.

S. No.	Rule	SIA
1.	If ((0 ≤ PPD ≤ 10) && (VR ≥ 15) && (AQI ≤ 100))	Good
2.	Else If ((10 ≤ PPD ≤ 25) && (13 < VR < 15) && (100 ≤ AQI ≤ 200))	Moderate
3.	Else If ((25 ≤ PPD ≤ 50) && (11 < VR < 13) && (201 ≤ AQI ≤ 300))	Poor
4.	Else If ((50 ≤ PPD ≤ 70) && (9 < VR < 11) && (301 ≤ AQI ≤ 400))	Very Poor
5.	Else ((PPD ≥ 70) && (VR < 9) && (AQI > 400))	Hazardous

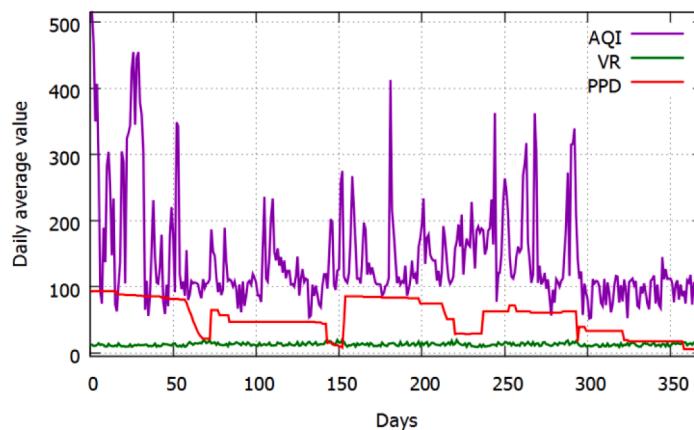


Fig. 15. Variation in average indoor VR, PPD and AQI on a daily scale.

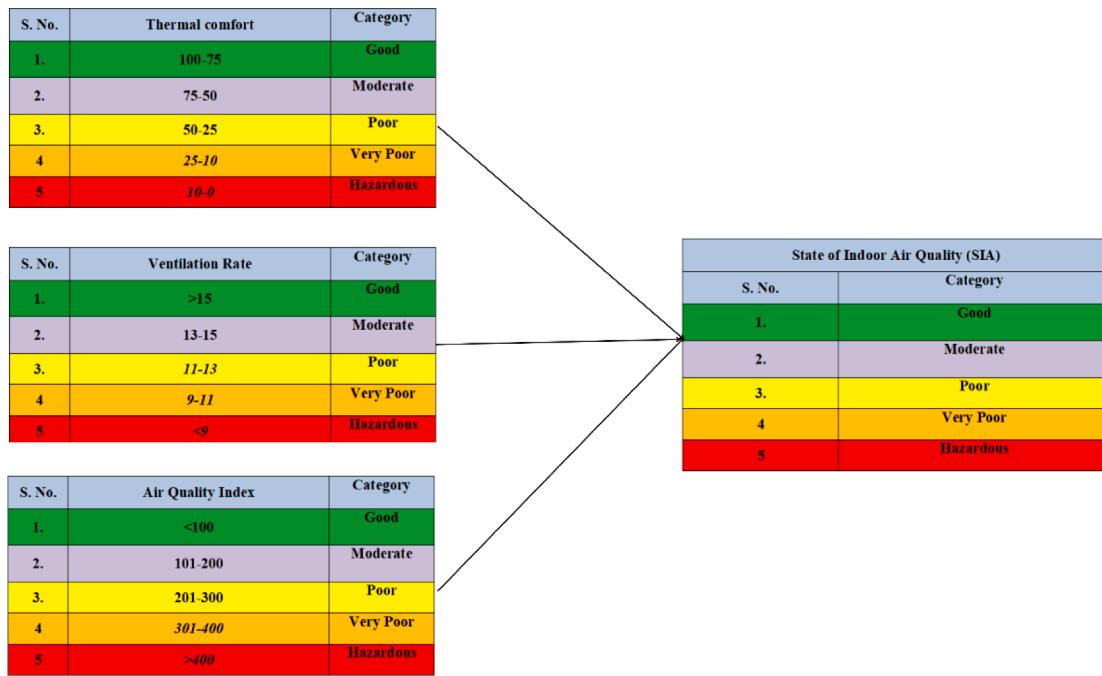


Fig. 16. Membership function for the formation of SIA.

2012; Lohani & Acharya, 2016):

$$cfm = \frac{\lambda * \text{Volume of room (cubic feet)}}{\text{no. of persons} * 60} \quad (17)$$

The volume of the classroom for this system under consideration is $49.51 \times 39.50 \times 9.58$ feet. For estimation of number of persons, the system consists of real-time analysing of data, the occupant's count is kept on changing and estimated using the regression-based models between dependant (RH and CO₂) and independent variables, the independent variable being the occupancy to be estimated (Shih, 2014). For indoor environments, air for ventilation is recommended as more than or equal to 15 cfm (ANSI/ASHRAE Standard 62–2001, 2003).

5.1.3. Indoor air quality index

The AQI was divided into five groups for this investigation and functioned as the output space for our investigation. The two AQI classifications of India's National Air Quality Index, Good (50) and

Satisfactory (51–100), have been combined into one AQI state, Satisfactory (100), to reduce the AQI state space and simplify the model. The Air Quality Index is a scale that measures the present state of quality of air indoors in terms of its influence on human health and the environment. It informs residents on how clear or contaminated the indoor air is, as well as the health implications of different levels of air quality and contaminants (Rastogi k., Barthwal A., Lohani D., Acharya D. et. al. 2020). When the AQI value is greater, the IAQ is worsened, and the health risks are increased. Furthermore, because National Air Quality Index does not specify a state for AQI values >500 , all air quality values exceeding 400 are deemed severe AQI. The output state space of the proposed model is shown in Table 3.

At least three of the eight pollutants, PM₁₀, PM_{2.5}, NO₂, SO₂, CO, O₃, NH₃, and Pb, are used to calculate the AQI. A pollutant's sub-index is a mathematical function of the pollutant's real time atmospheric level. The total AQI is the poorest sub-index value. Here, the ability to convert various indoor air pollutants into a single AQI index is advantageous since it eliminates the need to create forecast models for each pollutant. The following equation is used to compute a pollutant's sub-index:

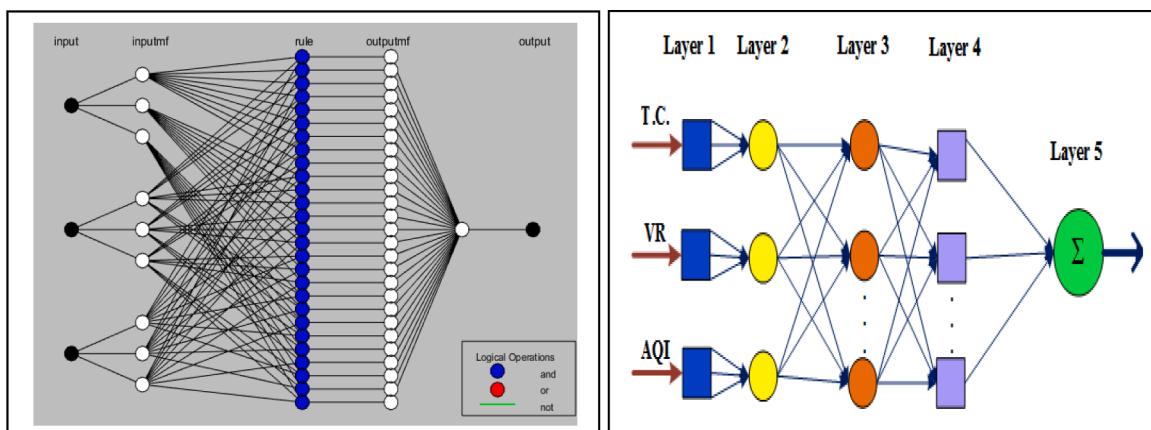


Fig. 17. Structure representation of ANFIS. (a) ANFIS structure generated by MATLAB, (b) Block diagram of ANFIS.

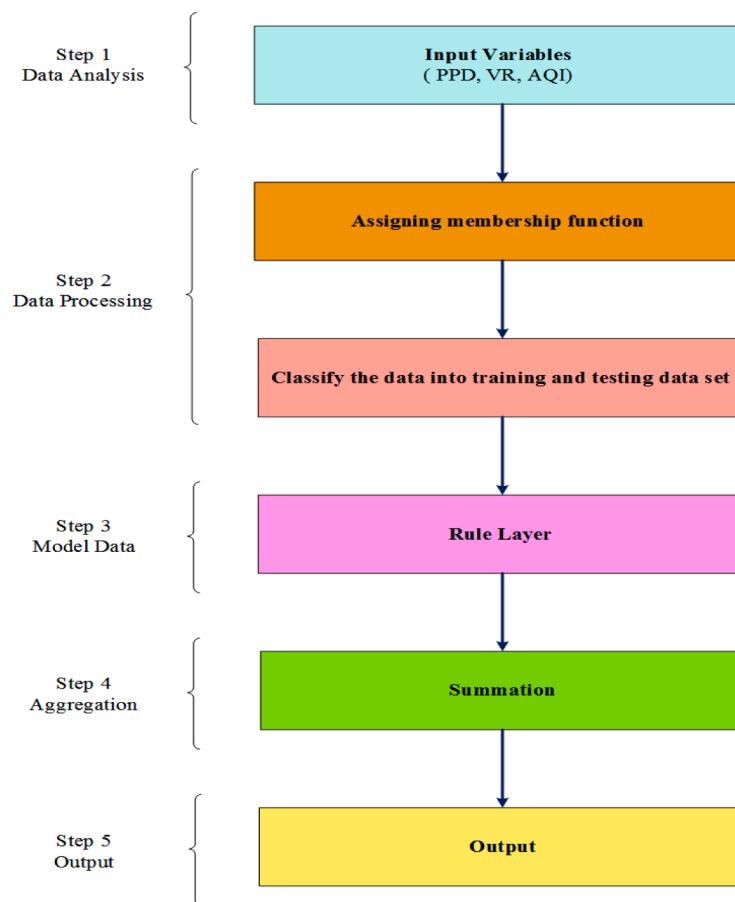


Fig. 18. Flow chart explaining the functioning of the ANFIS system.

$$I_p = C_p - BP_{LO} \times \frac{I_{HI} - I_{LO}}{BP_{HI} - BP_{LO}} \quad (18)$$

where,

I_p is the index for pollutant 'p', C_p is the index for the pollutant 'p', BP_{HI} is the breakpoint that is greater than or equal to C_p , BP_{LO} is the breakpoint that is less than or equal to C_p , I_{HI} is the AQI value corresponding to BP_{HI} , I_{LO} is the AQI value corresponding to BP_{LO} .

From August 2018 to October 2019, Fig. 15 depicts the variance of daily VR (calculated using CO_2), indoor PPD (calculated using temperature and relative humidity) and AQI (calculated using average indoor PM_{10} , $PM_{2.5}$ and CO concentrations).

5.3. Calculation of the proposed SIA index using an adaptive neuro-fuzzy inference system

The concentration is sent to the cloud for storage, and formation of a new index, ANFIS combines the air pollution, ventilation, and thermal comfort indicators AQI, VR, and PPD to bring out the state of indoor air (SIA). SIA expresses the state of indoor air in terms of five levels - good, moderate, poor, very poor, and hazardous. The computation of AQI, VR, and PPD, and the rules that ANFIS uses to determine SIA using these three parameters are described in detail in the following paragraphs.

An adaptive neuro-fuzzy inference system (ANFIS) is a hybrid system composed of a Fuzzy Inference System (FIS) with ANN features (Jang, 1993). ANFIS generates a stipulated input-output pair with a membership function based on FIS. The FIS part performs various functions of converting the value to a crisp fuzzy set and assigning the membership function, rules, and decision unit. The mathematical model of the process is not needed, and the process description with fuzzy rules emulates human thinking. Like FIS, the ANNs do not require a mathematical

model of the process. In addition to FIS, ANNs offer learning/adapting capabilities (Prasad et al., 2016). The resultant neuro-fuzzy system is capable of learning new rules or membership functions, to optimize the existing ones. A multilayer feed-forward network that uses input to map neural network learning algorithms and fuzzy reasoning into an output (Rastogi et al., 2020a). The architecture of a typical ANFIS with three inputs, rules, and one output using the Takagi-Sugeno-Kang (TSK) model, where each input is assumed to have membership functions (MFs) is shown in Fig. 16.

The sensor data is used to calculate the thermal comfort (TC), ventilation rate (VR) and air quality index (AQI) in real time. As shown in Fig. 16, the TC, VR and AQI values are fed to the Takagi-Sugeno-Kang inference model, which is a type of ANFIS architecture. The TC, VR and AQI values are aggregated in accordance with the rules defined in the Takagi-Sugeno-Kang (TSK) model and the output of the model is the state of indoor air (SIA).

The neural network model consists of the input, hidden, and the output layers. Corresponding to the three inputs, thermal comfort, AQI and ventilation rate, there are nine neurons in the input layer of ANN. Three training techniques, including Levenberg-Marquardt (LM), Resilient backpropagation (RP) and Scalar conjugate gradient (SCG) with activation functions including logsig, tangsig, and purelin, and a variety of neurons, including 2–20, were assessed for obtaining the best ANN model. The optimal number of neurons in the hidden layer is dependant upon how complex the system being approximated is Jang (1993). The LM 30-neuron method with tangsig activation function in the hidden layer and logsig function in the output layer had the lowest coefficient of correlation (R^2) values in terms of statistical performance. The hidden layer and the output layer have 30 neurons each.

Fig. 18 depicts the flowchart to explain the working of the ANFIS system. The indoor air parameters are used as inputs to obtain the



Fig. 19. Rules of the ANFIS logic controller and Simulink environment that dictate SIA output.

secondary variables PPD, VR, and AQI in the first step. The secondary variables PPD, VR, and AQI are used to express linguistic values in the form of fuzzy sets in Step 2, which are represented by the trapezoidal membership functions. The membership functions can be constructed with a range of the input and output variables classified into a different number of linguistic values as good, medium poor, very poor, and hazardous. The data sets are divided into two parts: training dataset and testing dataset of 80%–20% weightage. To generate FIS system, membership numbers and types were assigned to each input and output variable. For each input combination, the model data set is trained and test. Rules are defined in Step 3 according to the variables, and the logical AND is used in this case. Each function of the variable has rules defined for it. The rule values are then added to form the proposed SIA index value which is obtained by rules.

The working of each ANFIS layer is summarized and explained as follows.

The working of each ANFIS layer is summarized and explained as follows.

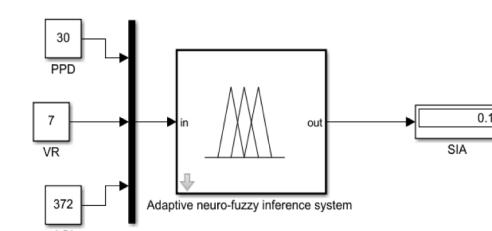
Fig. 18 shows the flow diagram for the data. In the first step, the sensor data is used to receive PPD, VR and AQI. Membership function is assigned to the inputs in the second step. The data is then divided into training and testing sets. The third step involves defining the rules of the fuzzy system. In the fifth step, the data is aggregated. The output is received in the form of ventilation class in the final step.

For layer 1, membership values for inputs can be generated by all nodes. The outputs of this layer are given by:

$$Q_{Ai}^1 = \mu_{Ai}(x), \quad i = 1, 2, 3, \dots, N \quad (19)$$

$$Q_{Bi}^1 = \mu_{Bi}(y), \quad i = 1, 2, 3, \dots, N \quad (20)$$

Where x and y are crisp inputs, MFs are characterized by hazardous, very poor, poor, medium, and good values of the trapezoidal function. The current study utilized trapezoidal type MFs:



$$0, (x < a) \text{ or } (x < b)$$

$$\frac{x-a}{b-a}, \quad a \leq x \leq b$$

$$\mu_{Ai}(x) = \begin{cases} \frac{x-b}{b-c}, & b \leq x \leq c \\ 1, & c \leq x \leq d \\ \frac{e-x}{e-d}, & d \leq x \leq e \end{cases} \quad (21)$$

$$0, (y < a) \text{ or } (y < b)$$

$$\frac{y-a}{b-a}, \quad a \leq y \leq b$$

$$\mu_{Bi}(y) = \begin{cases} \frac{b-y}{b-c}, & b \leq y \leq c \\ 1, & c \leq y \leq d \\ \frac{e-y}{e-d}, & d \leq y \leq e \end{cases} \quad (22)$$

For layer 2, the nodes act as a multiplier and are fixed. The outputs of this layer are represented by:

$$Q_{ij}^2 = w_{ij} = \mu_{Ai}(x)\mu_{Bi}(y), \quad i, j = 1, 2, 3, \dots, N \quad (23)$$

which represents the firing strength of each rule. The firing strength means the degree to which the antecedent part of the rule is satisfied.

For layer 3, the nodes are also fixed, indicating that they play a normalization role in the network. The outputs which are called normalized firing strengths from this layer can be represented as follows:

$$\bar{w_i} = \frac{w_i}{w_{11} + w_{12} + w_{21} + \dots + w_N} \quad (24)$$

For layer 4, the parameters in this layer are referred to as consequent parameters. Each node is an adaptive node, and its output is simply the product of the normalized firing strength and a first-order polynomial. The outputs of this layer are given by:

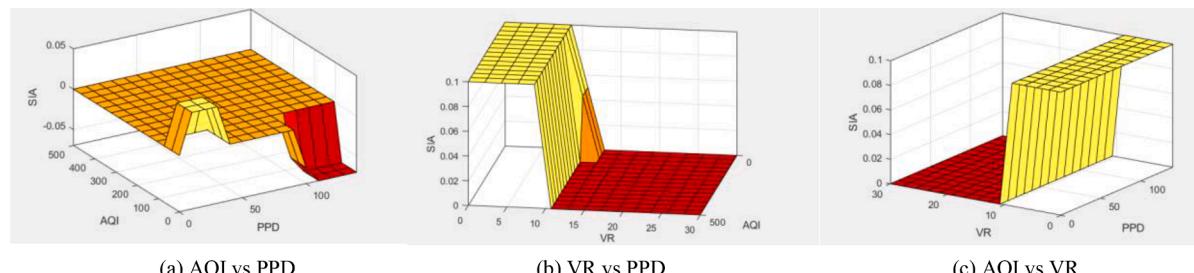


Fig. 20. Output logic of the ANFIS logic controller.

Table 5
Classification performance with 344 hourly test observations.

Indoor State	Hazardous	Very Poor	Poor	Moderate	Good
Hazardous	60 (93.75%)	2 03.13%	1 01.56%	1 01.56%	0 0%
Very Poor	1 02.13%	45 (95.74%)	1 02.13%	0 0%	0 0%
Poor	3 02.23%	2 01.53%	125 (93.284%)	3 02.23%	1 0.07%
Moderate	0 0%	0 0%	2 02.63%	73 (96.05%)	1 01.32%
Good	0 0%	0 0%	2 08.00%	1 04.34%	20 (86.95%)

Table 6
Classification accuracy corresponding to different ventilation states.

Ventilation State	Classification Accuracy (%)
Hazardous	93.75%
Very Poor	95.74%
Poor	93.284
Moderate	96.05%
Good	86.95%

$$0_1^4 = \overline{w_i} f_i = \overline{w_i}(px + qy +, \dots, +r) \quad (25)$$

where p_{ij} , q_{ij} and r_{ij} are consequent parameters of the first-order polynomial.

For layer 5, the single node which computes the summation of the entire incoming signal which is a fixed node labelled:

$$Z = \sum \overline{w_i} f_i \quad (26)$$

The VR, PPD and AQI values are used to measure the SIA index. The values of PPD from 0 to 100 are ideal for occupant comfort, though ASHRAE suggests a VR of more than or equal to 15 cfm for classrooms (Kumar Sai et al., 2019), and AQI values of 0 to 400 are taken into account. The ANFIS controller was chosen because of its versatility, low training data requirements, and ability to model complex non-linear functions with more variables to integrate.

Fig. 19 depicts the rules that dictate SIA based inputs on PPD, VR, and AQI. Based on the pollutants and their concentrations, the rule table is obtained and used in the controller design. The membership functions used for the inputs are also seen below using a rule-based design in the scenario below. It is observed from the output SIA values from the scope of the Simulink model. It shows the rule view of the neuro-fuzzy inference system for the calculation of SIA using PPD, VR and AQI as inputs. The first column on the left shows the PPD, the second the VR and the third the AQI values as they change with time. The red line passing through the column provides the mean value. The fourth column on the extreme right depicts the calculated SIA value using the neuro-fuzzy inference system.

The Simulink model is constructed using parameters VR, PPD, and AQI as input sources, multiplexers, output display the connectors. All these mentioned blocks are available in the Simulink modelling library. All the blocks must be pulled from the Simulink library into the model and must be built. Apart from these, the control system toolbox available in the Simulink library is being used, which uses the rules defined for the calculation of the SIA index value in the display output.

The 3-D representation of SIA based on ANFIS logic controller against VR, PPD, and AQI is shown in Fig. 20. These graphs use different colors to indicate the five different states of indoor air, good, moderate, poor, very poor, and hazardous. The yellow colour in the graph signifies state transition from good to moderate SIA, transaction from state poor to very poor state of the colour brown represents SIA, whereas the colour red represents SIA ranges from bad to dangerous.

An AQI v/s PPD plot is shown in Fig. 20(a). The state of indoor air improves with the decrease in the values of PPD and AQI, leading to moderate and good SIA. The rise in the graph represents a high value of AQI whereas a moderate value of thermal comfort is portrayed when the value is less. Fig. 20(b) shows a 3-D plot between VR and PPD, which demonstrates that a higher value of VR and thermal comfort is synonymous with a good state of indoor air. The plot indicates that the higher the value, the better the quality of indoor air corresponds to better quality. The 3D representation of AQI and VR in Fig. 20(c) shows that the rise in AQI makes the state of indoor air poor. For a good class category for SIA, a combination of good VR and good thermal comfort and AQI is needed with more than 15 cfm VR and AQI is less than 100 amounts. PPD values between 5 and 25 or VR between 13 and 15 cfm and AQI between 100 and 200 signify moderate SIA. Poor SIA is

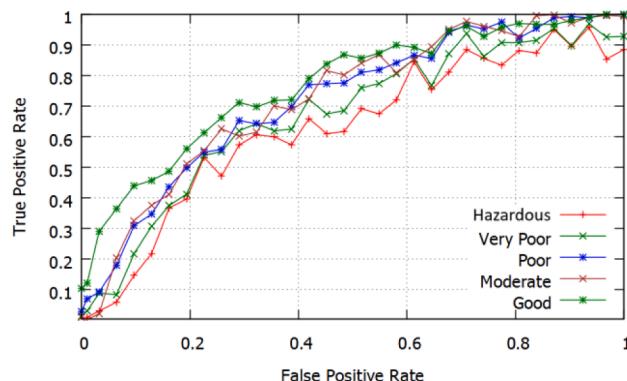


Fig. 21. ROC Curve.

Table 7
Precision table.

Category	Precision	Recall	F1 Score
Hazardous	0.93	0.93	0.93
Very Poor	0.95	0.91	0.93
Poor	0.93	0.95	0.94
Moderate	0.96	0.93	0.94
Good	0.86	0.90	0.88

indicated by VR less than 9 cfm, PPD values greater than 100, and AQI greater than 400.

5.2. Performance evaluation

The concentrations of indoor pollutants PM_{2.5}, PM₁₀, and CO are used to calculate the levels of ventilation of the indoor space. 1376 h of observation data out of a total 1720 h is used for training and the remaining 344 h of data is used for testing the accuracy of the ANFIS model. The most predominant ventilation state is poor, which occurs 39% of the time, or for 134 out of 344 test hours, followed by the moderate and hazardous states, which occur 22% and 18.6% of the time, respectively. The ANFIS output is in the form of a 5 × 5 confusion matrix (Table 5) which describes the accuracy of the algorithm in classifying the hazardous, very poor, poor, moderate, and good states, corresponding to the three attributes, AQI, VR and TC.

According to Table 5, 23 of the 344 observations refer to good SIA state, 76 to moderate state, 134 to a poor state, 47 to a very poor state and 64 are about hazardous states. With 73 of the 76 test outcomes correctly identified, the moderate class had the best identification performance. In the very poor class, the algorithm was able to accurately identify 45 of the totals of 47 test observations. The forecast algorithm correctly classified 125 of the 134 study observations as belonging to the poor state. The algorithm's accuracy for hazardous, very, poor, moderate, and good is 93.75 percent, 95.74 percent, 93.28 percent, 96.05 percent, and 86.95 percent, respectively. The classification accuracy on average is 93.62 percent.

5.3.1. To assess the classification performance of the proposed technique, precision, recall, F1 score, and receiver operating characteristic (ROC) curve is used

The diagnostic ability of the proposed classification technique is demonstrated with the help of the receiver operating characteristic (ROC) curve in Fig. 21. The ROC curve is a plot of the true positive rate

(sensitivity) against the false positive rate (1- specificity) with variation in the discrimination threshold. The area under the ROC curve (AUC) is a measure of how well a parameter can discriminate between the ventilation levels. AUC is graded on a scale from 0 to 1, 0.5 indicating no prejudice, 0.7 to 0.8 indicating suitable classification, 0.8 to 0.9 indicating excellent classification, and more than 0.9 indicating excellent classification.

The ROC curve in Fig. 21 indicates that the model's classification accuracy is extremely high. The proposed model has an average F1 score of 0.89 for three ventilation states, indicating that it is extremely precious.

- (i) Precision: It is the ratio of the number of true positive (TP) predictions to all positive predictions.

$$\text{precision} = \frac{\text{number of true positive samples}}{\text{number of positively predicted samples}} \quad (26)$$

- (ii) Recall: It is the ratio of the number of true positives to all the actual positive records.

$$\text{recall} = \frac{\text{number of true positive samples}}{\text{number of actual positive samples}} \quad (27)$$

- (iii) F1 score: It takes both precision and recall into consideration. It is the harmonic mean of precision and recall:

$$\text{F1 score} = 2 * \left(\frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \right) \quad (28)$$

As shown in the Table 7, the summarized performance model of the state. The value of precision, recall, and F1 score for the very poor and poor state are 0.93, 0.94, respectively. context-aware IoT system is used for modelling different situations and conditions of an enclosed environment. Factors that affect the indoor environment, such as IAQ, thermal comfort, and ventilation issues, are combined to develop a new index called the "state of indoor air" (SIA). The proposed system uses six indoor parameters to compute PPD, VR and AQI values. An adaptive neuro-fuzzy inference method (ANFIS) uses the PPD, VR, and AQI values to rate the current state of the indoor air as hazardous, very poor, poor, moderate, and satisfactory. Precision, recall, F1 ranking, and ROC were used to assess the proposed system's prediction accuracy. During the mild indoor state, accuracy, recall, and F1 score have values of 0.96, 0.93, and 0.94, respectively. For the state healthy, it's 0.86, 0.90, and 0.88, respectively. The model's F1 score for the weak state is 0.94, indicating that it works well in this state, followed by mild, very poor, and hazardous states. "The low ventilation condition" is more correctly defined as the most prevalent state in this proposed scheme, arising 39 percent of the time during the trial. The proposed model is highly predictive, with an average F1 score of 0.92 for all the five categories of indoor states.

6. Context prediction

Context prediction of future air quality levels is a crucial part of comprehensive monitoring and prediction for indoor air quality. The process of prediction will include past and present context concentration of pollutants values to forecast future values that will trigger different stages of air quality outcomes, thus enabling proactive state knowledge to end-users for taking healthy measures.

To model the different states of SIA, we employed Markov chains, constructed from ANFIS in this paper. The Markov chain is a probability model for describing the sequence of potential SIA levels, in which the

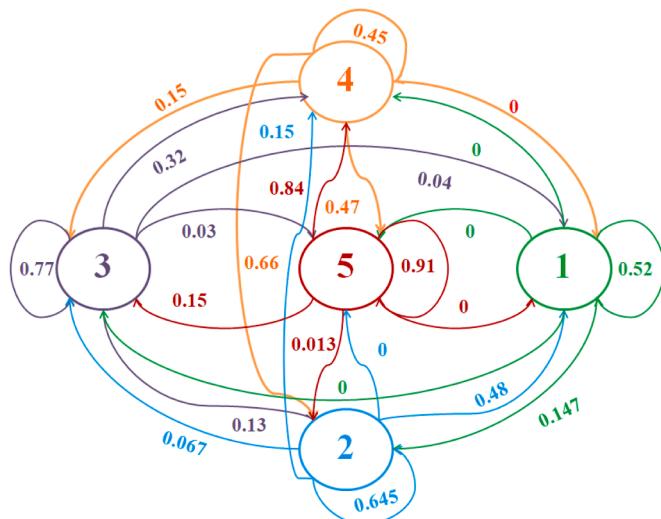


Fig. 22. The stochastic process moves from one SIA to other.

possibility of each SIA level occurrence is solely dependant on the condition of the previous SIA. As a result, it fulfills the Markovian property (memory-less-ness). It means that once you understand the present state of the stochastic mechanism, you do not need to know anything about prior IAQ states to provide the best possible forecast of the future state (Hoek & Elliott, 2012). Because the number of parameters required to assess a SIA process is considerably decreased as a result of this simplicity, the Markov chain method was utilized to forecast SIA states in this work.

6.1. Problem definition

Let the SIA be X_n at time t and X_{n+1} at time $t + 1$. As per the Markovian property, the future SIA (SIA at time $t + 1$) is exclusively reliant on the current SIA, i.e. X_n . Let $X = X_n$, where $n = 0, 1, \dots, n$ are random numbers with SIA state, $S = 1, 2, \dots, m$ in the Markov order of events. This process should meet the following property to be a discrete time Markov chain (Sericola et al., 2013):

$$P(X_{n+1}=j|X_n=i, X_{n-1}=i_{n-1}, \dots, X_0=i_0)=P(X_{n+1}=j|X_n=i) \quad (29)$$

The possibility of change between two states is not dependant on time in a discrete-time Markov chain, described with the help of the following equation:

$$P(X_{n+1}=j|X_n=i)=P(X_1=j|X_0=i)=p_{ij} \quad (30)$$

The one-step transition probability from state i to state j is given by p_{ij} . The following properties must be met by the one-step transfer probability p_{ij} :

$$\text{One : } p_{ij} \geq 0, \text{ and} \quad (31)$$

$$\text{Two : } \sum_{j=1}^m p_{ij} = 1 \quad (32)$$

The probability of transitioning from one SIA state to another is represented by each position in the transfer probability matrix.

States	Good	Moderate	Poor	V.Poor	Hazardous
Good	0.52	0.147	0.00	0.00	0.00
Moderate	0.48	0.645	0.067	0.15	0.00
Poor	0.04	0.13	0.77	0.32	0.03
V.Poor	0.00	0.66	0.15	0.45	0.47
Hazardous	0.00	0.013	0.15	0.84	0.91

In the transfer probability matrix, each position represents the probability of one state of transition to another. The transformation of

the stochastic SIA phase from one SIA state to another with marked probability is seen in Fig. 22 according to the probability matrix of trnsition. Five states, i.e. Good, Moderate, Poor, Very Poor and hazardous, form the state space of our Markov model. Each state represents a particular state of indoor air (SIA). State 1 denotes good SIA, State 2 stands for moderate SIA, State 3 represents poor SIA, State 4 is the very poor state and State 5 describes the hazardous state.

The steady SIA states have been calculated using the steady state and meand return duration in the following paragraph to calculate the percentage of time the DTMC process spends in each of the SIA states.

6.2. SIA state processes

π_i is the percentage of time for which the process stays in specific SIA state, and it is calculated by using the formulae below:

$$\pi_i = \sum_{k=1}^S \pi_k p_{ik}, \text{ and} \quad (33)$$

$$\sum_{i \in S} \pi_i = 1 \quad (34)$$

Where p_{ij} is the transfer probability from state i to state j . Higher value of π_i means that the occurrence probability of j^{th} state is high. Using the p_{ij} values from the transfer probability matrix, the steady state equations are derived:

$$\pi_1 = 0.52\pi_1 + 0.147\pi_2 + 0.0\pi_3 + 0.0\pi_4 + 0.0\pi_5 \quad (35)$$

$$\pi_2 = 0.48\pi_1 + 0.645\pi_2 + 0.067\pi_3 + 0.15\pi_4 + 0.0\pi_5 \quad (36)$$

$$\pi_3 = 0.04\pi_1 + 0.13\pi_2 + 0.77\pi_3 + 0.33\pi_4 + 0.03\pi_5 \quad (37)$$

$$\pi_4 = 0.0\pi_1 + 0.66\pi_2 + 0.15\pi_3 + 0.45\pi_4 + 0.47\pi_5 \quad (38)$$

$$\pi_5 = 0.02\pi_1 + 0.03\pi_2 + 0.15\pi_3 + 0.84\pi_4 + 0.91\pi_5 \quad (39)$$

Solving the above linear equations for π_1 , π_2 , and π_3 , we get the results as :

$$\pi = \begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \\ \pi_4 \\ \pi_5 \end{bmatrix} = \begin{bmatrix} 0.08 \\ 0.26 \\ 0.32 \\ 0.19 \\ 0.15 \end{bmatrix} \quad (40)$$

According to the forecasts, the most common indoor ventilation states are low and moderate (represented by 3 and 2). The IAQ stays in the good state only 0.08% of the time. The indoor ventilation is severe (poor, very poor, or hazardous) two-third of the time (66%) of the time, in very poor state 19%, and 15% time, there is a hazardous situation, indicating and alerting that there is a very high probability.

6.3. Mean return time

Mean return time (MRT) is the expected duration (in days) needed for coming back to the previous level i by the stochastic process.

Suppose in ' i ' the Markov chain starts. DTMC model then predict the mean return period for SIA level i is given as:

$$R_{ij} = \frac{1}{\pi_j} \quad (41)$$

The return period (days) to predict for each steady-state as:

$$R = \begin{bmatrix} 12.5 \\ 3.84 \\ 3.12 \\ 5.26 \\ 6.66 \end{bmatrix} \quad (42)$$

Table 7a
Actual and Model Predicted Return Periods (RP) Comparison.

State	Actual RP (days)	Predicted RP (days)	Error (days)	% Error
Good	12.42	12.5	-0.08	-0.65
Moderate	3.9	3.84	0.06	1.54
Poor	3.3	3.12	0.18	5.45
V. Poor	5.53	5.26	0.27	4.88
Hazardous	7	6.66	0.34	4.85

6.4. Results and discussion

This section describes the performance of the proposed index values from the classroom data during a lecture session to estimate the actual perceived comfort of the occupants. This paper explains the prediction of indoor air pollutants using the Markov chain forecasting model. Due to its memory less property and capabilities of random walk in transition matrix it generates results which are more reliable than any similar model. Earlier [Piccardi et al. \(2017\)](#), [Hazra et al. \(2017\)](#) and [Tserenjigmid et al. \(2019\)](#) studied and focussed in developing and solving the quality of air using simple prediction. They provide important information to improve air quality using those forecasting models. If the sampling rate is very high, the individual displacements are not random but rather deterministically coupled in time, and the DTMC model is incorrect. Hence, while using the DTMC model, care must be taken to space the readings sufficiently apart so that they do not affect each other and can be treated as random, else the model tends to give inaccurate results. Thus, context-based prediction is easy to understand and compute decision based on the data. Higher order Markov chains is the future scope of research to gain better insight in forecasting.

The accuracy of the system depends on the return period of SIA states and is estimated by comparing the recorded period of August 2018 to October 2019. For the recorded duration of August 2018 to October 2019, the DTMC model was used to forecast SIA levels using actual data. The average return duration is the time it takes for an event to happen again, such as a specified SIA value being exceeded. The actual return duration changes oppositely to the probability of exceedance as they are in inverse relationship of proportionality:

$$\text{Actual Return Time} = \frac{1}{P\{(SIA) > (SIA_{TH})\}} \quad (43)$$

where $P\{(SIA) > (SIA_{TH})\}$ is the probability of surpassing of the SIA. The probability of SIA surpassing is computed with the help of the relation:

$$P\{(SIA) > (SIA_{TH})\} = \frac{\text{days on which SIA exceeds threshold}}{\text{total no. of days}} \quad (44)$$

where SIA_{TH} = for an SIA state the threshold value.

The absolute forecast error is 3.47 percent on average. In [Table 7](#), the return period for the indoor SIA condition is best predicted. The error, percent error, and absolute percent error are calculated using the formulae below:

$$\text{Error} = \text{ActualRP} - \text{PredictedRP} \quad (45)$$

where, RP = return period

$$\% \text{ Error} = \left(\frac{\text{ActualRP} - \text{PredictedRP}}{\text{ActualRP}} \right) \times 100 \quad (46)$$

$$\text{Absolute \% Error} = \left(\frac{|(\text{ActualRP} - \text{PredictedRP})|}{\text{ActualRP}} \right) \times 100 \quad (47)$$

From [Table 7](#), the average absolute prediction error found to be 3.47% using [Eq. \(47\)](#). Amongst the 5 states, the best estimate of the return period for the good SIA state, amongst the model ([Table 7a](#)).

7. Conclusion

Here is an IoT framework that includes context-aware prediction for real-time monitoring of the indoor environment. The sensing system records, analyses, and processes real-time pollutant values present inside the indoor area, such as temperature, humidity, PM_{2.5}, PM₁₀, CO, and CO₂. It amalgamates the effects of indoor pollutant concentration, thermal comfort level, the state of ventilation and air quality index to represent the actual state of the indoor environment. The proposed index SIA is developed using ANFIS and can describe the overall condition of the environment inside the classroom. Thermal comfort is calculated with PPD using daily values of temperature and humidity, VR is calculated with the CO₂ concentration, using the daily average concentration of PM₁₀, PM_{2.5}, CO pollutants AQI is calculated. SIA in this system is classified into 5 states: Hazardous, Very Poor, Poor, Moderate, and Good. Using a yearly record of data SIA, the model is developed to forecast the return period for the defined SIA states using the DTMC model.

The most frequent interior ventilation states, according to the predictions, are poor and moderate (represented as state 3 and state 2). Only 0.08 percent of the time does the IAQ remain in good state. Two-thirds of the time (66%), the indoor ventilation is severe (poor, very poor, or hazardous), 19% of the time it is very bad, and 15% of the time it is hazardous, suggesting and warning that there is a very high probability of unhealthy AQI in educational institutions in the Delhi-NCR region. The performance is measured by the comparison between actual and forecasted return periods, and the forecast error for our system is low, with an absolute forecast error of 3.47% on an average.

Funding

No funding was received for this research.

Research involving human participants and/or animals

The research did not involve human or animal participants.

Informed consent

Not applicable (no human or animal participants were involved in the research).

CRediT authorship contribution statement

Krati Rastogi: Writing – original draft, Software, Validation, Visualization, Formal analysis. **Divya Lohani:** Conceptualization, Methodology, Investigation, Supervision.

Declaration of Competing Interest

The authors declare that they have no conflict of interest.

References

- ANSI/ASHRAE Standard 55-2010. (2010). *Thermal environmental conditions for human occupancy*. American society of heating, Refrigerating and Air-Conditioning Engineers, Inc [Online]. Available <http://arco-hvac.ir/wp-content/uploads/2015/11/ASHRAE-55-2010.pdf>.
- ANSI/ASHRAE Standard 62-2001. (2003). *Ventilation for acceptable indoor air quality*. American society of heating, Refrigerating and Air-Conditioning Engineers, Inc [Online]. Available https://www.ashrae.org/File%20Library/Technical%20Resources/Standards%20and%20Guidelines/Standards%20Addenda/62-2001/62-2001_Addendum-n.pdf.
- Becerra, J. A., Lizana, J., Gil, M., Barrios-Padura, A., Blondeau, P., & Chacartegui, R. (2020). Identification of potential indoor air pollutants in schools. *Journal of Cleaner Production*, 242. <https://doi.org/10.1016/j.jclepro.2019.118420>
- Cheng, Y., Niu, J., & Gao, N. (2012). Thermal comfort models: A review and numerical investigation. *Building and Environment*, 47, 13–22.
- Djamila, H. (2017). Indoor thermal comfort predictions: Selected issues and trends. *Renewable and Sustainable Energy Reviews*, 74, 569–580. ISSN 1364-0321.
- Dong, B., Prakash, V., Feng, F., & Neill, Z. O. (2019). A review of smart building sensing system for better indoor environment control. *Energy and Buildings*, 199, 29–46. <https://doi.org/10.1016/j.enbuild.2019.06.025>. ISSN 0378-7788.
- Dorizas, P. V., Assimakopoulos, M. N., Helmis, C., & Santamouris, M. (2015). An integrated evaluation study of the ventilation rate, the exposure and the indoor air quality in naturally ventilated classrooms in the Mediterranean region during spring. *Science of The Total Environment*, 502, 557–570. <https://doi.org/10.1016/j.scitotenv.2014.09.060>
- Environmental Protection Agency (EPA). (2016). *Environmental hazards in the home*. United States[Online]. Available <Https://www.epa.gov/report-environment/indoor-air-qualityw>.
- Fang, L., Clausen, G., & Fanger, P. O. (1998). Impact of temperature and humidity on the perception of indoor air quality. *Indoor air*, 8, 80–90. <https://doi.org/10.1111/j.1600-0668.1998.t01-2-00003.x>
- Feng, Y. H., Teng, T. H., & Tan, A. H. (2009). Modelling situation awareness for context-aware decision support. *Expert Systems with Applications*, 36(1), 455–463. <https://doi.org/10.1016/j.eswa.2007.09.061>. ISSN 0957-4174 <http://www.sciencedirect.com/science/article/pii/S0957417407004769>.
- Ha, Q. P., Metia, S., & Phung, M. D. (2020). Sensing data fusion for enhanced indoor air quality monitoring. *IEEE Sensors Journal*, 20(8), 4430–4441. <https://doi.org/10.1109/JSEN.2020.2964396>, 15 April15.
- Hazra, T., Nene, M. J., & Kumar, C. R. S. (2017). A strategic framework for searching mobile targets using mobile sensors. *Wireless Personal Communications*, 95, 1–16. <https://doi.org/10.1007/s11277-017-4113-7>
- Hoek, J., & Elliott, R. J. (2012). Asset pricing using finite state markov chain stochastic discount functions. *Stochastic Analysis and Applications*, 30, 5.
- Horr, Y. A., Arif, M., Katafygiotou, M., Mazroei, A., Kaushik, A., & Elsarrag, E. (2016). Impact of indoor environmental quality on occupant well-being and comfort: A review of the literature. *International Journal of Sustainable Built Environment*, 5(1), 1–11. Pages.ISSN 2212-6090.
- Jang, J. S. (1993). ANFIS adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man, and Cybernetics*, 23, 665–685. <https://doi.org/10.1109/21.256541>
- Kim, J. Y., Chu, C. H., & Shin, S. M. (2014). ISSAQ: An integrated sensing systems for real-time indoor air quality monitoring. *IEEE Sensors Journal*, 14(12), 4230–4244. <https://doi.org/10.1109/JSEN.2014.2359832>
- Kumar Sai, K. B., Mukherjee, S., & Sultana, H. P. (2019). Low cost iot based air quality monitoring setup using arduino and MQ series sensors with dataset analysis. *Procedia Computer Science*, 165, 322–327. <https://doi.org/10.1016/j.procs.2020.01.043>. ISSN 1877-0509.
- Li, L., Zhang, Y., Fung, J. C. H., Qu, H., & Lau, A. K. H. (2022). A coupled computational fluid dynamics and back-propagation neural network-based particle swarm optimizer algorithm for predicting and optimizing indoor air quality. *Building and Environment*, 207, Article 108533. <https://doi.org/10.1016/j.buildenv.2021.108533>. Part BISSN 0360-1323.
- Liu, B., Jin, Y., & Li, C. (2021). Analysis and prediction of air quality in Nanjing from autumn 2018 to summer 2019 using PCR-SVR-ARMA combined model. *Scientific Reports*, 11, 348. <https://doi.org/10.1038/s41598-020-79462-0>
- Liu, C., Park, E., & Jiang, F. (2020). Examining effects of context-awareness on ambient intelligence of logistics service quality: User awareness compatibility as a moderator. *Journal of Ambient Intelligence and Humanized Computing*, 11, 1413–1420. <https://doi.org/10.1007/s12652-018-1004-z>
- Ljung, L. (1979). Asymptotic behavior of the extended Kalman filter as a parameter estimator for linear systems. *IEEE Transactions on Automatic Control*, 24(1), 36–50. <https://doi.org/10.1109/TAC.1979.1101943>. February.
- Lohani, D., & Acharya, D. (2016). SmartVent: A context aware IoT system to measure indoor air quality and ventilation rate. In *Proceedings of the 17th IEEE international conference on mobile data management (MDM)* (pp. 64–69). <https://doi.org/10.1109/MDM.2016.91>
- Ma, N., Aviv, D., Guo, H., & Braham, W. W. (2021). Measuring the right factors: A review of variables and models for thermal comfort and indoor air quality. *Renewable and Sustainable Energy Reviews*, 135(2021), Article 110436. <https://doi.org/10.1016/j.rser.2020.110436>. ISSN 1364-0321.
- Massano, M., et al. (2020). An online grey-box model based on unscented kalman filter to predict temperature profiles in smart buildings. *Energies*, 13, 2097. <https://doi.org/10.3390/en13082097>
- Metia, S., Oduro, S. D., Ha, Q. P., Duc, H., & Azzi, M. (2013). Environmental time series analysis and estimation with extended kalman filtering. In *Proceedings of the 1st international conference on artificial intelligence, modelling and simulation* (pp. 235–240). <https://doi.org/10.1109/AIMS.2013.44>
- Norhidayah, A., Lee, C. K., Azhar, M. K., & Nurulwahida, S. (2013). Indoor air quality and sick building syndrome in three selected buildings". *Procedia Engineering*, 53, 93–98. <https://doi.org/10.1016/j.proeng.2013.02.014>, 2013.
- Nowak-Dzieszko, K., & Kisielewicz, T. (2020). Internal particulate matter pollution in educational building. *E3S Web of Conferences*, 172, 06008. http://doi.org/10.1051/e3sconf/202017206_08.
- Nwiabu, N., Allison, I., Holt, P., Lowit, P., & Oyeneyin, B. (2011). Situation awareness in context-aware case-based decision support. In *Proceedings of the IEEE conference on cognitive methods in situation awareness and decision support (CogSIMA 2011)* (pp. 9–16). Piscataway, New Jersey: IEEE, 22–24 February 2011.
- Piccardi, C., Riccaboni, M., Tajoli, L., & Zhu, Z. (2017). Random walks on the world input-output network. *Journal of Complex Networks*, 6, 187–205. <https://doi.org/10.1093/comnet/cnx036>
- Prasad, M.V., .Kumar, D.A., .Kishore, P.V., .Sastry, A.S., Harini, A., Raviteja, K. et al. (2016). Continuous sign language recognition from tracking and shape features using fsi and ann.
- Putra, J. C., Safrilah, S., & Ihsan, M. (2018). The prediction of indoor air quality in office room using artificial neural network. *AIP Conference Proceedings*, 1977, Article 020040. <https://doi.org/10.1063/1.5042896>
- Rastogi, K., Barthwal, A., & Lohani, D. (2019). AQCI: An IoT based air quality and thermal comfort model using fuzzy inference. In *Proceedings of the IEEE international conference on advanced networks and telecommunications systems (ANTS)* (pp. 1–6). <https://doi.org/10.1109/ANTS47819.2019.9118026>
- Rastogi, K., & Lohani, D. (2019a). IoT-based occupancy estimation models for indoor non-residential environments. In *Proceedings of the IEEE 16th India council international conference (INDICON)* (pp. 1–4). <https://doi.org/10.1109/INDICON47234.2019.9030326>
- Rastogi, K., & Lohani, D. (2020). An IoT-based Framework to forecast indoor air quality using ANFIS-DTMC model. *Source International Journal of Next-Generation Computing*, 11(1), p76–p97.
- Rastogi, K., & Lohani, D. (2020). An internet of things framework to forecast indoor air quality using machine learning. In: Thampi S., Trajkovic L., Li KC., Das S., Wozniak M., Berretti S. (eds). In *Machine learning and metaheuristics algorithms, and applications. Somma 2019. Communications in computer and information science*, 1203. Singapore: Springer. https://doi.org/10.1007/978-981-15-4301-2_8.
- Rastogi, K., Lohani, D., & Acharya, D. (2020a). An IoT-based discrete time Markov Chain Model for analysis and prediction of indoor air quality index. In *Proceedings of the IEEE sensors applications symposium (SAS)* (pp. 1–6). <https://doi.org/10.1109/SAS48726.2020.9220077>
- Rastogi, K., Lohani, D., & Acharya, D. (2020b). An IoT-based system to evaluate indoor air pollutants using grey relational analysis. In *Proceedings of the International Conference on COMmunication Systems & NETworks (COMSNETS)* (pp. 762–767). <https://doi.org/10.1109/COMSNETS48256.2020.9027308>
- Rastogi, K., Lohani, D., & Acharya, D. (2021). Context-aware monitoring and control of ventilation rate in indoor environments using internet of things. *IEEE Internet of Things Journal*, 8(11), 9257–9267. <https://doi.org/10.1109/JIOT.2021.3057919>
- Saad, S. M., Mohd Saad, A. R., Kamarudin, A. M. Y., Zakaria, A., & Shakaff, A. Y. M. (2013). Indoor air quality monitoring system using wireless sensor network (WSN) with web interface. In *Proceedings of the international conference on electrical, electronics and system engineering (ICEESE)* (pp. 60–64). <https://doi.org/10.1109/ICEESE.2013.6895043>
- Saini, J., Dutta, M., & Marques, G. (2022). ADFIST: Adaptive dynamic fuzzy inference system tree driven by optimized knowledge base for indoor air quality Assessment. *Sensors*, 22, 1008. <https://doi.org/10.3390/s22031008>
- Satish, U., Mendell, M. J., Shekhar, K., Hotchi, T., Sullivan, D., Streufert, S., et al. (2012). Is CO₂ an indoor pollutant? Direct effects of low-to-moderate CO₂ concentrations on human decision-making performance. *Environmental Health Perspectives*, 120(12), 1671–1677. <https://doi.org/10.1289/ehp.1104789>
- Sericola, B. (2013). *Markov chains: Theory, algorithms and applications*. London: ISTE Ltd and John Wiley and Sons Inc.
- Shih, H. C. (2014). A robust occupancy detection and tracking algorithm for the automatic monitoring and commissioning of a building. *Energy and Buildings*, 77, 270–280. <https://doi.org/10.1016/j.enbuild.2014.03.069>
- Sikora, A., Zielonka, A., & Woźniak, M. (2022). Minimization of energy losses in the BLDC motor for improved control and power supply of the system under static load. *Sensors*, 22(3), 1058. <https://doi.org/10.3390/s22031058> (Basel, Switzerland) PMID: 35161805; PMCID: PMC8839546.

- Spachos, P., & Hatzinakos, D. (2016). Real-time indoor carbon dioxide monitoring through cognitive wireless sensor networks. *IEEE Sensors Journal*, 16(2), 506–514. <https://doi.org/10.1109/JSEN.2015.2479647>. Jan.15.
- Stazi, F., Naspi, F., Ulpiani, G., & Perna, C. D. (2017). Indoor air quality and thermal comfort optimization in classrooms developing an automatic system for windows opening and closing. *Energy and Buildings*, 139, 732–746. ISSN 0378-7788.
- Sun, C., Huang, X., Zhang, J., Lu, R., Su, C., & Huang, C. (2021). The new model for evaluating indoor air quality based on childhood allergic and respiratory diseases in Shanghai. *Building and Environment*, 207. [\(https://www.sciencedirect.com/science/article/pii/S0360132321008076\)](https://doi.org/10.1016/j.buildenv.2021.108410). Part A, 108410, 03601323.
- Tian, X., Cheng, Y., & Lin, Z. (2021). Modelling indoor environment indicators using artificial neural network in the stratified environments. *Building and Environment*, 208, Article 108581. <https://doi.org/10.1016/j.buildenv.2021.108581>. ISSN 0360-1323.
- Tsang-Chu, Y., Lin, C. C., Chen, C. C., Lee, W. L., Lee, R. G., Tseng, C. H., et al. (2013). Wireless sensor networks for indoor air quality monitoring. *Medical Engineering & Physics*, 35(2), 231–235. <https://doi.org/10.1016/j.medengphy.2011.10.011>
- Tserenjigmid, G. (2019). On the characterization of linear habit formation. *Economic Theory*, 1–45. Available online <Https://doi.org/10.1007/s00199-019-01202-x>.
- Turanjanin, V., Vučićević, B., Jovanović, M., Mirkov, N., & Lazović, I. (2014). Indoor CO₂ measurements in Serbian schools and ventilation rate calculation. *Energy*, 77, 290–296. <https://doi.org/10.1016/j.energy.2014.10.028>
- Wong, L. T., Mui, K. W., & Tsang, T. W. (2022). Updating Indoor Air Quality (IAQ) assessment screening levels with machine learning models. *International Journal of Environmental Research and Public Health*, 19(9), 5724. <https://doi.org/10.3390/ijerph19095724>. PMID: 35565119; PMCID: PMC9104166.
- WSU, Energy Program, WSUEP13-005. (2013). *Measuring carbon dioxide inside buildings – why is it important?*. January.
- Wyon, D. P. (2004). The effects of indoor air quality on performance and productivity. *Indoor Air*, 14(7), 92–101. <https://doi.org/10.1111/j.1600-0668.2004.00278.x>. Supplement10.
- Zhao, H., & Wang, Z. (2011). Motion measurement using inertial sensors, ultrasonic sensors, and magnetometers with extended kalman filter for data fusion. *IEEE Sensors Journal*, 12(5), 943–953. <https://doi.org/10.1109/JSEN.2011.21166066>