Journal of Theoretical and Applied Information Technology
15th May 2026. Vol.104. No.9
© Little Lion Scientific

ISSN: **1992-8645** www.jatit.org E-ISSN: **1817-3195**

EDGE AI-DRIVEN AIR QUALITY FORECASTING USING KALMAN FILTERING: A LOW-LATENCY IOT FRAMEWORK FOR REAL-TIME ENVIRONMENTAL MONITORING

AYONIJA PATHRE [1], PARASA RAJYA LAKSHMI [2], KATAM NAGA LAKSHMAN [3], EERLA RAJESH [4], T.MURALIDHARA RAO [5], M. V. B. T.SANTHI[6], N.DHARANI KUMAR[7], Y.SUMANTH[8]

[1]Department of AIML , Jai Narain College of Technology, Bhopal. India
[2] Department of IT, PVP Siddhartha Institute of Technology, Vijayawada, Andhra Pradesh, India
[3]Department of CSE, Malla Reddy (MR) Deemed To Be University, Hyderabad, India
[4]Department of CSE, Madanapalle Institute of Technology & Science, Andhra Pradesh, India
[5]Department of Civil Engineering, CVR College of Engineering ,Hyderabad, India
[6]Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India
[7]Department of CSE(IOT), RVR&JC College of Engineering, Guntur, Andhra Pradesh, India
[8]Department of CSE(IOT), RVR &JC college of Engineering, Guntur, Andhra Pradesh ,India

Email:[1]Ayo.pathre@gmail.com,[2]parasa.raji@gmail.com,[3] katamnagalakshman@gmail.com
[4]erajesh557@gmail.com,[5]tmuralidhararao@cvr.ac.in,[6]santhi2100@gmail.com,[7]dharaninarne@gmail.com,[8] sumanth46@gmail.com

ABSTRACT

Air quality is a critical factor influencing public health and environmental sustainability. Machine learning (ML) has become a potent instrument in the field of air quality monitoring and prediction in recent years. A network of sensors is often used by machine learning-based air quality systems to gather data in real-time on a variety of air pollutants, including particulate matter, nitrogen dioxide, sulphur dioxide, ozone, and more. These sensors are strategically placed in urban areas to provide comprehensive coverage. Existing cloud-based air quality monitoring systems often suffer from latency, bandwidth dependency, and delayed decision-making, limiting their real-time effectiveness. The collected data, often characterized by its multidimensional nature, undergoes a process of feature extraction. ML algorithms are employed to train on historical datasets. Once trained, the ML models are capable of making real-time predictions of air quality based on current input data. These predictions can be continuously monitored and updated, providing valuable insights into the dynamic nature of air quality within a given area. This article presents a novel approach for air quality monitoring and forecast system using edge computing and IoT in environmental application, using the Raspberry Pi and the Kalman Filter technique. Using the machine learning method, this work improves real-time decision-making by minimizing data transmission lags brought on by network and bandwidth constraints. In contrast to standard cloud-based monitoring and forecasting of air quality systems. Air pollutants like SO2, PM10, PM2.5, and NO2 are immediately predicted using the RPi, as an edge device with significant computational capacity. Additionally, the KF algorithm boosts sensor accuracy by 30% in compared with to sensor observation data and improved projected value accuracy. The proposed framework demonstrates the feasibility of low-latency edge-based air quality forecasting with improved prediction accuracy and reduced communication overhead. The results indicate its practical applicability for scalable smart city environmental monitoring systems.

Keywords: *AI, Air Quality, Forecasting, Framework, Kalman Filtering*

1. INTRODUCTION

For the management of environmental health, air quality forecasting and monitoring are becoming more and more crucial. In this field, machine learning (ML) has become a potent instrument that can forecast trends in air quality and analyze large, intricate datasets. The performance of various ML algorithms in air quality monitoring and prediction systems [3]. Sulphur dioxide (SO2), Nitrogen dioxide (NO2), Particulate matter (PM), and ozone (O3) are among the air quality metrics that have been predicted using a variety of machine learning (ML) techniques, such as neural networks, support vector machines, decision trees and regression. The features of the dataset, the complexity of the model, and the particular needs of the application are some of the variables that affect how well these techniques perform.

Because it is easy to apply and comprehend, linear regression is a popular technique. It is straightforward and easy to understand. Its effectiveness, however, might be constrained when handling the non-linear correlations included in data on air quality. In order to deal with non-linearity, decision trees and ensemble techniques—such as Random Forests—build intricate decision-making structures from incoming feature data. These techniques may be prone to over-fitting, yet they are capable of capturing complex patterns.

This unique study aims to integrate Edge Computing and Internet of Things (IoT) to advance systems for monitoring and prediction of air quality. The processing, storing, and communicating of the real-time data contain several issues for traditional air quality monitoring systems[2]. By dispersing processing power closer to the data source and decentralising it, the integration of Edge Computing with IoT reduces latency and boosts efficiency in order to address these issues. IoT sensors are carefully placed around city of Visakhapatnam in this study to gather data on the current state of the air, including the quantities of pollutants like nitrogen dioxide, ozone, and particle matter. These sensors constitute a decentralised network that can process data locally by interacting with nearby Edge Computing devices.

Because Edge Computing does real-time data analytics on the gathered information, it is necessary to the system. This instant processing capability makes it possible to identify abnormalities in the quality of the air quickly and respond in a timely manner to possible pollution occurrences. Furthermore, Edge Computing optimises bandwidth utilisation and lowers latency by minimising the need to send massive amounts of raw data to centralised servers. The Edge Computing framework's incorporation of machine learning methods improves the predictive power of the system. These algorithms forecast future trends in air quality by examining past data patterns, weather, and other pertinent variables. Proactive decision-making for public health alerts, pollution control measures, and urban planning plans is made easier by this predictive modeling [13], [19], [36].

The next big thing in digital technology is the Internet of Things (IoT). It is a massive network that links a great number of information-detecting devices to the Internet. It explains situations where items or objects are more connected to the Internet than people.. Like people or cars, things are intangible objects that possess sensors and IP addresses that enable them to sense, collect, and transfer data. It is implemented in businesses from a variety of sectors. They can operate more effectively, offer superior customer service, make more informed decisions, and raise the value of the business thanks to it. Micro services, wireless, and MEMS technology played major roles in its emergence. It originated from machine-to- machine communication, when devices connect to each other's networks automatically and without the need for human interaction. It facilitates data collecting, management, and cloud connectivity [6],[11],[8].

Because more IoT terminal equipment is being used, more terminal data and connections are available, necessitating an effective IoT construction to enable speedy processing of data and analysis. Concurrently, the IoT is growing and proliferating across other domains, such as intelligent home, intelligent transportation, and intelligent agriculture. For a wide range of particular application situations, such as automated driving, road condition data collection, security, and others, the network needs to further minimise data transmission delays. In this case, data was recorded and sent over the network. The primary method for doing this is via using sensors. The temperature, motion, moisture, air quality, andlight may all be measured by the sensor. These kinds of sensors are grouped together to collect or transmit data. According to its description, edge computing breaks down the barrierto IoT development and is considered a key component of IoT. Combining network processing, storage, and core competencies at the edge of the net can create edge computing.

Numerous services, including application intelligence, security, privacy, and optimization as well as the essential requirements for the industry's digitalization are provided by this open platform. Similar to human nerve endings, edge computing features might have the ability to independently process normal impetuses and transmit the results back to the cloud[15],[16],[18].

As more people drive, so does the amount of pollutants released into the air. The six most common air pollutants, according to the report, are lead, ozone, PM, CO, SO_2, and NO_2. Since these pollutants are categorised as critical ones, an analysis of them is necessary to determine the level of air safety. Moreover, the CO, CO_2, and SO_2 released by moving vehicles contribute to air pollution. Thus, developing real-time monitoring systems and air quality forecasts is the most basic approach. The air quality prognosis is anticipated through the analysis of the monitored data. A forecast is somewhat impacted by the precision of the monitoring data. The Environment Agency set specific AQ criteria at the same time. But if the prediction is wrong, awful decisions will be taken[21].

There are already a number of methods available to gather monitoring data more precisely, including the use of high precision sensors, which are usually more expensive. There is a trade-off between cost and precision because a system as a whole is composed of multiple sets of sensors. Additionally, sensed data must be processed and analysed at the cloud computing platform's layer by a standard system for Air Quality monitoring using IoT. The incapacity to ensure both the input of data and the exploration of results affects how quickly decisions are made. Real-time processing on the edge alleviates a few of the bottlenecks associated with traditional cloud computing, particularly for environmental applications.

The goal of this project is to integrate edge computing with Internet of Things applications to improve real-time application performance at a reasonable cost. For instance, a Kalman filter algorithm implemented into forecasting and air quality monitoring systems may greatly increase overall accuracy at a minimal hardware cost. The setup continuously detects the concentrations of six air contaminants by placing inexpensive sensors in the monitoring area. Using 100 iterations, the KF technique creates a dynamic prediction model based on the attributes of air contaminants, with a brief lifespan. It can now estimate pollutant concentrations in real-time as a result, and accuracy may actually be improved at the algorithmic edge.

Therefore, an industrial PC is selected to act as both the carrier for edge-computing and the principal node of the set-up of sensors. A sensor mounted on an industrial PC collects monitoring data, runs the Kalman filter algorithm to generate precise forecasts, and then uploads the results to the cloud. The goal of this research is to create a real-time air monitoring system model that ensures accurate readings with minimal delay by leveraging edge computing and IoT architecture[33],[37].Unlike previous studies focusing primarily on prediction accuracy, this work emphasizes low-latency computation, lightweight processing, and edge deployment feasibility.

2. RELATED WORKS

The need for this study arises from the limitations of conventional centralized cloud-based frameworks, which are unsuitable for latency-sensitive environmental applications. Edge-based processing provides an efficient alternative by enabling faster computation, reduced communication overhead, and immediate response capabilities. As a result of reduced sensor and processing costs, now more and more "things" can be connected to the web, and computers are being shifted to the network's edge [14],[1].Edge computing is a crucial development towards making Internet of Things (IoT) systems more efficient and adaptive. There are various applications for edge computing, particularly in situations where cloud computing proves to be unproductive. Here are some advantages of edge-computing compared to cloud computing [34]. More effective and real-time data processing for big data support across all cloud situations. 2. Latency is decreased by processing data locally as opposed to at a distant cloud data centre. 3. Compared to data centres and cloud networks, local device data management solutions are less expensive. 4. When latency is decreased, applications are executed more efficiently and more quickly. 5. Different devices function without it, reducingthe possibility of failure [38].

IoT applications for edge-computing are diverse and include intelligent agriculture and autopilot. Businesses choose to provide IoT-enabled solutions using real-time edge computing [7],[41],[10].Furthermore, an embedded fog-enabled environmental monitoring system was shown by the authors. Intel andAVOB collaborated to produce an edge-aided remote management and observing for intelligent energy control using IoT. Build an IoT architecture and a platform for IoT servicesfor linked cars by utilising fog computing. [14],[4].The rise of intelligent industries necessitates high cloud computing and ubiquitous edge computing. Regardless of time lags, security concerns, or how well IoT apps work, edge-computing is the fundamental to the Internet of Things' allure.[5],[20],[28],[32].

Unlike conventional cloud-based and ML-driven air quality forecasting systems, the proposed framework performs lightweight computation at the edge using Kalman Filtering, reducing latency and computational complexity while maintaining prediction accuracy.

3. AIR QUALITY MONITORING SYSTEM

Diverse writers put forth various fixes for the AQ monitoring system. By combining wireless and sensor technologies, IoT establishes the groundwork for numerous AQ monitoring systems [40]. The perception layer detects and collects information on different air qualities. It is composed of multiple sensor devices. Before taking appropriate action based on the current AQ conditions, the network layer sends and evaluates data from the perception layer. Cloud computing, wireless networks, and other technologies make up its three components. The application layer gives the user access to pertinent information. To track the environment, it created and put into place a modular IoT- based infrastructure. It functions by having the client adopt the HTTP protocol so that the administration layer can receive the data and the interface can reply to client queries. As a result, the 4-tier architecture is in use.A method for monitoring pollution in major cities which uses wireless sensor network. An ATmega-32 microcontroller transfers the data from the analogue to the digital converter and then to the server. Bluetooth is used to create a communication connection between the server and the gas distribution system[29].

Internet of Things technologies to develop a greenhouse air monitoring and management system [35]. It was previously used to construct a low-cost Raspberry Pi air quality monitoring system. The information that the Raspberry Pi gathers and sends to the gateway layer is limited to that which has been identified as a node in the sensor network. Data travelling from the ESP 8266 which is a wireless device to the cloud- layer is analysed and anticipated before to being transferred via the gateway layer. This layer evaluates the collected information and replies to different client inquiries[9]; [12], [22],, [26], [27].

In a similar manner, using Raspberry Pi, web server is connected the sensor network (Jadhav et al., 2016). Because of its limited pricing and special card-like functionality, Raspberry Pi is used in the above described concepts. Even though earlier research has recommended the need for a sophisticated and comprehensive AQ monitoring plan, the majority of China's agricultural regions are situated in isolated areas with difficult environmental circumstances [21].Unfortunately, the cloud computing system's network layer is unable to provide the low latency needed for data processing and analysis. Consequently, when paired with SMA, the prediction outputs of the EWMA model are smoother and more accurate. Autoregressive Integrated Moving Average model abbreviated as ARIMA is a widely used model[42], [24], [36] [42].

Using the data difference method, the model will remove non-static components in the original order to improve prediction outcomes. The conventional statistically-based prediction model is constrained by the attribute expression and is not capable of handling intricate prediction problems like nonlinear processes, although performing better in terms of interpretation and processing cost. A lot of research has chosen to anticipate AQ utilising multi-mode fusion or ML techniques since ML has become so popular. Two distinct decision-support models are created using the J48 decision tree approach ,study in order to predict the absorption of PM2.5 in the two surrounding regions[30]. The gadget predicts certain areas and offers details on early warning systems for areas that are highly contaminated. An ANN-based multi-input, multi-output AQ forecasting model is offered by us. Based on the association between the most recent and prior 24-hour pollutant concentrations, an artificial neural network 24-hour prediction network is constructed[39].

Although existing studies have achieved promising results, most rely on centralized cloud processing or computationally intensive machine learning models, limiting their suitability for real-time edge deployment.

4. MATERIALS AND METHODS

Hardware:

As a result of the Internet of Things' advancements and the integration of wireless and tiny sensor technologies, many AQ monitoring systems are now constructed using Internet of Things architecture. The perception layer recognises and gathers data on various air quality parameters using a range of sensor devices. The network layer sends information from the perception layer and assesses it before drawing the proper conclusions based on the AQ conditions at the time. It is made up of wireless technologies, a network management system, and a cloud computing platform. Lastly, the application layer gives the user the necessary data.

This node's inexpensive temperature and humidity sensor measures the relative humidity and temperature in Celsius of the surrounding air. We employ a DHT-based sensor for this purpose, which offers the advantages listed below. The DHT Humidity & Temperature Sensor furthermore generates a calibrated digital signal. It provides exceptional long-term stability and dependability by utilising state-of-the-art technology for monitoring temperature and humidity as well as a novel digital signal collection method. This sensor also offers outstanding quality, fast response times, immunity to interference, and affordability. It consists of a potent 8-bit microprocessor that interfaces with resistive-style humidity measurement and NTC temperature monitoring components. Suburban areas are expanding as a result of urbanisation, which calls for an extension of the existing fossil fuel-based transportation infrastructure. As more people drive, so do the associated pollutant emissions. In addition to PM (particulate matter), O2 (ground-level ozone), CO (carbon monoxide), SO2 (sulphur oxides), and the six most common air contaminants, lead is one of the pollutants, according to the study.

Criteria ones is the classification given to these contaminants. To determine how safe the air is to breathe, they must be examined. Moreover, nitrogen oxides, carbon monoxide, and carbon dioxide produced by vehicle emissions all contribute to air pollution. AQI, a MQ- based gas sensor, is used to collect SO2 and NO2. Gas sensors, also called gas detectors, are electrical devices that identify and differentiate between different gases. They are used in houses to find gas leaks and in factories and manufacturing facilities to detect smoke and carbon monoxide. They are also widely used to measure the concentrations of explosive or dangerous gases. Gas sensors differ widely in terms of size, range, and detecting power. Usually employed as a component of more extensive embedded systems, such as security systems, are often linked to an audio alert or interface. Compared to many other sensors, gas sensors need to be calibrated more frequently since they are in constant contact with air and other gases.

In the sensor part, we are also employing a third type of sensor to retrieve the PM10 value. We are using a PM210-based sensor designed to detect dust particles. The dust smoke particle sensor has a diagonal orientation and is composed of a phototransistor and an infrared emitting diode (IRED). It can identify light reflected from particles in the air. Dust robots, air purifiers, fire alarms, dust removal equipment, and other related applications are among the many uses for this module. Particles such as dust mites, smoke spores, and others can be located using industrial equipment. In our system, we broadcast data to the cloud using an ESP-based Wi-Fi microcontroller. To do this, an ESP8266 board was employed. The ESP8266 Node MCU has 30 pins total, with 17 of those being GPIO pins. General-purpose input and output is abbreviated as GPIO. There are nine digital pins with the designations D0-D8, one analogue pin with the identifier A0, and a ten-bit ADC. The D0 pin can only read and write data, which is a limited capacity. The ESP8266 chip is activated when the pin of EN is pushed to level of peak. Consequently, when the chip is pulled down, it uses the least amount of power. The CP2102 USB to TTL converter board has a 2.4 GHz antenna that allows for long-distance network connectivity. The ESP-12E module, which includes an ESP8266 chip and a Tensilica Xtensa® 32-bit LX106 RISC CPU with programmable clock rate of 80 to 160 MHz and support for RTOS, is included with the development board. Hardware layers arrangement is shown in Fig.1.

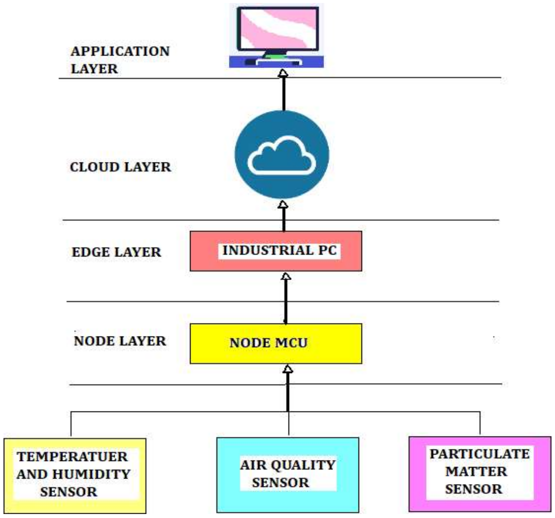

Figure.1. Hardware Layers

MQTT (Message Queuing Telemetry Transport)

A data transaction node is established using the most dependable MQTT service. A publish/subscribe messaging system with low bandwidth, Machine-to-Machine (M2M) telemetry is the purpose of MQTT. Another name for MQ Telemetry Transport is MQTT. MQTT is gradually gaining traction as an IoT installation protocol. The node operatio is shown in Fig.2.

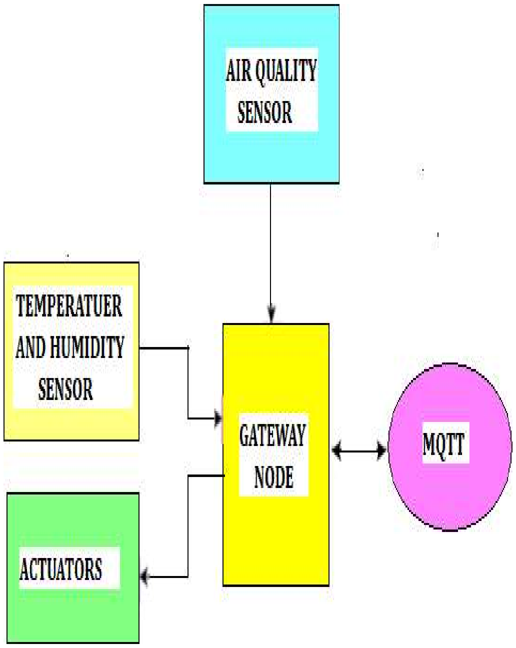

Figure.2 Node Operation

In addition, the node has an actuation section that can be utilised to perform different node-side predefined actions. Cutting-Edge Computing Traditional Internet of Things (IoT) design has been altered by the on-going decline in sensor and computing prices, which permits "things" to be connected to the Internet and processing to occur at the network's edge. Because edge computing is an essential development for extremely efficient and scalable sensor technology, it will be used in all aspects of life, particularly when cloud computing proves to be unusable. The advantages of edge computing over cloud computing are as follows. 1. An emphasis on rapid data analysis 2. Cloud apps that support big data 3. Real- time data processing and analysis 4. Lowest possible delay 5. Low price. 6. Less expensive alternatives to data centre and cloud networks for data management.

KALMAN Filter Algorithm

Two benefits of the model for conventional statistical prediction methods are its low processing cost and good interpretability. It bases its prediction theory on historical data that has been linearly fitted. As such, the expected conclusion can be predicted more accurately

when the changing trend is less outstanding. It loses relevance, though, if the concentrations of various air pollutants do not continue to climb gradually. When traffic is at its busiest, such as in the morning and evening, levels of pollutants like PM2.5, PM10, and SO2 can be swiftly absorbed and result in vehicle exhaust emissions. As a result, models like SMA, EWMA, and ARIMA are unable to correctly predict these shift points, despite the fact that the data have a higher predictability. (directly corresponding to measures).

The computing and storage capacity of the IoT device limits the prediction model's applicability to the AQ prediction application case. During training, a large number of Machine Learning (ML) models exhibit high temporal complexity. To improve forecast accuracy, they need to retain a large amount of past data (storage issue). For instance, the prediction accuracy of decision tree model can be increased by increasing the tree's depth andthe number of training samples; however, doing so raises the time complexity O (N x M x D) (in which N stands for the number of training samples). The ANN faces a comparable issue. Because of this, the ML model's forecasting accuracy has improved, but it still doesn't meet the needs of the lightweight model for edge computing node. This study uses the Kalman Filter (KF) method to develop an AQ prediction model. The KF (recursive filter model) is a reliable autoregressive filter model. Instead than having to store a lot of historical data, it simply needs to save data for a single system state. As a result, it requires very minimal memory.

Using the measured data, the predicted results are adjusted to more accurately match the actual findings. Because of its blazingly quick processing speed, the KF is ideal for handling problems in the present and works well with edge computing. The primary goal of the KF is to describe a dynamic system using a set of state-space equations and, utilising the best estimate, predict the state of the system at the next instant n (prediction) at time$\hat{y}$ n|n−1. Itis known as the prior estimate.

The concurrently expected states of the systems are represented by the observed values, zk and KF. The predicted value zk corrects the anticipated value y<n|n−1, as both zk and $\hat{y}$ n|n−1

depart from the actually exact state of system because of the observation error r. The system's state at time k is then best estimated, or forecasted, at that point. Compared to the conventional method of time prediction, the KF algorithm is distinct. First, it is not necessary for the error term to fit into a normal distribution. Secondly, it might make use of a range of erroneous observations (such missing times). The KF algorithm is unique when compared to the traditional time prediction method. First, it is not necessary for the error termto fit into a normal distribution. Second, it may estimate the system state (measurement error)using a variety of imprecise observations or noise. Third, by taking into account the combination of observations based on series of time data at various periods, the KF forecasts the unknown components that may affect the system, in contrast to models that use single observations. The KF forecast will be more precise as a result.

The Proposed System Architecture

More cellular bandwidth and affordable, mass-produced mobile sensing devices have made it possible to use air pollution sensors and acquire real-time data on air quality. Because of this, a lot of companies have created portable pollution measuring devices lately. These gadgets are small enough to be carried by pedestrians and are capable of detecting all the poisons from vehicle exhaust. Nevertheless, the real-time sensing capability on mobile platforms has not been assessed for any of these commercially accessible devices. On the other hand, the current AQM and prediction systems depend on precise chips in order to produce trustworthy data. However, because they are more expensive, greater precision sensors cannot be widely used in an IoT system with numerous sensors. The Proposed System Architecture is shown in the Fig.3.

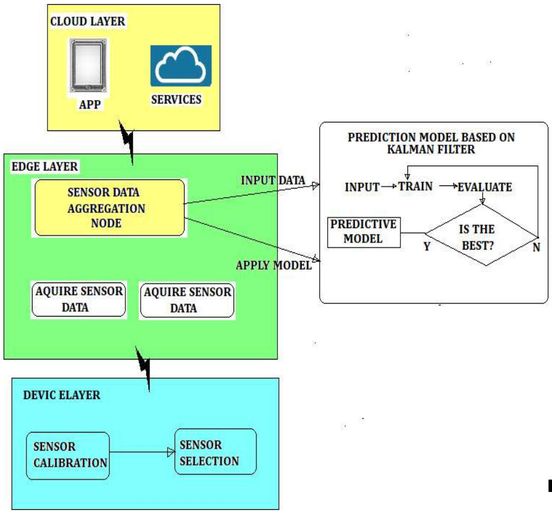

Figure.3. Proposed System Architecture

This research endeavour employs the suggested algorithm of an edge device to forecast the various concentration values of air pollutants based on real-time features of the Kalman Filter technique. By integrating the erroneous data, the KF can enhance sensor accuracy in the algorithm despite noise in the model processing and low-cost sensor measurement mistakes. Furthermore, the data of the previous instant can be used to infer the expected value of the subsequent instant, providing the structure with predictableness for various pollutants and enhancing the efficacy of decision-making.

Implementation of Kalman Filter Algorithm

An example of an autoregressive filtering paradigm is the Kalman Filter algorithm. Then, by combining the measurement of the system's present state with the most precise approximation of its prior state, the best approximation of the current status of the system moment can be obtained. The KF's initial state is represented by the first two variables:

As an approximate estimate of the system's condition at that moment k, y n|n.

The covariance matrix En|n, which measures the estimation's accuracy, represents the current state estimation error, k.

Through the estimation of the process condition at a given moment and the receipt of (noisy) observations, the KF gets input. The resulting KF equations are separated into time and measurement updating equations. By adding a fresh measurement to the ‘a priori’ estimate, the qualitative and quantitative assessment equations improve the ‘a posteriori’ estimation in response to the feedback. Measurement update equations are referred to as corrector equations, while time updates equations are also known as predictor equations. Figure 4 shows how similar the final estimate method is to a predictor-corrector methodologyfor solving numerical issues.

Figure. 4. Implementation of Kalman filter Algorithm

PREDICTION

The time update or projection is the first step in the suggested procedure. The following equations, which generate the projected error and the covariance matrix of the system's state at the most recent instant, are used to obtain the previous evaluation and the covariance matrix of an earlier estimate error in the current period:

$$\hat{y}n|n-1=Fn\hat{y}n-1|n-1+CnSn$$

$$En|n-1=FnEn-1|n-1F^{T}n+Zn \tag{1}$$

where Cn is the input matrix(control), Zn is the noise matrix(covariance) for the process, and Fn represents the transition matrix of various states. Correcting in tandem with sensor observation $\hat{y}$ n|n−1 is crucial, as the 'a-priori' approximation $\hat{y}$ n|n−1 is not finest at this moment, n.

Correction

The following three values mark the start of the algorithm's corrective stage (Measurementupdate).

$$\emptyset n=InEn|n-1Ik^{T}+On$$

$$Kn=En|n-1\,In^{T}+\emptyset^{-1}n$$

$$\hat{R}k=Rk-En\,\hat{y}n|n-1 \tag{2}$$

Through the use of the observation transfer matrix, Fn, in the preceding equation, the actual space of the dynamic system, $y\,|n-1$, is mapped into the observation space. The variance between the observation and the actual value derived from the a priori estimation is represented by the covariance matrices for the observed noise and margin, which are On and, $F\mathbf{k}$ respectively, as the observation at time k occurred is represented by the Kalman gain, which is $E\,n|n-1\,\hat{R}k$. The three results of the computation above are used to update the filter variables and $\hat{y}$ $n|n-1$ in order to obtain the best estimate $En|n-1$ and the covariance matrix $E\,n|n-1$ of the time-based state of system estimation error k.

$$\hat{y}k|k=\hat{y}k|k-1+Kn\hat{R}n$$

$$En|n=(1-KnEn)\,En|n-1 \tag{3}$$

RESULTS

Analysis of Accuracy Improvement

One way that the Kalman Filter technique varies from more time series forecasting models is its ability to incorporate statistical data, like process noise and observation error, efficiently while making predictions to change the model is a priori estimation. As a result, in order to improve prediction accuracy from a foretelling viewpoint, the KF employs measured data produced by the sensor network. By eliminating measurement data inaccuracies at the algorithm level, the KF may increase sensor accuracy side-by-side even with cheap and imperfect monitor sensors.

The study's goal is to show that the KF method may be used to successfully raise the sensor's data collection accuracy. Lastly, a comparison analysis is suggested as a way to calculate the error size. Additional research compares the data from the official monitoring station with the estimated KF value and the actual reading from the sensor. Mean Absolute Error(MAE), Mean Square Error(MSE), and Root Mean Square Error(RMSE) are the three most important error indicators.

Analysis of Predictability:

The KF technique does not explain how it improves performance prediction, even if it can increase sensor accuracy. Most machine learning (ML)-based prediction models are not appropriate for use in Internet of Things (IoT) scenarios because of the limited storage capacity and processing power of computing nodes (ML-based prediction procedures require high computing components like CPUs and a lot of processing, and ML-based prediction models need more training data to produce accurate predictions). In order to compare the existing algorithm with other statistical algorithms-Simple Moving Average (SMA), Exponentially Weighted Moving Average (EWMA), and ARIMA-this work designs an experiment (Autoregressive Integrated Moving Average).

Prediction of the Kalman Filter

Establish the KF algorithm's hyper parameters first. The hourly observing data of the air pollutant absorption at Quantancis Pvt Ltd. (from 0:00 to 23:00) are the monitoring data inthe corrected state used by the KF during the actual prediction process. Every forecast follows the process outlined below: The KF approach is updated iteratively using the sample findings, and the progressive convergence results serve as the expected values for each contaminant absorption, during this span of time. During the first ten minutes of each forecastphase, the system's sensors collect 100 samples of the concentration of each air pollutant.

SMA Prediction

By averaging future time, one can forecast SMA by computing the average mean value of historical data values and using that value for the upcoming period. One way to write it is:

$$Pt = (m\ t{-}\mathbf{1} + m\ t{-}\mathbf{2} + m\ t{-}\mathbf{3} + \cdots + m\ t{-}n)\ /\ n$$

In the equation above, Pt represents the foretold value for a specified period, and n being the total number of calculations used to determine the average periods, which is normally between 4 and 200.The true value of the preceding time period is $m\ t{-}1$, the real values of the first two periods are $m\ t{-}2$, the first three intervals are $mt{-}3$, and the initial n intervals are $m\ t{-}n$.

EWMA Prediction

An approach to sequence processing called EWMA improves SMA. Time series data are given a non-uniform weighting using this method, which makes it possible to use a large amount of data while assigning a higher weight to recent data. Weights are determined using the exponential function, as the name implies. It is said as follows:

$$Et = \lambda\gamma t + \mathbf{1} - \lambda\ E\ t{-}\mathbf{1}\ \ (t = \mathbf{1,2, \dots .}\ n)$$

where,Et is the predict assessment for the subsequent period Where t is the no of Calculating average periods, which is commonly within range 4 and 200, and real first period intervals, $(1-\lambda)$, the first 2-intervals, the first 3-intervals, and the first n- intervals , respectively, are $E\ t{-}1$ $(t = 1{,}2, \dots .\ n)$.

ARIMA Prediction

ARIMA uses past data to predict future patterns, which transforms non-stationary time series into stationary ones. It can be used to forecast future events. An expansion of the ARMA (p, q) model is the ARIMA model, often known as ARIMA (p, d, q). The sequence and degree of variance between the models at which the moving-average model's order for the time series becomes stable are indicated by the non-negative numbers p, d, and q in the formula ARIMA (p, d, q). ARIMA (p, d, q) using lag operator L as follows:

$$(\mathbf{1} - \textstyle\sum^{p} i{=}\mathbf{1}\quad \emptyset i\ \ )(\mathbf{1} - P)\ {}^{d}Y t = (\mathbf{1} - \textstyle\sum^{q} i{=}\mathbf{1}\emptyset\ \ )\ E\ t \tag{4}$$

Because before creating an ARIMA forecast, each model difference's parameters must be specified. The following is the specific training protocol: 1.It is necessary to determine the value of D (minimum difference order), which is used to transform the original data into a stationary sequence; 2. To determine p and q, find the coefficients of the sequence s's partial autocorrelation and autocorrelation. 3. After the ARIMA model is finished, assess each difference's precision. The following table shows the error indicators for the four algorithms (KF, ARIMA, EWMA, SMA) for the four pollutants (SO2, NO2, PM10, and PM2.5). The bar charts in Figures 5-8 below comparison of the error parameters for the various algorithms.

Table.1 Performance characteristics of different algorithms

(5)

POLLUTANT	ALGORITHM	MSE	RMSE	MAE
SO2	KF	0.0833	0.2887	0.2291
	ARIMA	0.4381	0.661	0.6482
	EWMA	0.4201	0.6482	0.4695
	SMA	1.2254	1.1072	0.5977
NO2	KF	2.0522	1.4325	1.1996
	ARIMA	10.6013	3.255	2.4727
	EWMA	12.8008	3.5777	2.9708
	SMA	19.7064	2.9708	3.576
PM10	KF	2.661	3.086	4.0672
	ARIMA	4.2966	5.1446	5.3236
	EWMA	36.6993	6.058	2.5465
	SMA	111.3151	5.5465	3.369
PM2.5	KF	1.711	2.843	2.982
	ARIMA	2.2673	3.5176	3.2831
	EWMA	24.3083	4.9302	2.0414
	SMA	73.336	5.0414	3.5652

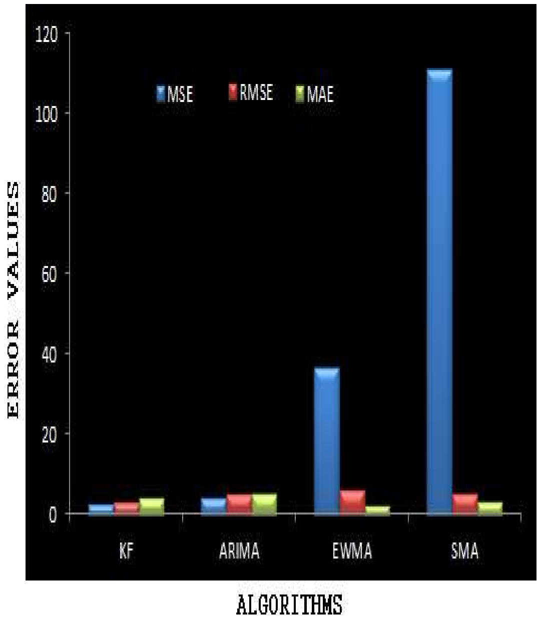

Figure 5. Comparison of SO2 chart errors using different algorithms

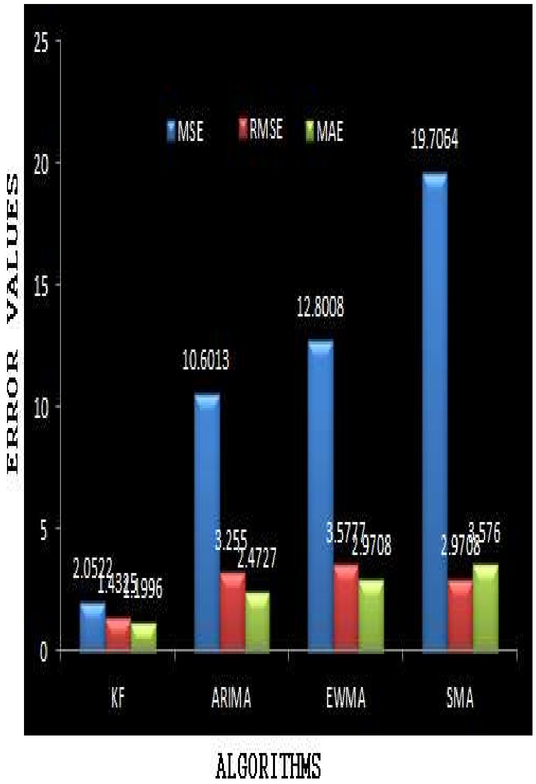

Figure 6. Comparison of NO2 chart errors using different algorithms

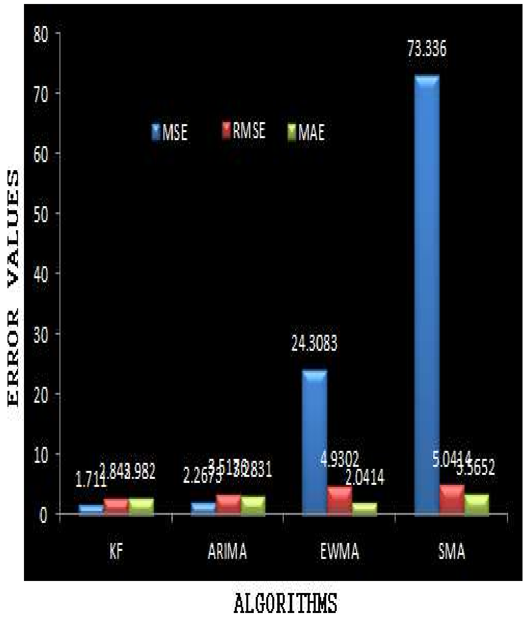

Figure 7. Comparison of PM10 chart errors using different algorithms

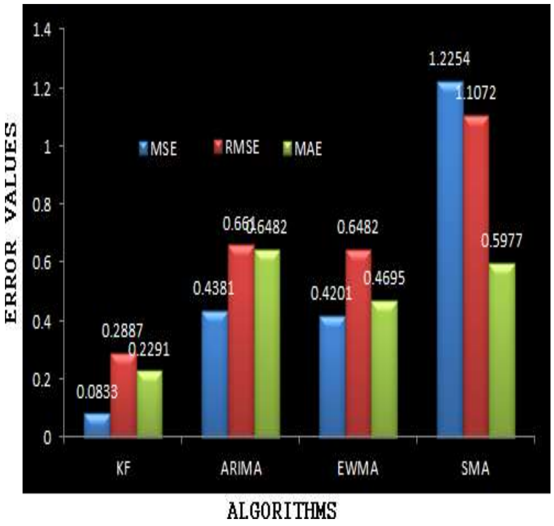

Figure 8. Comparison of PM2.5 chart errors using different algorithms

From results of predicted and observed values, the error indicator givenin the above table. From the table it can be concluded that KF algorithm yields much less error in the out put for all the pollutants data analysis. Except MAE, both MSE and RMSE were in very low range when KF is used. This indicates that out of all the four algorithms selected and implemented in the system, KF shows the better performance.

LIMITATIONS OF THE STUDY

Despite achieving improved forecasting accuracy, the proposed system has certain limitations. The dependency on low-cost sensors may introduce measurement inconsistencies under varying environmental conditions. Additionally, the system has been validated in a limited geographical region, which may affect generalizability. Future large-scale deployment may also require enhanced synchronization and calibration mechanisms.

FURURE SCOPE

Future work may explore hybrid Kalman-Deep Learning models, multi-modal data integration, and federated edge learning for secure large-scale environmental monitoring.

CONCLUSION

The major scientific contribution of this work lies in integrating edge computing with Kalman Filter-based forecasting for low-cost and real-time air quality prediction. This work constructed an edge computing and IoT system which monitors and forecast air quality using the Raspberry Pi and the Kalman Filter technique. This work uses machine learning instead of the traditional cloud-based monitoring and forecasting approach for air quality, which eliminates data transfer delays. Six different categories of contaminants of air viz., SO2, NO2, PM10, and PM2.5 are included in the real-time forecast data that are produced by running the Kalman Filter on powerful edge device, the Raspberry Pi. Comparing SMA, EWMA, and ARIMA algorithms reveals that, despite using low-accuracy sensors, the forecast results produced on the KF had the least amount of inaccuracy. In addition, the RMSE figure dropped to 68.3 per cent. In a similar vein, the KF algorithm improves sensor accuracy from the algorithm's point of view by 30% when compared to sensor observation data. The findings demonstrate that integrating edge computing with Kalman filtering provides a practical and computationally efficient framework for real-time air quality prediction. The proposed approach offers significant benefits in reducing latency, improving sensor accuracy, and enabling scalable deployment for smart city environmental monitoring applications. This study contributes to existing knowledge by demonstrating that lightweight recursive filtering techniques can effectively replace computationally intensive prediction models in resource-constrained IoT environments.

REFERENCES

[1] Al-Ali, AR, Zualkernan, I & Aloul, F 2010, 'A mobile GPRS-sensors array for air pollution monitoring', IEEE Sensors Journal, vol. 10,no. 10, pp. 1666-1671.

[2] Atzori, L, Iera, A & Morabito, G 2010, 'The internet of things: A survey', Computer Networks, vol. 54, no. 15, pp. 2787-2805.

[3] Bacco, M, Delmastro, F, Ferro, E & Gotta, A 2017,'Environmental monitoring for smart cities', IEEE Sensors Journal, vol. 17, no. 23, pp. 7767-7774.

[4] Baker, KR & Foley, KM, 2011, 'A nonlinear regression model estimating single source concentrations of primary and secondarily formed PM2:5', Atmos. Environ., vol. 45, no. 22, pp. 3758-3767.

[5] Benammar, M, Abdaoui, A, Ahmad, S, Touati, F & Kadri, A 2018, 'A modular IoT platform for real-time indoor air quality monitoring', Sensors, vol. 18, no. 2, p. 581.

[6] Bengio, Y, Simard, P & Frasconi, P 1994,'Learning long-term dependencies with gradient descent is difficult', IEEE Trans. Neural Netw., vol. 5, no. 2, pp. 157-166.

[7] Bilal, K, Khalid, O, Erbad, A& Khan, SU 2018, 'Potentials, trends, and prospects in edge technologies: Fog, cloudlet, mobile edge, and micro data centers', Computer Network, vol. 130, pp. 94-120.

[8] Choi, S, Kim, N, Cha, H & Ha, R 2009, 'Micro sensor node for air pollutant monitoring: Hardware and software issues', Sensors, vol. 9, no. 10, pp. 7970-7987.

[9] Devarakonda, S, Sevusu, P, Liu,H, Liu, R & Badrinath 2013,'Real-time air quality monitoring through mobile sensing in metropolitan areas', in Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing, pp. 1-8.

[10] Dhingra, S, Madda, RB, Gandomi, AH, Patan, R & Daneshmand, M 2019, 'Internet of Things mobile–air pollution monitoring system (IoT-Mobair)', Internet of Things Journal, vol. 6, no. 3,pp. 5577-5584.

[11] Esling, P & Agon, C 2012,'Time series data mining', ACM Journal of Computing Surveys', vol. 45, no. 1, pp. 12:1-12:34.

[12] Ferdoush, S & Li, X 2014, 'Wireless sensor network system design using raspberry PI and Arduino for environmental monitoring applications', Procedia Computer Science, vol. 34, pp. 103-110.

[13] Gulia, S, Khanna, I, Shukla, K & Khare, M 2020, 'Ambient air pollutant monitoring and analysis protocol for low and middle-income countries: An element of comprehensive urban air quality management framework', Atmospheric Environment, vol. 222, pp. 1-25.

[14] Hajji, W& Tso, F 2016, 'Understanding the performance of low power raspberry Pi cloud for big data', Journal of Electronics, vol. 5, P. 29.

[15] Hochreiter, S & Schmidhuber, J 1997,'Long short-term memory', Neural Computer, vol. 9, no. 8, pp. 1735-80.

[16] Hoek, G, Beelen, R, Hoogh, K, Vienneau, D, Gulliver, J, Fischer, P & Briggs, D 2008,'A review of land-use regression models to assess spatial variation of outdoorair pollution', Atmos. Environ., vol. 42,pp. 7561-7578.

[17] Jadhav, G, Jadhav, K & Nadlamani, K 2016, 'Environment monitoring system using raspberry-Pi', International Journal of Engineering Technolgy (IRJET), pp. 2- 4.

[18] Jeya, S & Sankari, L 2020, 'Air pollution prediction by deep learning model', in 2020 4th International Conference on Intelligent computing and Control Systems (ICICCS), IEEE, pp. 736-41.

[19] Kularatna, N & Sudantha, BH 2008, 'An environmental air pollution monitoring system based on the IEEE 1451 standard for low-cost requirements', IEEE Sensors Journal, vol. 8, no. 4, pp. 415-422.

[20] Kumar, A & Hancke, GP 2014, 'Energy efficient environment monitoring system based on the IEEE 802.15.4 standard for low cost requirements', IEEE Sensors J., vol.14, no. 8, pp. 2557-2566.

[21] Li, X, Peng, L, Yao, X, Cui, S, Hu, YC & Chi, T 2017, 'Long short-term memory neural network for air pollutant concentration predictions: method development and evaluation', Environ. Pollution, vol. 231, no. 1, pp. 997-1004.Linh An, PM & Kim, T 2018, 'A study of the Z-Wave protocol: Implementing your own smart home gateway', in 3rd International Conference on Computer and Communication Systems (ICCCS), Nagoya, pp. 411-415.

[22] Ma, Z, Li, C, Cheng, PJ, Ding, Y, Lin & Xu, Z 2020, 'Air quality prediction at new stations using spatially transferred bi-directional long short-term memory network', Science of The Total Environment, vol. 705, no. 13, pp. 57-71.

[23] Manuel, A & Michel 2012, 'Review of small commercial sensors for indicative monitoring of Ambientgas', Chem. Eng. Trans., vol. 30, pp. 169-174.

[24] Nimesh, R, Arora, S, Mahajan, K & Gill, A 2014, 'Predicting air quality using ARIMA, ARFIMA and HW smoothing', Model Assisted Statistics and Application, vol. 9, no. 2, pp. 137-149.

[25] Pascanu, R, Mikolv, T & Bengio, Y 2013, 'On the difficulty of training recurrent neural networks', in Proceedings of the 30th International Conference on International Conference on Machine Learning, Atlanta, Georgia, USA, vol. 28, pp. III–1310-III–1318.

[26] Qi, Z, Wang, T, Song, G, Hu, W, Li, X & Zhang, Z 2018, 'Deep air learning: Interpolation, prediction, and feature analysis of fine-grained air quality', IEEE Transactions on Knowledge and Data Engineering, vol. 30, no. 12, pp. 2285-2297.

[27] Qin, D, Yu J, Zou G, Yong, R, Zhao, Q & Zhang, B 2019, 'A novel combined prediction scheme based on CNN and LSTM for urban PM2.5 concentration', IEEE Access, vol. 7, pp. 20050-20059.

[28] Qiu, T, Chen, N, Li, K, Atiquzzaman, M & Zhao, W 2018, 'How can heterogeneous internet of things build our future: A survey', in IEEE Communications Surveys & Tutorials, vol. 20, no. 3, pp. 2011-2027.

[29] Raipure, S & Mehetre, D 2015, 'Wireless sensor network based pollution monitoring system in metropolitan cities', IEEE: Piscataway, NJ, USA, pp. 1835-1838.

[30] Rybarczyk, Y & Zalakeviciute, R 2016, 'Machine learning approach to forecasting urban pollution', IEEE: Piscataway, NJ, USA, pp. 1-6.

[31] Rybarczyk, Y & Zalakeviciute, R 2016, 'Machine learning approach to forecasting urban pollution', IEEE: Piscataway, NJ, USA, pp. 1-6.

[32] Salamone F, Belussi L, DanzaL, Galanos T, Ghellere M & MeroniI 2017,'Design and development of a wearable wireless system to control indoor air quality and indoor

lighting quality', Sensors, vol. 17, no. 5, P. 1021.

[33] Salman G, Heryadi Y, Abdurahman, E & Suparta, W 2018, 'Single layer & multi-layer Long Short-Term Memory (LSTM) model with intermediate variables for weather forecasting', Procedia Computer. Sci., vol. 135, pp. 89-98.

[34] Satyanarayanan, M 2017, 'The emergence of edge computing', Computer, vol. 50, pp. 30-39.

[35] Shinde, D & Siddiqui, N 2018, 'IoT Based environment change monitoring & controlling in greenhouse using WSN', IEEE: Piscataway, NJ, USA, pp. 1-5.

[36] Siew, LY, Chin, LY, Mah, P & Wee, J 2008, 'Arima and integrated affima models for forecasting air pollution index in shah alarm, selangor', Malaysian Journal of Analytical Science, vol. 12, no. 1, pp. 257-263.

[37] Sikary, A, Dixit, S & Murty, OP 2017, 'Fatal carbon monoxide poisoning: A lesson from a retrospective study at all India Institute of Medical Sciences, New Delhi', Journal of Family Medicine and Primary Care, vol. 6, no. 4, pp. 791-794.

[38] Sinnott, RO & Guan, Z 2018, 'Prediction of air pollution through machine learning approaches on the cloud', in Proceedings of 2018 IEEE/ACM 5th International Conference on Big Data Computing Applications and Technologies (BDCAT), pp. 51-60.

[39] Srivastava, N, Hinton G, Krizhevsky, A, Ilya, S & Salakhutdinov, R 2014, 'Dropout: A simple way to prevent neural networks from over fitting', Journal of Machine Learning Research, vol. 15, pp. 1929-1958.

[40] Zhang, Y & Yu, J 2013, 'A study on the fire IOT development strategy', Procedia Eng., vol. 52, pp. 314-319.

[41] Zhao, L, Wu, W & Li, S 2019, 'Design and implementation of an IoTbased indoor air quality detector with multiple communication interfaces', in IEEE Internet of Things Journal, vol. 6, no. 6, pp. 9621-9632.

[42] Zhu, J, Ruoling Zhang, Binbin Fu & Renhao Jin 2015, 'Comparison of ARIMA model and exponential smoothing model on 2014 air quality index in Yanqing county,Beijing, China', Applied and Computational Mathematics, vol. 4, no. 6, p. 456.

[43] Zhu, J, Ruoling Zhang, Binbin Fu & Renhao Jin 2015, 'Comparison of ARIMA model and exponential smoothing model on 2014 air quality index in Yanqing county,Beijing, China', Applied and Computational Mathematics, vol. 4, no. 6, p. 456.