1. INTRODUCTION

This chapter briefly introduces air pollutions and its hazardous effects, the role of the Internet of Things in implementing a system for air quality monitoring, and the employability of deep learning in predicting air quality parameters. It also tells the motive behind this work, the problem statement, and the research contributions.

1.1 AIR POLLUTION AND HAZARDOUS EFFECTS

In the current period of a new industrial revolution broadly recognized as Industry 4.0, the internet of things(IoT), big data, and artificial intelligence have enabled a range of industrial sectors and started concentrating majorly on areas such as Smart Agriculture, Smart Transportation, Smart Environmental Monitoring, Smart Manufacturing and many more. Such industrial sectors have revamped the practices of traditional industries by facilitating 24 * 7 interconnectivity, automation, real-time data analytics, and IoT-based predictive maintenance. With the latest advancements and innovations in the new phase of the industrial revolution, air pollution has emerged as one of the major concerns of the 21st century. Due to the rapidly emerging economy, an upsurge in manufacturing zone development, and industrial progressions, air quality or pollution monitoring has emerged as a worrying issue for society.

Air pollution is defined as the contamination of the nearby atmosphere by several gaseous, fluid, and concrete wastes generating hazardous effects on health of humans, the well-being of plants and wild life worldwide [1]. These air pollutants can potentially produce numerous health risks in multiple illnesses like sinus infection, asthma, allergy syndrome, toxic dust syndrome, nervous system related threats, and several more.

According to [5], the substantial worry of the air quality is a high concentration of particulate matter in the surrounding atmosphere. Furthermore, incomplete coal combustion and vehicle emissions may create detrimental effects on health of human beings because they release particulate matter (PM2.5) directly into the atmosphere. Carbon monoxides(CO) are also gas produced mainly through the incomplete combustion of fuel. Automobile engines contribute 80% of Carbon monoxide emissions in urban areas [14]. In urban areas, smog is an irresistible environmental problem, especially in big metro cities. The primary constituent of smog is ground-level ozone, generated due to the chemical reaction of Carbon monoxide, VOCs, and heat produced from the sun. So the Carbon monoxide is the primary pollutant in forming smog in urban areas [15]. So Particulate matter and carbon monoxide has emerged as a prime air pollutant in urban areas. PM 2.5 or 10 exposure for the long term can be the reason for increasing circulatory mortality, cardiac danger, lung harm, and subclinical atherosclerosis. The outermost layer of skin also can be damaged due to such long-term exposures [7]. Short-term exposures to particulate matter are associated with cardiovascular death, stroke mortality, and hospital admissions [6,10]. Carbon monoxide (CO) present in the atmosphere is responsible for an enormous proportion of the poisonings and life losses reported all around the globe. Although the medical symptoms of CO poisoning are nonspecific, Dizziness, shakiness, and nausea, or vomiting are the recorded appearances. Neurological signs include consciousness fluctuations, neck stiffness, tremor, and a positive Babinski's indication (condition of the nervous system causing reflexes to respond unusually). Children may be more vulnerable to CO exposure [12]. In various cases, Carbon monoxide levels increase to a level that sometimes causes coma and even death [13].

As per the reports [2], in-house air pollution and outdoor air pollution had caused four million and three million pollution-related demises. Most indoor and outdoor pollution-related deaths occurred due to heart, lung, and respiratory diseases [9,11]. The WHO report [2] also suggests that about 91% of the world's population lives in places where air quality levels have exceeded WHO prescribed limits. Fresh and clean air is seeing to be the essential and must requirement of human health. However, air pollution continues to create a significant threat to health worldwide. The WHO report [2,5] also indicates that more deaths have happened due to pollution-related issues than road accidents in the last decade. Every year, greater than 2 million premature demises occur due to the effects of urban outdoor air pollution and indoor air pollution, as per WHO assessment of the liability of illness due to air pollution. The people of developing countries tolerate more than fifty percent of this illness problem. Due to quick-growing ecological issues in developing countries like India, air quality index monitoring has risen severely. Due to gas emissions from vehicles and industries, the air quality is substandard in urban areas and metro cities like Delhi, Ahmedabad, Kolkata, etc. In the city Delhi, the air quality index has recently gone across the bar of thousand, which may cause enormous breathing problems, cancer-like diseases, and chronic respiratory conditions to the country's citizens. In thickly populated and rapidly urbanizing areas, deaths connected with surrounding air contamination have expanded. Such air pollution-related issues have arisen due to rapid urbanization, enormous population growth and the shortfall in the availability of the required infrastructure, and a lack of a pre-defined health framework for air quality monitoring in such developing countries [3,4].

The combined study of the World Bank and the Institute for Health Metrics and Evaluation (IHME) considers assessing the costs of unexpected demises recognized with air contamination. An estimated 5.5 million life losses occurred in 2013 to diseases linked with outside and in-house air contamination, causing decreasing physical and financial asset development. These deaths strike primarily young kids and the elderly, resulting in missing labour income for working-age group men and women. Finding [16, 17] says that the cost of income loss due to such deaths is about 1 percent of Gross Domestic Product(GDP) in southern Asia and around 0.25 percent in Pacific and East Asia. In Africa, where air contamination damages the potential of young people's earnings, such losses are about 0.60 percent of Gross Domestic Products.

1.2 ROLE OF IOT IN ENVIRONMENT MONITORING

Internet of Things(IoT) is emerging as an essential technology in developing smart cities across the globe. The government is also using the Internet of Things to provide various public sector services to the citizens. "Internet of Things" (IoT)is a system that extends communication to "things" or physical objects. Interrelated machines or everyday devices embedded with sensors or other hardware/electronic devices and software can communicate and provide identity to other devices over the internet without active human interaction. Any "Things" (machines) connected in the IoT are having the following characteristics [18].

- Things positioned in an environment, which has the competencies of transferring the data over the internet to the other Things
- Can program Things to act consequently
- The Things obtain information from the internet
- The Thing is part of a group of devices that can interconnect with other nodes in the same network

These provide an opportunity to integrate computer(controller/processor) based systems into the physical methods and ambiances in the real world. Internet of Things is being empowered by prominent technologies such as Big data and Hadoop for scaling and data handling, machine learning, data mining, and deep learning for application-specific outcomes and analysis.

Internet of Things can bring vast revolutions in the history of computing, even though it is a new technology. The IoT can develop an enormous quantity of applications in various domains, out of which only a limited portion is available to society. The Internet of Things is a technological paradigm expected to grow the connectivity of day-to-day devices. In the subsequent years, the progression and usage of the technology will rise exponentially in multiple fields. The projected growth, the presently assessed 5 billion IoT devices in practice, will increase to 25 billion by the end of 2021[19]. Internet of Things is all about the rise in machine-to-machine (M2M) communication and wireless embraces, integrated sensors, and actuators that assist in monitoring and regulating devices remotely and efficiently. Still, many domains have the scope for development of the application that can improve the living quality of human beings: when roaming, when staying at home, when at work, when jogging or at the gym, or when at the hospital, these are just a few to be cited. These ambiances with Things(objects) with some essential intellect and some communication capabilities allow these things to communicate with others and convey the data or information from surroundings continuously or at a particular event. Surroundings infer having various domain-specific environments in which developers can deploy many applications. Such applications can be grouped in the following significant domains but not limited to the given list [20, 21]:

- Transportation and logistics
- Health care
- Smart environment
- Smart city (traffic management, water leakage detection, parking management, etc.)
- Personal and social domain
- Home automation
- Smart farming and agriculture
- Smart Inventory management
- Energy consumption and tracking

Nowadays, metropolises are experiencing essential functioning alterations. These novel models recognized as Smart Cities target to improve the quality of life of people utilizing collected data about the immediate environment through the use of IoT. Conventionally, the immobile fixed air quality monitoring stations built mainly by environmental or governmental establishments measure the air pollution or air quality parameters concentrations. The key advantages of such air quality monitoring stations lie in assessing accessibility for a range of air pollutants and the measurement reliability. This nature of measurement can be beneficial for long-term pollution estimation at certain spots or regions. Though air pollution or air quality monitoring using such fixed stations are suffering from the low spatial determination of the data, that may contribute to inaccurate assessment over the whole sizeable targeted area of the study. One more disadvantage is the very high cost of installation and functioning of such air quality monitoring stations. It is necessary to develop tools and technology by which equipment deployed in small clusters can sense changes in the air pollutant in the surroundings and then report for monitoring and action based on the changes. It is entirely possible with the Internet of Things; when Things- sensors are interfaced and programmed, we can fetch data from the environment. Nowadays, much research is going on in air sensing sensor technologies based on electrochemical, resistive, photoionization, and dispersive infrared radiation absorption technologies. With the development of these technologies and the invention of low-cost air quality sensing technology and the IoT suits as the economical alternate to develop air quality parameters monitoring systems against the costly and fixed air quality or pollution monitoring stations. The IoT-based air quality monitoring system can better notify and monitor the influence of guidelines to control air quality. IoT-based solutions can be beneficial due to the feasibility of Planning the deployment at vulnerable sites. IoT-based solutions authorize city inhabitants to make more informed choices to restrict their exposure to harmful air pollution.

1.3 DEEP LEARNING IN AIR QUALITY PREDICTION

The air quality forecasts fetched numerous researches in the early era, and two approach types are typically used, that are deterministic and statistics based approach [22]. The deterministic technique applies theoretical atmospheric emissions and chemical models to pretend air pollutant release, transfer, and dispersal processes with the use of dynamic data of limited air quality monitoring stations in a model-driven approach. Though, because of unreliable emission data of pollutants, complex original surface circumstances, and an imperfect theoretical basis, such model outcomes undergo low prediction accuracy. Compared to the complicated theoretical approaches, statistics-based approaches employ the statistical modelling method to forecast air pollutants in a manner that is data-driven. The techniques such as the Gaussian model, multiple linear regression, and the autoregression moving average (ARMA) are used for air quality prediction. These approaches typically produce partial accuracy due to the incompetence to model nonlinear behaviours and patterns.

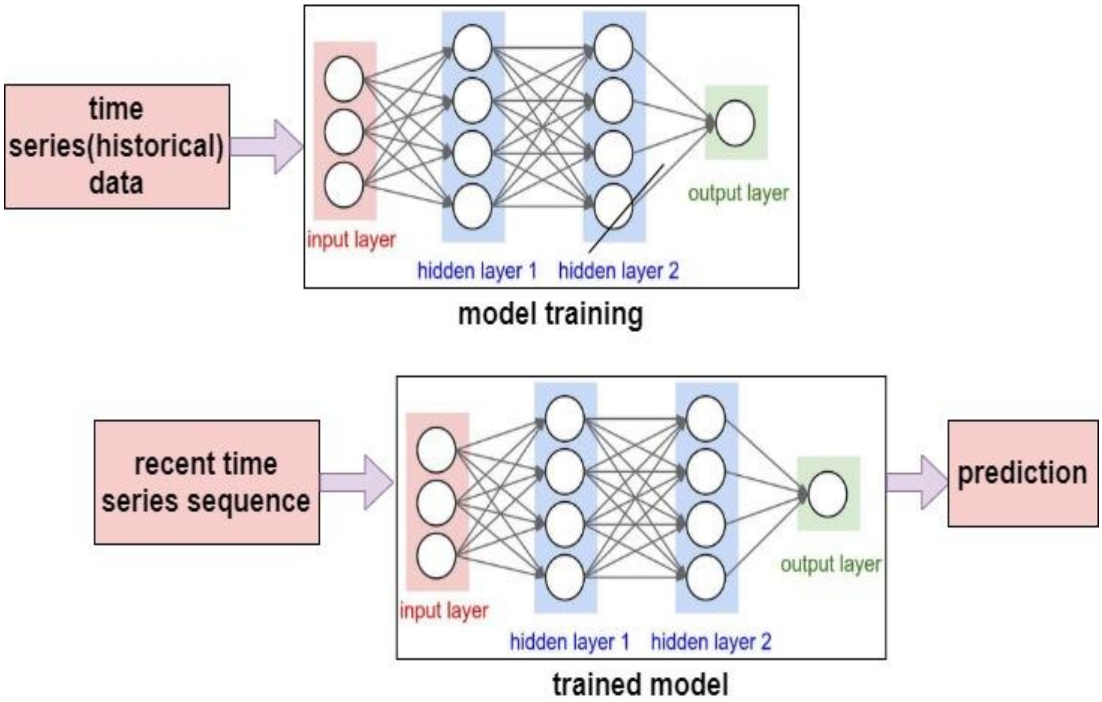

Figure 1.1. Deep learning model

Recently, deep learning, an innovative perspective machine learning methodology, has gained significant academic and research attention. Deep learning, also termed deep structured learning, is a portion of a broader family of machine learning methods. Deep learning has been effectively employed for the classification of images, processing of natural language, forecasting or prediction job, object recognition, artificial intelligence problem solving, gesture as well motion modelling, etc. Deep learning approaches utilizes multi-layer based architectures to fetch the integral features from the data, layer after layer starting with the bottom to the top level. They can recognize descriptive structure in the input data. So, in deep learning, every level studies to convert input data into additional abstract and merged representation. As the air quality prediction process is complex, time-based or temporal tendencies and spatial spreading are influenced by several causes, such as discharge of air pollutants, climate situations, traffic rush, human actions, etc. This condition has enlarged the inefficiency of using conventional models, particularly for simulating a better representation of air pollutant features. The recently appreciated learning method called Neural Networks (NN) or Artificial Neural Networks (ANN) is utilized to classify data mostly. Recurrent Neural Network (RNN) contributes to the application in time-series prediction because of the back-propagation theory and the concept incorporated in the underlying model or architecture. Time series analysis for forecasting [23] is the division of temporal data mining. The acquirement of an enormous quantity of data over a continuous time duration is named Time-series data. The frequency of collection can be years, weeks, days, hours, or minutes. The changes in historical time series data are inspected and analysed by building the training model. Then the trained model is utilized to estimate(prediction) the future patterns, as shown in figure 1.1.

1.4 MOTIVATION FOR THIS WORK

Air pollution is the reason for critical concern for the current century. IoT technologies have the power to make it possible to monitor air quality with real-time reporting of updates of sudden deviations in the air quality [8]. Various researchers have provided solutions for air quality monitoring using IoT-based technologies. The IoT-based air quality monitoring remains an open research domain due to the issues and challenges of the IoT field such as architectural complexity, lack of standardization efforts, low memory capacity, power consumption, sensor interfacing, security, privacy as well reliable delivery implementation and assessment.

- Air quality monitoring and real-time reporting with handy tools like the mobile application can help avoid health hazards and ease living in an urban area.
- The logged (historical) data can help in the identification of characteristics, periodic patterns, causes(sources) of air pollution, distribution of pollutants at various regions (intracity and intercity) of air pollutants.
- Forecasting helps the user in providing early warnings about the pollutants. Developing an accurate air quality (air pollution) parameter forecasting model or system in urban areas can help people avoid health hazards by canceling to be outdoors or visiting certain critical places.
- Also, the forecasting results inform government agencies to execute traffic control or any such policy implementation looking at the critical levels of air pollution in the future.

1.5 PROBLEM STATEMENT AND OBJECTIVES

1.5.1 Problem Statement

- To build IoT based system that collects air quality parameters from any remote resource-constrained site, processes and transmits the observation in real-time to the server by addressing various issues of IoT architecture, sensor interfacing, reliable delivery, and power consumption.

- To process, store, and visualize the fetched parameters and do the prediction using deep learning from the available historical observation data.

1.5.2 Objectives

- To identify critical air quality parameters and prepare IoT-based intelligent node that comprises parameter-specific sensors, interfaced to the processing board by considering the economic aspect.
- To design or develop some strategy for power consumption optimization to increase the intelligent node's battery-lasting period.
- To standardize the IoT layered architecture and propose the complete framework for monitoring (processing, transmission, logging, and providing GUI) identified air quality parameters from the resource-constrained remote site.
- To test the system performance over parameters like throughput and reliability.
- To design a methodology or model based on deep learning for forecasting/predicting air quality parameters from the logged historical parameters.
- To transform raw data collected into the form suitable for such prediction model and test the system performance for prediction accuracy. Also, design strategy to improve performance further.

1.6 RESEARCH CONTRIBUTIONS

1. IoT-based systems suffer from challenges like no standardization and complex architectural design. In the proposed work, we standardize the layered architecture to provide the systematic framework comprising the data (air pollutants) collection from the site (where the sensing node deployed) to the logging. The framework also includes the rendering of the sensed parameters with GUI on various platforms like server (python scripts) and mobile interface(android) at the same time by availing the semantic interoperability and data harmonization among various objects. The underlying architecture is benefited from the following unique features.

- No relaying nodes (i.e., in the case of ZigBee-based implementation) or network setup is required.

- The system is implemented with broker-based architecture, and cloud broker HiveMQ is utilized for MQTT communication protocol implementation rather than any local broker. So, the system can be scaled easily.
- The system implements authentication of publisher or subscriber node at broker during registration phase by use of unique ClientID.
- The system implements an integrity check while reading parameters from sensors to avoid any physical level sensing error.

2. Power consumption is one of the significant challenges for the IoT-based monitoring system because of the battery-operated node deployments at resource-constrained sites. In the proposed research work, we provide a power consumption optimization scheme. The sensing operation of the smart node is divided into five phases. The method utilizes the sleep mode of the controller and particulate matter sensor with the help of customized hardware design for power saving during the sampling period. The proposed scheme can reduce the power consumption by almost 64% per sampling period, and the battery lasting period is increased 1.7 times, as exhibited in the experiment.
3. The average power consumption of the smart node depends on the number of readings taken frequently(periodically) and the actual transmission readings. The proposed system was also implemented and tested with a novel event-based transmission scheme to decrease the total transmissions and eventually save power usage. The proposed aperiodic transmission approach informs or reports the subscriber (transmits data) only when the substantial alteration in current (recent) reading is realized compared to the previous measurement. Under the employed scheme with the set threshold criterion, the transmissions are reduced from 240 to 64 (on an average) over 6 hour's transmission. Also, the implemented scheme takes care of the no reporting for a longer period to balance the tradeoff between the number of transmissions and tolerance of parameter changes for reporting.
4. The IoT-based monitoring systems for air quality can provide real-time sensed parameters to the subscriber or the remote server. Reliable delivery of the messages to the other end is one of the essential parameters in such a system. Reliable delivery or accuracy represents the Quality of Service delivered by such a diverse reporting system. So, Quality of Service implementation can add or enhance the value to the system by

availing the performance visibility and usability of the reporting services provided. Very few efforts have been presented to implement and assess the IoT-based monitoring system's performance, designed and developed under complex architecture. We have implemented and tested the system with two QoS level schemes. QoS level 0 is transmitted and forgets communication. In QoS 0, the publisher sends a message but does not worry about the actual delivery of the message at the other end. QoS level 1 guarantees MQTT packet delivery at least once. Even though the subscriber is temporarily unavailable due to low bandwidth or other issues, it will get the published messages from the broker once the subscriber is UP(available) again. Under the employed architecture with QoS level 0 also we can achieve 98% accuracy. QoS level 1 increases the accuracy performance at the cost of increased end-to-end delay.

5. RNN is a widely used approach for time series prediction due to the backpropagation with time algorithm usage for learning. We have used TensorFlow backend and Keras for timestep-based moving window training and prediction of future timestep value for three parameters: CO, PM 2.5, and PM 10. RNN suffers from a vanishing gradient problem as the sequence sample size increases, which also can be observed in the performance evaluation.
6. To overcome the problem and further minimize the loss function (increase performance) for longer sequences, we use simple LSTM and stacking with the forward LSTM +backward LSTM based model. The model is checked with four merge modes, and the prediction performance is further improved with the applied model. The proposed approach outperforms both RNN and Simple LSTM based prediction.
7. LSTM based cascaded model improves the performance but is found to be suffering from overfitting. We experimented with the model with two regularization techniques and found good performance with the dropout-based technique to overcome the issue. A proper hyperparameter setting also fine-tunes the model to improve the performance. We applied dropout layers in the model between hidden layers. Adding dropout layers makes the model dynamic by changing the number of LSTM cells while backpropagation is applied for a particular batch. We tested the model with varying dropout values and found good convergence for validation data at a 0.3 dropout value. The final loss for validation was also found to be further minimized and stabilized at around 300 epochs.

8. The attention-based mechanism is adopted to focus on timesteps that are more significant for prediction purposes. The addition of the self-attention mechanism improves the performance further and works well for longer sequences and extended time horizons. The applied self-attention mechanism is also one of the first attempts to the best of our knowledge in the field of sequence to sequence air quality prediction.

1.7 THE OVERALL ORGANIZATION OF THE THESIS

The overall thesis has been divided into seven chapters. The **first chapter**, as discussed above, was about air pollution and its hazardous effect, the role of IoT in air quality monitoring, deep neural network employment for air quality prediction. Moreover, the chapter discusses motivation, problem statement, objectives, and research contributions. The rest of the chapters are organized as under:

Chapter 2 gives a detailed review of the approaches carried out so far for air quality monitoring purposes across the globe, in the form of a literature study. It also discusses the literature review of various approaches attended by researchers for air quality prediction.

Chapter 3 gives an introduction to the necessary concepts of the internet of things ecosystem. It provides essential details of various protocols in the layered architecture of the internet of things and the current state of art sensing technology. It also discusses the necessary concepts of the primary neural network, recurrent neural network, and long short-term memory network.

Chapter 4 discusses the detailed architecture and implementation of the proposed air quality monitoring system and the proposed methodology employed for the air quality prediction system.

Chapter 5 gives the details of the performance of the proposed air quality parameter monitoring and prediction system.

Chapter 6 concludes the whole work and states the possibility of future work.

Chapter 7 lists out the publications that emerged through this research work.

Chapter 8 lists the references for the entire content.