iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.53176

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

Air Quality Forecasting Using Deep Learning Framework

Apurva[1], Sakshi Mahadik[2], Priya Pawar[3], Komal Rayrikar[4], Siddhi Argade[5]
Department of Computer Engineering, Zeal College of Engineering and Research, Pune, India.

Abstract: ***Air quality is a crucial environmental factor that affects human health and well-being. Accurate forecasting of air quality can aid in mitigating the adverse effects of air pollution. In this research paper, we propose a deep learning-based approach for air quality forecasting. We utilize Convolutional Neural Networks (CNN) and Bidirectional Long Short-Term Memory (BILSTM) models to capture spatial and temporal dependencies in air quality data. The existing system is analyzed, and its limitations are identified. Our proposed solution overcomes these limitations by incorporating deep learning techniques. Experimental results demonstrate the effectiveness of our approach in forecasting air quality, providing valuable insights for environmental management and public health initiatives.***
Keyword: air quality forecasting, deep learning, CNN, BILSTM, spatiotemporal data, environmental management, public health initiatives.

I. INTRODUCTION

Air pollution has become a major global concern, with significant impacts on human health and the environment. Accurate forecasting of air quality plays a vital role in developing proactive measures for pollution control and public health management. Traditional forecasting models often struggle to capture complex spatiotemporal patterns inherent in air quality data. In this research, we propose a deep learning-based approach using CNN and BILSTM models to address these challenges. The proposed models aim to leverage the inherent capabilities of deep learning in capturing both spatial and temporal dependencies, enabling more accurate air quality predictions. In the big data era, with the rapid development and application of the Internet of Things and sensor technology, air quality forecasting is increasingly dependent on a variety of sensors and related data acquisition equipment to collect the urban air big data, e.g., PM2.5, NO2, PM10, weather condition data and traffic data, etc. Since traditional shallow learning models still have bottlenecks in handling big data, new air quality forecasting methods need data-driven model support . Deep learning is currently the most popular data driven method , which can extract and learn the inherent features of various air quality data automatically. Since 2012, deep learning has made great progress in research and applications of image processing, audio processing, and natural language understanding . Although air quality forecasting task usually adopts the traditional shallow machine learning methods, the deep learning method for time series analysis and air quality prediction is getting more and more attention from researchers . In the issue of air quality forecasting, which is a typical multivariate time series analysis problem , it's a useful exploration of learning various implicit features and long temporal dependencies of multivariate air quality time series data based on the hybrid deep learning model.

According to World Health Organisation (WHO), every year more than seven million persons are dying because of this problem and more than 80% of urban areas population lives in hplaces where air quality rises over WHO guideline limits . As reported by Apte et al., the global and national life expectancy has been reduced because of air pollution. The study shows that in 2016 particulate matter with a diameter equal to 2.5 micrometres (PM2.5) reduced global life expectancy about 1.2–1.9 years in some polluted countries of Asia and Africa. According to the following research PM2.5 has severe effects for human life, becoming the reason of about 3% of mortality from cardiopulmonary disease, 5% of mortality from cancer of the trachea, bronchus, and lung, and about 1% of mortality from acute respiratory infections in children under five year. This study presents that PM2.5 in 2015 was the fifth-ranking mortality risk factor. Therefore, it is a crucial problem to prevent or reduce consequences caused by air pollution. Having information about air quality will induce us to make protective measures; it can lead the population to apply their daily activities in the places which are less polluted (by escaping high polluted areas). However, analysing the data, giving smart solutions remains as a challenging task. Thus, it is essential to apply productive methods and techniques for more effectively and more efficiently analysing big data, converting the invisible to visible, and extracting information hidden behind data.

©IJRASET: All Rights are Reserved

II. LITERATURE SURVEY

A. An Air Quality Prediction Model Based on a Noise Reduction Self-Coding Deep Network -May 2020

For this study, based on an LSTM network, a denoising autoencoder deep network (DAEDN) model was designed to solve the low prediction accuracy of existing air pollutant prediction models. Here, PM2.5 Concentrated Open Source Air Quality dataset is collected by Beijing's 12 air qualitymonitoring stations over 5 years, the prediction model of the study was analysed and verified.It contains the main input of the model, including historical air quality data, primarily the air quality index AQI and PM2.5, PM10, SO2, NO2, O3, CO, and other pollutant concentrations. The model's Bi LSTM structure makes good use of hhistorical and future information to eliminate thelag and improve prediction accuracy. The prediction p erformance of the study's surface DAEDN prediction model can be superior to the prediction r esults of Techniques Used:

1) DAEDN (Denoising AutoEncoder Deep Network)
2) PM2.5 Concentrated Dataset Processing
3) Air Quality Monitoring System

B. Real-time air quality forecasting, part II: State of the science, current research needs, and future prospects-February 2012

Part II specifies about RT-AQF models to address model deficiencies and improve forecast accuracies. This paper is giving idea about meteorological forecasts, chemical inputs, and model treatments of atmospheric physical, dynamic, and chemical processes in real-time air quality forecasting (RT-AQF) models.

Nationwide their are many cities which are implementing the Real time Air Quality Forecasting System that are based on tools and models with varying degrees of sophistication's ranging from the simplest rule of thumb to the most advanced 3D online-coupled meteorology and chemistry models. This system will be equipped with many modern technologies to reduce forecasting biases and enhance computational efficiencies including advanced techniques for multi-scale data assimilation The realization ofthis new generation of RT-AQF system will represent a significant landmark in thehistory of operational RT-AQF.

Technologies Used :

1) RT-AQF
2) Forecasting Model
3) 3D Online-Coupling meteorology

C. Modelling Air Quality in Street Canyons - Mar 2014

As per the survey carried out, it is understood that there is a vast decrease in Natural Ventilation inthe Urban Street Canyons. This paper includes two main datasets Traffic Data and Emissions. This discussion carries out a Mathematical Model which Calculate the Pollutant Concentrations by solving the Parametric Equations or by applying the algorithm numerically on a set of differential equations that describe in detail the Wind Flow and the Pollutant Dispersion as well as Pre-Concentration Monitoring and Sampling Techniques.

D. PM2.5 Concentration Prediction Using HSSM - Mar 2008

According to the Study, Prediction of Particular Matter in the air is an Important issue in Control reduction of Pollutants in the air. This Study generally involves Prediction of PM2.5 Concentration using Hidden Semi-Markov Model based " Time Series Data Mining. In this Paper, they Presented Hidden Semi-Markov Models for the prediction of High PM2.5 Concentration and its affect on the Air Quality Contents. Trained HSMM's can be used to obtain High PM2.5 Concentration level, which can be further used for PM2.5 Predictions. Also, in the dataset used, Vector Quantization is used to Discretize the Continuous data.

Techniques Used :

1) Hidden Semi-Markov Modeling
2) Vector Quantization
3) Time-Series Data Minin

III. EXISTING SYSTEM

The existing air quality forecasting systems often rely on statistical models or traditional time-series forecasting techniques. While these approaches have achieved some level of accuracy, they struggle to capture the intricate relationships present in complex spatiotemporal data

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

IV. PROPOSED SOLUTION

Our proposed solution overcomes these limitations by integrating CNN and BILSTM models. The CNN extracts spatial features from air quality data, while the BILSTM captures temporal dependencies. By combining these models, we can effectively capture both the local patterns and long-term dependencies in the data, resulting in improved air quality forecasting accuracy.

V. METHODOLOGY

A. Dataset Pre-processing

Very first task is Dataset Analysis. Under dataset analysis, their are multiple phases like Dataset Preprocessing, Data Preparation, feature extraction, etc. "Dataset Pre-processing" involves loading the dataset, analysing the dataset and knowing the essentials of the dataset. As, our Dataset is PM2.5 Dataset which consists multiple parameters w.r.to Air Quality which will predict/forecast the Quality of the air in our surrounding. The dataset contains of the attributes like Humidity, Wind pressure, Temperature, Contents of NO2, Sox, CO2, etc. Studying all these contents and finding their co-relations with each other will help to determine the contents affecting the Air Quality and their extent of damage to the Air Quality. Studying all these contents and finding their co-relations with each other will help to determine the contents affecting the Air Quality and their extent of damage to the Air Quality.

"Dataset Preparation " involves dealing with the analysed dataset in previous step. This generally will check for any inconsistencies in the dataset to be carried out. If their are any null values, missingvalues, Outliers, then this phase / Step is used to deal with such inconsistencies. Data Preparation will make the dataset suitable to be processed in further phases and make it in range in order to reduce the complexities for performing an operation to it. "Feature Extraction" is the last phase where the dataset is directly dealt with. In this phase, important features from the dataset which can be helpful for the predictions are been extracted inan efficient way in order to reduce the attributes for the study and reduce the time complexity. Feature Extraction is also called as Feature Selection, which involves the selection of best features or attributes which are alone capable to perform the analysis or predictions. During this phase or process most ofthe irrelevant attributes or features are been ignored or not considered for furtherprocessing. Actual Model which is applied is unaware of the rest features.

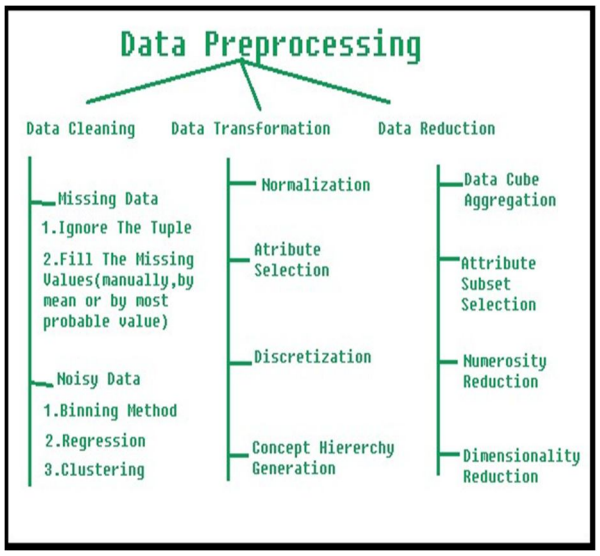


Figure 5.1: Data Preprocessing

B. 1D CNN

Convolutional Neural Networks (CNNs) are the Standardised Algorithms for multiple Computer Vision as well as Machine Learning Operations. It Consists ofspecial Artificial Neural Network (ANN) which take help of Subsampling layers to provide an efficient solution with the help of samples. It also consists of multiple deep hidden layers which can be trained in an efficient way in order to perform multiple simultaneous tasks which provides the feature of learning the complex objects and their respective patterns on the massive size from the visual database with their respective truth tables. This Property makes 1D-CNN capable of performing Multiple Engineering Applications. These applications can be used for 2D signals like images and their respective video frames. One of the most advantage of using 1D-CNN is that is considered as a real time, low-cost hardware whose implementation is much feasible due to its simpler and compact configuration.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

C. Bi-LSTM

Bi-LSTM is an abbreviation for Bi-directional Long Short term Algorithm. Bi-LSTM is used in this studywhich will focus on particular sequence which is both either it can be from front-to-back or from back-to-front. It will work same as a unidirectional counter with some additional features like the network not only connects with past but also connects with the future. Bi-LSTM in the study is used to detect the long temporal dependencies and the patterns from the dataset. The traditional statistical methods of analysis like ARIMA can't be considered to be useful in Processing Time Series Data because it does not take Long term temporal dependencies into account. Also, the Efficiency of such statistical models is low. To overcome all this problem, LSTM isbeen applied which is known to be a popular dynamic model for handling Sequence Tasks. According to the analyst it issaid that Bi-LSTM isthe good option for Popular dynamic Model which can handle Multiple Sequence tasks. Using this algorithm or Layer in the model is able to generate predictive output from the past and future contexts. For summary, it can be concluded that the Bi-Directional LSTM is used to process the Time Series dataset for Air Quality in two directions with an iterative process, i.e. Forward layer with t = 1 to t = T and the Backward Layer with t = T to t = 1

VI. RESULT

Key	Category	PM2,5 µg/m³	PM10 µg/m³	NO2 ppb	O3 ppb	SO2 ppb
1	Good	0-10	0-20	0-40	0-50	0-100
2	Fair	10-20	20-40	40-90	50-100	100-200
3	Moderate	20-25	40-50	90-120	100-130	200-350
4	Poor	25-50	50-100	120-230	130-240	350-500
5	Very Poor	50-75	100-150	230-340	240-380	500-750
6	Extremely Poor	75-800	150-1200	340-1000	380-800	750-1250

We will use key numbers for each category:

Good: 1 Fair: 2 Moderate: 3 Poor: 4 Very Poor: 5 Extremely Poor: 6

Figure 6.1: EEA AQI Table

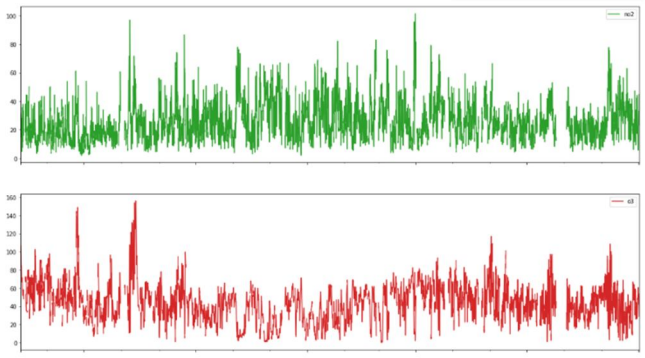


Figure 6.2: Concentration of no2 & o3

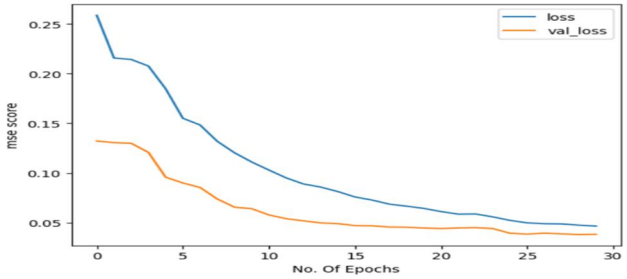


Figure 6.3: Epoch-wise MSE scores for pollutant

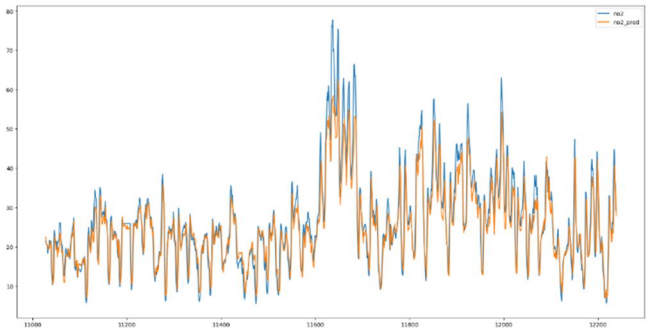


Figure 6.4: Original and Predicted Values of no2

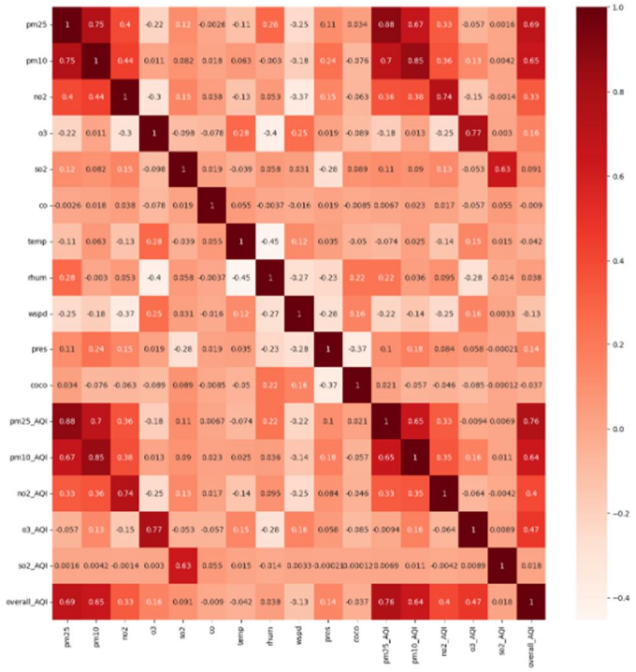


Figure 6.5: Original and Predicted Values of o3

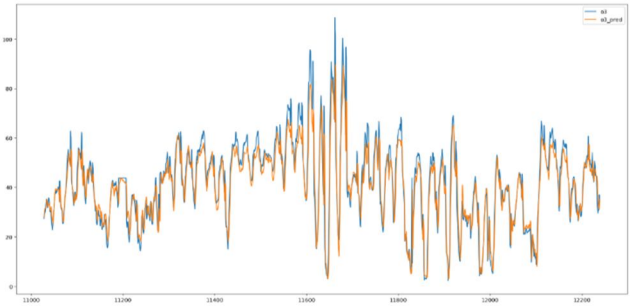


Figure 6.6: Air Pollutant Heatmap

VII. CONCLUSION

So, by considering all the study work and the detailed analysis, we prefer to choose the studying of Forecasting the Air Quality in the Surrounding relating with the Computer Science Domain which is Deep Learning. So, this Study will include the New Air Quality Forecasting Framework which will discuss about CNN Bi-LSTM levels for PM2.5 single step forward and Multi-Step forward Prediction, which is based on Hybrid Deep Learning Model.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

It demonstrates the Effectiveness of the model and helps to forecast about the contents in the Air to Predict the Air Quality and discusses about the Approach and the applications of Deep Learning with Neural Networks.

The Study will be fulfilling legal, operational, technical and Economical feasibility. There is wide range of scope for Air Quality hence surely it will be worth to predict Air Quality using the Model. And contains very few risks while developing the model.

This Study l include the New Air Quality Forecasting Framework which will discuss about CNN Bi-LSTM levels for PM2.5 single step forward and Multi-Step forward Prediction, which is based on Hybrid Deep Learning Model. It demonstrates the Effectiveness of the model and helps to forecast about the contents in the Air to Predict the Air Quality and discusses about the Approach and the applications of Deep Learning with Neural Networks.

A new hybrid deep learning framework which can deal with hierarchical feature representation and multi-scale spatial-temporal dependency fusion learning in an end-to-end process for air quality forecasting. This study was the first attempt to combine multiple one-dimensional CNNs and bi-directional LSTM for hybrid fusion learning of air quality related multivariate time series data, which can extract spatial-temporal dependency .

This include the New Air Quality Forecasting Framework which will discuss about CNN Bi-LSTM levels for PM2.5 single step forward and Multi-Step forward Prediction, which is based on Hybrid Deep Learning Model.

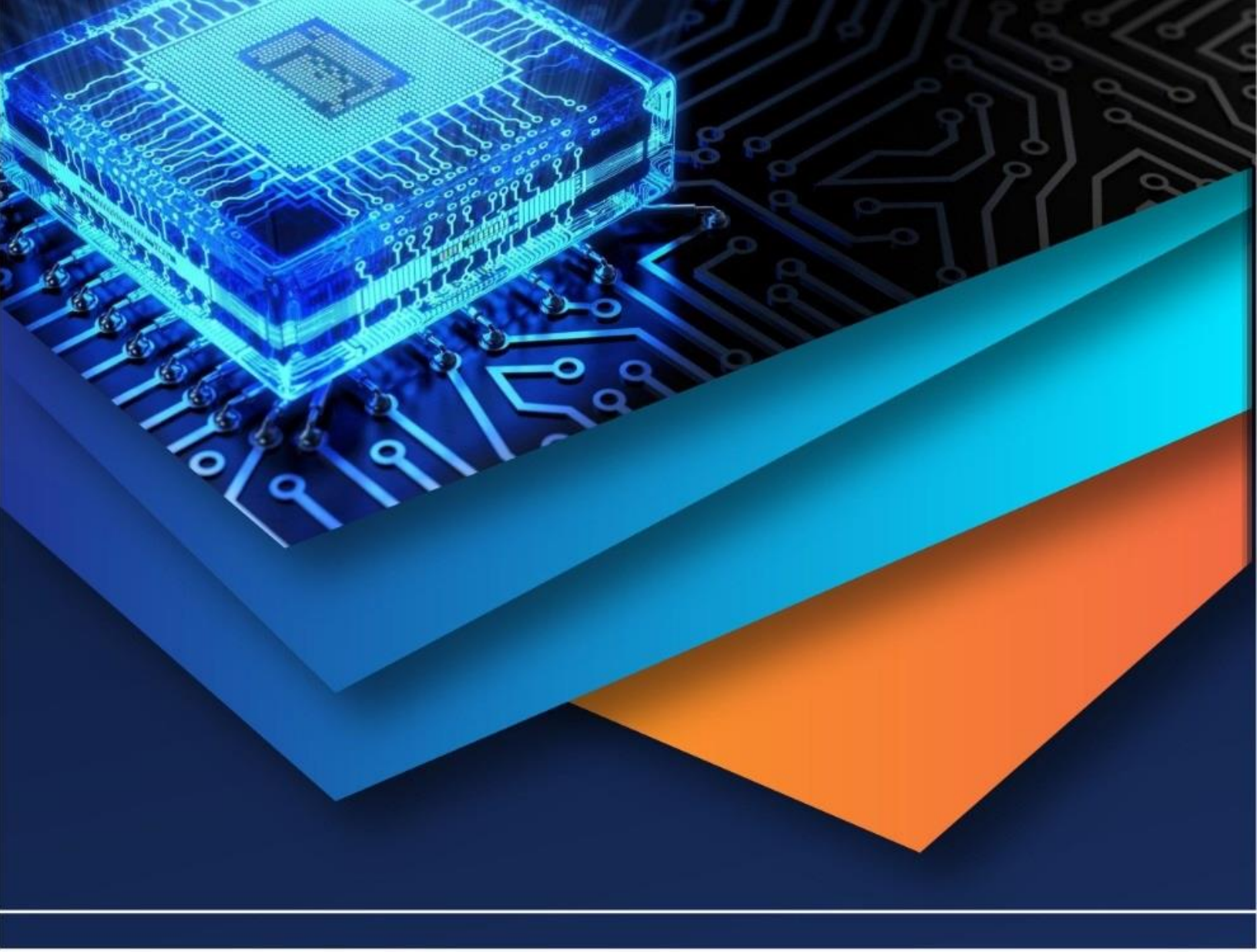
Hence we have drawn Activity diagram and deployment diagram for air quality forecasting using deep learning framework.

It demonstrates the Effectiveness of the model and helps to forecast about the contents in the Air to Predict the Air Quality and discusses about the Approach and the applications of Deep Learning with Neural Networks.

This include Air Quality Forecasting but we add one feature that can reduce the amount of pollutants from air and maintain the quality of air.

REFERENCES

[1] S. Vardoulakis et al., “Modelling air quality in street canyons: A review,” Atmos. Environ., vol. 37, no. 2, pp. 155–182, 2003.
[2] M. Dong, D. Yang, Y. Kuang, D. He, S. Erdal, and D. Kenski, “PM2.5 concentration prediction using hidden semi-Markov model-based times series data mining,” Expert Syst. Appl., vol. 36, no. 5, pp. 9046–9055, 2009.
[3] Xiuwen Yi , Zhewen Duan , Ruiyuan Li, Junbo Zhang , Member, IEEE,Tianrui Li , Senior Member, IEEE, and Yu Zheng Predicting Fine-Grained Air Quality Based on Deep Neural Networks, IEEE TRANSACTIONS ON BIG DATA, VOL. 8, NO. 5, SEPTEMBER/OCTOBER 2022.
[4] Y. Zheng, F. Liu, and H.-P. Hsieh, “An Air Quality Prediction Model Based on a Noise Reduction Self-Coding Deep Network -May 2020” in Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2013, pp. 1436–1444.
[5] J. Schmidhuber, “Deep learning in neural networks: An over- view,” Neural Netw., vol. 61, pp. 85–117, 2015.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)