iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** VII **Month of publication:** July 2021

DOI: https://doi.org/10.22214/ijraset.2021.36655

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VII July 2021- Available at www.ijraset.com

Analysis of Fine Gained Air Quality Using Deep Learning

Shivangi Singh[1], Ritika Tiwari[2], Dr. Atul Verma[3]
[1, 2, 3]*Department of Computer Science and Engineering, Shri RamSwaroop Memorial Group of Professional Colleges, Lucknow*

Abstract: With the rapid development of industrialization and urbanization, air pollution is increasing at an alarming rate in many developing countries. The four air pollutants which are becoming a concerning threat to human health are namely respirable particulate matter, nitrogen oxides, particulate matter and sulphur dioxide. The models which are currently used for air quality prediction by comparing to the AQI indexes do not give satisfactory results, which inspired us to examine the methods of predicting air quality based on deep learning using the K-Mean Clustering algorithm and Image Processing Technique. The interpolation, prediction, and feature analysis of fine-gained air quality are three important topics in the area of air computing. A good interpolation helps to estimate the limited air quality monitoring stations whose distribution is uneven in a city; an accurate prediction provides valuable information to protect humans and take necessary measures to reduce the effect of air pollution; a reasonable feature analysis is used to provide a more effective and general model. Overall, finding solutions to these topics can bring out extremely useful information to support air pollution control, and consequently generate great societal and technical impacts.
Keywords- Air quality prediction, Interpolation, Prediction, Feature Analysis, Deep Learning, Air Quality Index[AQI], K- Mean Clustering, Image Processing Technique

I. INTRODUCTION

Air pollution is an alarming environmental issue that is attracting attention globally. The global warming and depletion of the ozone layer has caused adverse effects like rise in sea level, fast melting of glaciers, rise in the temperature and unforeseen climatic changes. The pollution in air has seen a drastic increase after the industrial revolution. One of the dangerous parts of the air pollution is particulate matter, which is a composition of both liquid and solid particles. It may occur naturally due to volcanoes and fire bursts or by human involvement such as power plants and vehicle emission. The transition of population from rural to urban shifting is also seen as a leading factor in increase in pollution level. The three important topics in the area of air quality analysis are interpolation, prediction, and feature analysis. A good interpolation helps to estimate the uneven distribution of limited air quality monitoring stations in a city; an accurate prediction provides valuable information to protect humans and take necessary measures to reduce the effect of air pollution; a reasonable feature analysis is used to provide a more effective and general model. Overall, finding solutions to these topics can bring out extremely useful information to support air pollution control, and consequently generate great societal and technical impacts. However, there still exist few challenges in urban air computing such as insufficient monitoring stations due to the excessive cost of building and maintenance, incomplete and missing labels of historical data and also it is difficult to comprehend what sorts of information are the main applicable points for interpolation and prediction. In this paper, we introduced a deep learning based method for air quality prediction. We inferred the fine-grained air quality information throughout a city using historical air quality data reported by a limited number of existing monitor stations and a variety of data sets observed in the city. Deep learning can be successfully applied to image classification, prediction tasks and object detection processes. We have used K-means clustering, image processing techniques and deep learning in our model for prediction and feature selection. Our study will provide important information of air pollution levels and its quality in a particular city by using AQI, so that it will be helpful for the government to take necessary measures for pollution control.

II. LITERATURE REVIEW

Air pollution is one of the serious problems in the urban cities, where particulate matter (PM2.5) has a greater effect on humans than any other impure substances. This PM2.5 is a dangerous part of the air pollution due its adverse effects on humans as well as other living things. The proposed work aims at predicting air quality by using Random forest algorithm with air quality index provided by the respected government agencies[1]. In this work we make use of feature analysis for prediction processes to provide a more effective and general model. The proposed model considers various data taken for a number of hours and days from Bangalore, India.

This dataset will be helpful for training the system, later we compute real time datasets for predicting the pollution levels in that city, this predicted information will provide crucial information which will be helpful for pollution control and management. In this paper, an air quality prediction system using the Air quality Index and random forest algorithm is proposed.[2] The basic idea is to find out the amount of particulate matter present in the air which will be further helpful to find out pollution levels in a particular city. For this large number of datasets is gathered from a particular city for example, Bengaluru. Feature selection is used to reduce the number of attributes and obtain only relevant attributes which will be helpful for prediction. We use data mining techniques for extracting and cleaning the datasets and AQI is used for labelling it. The datasets are used for training the random forest system and a new set of datasets is used for testing the system. Finally, real time data is used for predicting the pollution levels.[3] The predicted output will provide crucial information which will be helpful for pollution control in urban cities. This proposed system provides highly reliable and accurate predictions. Further research can be done to improve the system to predict more accurately and user interfaces can be implemented to provide the predicted information to individuals for pollution control and management. Furthermore, studies will be able to find out the origin of the polluted particles and experiments will provide solutions to the pollution control in urban cities[4].

This paper discusses three vital matters in the area of urban air computing, namely, the interpolation, prediction, and feature analysis of fine-gained air quality. The solutions to these important topics can provide crucial data to guide air pollution control, and consequently generate great societal and technical impacts.[5] Most of the current models focus on solving the three issues one by one with the help of setting up different models for each. In this project, we improve an everyday and superb approach called DAL[6] to unify the interpolation, prediction, feature selection and evaluation of the fine-grained air pleasant into one model. To improve the performance of interpolation and prediction, we utilize the inherent qualities of the spatio-temporal facts and the data contained in the unlabeled records by embedding spatio-temporal semi-supervised learning on the external layer of the neural network. We additionally suggest a different approach to function selection[7] in the input layer of the neural network, whose optimization is easy to clear up and performance is good in eliminating the unnecessary information or irrelevant features. The proposed characteristic choice and analysis method has the capacity to show some internal working of the black box deep network models. Extensive experiments are conducted on real records sources showing that DAL is ultimate to the contrast opponents when solving the matters of interpolation, prediction, and feature analysis of fine-gained air quality.[8]

This work focuses on building an XGBoost model that considers meteorological data of Velachery, a commercial station in Tamil Nadu collected from the database of the Central Control room for Air Quality for Air Quality Management. The model is built by considering the following elements in air such as Carbon Monoxide, Benzene, Eth benzene, Nitrous oxide, Nitrogen Oxides, Ozone, PM2.5, SO2 and Toluene.[9] Other than the above mentioned pollutants, this model also considers the highly fluctuating meteorological parameters like pressure, relative humidity, wind speed, temperature and wind direction of the geographical region. Also, this paper attempts to rank the most influential meteorological factors that highly contribute to air quality. [10]The freshness of this model is that the decision trees that were integrated for a complete tree will be weighed based on the importance of the feature being considered. The meteorological factors of the model presented in this paper are ranked based on their contribution to the pollutant level. The factors are assigned weights between 0 and 1 based on the percentage of contribution. The weighted or ranked features are spanned as decision trees in the ensemble XGBoost. Since XGBoost spans shallow interpretable trees, the prediction could be made very easily.[11] This correlation among the factors also plays a crucial role, since the weakly correlated features are given very lower significance. This proposed model is deployed to predict the air quality index of the data obtained from the Central Control room for Air Quality Management. Velachery which is a commercial station in Tamil Nadu served as a testbed for the model. This model shows a reduced error rate when compared with other machine learning algorithms and also ranks the meteorological factors based on their order of importance. This model resulted in lower RMSE values which makes it suitable for real me AQI prediction.[12]

Air quality information, such as the amount of NO2, PM2.5, and PM10 in the air, for the real-time prediction, is very useful and important to control air pollution and protect human beings. In this paper, authors have infer the real-time and fine- grained air quality information throughout a city using (historical and real-time) air quality data reported by the monitoring stations[13] in the city and a variety of data sets were also collected that was observed in the city, like meteorology, traffic flow, human mobility, structure of road networks etc. The contribution of this paper lies in the following three aspects for prediction of air quality: The Authors have proposed a co-training-based semi-supervised learning approach, which leverages unlabeled data to improve the inference accuracy. Additionally, this approach consists of two classifiers respectively modeling the spatial and temporal factors that influence air qualities. Secondly, The Authors have also identified spatially-related and temporally- related features (e.g., humidity, temperature, and traffic flow), which contributes to not only this application but also the general problem of air quality inference.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VII July 2021- Available at www.ijraset.com

[14] The Authors have first calculated a co-training-based approach using the information obtained in Beijing, resulting in an overall (Precision=0.828, Recall=0.826) for PM10 and (Precision=0.808, Recall=0.798) for NO2. The results obtained from this paper is that of linear interpolation, a classical dispersion model, and some well-known supervised learning models like Decision Tree and CRF. So, according to the authors, applying all features to the SC or TC is worse than our co- training-based approach. Beijing (about 0.76). These results demonstrated the approach is applicable to different city environments and seasons. It is very useful and important to control air pollution and protect human beings. In this paper, authors have infer the real-time and fine-grained air quality information throughout a city using (historical and real-time) air quality data reported by the monitoring stations in the city and a variety of data sets were also collected that was observed in the city, like meteorology, traffic flow, human mobility, structure of road networks etc. The contribution of this paper lies in the following three aspects for prediction of air quality:[15] The Authors have proposed a co-training-based semi-supervised learning approach, which leverages unlabeled data to improve the inference accuracy. Additionally, this approach consists of two classifiers respectively modeling the spatial and temporal factors that influence air qualities. Secondly, The Authors have also identified spatially-related and temporally- related features (e.g., humidity, temperature, and traffic flow), which contributes to not only this application but also the general problem of air quality inference. The Authors have first calculated a co-training-based approach using the information obtained in Beijing, resulting in an overall (Precision=0.828, Recall=0.826) for PM10 and (Precision=0.808, Recall=0.798) for NO2. The results obtained from this paper is that of linear interpolation, a classical dispersion model, and some well-known supervised learning models like Decision Tree and CRF. So, according to the authors, applying all features to the SC or TC is worse than our co- training-based approach. Beijing (about 0.76). These results demonstrated the approach is applicable to different city environments and seasons.[16]

III. PROPOSED METHODOLOGY

Our proposed work consists of Air Parameters, Data Pre-Processing, K Means Clustering and Air quality index, air Quality prediction architecture is shown in figure below:

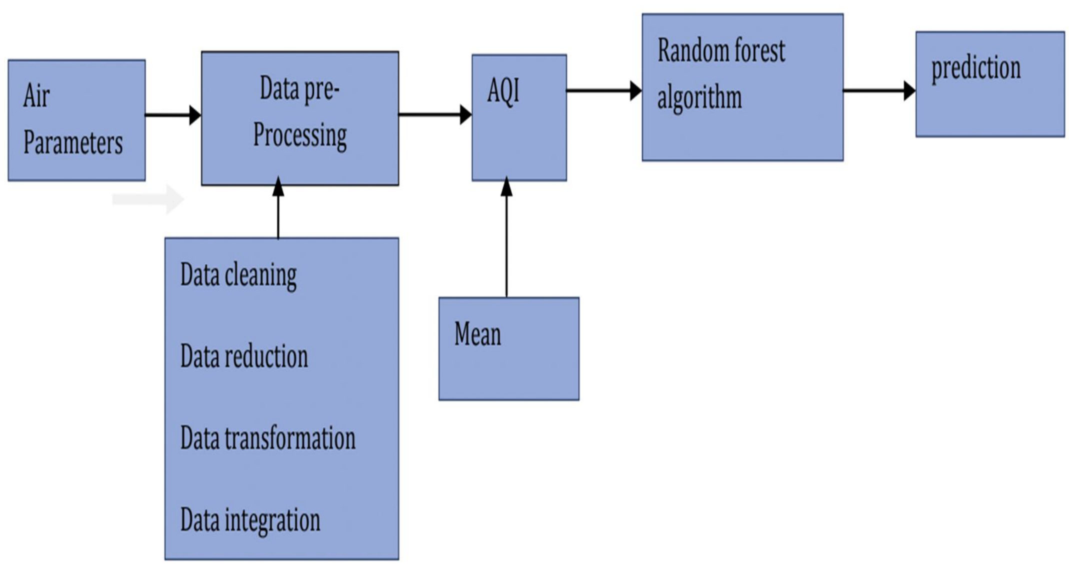


Fig.1

We introduced a deep learning based method for air quality prediction. We inferred the fine-grained air quality information throughout a city using historical air quality data reported by a limited number of existing monitor stations and a variety of data sets observed in the city. We utilize the characteristics of the spatiotemporal information, and also the information contained within the unlabelled information, to attain spatiotemporal semi-supervised learning within the neural network. Deep learning can be successfully applied to image classification, prediction tasks and object detection processes. After the collection of the dataset, data pre-processing methods are applied to the air parameters. Data pre-processing methods include Data Cleaning, Data Reduction, Data Transformation and Data Integration. We have used K-means clustering, image processing techniques and deep learning in our model for prediction and feature selection.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VII July 2021- Available at www.ijraset.com

A. Data Pre-Processing

The datasets collected for the prediction are in raw data format in which contain some unwanted data which aren't necessary for analysis and is not feasible. To overcome this problem data pre-processing is used where the dataset is cleaned and obtained from the raw datasets, this process is referred to as data pre-processing. This process involves several steps in order to convert raw data into a reduced clean data of small datasets.

Data pre-processing involves four steps, it includes:

1) Data Cleaning: This is the process of identifying and removing null values or unwanted values.
2) Data Integration: It is the process where many different data are compared to get a combined view.
3) Data Transformation: This is the process where data is transformed from one format to another.
4) Data Reduction: This is the technique which is applied to reduce the no. of parameters in the dataset, so that only the required attributes are used for the prediction.

B. K-Means Clustering Algorithm

K-means clustering is a method used for vector quantization. It aims to partition n observations into k clusters where each observation belongs to the cluster to the nearest mean, giving as a prototype of the cluster. The algorithm will categorize the items into k groups of similarity.This results in a partitioning of the data space into cells. This algorithm tends to find clusters within comparable spatial extent. This algorithm has a loose relationship with the nearest neighbor classifier, the machine learning technique used for classification is often confused with k-means due to the k in the name.

C. Air Quality Index (AQI)

Air Quality Index (AQI) is tables which categorise and has its own range which is used to predict air pollution level and its health impacts on humans and animals. AQI is implemented by the government in order to tell people how much the air is polluted. The increase in AQI results in a hazardous impact on human health which may lead them to severe health issues. The higher the AQI value, the more air is polluted and greater the health concern. When AQI values reach above 100, air quality is unhealthy. The AQI is divided into six categories. Each category corresponds to a different level of health concern. Each category also has its own specific color. These colors make it easy to quickly determine whether air quality is reaching unhealthy levels in their communities. Table III shows the adverse effect of air pollution:

Table III

Air Quality Index (AQI) Values	Levels of Health Concern	Colors
When the AQI is in this range:	*..air quality conditions are:*	*...as symbolized by this color:*
0-50	Good	Green
51-100	Moderate	Yellow
101-150	Unhealthy for Sensitive Groups	Orange
151 to 200	Unhealthy	Red
201 to 300	Very Unhealthy	Purple
301 to 500	Hazardous	Maroon

D. Image Processing Techniques

The image processing technique refers to processing of images which are digital, i.e. removing the noise and irregularities present in an image uploaded using the digital computers. All these noise or irregularities may enter into the image either during formation or during transformation taking place etc. For analysis mathematically, an image may be defined in two dimensional functions x and y where x and y are plane coordinates, and the amplitude of any of these pairs of coordinates x and y is called the intensity or gray level of the image at that point. When x, y, and the intensity values of a function are all finite, discrete quantities, it's called a digital image. a digital image is composed of a finite number of elements, and the element has a particular location and value. These finite elements are called picture elements or image elements . Pixel is the most widely used term for referring to digital images.

IV. RESULT AND DISCUSSION

The experiments demonstrate our proposed model which is an effective and promising solution to the topics of interpolation, prediction, and feature analysis of fine-gained air quality. The performance of the model is superior to the performances of the compared methods for interpolation and prediction. The main relevant and irrelevant features to the variation of PM2.5's concentration in Andhra Pradesh are revealed by feature selection and association analysis in. Further, the location of the pollution sources is also revealed by analysis of the interpolation results of PM2.5's concentration using. Fig.2 shows our dataset used while performing the experiment for the prediction of AQI and status of the city with the help of image processing technique. We are obtaining datasets from a government organization website known as "Central Pollution Control Board". From this we can obtain different datasets of various air parameters which exist in the air. Only specific parameters are considered to predict the pollution, where individual parameters have their own distinct range of values for predicting pollution.

View Dataset

Stn_code	Sampling Date	State	Location	Agency	Type	SO2	NO2	RSPM	SPM	Location Monitoring Station	PM25	Date
150	February - M021990	Andhra Pradesh	Hyderabad	NA	Residential, Rural and other Areas	4.8	17.4	NA	NA	NA	NA	2/1/1990
151	February - M021990	Andhra Pradesh	Hyderabad	NA	Industrial Area	3.1	7	NA	NA	NA	NA	2/1/1990
152	February - M021990	Andhra Pradesh	Hyderabad	NA	Residential, Rural and other Areas	6.2	28.5	NA	NA	NA	NA	2/1/1990
150	March - M031990	Andhra Pradesh	Hyderabad	NA	Residential, Rural and other Areas	6.3	14.7	NA	NA	NA	NA	3/1/1990
151	March - M031990	Andhra Pradesh	Hyderabad	NA	Industrial Area	4.7	7.5	NA	NA	NA	NA	3/1/1990
152	March - M031990	Andhra Pradesh	Hyderabad	NA	Residential, Rural and other Areas	6.4	25.7	NA	NA	NA	NA	3/1/1990
150	April - M041990	Andhra Pradesh	Hyderabad	NA	Residential, Rural and other Areas	5.4	17.1	NA	NA	NA	NA	4/1/1990
151	April - M041990	Andhra Pradesh	Hyderabad	NA	Industrial Area	4.7	8.7	NA	NA	NA	NA	4/1/1990

Fig.2

Further, after prediction our model can be used on any trained data and images to know the AQI and the status of its pollution. Due to being based on small training datasets our model first needs to train the images and datasets for prediction.

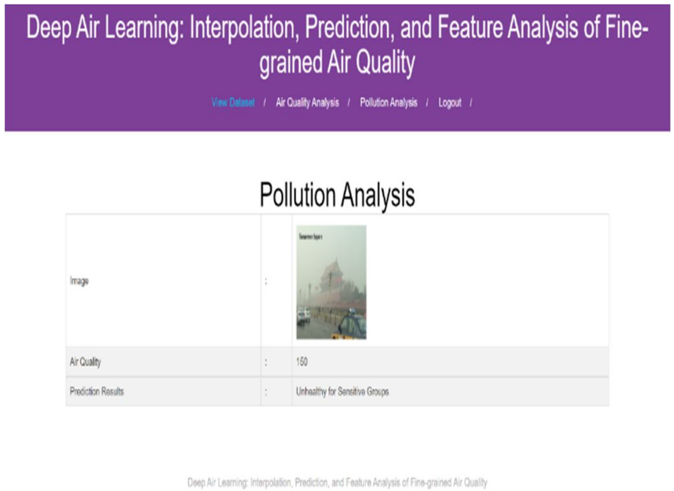


Fig.3

Later, using the basemap matplot library and using K-Mean clustering, our model shows the graphical representation between the actual vs predicted AQI which is shown below Fig.3. We can analyse different parameters of the air by comparing by using graphs, several graphs will be obtained. All the scatter plots from the graphs are combined which will give comparative results between parameters.

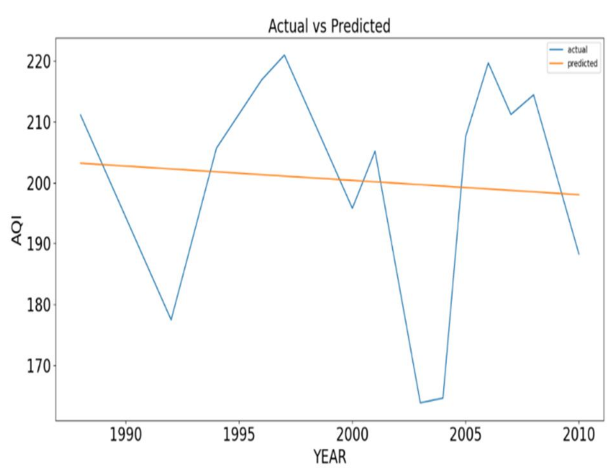


Fig.4

V. CONCLUSION

Our proposed model studies the three important topics in the area of urban air computing for prediction: the interpolation, prediction, and feature analysis of fine-gained air quality. The solutions to these three topics which are interpolation, feature analysis, spatiotemporal provides us the crucial information to support air pollution control in different cities, and consequently generating the great societal and technical impacts on the cities. We have inferred the fine-grained air quality information throughout a city of Andhra Pradesh which include different industrial, rural and urban areas using historical air quality data reported by a limited number of existing monitor stations and a variety of data sets which was observed in the city. Extensive experiments are conducted on real data sources collected from different cities of Andhra Pradesh which include urban, rural and industrial areas. These dataset is between years 1990 to 2005. There are some limitations of this model like a large amount of training samples has to be studied for accurate prediction, lack of the data and historical information, obtaining higher accuracy is challenging, prediction and feature selection may vary from place to place due to change in air humidity and climatic changes. Future more modification in the model can work on real time prediction and more accurate predictions can be made within less time.

VI. ACKNOWLEDGEMENT

This work was mainly inspired from the different researches done on air quality prediction. The proposed model is deployed to predict the air quality index of the data obtained from the Central Control room for Air Quality Management. This work is performed at the different cities of Andhra Pradesh in collaboration with Dr. Atul Kumar as our project guide at SRMGPC,LUCKNOW.

REFERENCES

[1] Ditsuhi Iskandaryan, Francisco Ramos and Sergio Trilles,"Air Quality Prediction in Smart Cities Using Machine Learning Technologies Based on Sensor Data:A Review",Institute of New Imaging Technologies (INIT), Universitat Jaume I, Av. Vicente Sos Baynat s/n,12071 Castelló de la Plana, Spain;

[2] Giffinger, R.; Fertner, C.; Kramar, H.; Kalasek, R.; Pichler-Milanovi´c, N.; Meijers, E. Shrewd urban communities: Ranking of european medium-sized urban communities. Vienna, austria: Center of territorial science (srf), Vienna college of innovation. Accessible (got to on 31 March 2020) .

[3] Wan, J.; Li, D.; Zou, C.; Zhou, K. M2M correspondences for savvy city: An occasion based design. In Proceedings of the 2012 IEEE twelfth International Conference on Computer and Information Technology, Chengdu, China, 27–29 October 2012; pp. 895–900.

[4] Trilles, S.; Calia, A.; Belmonte, Ó.; Torres-Sospedra, J.; Montoliu, R.; Huerta, J. Organization of an open sensorized stage in a shrewd city setting. Future Gener. Comput. Syst. 2017, 76, 221–233. [CrossRef]

[5] Miss. Vadaka Keerthi, Mrs. Kavitha Juliet,"An Effective Method to Predict Air Pollutants utilizing Random Forest Algorithm", International Journal for Research in Applied Science and Engineering Technology (IJRASET) ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177 Volume 7 Issue VI, June 2019

[6] M. Yuan and Y. Lin, "Model choice and assessment in relapse with gathered factors," Journal of the Royal Statistical Society: Series B (Statistical Methodology), vol. 68, pp. 49–67, 2006.

[7] L. Li, X. Zhang, J. Holt, J. Tian, and R. Piltner, "Spatiotemporal interjection techniques for air contamination openness," in Symposium on Abstraction, Reformulation, and Approximation, 2011.

[8] Y. Zheng, F. Liu, and H.- P. Hsieh, "U-air: When metropolitan air quality induction meets large information," in Proceedings of the nineteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ser. KDD '13, 2013, pp. 1436–1444..

[9] NandigalaVenkatAnurag, YagnavalkBurra, S.Sharanya,"Air Quality Index Prediction with Meteorological Data Using Feature Based Weighted Xgboost",International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8, Issue-1, May 2019.

[10] JaakkoKukkonen, LeivHårvardSlordal, RanjeetSokhi, " Analysis and assessment of European air contamination scenes", Meteorology applied to metropolitan air contamination issues, ISBN 954-9526-30-5, Demetra Ltd Publishers, Bulgaria, pp. 99-114, 2005.

[11] Perez, P., Trier, A., and Reyes, " Prediction of PM2.5 fixations a few hours ahead of time utilizing neural organizations in Santiago, Chile", Atmospheric Environment, 34:1189–1196, 2005.

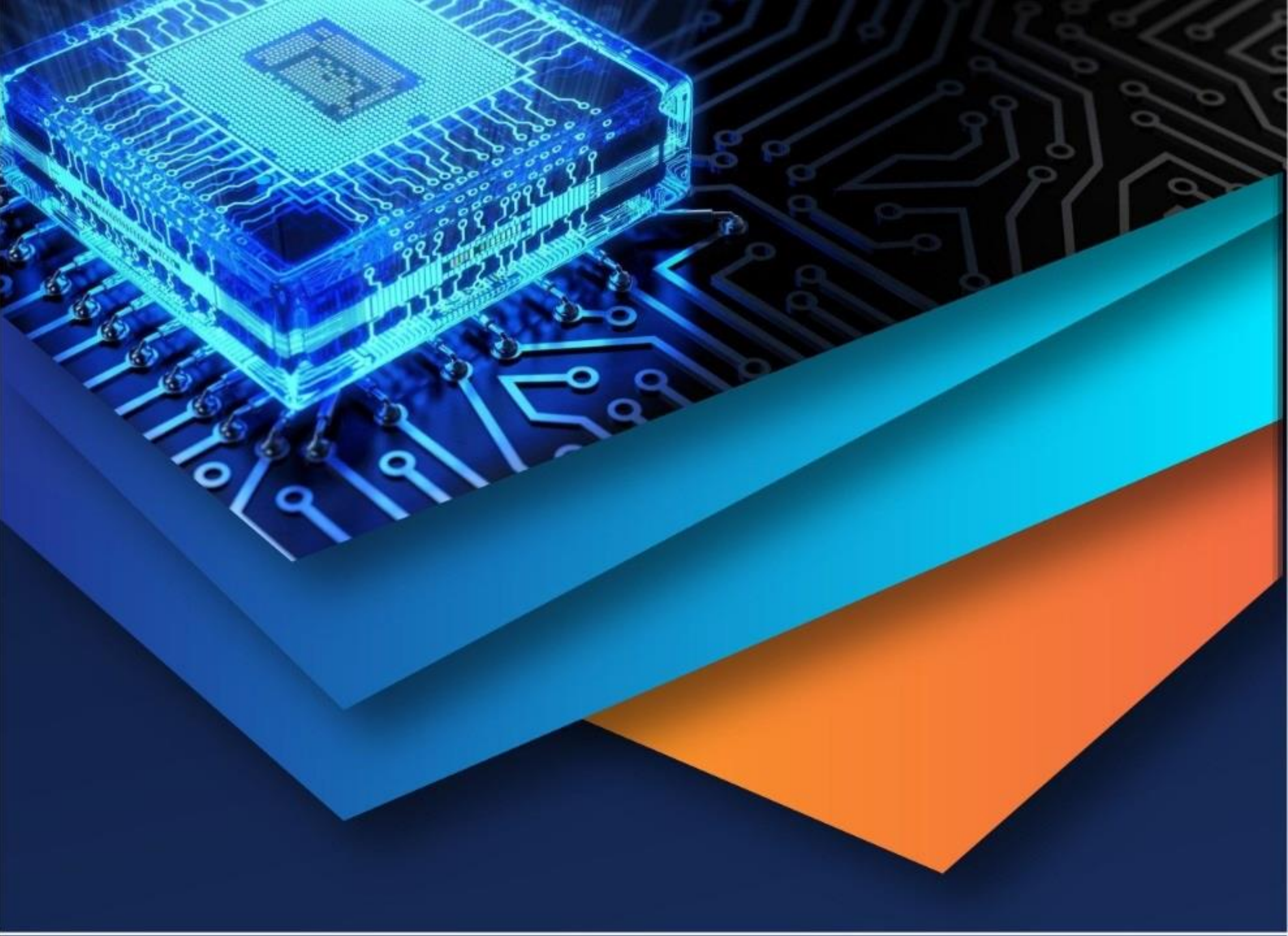
[12] Gardner, Dorling and S.R. ,"Neural organization displaying and forecast of hourly NOXand NO2 focuses on metropolitan air in London", Atmospheric Environment 33(5), 709-719, 1999.

[13] Yu Zheng, Furui Liu, Hsun-Ping Hsieh."U-Air: When Urban Air Quality Inference Meets Big Data".

[14] A. V. Donkelaar, R. V. Martin, and R. J. Park (2006), Estimating ground-level PM2.5 utilizing vaporized optical profundity decided from satellite distant detecting, J. Geophys. Res., 111, D21201.

[15] D. Hasenfratz, O. Saukh, S. Sturzenegger, and L. Thie le. Participatory Air Pollution Monitoring Using Smartphones. In the second International Workshop on Mobile Sensing.

[16] Y. Jiang, K. Li, L. Tian, R. Piedrahita, X. Yun, O. Mansata, Q. Lv, R. P. Dick, M. Hannigan, and L. Shang. Maqs: A customized versatile detecting framework for indoor air quality. In Proc. of UbiComp 2011.



10.22214/IJRASET

INDEX COPERNICUS
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)