IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, August 2023

Air Quality Index (AQI) Classification using Sensor Data with Neural Network

Dr. Murali G[1], Mr. Chandan M C[2], Ms. Prabodhana K[3], Ms. C B Manogna Reddy[4], Mr. Abhinav Pavan[5]
Professor & Head, Department of Computer Science & Engineering[1]
Students, Department of Computer Science & Engineering[2,3,4,5]
R. L. Jalappa Institute of Technology, Doddaballapura, Karnataka, India

Abstract: *At present the interpolation, prediction, and feature analysis of fine-gained air quality are three important topics in the area of urban air computing. The solutions to these topics can provide extremely useful information to support air pollution control, and consequently generate great societal and technical impacts. Most of the existing work solves the three problems separately by different models. In this paper, we propose a general and effective approach to solve the three problems in one model called the Deep Air Learning (DAL). The main idea of DAL lies in embedding feature selection and semi-supervised learning in different layers of the deep learning network.*

The proposed approach utilizes the information pertaining to the unlabeled spatio- temporal data to improve the performance of the interpolation and the prediction, and performs feature selection and association analysis to reveal the main relevant features to the variation of the air quality. We evaluate our approach with extensive experiments based on real data sources obtained in Beijing, China. Experiments show that DAL is superior to the peer models from the recent literature when solving the topics of interpolation, prediction, and feature analysis of fine-gained air quality

Keywords: Machine Learning (ML), AQI, ANN

I. INTRODUCTION

Machine learning is to predict the future from past data. Computer studying (ML) is a style of artificial intelligence (AI) that delivers computers the capability to gain knowledge of without being explicitly programmed. Machine finding out makes a specialty of the progress of pc applications that can alternate when exposed to new information and the basics of laptop studying, implementation of an easy laptop finding out algorithm utilizing python. Process of coaching and prediction involves use of specialized algorithms. It feed the training data to an algorithm, and the algorithm uses this training knowledge to offer predictions on a brand-new test information. Machine finding out can be roughly separated in to three classes. There is supervised learning, unsupervised finding out and reinforcement finding out. Supervised studying software is each given the input knowledge and the corresponding labelling to be trained data must be labelled with the aid of a person previously. Unsupervised learning isn't any labels.

It provided to the learning algorithm. This algorithm has to figure out the clustering of the input knowledge. Subsequently, Reinforcement learning dynamically interacts with its environment andit receives positive or bad suggestions to toughen its efficiency. Data scientists use many one-of- a-kind types of computing device learning algorithms to observe patterns in python that lead to actionable insights. At a high stage, these specific algorithms can also be labelled into two companies situated on the way they "gain knowledge of" about data to make predictions: supervised and unsupervised learning. Classification is the method of guess ing the class of given information points. Lessons are in many instances referred to as goals/ labels or classes. Classification predictive modelling is the task of approximating a mapping function from enters variables(X) to discrete output variables(y). In computer studying and facts, classification is a supervised studying technique in which the pc software learns from the information input given to it after which makes use of this studying to classify new statement. This data set could without problems be bi-classification or it may be multi classification too. Some examples of classification problems

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, August 2023

are: speech consciousness, handwriting awareness, bio metric identification, file classification and so forth Air pollution is a significant environmental issue that affects the health and well-being of individuals and communities worldwide. The rapid industrialization and urbanization in recent years have led to a substantial increase in air pollution levels, resulting in adverse effects on humanhealth and the environment. To mitigate the impact of air pollution, it is crucial to monitor and assess air quality accurately.

The Air Quality Index (AQI) is a standardized measure used by environmental agencies to quantifyand communicate the level of air pollution in a specific area. The AQI provides information aboutthe concentrationof common air pollutants, such as particulate matter (PM2.5 and PM10), nitrogendioxide (NO2), ozone (O3), carbon monoxide (CO), and sulphur dioxide (SO2). By classifying theair quality into different categories, the AQI helps individuals and authorities make informed decisions regarding outdoor activities and public health measures.

In recent years, advancements in sensor technology and the Internet of Things (IoT) have enabled the development of low-cost, portable air quality sensors that can bedeployed across different locations providing a more comprehensive and dynamic view of air pollution.

In this context, the use of neural networks for air quality classification based on sensor data has gained significant attention. Neural networks, a subfield of machine learning, have shown remarkable capabilities in modelling complex relationships between input data and output categories. By training a neural network on historical sensor data and their corresponding AQI values, it is possible to develop a model that can accurately predict the air quality class given the sensor readings.

The objective of this study is to explore the application of neural networks for AQI classification using sensor data. By leveraging the power of neural networks, we aim to develop a model that caneffectively classify air quality based on sensor readings, providing a more efficient and cost- effective approach to air pollution monitoring. The developed model can be deployed in various settings, including smart cities, industrial areas, and personal air quality monitoring devices, to provide real-time information about air pollution levels and enable timely interventions to protect public health.

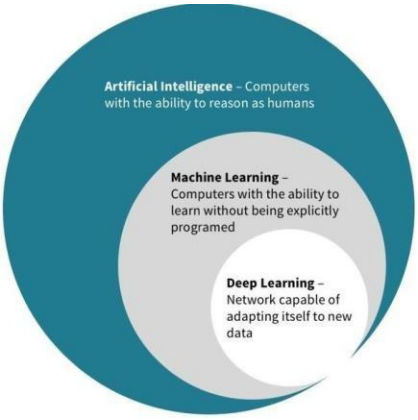

Figure 1: AL, ML and Deep learning Architecture

Machine Learning involves the examination of computer algorithms that can enhance their performance automatically based on experience and the utilization of data. Machine learning, a subfield of artificial intelligence, refers to the process of training computer systems to learn from data and improve their performance without explicit programming. Machine learning algorithms are designed to analyse data, recognize patterns, and make predictions or decisions based on the information available.

II. METHODOLOGIES

Artificial Neural Networks (ANN) are a type of machine learning model that is inspired by the structure and function of the human brain. ANNs are composed of interconnected nodes or artificial neurons, which process and transmit information through weighted connections. ANNs are capable of learning from data and can be trained to perform a variety of tasks, such as classification, regression, and pattern recognition. The learning process involves adjusting the weights of the

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, August 2023

connections between neurons in response to the inputdata and desired output.

There are different types of ANN architectures, such as feed forward networks, recurrent networks, and convolutional networks, each suited to different types of tasks. ANN architectures are often organized in layers, with input layers, hidden layers, and output layers. ANNs have been used in a wide range of applications, including image and speech recognition, natural language processing, autonomous vehicles, and finance. They have also been used to create advanced AI systems, such as deep learning models.

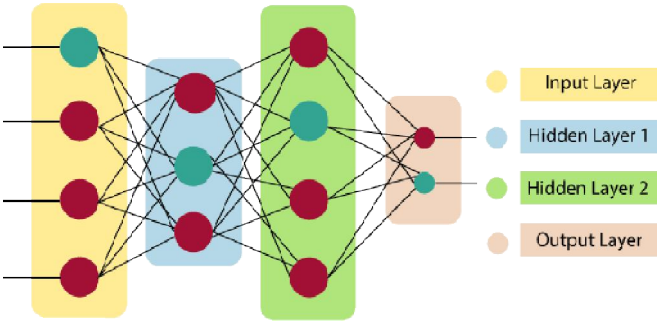

Figure 2: Architecture of Arificial Neural Network

Evaluation MatricsConfusion Matrix:

A confusion matrix, sometimes known only as an error matrix, is a specific type of table structure that is used to analyse the performance of algorithms. Such a technique is frequently used it for supervised methods

Confusion Matrix

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positives (TPs)	False Positives (FPs)
Predicted Negative (0)	False Negatives (FNs)	True Negatives (TNs)

$$\text{True Positive Rate} = \frac{TP}{TP + FN}$$

$$\text{True Negative Rate} = \frac{TN}{TN + FP}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

III. SYSTEM DESIGN

A system architecture diagram would be used to show the relationship between different components. Usually they are created for systems which include hardware and software and these are represented in the diagram to show the interactionbetween them.

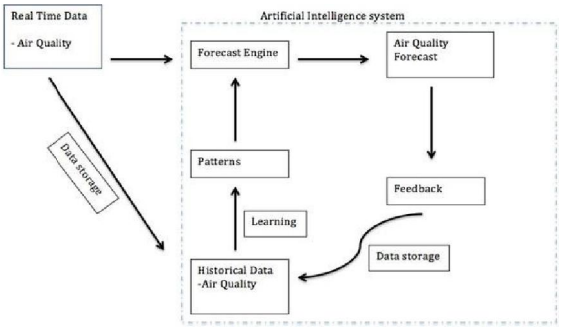

Figure 3: System Architecture

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, August 2023

Flow diagram

It is important to complete all tasks and meet deadlines. There are many project management tools that are available to help project managers manage their tasks and schedule and one of them is the flowchart.

A flowchart is one of the seven basic quality tools used in project management and it displays the actions that are necessary to meet the goals of a particular task in the most practical sequence.

Also called as process maps, this type of tool displays a series of steps with branching possibilities that depict one or more inputs and transforms them to outputs.

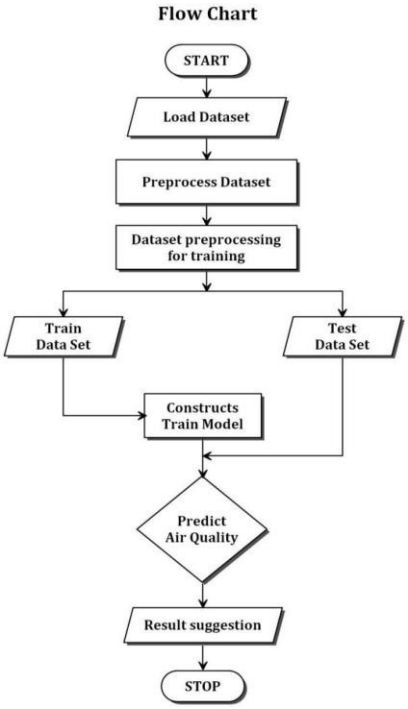

Figure 4: Flow Diagram

IV. RESULTS

TABLE: Metrics of air quality levels

Metrics	Definition
Precision	Precision is defined as the ratio of positive examples to the sum of such actual andfalse positives.
Recall	Recall is defined as the ratio of correct positives to all true negatives and false negatives.
F1 Score	A weighted harmonic average of such recall and precision is known as the F1. The projected capacity for the model is higher the closer the F1 score value is near 1.0.
Support	An amount of instances of a class that truly exist in the dataset constitutes the number of supports. It does not differentiatebetween kinds; it only improves theperformancee valuation process.

The above table shows the precision, recall, and f1- score for six categories of air quality levels: Good, Moderate, Poor, Satisfactory, Severe, and Very Poor The accuracy of the model is 0.99, indicating that the model performs well in predicting the correct air quality level. The f1-score for each category ranges from 0.97 to 1.00, indicating a high level of precision and recall for each category.The macro average and weighted average of precision, recall, and f1-score are all around 0.99, indicating that the model has a high overall performance. The answer is accurate and precise in summarizing the key metrics for evaluating the performance of the air quality monitoring model

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, August 2023

	precision	recall	f1-score	support
Good	0.99	1.00	0.99	1184
Moderate	0.98	0.98	0.98	1215
Poor	1.00	1.00	1.00	1164
Satisfactory	0.97	0.98	0.97	1155
Severe	1.00	1.00	1.00	1206
Very Poor	0.99	1.00	0.99	1206
accuracy			0.99	7130
macro avg	0.99	0.99	0.99	7130
weighted avg	0.99	0.99	0.99	7130

Figure 5: Classification Report for Random Forest

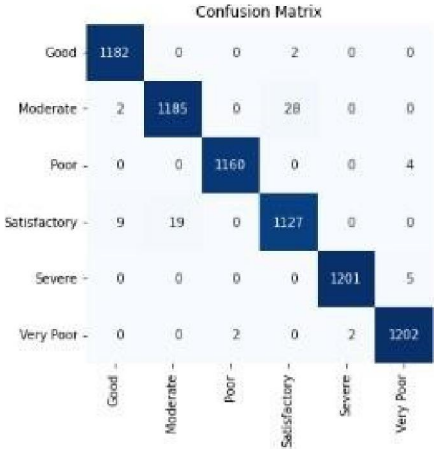

Figure 6: Confusion Matrix for Random Forest

IV. CONCLUSION

Air quality monitoring is a crucial process that involves the collection and analysis of data on the level of pollutants present in the atmosphere. It is essential for governments, organizations, and individuals to monitor air quality to protect human health, the environment, and ecosystems.

The process of air quality monitoring has advanced significantly in recent years with the development of new technologies and sensors that can detect even the smallest pollutants. Real- time air quality monitoring systems have also been developed, which enable continuous monitoring of air quality, allowing for immediate response in case of a pollution event.

Despite the advancements in air quality monitoring, challenges still exist, such as the need for greater standardization in monitoring methods and the lack of access to monitoring equipment in some regions. Nonetheless, air quality monitoring remains a critical tool in the fight against air pollution, and continued efforts are necessary to improve the accuracy and accessibility of air quality data.

REFERENC ES

[1] Jingyang Wang, Xiaolei Li, Lukai Jin, Jiazheng Li, Qiuhong Sun1 & HaiyaoWang "An air quality index prediction model based on CNN-ILSTM" Scientifc Reports (2022) 12:8373 https://doi.org/10.1038/s41598-022-12355-6.

[2] WHO. (2019) 7 Million Premature Deaths Annually Linked to Air Pollution. World Health Organization.

[3] Kiruthika, R., & Umamakeswari, A. (2018). Low cost pollution control and air quality monitoring system using Raspberry Pi for Internet of Things. 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing, ICECDS 2017, 2319–2326. https://doi.org/10.1109/ICECDS.20178389867.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, August 2023

[4] Hsieh, H. P., Lin, S. De, & Zheng, Y. (2015). Inferring air quality for station location recommendation based on urban big data. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2015-August, 437–446. https://doi.org/10.1145/2783258.2783344

[5] Malalgoda, C., Amaratunga, D., & Haigh, R. (2015). Local governments and disaster risk reduction: a conceptualframework. 6th International Conference on Building Resilience, 699–709. http://eprints.hud.ac.uk/id/eprint/30327/