ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 03 (11), November 2024, pp. 33-49

Article Title: A Data-Driven Approach to Air Pollution Management with Self-Tuning Deep Regression Neural Network

A Data-Driven Approach to Air Pollution Management with Self-Tuning Deep Regression Neural Network

S. Margret Beaula[1,*]

[1]Research Scholar, Department of Electronics and Communication Engineering, St. Xavier's Catholic College of Engineering, Nagercoil, Tamil Nadu. margretbeaulas@gmail.com

ABSTRACT

Air Qualities (AQ) are developed to be a serious environmental and health concern as a result of worldwide population growth. A predictive model for air quality is essential for the timely prevention and management of air pollution. This study proposes a novel Self-Tuning Deep Regression Neural Network (STDRNN) for identifying air pollution in the environment using deep neural network. The Air Quality Index (AQI) dataset is employed as the primary source of input data. The methodology begins with a data pre-processing phase where the input data finds the missing data and shows the duplicate AQ values using data imputation and remove outlier method. This is followed by feature construction where the each AQ label is converted into an integer AQ value and scaling the features to a similar range using feature encoding and feature scaling method. The AQI data are then given to the STDRNN model for training and testing, enabling accurate prediction. The performance of the proposed model is evaluated using standard metrics such as RMSE, MSE and R2 coefficient. Using the python software AQI dataset shows the comparative analysis with existing methods on similar datasets demonstrates the superiority of the proposed approach in terms of selection performance. This work highlights the potential of STDRNN in advancing the accuracy of the model as well as its promising potential applications.

Keywords: Air Quality Index (AQI), data imputation, feature encoding, feature scaling, Self-Tuning Deep Regression Neural Network (STDRNN)

1 Introduction

According to estimates from the World Health Organization (WHO), air quality problems cause over 7 million deaths worldwide each year [1]. Because of the abundance of biological components, toxic substances, and particulate matter in the Earth's atmosphere, air pollution is a complicated issue. The effects of this pollution are extensive, affecting not only food crops but also humans, animals, and other living forms. [2]. The harmful consequences of fine Particulate Matter (PM) are extensive and range from allergies to serious illnesses like lung cancer, strokes, and cardiovascular and respiratory disorders even death occurs. Global attention takes slowly towards the threat that environmental pollution poses to human and ecological health [3]. The system gauges factors such as humidity, elevation, atmospheric pressure, and the levels of hazardous chemicals, including NO2 and CO2, present in the atmosphere [4]. Other experiments have analyzed air pollution using a technique known as

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 03 (11), November 2024, pp. 33-49

Hyper Spectral Imaging (HSI) technology, in which data is gathered from the aerial camera's visible-light HSI technology. The drone camera's captured image contains hyperspectral information [5, 6]. Thereby, in order to find the air quality the below Machine Learning (ML) method are used.

Machine learning technique called Artificial Neural Networks (ANN) and gradient boost used to collect real-time traffic on a moving camera. Models developed for fine particulate matter performed better than those for black carbon [7]. Continuously, ML method called Extreme Gradient Boosting (XGBoost), CatBoost, Light Gradient-Boosting Machine (LightBGM), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU) to forecast the PM_{10} levels in Colombia [8]. AQI takes 171 features ranging from environmental, demographical, economic, meteorological, and energy, were collected and analysed using XGBoost with Bayesian Optimization [9]. A random forest method is to remove meteorological effect for air quality data [10].

Since, air quality predicting on machine learning models have advanced quickly and are now frequently utilized on the Gaussian Naive Bayes [11]. It uses the Gravity Search Algorithm (GSA) to optimize the unknown parameters of the Least Squares Support Vector Regression (LSSVR) in order to construct the corresponding hybrid model, LSSVRGSA, which effectively combines machine learning (ML) with accuracy to improve the model's performance [12]. A Convolutional Neural Network-Long Short-Term Memory (CNN-LSTM) is a hybrid deep neural network that estimates air pollution concentrations in various urban locations by utilizing spatiotemporal correlations. The model's dependence on data, sensitivity to data quality, and extensive hyperparameter optimization present serious difficulties despite its complex architecture [13, 14]. Whereas, Artificial Neural Networks (ANNs), RF, SVM, Adaptive Boosting (AdaBoost), and stacking ensemble all demonstrated positive outcomes in the AQI forecast; stacking ensemble produced results for R2 and RMSE, while AdaBoost shown performance for MAE [15]. Consequently, the developed self-tuning deep regression neural network is used as the model selection (regression) approach in this work. The hybrid algorithm aims to find the air quality that enhances the network performance of the deep learning-based neural network. The deep neural networks have higher proficiency in handling large data and they are more adept in model selection than ML techniques.

The contributions of the work are as follows:

- Handles missing data effectively through imputation and ensures data consistency by removing outliers.
- Transforms raw data into a machine-learning-ready format using feature encoding and feature scaling.
- Normalizes features to ensure balanced contributions during model training.
- Systematically splits the dataset into training (80%) and testing (20%) subsets for proper evaluation and reduced overfitting.

- Employs a STDRNN optimized for regression tasks. Also, integrates self-tuning mechanisms within STDRNN for automated hyperparameter optimization.
- Provides precise and robust predictions for AQI related metrics.

2 Related Work

Jang *et al* [16] (2020) have invented a machine learning model to find surface topography, the chemical origins of tiny dust, traffic, and human activity in smaller regions. The topological feature on the surface is captured using linear regression, Autoregressive Integrated Moving Average (ARIMA) and Gaussian process regression. Where Gaussian process regression maps the input space using the input features with a modified kernel design and the PM is predicted using the radial basis kernel function, which captures the spatial pattern. The drawback is Gaussian process regression improves the performance and linear regression is low.

Harishkumar *et al* [17] (2023) have proposed a Neural Basis Expansion Analysis for Time Series (N-BEATS) neural network-based architecture to calculate the error rates and tune hyper parameters. A data pre-processing technique called resampling is used to find the missing values in the air quality data. It is categorized using the N-BEATS neural network-based architecture, which consists of a deep stack of fully connected layers with backward and forward residual linkages. Because of the effectiveness this method it fails to achieve the necessary level on hyper parameters.

Mogollón-Sotelo *et al* [18] (2021) have proposed a Support Vector Machine (SVM) to forecast ground-level PM2.5 in populated city with complex topography. Data cleaning is done in the data pre-processing stage and using structural risk minimization the boundaries are captured. Using the kernel SVM AQ data is to hyperplane the maximum margin is forecasted. It is accurate and active in high-dimensional spaces .The drawback is the minimum margin with complex topography is not captured.

Liu *et al* [19] (2024) have developed a Light Gradient Boosting Machine (LGBM) classifier with Long Short Term Memory (LSTM) model to find the errors. Air quality index problems of gradient explosion and disappearance during model training are resolved by using the classification method known as LGBM with LSTM to alter the output gate. However, for AQI, regular LGBM with LSTM proved slightly less in RMSE.

Maltare *et al* [20] (2023) have proposed machine learning technique called SVM, LSTM and SARIMA (Seasonal Auto-Regressive Integrated Moving Average) to find the auto regressive values and moving average values. Whereas, SARIMA, SVM through various kernels in AQI, and LSTM by hyper parameter are estimated to as the most effective models for AQI prediction. With the aid of performance metrics like R2 score and RMSE, the trained value using SVM by Radial Basis Function (RBF) kernel yields larger outcomes for AQI prediction. Due to these approaches inability to reach the required RMSE level for AQI.

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 03 (11), November 2024, pp. 33-49

3 Proposed Method

STDRNN is a new DL-based model for environmental analysis that uses progressive learning to efficiently predict from AQI dataset for air quality. Here, the data pre-processing technique uses data imputation and remove outliers to find the missing AQ values and already used AQ values. Subsequently, the feature construction uses the feature encoding and feature scaling technique their each AQ label is converted into an integer AQ value and scaling the features to a similar range as (0, 1) which is a minimax scalar, Following which the Self-Tuning data efficient deep learning technique and the predicted data are given to the hidden layers were the neural networks using the deep regression neural network the output are predicted.

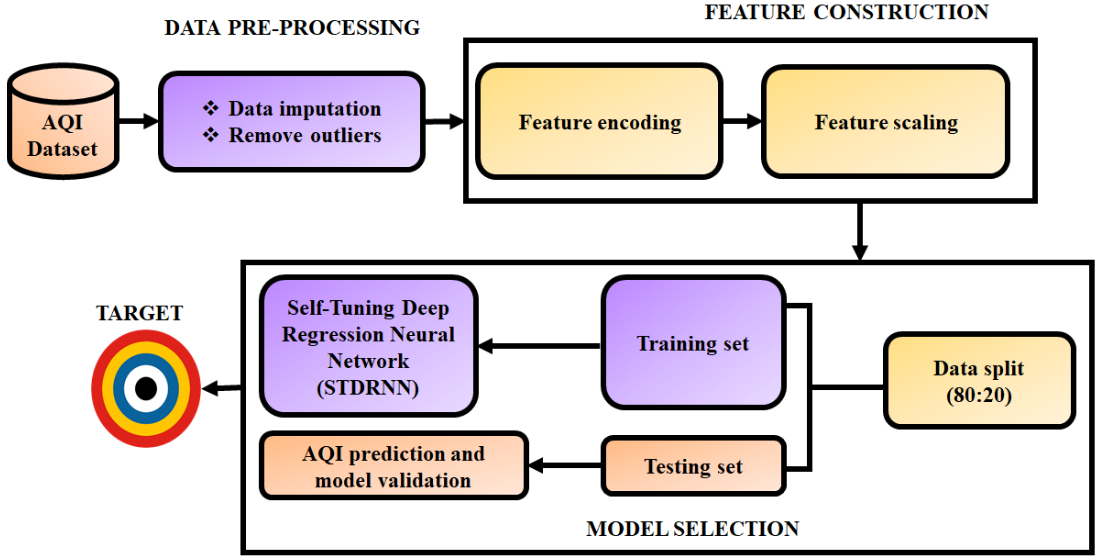

Figure 1: *Block diagram of proposed air quality*

3. 1 Data Pre-processing for AQ

The first step in the data pre-processing for air quality is data imputation and removes outliers.

3. 1. 1 Data imputation for AQ

Imputation is crucial for addressing the challenge of missing data from AQ. This technique is advantageous as it creates a complete dataset by substituting missing values with valued ones. In the single imputation AQ approach, the air quality estimates value as the true value. This method produces a single replacement AQ value for each data point that is missing from AQ.

3. 1. 2 Remove outliers for AQ

Outliers show the duplicate AQ values or already used AQ values. Outliers are individual AQ values that fall outside of the overall pattern from the rest of the original AQ data. For every AQ data, the appearance of outliers cannot be ignored. An essential component of data

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 03 (11), November 2024, pp. 33-49

cleaning is outlier detection, which is especially pertinent for inexpensive air quality networks. AQ outlier detection is the process of identifying AQ values that deviate statistically from the average AQ value at a given time and location. Followed by the data pre-processing input AQI data is given to the feature construction.

3. 2 Feature Construction for AQ

Feature construction is the second stage here the pre-processed AQI data is used as the input to feature the AQI data using feature encoding and feature scaling.

3. 2. 1 Feature encoding For AQ

In Label encoding, each AQ label is converted into an integer AQ value. In supervised machine learning, a set of AQ data is labeled as the output before being sent into the prediction algorithm. In detail, the air quality methods to label the quality of the air in a particular environment. Typically, the AQ is divided into a few ranges, and each range is given as a description and an integer.

3. 2. 2 Feature scaling for AQ

Feature scaling involves scaling the features to a similar range, typically between 0 and 1, so that the model is more effective.

Mini-max scalar for AQ

Mini-max scalar value from various ranges to [0, 1], in AQ, not every dataset requires normalization. Removing the effect of AQ data with dissimilar orders of magnitude and circumventing large values dominance over little values is an issue in AQ data. Additionally, normalization boosts convergence efficiency and solution speed. Equation (1) provides the mini-max formula, wherex_{max} and x_{min} stand for the minimum and maximum values in the AQ dataset, respectively, and x and x_i' for the original and normalized values.

$$x_i' = \frac{x_i - x_{min}}{x_{max} - x_{min}} \qquad (1)$$

Followed by the feature construction AQI data is given to the data splitting.

3. 3 Data Splitting for AQ

Data splitting involves dividing the AQ dataset into training and testing sets. The ML algorithm is trained on the training set, and the model's performance is assessed on the testing set. However, the AQ data is splitting as 80 trained and 20 testing samples. When the testing subset is altered, the AQ dataset loses its validity. In order to get around this issue, every model is constructed 20 periods using various random subsets of training and testing data. The 80:20 splitting proportion is unaffected. Followed by the data splitting AQI data is given to the proposed STDRNN.

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 03 (11), November 2024, pp. 33-49

3. 4 Model Selection (Self tuning deep regression network) for AQ

In this proposed work the proposed Deep Learning method called Self-Tuning Deep Regression Neural Network is utilized for effective neural network. A Self-Tuning data Efficient Deep Learning is trained and labelled using the air quality data.

3. 4. 1 Self-Tuning Data Efficient Deep Learning for AQ

In this method, a pre-trained AQ model$M$, a labelled AQ dataset $L = (\{(x_i^L, y_i^L)\})_{i=1}^{nL}$ and an unlabelled AQ data set $U = \{(x_i^U)\}\}_{i=1}^{nU}$in the object area are set. Initial as a deep network M is collected by a pre-trained backbone $f_0$for feature removal and a pre-trained head g_0 , whereas fine-tuned ones are represented by f and g individually. f Is generally modified as f_0 although $g$is casually adjusted, then the target AQ dataset frequently.

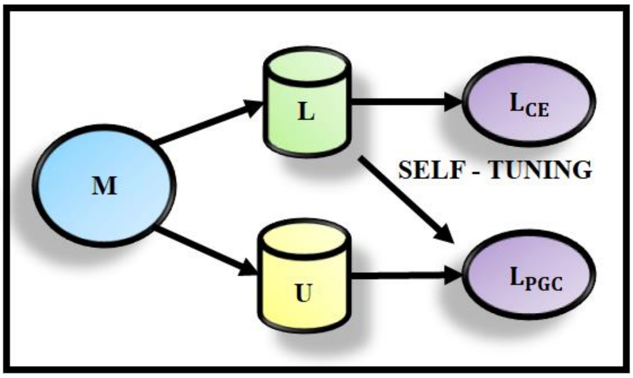

Figure 2: *Self-tuning for AQ*

First, an AQ data is modified after an accurate pre-trained AQ model, When it comes to implicit regularization, self-tuning is a well beginning point than a model that is trained from scratch on the target AQ dataset. Additionally, the pre-trained AQ model's knowledge moves into the labeled and unlabeled AQ data in parallel, unlike the progressive form that over fits the restricted AQ labelled data. However, gradients from the labelled AQ data $L$and the unlabelled AQ data Uwill simultaneously change the model's parameters. The air quality is tuned by simultaneously examining the intrinsic structure of $U$and the AQ label information of Lin a single, cohesive form.

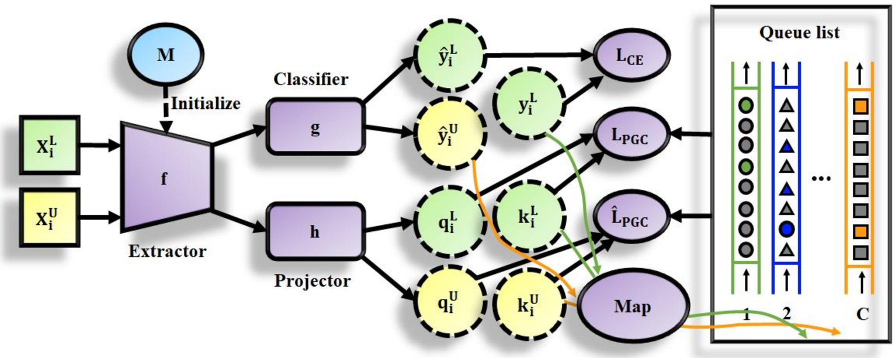

Figure 3: *shared queue list for AQ*

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 03 (11), November 2024, pp. 33-49

The ground-truth AQ labels of a shared queue list between L and U are easily accessible, given a labelled AQ data set L from C categories. The encoded query $q_i^L = h(f(aug_1(X_i^L)))$ & encoded $K_i^L = h(f(aug_2(X_i^L)))$ keys are created similar for an AQ data sample (X_i^L, y_i^L) in$L$. Also it eliminates the subscript and superscript$L$ for clarity and concentrates on a specific AQ data example(x, y). The labelled AQ data is used to forecast the ground-truth versions of PGC. The formal summary of PGC loss for every AQ data point (x, y)on$L$ is follows as:

$$L_{PGC} = \frac{1}{D+1}\sum_{d=0}^{D} \log[\exp(q.K_d^y/T)]/[Pos + Neg] \quad (2)$$

3. 4. 2 Deep regression neural network for AQ

To model all pollutant in AQ, multi-layer perceptron's with the same architecture were employed. The prediction error in AQ is estimated by comparing the observed and anticipated AQ concentrations. In order to evaluate the prediction error for AQ across the entire concentration range, the estimate errors for different sub-models were computed. A comparison is made between this error and the approximation error derived from a single model that encompasses the complete attention series. A regression modeling is conducted using multi-layer perceptron's, which are artificial neural networks for AQ. In AQ, each perceptron is made up of one output neuron and ten input neurons in one hidden layer. The AQ network repeatedly runs the cycles during the learning process to reduce the learning error in AQ. The function of activating hidden and output neurons in a logistic function measured on AQ. The scales were initiated randomly. The AQ network initialization is random (Gaussian). AQ is a logistic function that measures how hidden and output neurons are activated. The initialization of the AQ network is random (Gaussian).

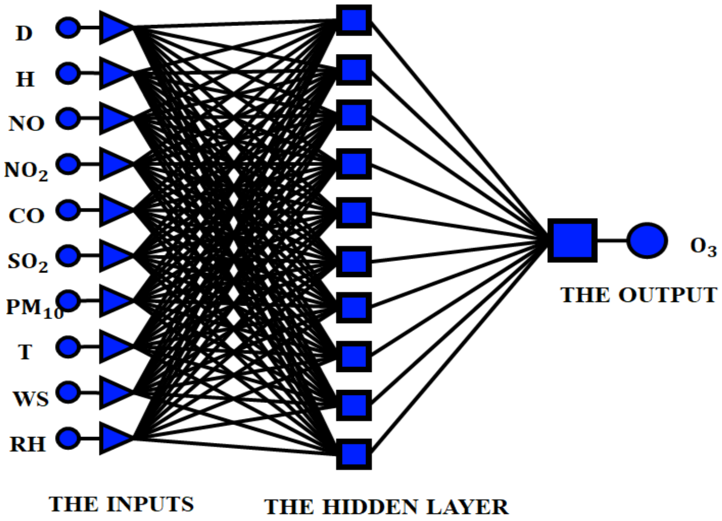

Figure 4: *Hidden layers of AQ*

Each AQ prediction implemented five times. AQ neural network training is repeated, but it is standard training to keep the learning process's parameters constant. The learning process for

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 03 (11), November 2024, pp. 33-49

each AQ model is a little random and results in a distinct network. Although the neurons in the generated AQ networks share the same structure, their weights and levels of individual neuron activation vary. In general, the modelling error for AQ varies slightly. Out of the five models developed, the most accurate one is chosen for reporting. They rejected the other AQ models. It is assumed that the miscalculation task is the Sum Of Squares (SOS). The sum of the squared deviations between all expected and real AQ values is known as SOS.

Performance Criteria for AQ

The model performance is measured using a statistical evaluation for AQ, including coefficient of determination (R2), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE).The AQ criterion formulas are shown below.

$$MSE = \frac{\sum_{i=1}^{m}[x_i - \widehat{x_i}]}{m} \tag{3}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{m}[x_i - \widehat{x_i}]2}{m}} \tag{4}$$

$$R^2 = [\frac{1}{M}\frac{\sum_{j=1}^{M}[(Y_j - \bar{Y})(X_j - \overline{X})]}{\sigma_y \sigma_x}]^2 \tag{5}$$

4 Result and discussion

All of the data were divided into two sets: 20% were used for testing, and the remaining 80% were used for training. AQI datasets were used in the models' execution. Moreover, the model's computation time for future prediction algorithm implementation in performance metrics is another crucial factor. Furthermore, testing in Python applications helps to understand the regression model's capabilities.

Distribution of AQI

Figure 5 displays the frequency distribution histograms of the air quality index concentrations. In order to minimize the work involved in creating several models for each pollutant, as well as in maintaining and monitoring the models and building a deployment pipeline, the major objective of creating a global model is to predict multiple pollutants simultaneously.

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 03 (11), November 2024, pp. 33-49

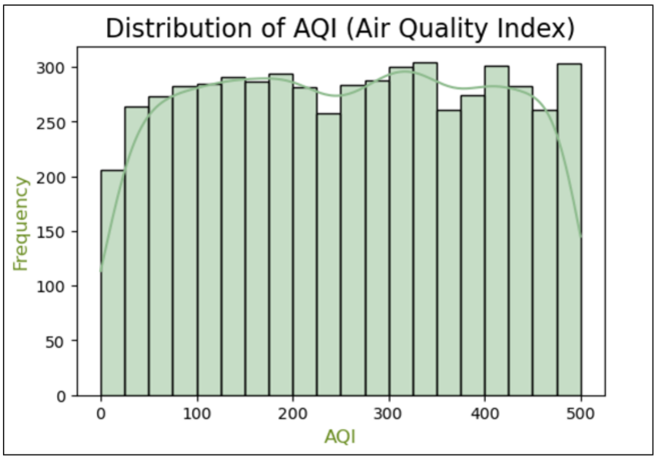

Figure 5: *Distribution of AQI*

In the above figure 5 shows the distribution of air quality index is calculated between frequency and air quality data it is going on increasing when the health impact is increasing and then decreasing. At first in the 0^{th} stage health impact is increasing and in between 200 to 300 the health impact is decreasing at 300 it is going on increasing at last in-between 400 to 500 the health impact is decreasing.

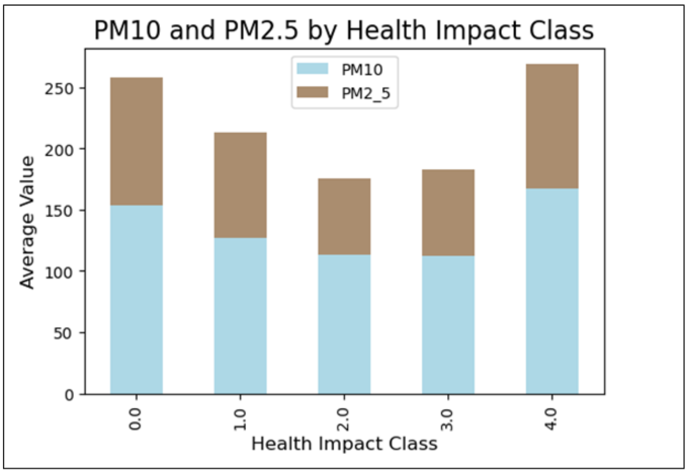

Figure 6: *PM10 and PM2.5 by health impact class for AQ*

Health impact class for PM10 and PM2.5 is calculated using the air quality dataset here the average values ranges from 0 to 300 and the health impact class ranges from 0.0 to 4.0. At first the average value of the health impact class PM10 is 0.0 to 150 and at last the PM10 is 4.0 to 180. However the health impact is low in the sample 2.0.

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 03 (11), November 2024, pp. 33-49

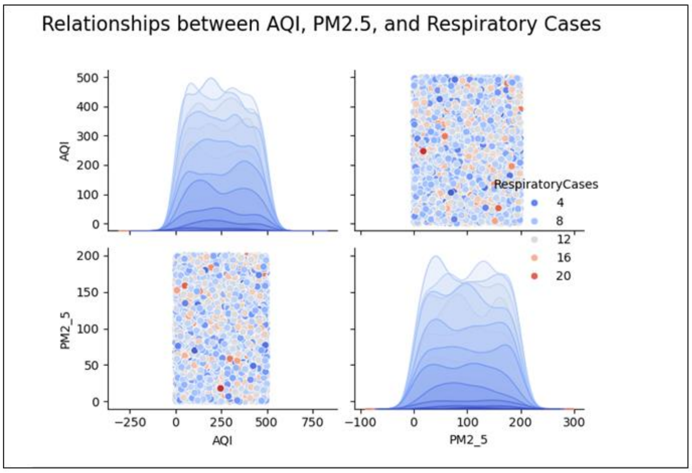

Figure 7: *Relationships between AQI, PM2.5 and respiratory cases for AQ*

In the above figure 7 the relationships between AQI, PM2.5 and respiratory cases are predicted using the air quality data. The respiratory cases have five cases which are 4, 8, 12, 16, and 20. The AQI data for respiratory cases in PM2.5 the respiratory cases are spread over the samples of 0 to 500 (8, 12 and16 are spreading very high)

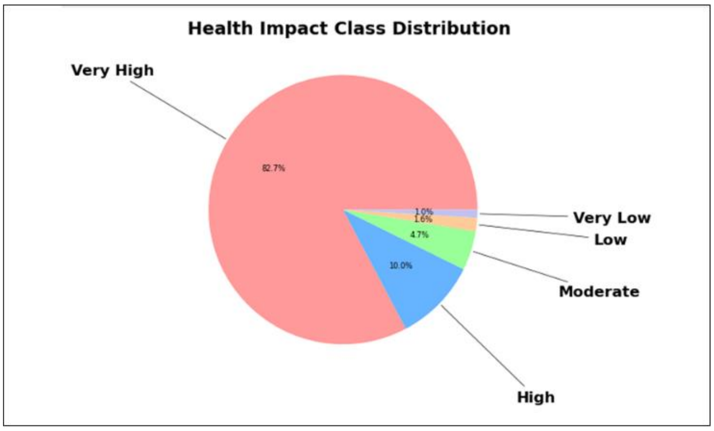

Figure 8: *Health impact class distribution for AQ*

In Figure 8 the health impact class distribution the classes for respiratory cases are split as very low, low, moderate, high and very high. In very low it is 1.0%, low it is 1.6%, moderate it is 4.7%, high it is 10% and very high it is 82.7%.

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 03 (11), November 2024, pp. 33-49

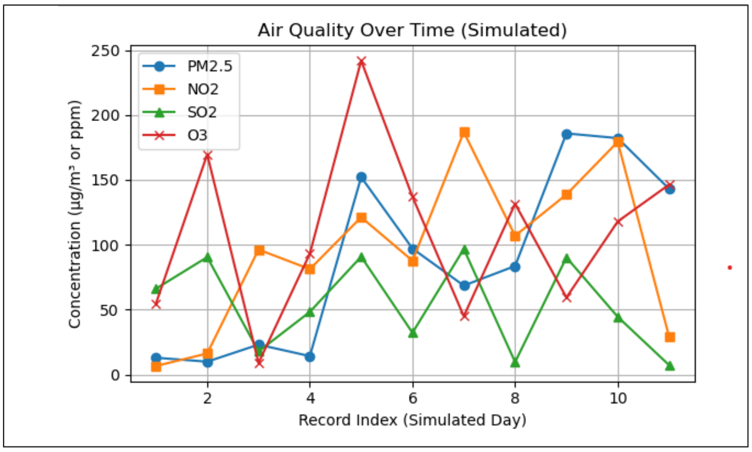

Figure 9: *Air quality over time*

During day time the air quality is simulated on concentration and record index for O3, SO2, NO2 and PM2.5. The air quality data is calculated over time and it varying their concentration and record index. Here O3 is highly varying in the sample 4 to 8 and then the PM2.5 is slightly varying in between the sample 0 to 4 and then going on increasing in-between 4 and 6 and slightly lower in between 6 and 8 and then high in between 8 and 10 is shown in Figure 9.

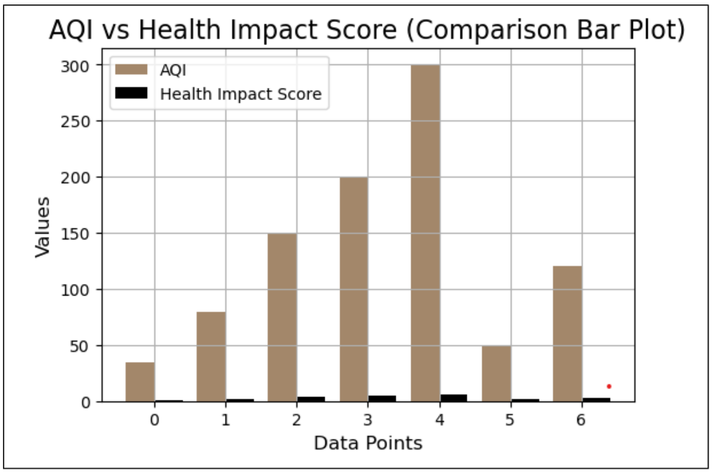

Figure 10: *AQI vs health impact score*

In the above figure 10 AQI vs health impact score is plotted here values ranges from 0 to 300 and the data point is 0 to 6. Whereas the AQI is high on sample 4 which means it varies 300 and the health impact score is 4 to 10.

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 03 (11), November 2024, pp. 33-49

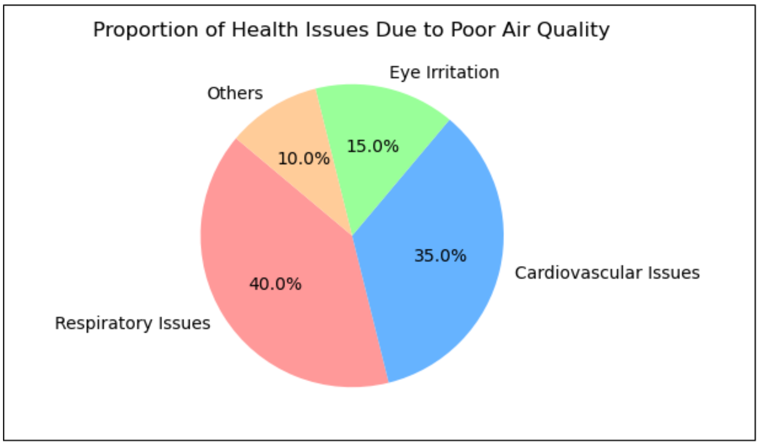

Figure 11: *Proportion of health issues due to poor air quality*

In the above figure 11shows the proportion of health issues due to poor air quality the eye irritation is caused about 10%, cardiovascular issues is about 35%, respiratory issues is about 40% and other issues are about 10%. Here more where affected by the respiratory issues.

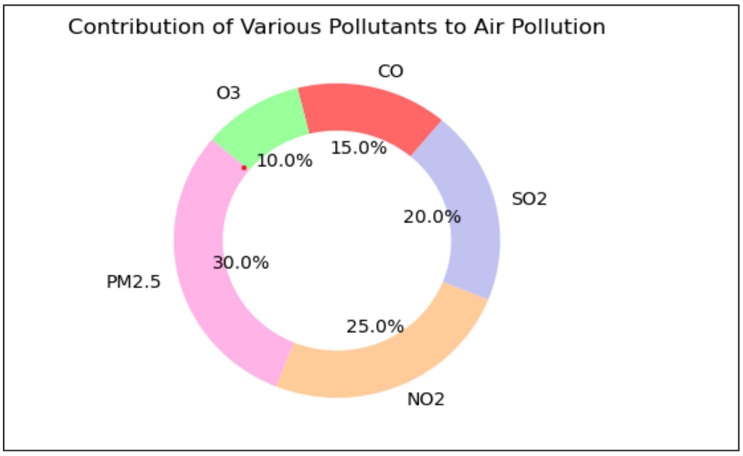

Figure 12: *Contribution of various pollutants to air pollution*

In the above figure 12 contribution of various pollutants to air pollution is calculated for O3, CO, SO2, NO2 and PM2.5. Whereas air pollution of O3 is about 10%, CO is about 15%, SO2 is about 20%, NO2 is about 25% and PM2.5 is about 30% like this the various pollutants are contributed.

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 03 (11), November 2024, pp. 33-49

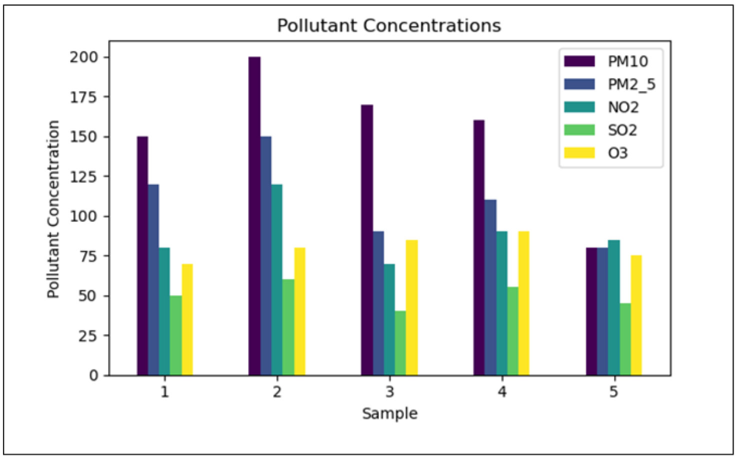

Figure 13: *Pollutants concentrations*

In the above figure 13 the pollutant concentration is calculated for the 5 samples and the pollutant concentration ranges from 0 to 200. In sample 1 the PM10 pollution is about 0 to 150, PM2.5 pollution is about 0 to 120, NO2 is about 0 to 80, SO2 is about 0 to 50 and O3 is about 0 to 70. In sample 2 the PM10 pollution is about 0 to 200, PM2.5 pollution is about 0 to 150, NO2 is about 0 to 120, SO2 is about 0 to 70 and O3 is about 0 to 75. In sample 3 the PM10 pollution is about 0 to 175, PM2.5 pollution is about 0 to 80, NO2 is about 0 to 60, SO2 is about 0 to 40 and O3 is about 0 to 75. In sample 4 the PM10 pollution is about 0 to 160, PM2.5 pollution is about 0 to 120, NO2 is about 0 to 80, SO2 is about 0 to 50 and O3 is about 0 to 80. In sample 5 the PM10 pollution is about 0 to 75, PM2.5 pollution is about 0 to 75, NO2 is about 0 to 80, SO2 is about 0 to 40 and O3 is about 0 to 70.

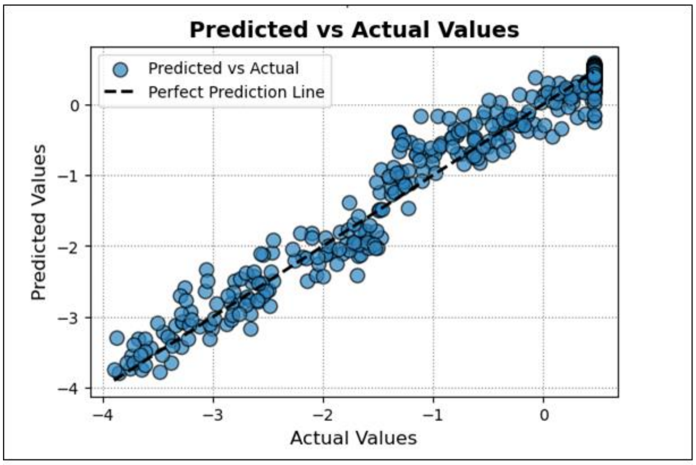

Figure 14: *Predicted vs actual values of AQ*

In the above figure 14 the predicted value vs actual value is calculated to find the perfect predicted line using the air quality data .Where the value which is to calculate the RMSE and MSE value using the predicted value.

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 03 (11), November 2024, pp. 33-49

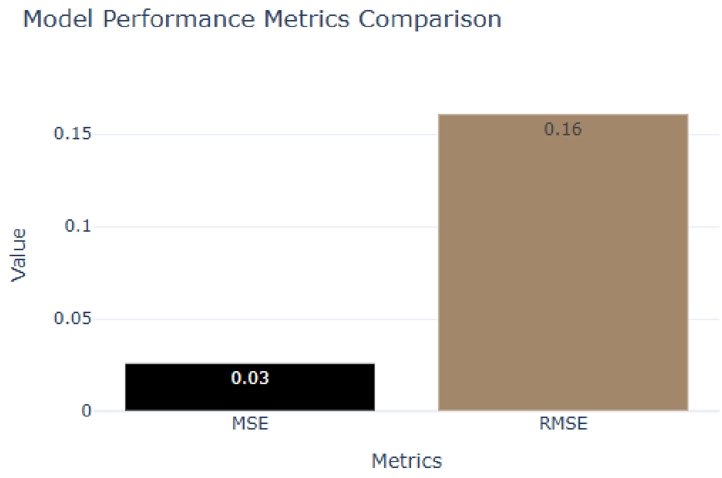

Figure 15: *Model performance metrics comparison for AQ*

In the above figure 15 shows model performance metrics comparison the Mean Square Error (MSE) and Root Mean Square Error (RMSE) is calculated Where the MSE value is 0.03, RMSE value is 0.16 and R2 is 98%.

Comparison for R2 coefficient

In contrast, several other studies [19, 21] in figure 16 depend on baseline data, ignoring the training data's component. These studies did not sufficiently explain their findings and lacked training data for prediction analysis.

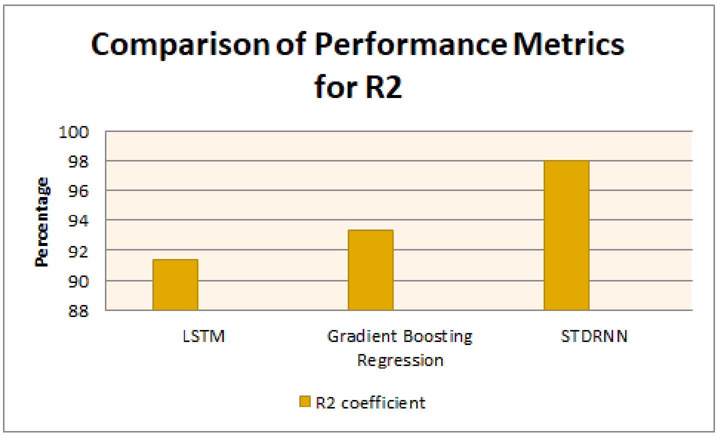

Figure 16: *Comparison of performance metrics for R2 coefficients*

5 Conclusion

In this work a comprehensive and efficient frame work for the prediction of air quality using the proposed STDRNN. The integration of advanced data preprocessing techniques, such as missing data and duplicate AQ values, ensures that the input AQI data are optimized for feature construction. The use of approximate feature encoding and feature scaling for each

AQ label is converted into an integer AQ value and scaling the features to a similar range in the AQI data. Additionally, labelled data are predicted and shared queue list in the ground-truth labels are predicted using the Self-tuning data efficient deep learning technique and the predicted data are given to the hidden layers were the neural networks using the deep regression neural network the output is stored. The best selection is then subsequently selected using the proposed self-tuning deep regression neural network to identify the AQ. The performance evaluation, conducted on the AQI dataset, demonstrates the superior capabilities of the STDRNN model compared to existing state-of-the-art methods. By using regression mechanism, the model focuses on critical features, resulting in improved MSE value of 0.03 and RMSE value of 0.16 and R2 coefficient of 98 %.

References

1. Jindong Han; Hao Liu; Haoyi Xiong; Jing Yang, Year: 2022 "Semi-supervised air quality forecasting via self-supervised hierarchical graph neural network", IEEE Transactions on Knowledge and Data Engineering, Vol.35, no: 5, pp. 5230-5243.

2. Hanqing Xu; Yang Jia; Zhendong Sun; Jiahui Su; Qian S. Liu; Qunfang Zhou; Guibin Jiang, Year: 2022, "Environmental pollution, a hidden culprit for health issues", Eco-Environment & Health, Vol. 1, no.1, pp. 31-45.

3. Khaiwal Ravindra; Tanbir Singh; Shikha Vardhan; Aakash Shrivastava; Sujeet Singh; Prashant Kumar; Suman Mor, Year: 2022, "COVID-19 pandemic: What can we learn for better air quality and human health", Journal of infection and public health, Vol. 15, no. 2, pp. 187-198.

4. Md Milon Islam; Ashikur Rahaman; Md Rashedul Islam, Year: 2020, "Development of smart healthcare monitoring system in IoT environment", SN computer science, Vol. 1, pp.1-11.

5. Chi-Wen Chen; Yu-Sheng Tseng; Arvind Mukundan; Hsiang-Chen Wang, Year: 2021, "Air pollution: Sensitive detection of PM2. 5 and PM10 concentration using hyperspectral imaging", Applied Sciences, Vol. 11, no. 10, pp. 4543.

6. Arvind Mukundan; Chia-Cheng Huang; Ting-Chun Men; Fen-Chi Lin; Hsiang-Chen Wang, Year: 2022, "Air pollution detection using a novel snap-shot hyperspectral imaging technique", Sensors, Vol. 22, no. 16, pp. 6231.

7. An Wang; Junshi Xu; Ran Tu; Marc Saleh; Marianne Hatzopoulou, Year: 2020, “Potential of machine learning for prediction of traffic related air pollution”, Transportation Research Part D: Transport and Environment, Vol. 88, pp.102599.

8. Lakindu Mampitiya; Namal Rathnayake; Lee P. Leon; Vishwanadham Mandala; Hazi Md Azamathulla; Sherly Shelton; Yukinobu Hoshino; Upaka Rathnayake, “Machine learning techniques to predict the air quality using meteorological data in two urban areas in Sri Lanka”, Environments, Vol.10, no. pp.141.

9. Jun Ma; Yuexiong Ding; Jack CP Cheng; Feifeng Jiang; Yi Tan; Vincent JL Gan; Zhiwei Wan, Year: 2020, “Identification of high impact factors of air quality on a national scale using big data and machine learning techniques”, Journal of Cleaner Production, Vol. 244, pp. 118955.

10. Yong Guo; Kangwei Li; Bin Zhao; Jiandong Shen; William J. Bloss; Merched Azzi; Yinping Zhang, Year: 2022, “Evaluating the real changes of air quality due to clean air actions using a machine learning technique: Results from 12 Chinese mega-cities during 2013–2020”, Chemosphere, Vol. 300, pp.134608.

11. K. Kumar; and B. P. Pande, Year: 2023, “Air pollution prediction with machine learning: a case study of Indian cities”, International Journal of Environmental Science and Technology, Vol. 20, no. 5, pp. 5333-5348.

12. Rana Muhammad Adnan; Zhihuan Chen; Xiaohui Yuan; Ozgur Kisi; Ahmed El-Shafie; Alban Kuriqi, Misbah Ikram,Year: 2020 “Reference evapotranspiration modeling using new heuristic methods”, Entropy, Vol. 22, no. 5, pp. 547.

13. Aysenur Gilik; Arif Selcuk Ogrenci; Atilla Ozmen, Year: 2022, “Air quality prediction using CNN+ LSTM-based hybrid deep learning architecture”, Environmental science and pollution research, pp. 1-19.

14. Bowen Cui; Minyi Liu; Shanqiang Li; Zhifan Jin; Yu Zeng; Xiaoying Lin, Year: 2023, “Deep learning methods for atmospheric PM2. 5 prediction: A comparative study of transformer and CNN-LSTM-attention”, Atmospheric Pollution Research, Vol.14, no. 9, pp. 101833.

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 03 (11), November 2024, pp. 33-49

15. Yun-Chia Liang; Yona Maimury; Angela Hsiang-Ling Chen; Josue Rodolfo Cuevas Juarez, Year: 2020, “Machine learning-based prediction of air quality”, applied sciences, Vol. 10, no. 24, pp. 9151.

16. JoonHo Jang; Seungjae Shin; Hyunjin Lee; Il-Chul Moon, Year: 2020 “Forecasting the concentration of particulate matter in the seoul metropolitan area using a gaussian process model”, Sensors, Vol. 20, no. 14, pp.3845.

17. Rajnish Rakholia; Quan Le; Bang Quoc Ho; Khue Vu; Ricardo Simon Carbajo, Year: 2023, “Multi-output machine learning model for regional air pollution forecasting in Ho Chi Minh City, Vietnam”, Environment international, Vol. 173, pp. 107848.

18. Caroline Mogollón-Sotelo; Alejandro Casallas; Sergio Vidal; Nathalia Celis; Camilo Ferro; Luis Belalcazar, Year: 2021, “A support vector machine model to forecast ground-level PM 2.5 in a highly populated city with a complex terrain”, Air Quality, Atmosphere & Health, Vol. 14, pp. 399-409.

19. Qian Liu; Bingyan Cui; Zhen Liu, Year: 2024, “Air Quality Class Prediction Using Machine Learning Methods Based on Monitoring Data and Secondary Modeling”, Atmosphere, Vol.15, no. 5, pp.553.

20. Nilesh N Maltare; Safvan Vahora, Year: 2023, “Air Quality Index prediction using machine learning for Ahmedabad city”, Digital Chemical Engineering, Vol. 7, pp.100093.

21. K. S. Harishkumar; K. M. Yogesh; Ibrahim Gad, Year: 2020, “Forecasting air pollution particulate matter (PM2. 5) Using machine learning regression models”, Procedia Computer Science, Vol. 171, pp. 2057-2066.