Grenze International Journal of Engineering and Technology, June Issue

Leveraging Machine Learning and Deep Learning Percepts for Air Pollution Analysis

Mansi Gahlot[1], Maitreyi Katiyar[2], Maitri Katiyar[3] and Ayushi Agarwal[4]
[1-3]Computer Science and Engineering, Abes Engineering College,
Email: Mansi.20b0101158@abes.ac.in, Maitreyi.20b0101067@abes.ac.in, Maitri.20b0101136@abes.ac.in
[4]Assistant Professor, Abes Engineering College,Computer Science And Engineering
Email: Ayushi.agarwal@abes.ac.in

***Abstract*— The main aim of this project is to teach the computer to recognize patterns in air quality data, such as how pollution levels change over time. We use the LSTM network to make sense of different things like pollution amounts, weather conditions, and past AQI numbers.**

What makes this model smart is that it improves at understanding the air by looking at a lot of data. It can handle the tricky and ever-changing nature of the atmosphere. The computer learns from what happened before and adjusts to what's going on right now, which is super useful for predicting air quality in the future.

Utilizing deep learning's predictive powers to improve AQI forecast accuracy is the main goal of this research. The LSTM model exhibits higher adaptability to the dynamic character of meteorological circumstances by assimilation of complex patterns included in the air quality data. The model learns the intricate correlations between different contaminants, climatic variables, and AQI through extensive training. But this study isn't just about computers and data; it's about making our cities and lives better. The computer predictions can help our leaders make wise decisions to protect our health, especially in busy cities. It's like having a friend that always watches out for us, letting us know when things might not be great so we can make changes and live in a cleaner and safer environment.

I. INTRODUCTION

Air pollution is the global concern that affects the health of the humans, animals, environment, and various other living things on earth. As today we see the population of the world rapidly increase due to this more and more people moving to the cities from the urban areas. The expansion of cities and industries leads to various issues , and one of these is air pollution. It is a very complex issue driven by the presence of various particles and gasses in the air. The Pollutants present is air is not only decrease the quality of air but also harm our well-being and ecosystem around us. The most significant particles present in our air is Particulate Matter(PM),Nitrogen Dioxide(NO_2), Sulphur Dioxide(SO_2), Volatile Organic Compound(VOCs) etc.As we discuss about Particulate Matter, It refers to tiny solid particles or liquid droplets present in air. This particle comes in different sizes as PM2.5 and PM10. Sources of PM include vehicle exhaust, industrial emission, construction activity and even natural processes like dust storms. Tiny particle known as PM2.5 are a big worry because they can sneak into deep our lungs and cause health issues like asthma and heart problems. Moreover particulate matter reduce visibility, creating hazardous condition for transportation and affecting the aesthetic appeal to our surroundings. Nitrogen dioxide(NO_2) and Sulphur Dioxide(SO_2) are the gasses mainly release from burning fossil fuels.NO_2

Grenze ID: 01.GIJET.10.2.130
© Grenze Scientific Society, 2024

cause smog formation and also responsible for respiratory problems. It's a key player in the creation of ground level ozone, a component of smog that poses serious health risks.On the other hand, SO_2 not only harm our respiratory system but also play a role in acidic rain formation. Acid rain damage ecosystems, corrodes buildings, and affect water quality. Volatile Organic Compounds (VOCs) are the bunch of chemicals that easily turned into the vapor and float into the air. They originate from sources like vehicle emission and industrial process and even natural sources such as plants. It also contribute in the formation of ground level ozone and smog, affecting human health and the air quality.

A. Motivation

Understanding the cause and effect of air pollution is essential for creating awareness in the community as by educating ourselves about the sources and the impacts of air pollution, we can take steps to reduce our own contribution. Mankind cannot be imagined surviving without air. The ongoing developments in various fields of modern human society has impacted the air quality adversely. Daily domestic and industrial activities are the major cause of these hazardous air pollutants. Therefore, the need to monitor as well as predict the quality of air has become essential in this era, especially in a rapidly developing nation like India.

The traditional methods to forecast air quality majorly use historical data for learning amongst which Autoregressive Moving Average ARMA and Autoregressive Integrated Moving Average ARIMA were widely used. But due to the increased amount and complexity of data they could not keep up with the required demand. Compared to these conventional methods, the machine learning based prediction techniques have proven to be more effective in studying the air pollutants and their hazards. Therefore, in the treatment of air pollution in urban areas, accurate air pollution prediction is of the most essential tasks.

B. Scope

Our idea is basically to predict the AQI (air quality index) in the air by analysing the past data. It will definitely help in estimating the pollution and will address the current issue regarding the pollution. By using the pollutant analysis, we can also mitigate the level of pollution by taking proper steps and measures. We can further collect the data of all routes and this approach can be expanded by incorporating a simulation mechanism capable of emulating congestion and pollution patterns within an urban area, and thereby extracting more meaningful insights from the available data. By deploying such a system and centralizing the collected data, both real-time pollution tracking, and comprehensive long-term pollution analysis can be conducted with ease. This proposed approach addresses the current issue of insufficient data points for estimating pollution along specific routes or roads within the city.

II. LITERATURE SURVEY

There are various methods and techniques that are used to implement a prediction model for AQI.Some of them are shown with their result below by a table-

TABLE I: LITERATURE SURVEY

S.NO	TITLE	AUTHOR	YEAR OF PUBLISHING	ALGORITHM USED	ADVANTAGE	JOURNAL NAME	RESULT PERCENTAGE
1	Multi-step forecast of PM2.5 and PM10 concentrations using convolutional neural network integrated with spatial–temporal attention and residual learning	Zhang, K., Yang, X., Cao, H., Thé, J., Tan, Z., & Yu, H.	2023	a hybrid forecasting system based on convolutional neural network integrated with spatial–temporal attention and residual learning (STA-ResCNN) for multi-step prediction of PM2.5 and PM10 concentrations	achieved accurate prediction results by integrating CNN with residual learning and also prevent model performance from degrading.	Environment International	average improvement rates of the proposed STA-ResCNN model compared to ANN, LSTM, GRU, CNN and TA-ResCNN model are 15.247 %, 12.781 %, 11.142 %, 9.604 % and 5.595 %, respecively.
2	Multi-output machine learning model for regional air pollution forecasting in Ho Chi Minh City, Vietnam	Rakholia, R., Le, Q., Ho, B. Q., Vu, K. H. N., & Carbajo, R. S	2023	Using an N-BEATS neural network-based architecture consisting of a deep stack of fully connected layers with backward and forward residual links, a global model was developed and trained on multiple series at the same time	to reduce the effort required to develop several models for each pollutant, as well as to create a deployment pipeline, maintain, and monitor the models.	Environment International	average RMSE values of model were lower than the average standard deviation of O3, SO2, CO, and NO2 across all stations, and the correlation score between actual and predicted values is from 0.81 to 0.85, except for CO, which has a score of0.70
3	Long-term exposure to air pollution and COVID-19 severity: A cohort study in Greater Manchester, United Kingdom	Hyman, S., Zhang, J., Andersen, Z. J., Cruickshank, S. M., Möller, P., Daras, K., Williams, R., Topping, D., & Lim, Y.	2023	To visualize the relationship between air pollution exposure and COVID-19 severity they used generalized additive models with a regres-sion spline function with 5 degrees of freedom for air pollution exposure. The association between long-term exposure to air pollution of PM2.5, PM10, O3, NO2, SO2, and benzene (2019 annual levels) and COVID-19 severity (hospitalizations and mortality in separate	This study, using individual-level data on 313,657 residents of Greater Manchester, United Kingdom, who tested positive for SARS-CoV-2, showed significant positive associations between residential exposure to air pollution and severe COVID-19 requiring hospitalization or	Environmental Pollution	Found that older people, obese individuals, current smokers, and individuals with historic cardiovascular disease (hypertension, CHD, stroke), type 2 diabetes, and COPD had greater risk of developing severe COVID-19 related to air pollution.
4	Air pollution prediction using LSTM deep learning and metaheuristics algorithms	Drewil, G. I., & Al-Bahadili, R. J.	2022	Their model is based upon Genetic Algorithm(optimization algorithm) and LSTM deep learning algorithm.It finds the best parameters for LSTM and the level of pollution for the next day by using these pollutants- PM2.5, PM10, NOX and CO	Their model uses optimization algorithms(GA) and gives result with more accuracy with less experience and has more speed compared to the machine learning models as well as LSTM models.	Measurement: Sensors	It was noted that the Genetic optimization algorithm has brought a major improvement in the prediction of LSTM with RMSE value 9.5820 and MAE value 19.164

5	A new hybrid optimization prediction model for PM2.5 concentration considering other air pollutants and meteorological conditions	Yang, H., Liu, Z., & Li, G.	2022	a new hybrid optimization prediction model for PM2.5 concentration based on CEEMDAN, COOT-VMD and JAYA-LSSVM, named CEEMDAN-COOT-VMD-JAYA-LSSVM, is proposed. Firstly, the original sequence of PM2.5 concentration is decomposed by CEEMDAN. Secondly, the high complexity component with low prediction accuracy after once decomposition is decomposed by COOT-VMD. Thirdly, a prediction model of optimized LSSVM by JAYA optimization algorithm, named JAYA-LSSVM, is proposed.	Diebold Mariano test results display the proposed prediction model is superior to all comparison models at 99% confidence level.	Chemosphere	The results of experiment in Xi'an show that the RMSE, MAE, MAPE and r2 ar 2.843, 1.8344, 2.94%, and 0.99525 respectively. The results of experiment in Shenyang show that the RMSE, MAE, MAPE and r2 are 2.2714, 1.673, 3.13%, and 0.99573 respectively. Compared to other single and hybrid models, the proposed model can accurately predict PM2.5 concentration.
6	Hybrid interpretable predictive machine learning model for air pollution prediction	Gu, Y., Li, B., & Meng, Q.	2022	proposed HIP-ML model is to design a hybrid model that combines an interpretable model and a deep neural network.	the hybrid model can achieve both transparency and explainability with one single model framework.	Neurocomputing	experimental results show that the overall prediction accuracy (measured by correlation coefficient) of 1, 3 and 6 hours ahead prediction is 0.9855, 0.9253 and 0.8294, respectively
7	Air pollution prediction with machine learning: a case study of Indian cities	Kumar, K. S., & Pande, B. P.	2022	correlation-based feature selection technique is exercised to flter AQI afecting pollutants for further study and logarithmic transformations are applied to the skewed features. The exploratory data analysis methods are exercised to fnd various hidden patterns present in the dataset. The data imbalance problem is addressed by the SMOTE analysis. The dataset is split into train-test subsets by the ratio of 75–25% respectively. ML-based AQI prediction is carried out with and without SMOTE resampling technique and a comparative analysis is presented.	The performances of these models are evaluated and compared through established performance parameters. The XGBoost model performed the best among the other models and gets the highest linearity between the predicted and actual data.	International Journal of Environmental Science and Technology	results of ML models for both the train-test subsets are presented in terms of standard metrics like accuracy, precision, recall, and F1-Score. For both the train-test sets, the XGBoost model attained the highest accuracy and the SVM model exhibited the lowest accuracy
8	RCL-Learning: ResNet and convolutional long short-term memory-based spatiotemporal air pollutant concentration prediction model	Zhang, B., Zou, G., Qin, D., Ni, Q., Mei, H., & Li, M.	2022	a big data correlation principle and deep learning technology are used for a proposed model of predicting PM2.5 concentration. The model comprises a deep learning network model based on a residual neural network (ResNet) and a convolutional long short-term memory (LSTM) network (ConvLSTM). ResNet is used to deeply extract the spatial distribution features of pollutant concentration and meteorological data from multiple cities. The output is used as input to ConvLSTM, which further extracts the preliminary spatial distribution features extracted from the ResNet, while extracting the spatiotemporal features of the pollutant concentration and meteorological data.	The model combines the two features to achieve a spatiotemporal correlation of feature sequences, thereby accurately predicting the future PM2.5 concentration of the target city for a period of time.Compared with other neural network models and traditional models, the proposed pollutant concentration prediction model improves the accuracy of predicting pollutant concentration.	Expert Systems With Applications	For 1- to 3-hours prediction tasks, the proposed pollutant concentration prediction model performed well and exhibited root mean square error (RMSE) between 5.478 and 13.622. In addition, we conducted multiscale predictions in the target city and achieved satisfactory performance, with the average RMSE value able to reach 22.927 even for 1- to 15-hours prediction tasks.
9	A Hybrid Spatiotemporal Deep Model Based on CNN and LSTM for Air Pollution Prediction	Tsokov, S., Lazarova, M., & Aleksieva-Petrova, A.	2022	The suggested deep spatiotemporal model for predicting air pollution with the automatic selection of input variables uses a genetic algorithm for the optimization of network architecture and hyperparameters. In addition, a hybrid strategy for filling in the missing values in time series is proposed	The used hybrid strategy for missing value imputation enhances the results, especially for data with clear seasonality, and produces better MAE compared to the strategy using average values for the same hour of the same day and month in other years.	Sustainability	the result shows that that random searching is a simple and effective strategy for selecting the input variables
10	Prediction of Air Pollution Interval Based on Data Preprocessing and Multi-Objective Dragonfly Optimization Algorithm	Wang, J., Li, J., & Li, Z.	2022	propose a combined prediction system based on fuzzy granulation, multi-objective dragonfly optimization algorithm and probability interval, which can effectively analyze the volatility and uncertainty of pollution.	for data denoising strategy, the combined system computes the data series without fluctuation and uncertainty by FIG, which yields better performance compared with single mode by decomposing and reconstructing the initial dat	Frontiers in Ecology and Evolution	the combined prediction system can not only effectively predict the changing trend of pollution data and analyze local characteristics but also provide strong technical support for the early warning of air pollution
11	Air-pollution prediction in smart city, deep learning approach	Bekkar, A., Hssina, B., Douzi, S., & Douzi, K.	2021	a hybrid model which is based upon CNN-LSTM, was used in prediction of PM2.5 as air pollutant in Beijing. Seven other deep learning predictive models were also made for compari-son with their proposed model.	it is expected to generate high accuracy of PM2.5 by learning the Spatio-temporal information from the big data. Also it can learn more effectively by using the data from nearby monitoring stations	Journal of Big Data	RMSE is 12.921 and MAE is 6.742. the R^2 value of CNN-LSTM comes out to be 0.989
12	Ensemble multifeatured deep learning models for air quality forecasting.	Lin, C., Chang, Y., & Abimannan, S	2021	proposes an Ensemble Hybrid Models based on Multiple Linear Regression. we use statistical multiple linear regression technology to consider the five basic GRU prediction models together to produce an ensemble hybrid deep learning model called MLEGRU, which integrates these five GRU models.	The three evaluation indices demonstrate that the proposed MLE-GRU model is superior to other models either short-term or long-term forecasting.	Atmospheric Pollution Research	The proposed models were evaluated using three well-known factors; namely MAE, RMSE, and Absolute Error less than 3
13	Integrated Multiple Directed Attention-based Deep Learning for Improved Air Pollution Forecasting	Dairi, A., Harrou, F., Khadraoui, S., & Sun, Y	2021	The paper introduces first the traditional Variational AutoEncoder (VAE) and the attention mechanism to develop the forecasting modeling strategy based on the innovative Integrated Multiple Directed Attention Deep Learning architecture (IMDA)to forecast concentrations of ambient pollutants. To assess the performance of the proposed forecasting methodology, experimental validation is then performed using air pollution data from four US states.	this is the first study introducing the traditional VAE and the attention mechanism to accurately forecast various air pollutants concentrations and leverage air pollution complexity. proposed method is validated through three different forecasting ex-perimentations: uni-variate, multi-variate with a single output, and multi-variate with multiple outputs.	IEEE Transactions on Instrumentation and Measurement	results indicate that the proposed IMDA-VAE model can effectively improve air pollution forecasting performance and outperforms the deep learning models, namely VAE, Long Short-Term Memory (LSTM), Gated Recurrent Units (GRU), bidirectional LSTM, bidirectional GRU, and ConvLSTM
14	Forecasting Air Pollution Particulate Matter (PM2.5) Using Machine Learning Regression Models	Doreswamy, Harishkumar, K. S., Km, Y., & Gad, I.	2020	Analize different algirithms auch as Linear Regression,Lasso,Ridge,Random Forest Regressor,Gradient Boosting Regressor,K Neighbors Regressor,Decision Tree Regressor CART	Analyze best algorithm with least RMSE value	Procedia Computer Science	gradient boosting regressor model is best for forecasting air pollution on the TAQMN data.
15	Predicting Daily Air Pollution Index Based on Fuzzy Time Series Markov Chain Model	Alyousifi, Y., Othman, M., Sokkalingam, R., Faye, I., & De Lima E Silva, P. C.	2020	In this, a Fuzzy Time Series Markov chain (FTSMC) model based on a grid method with an optimal number of partitions was used to predict the daily Air Pollution Index (API).	This is considered the first study that has ever properly defined the number of partitions in the FTSMC model. In forecasting the daily API data, it shows that the proposed model has produced a higher prediction accuracy as compared to some FTS models.	Symmetry	The FTSMC model has been validated using 3 statistical criteria, with a RMSE of 17.01, MAPE of 17.32 and Theils-U of 0.80

16	Air pollution prediction using semi-experimental regression model and Adaptive Neuro-Fuzzy Inference System	Zeinalnezhad, M., Chofreh, A. G., Goni, F. A., & Klemeš, J. J.	2020	This study applies a regression technique as well as ANFIS modelling to conduct air pollution forecasting dealing with data for chaotic time-series schemes' prediction. According to these models, the concentration of a pollutant in the current time depends on the concentration of pollutants in the past 5 d.	Through comparison with the reviewed models on the same data scheme, it was found that the ANFIS model with time-series data has acceptable accuracy.	Journal of Cleaner Production	Adaptive Neuro-Fuzzy Inference System is more accurate in predicting time-series data than regression models.
17	A new method for prediction of air pollution based on intelligent computation	Al-Janabi, S., Mohammad, M., & Al-Sultan, A. Y.	2019	the predictor based on the LSTM method by identifying the best structure and parameter values (weight, bias, number of hidden layers, number of nodes in each hidden layer, and activation function) for the network using the functional particle swarm optimization(PSO) algorithm	The combination of LSTM and SPO reduces the training time for the network because the SPO algorithm provides the best function for activation and identifies the number of hidden layers and number of nodes in each hidden layer	Soft Computing	The combination of LSTM and SPO accurately predict the concentration of particles.
18	Air Pollution Forecasting Using a Deep Learning Model Based on 1D Convnets and Bidirectional GRU	Qing, T., Liu, F., Li, Y., & Sidorov, D.	2019	a short-term forecasting model based on deep learning is proposed for PM2.5 concentration, and the convolutional-based bidirectional gated recurrent unit (CBGRU) method is presented, which combines 1D convnets (convolutional neural networks) and bidirectional GRU (gated recurrent unit) neural networks.	Comparing the prediction results with the traditional ones, it is proved that the error of the CBGRU model is lower and the prediction performance is better.	IEEE Access	It showed at a RMSE value 14.5319, MAE 10.4798 and SMAPE 0.2055
19	Real-time Air Pollution prediction model based on Spatiotemporal Big data	Le, D. C.	2018	CNN algorithm for imagelike Spatial distribution of air pollution. For Temporal data LSTM is used and a Neural Network model for other air pollution impact factors	Hybrid prediction model is used. This model is simple in architecture but still brings good prediction ability.	arXiv (Cornell University)	75% accuracy achieved
20	A Comprehensive Evaluation of Air Pollution Prediction Improvement by a Machine Learning Method	Xi, X., Wei, Z., Xiaoguang, R., Yijie, W., Xinxin, B., Wenjun, Y., & Jin, D	2015	design a comprehensive evaluation framework to improve the prediction performance.5 classify algorithms in machine learning methods(random forest model, gradient booting model, SVM model,decision tree model, combined model of the four above models) are adopted with different features groups which come from WRF-Chem models forecast results.	The best result can be obtained by different group of feature selection and model selection. Experimental results indicate that the more feature we used, the more possibility to enhance the accuracy.	IEEE	the result from combined model is better than the unique model

III. METHODOLOGY

The proposed system is classified into these modules to predict the AQI.

A. Data Collection

The following meteorological data was obtained from TuTiempo.net- average relative humidity (%), total rainfall (in mm), visibility (in km), average wind speed (in km/h), maximum sustained wind speed (in km/h), maximum temperature (in CC), minimum temperature (in CC), and average visibility (in Km).
The air pollutants concentrations, including those of PM2.5, PM10, NO, NO2, NOx, NH3, CO, SO2, and O3 were gathered via Kaggle.

B. Data Cleaning

Data cleaning includes finding missing, erroneous, imprecise, or unnecessary portions of the data and then removing, altering, or replacing the unclean or coarse data. The concentrations of air contaminants were obtained from the Kaggle dataset, and meteorological parameters were web-scraped from WEBSITE. The data has been cleaned up to eliminate numerous features including "NH3," "NO," "Benzene," "Toluene," "Xylene," and "AQI Bucket" that are not helpful in AQI prediction. Lastly, PM2.5, PM10, NO2, NOx, CO, SO2, O3, average temperature, maximum temperature, minimum temperature, humidity, visibility, and wind speed are the independent features that are used. While the missing values in independent features were interpolated, rows with missing AQI values were removed. Standard Scalar was used to carry out the feature scaling.

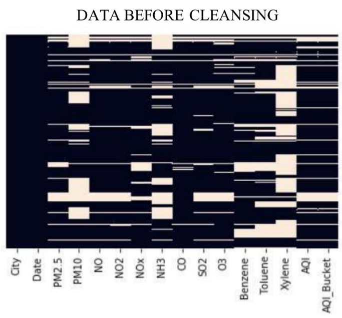

Fig 1: Data Before Cleaning

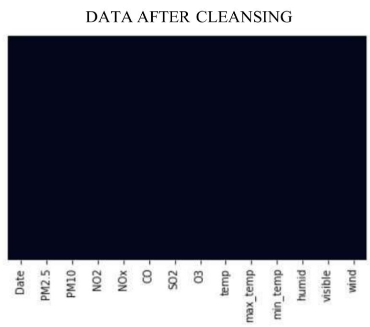

Fig 2 : Data After Cleaning

C. Data Pre-processing

Data quality and representativity is the first and most crucial need to guarantee the efficient building of forecasting models. A machine learning algorithm's ability to generalize is frequently impacted by the data preparation stage. Examples of data preparation include feature engineering, data transformation (usually normalization and standardization), removing or modifying outlier observations, and imputation of missing data. The third step is frequently used to collect data that is more regularly distributed and to reduce data variability, even though the first two phases are crucial for generating more complete and precise data sets. Eventually, a new dataset that is frequently smaller and more informative is produced using the fourth phase. Feature selection and extraction are typically part of the last phase.

D. Data Visualisation

After being cleaned, the data was collected and visualized to verify if its proper or not.

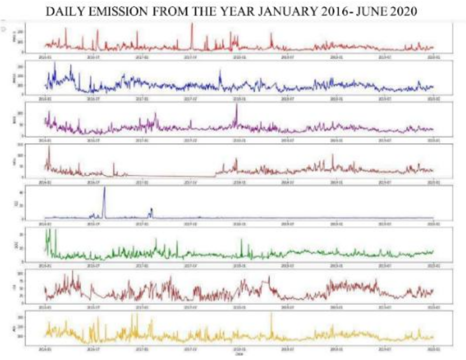

Fig 3: Data visualization

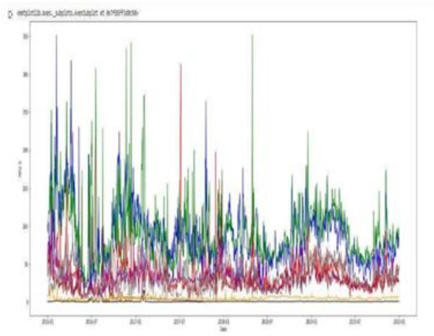

Fig 4: Data visualization

E. Training dataset formation

Data from 2016 to 2019 will be used to train the model, and the AQI values gathered for 2020 will be compared with the values predicted by the LSTM model and several other machine learning models. Data is collected for the years 2016 to 2020.

F. Architecture

Long Short-Term Memory is a type of recurrent neural network. The output of the previous step of an RNN is fed into the current step. Hochester & Schmid 1--luber developed LSTM. It addressed the issue of RNN long-term dependency, wherein the RNN can make more accurate predictions based on current data but is unable to predict words stored in long-term memory. When the gap length grows, RNN performance becomes less effective. By default, LSTM has a long retention time for the data. It is applied to time series data processing, prediction, and classification.

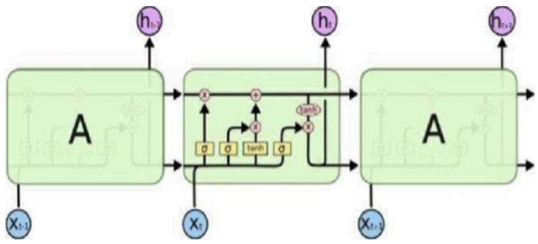

Fig 5: Model Diagram

The given dataset is translated so that a row contains the data for the previous seven days.The target feature is the AQI value for the eighth day. This means that the AQI values for the past seven days will be utilized to forecast the AQI for the next day.

Four models with fully linked and LSTM layers were created. The ReLU activation function receives the output from the fully linked layer and LSTM. To optimize the weights and learning rate values, the Adam optimizer is employed. The loss function that we employed was Mean Squared Error. The learning rate is set at 0.001 at the beginning.

One LSTM layer, one fully linked layer, and an output layer make up the first model.

The second model is made up of an output layer, a fully linked layer, and two LSTM layers. The ReLU activation function receives the output from the LSTM and fully linked layer.

Two LSTM layers, two fully connected layers, and an output layer make up the third model. ReLU processes the output from the layers.

Lastly, by incorporating a dropout layer with a dropout rate of 0.5 into the third model, the fourth and final model is created. The best AQI value forecasting results were produced by this model.

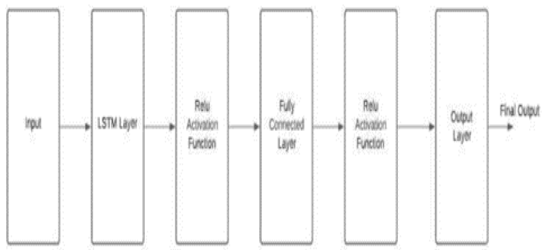

Fig 6: LSTM Model 1

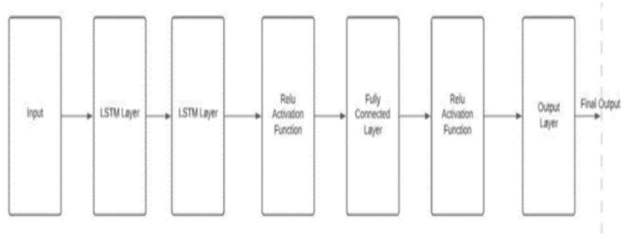

Fig 7: LSTM Model 2

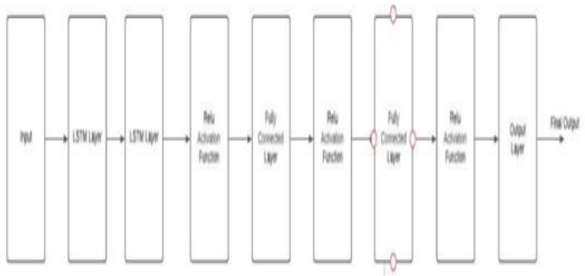

Fig 8: LSTM Model 3

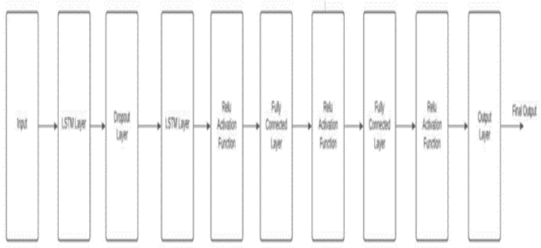

Fig 9: LSTM Model 4

IV. RESULT

There are four LSTM models built to find the best result. The value of absolute error, Root mean square error and mean square error is shown below by a table.

MODEL	R2 SCORE	Mean absolute error	Root mean squared error
LSTM MODEL 1	0.766110725303	0.19053435	0.49583682
LSTM MODEL 2	0.773016977716	0.18732437	0.49213535
LSTM MODEL 3	0.771508592511	0.18938948	0.49295092
LSTM MODEL 4	0.802407929275	0.1743705	0.4753665

Fig 10: Comparison between various LSTM models

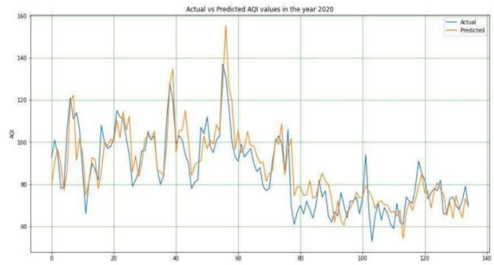

Fig 11: Result Graph

IV. CONCLUSION

Promising forecasting skills are shown by the LSTM model created to predict the Air Quality Index (AQI). The model performs exceptionally well in identifying complex patterns and dependencies within temporal sequences of air quality data by utilizing the capabilities of long short-term memory networks. By means of extensive training on past datasets, the LSTM model has acquired the ability to identify the intricate correlations between diverse atmospheric parameters and AQI, facilitating precise forecasts. This model's potential to improve environmental and public health monitoring initiatives is what makes it significant. Accurate AQI forecasts enable authorities and communities to take proactive measures to address air pollution issues, carry out focused interventions, and reduce health risks. Furthermore, the model's flexibility in responding to changing environmental circumstances makes it an invaluable instrument for real-time monitoring.

FUTURE SCOPE

The AQI prediction model based on LSTM has a lot of potential and can be applied in several scenarios. Future research and development efforts should primarily focus on enhancing the model's adaptability to diverse

geographical and climatic conditions. To train the LSTM network to perform well in a range of environmental settings, this involves utilizing datasets from several sources.
When combined with emerging technologies like Internet of Things (IoT) gadgets and remote sensing platforms, future development has a lot of promise. Real-time sensor data and satellite imagery provide a more comprehensive input to the model, resulting in more accurate and timelier AQI projections. It can be simpler to collaborate with agencies and authorities that keep an eye on the environment.

REFERENCE

[1] Zhang, K., Yang, X., Cao, H., Thé, J., Tan, Z., & Yu, H. (2023). Multi-step forecast of PM2. 5 and PM10 concentrations using convolutional neural network integrated with spatial–temporal attention and residual learning. Environment International, 171, 107691.

[2] Rakholia, R., Le, Q., Ho, B. Q., Vu, K., & Carbajo, R. S. (2023). Multi-output machine learning model for regional air pollution forecasting in Ho Chi Minh City, Vietnam. Environment International, 173, 107848.

[3] Hyman, S., Zhang, J., Andersen, Z. J., Cruickshank, S. M., Möller, P., Daras, K., Williams, R., Topping, D., & Lim, Y. (2023). Long-term exposure to air pollution and COVID-19 severity: A cohort study in Greater Manchester, United Kingdom. Environmental Pollution, 327, 121594.

[4] Drewil, G. I., & Al-Bahadili, R. J. (2022). Air pollution prediction using LSTM deep learning and metaheuristics algorithms. Measurement: Sensors, 24, 100546.

[5] Yang, H., Liu, Z., & Li, G. (2022). A new hybrid optimization prediction model for PM2. 5 concentration considering other air pollutants and meteorological conditions. Chemosphere, 307, 135798.

[6] Gu, Y., Li, B., & Meng, Q. (2022). Hybrid interpretable predictive machine learning model for air pollution prediction. Neurocomputing, 468, 123–136.

[7] Kumar, K. S., & Pande, B. P. (2022). Air pollution prediction with machine learning: a case study of Indian cities. International Journal of Environmental Science and Technology, 20(5), 5333–5348.

[8] Zhang, B., Zou, G., Qin, D., Ni, Q., Mei, H., & Li, M. (2022). RCL-Learning: ResNet and convolutional long short-term memory-based spatiotemporal air pollutant concentration prediction model. Expert Systems With Applications, 207, 118017.

[9] Tsokov, S., Lazarova, M., & Aleksieva-Petrova, A. (2022). A hybrid spatiotemporal deep model based on CNN and LSTM for air pollution prediction. Sustainability, 14(9), 5104.

[10] Bekkar, A., Hssina, B., Douzi, S., & Douzi, K. (2021). Air-pollution prediction in smart city, deep learning approach. Journal of Big Data 8(1).

[11] Wang, J., Li, J., & Li, Z. (2022b). Prediction of air pollution interval based on data preprocessing and Multi-Objective Dragonfly Optimization Algorithm. Frontiers in Ecology and Evolution, 10.

[12] Lin, C., Chang, Y., & Abimannan, S. (2021). Ensemble multifeatured deep learning models for air quality forecasting. Atmospheric Pollution Research, 12(5), 101045.

[13] Dairi, A., Harrou, F., Khadraoui, S., & Sun, Y. (2021). Integrated Multiple Directed Attention-Based Deep Learning for improved air pollution forecasting. IEEE Transactions on Instrumentation and Measurement, 70, 1–15.

[14] Doreswamy, Harishkumar, K. S., Km, Y., & Gad, I. (2020). Forecasting air pollution particulate matter (PM2.5) using machine learning regression models. Procedia Computer Science, 171, 2057–2066.

[15] Alyousifi, Y., Othman, M., Sokkalingam, R., Faye, I., & De Lima E Silva, P. C. (2020). Predicting daily air pollution index based on Fuzzy Time Series Markov Chain model. Symmetry, 12(2), 293.

[16] Zeinalnezhad, M., Chofreh, A. G., Goni, F. A., & Klemeš, J. J. (2020b). Air pollution prediction using semi-experimental regression model and Adaptive Neuro-Fuzzy Inference System. Journal of Cleaner Production, 261, 121218.

[17] Al-Janabi, S., Mohammad, M., & Al-Sultan, A. (2019). A new method for prediction of air pollution based on intelligent computation. Soft Computing.

[18] Qing, T., Liu, F., Li, Y., & Sidorov, D. (2019). Air pollution forecasting using a deep learning model based on 1D Convnets and bidirectional GRU. IEEE Access, 7, 76690–76698.

[19] Le, D. C. (2018). Real-time Air Pollution prediction model based on Spatiotemporal Big data. arXi(Cornell University).

[20] Xi, X., Wei, Z., Xiaoguang, R., Yijie, W., Xinxin, B., Wenjun, Y., & Jin, D. (2015, November). A comprehensive evaluation of air pollution prediction improvement by a machine learning method. In 2015 IEEE international conference on service operations and logistics, and informatics (SOLI) (pp. 176-181). IEEE.