International Refereed Journal of Engineering and Science (IRJES)
ISSN (Online) 2319-183X, (Print) 2319-1821
Volume 13, Issue 2 (Mar.-Apr. 2024), PP. 247-257

Urban Air Pollution: A Comparision of Statistical and Deep Learning Models

[1.] Arepalli Pavani Naga Pravallika, *B.Tech, Department of CSE, DNR COLLEGE OF ENGINEERING AND TECHNOLOGY,pravallika6931@gmail.com*
[2.] Maduthuri Purna Pradeep, *B.Tech, Department of CSE, DNR COLLEGE OF ENGINEERING AND TECHNOLOGY,pradeepmb1379@gmail.com*
[3.] Pathapati Raja Rajesh Varma, *B.Tech, Department of CSE, DNR COLLEGE OF ENGINEERING AND TECHNOLOGY,rajeshvarma1285@gmail.com*
[4.] Garikipati Chandra Sekhar, *B.Tech, Department of CSE, DNR COLLEGE OF ENGINEERING AND TECHNOLOGY,garikipatichandu99@gmail.com*
[5.] Mrs.N.Bharathi, *M.Tech, Assistant Professor, Department of CSE, DNR COLLEGE OF ENGINEERING AND TECHNOLOGY,kbharathi.1224@gmail.com*

Abstract: *A quiet but severe public health catastrophe, air pollution has worsened due to the development of industry and urbanization. Stakeholders must prioritize accurate air quality forecast if they are to successfully address this growing challenge. The purpose of this research is to assess the performance of statistics and deep learning models for predicting urban air pollution levels. We investigate their prediction capacities by means of state-of-the-art methodologies—including Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), Convolutional Neural Networks (CNN), and combinations thereof. Our results suggest that ensemble approaches, especially CNN and CNN + LSTM, are better, with an accuracy of more than 90%. These ensemble approaches show improved performance and build on the underlying model's effectiveness in air quality forecasting. They also highlight the possibility for future development. By combining several model projections, our method provides a solid and precise foundation for everyone involved to tackle air quality concerns head-on, reducing the negative effects of pollution on people's health and the environment. Air quality, statistical approaches, prediction models, deep learning, and machine learning are index terms.*

I. INTRODUCTION

The detrimental effects of air pollution on ecosystems and human health have made it one of the world's most pressing environmental concerns in recent years. The need to reduce air pollution is more pressing than ever before because of the wide variety of health problems it causes, including asthma, lung illness, cognitive impairment, and even death [1][2]. Recent studies have shown that air pollution is responsible for approximately 3 million fatalities each year, with the majority of these casualties occurring in countries with low or medium economic levels [3]. Global efforts like the United Nations Sustainable Development Goals (SDGs) have set goals for 2030 to improve air quality in cities in order to reduce fatalities, diseases, and negative environmental consequences [4]. This is in response to the severity of the problem.

Along with these worldwide initiatives, certain nations, like the UK, have established lofty goals to reduce air pollution. The government of the United Kingdom has committed to a 35% reduction in air pollution by 2040 [5]. These goals highlight the increasing awareness of air pollution as a major threat to public health and the need of working together to reduce its impacts.

Air pollution is caused by a multitude of things, but the most common ones are transportation-related activities, dust, and the burning of coal [6]. When dangerous substances, such as gases and materials, are released into the air, this is called air pollution. At high enough quantities, particulate matter (PM2.5), one of the most prevalent contaminants, may cause serious harm to human health [7][8][9]. These contaminants diminish environmental quality and endanger the health of people and other creatures when they build up in the air. Research, legislative initiatives, and technical improvements must all be part of the solution to the complicated problem of air pollution. In order to create successful methods to reduce the negative consequences of air pollutants on human health and the environment, it is crucial to understand where these pollutants come from, how they spread, and what effects they have. To ensure stakeholders and decision-makers have access to up-to-date information, developments in air quality monitoring and forecasting are vital.

Predictive models and monitoring methods for air quality evaluation have come a long way in the last few years. These models evaluate trends in air quality and predict amounts of pollutants using cutting-edge methods including statistical analysis, deep learning, and machine learning. Through the use of extensive data

sets and advanced algorithms, these models provide valuable insights into the intricate dynamics of air pollution and allow proactive actions to reduce its impacts.

Although air quality modeling has come a long way, there are still a number of obstacles to overcome before we can effectively forecast and control air pollution. Among these difficulties are the following: the need for high-resolution data for model validation and calibration; the impact of several interacting variables on pollution dispersion; and the ever-changing character of atmospheric processes. In order to tackle these difficulties and find novel solutions for managing air quality, scientists, lawmakers, and industry stakeholders must work together across disciplines.

Air pollution is a worldwide problem, and this article will examine its many facets, effects on people and the environment, present mitigation strategies, and the part played by predictive modeling in tackling this issue. Contributing to the continuing conversation on air quality management and providing insights for future research paths, this study attempts to conduct a complete overview of existing research and methodology.

II. LITERATURE SURVEY

There are many different aspects to air pollution, and it has serious consequences for both human health and the ecosystem. Air pollution has been the subject of much study throughout the years in an effort to identify its causes, effects, and ways to lessen it. Focusing on its impacts on cardiovascular health, household consequences, public health, and the creation of prediction models for air quality forecasts, this literature review seeks to provide a thorough overview of the current body of information on air pollution.

According to Brook (2008), there is a strong correlation between breathing in particulate matter and an elevated risk of cardiovascular disease and death [1]. Both short-term and long-term exposure to air pollution has been linked to negative cardiovascular outcomes, such as myocardial infarction, stroke, and hypertension, according to studies. These results highlight the need for further research into the cardiovascular consequences of pollution in the air and for action to reduce this threat to public health.

In poor and medium income nations, home air pollution is a major health hazard, according to the World Health Organization (WHO) [3]. Indoor air pollution is a major public health concern since it may lead to respiratory illnesses, cardiovascular issues, and even negative pregnancy outcomes. Some of the causes of indoor air pollution include cooking fuels, biomass burning, and insufficient ventilation. The World Health Organization has highlighted the need of taking action to reduce indoor air pollution and its negative impacts on the health of vulnerable groups, including children and pregnant women.

Air pollution is a major environmental risk factor for illness burden and early death, as discussed by Landrigan (2017), who also examines the wider public health implications of this issue [6]. Respiratory infections, lung cancer, and neurodevelopmental issues are just a few of the many health problems linked to air pollution. Children, the elderly, and those with preexisting health issues are more likely to be impacted by the negative health impacts of air pollution. Comprehensive public health policies should be implemented to reduce emissions, improve air quality monitoring, and promote sustainable urban growth in order to combat air pollution.

The intricate relationships between ecosystems and human health are highlighted in the extensive assessment of air pollution's effects on the environment and human health that Manisalidis et al. (2020) provide [9]. Exacerbating the worldwide burden of illness, air pollution adds to environmental degradation, biodiversity loss, and climate change. In order to identify the sources of air pollution and create effective solutions to reduce it, the study stresses the need of multidisciplinary research and teamwork.

Forecasts of air pollution levels are now more precise and up-to-date than ever before, made possible by developments in deep learning and machine learning. The use of machine learning regression models to estimate amounts of particulate matter (PM2.5) in the air is investigated by Doreswamy et al. (2020), who show encouraging findings in this area [13]. Similarly, to anticipate air pollution, Chang et al. (2020) suggest an aggregated model based on long short-term memories (LSTM) that uses recurrent neural networks to include temporal dependencies in pollutant data [14]. These studies demonstrate the potential of machine learning and deep learning techniques to improve air quality forecasting and to aid in well-informed decision-making regarding air pollution mitigation initiatives. For example, Tao et al. (2019) introduce a deep learning model for air pollution forecasting that uses 1D convolutional neural networks (ConvNets) and bidirectional gated recurrent units (GRU) to achieve substantial gains in prediction accuracy [20].

In conclusion, the literature review shows how air pollution, human health, and ecological sustainability are all interdependent. The causes and consequences of air pollution have been better understood, but there is still a long way to go before we can devise efficient methods to lessen its negative effects. To tackle the underlying causes of air pollution and encourage sustainable development, future research should center on collaborative efforts across disciplines, new technology, and legislative initiatives.

III. METHODOLOGY

a) Proposed Work:

As part of the planned study, we will compare and contrast the accuracy of air pollution forecasts made using classic statistical models and those made using state-of-the-art deep learning techniques, such as Gated Recurrent Unit (GRU)[19] and Long Short-Term Memory (LSTM)[19] neural networks. Because of their design to process sequential data, LSTM and GRU networks are ideal for time-series prediction problems like air quality modeling, which include temporal relationships. This makes them a good alternative to traditional statistical methods. We want to circumvent the shortcomings of conventional statistical approaches by making use of these state-of-the-art deep learning models and capitalizing on their capacity to detect intricate patterns and time-dependent dynamics in air pollution data.

The project also introduces two high-accuracy models for air pollution prediction, one using a Convolutional Neural Network (CNN) and the other a hybrid CNN+LSTM, which achieves 97% accuracy. This further expands the research's scope. Significant gains in predicting accuracy are available with these state-of-the-art deep learning approaches. The smooth register and signin procedures for user testing are made possible by integrating the user-friendly Flask framework with SQLite. This enhances the practical usability and accessibility of the deep learning models. Users are encouraged to participate and provide critical input to refine and optimize the models for real-world deployment in air quality control. This simplified integration speeds up testing processes.

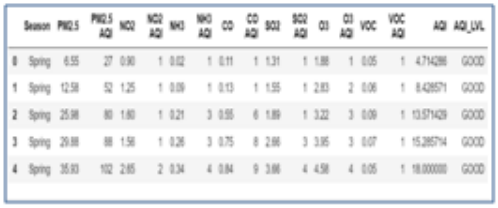
b) System Architecture:

Fig 1 Proposed Architecture

a system architecture project that has several interdependent parts. After the data has been processed and explored, the next step is to visualize the data in order to draw conclusions. Traditional statistical approaches like ARIMA[15] and sophisticated deep learning techniques like LSTM[19], GRU[19], CNN, and CNN+LSTM are used in the model development process, which involves train-test splitting of the dataset. Model performance is evaluated using criteria such as F1 score, recall, accuracy, and precision. In the end, the system uses the chosen models to provide air quality projections, which are helpful for managing and making decisions about urban air quality.

c) Collection:

Air quality monitoring stations in Northern Ireland provided the publicly accessible dataset used in this investigation [30]. Nitrogen Dioxide (NO2), ozone (O3), sulfur dioxide (SO2), and particulate matter (PM2.5 and PM2.10) are among the many air quality metrics included by this extensive dataset, which covers hourly readings. Temperature, wind speed, and wind direction are some of the meteorological variables included in the dataset. These readings were taken in the heart of Belfast from 2015 to 2020, giving a wealth of geographical and chronological information for analyses of air quality. With more than 50,000 samples, the dataset provides a strong basis for investigating the interplay between air pollution and weather conditions. In addition, the dataset contains statistical information specific to weather data, such as the overall sample size, average, standard deviation, and range of values for each parameter. There are more than 50,000 samples in all, and the standard deviations and mean values for all parameters range from 2.77 to 84.87, with a total of 213.19. The dataset shows a wide range of pollutant concentrations, with NO2 readings ranging from 1 to 203, mean 26.11 and standard deviation 17.87. On the other hand, SO2 has the lowest mean and standard deviation among the air pollutants, highlighting how levels of these pollutants vary across various air quality metrics [30].



	Season	PM2.5	PM2.5 AQI	NO2	NO2 AQI	NH3	NH3 AQI	CO	CO AQI	SO2	SO2 AQI	O3	O3 AQI	VOC	VOC AQI	AQI	AQI_LVL
0	Spring	6.55	27	0.90	1	0.02	1	0.11	1	1.31	1	1.88	1	0.05	1	4.714286	GOOD
1	Spring	12.58	52	1.25	1	0.09	1	0.13	1	1.55	1	2.83	2	0.06	1	8.428571	GOOD
2	Spring	25.98	80	1.60	1	0.21	3	0.55	6	1.89	1	3.22	3	0.09	1	13.571429	GOOD
3	Spring	29.88	88	1.58	1	0.26	3	0.75	8	2.66	3	3.85	3	0.07	1	15.285714	GOOD
4	Spring	35.93	102	2.65	2	0.34	4	0.84	9	3.66	4	4.58	4	0.05	1	18.000000	GOOD

Fig 2 Sample Dataset

d) Data Processing:

The first thing to do when processing data is to import the information into a pandas dataframe. Pandas is a powerful Python framework that can handle data manipulation. This makes the dataset's contents easily accessible, which in turn makes data operations more efficient. Researchers may examine the dataset for characteristics, structure, and missing or inconsistent values once it has been loaded. Consequently, the dataset has to be prepared according to Keras' specifications if it is going to be used with the well-known deep learning package. Data rearrangement may be necessary to conform to Keras' input requirements, such as standardizing numerical features for uniform scaling across attributes or transforming categorical variables into one-hot encoded vectors. The compatibility with Keras' neural network models is guaranteed by this format.

After the data has been formatted, any columns that aren't needed for the modeling assignment may be removed. This improves the model's performance, decreases computational cost, and simplifies the dataset by eliminating unnecessary or duplicate characteristics. Based on the analytical goals, columns that include IDs or information that do not contribute to the prediction job may be evaluated for elimination. Data may be prepared for analysis and modeling by following these steps: importing dataset into pandas dataframe, converting to Keras-compatible format, and removing extraneous columns. This methodical procedure guarantees that the dataset is well-organized, suitable for ML algorithms, and free of irrelevant material that can impair model accuracy or understanding.

2) Visualization

By combining Seaborn with Matplotlib, a powerful data visualization tool, analysts and researchers can make eye-catching and informative visuals out of complicated datasets. Statistical visualizations may be made both visually appealing and educational with the help of Seaborn, an interface that is developed on top of Matplotlib. Seaborn, with its user-friendly syntax and visually beautiful default settings, makes it easy to generate scatter plots, line plots, and heatmaps, which improves data exploration. One of the most important libraries for plotting data is Matplotlib. It allows users to create complex visualizations with granular control over plot customisation. With Seaborn and Matplotlib working in tandem, users may take use of Seaborn's ease of use for rapid visualizations and Matplotlib's flexibility for fine-grained customization.

A combination of these libraries allows for the creation of aesthetically pleasing charts, which in turn help to communicate trends, patterns, and correlations in data. The combination of Seaborn with Matplotlib offers researchers a versatile and complete toolbox for accurately communicating complicated discoveries, whether they are investigating variable distributions, displaying correlations, or graphing trends across time.

f) Coding Names:

One of the most basic ways to convert categorical data to numerical representations is label encoding, which the LabelEncoder function makes easy to do. This procedure transforms labels that are not numerical into numerical form by assigning a different integer to each category in a categorical feature. Label encoding allows for the incorporation of categorical variables into predictive models, which is especially helpful in machine learning workflows when algorithms need numerical inputs. The ordinality of categorical data is preserved via label encoding, which may unintentionally create a hierarchical connection among categories, despite its basic implementation. Use extreme care when combining label encoding with algorithms that treat numerical values as ordinal; doing so can cause them to draw erroneous conclusions. Even though it's easy to overlook, label encoding is nevertheless an important part of data preparation since it allows machine learning models to include categorical information without a lot of feature engineering.

g) Selecting Features:

A vital part of data analysis and machine learning is feature selection, which seeks to isolate and keep the most important properties from a dataset while removing any superfluous or unneeded ones. Model

performance, interpretability, and reduction of overfitting are all improved by feature selection, which involves picking a subset of characteristics that contribute most substantially to the predicting job. To evaluate the significance of features and choose the most informative qualities, many strategies are used, including filter methods, wrapper methods, and embedding methods. In contrast to prediction models, filter approaches use statistical metrics or correlation analysis to assess characteristics on their own. To evaluate feature subsets, wrapper approaches iteratively train and evaluate candidate models, with performance serving as the selection criteria. To maximize the relevance of features and model parameters simultaneously, embedded techniques include feature selection into the model training process. Feature selection allows for more robust and interpretable machine learning solutions by systematically evaluating and prioritizing features. This improves model efficiency and generalization.

h) Evaluation and Instruction:

An essential step in deep learning for evaluating a model's performance and generalizability is to divide the data into a training set and a testing set. Before training the deep learning model, the dataset is divided into two parts: the training set and the testing set. The training set is used for training purposes, while the testing set is used to evaluate the model's performance on unknown data. The model's parameters are optimized using iterative optimization methods like gradient descent on the training set, and the model's performance on fresh, unknown samples is estimated on the testing set, which also acts as a validation set. Data splitting aids in avoiding overfitting and gives a credible assessment of the model's generalizability to new data by making sure the training and testing sets are separate. To further evaluate the performance of the model and make the assessment process more reliable, methods like cross-validation may be used. Practitioners of deep learning may guarantee robust model assessment and offer more precise and trustworthy machine learning solutions via thorough data separation.

a) Computer programs that use algorithms:

ARIMA, which stands for "AutoRegressive Integrated Moving Average," is a statistical mainstay when it comes to time series forecasting. It successfully models data connections by combining autoregression, differencing, and moving average components. [15] Predicting future values from previous observations is made easy with its power in identifying linear patterns within sequential data.

Fig 3 ARIMA

LSTM (Long Short-Term Memory): Long short-term memory (LSTM) recurrent neural networks (RNNs) are very good at handling complex sequential input and long-range relationships. The year 19 Time series forecasting tasks are well-suited to its design, which includes memory cells that can store information over lengthy periods. It is good at recognizing and learning from temporal trends.

Fig 4 LSTM

GRU (Gated Recurrent Unit): Another simpler variation of RNN architecture, GRU is similar to LSTM. It has gating mechanisms that allow for the selective updating or removal of information. Time series prediction and language modeling are only two of the many uses for GRU[19] networks, which are tasked with effectively collecting sequential data relationships.

Fig 5 GRU

CNN Algorithm: By digging into spatial and temporal feature extraction from air quality data, the comparison analysis project is enhanced with the use of Convolutional Neural Networks (CNNs). By effectively detecting hierarchical patterns and correlations in sequences of pollutant concentrations, CNN improves our comprehension of spatial dependencies, which are vital for making precise predictions. This study aims to enhance the accuracy of air pollution predictions by incorporating CNN into deep learning models. CNN will supplement the temporal characteristics recorded by recurrent models.

```
X_train = X_train.reshape(-1, X_train.shape[1],1)
X_test = X_test.reshape(-1, X_test.shape[1],1)

Y_train=to_categorical(y_train)
Y_test=to_categorical(y_test)
```

```
def CNN():

    cnnmodel = Sequential()
    cnnmodel.add(Conv1D(filters=128, kernel_size=2, activation='relu',input_shap
    cnnmodel.add(MaxPooling1D(pool_size=2))
    cnnmodel.add(Dropout(rate=0.2))
    cnnmodel.add(Flatten())
    cnnmodel.add(Dense(3, activation='softmax'))
    cnnmodel.compile(optimizer='adam', loss='categorical_crossentropy',metrics=[
    cnnmodel.summary()
    return cnnmodel

cnnmodel = CNN()
```

CNN+LSTM: When it comes to air quality forecasting, the hybrid CNN+LSTM model is a powerful combination of spatial and temporal learning. Using convolutional neural networks (CNNs) for spatial feature extraction and long short-term memories (LSTMs)[19] for temporal dependency capture, this model provides a holistic view of complicated patterns in pollutant concentrations. The goal of the CNN+LSTM model is to provide a comprehensive picture of the dynamics of air quality by integrating the two architectures; this will help shed light on how spatial and temporal learning work together to improve urban forecast accuracy.

CNN + LSTM

Fig 7 CNN + LSTM

4. EXPERIMENTAL RESULTS

Accuracy: A test's accuracy is defined by how well it distinguishes between healthy and sick samples. We can determine a test's accuracy by calculating the percentage of reviewed instances with true positives and true negatives. If we express this mathematically, we get: Accuracy = TP + TN TP + TN + FP + FN.

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

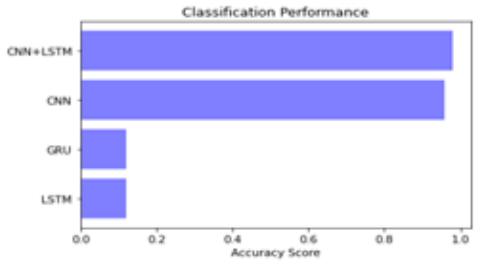


Fig 8 Accuracy Comparison Graph

F1-Score:

$$\textbf{F1 Score} = \frac{2}{\left(\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}\right)}$$

$$\textbf{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

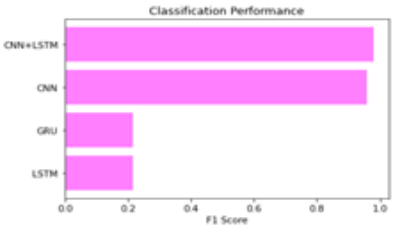


Fig 9 F1 Score Comparison Graph

Precision:

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

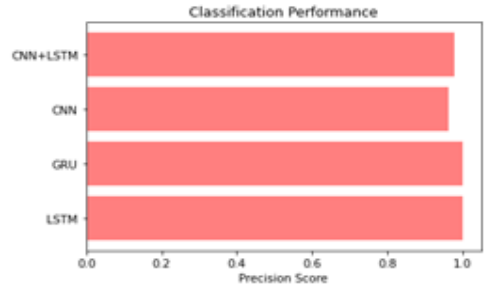


Fig 10 Precision Comparison Graph

Recall: The capacity of a model to detect all significant occurrences of a given class is measured by recall, a statistic in machine learning. The completeness of a model in capturing instances of a particular class is shown by the ratio of properly predicted positive observations to the total actual positives.

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

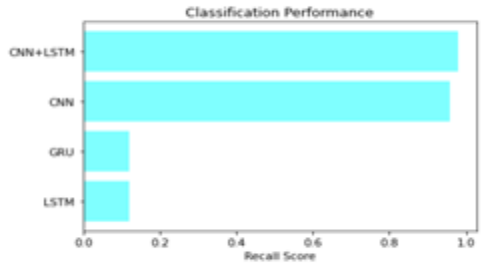


Fig 11 Recall Comparison Graph

ML Model	Accuracy	Precision	Recall	F1-Score
LSTM	0.120	1.000	0.120	0.214
GRU	0.120	1.000	0.120	0.214
EXTENSION CNN	0.960	0.964	0.960	0.960
EXTENSION CNN+LSTM	0.979	0.979	0.979	0.979

Fig 12 Evaluation Table

Fig 13 Home Page

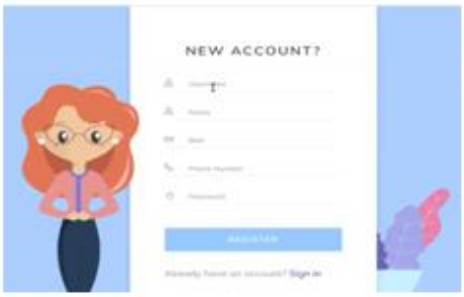


Fig 14 Registration Page

Fig 15 Login Page

Fig 16 Upload Input Values

Fig 17 Predicted Results

IV. CONCLUSION

Ultimately, protecting global health and environmental sustainability hinges on the precision with which urban air pollution levels can be predicted. This study successfully established the superiority of sophisticated deep learning models over standard statistical techniques in predicting air quality via the examination of multiple forecasting methodologies, including LSTM[19] and GRU [19]. Air pollution predictions have the potential to become even more accurate and versatile as a result of ongoing methodological refinement that prioritizes feature engineering, parameter optimization, and multi-step prediction methodologies. It is quite probable that these developments will enable health groups, environmental organizations, and legislators to make educated judgments and execute focused initiatives to lessen the negative effects of air pollution. These developments help improve the accuracy of air quality predictions, which in turn helps with larger public health and environmental sustainability programs, leading to better city environments over the long run.

VI. OUTLINE FOR THE REST OF THE WORK

We want to improve the performance of deep learning (DL) models and get closer to multi-step prediction in the future by using novel feature engineering approaches and hyperparameter tuning. We aim to provide more thorough insights into future levels of air pollutants by focusing on multi-step prediction, which extends the forecasting horizon beyond single time steps. To further improve the DL models' ability to grasp intricate patterns and connections, we want to investigate new feature engineering techniques that extract more relevant features from the data. To get the most out of your models in terms of predicted accuracy and resilience, you should optimize their hyperparameters. In addition to improving DL models' effectiveness in air quality forecasting, these developments will help us understand environmental dynamics better and make better decisions about how to reduce the negative impacts of air pollution.

REFERENCES

[1]. R. D. Brook, ‘‘Cardiovascular effects of air pollution,’’ Clin. Sci., vol. 115, no. 6, pp. 175–187, Sep. 2008.
[2]. M. Stafoggia and T. Bellander, ‘‘Short-term effects of air pollutants on daily mortality in the Stockholm county—A spatiotemporal analysis,’’ Environ. Res., vol. 188, Sep. 2020, Art. no. 109854.
[3]. WHO. Household Air Pollution and Health. Accessed: Dec. 29, 2022. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/ household-air-pollution-and-health

[4]. WHO. Air Quality and Health. Accessed: Jan. 10, 2023. [Online]. Available: https://www.who.int/teams/environment-climate-change-andhealth/air-quality-and-health/policy-progress/sustainable-developmentgoals-air-pollution
[5]. B. Paul and S. Louise. (2022). Air Quality: Policies, Proposals and Concerns—House of Commons Library. Accessed: Jan. 10, 2023. [Online]. Available: https://commonslibrary.parliament.uk/researchbriefings/cbp-9600/
[6]. P. J. Landrigan, ''Air pollution and health,'' Lancet Public Health, vol. 2, pp. e4–e5, Jan. 2017.
[7]. K. Abutalip, A. Al-Lahham, and A. El Saddik, ''Digital twin of atmospheric environment: Sensory data fusion for high-resolution PM2.5 estimation and action policies recommendation,'' IEEE Access, vol. 11, pp. 14448–14457, 2023.
[8]. J. Ma, Y. Ding, J. C. P. Cheng, F. Jiang, Y. Tan, V. J. L. Gan, and Z. Wan, ''Identification of high impact factors of air quality on a national scale using big data and machine learning techniques,'' J. Cleaner Prod., vol. 244, Jan. 2020, Art. no. 118955.
[9]. I. Manisalidis, E. Stavropoulou, A. Stavropoulos, and E. Bezirtzoglou, ''Environmental and health impacts of air pollution: A review,'' Frontiers Public Health, vol. 8, p. 14, 2020.
[10]. S. Ameer, M. A. Shah, A. Khan, H. Song, C. Maple, S. U. Islam, and M. N. Asghar, ''Comparative analysis of machine learning techniques for predicting air quality in smart cities,'' IEEE Access, vol. 7, pp. 128325–128338, 2019.
[11]. Q. Chen, W. Wang, F. Wu, S. De, R. Wang, B. Zhang, and X. Huang, ''A survey on an emerging area: Deep learning for smart city data,'' IEEE Trans. Emerg. Topics Comput. Intell., vol. 3, no. 5, pp. 392–410, Oct. 2019.
[12]. Y. Zhang, Y. Wang, M. Gao, Q. Ma, J. Zhao, R. Zhang, Q. Wang, and L. Huang, ''A predictive data feature exploration-based air quality prediction approach,'' IEEE Access, vol. 7, pp. 30732–30743, 2019.
[13]. Doreswamy, K. S. Harishkumar, Y. Km, and I. Gad, ''Forecasting air pollution particulate matter (PM2.5) using machine learning regression models,'' Proc. Comput. Sci., vol. 171, pp. 2057–2066, Jan. 2020.
[14]. Y.-S. Chang, H.-T. Chiao, S. Abimannan, Y.-P. Huang, Y.-T. Tsai, and K.-M. Lin, ''An LSTM-based aggregated model for air pollution forecasting,'' Atmos. Pollut. Res., vol. 11, no. 8, pp. 1451–1463, Aug. 2020.
[15]. J. Ma, Y. Ding, J. C. P. Cheng, F. Jiang, V. J. L. Gan, and Z. Xu, ''A lag-FLSTM deep learning network based on Bayesian optimization for multi-sequential-variant PM2.5 prediction,'' Sustain. Cities Soc., vol. 60, Sep. 2020, Art. no. 102237.
[16]. R. Yan, J. Liao, J. Yang, W. Sun, M. Nong, and F. Li, ''Multi-hour and multi-site air quality index forecasting in Beijing using CNN, LSTM, CNN-LSTM, and spatiotemporal clustering,'' Expert Syst. Appl., vol. 169, May 2021, Art. no. 114513.
[17]. S. Du, T. Li, Y. Yang, and S. Horng, ''Deep air quality forecasting using hybrid deep learning framework,'' IEEE Trans. Knowl. Data Eng., vol. 33, no. 6, pp. 2412–2424, Jun. 2021.
[18]. N. Zaini, L. W. Ean, A. N. Ahmed, M. A. Malek, and M. F. Chow, ''PM2.5 forecasting for an urban area based on deep learning and decomposition method,'' Sci. Rep., vol. 12, no. 1, p. 17565, Oct. 2022.
[19]. B. Wang, W. Kong, H. Guan, and N. N. Xiong, ''Air quality forecasting based on gated recurrent long short term memory model in Internet of Things,'' IEEE Access, vol. 7, pp. 69524–69534, 2019.
[20]. Q. Tao, F. Liu, Y. Li, and D. Sidorov, ''Air pollution forecasting using a deep learning model based on 1D convnets and bidirectional GRU,'' IEEE Access, vol. 7, pp. 76690–76698, 2019.
[21]. A. Bekkar, B. Hssina, S. Douzi, and K. Douzi, ''Air quality forecasting using decision trees algorithms,'' in Proc. 2nd Int. Conf. Innov. Res. Appl. Sci., Eng. Technol. (IRASET), Mar. 2022, pp. 1–4.
[22]. A. Bekkar, B. Hssina, S. Douzi, and K. Douzi, ''Air-pollution prediction in smart city, deep learning approach,'' J. Big Data, vol. 8, no. 1, pp. 1–21, Dec. 2021.
[23]. Y. Liang, Y. Xia, S. Ke, Y. Wang, Q. Wen, J. Zhang, Y. Zheng, and R. Zimmermann, ''Airformer: Predicting nationwide air quality in China with transformers,'' Nov. 2022, arXiv:2211.15979.
[24]. S. Wang, Y. Li, J. Zhang, Q. Meng, L. Meng, and F. Gao, ''PM2.5-GNN: A domain knowledge enhanced graph neural network for PM2.5 forecasting,'' in Proc. 28th Int. Conf. Adv. Geographic Inf. Syst. (SIGSPATIAL), Nov. 2020, pp. 163–166.
[25]. G. Box, G. Jenkins, G. Reinsel, and G. Ljung, Time Series Analysis: Forecasting and Control (Wiley Series in Probability and Statistics). Hoboken, NJ, USA: Wiley, 2015. [Online]. Available: https://books.google.co.uk/books?id=rNt5CgAAQBAJ
[26]. Z. Yang, K. Wang, J. Li, Y. Huang, and Y. Zhang, ''TS-RNN: Text steganalysis based on recurrent neural networks,'' IEEE Signal Process. Lett., vol. 26, no. 12, pp. 1743–1747, Dec. 2019.
[27]. G. Gelly and J. Gauvain, ''Optimization of RNN-based speech activity detection,'' IEEE/ACM Trans. Audio, Speech, Language Process., vol. 26, no. 3, pp. 646–656, Mar. 2018.
[28]. S. Hochreiter and J. Schmidhuber, ''Long short-term memory,'' Neural Comput., vol. 9, no. 8, pp. 1735–1780, Nov. 1997.
[29]. K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, ''Learning phrase representations using RNN encoder–decoder for statistical machine translation,'' in Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP). Doha, Qatar: Association for Computational Linguistics, Oct. 2014, pp. 1724–1734.
[30]. Northern Ireland Air. Download Air Quality Data—Northern Ireland. Accessed: Dec. 1, 2022. [Online]. Available: https://www.airqualityni. co.uk/data
[31]. L. Li, K. Jamieson, G. DeSalvo, A. Rostamizadeh, and A. Talwalkar, ''Hyperband: A novel bandit-based approach to hyperparameter optimization,'' J. Mach. Learn. Res., vol. 18, no. 1, pp. 6765–6816, Jan. 2017.

Dataset Link:

[32]. https://www.kaggle.com/datasets/cpluzshrijayan/air-quality-prediction-harbor