Developing a Machine Learning model to Analyze the effect of Nitrogen Dioxide (NO_2) on Air Pollution in WHO countries

YGGD Ranasinghe[1]
Department of Languages
Faculty of Management, Social sciences and Humanities
General Sir John Kotelawala Defence University
Rathmalana, Sri Lanka
40-adc-0025@kdu.ac.lk

DLS Nadavi[2]
Department of Languages
Faculty of Management, Social sciences and Humanities
General Sir John Kotelawala Defence University
Rathmalana, Sri Lanka
40-adc-0028@kdu.ac.lk

AGAUS Gunasekara[3]
Department of Languages
Faculty of Management, Social sciences and Humanities
General Sir John Kotelawala Defence University
Rathmalana, Sri Lanka
40-adc-0034@kdu.ac.lk

MLBTS Perera[4]
Department of Languages
Faculty of Management, Social sciences and Humanities
General Sir John Kotelawala Defence University
Rathmalana, Sri Lanka
40-adc-0047@kdu.ac.lk

***Abstract* - This research analyzes the effect of Nitrogen Dioxide (NO_2) pollution on some of World Health Organization (WHO) member countries from 2010 to 2019 through Machine Learning model and WHO Air Quality Database 2022 information. NO_2 operates as a primary air contaminant generated by industrial facilities, traffic systems and electric power plants thus causing respiratory disorders while damaging the environment and enabling Ozone formation and $PM_{2.5}$ particle generation. Applying a Random Forest Regressor model produces exceptional predictive ability (MSE: 0.00545, R^2: 0.99996) when evaluating NO_2 concentrations and their changes in the study. The study shows that NO_2 pollution decreased during the period but some countries still display different pollution levels as legal requirements and technological innovations reduced NO_2 emissions. The Countries Bahrain and Lebanon hold the highest NO_2 concentration levels so both nations need specialized air quality intervention measures. The analysis faces three main constraints which stem from using secondary data while the model could show signs of overfitting and omits information about the COVID-19 pandemic effects. Future recommendations include using data from different sources as well as understanding pollutant effects on each other and developing models that function across varied conditions.**

Keywords - Nitrogen Dioxide (NO_2), Air Pollution, WHO countries

I. INTRODUCTION

The environment faces severe pollution issues that negatively impact everyone from human beings to natural ecosystems and the atmosphere. The main air pollutant NO_2 comes from exhaust sources, industrial activities and power stations. NO_2 participates in producing hazardous ozone molecules and small $PM_{2.5}$ particles which negatively affect human health. NO_2 exposure brings about respiratory symptoms while making asthma worse and making people more vulnerable to respiratory infections. The pollutant NO_2 allows the development of acid rain and creates water body eutrophication which results in additional ecological damage. This paper's objective is to analyze the effect of Nitrogen Dioxide (NO_2) on Air Pollution in WHO countries using a Machine Learning Model. In so doing, the study aims to evaluate NO_2 concentrations and their impact on air quality trends from 2010 to 2019 based on data from the World Health Organization (WHO) Air Quality Database 2022. Knowledge of these patterns is highly important for the design of more effective preventive measures that may reduce the effect of NO2 on Air Pollution.

II. METHODOLOGY

In this research, the methodology mainly emphasizes Secondary data. This data has been collected from a source that is easily accessible to the public, having been collected by the official World Health Organization (WHO) website and is titled "Air Quality Database: Update 2022" (Available on : Air quality database 2022). Since this data set covers several years, it does aid in patterns of the variation of No2 impact in some WHO countries between 2010 to 2019. It captures ground measurements of the annual mean concentrations of Nitrogen Dioxide (NO2), particulate matter of a diameter equal to or smaller than 10 μm (PM10) or equal to or smaller than 2.5 μm (PM2.5) which aim at representing an average for the city or town, rather than for individual stations. The database contains 32,191 entries spread across 15 columns with fields such as WHO Region, ISO3, WHO Country Name, City or Locality, Measurement Year, $PM_{2.5}$ (μg/m³), PM_{10} (μg/m³), NO_2 (μg/m³), $PM_{2.5}$ Temporal Coverage (%), PM_{10} Temporal Coverage (%), NO_2 Temporal Coverage (%), Reference, Number and Type of Monitoring Stations, Version of the Database, and Status. As for the analysis, the raw data was cleaned using Python Programming Language on Google Colab. Cleaning of this data will entail screening for missing values, null values, and duplicity. For the analysis, Python helped to build the Machine Learning model and visualize some graphical charts such as line graphs or bar charts according to the cleaned data set. Descriptive analysis will also be carried out to investigate any trends or differences over time in the cases.

III. LITRERATURE REVIEW

A. Introduction to Air Pollution and Nitrogen Dioxide (NO_2)

According to Huang et al., 2021, air pollution stands as a major environmental health danger while nitrogen dioxide (NO_2) emerges as a significant pollutant released from traffic. Air pollution shows different chemical composition patterns which depend on geographic positioning weather conditions population distribution and human-made activities yet industrial exhausts remain the primary source of environmental destruction and climate change (Environmental Pollutants PM2.5, PM10, Carbon Monoxide (CO), Nitrogen Dioxide (NO2), Sulfur Dioxide (SO2), and Ozone (O3) Impair Human Cognitive Functions, n.d.). The emissions generated by manufacturing combined with power generation and chemical production release dangerous particles such as PM2.5 and PM10 alongside CO, SO_2, O_3, VOCs and NO_2 (Environmental Pollutants PM2.5, PM10, Carbon Monoxide (CO), Nitrogen Dioxide (NO2), Sulfur Dioxide (SO2), and Ozone (O3) Impair Human Cognitive Functions, n.d.). The presence of NO_2 poses serious risks for urban residents because transportation and stationary fuel combustion act as their dominant emission sources(Restrepo, 2021). The measurement of NO_2 concentrations is a difficult task because its concentrations vary with the amount of traffic and the positioning of monitoring devices (Restrepo, 2021). The short atmospheric lifetime and spatial and temporal variations distinguish NO_2 as one of the criteria for air pollution components (Larkin et al., 2023). According to Kashtan et al., 2024, the studies show that NO_2 pollution levels from gas and propane stove combustion interfere with indoor air quality through harmful emissions that surpass health standards Environmental and human health evaluation reveals NO_2 pollution exists at wide-reaching levels throughout the atmosphere.

B. Health and Environmental imppacts of NO_2

Multiple studies prove that traffic-originating pollutant Nitrogen dioxide (NO_2) creates substantial effects on health and the environment. The study by Huang et al. (2021) establishes that NO_2 exposure leads to health problems that primarily affect the respiratory system. Exposure to NO_2 at any duration creates respiratory issues and worsens asthma conditions thus becoming a significant health problem for the public (Restrepo, 2021). The chemical reaction between NO_2 molecules contributes vitally to ozone layer development at ground level (O_3) thereby worsening respiratory diseases (Restrepo, 2021). Long-term NO_2 exposure has been established as a factor increasing chronic obstructive pulmonary disease (COPD) incidence rates and mortality rates and asthma incidence in children and lung cancer and preterm birth and diabetes mellitus occurrence and mortality (Kashtan et al., 2024). Scientific studies have established that exposure to NO_2 along with its byproduct N_2O_3 induces DNA damage resulting in breast cancer development because of these toxic substances' mutagenic and carcinogenic properties (Keramatinia et al., 2016). Research explores the connection between air pollution and breast cancer through the specific proxy indicator provided by NO_2 because of its special quality as an exposure marker (Keramatinia et al., 2016). The findings emphasize the importance of measuring NO_2 along with assessing its health effects including those attributable to gas and propane stoves to recognize and decrease public health risks (Kashtan et al., 2024).

C. Previous Studies and Models on NO_2 and Air Pollution

Multiple research methods have been used by previous investigators to both measure and model exposure levels of NO_2 and air pollution. The research by Environmental Pollutants PM2.5, PM10, Carbon Monoxide (CO), Nitrogen Dioxide (NO2), Sulfur Dioxide (SO2), and Ozone (O3) Impair Human Cognitive Functions (n.d.) collected 24 hours of air pollutant data through integrated sampling systems on hourly bases. The specialized MP101M (2.5) and MP101M (10) devices combined with CO12e and AC32e and AF22e and O342e monitors were used to measure PM2.5 and PM10 and CO and O_3 and NO_2 and SO_2 (Environmental Pollutants PM2.5, PM10, Carbon Monoxide (CO), Nitrogen Dioxide (NO2), Sulfur Dioxide (SO2), and Ozone (O3) Impair Human Cognitive Functions, n.d.). The study acquired data from the National Center for Environmental Compliance (NCEC) in Riyadh, Saudi Arabia together with daily records obtained from the Air Quality Index (AQI) and the established methods delivered complete pollution data suitable for measuring pollutant activity and its consequences (Environmental Pollutants PM2.5, PM10, Carbon Monoxide (CO), Nitrogen Dioxide (NO2), Sulfur Dioxide (SO2), and Ozone (O3) Impair Human Cognitive Functions, n.d.). Kashtan et al. (2024) created a population-level model to analyze NO_2 exposure from gas and propane stoves by using the National Institute of Standards and Technology's (NIST) CONTAM multizone indoor air quality model. The study used direct field measurements from more than 70 homes to determine NO_2 emission rates together with information about residence size ventilation habits and range hood use (Kashtan et al., 2024). The CONTAM model received validation through its successful comparison of predicted NO_2 concentrations against actual measurements within 18 test houses making it possible to estimate both time-period and group-specific NO_2 exposure accurately(Kashtan et al., 2024). Scientists evaluated pediatric asthma and adult mortality linked to NO_2 emissions from stove combustion by analyzing the research findings which showed the public health threats from indoor air pollution(Keramatinia et al., 2016).

D. Limitations in NO_2 Monitoring and Anlysis

Various studies show that limitations exist in both the process of monitoring and analyzing nitrogen dioxide (NO_2) levels.

The COVID-19 pandemic created major disruptions to research activities by causing the closure of educational institutions together with limitations on public activities during lockdowns(Environmental Pollutants PM2.5, PM10, Carbon Monoxide (CO), Nitrogen Dioxide (NO2), Sulfur Dioxide (SO2), and Ozone (O3) Impair Human Cognitive Functions, n.d.). Actual pollutant impact on cognitive functions decreased because of reduced industrial production lower vehicle traffic and decreased power usage which reduced pollution from NO_2 and other pollutants (Environmental Pollutants PM2.5, PM10, Carbon Monoxide (CO), Nitrogen Dioxide (NO2), Sulfur Dioxide (SO2), and Ozone (O3) Impair Human Cognitive Functions, n.d.). Keramatinia et al. (2016) recognized two main weaknesses in their research which included their failure to use individually adjusted measures and their use of ecological correlations that might not capture personal risk adequately. The research included few individual-based studies because the authors limited their investigation to traditional white literature (Keramatinia et al., 2016). The NO_2 model predictions from Larkin et al. (2023) rely on accurate OMI measurements yet they did not assess the quality of air monitor sampling since OpenAQ records had insufficient sampling details. Kashtan et al. (2024) described several unknown variables that affect their modeling process from NO_2 emissions to indoor concentration calculations to person behavior assumptions and disease burden conversion. Multiple obstacles diminish the ability to precisely determine the amounts of NO_2 and their effects (Kashtan et al., 2024).

E. Conclusion

According to Huang et al. (2021), long-term NO2 exposures linked with traffic-related pollution present substantial health hazards that lead to death risks for overall mortalities and cardiovascular diseases and lung conditions while appearing unaffected by coexisting pollutants. Science shows that environmental pollutants specifically NO2 impair various cognitive abilities of students who attend schools near heavily polluted areas leading to diminished attention function decision-making capacity and thinking efficiency (Environmental Pollutants PM2.5, PM10, Carbon Monoxide (CO), Nitrogen Dioxide (NO2), Sulfur Dioxide (SO2), and Ozone (O3) Impair Human Cognitive Functions, n.d.). NO2 exposure shows a slight but substantial relation to breast cancer occurrence at the individual level along with more significant links observed at the aggregate level according to Keramatinia et al. (2016). According to Restrepo (2021), the adoption of transportation sector policies that lower fossil fuel use requires improvements in fuel efficiency and cleaner fuel standards with electric vehicle adoption and promotion of walking and cycling. The COVID-19 pandemic offers useful opportunities to redirect street areas in urban parts and elevate public transit standards because these steps help decrease pollutant emissions and make the air cleaner (Restrepo, 2021). City authorities need to assess these measures thoroughly along with maintaining them to reach environmental and social sustainability targets that guarantee better health outcomes for urban residents in the future.(Restrepo, 2021)

IV. RESULTS AND DISCUSSSION

The trained Random Forest Regressor with 100 estimators and a set random state value of 42 presents outstanding results according to evaluation metrics. Fig.1 illustrates that the model predicts actual values with high accuracy because its Mean Squared Error (MSE) reaches 0.00545. MSE demonstrates the average quantified difference between predicted values and actual observations. The low MSE value signifies the high accuracy of the model's predictions since the measurement result is exceptionally small. And the model performance is exceptional because its R^2 score reaches 0.99996 which demonstrates it detects almost all target variable variation. The closeness of an R^2 score to one indicates that the model achieves almost perfect data fit by accurately detecting underlying patterns. The Random Forest Regressor demonstrates effective performance for this task because it applies ensemble methodology to reduce errors and increase its predictive capability. The model shows exceptional predictive accuracy because of its minimal MSE and perfect R^2 statistical measurement. The Random Forest Regressor demonstrates reliable performance as a robust forecasting model due to its current configuration for this particular dataset. The model's high-performance metrics require proper assessment of overfitting potential as well as tests for generalization ability on new data.

According to Fig.2, all cross-validation R^2 scores produced exceptional results showing model performance across each fold with scores between 0.99503289 and 0.99993652. The model demonstrates superior predictive capabilities because it explains almost all target variable variance according to the high values. The model exhibits reliable characteristics along with strong consistency because its mean cross-validation R^2 score of 0.9988182117861705 demonstrates near-perfect predictive ability on different data subsets. NO2 (µg/m3) emerges as the key predictive variable according to the importance scores which reach 0.999640. The Model shows NO2 (µg/m3) as its strongest predictor because it has maximum influence on predictive results. Most of the examined features such as PM10 (µg/m3) and PM2.5 (µg/m3) and Measurement Year along with temporal metrics for PM10, NO2, and PM2.5 have tiny predictive values according to the analysis with scores between 0.000026 and 0.000113. The predictive power of the model primarily relies on NO2 rather than the other less significant features. The model achieves remarkable predictive success which is confirmed by the high scores obtained from cross-validation R^2 calculations. The model heavily depends on NO2 measurements for its predictions which surpass the importance of all other features by a wide margin. The overfitting to NO2 as well as the meaningless predictive power of other features within this dataset could be two possible explanations for this result. A deeper examination of the relationship between NO2 data with target variables should be combined with possible feature selection processes to optimize model generalizability.

Fig.3 shows statistical details consisting of mean, median, standard deviation, minimum value, and maximum value with range calculation delivered across some WHO countries between 2010 – 2020 using the cleaned dataset. These statistics offer insights into the distribution and variability of the data for each country. Statistical measurements of this metric demonstrate wide intercountry variation according to the dataset analysis. The countries of Bahrain and Lebanon show predominantly higher measurement levels based on their mean values of 51.63 and 53.00 respectively. The results from Estonia and Iceland indicate lower measurement figures through their mean values which stand at 8.79 and 7.94 respectively. The spread of data across different countries is measured through standard deviations while these measurement variations differ significantly between countries. The measurement variability in Germany is high with a standard deviation of 14.76 but Costa Rica, Croatia, Greece, Ireland, Jordan, Latvia, and Singapore demonstrate no variation because their data points remain constant. The range shows data distribution diversity through its calculation of maximum value subtraction from the minimum value. Bangladesh and India record wide measurement dispersion because their ranges reach 67.83 and 68.67, respectively. The measurement analysis from Costa Rica as well as Croatia Greece, Ireland, Jordan, Latvia, and Singapore shows no variation since their standard deviation amounts to zero. The statistical information from certain nations including Denmark and Singapore shows that their median values match their mean because their data exhibits symmetrical dispersion. The measured values in Slovenia show median results (1.99) that are much lower than mean values (14.86) because of a distribution pattern that includes outliers. The statistical values in the dataset document widespread variability in the metric between countries due to some exhibiting consistent performance while others demonstrate large variations. The collected data holds potential value for recognizing patterns and exceptions as well as detecting trends according to the data conditions. The data shows that countries having both high performance levels and minimal variation reflect stable results but nations with great performance swing reflect either significant transformations or unreliable measurements over time. Extended research should identify the attributes that lead to these observed value variances between states.

Fig.4 presents data about Average Nitrogen Dioxide (NO_2) concentrations from 2010 to 2020 using micrograms per cubic meter (µg/m³) as the measurement units according to the cleaned dataset. The average NO_2 data reveals a decreasing pattern throughout these ten years since 2010 when the NO_2 value reached 26 µg/m³ until 2020 when it measured at 18.3 µg/m³. The declining values demonstrate that environmental conditions have improved most likely because of stricter regulations technological progress and shifting industrial and transport sector operations. The downward trend of NO_2 emissions showed temporary variations between specific years because environmental conditions and emission source dynamics appeared to change. Public health together with environmental quality benefits from NO_2 level reductions since NO_2 appears as a dangerous pollutant linked to respiratory problems and environmental damage. The study should examine particular elements causing these patterns along with evaluating air pollution control methodologies.

Fig.5 contains a list of top 10 countries with their measured concentrations of NO_2 presented in micrograms per cubic meter (µg/m³) according to the cleaned dataset. Among the nations displayed in the chart Lebanon stands as the highest contributor of NO_2 pollution. Bahrain follows as the second country in the ranking then comes Latvia next to Costa Rica, Turkey, Kuwait, and Qatar then Mexico followed by the Republic of Korea and Bangladesh to finish the list. Three nations including Lebanon, Bahrain, and Kuwait occupy top positions in the list indicating their substantial industrial zones and urbanized areas and transportation networks resulting in elevated NO_2 pollution levels. High concentrations of NO_2 extend across every part of the world since both developed nations and developing ones participate in the study. These findings emphasize the necessity to direct air quality management strategies and pollution control operations within these countries because they face serious health risks from elevated NO_2 concentrations. Additional studies must investigate what generates NO_2 pollution in each country together with analyzing the current policies that aim to minimize emissions levels.

Fig.6 presents data on NO2 Quantity over Years in the selected country to display Nitrogen Dioxide (NO_2) concentration measurements expressed in micrograms per cubic meter (µg/m³) according to the cleaned dataset. In here, for an example this line chart represents the " NO2 Quantity over Years in the Bangladesh". NO_2 air pollution measurements exhibit an overall upward pattern from 28.5 µg/m³ in 2013 until they reached 38 µg/m³ in 2015. The increase of NO_2 levels across these two years reflects worsening air quality mainly due to enhanced industrial operations and urban population shifts together with vehicle exhaust. Air pollution in Bangladesh continues to worsen due to the steady upward trend of measurements which will create major health risks for both people and environmental systems. Future air quality management and pollution control policies must be implemented immediately because the data strongly demonstrates the growing NO_2 contamination. Researchers should investigate the particular emission sources of NO_2 and evaluate the effectiveness of current policies that aim to reduce air pollution throughout Bangladesh.

The code in Fig.7 checks the availability of data for the selected countries in the cleaned dataset to build the Fig. 6 chart. If there is no data available for a relevant country, a message will be displayed on the screen. For an example, there is no available data of Vietnam in the cleaned dataset. Therefore, it shows a message about the unavailability of data.

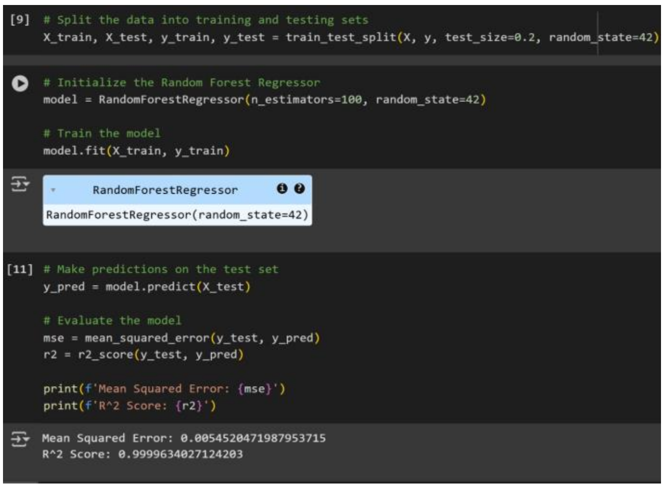

Fig.1. Random Forest Regressor
Source: "Air Quality Database: Update 2022."
(Available at: Air quality database 2022)

```
[19] from sklearn.model_selection import cross_val_score

     # Perform cross-validation
     cv_scores = cross_val_score(model, X, y, cv=5, scoring='r2')

     print(f'Cross-Validation R² Scores: {cv_scores}')
     print(f'Mean Cross-Validation R² Score: {cv_scores.mean()}')

     # Get feature importances
     importances = model.feature_importances_  # Correct attribute name

     # Create a DataFrame to display feature importances
     feature_importance_df = pd.DataFrame({
         'Feature': X.columns,
         'Importance': importances
     }).sort_values(by='Importance', ascending=False)

     print(feature_importance_df)
```

```
Cross-Validation R² Scores: [0.99993652 0.99955685 0.99970667 0.99985813 0.99503289]
Mean Cross-Validation R² Score: 0.9988182117861705
                      Feature  Importance
3                NO2 (µg/m3)    0.999640
2               PM10 (µg/m3)    0.000113
1              PM2.5 (µg/m3)    0.000092
0           Measurement Year    0.000050
5  PM10 temporal coverage (%)    0.000046
6   NO2 temporal coverage (%)    0.000033
4  PM25 temporal coverage (%)    0.000026
```

Fig.2. Cross-Validation R^2 scores and Mean Cross Validation R^2 score
Source: "Air Quality Database: Update 2022."
(Available at: Air quality database 2022)

	WHO Country Name	mean	median	std	min	max	range
0	Australia	8.979787	8.270	4.130846	3.50	20.46	16.96
1	Austria	27.486364	31.360	11.005448	7.37	39.61	32.24
2	Bahrain	51.625000	51.625	5.932626	47.43	55.82	8.39
3	Bangladesh	33.875676	35.700	17.394212	5.87	73.70	67.83
4	Belgium	25.269474	28.660	9.131774	8.20	40.02	31.82
5	Brazil	24.506250	19.975	12.739226	7.55	45.34	37.79
6	Bulgaria	25.191429	28.910	12.593077	1.17	40.93	39.76
7	Canada	13.574156	13.160	6.145999	2.00	27.00	25.00
8	Chile	16.460000	15.570	9.728876	7.10	43.77	36.67
9	Colombia	26.040000	24.050	10.282758	13.10	42.48	29.38
10	Costa Rica	40.710000	40.710	NaN	40.71	40.71	0.00
11	Croatia	11.760000	11.760	NaN	11.76	11.76	0.00
12	Cyprus	17.716000	17.230	10.828468	3.02	33.60	30.58
13	Czechia	18.072132	17.000	6.377071	2.54	38.72	36.18
14	Denmark	33.570000	33.570	5.317443	29.81	37.33	7.52
15	Estonia	8.793684	9.295	4.507057	1.82	17.14	15.32
16	Finland	20.992000	20.925	8.384330	5.66	35.73	30.07
17	France	26.910500	26.355	9.161944	2.99	49.09	46.10
18	Germany	29.848846	30.870	14.755895	5.91	69.84	63.93
19	Greece	21.040000	21.040	NaN	21.04	21.04	0.00
20	Hungary	27.607500	28.925	6.934293	18.19	34.39	16.20
21	Iceland	7.940000	7.940	6.392245	3.42	12.46	9.04
22	India	21.043047	19.000	10.911389	5.00	73.67	68.67
23	Ireland	28.890000	28.890	NaN	28.89	28.89	0.00
24	Italy	31.876857	31.790	12.524409	3.32	66.59	63.27
25	Jordan	27.040000	27.040	NaN	27.04	27.04	0.00
26	Kuwait	35.970000	30.460	16.420680	20.39	71.99	51.60
27	Latvia	41.660000	41.660	NaN	41.66	41.66	0.00
28	Lebanon	53.000000	53.000	8.485281	47.00	59.00	12.00
29	Lithuania	19.557333	21.060	3.713188	13.33	24.50	11.17
30	Luxembourg	30.723333	28.830	16.720590	15.03	48.31	33.28
31	Malta	17.894000	13.700	15.956243	3.02	34.85	31.83
32	Mexico	35.073056	29.810	15.026164	23.57	83.94	60.37
33	Netherlands	29.660625	34.870	9.855485	10.32	41.03	30.71
34	Norway	16.767500	17.235	14.294530	0.35	51.57	51.22
35	Poland	20.660370	21.010	10.478133	3.39	50.45	47.06
36	Portugal	17.572500	12.480	13.371794	2.69	37.56	34.87
37	Qatar	35.200000	30.000	8.700575	28.00	47.00	19.00
38	Republic of Korea	34.669412	34.780	9.700730	15.04	60.16	45.12
39	Romania	27.323000	24.430	8.801751	16.77	44.84	28.07
40	Senegal	24.676667	21.215	8.421588	15.68	36.16	20.48
41	Singapore	26.000000	26.000	NaN	26.00	26.00	0.00
42	Slovakia	26.703333	32.920	12.554323	8.08	37.97	29.89
43	Slovenia	14.856000	1.990	17.874649	1.59	34.70	33.11
44	South Africa	21.407308	18.505	13.104694	3.70	70.69	66.99
45	Spain	14.578226	10.995	10.670379	1.51	46.94	45.43
46	Sweden	10.608571	7.410	10.296860	0.48	37.57	37.09
47	Thailand	20.014583	21.000	9.383786	6.00	36.83	30.83
48	Trinidad and Tobago	12.400000	12.400	5.444722	8.55	16.25	7.70
49	Turkey	40.660000	39.810	18.196632	13.90	64.24	50.34
50	United Kingdom	32.685625	33.460	12.719567	8.68	58.80	50.12

Fig.3. Statistics for NO_2 levels
Source: "Air Quality Database: Update 2022."
(Available at: Air quality database 2022)

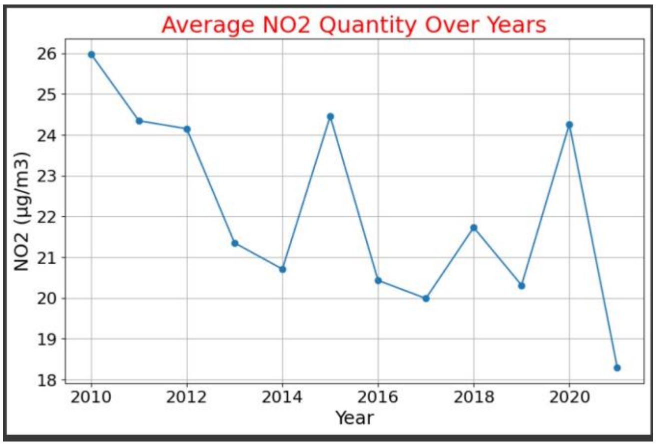

Fig.4. Average NO_2 Quantity over years
Source: "Air Quality Database: Update 2022."
(Available at: Air quality database 2022)

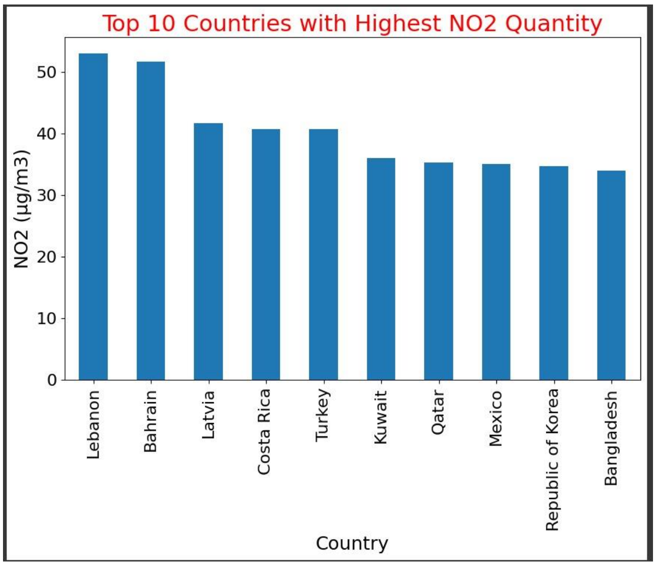

Fig.5. Top 10 countries with Highest NO2 Quantity
Source: "Air Quality Database: Update 2022."
(Available at: Air quality database 2022)

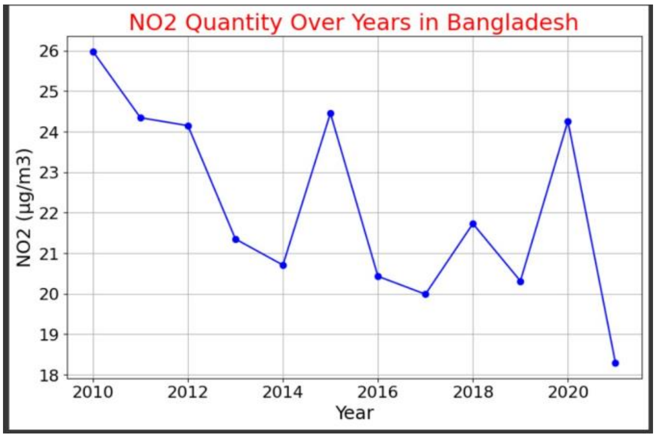

Fig.6. NO_2 Quantity over years in Bangladesh
Source: "Air Quality Database: Update 2022."
(Available at: Air quality database 2022)

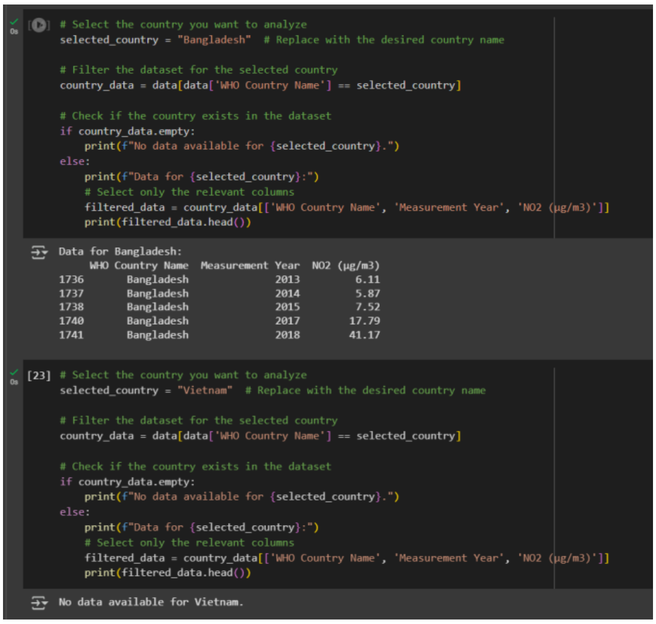

Fig.7. Check the availability of data
Source: "Air Quality Database: Update 2022."
(Available at: Air quality database 2022)

V. LIMITATIONS

Extra constraints within this study exist that researchers need to recognize before understanding the obtained results. This study faces limitations because it uses second-hand data from the WHO Air Quality Database 2022 which depends on the quality of specific monitoring techniques along with station positions and reporting criteria of each country. The reliability of obtained results might be influenced by differences in these factors. The research fails to examine the combined effects of pollutants because it distinguishes NO_2 concentrations from $PM_{2.5}$ and ozone despite their potential combined impact on air quality and health status. The narrow research boundaries limit how complete the study's results become. The machine learning model performs well when predicting but its dependence on NO_2 data combined with weak predictive power of additional features creates challenges in overfitting and universal prediction capabilities. The 2010–2019 study duration fails to consider the impacts of the COVID-19 pandemic because it occurred after this period thereby omitting the major changes in air pollution levels caused by lockdowns and decreased industrial operations. The study omits an analysis of NO_2 trends which would significantly limit researchers' ability to determine complete air quality implications from recent global events.

VI. RECOMMENDATIONS

Multiple recommendations exist to improve this study by addressing existing limitations while advancing the identified findings. Future research needs to integrate several additional monitoring sources that combine satellite data with ground platforms to enhance the accuracy along with wide-area coverage when measuring NO_2 concentrations. The research would achieve a better overview of NO_2 spatial patterns and emission origins by utilizing additional data collection methods. The research must investigate NO_2 interactions with $PM_{2.5}$ alongside ozone to develop a complete understanding of the environmental and health consequences of multi-pollutant exposure. The approach that considers all aspects would reveal expansive details about air pollution processes. The machine learning model requires improvement to minimize its overfitting problem while enhancing generalization for new data. Attention from policymakers should focus on establishing robust air quality benchmarks in Lebanon and Bahrain because these nations have elevated NO_2 readings along with other countries. The implementation of public awareness campaigns must occur to inform communities about the medical risks from NO_2 exposure together with the significance of implementing cleaner air. Such concerted actions will lead communities toward a better environment and enhanced public wellness.

VII. CONCLUSION

The research establishes NO_2 as a vital element that causes considerable air pollution problems which worsens public wellness alongside environmental conditions. The examination points to a wide distribution of NO_2 intensity measurements within WHO member states between 2010 and 2019 as a result of changed regulatory parameters as well as developing technologies and industrial production along with transportation methods. The continued high levels of NO_2 pollution endanger the public health of Lebanon and Bahrain because these countries need specific intervention strategies. The machine learning model in this research examined NO_2 effects across WHO countries from 2010 to 2020 through analysis and prediction insights. Further research should focus on resolving the identified data quality and model generalization issues to improve future work. The combination of introduced measures enables nations to decrease NO_2 emissions which leads to better air quality standards while protecting public health thus creating a more sustainable environment for everyone.

ACKNOWLEDGMENT

This research paper achieved success because of the deep and sincere support from all contributors. At the beginning we express deep appreciation to our lecturers Dr. Charith Silva and Ms. BCT Wickramasinghe for guiding us by providing essential direction and consistent backing as well as constructive feedback throughout research. The group members deserve our most sincere gratitude because they dedicated themselves to the project through teamwork and diligent work. The completion of this research required the devoted work of every person who participated together.

REFERENCES

[1] "Air Quality Database: Update 2022." (Available at: Air quality database 2022)

[2] Huang, S., Li, H., Wang, M., Qian, Y., Steenland, K., Caudle, W.M., Liu, Y., Sarnat, J., Papatheodorou, S. and Shi, L., 2021. Long-term exposure to nitrogen dioxide and mortality: A systematic review and meta-analysis. *Science of The Total Environment*, *776*, p.145968.

https://www.sciencedirect.com/science/article/abs/pii/S0048969721010354

[3] Kashtan, Y., Nicholson, M., Finnegan, C.J., Ouyang, Z., Garg, A., Lebel, E.D., Rowland, S.T., Michanowicz, D.R., Herrera, J., Nadeau, K.C. and Jackson, R.B., 2024. Nitrogen dioxide exposure, health outcomes, and associated demographic disparities due to gas and propane combustion by US stoves. *Science Advances*, *10*(18), p.eadm8680. https://www.science.org/doi/full/10.1126/sciadv.adm8680

[4] Keramatinia, A., Hassanipour, S., Nazarzadeh, M., Wurtz, M., Monfared, A.B., Khayyamzadeh, M., Bidel, Z., Mhrvar, N. and Mosavi-Jarrahi, A., 2016. Correlation between nitrogen dioxide as an air pollution indicator and breast cancer: a systematic review and meta-analysis. *Asian Pacific Journal of Cancer Prevention*, *17*(1), pp.419-424. https://koreascience.kr/article/JAKO201608160154008.page

[5] Larkin, A., Anenberg, S., Goldberg, D.L., Mohegh, A., Brauer, M. and Hystad, P., 2023. A global spatial-temporal land use regression model for nitrogen dioxide air pollution. *Frontiers in Environmental Science*, *11*, p.1125979. https://doi.org/10.3389/fenvs.2023.1125979

[6] *Meo, S.A., Salih, M.A., Al-Hussain, F., Alkhalifah, J.M., Meo, A.S. and Akram, A., 2024. Environmental pollutants PM2. 5, PM10, carbon monoxide (CO), nitrogen dioxide (NO). Eur Rev Med Pharmacol Sci, 28, pp.789-96. https://www.europeanreview.org/wp/wp-content/uploads/789-796.pdf*

[7] Restrepo, C.E., 2021. Nitrogen dioxide, greenhouse gas emissions and transportation in urban areas: lessons from the covid-19 pandemic. *Frontiers in Environmental Science*, *9*, p.689985. https://doi.org/10.3389/fenvs.2021.689985