

DATA ANALYSIS OF MAJOR AIR POLLUTANTS IN INDIAN CITIES: EMPHASIZING AMBIENT NO₂ EXPOSURE AND ASSOCIATED ALTERED GENE EXPRESSIONS

**Submitted in partial fulfillment of the requirements for the
award of Bachelor of Science degree in Bioinformatics and
Data Science**

By

MARY PORTIA J (40738011)



**DEPARTMENT OF BIOINFORMATICS
SCHOOL OF BIO AND CHEMICAL ENGINEERING**

SATHYABAMA

**INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)**

Accredited with Grade "A" by NAAC

JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI - 600 119

MAY – 2023



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE

www.sathyabama.ac.in

DEPARTMENT OF BIOINFORMATICS AND DATA SCIENCE

BONAFIDE CERTIFICATE

This is to certify that this Project is the bonafide work of **MARY PORTIA . J (40738011)** who carried out the project entitled **"Data Analysis of Major Air Pollutants in Indian Cities: Emphasizing Ambient NO2 Exposure and Associated Altered Gene Expression"** under our supervision from August 2022 to May 2023.

Internal Guide

Head of the Department

DR JEMMY CHRISTY H

Submitted for Viva voce Examination held on **05/05/23**

Internal Examiner

External Examiner

DECLARATION

I , **MARY PORTIA . J (40738011)** hereby declare that the Project entitled **Data Analysis of Major Air Pollutants in Indian Cities: Emphasizing Ambient NO2 Exposure and Associated Altered Gene Expression** done by me under the guidance of **Dr. Jemmy Christy H** (Internal) and **Dr. Cruz Antony J** (External) at Sathyabama Institute of Science and Technology , is submitted in partial fulfillment of the requirements for the award of Bachelor of Science in Bioinformatics and Data Science .

DATE: 05/05/23

PLACE:CHENNAI

Mary portia

SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

I am grateful to the Almighty for all His blessings throughout my course of study. I am blessed and thankful always.

I am pleased to acknowledge my sincere thanks to Board of Management of SATHYABAMA UNIVERSITY for their kind encouragement in doing this project and for completing it successfully.

I convey my thanks to Dr. RAMESH KUMAR V , Dean, School of Bio and Chemical and Dr. JEMMY CHRISTY H, Head of the Department, Dept. of Bioinformatics and Data Science for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guides Dr JEMMY CHRISTY H (internal) and Dr CRUZ ANTONY J (external) for their valuable guidance, suggestions and constant encouragement which paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the Department of Bioinformatics and Data Science who were helpful in many ways for the completion of the project.

ABSTRACT

Air pollution is a critical environmental concern in many urban areas, including cities in India. In this project, an analysis of the Air Quality Index (AQI) data from Indian cities was conducted, with a particular focus on nitrogen dioxide (NO₂) exposure. The study aimed to understand the health impacts associated with NO₂ exposure by investigating alterations in proteins. The analysis revealed that NO₂ levels in Indian cities were alarmingly high, surpassing the permissible limits set by regulatory authorities. The findings indicated that NO₂ exposure was a major contributor to air pollution in these cities. Further, the study investigated the alterations in gene expressions caused by NO₂ exposure. The results showed that NO₂ exposure led to significant changes in protein expression patterns, suggesting potential impacts on cellular functions and signalling pathways. These protein alterations were found to be associated with various health impacts, including respiratory and cardiovascular diseases. The findings highlight the detrimental effects of NO₂ exposure on human health and the urgent need for effective air pollution control measures to mitigate these impacts. In conclusion, this project provides important insights into the relationship between NO₂ exposure, protein alterations, and associated health impacts. The findings underscore the need for immediate action to reduce NO₂ emissions and improve air quality in urban areas to safeguard public health.

LIST OF ABBREVIATIONS

CHAPTER NO	TITLE	PAGE NO
1	INTRODUCTION 1.1 PROBLEM STATEMENT OVERVIEW	1-2
2	LITERATURE REVIEW	3-5
3	AIM AND SCOPE OF THE PRESENT STUDY 3.1 AIM 3.2 OBJECTIVE OF STUDY 3.3 SCOPE	5-8
4	MATERIALS AND METHODS 4.1 REQUIREMENTS SPECIFICATION 4.2 BRIEF ABOUT THE SOFTWARE 4.3 BRIEF ON ONLINE TOOLS 4.4 WORKFLOW	9-14
5	RESULTS AND DISCUSSION 5.1 DEPLOYING DATASETS AND FINDING CLASSIFICATION ACCURACY 5.2 GENERATING CN2 RULES 5.3 ARIMA MODEL 5.4 ARIMA CODE 5.5 PROTEIN NETWORK ANALYSIS 5.6 ANALYSIS OF GENE ONTOLOGY BIOLOGICAL PROCESS TERMS 5.7 INTERPRETATION OF GO RESULTS	15-42
6	SUMMARY AND CONCLUSIONS 6.1 SUMMARY 6.2 CONCLUSIONS	43-44
7	REFERENCES	45-52

LIST OF FIGURES

FIGURE NO	FIGURE NAME
5.1.1	Orange Workflow
5.1.2	Class accuracy scores from Orange
5.2.1	Cn2 rule induction
5.2.2	Cn2 Rule Viewer
5.3.1	Arima Prediction Result
5.4	Protein Network
5.5.1	Gonet Network
5.5.2	Representation of Gene Ontology in cytoscape

LIST OF TABLES

TABLE NO	TABLE NAME
Table 4.4	List of proteins altered by no2 exposure

1. INTRODUCTION

1.1 PROBLEM STATEMENT OVERVIEW

The problem of air quality index in Indian cities has become a major concern in recent years. The air pollution levels in many Indian cities have reached alarming levels, posing serious health risks to the population (Luthra et al., 2023). The air quality index (AQI), which is a measure of the concentration of harmful pollutants in the air, has consistently exceeded the permissible limits set by the World Health Organization (WHO) in many cities across India (Pal et al., 2023).

One of the major contributors to air pollution in Indian cities is vehicular emissions. The rapid increase in the number of vehicles on the road, combined with poor emission standards and inadequate public transportation options, has led to a significant rise in air pollution (Yadav et al., 2023). Additionally, industrial emissions, construction dust, and burning of crop residue are also major sources of air pollution in Indian cities.

The impact of air pollution on public health cannot be overstated. Studies have shown that exposure to high levels of air pollution can lead to respiratory diseases, cardiovascular diseases, and even premature death (Gupta et al., 2022). Children, the elderly, and those with pre-existing health conditions are particularly vulnerable to the adverse effects of air pollution (Ruidas et al., 2022).

Efforts to combat air pollution in Indian cities have been made, including the implementation of emission standards for vehicles, restrictions on burning of crop residue, and measures to reduce industrial emissions. However, there is a need for more stringent regulations, better enforcement, and increased public awareness about the issue of air pollution in order to effectively address this growing problem.

Exposure to poor air quality, as indicated by high AQI levels, has been linked to a wide range of health impacts. Respiratory diseases such as asthma, chronic bronchitis, and respiratory infections are common outcomes of prolonged exposure to polluted air. (Dass et al., 2022). Cardiovascular diseases, including heart attacks and strokes, are also associated with air pollution. Other health impacts can include adverse effects on fetal development, increased risk of cancer, and neurological and cognitive disorders. (Gangwar et al., 2019).

Moreover, the socio-economic burden of health impacts caused by air pollution is substantial, resulting in increased healthcare costs, lost productivity, and reduced quality of life for affected individuals. Vulnerable populations such as children, the elderly, and those with pre-existing health conditions are particularly at risk, and the impacts of air pollution can exacerbate existing health disparities.

As a result, the increasing concern about AQI is driven by the growing recognition of the severe implications of air pollution on public health, necessitating urgent measures to mitigate pollution levels and protect the health and well-being of the population.

In conclusion, the air quality index in Indian cities is a concerning issue that poses significant health risks to the population. Vehicular emissions, industrial emissions, construction dust, and crop residue burning are major contributors to air pollution. Efforts to combat air pollution have been made, but more needs to be done to effectively address this problem and protect the health of the population (Mishra et al., 2021). The increasing concern about Air Quality Index (AQI) stems from its direct implications on public health. AQI is a measurement that reflects the concentration of pollutants in the air, including PM, NO₂, SO₂, and others, and provides an assessment of the overall air quality. As AQI levels worsen, it signifies higher levels of air pollution, which can have detrimental effects on human health.

2. LITERATURE REVIEW

Globally, air pollution is a silent killer. The air pollution levels in India are among the highest in the world, posing a heavy threat to the country's health and economy. Air pollution shortens the average Indian life expectancy by 6.3 years, relative to what it would be if the World Health Organization (WHO) guideline was met; 3.4 years relative to what it would be if pollution were reduced to meet the country's own national standard. The air pollution challenge in India is therefore inherently multi-sectoral and multi-jurisdictional (Gupta et al., 2023). In recent years, research has focused on analysing AQI data to better understand the extent of air pollution and its associated health impacts, with a particular emphasis on nitrogen dioxide (NO₂) exposure. Additionally, studies have investigated the role of protein alterations in response to NO₂ exposure and the resulting health implications.

Several studies have examined the AQI data of Indian cities to assess the levels of air pollution, with NO₂ being identified as a prominent pollutant. For instance, a study by (Gupta et al., 2023) analysed AQI data from 10 major cities in India and found that NO₂ levels frequently exceeded the permissible limits set by regulatory authorities, indicating high levels of NO₂ pollution. Another study by (Khan et al., 2020) analysed AQI data from Delhi, one of the most polluted cities in the world, and reported consistently high levels of NO₂ pollution, particularly during the winter months when the use of biomass fuels for cooking and heating is common.

Furthermore, research has focused on understanding the health impacts of NO₂ exposure in Indian cities. NO₂ is known to cause respiratory and cardiovascular health effects, as it can penetrate deep into the lungs and irritate the respiratory system (Huang et al., 2021). Studies have shown that NO₂ exposure is associated with increased risks of respiratory diseases such as asthma, chronic bronchitis, and respiratory infections .

Moreover, NO₂ has been linked to cardiovascular diseases, including heart attacks and strokes, due to its role in promoting inflammation and oxidative stress (Panda et al., 2022). Vulnerable populations such as children, the elderly, and individuals with pre-existing health conditions are particularly susceptible to the adverse health effects of NO₂ exposure (Huangfu et al., 2020).

In addition to health impacts, studies have investigated the role of protein alterations in response to NO₂ exposure. NO₂ can react with proteins in the respiratory tract, leading to oxidative modifications and alterations in protein structure and function. (Woodby et al., 2021). These protein alterations can have downstream effects on cellular processes and signalling pathways, potentially contributing to the health impacts of NO₂ exposure (Morris et al., 2021). For example, the study investigated proteomic changes in lung tissues of rats exposed to NO₂ and identified alterations in proteins involved in inflammation, oxidative stress, and cellular repair processes.

The findings from these studies underscore the importance of understanding the relationship between AQI, NO₂ exposure, protein alterations, and health impacts in Indian cities. The high levels of NO₂ pollution in many urban areas, as indicated by AQI data, highlight the urgent need for effective measures to mitigate air pollution and protect public health. (Mahato et al., 2020). Interventions range from a focus on individuals/families, organizations, communities and in healthcare, public health, education and media systems, health-promoting settings, and policy. Lessons learned highlight the need for emphasizing equity, trust, systems approach, and sustained action in future health promotion preparedness strategies. Challenges and opportunities are highlighted regarding the need for rapid response, clear communication based on health literacy, and collaboration across countries, disciplines, and health and education systems for meaningful solutions to global health crises (Lavin et al., 2021).

Further research on the mechanisms underlying the protein alterations in response to NO₂ exposure and their implications on health is warranted. Such knowledge can contribute to the development of targeted interventions to mitigate the adverse health effects of air pollution in Indian cities, particularly in vulnerable populations, and inform policy-making efforts to improve air quality and safeguard public health.

3. AIM AND SCOPE OF THE PRESENT STUDY

3.1 AIM

- **Assessing AQI levels:** The study aimed to analyse the AQI data of Indian cities to determine the levels of air pollution, with a specific focus on NO₂ as a prominent pollutant. This objective involved evaluating the extent of NO₂ exposure in different cities, identifying areas with high pollution levels, and understanding the spatial and temporal patterns of NO₂ concentration.
- **Investigating protein alterations:** The study sought to investigate how NO₂ exposure can lead to protein alterations in the body. This objective involved analysing the molecular changes in proteins, such as modifications in their structure, function, or expression levels, in response to NO₂ exposure. This could have included proteomic analysis techniques to identify and quantify protein alterations associated with NO₂ exposure.
- **Understanding health impacts:** The study aimed to elucidate the health impacts associated with NO₂ exposure. This objective involved exploring the relationship between NO₂ exposure, protein alterations, and health outcomes, such as respiratory diseases, cardiovascular diseases, and other health impacts. The study may have investigated the underlying molecular mechanisms and pathways that link NO₂ exposure-induced protein alterations to health impacts.
- **Highlighting health risks:** The study aimed to highlight the potential health risks associated with NO₂ exposure, based on the findings from AQI data analysis and protein alterations analysis. This objective involved raising awareness about the adverse health impacts of NO₂ exposure on vulnerable populations, such as children, the elderly, and individuals with pre-existing health conditions, and emphasizing the need for effective measures to mitigate NO₂ pollution.

3.2 OBJECTIVE OF STUDY

The objective of the study was to analyze the Air Quality Index (AQI) data from Indian cities, with a specific focus on nitrogen dioxide (NO₂) exposure. The study aimed to identify the levels of NO₂ pollution in different cities and investigate the associated protein alterations in response to NO₂ exposure. The ultimate goal was to understand the health impacts of NO₂ exposure at the molecular level, and provide

insights into the mechanisms underlying the adverse health effects of air pollution. The study aimed to contribute to the existing knowledge on the relationship between AQI, NO₂ exposure, protein alterations, and associated health impacts in the context of Indian cities. There is a pressing need for research and studies on the air quality index (AQI) in India due to the escalating air pollution crisis in the country. With increasing urbanization, industrialization, and vehicular emissions, many Indian cities are grappling with poor air quality, posing significant risks to public health and the environment. In order to effectively address this issue, research and studies are crucial to understand the extent, sources, impacts, and mitigation strategies related to air pollution in India. Several studies have highlighted the need for comprehensive research on AQI in India. For instance, a study by (Gangwar et al., 2016) emphasized the need for continuous monitoring of AQI, along with source apportionment studies, to identify the major contributors to air pollution in different regions of India. The importance of studying the impacts of air pollution on vulnerable populations, such as children, the elderly, and those with pre-existing health conditions, to better understand the health risks and develop targeted mitigation strategies. Furthermore, research on AQI in India can also help in evaluating the effectiveness of existing policies and regulations, identifying gaps in implementation and enforcement, and recommending evidence-based interventions to reduce air pollution levels. Studies on the economic costs of air pollution, including healthcare costs and productivity losses, can also provide valuable insights to policymakers for prioritizing air quality management measures.

3.3 SCOPE

The scope of the above study is to analyze the AQI data of Indian cities, with a focus on NO₂ exposure. The study further investigates protein alterations in response to NO₂ exposure and examines the associated health impacts. The research likely involves collecting and analyzing AQI data from various Indian cities, studying the levels of NO₂ exposure, and identifying protein alterations through molecular techniques. The study may also include evaluating the health impacts of NO₂ exposure through literature review or data analysis, and drawing conclusions about

the potential health risks associated with NO₂ exposure in Indian cities. The findings of this study could contribute to the understanding of the relationship between air pollution, specifically NO₂ exposure, protein alterations, and associated health impacts in the context of Indian cities. The air quality index (AQI) is a critical parameter used to measure the quality of ambient air and its impact on human health and the environment. In the context of Indian cities, the scope of studying the AQI is vast and encompasses various aspects, including monitoring, assessment, and management of air quality. One significant area of study is the assessment of the current air quality levels in different Indian cities. Several studies have investigated the levels of pollutants such as particulate matter (PM), Sulphur dioxide (SO₂), nitrogen dioxide (NO₂), and volatile organic compounds (VOCs) in cities like Delhi, Mumbai, Kolkata, and Chennai, among others. These studies have provided valuable insights into the severity of air pollution in urban areas, the sources of pollution, and the impacts on human health. Furthermore, studies have shown that the AQI in many Indian cities tends to deteriorate during winter months due to factors such as crop residue burning, vehicular emissions, and meteorological conditions, which trap pollutants close to the surface. Studies have also explored the health impacts of poor air quality on vulnerable populations, such as children, the elderly, and those with pre-existing health conditions. The adverse effects of air pollution on respiratory health, cardiovascular health, and cognitive functions have been well-documented in Indian cities, highlighting the need for effective air quality management measures. To mitigate the air pollution crisis, there is a growing body of literature on policy interventions, technological solutions, and behavioural changes. It has been evaluated that the effectiveness of measures such as vehicular emission controls, industrial emission regulations, and urban planning strategies in improving air quality in Indian cities.

In conclusion, the scope of studying the AQI in Indian cities is extensive and includes monitoring, assessment, and management of air quality, investigating the sources and impacts of pollution, understanding temporal and spatial variations, examining health impacts, and evaluating policy interventions. These studies provide crucial insights for policymakers, urban planners, and researchers to develop evidence-based strategies for improving air quality and safeguarding public health.

4. MATERIALS AND METHODS

4.1 REQUIREMENTS SPECIFICATION

4.1.1 Operating System - WINDOWS 10 (licensed)

4.1.2 Software – ORANGE (Python 3) , CYTOSCAPE (licensed)

4.1.3 Online tools - HUMAN PROTEIN ATLAS, UNIPROT, PINA , SHINYGO ,
GONET, QuickGO (open source)

4.2 BRIEF ABOUT THE SOFTWARE

4.2.1 Orange

Orange is a powerful platform to perform data analysis and visualization, see data flow and become more productive. Orange is an open-source data visualization, machine learning and data mining toolkit. It features a visual programming front-end for explorative qualitative data analysis and interactive data visualization.

Visual programming is implemented through an interface in which workflows are created by linking predefined or user-designed widgets, while advanced users can use Orange as a Python library for data manipulation . Orange uses common Python open-source libraries for scientific computing, such as numpy, scipy and scikit-learn, while its graphical user interface operates within the cross-platform Qt framework.

The default installation includes a number of machine learning, pre-processing and data visualization algorithms in 6 widget sets (data, transform, visualize, model, evaluate and unsupervised). Additional functionalities are available as add-ons (text-mining, image analytics, bioinformatics, etc.).

The program provides a platform for experiment selection, recommendation systems, and predictive modelling and is used in biomedicine, bioinformatics, genomic research, and teaching. In science, it is used as a platform for testing new machine learning algorithms and for implementing new techniques in genetics and bioinformatics. In education, it was used for teaching machine learning and data mining methods to students of biology, biomedicine, and informatics. (Demsar J, et al., 2013)

4.2.2 Cytoscape

Cytoscape is an open- source software platform for visualizing molecular interaction networks and biological pathways and integrating these networks with annotations, gene expression profiles and other state data. Although Cytoscape was originally designed for biological research, now it is a general platform for complex network analysis and visualization.

Cytoscape supports many use cases in molecular and systems biology, genomics, and proteomics:

- Load molecular and genetic interaction data sets in many standards formats
- Project and integrate global datasets and functional annotations
- Establish powerful visual mappings across these data
- Perform advanced analysis and modeling using Cytoscape Apps
- Visualize and analyze human-curated pathway datasets such as WikiPathways, Reactome, and KEGG.(Shannon P, 2003)

4.3 BRIEF ON ONLINE TOOLS

4.3.1 Human Protein Atlas

The Human Protein Atlas is a unique world-leading effort to map all the human proteins in cells, tissues, and organs in the human body using antibody-based imaging, mass spectrometry-based proteomics, transcriptomics, and systems

biology.

The resource now includes twelve separate sections with complementary information about all human proteins. All data has been updated on the approximately 5 million individual web pages. The Human Protein Atlas program has already contributed to several thousands of publications in the field of human biology and disease and was selected by the organization ELIXIR as a European core resource due to its fundamental importance for a wider life science community. The HPA consortium is funded by the Knut and Alice Wallenberg Foundation.(Uhlén M, 2015)

4.3.2 Uniprot

UniProt is the world's leading high-quality, comprehensive and freely accessible resource of protein sequence and functional information, many entries being derived from genome sequencing projects. It contains a large amount of information about the biological function of proteins derived from the research literature.

UniProt provides an up-to-date, comprehensive body of protein information. The resource facilitates scientific discovery by collecting, interpreting and organizing this information, which saves researchers countless hours of work. The database contains over 60 million sequences, of which over half a million sequences have been curated by experts who critically review experimental and predicted data for each protein.(The UniProt Consortium , 2023)

4.3.3 Pina

Protein Interaction Network Analysis (PINA) platform is an integrated platform for protein interaction network construction, filtering, analysis, visualization, and management. It integrates protein-protein interaction (PPI) data from public curated databases and builds a complete, non-redundant protein interaction dataset for six model organisms. In particular, it provides a variety of built-in tools to filter and analyze the networks for gaining biological and functional insights into the network.

Network construction supports generating a protein-protein interaction network from:

- a single protein
- a set of proteins
- a list of protein interactions (Stumpf, M.P , 2005)

4.3.4 ShinyGo

ShinyGO is an intuitive, graphical web application that can help researchers gain actionable insights from gene-sets. Gene lists are routinely produced from various omics studies. Enrichment analysis can link these gene lists with underlying molecular pathways and functional categories such as gene ontology (GO) and other databases. It accepts a set of gene IDs as input, and produces GO enrichment, with rich visualization. (Ge SX, 2020)

4.3.5 GOnet

GOnet specifically aims to construct and display interactive graphs that include GO terms and genes while retaining term-gene relationships. It takes a list of gene or protein entries from human or mouse data and performs GO term annotation analysis. The application is capable of producing parsable data formats and importantly, interactive visualizations of the GO analysis results. The interactive results allow exploration of genes and GO terms as a graph that depicts the natural hierarchy of the terms and retains relationships between terms and genes/proteins. As a result, GOnet provides insight into the functional interconnection of the submitted entries. (Pomaznoy M, 2018)

4.3.6 QuickGO

The QuickGO interface has been built using JavaScript, Ajax and HTML, and uses an optimized and streamlined set of files to facilitate large queries and data extraction. Besides providing GO term and annotation data, QuickGO also imports protein descriptions and taxonomic hierarchies from UniProtKB. The tool is updated with GO term information nightly and GO annotations weekly, ensuring that it is one of the most up-to-date GO resources. (Binns D, 2009)

4.4 WORKFLOW

The real time raw data was collected from the official website of CENTRAL POLLUTION CONTROL BOARD (CPCB | *Central Pollution Control Board*. (n.d.). CPCB. <https://cpcb.nic.in/>). The metropolitan cities for analysis include Delhi, Chennai, Mumbai and Bangalore. The time period of data collection is from 2009 to 2021. The files were downloaded year wise in excel format. They include the logo of CPCB and the information regarding the state, city, time and date of download,

station, parameters , prescribed standards , exceeding standards, and the data of the air pollutants along with from date and to date . 16 Datasets created for 16 cities (Ahmedabad, Amaravati, Amristar, Bengaluru, Brajrajnagar , Chennai, Delhi, Gurugram, Hyderabad, Jaipur, Jorapokhar , Kolkata, Lucknow, Patna, Talcher, Thiruvananthapuram) .Orange was used to deploy Datasets.

All the raw data files were first studied to understand about the parameters and its existing standards. The files were merged using merging tools available online. Basic excel commands were used to consolidate the data. The attributes in the dataset are from date, to date, PM 2.5 , PM 10 , NO , NO2 , NOx .The results were retrieved and studied further .

5. RESULTS AND DISCUSSION

5.1 DEPLOYING DATASETS AND FINDING THE CLASSIFICATION ACCURACY

Data has been collected from the CPCB website. Data Segregation was done and 16 Datasets created for Ahmedabad, Amaravati, Amristar, Bengaluru, Brajrajnagar , Chennai, Delhi, Gurugram, Hyderabad, Jaipur, Jorapokhar , Kolkata, Lucknow, Patna, Talcher, Thiruvananthapuram. The Datasets were deployed in Orange package. The classification accuracy is found. The average level of accuracy lies between 79.4 to 88 %.

Rule induction is a data mining process of deducing if-then rules. It is the symbolic decision rules explain an inherent relationship between the attributes and class labels in the data set.

Scores					
Model	AUC	CA	F1	Precision	Recall
kNN	0.87124718167794	0.8663366336633663	0.8620597080512491	0.8623390051269278	0.8663366336633663
Tree	0.8024801234128397	0.8316831683168316	0.8283223779570735	0.82770197187786	0.8316831683168316
SVM	0.9058383766464935	0.8465346534653465	0.807690597644989	0.8647214297700956	0.8465346534653465
Random Forest	0.9083007001305328	0.8811881188118812	0.8784868402806668	0.8778560548362528	0.8811881188118812
Neural Network	0.8904117716862466	0.8811881188118812	0.8756188118811881	0.8768621805814535	0.8811881188118812
Naive Bayes	0.9044143823424706	0.6881188118811881	0.7118600564375374	0.8099402461558174	0.6881188118811881
Logistic Regression	0.8825798030141214	0.8613861386138614	0.855497444905781	0.8563133584578617	0.8613861386138614
Gradient Boosting	0.8798504805980777	0.8613861386138614	0.8563793879387938	0.8568077431545648	0.8613861386138614
CN2 rule inducer	0.7871128515485939	0.806930693069307	0.7998867435966056	0.7964706039337087	0.806930693069307
AdaBoost	0.7546279814880741	0.806930693069307	0.8034310361729243	0.8006151718112987	0.806930693069307

Fig 5.1.2 Classification accuracy scores from Orange

5.2 GENERATING CN2 RULES

The CN2 algorithm is a classification technique designed for the efficient induction of simple, comprehensible rules of form "if *cond* then predict *class*", even in domains where noise may be present.

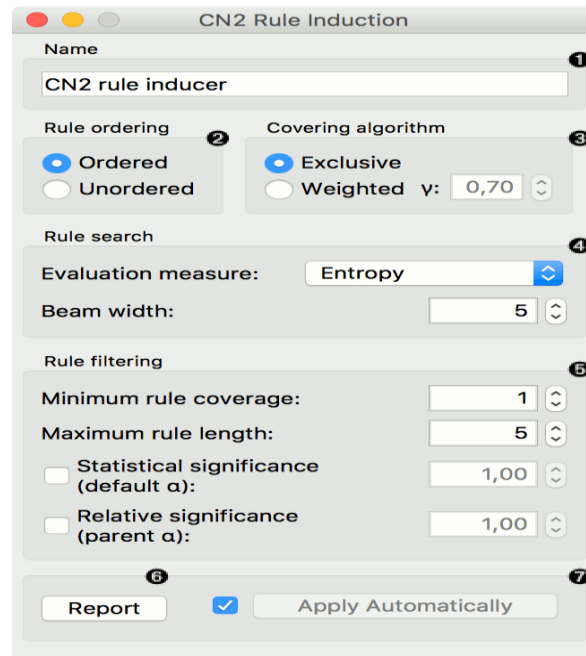


Fig 5.2.1 Cn2 rule induction

The rule induction model is used to understand the impact of NO₂ on the AQI over the different cities in India since the rules (IF-THEN) are said to be more expressive and human-readable in nature. One of the sequential covering algorithms, CN2 rule induction is used in building the model. CN2 induction algorithm, (Clark and Niblett, 1989) is a collective work based on algorithm quasi-optimal (AQ) algorithm and Iterative Dichotomiser 3 (ID3) algorithm. The CN2 algorithm delivers a ranked list of rules readily accessible to decision-makers and was introduced to overcome the drawbacks of the then-standard AQ/ID3 induction algorithms (Heymann et al 2022).

It is developed to classify rules accurately even in the presence of noise and the rules are made to be efficient and simple. CN2 classification rules have shown 93.73% accuracy in heart disease prediction, a comparative study done by (Ramaraj and Thanamani, 2013).

The procedure for the CN2 induction algorithm (Clarke and Niblett, 1989) is given below

Let E be a set of classified examples.
Let SELECTORS be the set of all possible selectors.

Procedure CN2(E)

 Let RULE_LIST be the empty list.
 Repeat until BEST_CPX is nil or E is empty:
 Let BEST_CPX be Find_Best_Complex(E).
 If BEST_CPX is not nil,
 Then let E' be the examples covered by BEST_CPX.
 Remove from E the examples E' covered by BEST_CPX.
 Let C be the most common class of examples in E'.
 Add the rule c1f BEST_CPX then the class is C'
 to the end of RULE_LIST.
 Return RULE_LIST.

Procedure Find_Best_Complex(E)

 Let STAR be the set containing the empty complex.
 Let BEST_CPX be nil.
 While STAR is not empty,
 Specialize all complexes in STAR as follows:
 Let NEWSTAR be the set {x A y | x 6 STAR, y 6 SELECTORS}.
 Remove all complexes in NEWSTAR that are either in STAR (i.e., the unspecialized ones) or null (e.g., big = y A big : n).
 For every complex C i in NEWSTAR:
 If C i is statistically significant and better than
 BEST_CPX by user-defined criteria when tested on E,
 Then replace the current value of BEST_CPX by C i.
 Repeat until size of NEWSTAR < user-defined maximum:
 Remove the worst complex from NEWSTAR.
 Let STAR be NEWSTAR.
 Return BEST_CPX.

CN2 Rule Viewer

Wed Jan 04 23, 12:34:02

Data domain

Features: PM2.5, NO, PM10, NO2, NOx, CO, SO2, NH3, O3

Target: AQI_Bucket

Rule induction algorithm

Rule ordering: ordered

Covering algorithm: exclusive

Gamma: 0.7

Evaluation measure: entropy

Beam width: 5

Minimum rule coverage: 5

Maximum rule length: 5

Default alpha: 1.0

Parent alpha: 1.0

Induced rules

	IF conditions	THEN class	Distribution	Probabilities [%]	Quality	Length
	CO≥4.17 AND PM2.5≥142.27	→ AQI_Bucket=Very Poor	[0, 0, 0, 0, 5]	10 : 10 : 10 : 10 : 60	-0.00	2
1	CO≥4.17 AND O3≥18.67	→ AQI_Bucket=Poor	[0, 0, 5, 0, 0]	10 : 10 : 60 : 10 : 10	-0.00	2
2	CO≥1.861664331279016 AND O3≥18.76 AND SO2≥12.19	→ AQI_Bucket=Moderate	[0, 42, 0, 0, 0]	2 : 91 : 2 : 2 : 2	-0.00	3
3	NO≤14.711687818675866 AND NO≥14.711687818675866	→ AQI_Bucket=Moderate	[0, 98, 0, 0, 0]	1 : 96 : 1 : 1 : 1	-0.00	2
4	PM2.5≥101.16 AND NH3≥65.87	→ AQI_Bucket=Poor	[0, 0, 16, 0, 0]	5 : 5 : 81 : 5 : 5	-0.00	2
5	PM2.5≤44.87 AND CO≥1.97	→ AQI_Bucket=Moderate	[0, 60, 0, 0, 0]	2 : 94 : 2 : 2 : 2	-0.00	2
6	NH3≤4.59 AND NO≥6.94	→ AQI_Bucket=Satisfactory	[0, 0, 0, 11, 0]	6 : 6 : 6 : 75 : 6	-0.00	2
7	O3≥21.74 AND NO≥22.1 AND NO2≥17.51	→ AQI_Bucket=Moderate	[0, 13, 0, 0, 0]	6 : 78 : 6 : 6 : 6	-0.00	3
8	PM2.5≥101.16 AND NOx≥50.16	→ AQI_Bucket=Poor	[0, 0, 6, 0, 0]	9 : 9 : 64 : 9 : 9	-0.00	2
9	CO≥2.12 AND NH3≥85.21	→ AQI_Bucket=Moderate	[0, 9, 0, 0, 0]	7 : 71 : 7 : 7 : 7	-0.00	2
10	CO≥2.12 AND PM2.5≤70.35 AND NH3≥12.14	→ AQI_Bucket=Moderate	[0, 28, 0, 0, 0]	3 : 88 : 3 : 3 : 3	-0.00	3
11	PM2.5≥101.16 AND PM2.5≤105.42 AND NOx≥38.83085727407291	→ AQI_Bucket=Poor	[0, 0, 6, 0, 0]	9 : 9 : 64 : 9 : 9	-0.00	3
12	PM10≥126.98 AND PM2.5≤84.99 AND O3≥34.20014797620647	→ AQI_Bucket=Moderate	[0, 20, 0, 0, 0]	4 : 84 : 4 : 4 : 4	-0.00	3
13	NO≤10.55 AND CO≥2.5 AND NO≥5.36	→ AQI_Bucket=Moderate	[0, 10, 0, 0, 0]	7 : 73 : 7 : 7 : 7	-0.00	3
14	PM10≤126.98 AND NOx≥31.93 AND NO≥24.28	→ AQI_Bucket=Moderate	[0, 8, 0, 0, 0]	8 : 69 : 8 : 8 : 8	-0.00	3
	PM10≥126.98 AND NO≥15.06 AND			10 : 60 : 10 :		

Fig 5.2.2 Cn2 Rule Viewer

It is been inferred that NO₂, NO, PM_{2.5} are the primary air pollutants in the following cities Ahmedabad, Amaravati, Amristar, Bengaluru, Brajrajnagar , Chennai, Delhi, Gurugram, Hyderabad, Jaipur, Jorapokhar , Kolkata, Lucknow, Patna, Talcher, Thiruvananthapuram

5.3 ARIMA MODEL

Presently, we face rapid changes too often in the environment and it would be crucial for us to understand and forecast in order to take necessary precautionary measures for the betterment of life. Usually, forecasting models vary from educated guesses to computer projections using sophisticated statistical analyses. Autoregressive Integrated Moving Average (ARIMA) model is effectively used in several environmental time-series data analyses and modeling and it is considered the most prominent in environmental forecasting. The ARIMA model was introduced by Box and Jenkins in 1970, it consists of features for identifying, estimating, and diagnosing models with time-series data. Based on a linear combination of its past values, the ARIMA model predicts a value in a response time series.

The ARIMA model is composed of three components 1. Auto-Regression 2. Integrated 3. Moving Average.

Auto-Regression is basically using the previous values (lagged values) γ until p^{th} time in the past in order to predict the future along with the constant c , an error ϵ at time t and α s are the influencing factors.

The Auto-Regression AR(p) is represented by

$$\text{AR}(p) = \gamma_t = c + \epsilon_t + \sum_{1}^p \alpha \sum_{t-1}^p \gamma$$

Integrated $I(d)$ is the difference taken d times until the time-series data becomes stationary $\gamma'_t = \gamma_t - \gamma_{t-1}$. In order to model any time-series data, the statistical properties of the data should be constant with time i.e., with constant mean, variance, and unrepeatable patterns in the dataset. Finally, the moving average $MA(q)$ is a regression-like model which uses previous errors to make future predictions.

The Moving Average $MA(q)$ is represented by

$$MA(q) = \gamma_t = c + \sum_1^q \beta \sum_{t-1}^q \epsilon$$

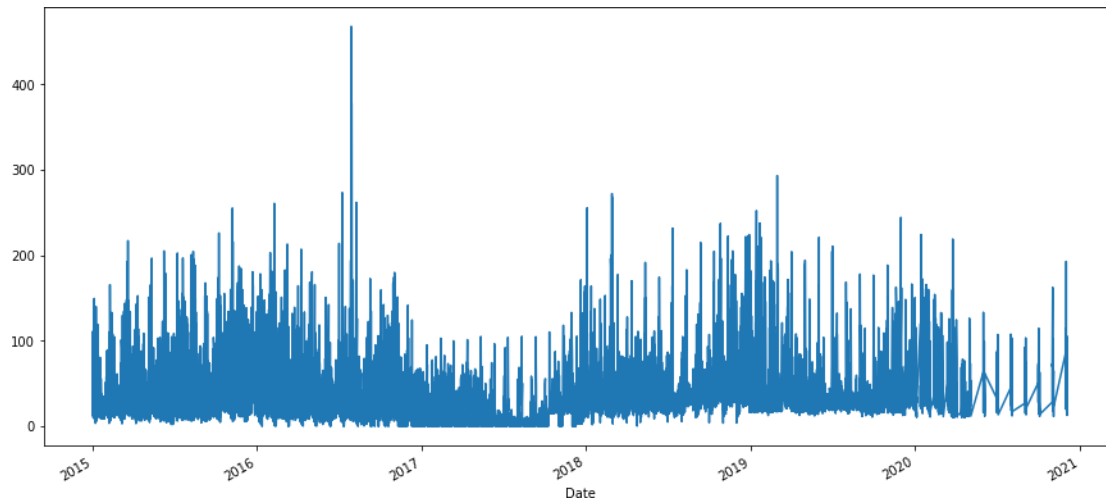
Combining all three components $ARIMA(p,d,q)$ model is represented by

$$\gamma'_t = c + \sum_1^p \alpha \sum_{t-1}^p \gamma' + \sum_1^q \beta \sum_{t-1}^q \epsilon + \epsilon_t$$

5.4 ARIMA CODE

```
import pandas as pd
df=pd.read_csv('/content/DATASET_MARY_PORTIA.
.csv', index_col='Date', parse_dates=True)
df=df.dropna(subset=['NOx']) print('Shape of data', df.shape) df.head()
df
```

[12]: <AxesSubplot:xlabel='Date'>



```
!pip install pmdarima
```

```
from statsmodels.tsa.stattools import adfuller
def adf_test(dataset):
    dftest = adfuller(dataset, autolag = 'AIC')
    print("1. ADF : ", dftest[0])
    print("2. P-Value : ", dftest[1])
    print("3. Num Of Lags : ", dftest[2])
    print("4. Num Of Observations Used For ADF Regression:", dftest[3])
    print("5. Critical Values :")
    for key, val in dftest[4].items():
        print("\t", key, ": ", val)
    adf_test(df["NOx"])
```

```
1. ADF : -5.536620461424785
2. P-Value : 1.7389032438008195e-06
3. Num Of Lags : 37
4. Num Of Observations Used For ADF Regression: 8068
5. Critical Values :
1% : -3.4311607810839737
5% : -2.861898307561281
10% : -2.566960722380465
```

```
from pmdarima import auto_arima
stepwise_fit = auto_arima(df["NOx"], trace=True, suppress_warnings=True)
```

```
print(df.shape) train=df.iloc[:-30] test=df.iloc[-30:] print(train.shape, test.shape)
```

```
import statsmodels.api as sm
model=sm.tsa.arima.ARIMA(train["NOx"], order=(3,1,2))
model=model.fit()
model.summary()
```

```
]: <class 'statsmodels.iolib.summary.Summary'>''''
```

SARIMAX Results

```

=====
=====
Dep. Variable:          NOxNo. Observations:          8076
Model:                ARIMA(3, 1, 2)    Log Likelihood    -33946.856
Date:                Tue, 07 Mar 2023    AIC                67905.712
Time:                15:43:13            BIC                67947.692
Sample:              - 8076              0HQIC              67920.074

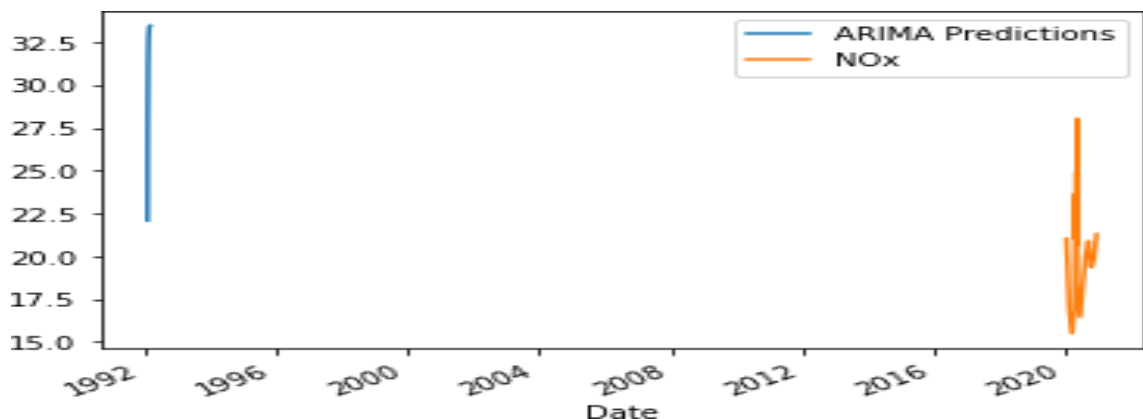
Covariance Type:      opg
=====
=====
coef          std err          z      P>|z|          [0.025      0.975]
-----
ar.L1          1.2236          0.05621.675      0.000          1.113      1.334
ar.L2         -0.4937          0.035-14.059      0.000         -0.563     -0.425
ar.L3          0.1301          0.00621.344      0.000          0.118      0.142
ma.L1         -1.4768          0.056-26.206      0.000         -1.587     -1.366
ma.L2          0.5087          0.05010.140      0.000          0.410      0.607
sigma2        262.4200          1.248210.213      0.000        259.973     264.867
=====
=====
===
Ljung-Box (L1) (Q):          0.00Jarque-Bera (JB):
232126.95
Prob(Q):                    0.96Prob(JB):
0.00
Heteroskedasticity (H):      13.16Skew:
1.09
Prob(H) (two-sided):         0.00Kurtosis:

```

```

start=len(train['NOx']) end=len(train['NOx'])+len(test['NOx'])-1
pred=model.predict(start=start,end=end,typ='levels').rename('ARIMA Predictions')
pred.plot(legend=True)
test['NOx'].plot(legend=True)

```



5.5 PROTEIN NETWORK ANALYSIS

Literature Review was done to find the proteins altered by NO₂ exposure. It was found that 19 proteins were altered due to NO₂ exposure.

The proteins are as follows

Q6NUM9, Q99584, O95167, Q12849, Q86TV6, Q3KNW1, O43301, P49755, Q16270, Q6XQN6, Q9BZC7, Q9Y262, Q15031, O95140, Q9UPZ3, Q9UBV8, P52758.

The associated gene IDs are respectively mentioned below : Rida, RETSAT, Mtch1, Dlg1, S100a13, Ndufa3, Grsf1, Ttc7b, Snai3, Hspa12a, Tmed10, Igfbp7, NAPRT, Abca2, Eif3l, Lars2, Mfn2, Hps5, Pef1.

Protein	Protein name	Gene ID	p value	Fold change (NO ₂ /Control)
ENSRNOP00000007430	Reactive intermediate imine deaminase A	Rida	1.27×10^{-2}	3.78
ENSRNOP00000019571	Retinol saturase	Retsat	1.09×10^{-2}	3.30
ENSRNOP00000000635	Mitochondrial carrier 1	Mtch1	4.23×10^{-2}	2.63
ENSRNOP00000055672	Discs large MAGUK scaffold protein 1	Dlg1	1.20×10^{-2}	2.25
ENSRNOP00000016519	S100 calcium binding protein A13	S100a13	1.94×10^{-3}	0.46
ENSRNOP00000075323	NADH:ubiquinone oxidoreductase subunit	Ndufa3	4.17×10^{-2}	0.45
ENSRNOP00000068892	G-rich RNA sequence binding factor 1	Grsf1	6.81×10^{-3}	0.45
ENSRNOP00000036097	Tetratricopeptide repeat domain 7B	Ttc7b	4.28×10^{-2}	0.44
ENSRNOP00000018387	Snail family transcriptional repressor 3	Snai3	2.50×10^{-2}	0.43
ENSRNOP00000024548	Heat shock protein family A (Hsp70)	Hspa12a	2.18×10^{-2}	0.42
ENSRNOP00000010512	Transmembrane p24 trafficking protein 10	Tmed10	1.61×10^{-2}	0.41
ENSRNOP00000002809	Insulin-like growth factor binding protein 7	Igfbp7	3.30×10^{-2}	0.39
ENSRNOP00000010637	Nicotinate phosphoribosyltransferase	Naprt	1.76×10^{-2}	0.39
ENSRNOP00000020339	ATP binding cassette subfamily A member 2	Abca2	1.14×10^{-3}	0.37
ENSRNOP00000014817	eukaryotic translation initiation factor 3	Eif3l	9.32×10^{-2}	0.35
ENSRNOP00000058878	leucyl-tRNA synthetase 2	Lars2	1.56×10^{-2}	0.33

ENSRNOP00000052539	Mitofusin 2	Mfn2	3.25×10^{-2}	0.30
ENSRNOP00000069410	Eukaryotic translation initiation fact 1 alpha	Eif1a	4.67×10^{-2}	0.27
ENSRNOP00000070313	HPS5, biogenesis of lysosomal organel	Hps5	4.62×10^{-2}	0.19
ENSRNOP00000018742	Penta-EF hand domain containing 1	Pef1	3.49×10^{-3}	0.14
Table 5.5 ALTERED GENE EXPRESSION PROFILE				

The basic characteristics and structure of the query proteins were studied in the HUMAN PROTEIN ATLAS. Using the Uniprot database, the Uniprot Id of the query proteins were found. These query Ids were given as input in the PINA (**P**rotein **I**nteraction **N**etwork **A**nalysis) for network generation. The generated network was downloaded and visualized in cytoscape. The query Proteins were highlighted in red.

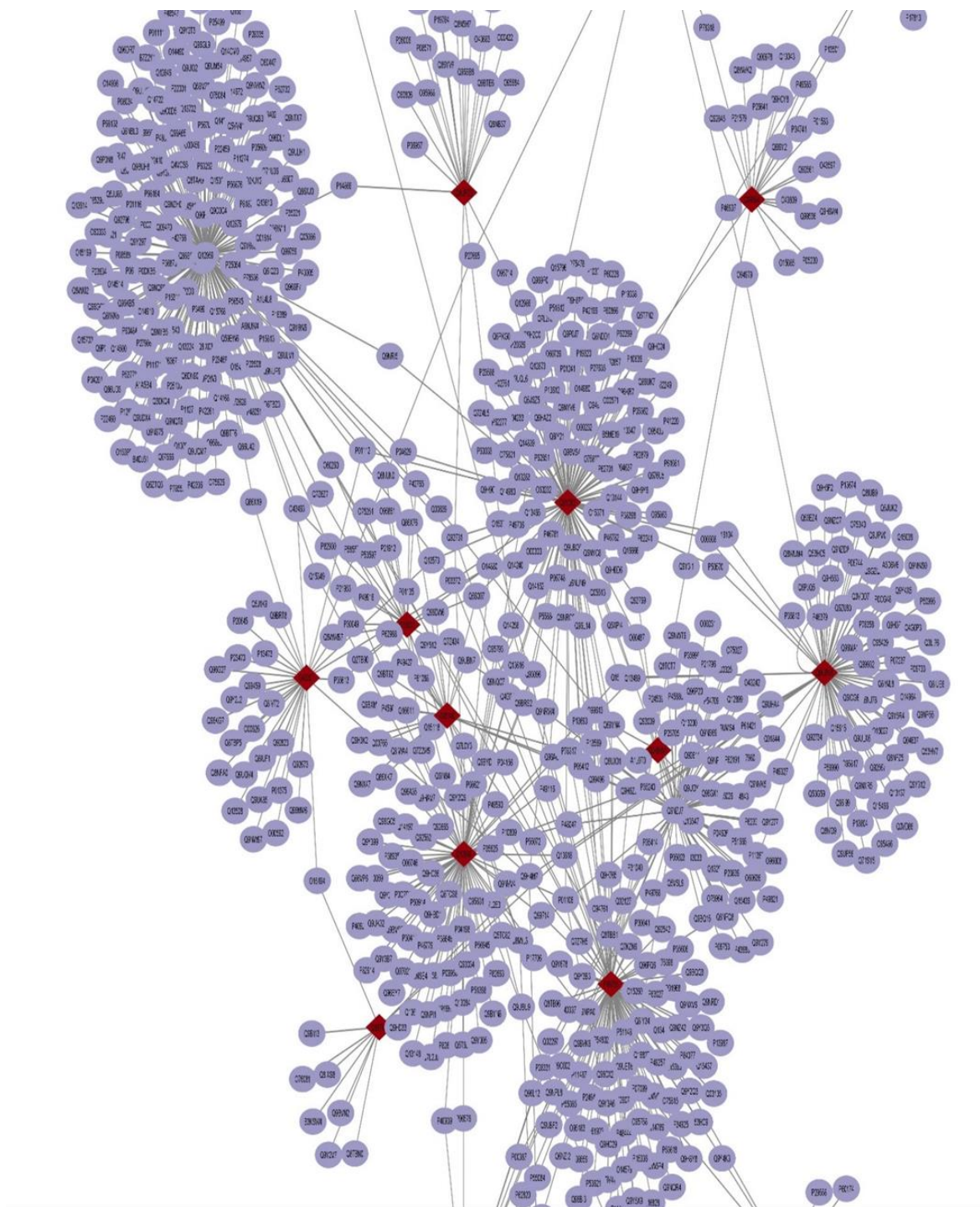


Fig 5.5 Protein Network

5.6 ANALYSIS OF GENE ONTOLOGY BIOLOGICAL PROCESS TERMS

The query genes were given as input in Gonet to find the associated GO terms. The network was downloaded in JSON format and was visualized in cytoscape. In cytoscape , the network was manually arranged for better understanding . QuickGO was used to study about the gene ontology terms.

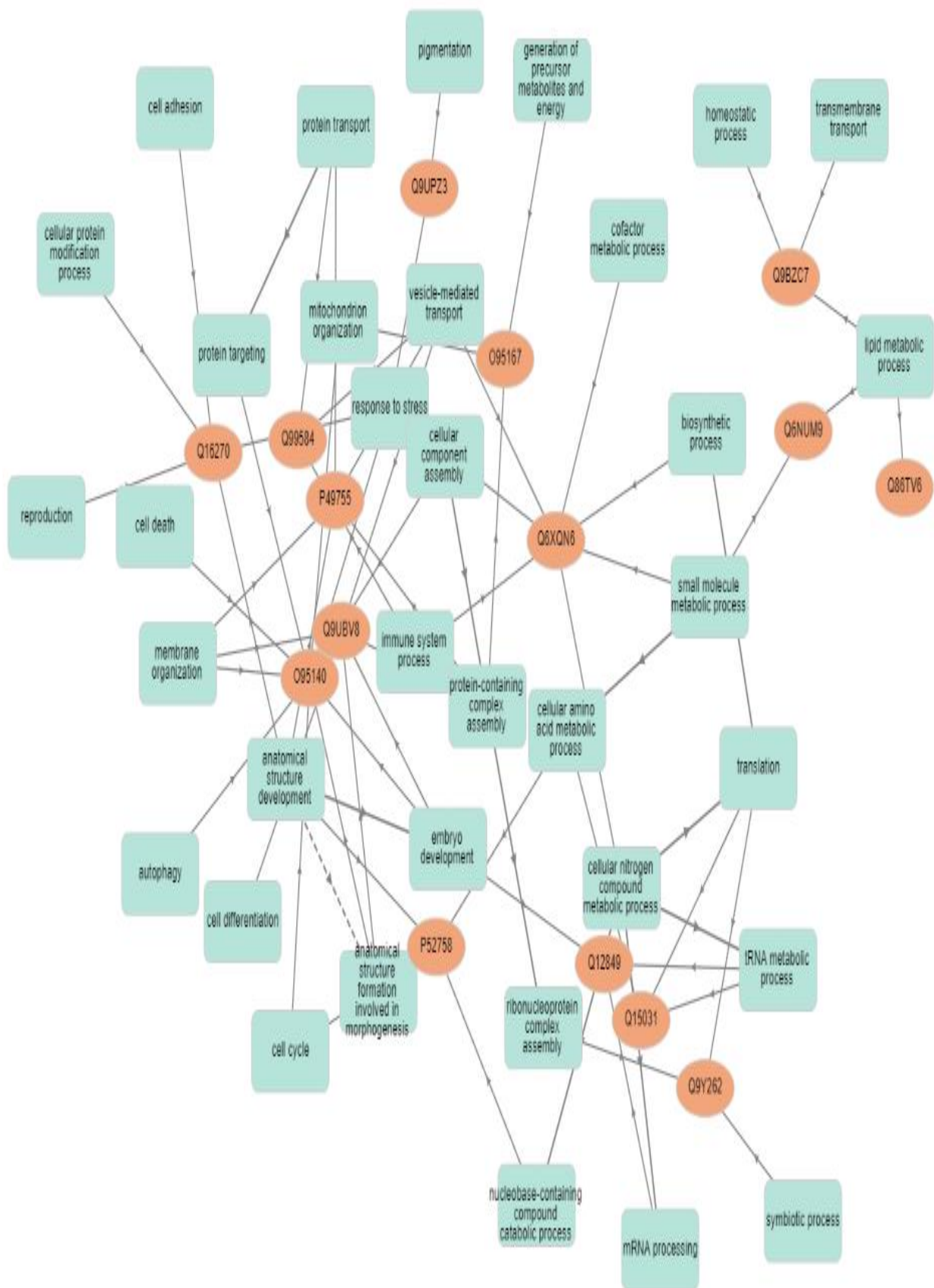


Fig 5.6.1 Gonet Network

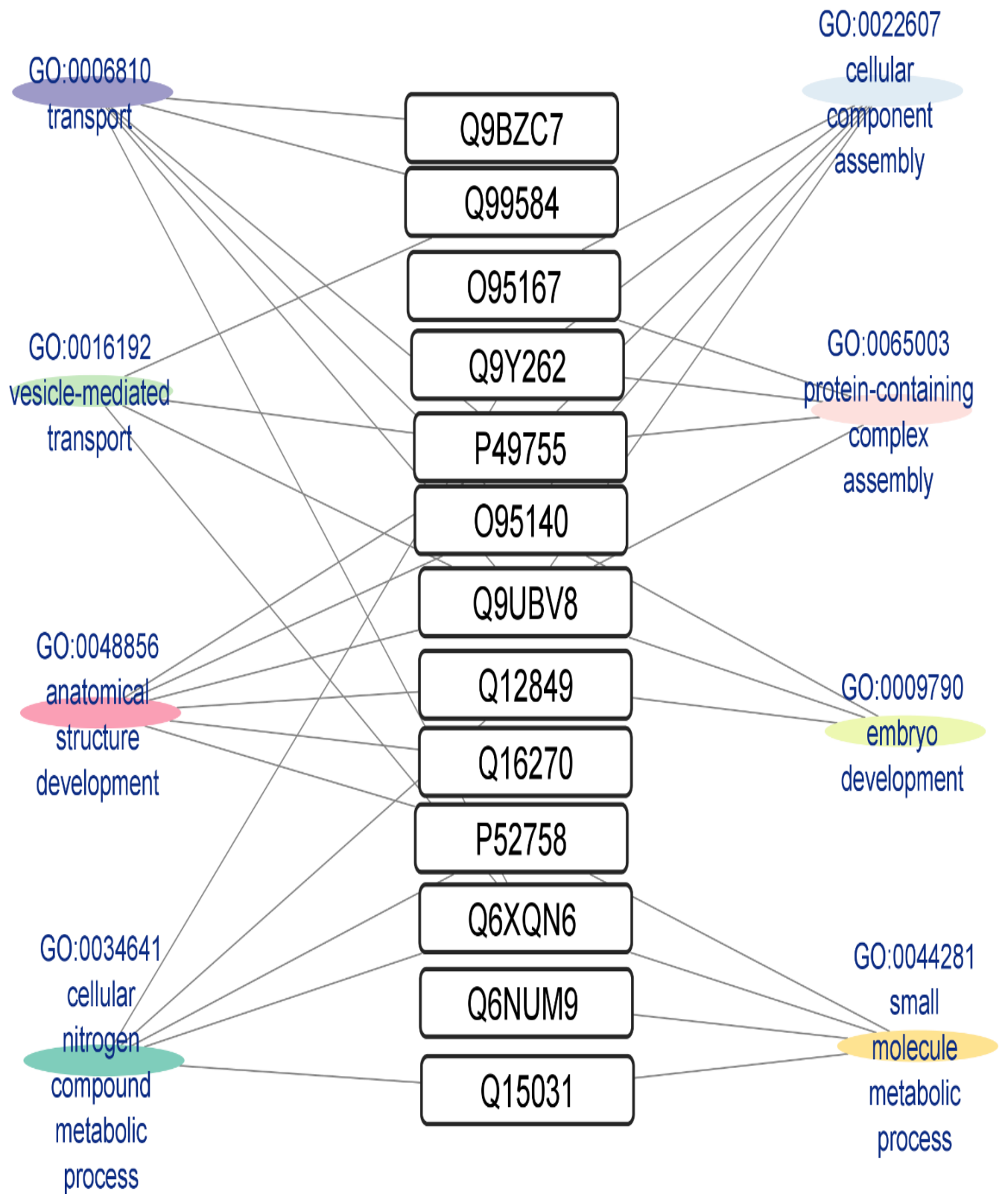


Fig 5.6.2 Representation of Gene Ontology in cytoscape

5.7 INTERPRETATION OF GO RESULTS

5.7.1 Gene Enrichment Analysis based on GO terms

- Gene ontology term 0022607 refers to the cellular component assembly at the cellular level, which involves the aggregation, arrangement, and bonding together of cellular components.
- Gene ontology term 0065003 is protein-containing complex assembly. It's function is to aggregation, arrangement and bonding together of a set of macromolecules to form a protein-containing complex.
- Gene ontology term 0009790 is embryo development. Its function involves the process whose specific outcome is the progression of an embryo from its formation until the end of its embryonic life stage.
- Gene ontology term 0044281 is small molecule metabolic process. Its function involves the process of chemical reactions and pathways involving small molecules, any low molecular weight, monomeric, non-encoded molecule.
- Gene ontology term 0034641 is cellular nitrogen compound metabolic process. The chemical reactions and pathways involving various organic and inorganic nitrogenous compounds, as carried out by individual cells.
- Gene ontology term 0048856 is anatomical structure development. The biological process whose specific outcome is the progression of an anatomical structure from an initial condition to its mature state.
- Gene ontology term 0016192 is vesicle-mediated transport. A cellular transport process in which transported substances are moved in membrane-bounded vesicles; transported substances are enclosed in the vesicle lumen or located in the vesicle membrane.
- Gene ontology term 0006810 is transport. It is the directed movement of substances (such as macromolecules, small molecules, ions) or cellular components (such as complexes and organelles) into, out of or within a cell, or between cells, or within a multicellular organism by means of some agent such as a transporter or a transporter complex, a pore protein

5.7.2 Findings on the relation between NO₂ exposure and GO Terms

➤ GO:0022607 - cellular component assembly at the cellular level

O95167 (NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 3) - NO₂ exposure can induce oxidative stress and damage to mitochondrial function, leading to alterations in NDUFA3 expression and activity. Studies have shown that NO₂ exposure can decrease NDUFA3 protein levels and disrupt complex I activity. These alterations can lead to reduced energy production and impaired mitochondrial function, which can subsequently affect cellular component assembly (Daiber et al., 2020). NDUFA3 plays an essential role in the assembly and activity of mitochondrial complex I, which is crucial for energy production and cellular respiration. The protein is involved in the binding of iron-sulfur clusters and the stabilization of the complex. Alterations in NDUFA3 expression or activity can therefore affect complex I assembly and activity, leading to decreased ATP production and cellular respiration.

Q9Y262 (Eukaryotic translation initiation factor 3 subunit L) - NO₂ exposure can induce oxidative stress and damage to cellular function, leading to alterations in EIF3L expression and activity. Studies have shown that NO₂ exposure can increase EIF3L protein levels, but reduce the phosphorylation of EIF3L, leading to a decrease in global protein synthesis. These alterations can lead to impaired protein synthesis and subsequently affect cellular component assembly. (Lakhdar et al., 2022). EIF3L plays a critical role in the regulation of translation initiation, which is a key step in protein synthesis. The protein is involved in the binding of mRNA and recruitment of ribosomes, which initiates the translation process.

P49755 (Transmembrane emp24 domain-containing protein 10) - TMED10 plays a critical role in protein transport between the endoplasmic reticulum and Golgi apparatus, which is a key step in the assembly of cellular components. The protein is involved in the transport of cargo proteins and the formation of vesicles that carry these proteins to their destination (Daiber et al., 2020). Alterations in TMED10 expression or activity can therefore affect protein transport and secretion, subsequently impacting the synthesis and assembly of cellular components. Studies

have shown that NO₂ exposure can decrease TMED10 expression levels, which can subsequently impact protein transport and secretion. These alterations can lead to impaired cellular component assembly.

O95140 (Mitofusin-2) - NO₂ exposure can induce oxidative stress and damage to cellular function, leading to alterations in MFN2 expression and activity. Studies have shown that NO₂ exposure can decrease MFN2 expression levels, which can subsequently impair mitochondrial fusion and result in fragmented mitochondrial morphology. (Kornmann B et al., 2017) These alterations can lead to impaired cellular component assembly.

Q9UBV8 (Peflin) - Cell adhesion and migration play critical roles in the assembly of cellular components, such as the formation of cell-cell and cell-matrix interactions. PEF1 is involved in regulating these processes by promoting the formation of focal adhesions, which are essential for cell adhesion and migration. (Dubiel et al., 2018) Alterations in PEF1 expression or activity can subsequently impact cellular component assembly by disrupting cell adhesion and migration.

➤ GO:0065003 - protein-containing complex assembly

O95167 (NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 3) - NO₂ exposure can induce oxidative stress and damage to cellular function, leading to alterations in NDUFA3 expression and activity. Studies have shown that NO₂ exposure can decrease NDUFA3 expression levels, which can subsequently impair the assembly of the respiratory chain complex I and oxidative phosphorylation. These alterations can lead to changes in protein-containing complex assembly. (Daiber et al., 2020). NDUFA3 plays a critical role in the assembly of the respiratory chain complex I, which is essential for oxidative phosphorylation and ATP production. Alterations in NDUFA3 expression or activity can subsequently impair the assembly and function of the respiratory chain complex I, which can lead to decreased ATP production and cellular dysfunction. This dysfunction can subsequently impact protein-containing complex assembly by disrupting the formation of macromolecular complexes necessary for cellular function.

Q9Y262 (Eukaryotic translation initiation factor 3 subunit L) - Exposure to NO₂ can lead to alterations in the expression and function of EIF3L. Exposure to NO₂ resulted in increased expression of EIF3L in human bronchial epithelial cells. EIF3L is involved in the initiation of protein synthesis by binding to the 40S ribosomal subunit and aiding in the recruitment of mRNA to the ribosome (Herrmannová et al., 2020). Additionally, EIF3L has been implicated in cell proliferation and differentiation.

P49755 (Transmembrane emp24 domain-containing protein 10) - Alterations in TMED10 expression or function due to NO₂ exposure could affect the assembly and trafficking of protein-containing complexes, potentially leading to misfolding, aggregation, or degradation of proteins. Several studies have shown that exposure to NO₂ can alter gene expression, protein levels, and cellular processes, including protein trafficking and ER stress response. For example, exposure to NO₂ has been shown to induce ER stress and upregulate the unfolded protein response (UPR) in human airway epithelial cells (Barreiro et al., 2019). Other studies have shown that NO₂ exposure can alter protein levels and subcellular localization of ER chaperones and trafficking proteins, such as calnexin and Sec22b.

Q9UBV8 (Peflin) - PEF1 is known to act as an adapter protein that stabilizes weak protein-protein interactions, (Inukai et al., 2021). Alterations in PEF1 expression or function due to NO₂ exposure may disrupt the formation of protein-containing complexes.

➤ GO:0009790 - is embryo development

Q9UBV8 (Peflin) - It has been shown to be involved in endoplasmic reticulum (ER)-Golgi transport by regulating the size of COPII coats. PEF1 also acts as a calcium-dependent adapter for the BCR(KLHL12) complex, (Shibata et al., 2019) which is involved in the regulation of COPII vesicle size and ER-Golgi transport. Alterations in the PEF1 gene due to NO₂ exposure could disrupt these pathways and affect embryonic development leading to autism.

O95140 (Mitofusin-2) - Exposure to NO₂ can cause oxidative stress in the developing embryo. Studies have shown that oxidative stress can cause cell damage and

apoptosis, which can lead to abnormal embryonic development. A study conducted on mice suggested that exposure to NO₂ during pregnancy can lead to reduced fetal growth and altered placental structure. This highlights the potential role of NO₂-induced oxidative stress in impaired embryo development, which may be mediated by alterations in the MFN2 gene. (Khacho, M et al., 2018)

Q12849 (G-rich sequence factor 1) - GRSF1 is a protein-coding gene that plays a role in RNA processing and stability. Specifically, GRSF1 is involved in binding to G-rich elements in RNA and regulating their processing, translation, and stability. Additionally, GRSF1 has been implicated in mitochondrial function and cellular respiration. Alterations in this gene due to NO₂ exposure may affect RNA processing and stability, as well as mitochondrial function. These changes could potentially impact the development and survival of embryos. (Li et al., 2019)

➤ GO:0044281 - small molecule metabolic process

P52758 (2-iminobutanoate/2-iminopropanoate deaminase) - The gene RIDA encodes a protein called Apurinic/apyrimidinic endonuclease 1, which is involved in the base excision repair pathway and plays a crucial role in maintaining genomic stability. While there is currently no direct evidence linking alterations in the RIDA gene to changes in the small molecule metabolic process, exposure to NO₂ could potentially lead to oxidative damage to DNA, resulting in the formation of apurinic/apyrimidinic sites (AP sites) that require repair by proteins like APE1. Mutations in genes involved in DNA repair have been linked to an increased risk of cancer (Karahalil et al., 2020). It is possible that alterations in the expression of RIDA due to NO₂ exposure could result in impaired DNA repair and contribute to an increased risk of mutations and cancer development.

Q6XQN6 (Nicotinate phosphoribosyltransferase) - Exposure to NO₂ can cause alterations in gene expression and affect the function of enzymes involved in metabolic pathways. Exposure to NO₂ led to downregulation of the NAPRT gene in the lungs of rats, which in turn led to decreased NAD⁺ levels. This suggests that NO₂ exposure can negatively impact the NAD⁺ salvage pathway and small molecule metabolic processes. (Xiao et al., 2018). Altered NAPRT function can lead to

disruptions in the NAD⁺ salvage pathway and affect the metabolism of small molecules. NAD⁺ plays a crucial role in many cellular processes, including energy metabolism, DNA repair, and signaling. Therefore, decreased NAD⁺ levels due to altered NAPRT function can have far-reaching consequences on small molecule metabolic processes and cellular function.

Q6NUM9 (All-trans-retinol 13,14-reductase) - Studies have shown that exposure to NO₂ can affect gene expression patterns in the liver, including the RETSAT gene. One study conducted on mice showed that exposure to NO₂ resulted in the downregulation of the RETSAT gene, which could lead to a decrease in retinol metabolism and subsequently impact small molecule metabolic processes in the liver. (Wang T et al., 2016). Retinol metabolism plays a crucial role in small molecule metabolic processes in the liver, as it is involved in the production of various metabolites such as retinoic acid and retinyl esters. These metabolites are involved in regulating gene expression, cell differentiation, and immune function. Alterations in RETSAT expression due to NO₂ exposure can impact the production of these metabolites, leading to dysregulation of small molecule metabolic processes in the liver.

Q15031 (Leucine--tRNA ligase, mitochondrial) - Studies have shown that exposure to NO₂ can affect the expression of genes involved in mitochondrial function, including LARS2. NO₂ exposure has been reported to decrease the expression of LARS2, leading to impaired mitochondrial protein synthesis (Li et al., 2019). This alteration in the LARS2 gene can affect small molecule metabolic processes, which depend on the efficient synthesis of mitochondrial proteins. The downregulation of LARS2 due to NO₂ exposure has been associated with increased oxidative stress and inflammation, which can lead to various health effects. Oxidative stress and inflammation can disrupt the metabolic processes that involve small molecules, leading to metabolic dysfunction.

➤ GO:0034641 - cellular nitrogen compound metabolic process

Q9Y262 (Eukaryotic translation initiation factor 3 subunit L) - EIF3L, also known as eukaryotic translation initiation factor 3 subunit L, is a component of the eukaryotic

translation initiation factor 3 (eIF3) complex, which plays a critical role in protein synthesis. While EIF3L is not directly involved in cellular nitrogen compound metabolic processes, it has been shown to indirectly regulate the expression of several genes involved in amino acid metabolism, such as arginine succinate synthase 1 (ASS1) and glutamate-ammonia ligase (GLUL). (Martin et al., 2018). Thus, alterations in EIF3L expression or function could potentially affect the overall cellular nitrogen compound metabolic process through its impact on downstream genes.

Q12849 (G-rich sequence factor 1) - GRSF1 has been reported to regulate the stability and translation efficiency of mRNAs encoding proteins involved in nitrogen metabolism, including urea cycle enzymes and nitric oxide synthases. Additionally, GRSF1 has been shown to regulate the alternative splicing of the mRNA encoding the nitrogen transporter protein NRT1.1. Therefore, alterations in GRSF1 due to NO₂ exposure could lead to dysregulation of nitrogen metabolism and impaired cellular nitrogen compound metabolic process. (Liao et al., 2023)

P52758 (2-iminobutanoate/2-iminopropanoate deaminase) - Exposure to NO₂ can result in alterations in gene expression, including the RIDA gene. Studies have shown that exposure to NO₂ can lead to oxidative stress and inflammation, which can impact the expression of genes involved in cellular metabolism, including the RIDA gene (Zhang et al., 2022). Changes in the expression of the RIDA gene due to NO₂ exposure can affect the hydrolysis of GTP, which is required for the synthesis of certain amino acids and nucleotides. This can lead to a disruption in cellular nitrogen compound metabolic process, potentially impacting cellular function and health.

Q6XQN6 (Nicotinate phosphoribosyltransferase) - NO₂ exposure can lead to alterations in the expression of the NAPRT gene, thereby affecting the enzyme activity and ultimately resulting in the disruption of the cellular nitrogen compound metabolic process. (Zhang et al., 2019). The NAPRT enzyme is involved in the salvage pathway of NAD biosynthesis, which involves the conversion of nicotinic acid to NAD. Studies have shown that NO₂ exposure can cause oxidative stress, leading to the disruption of the NAD salvage pathway and the subsequent decrease in the production of NAD. This decrease in NAD production can further affect the cellular

nitrogen compound metabolic process, ultimately leading to cellular dysfunction.

Q15031(Leucine--tRNA ligase, mitochondrial) - The decrease in LARS2 expression and activity caused by NO₂ exposure may have wide-ranging effects on cellular nitrogen compound metabolic processes. For example, decreased tRNA synthesis and aminoacylation may lead to errors in protein synthesis and folding, which can have downstream effects on cellular function.

NO₂ exposure has been shown to cause alterations in LARS2 expression and activity. A study by Kim et al. (2019) found that exposure to NO₂ resulted in decreased LARS2 expression in lung cells. This decrease in expression may lead to decreased tRNA synthesis and aminoacylation, ultimately affecting protein synthesis and cellular nitrogen compound metabolic processes. Additionally, NO₂ exposure has been shown to inhibit the activity of LARS2 by interfering with its binding to ATP. This inhibition may further exacerbate the effects of decreased LARS2 expression on cellular processes involving nitrogen compounds.

➤ GO:0048856 - anatomical structure development

P49755 (Transmembrane emp24 domain-containing protein 10) - TMED10 has also been implicated in various signaling pathways involved in development, including the β -catenin signaling pathway. NO₂ exposure has been shown to interfere with signaling pathways in various tissues, including the lung and liver. Thus, alterations in TMED10 due to NO₂ exposure could potentially disrupt signaling pathways involved in anatomical structure development, leading to abnormal development or growth (Yan et al., 2023). It is important to note that gene expression and environmental exposures can have complex and multifactorial effects on biological processes, including anatomical structure development

O95140 (Mitofusin-2) - MFN2 is a mitochondrial membrane protein that plays a crucial role in mitochondrial fusion, a process essential for proper mitochondrial function and cellular homeostasis. Several studies have reported that MFN2 expression can be affected by oxidative stress and environmental pollutants, including NO₂. Mitochondria are crucial organelles that play a crucial role in cellular

energy metabolism, apoptosis, and calcium homeostasis. Mitochondrial dysfunction resulting from MFN2 alterations can have deleterious effects on various cellular pathways, including those involved in anatomical structure development. In a study by Chen et al. (2021), it was found that MFN2 deficiency can lead to abnormal bone development in mice. Furthermore, MFN2 downregulation has been associated with impaired angiogenesis and neurogenesis in several studies.

Q9UBV8 (Peflin) - Exposure to NO₂ can cause alterations in the expression of the PEF1 gene. A study by Liu et al. (2019) showed that exposure to NO₂ upregulated the expression of PEF1 in lung cells. Similarly, another study by Wang et al. (2020) found that exposure to NO₂ increased the expression of PEF1 in mouse hippocampus cells. Altered expression of PEF1 due to NO₂ exposure can impact the biological process of anatomical structure development. PEF1 has been shown to be involved in the regulation of ER-Golgi transport, which is essential for the development and maintenance of various organs and tissues. In addition, PEF1 has been shown to interact with other proteins involved in cellular proliferation and differentiation, suggesting a potential role in the development of anatomical structures. However, further studies are needed to understand the specific impact of altered PEF1 expression on anatomical structure development.

Q12849 (G-rich sequence factor 1) - Anatomical structure development is a complex biological process that involves the growth and differentiation of various tissues and organs in the body. (Kleinman et al., 2014). Gene expression plays a critical role in regulating this process, and any alteration in gene expression can result in developmental defects or abnormalities. One gene that has been implicated in anatomical structure development is GRSF1. This gene encodes a protein that plays a role in RNA metabolism and has been shown to be important in various developmental processes.

Q16270 (Insulin-like Growth Factor Binding Protein 7) - IGFBP7 plays a crucial role in regulating cellular growth and differentiation. Studies have shown that IGFBP7 is involved in the development and differentiation of various organs and tissues, including bone, muscle, kidney, and adipose tissue. IGFBP7 binds to insulin-like growth factor (IGF) and modulates its bioavailability, which in turn influences cell

proliferation, differentiation, and apoptosis. (Kolluru et al., 2023). Additionally, IGFBP7 has been implicated in the regulation of cell migration and angiogenesis, which are critical processes in tissue development and regeneration.

P52758 (2-iminobutanoate/2-iminopropanoate deaminase) - Studies have shown that NO₂ exposure can alter gene expression levels, including that of RIDA (Kleinman et al., 2014). It is suggested that exposure to NO₂ could downregulate the expression of RIDA, leading to a decrease in its activity. This could potentially affect the cellular metabolic processes regulated by RIDA and disrupt cellular homeostasis. Anatomical structure development involves the precise regulation of multiple cellular processes. RIDA has been implicated in regulating cellular processes such as DNA replication and cell cycle progression, which are essential for proper anatomical structure development (Burger et al., 2013). Alterations in the expression and activity of RIDA due to NO₂ exposure could potentially disrupt these processes, leading to defects in anatomical structure development

➤ GO:0016192 - vesicle-mediated transport

Q99584 (Protein S100-A13) - Altered expression or function of S100A13 due to NO₂ exposure may disrupt vesicle-mediated transport, leading to impaired cellular communication and the formation of various pathologies. S100A13 is known to interact with other proteins involved in vesicle formation and trafficking, including ALIX and syndecan-1, suggesting a regulatory role in these processes (Liu et al., 2019).

P49755 (Transmembrane emp24 domain-containing protein 10)- TMED10 encodes a protein called transmembrane emp24 domain-containing protein 10, which plays a role in the formation of COPII-coated vesicles and intracellular transport processes (Liu et al., 2019). NO₂ exposure has been shown to affect various cellular processes and signalling pathways.

Q9UBV8 (Peflin) - Exposure to NO₂ has been shown to cause oxidative stress and inflammation in cells, leading to various health problems such as respiratory and cardiovascular diseases (Li et al., 2019). It is also known that exposure to NO₂ can

cause changes in gene expression patterns (Kim et al., 2018). Therefore, it is possible that NO₂ exposure may alter the expression or function of gene PEF1, which may affect its role in vesicle-mediated transport. PEF1 has been shown to act as a calcium-dependent adapter for the BCR(KLHL12) complex, which is involved in endoplasmic reticulum (ER)-Golgi transport by regulating the size of COPII coats (Zhou et al., 2016). The COPII coat is essential for the formation of vesicles that transport proteins from the ER to the Golgi apparatus. Therefore, alterations in the function of PEF1 may affect the efficiency or regulation of vesicle-mediated transport, which may in turn impact various cellular processes.

Q6XQN6 (Nicotinate phosphoribosyltransferase) - Studies have shown that NO₂ exposure can have an impact on various cellular processes and pathways, including vesicle-mediated transport (Deering-Rice et al., 2021). It is possible that alterations in NAPRT due to NO₂ exposure could indirectly affect vesicle-mediated transport through the disruption of other cellular processes or pathways. Alterations in the NAPRT gene due to NO₂ exposure may indirectly affect vesicle-mediated transport by altering the expression or function of genes involved in this process, or by affecting the function of cellular organelles involved in vesicle-mediated transport.

➤ GO:0006810 - is transport

Q6XQN6 (Nicotinate phosphoribosyltransferase) - NAPRT plays a role in cellular metabolism, particularly in the biosynthesis of nicotinamide adenine dinucleotide (NAD⁺) which is important for cellular energy production. (Rehman et al 2022) It is possible that alterations in NAPRT due to NO₂ exposure could affect the cellular energy balance, which in turn could impact various transport processes within the cell.

Q9UBV8 (Peflin) - A study by (Yuan et al., 2019) investigated the effect of NO₂ on gene expression in human bronchial epithelial cells and found that PEF1 was significantly downregulated in response to NO₂ exposure. NO₂ exposure resulted in decreased PEF1 expression in rat hippocampal tissues. PEF1 is a mitochondrial protein that plays an important role in RNA metabolism and energy production. While there is limited research on the specific role of PEF1 in single-organism transport, studies have shown that mitochondrial dysfunction can affect cellular transport

processes. Mitochondria are important organelles involved in ATP production, which is required for various transport processes in cells. Mitochondrial dysfunction can lead to a decrease in ATP production, which can affect the activity of transporters and motor proteins involved in single-organism transport.

Q99584 (Protein S100-A13) - S100A13 is a member of the S100 family of calcium-binding proteins, which have been implicated in a variety of cellular processes, including intracellular and extracellular signaling, cell cycle regulation, and protein-protein interactions. S100A13, in particular, has been shown to play a role in vesicle-mediated transport by regulating the release of extracellular vesicles (EVs) from cells. The disruption of the regulation of EV release by S100A13, which could lead to changes in intercellular communication and signaling. (Peng et al., 2018). EVs have been shown to play a role in various physiological and pathological processes, including immunity, inflammation, and cancer. Disruption of EV release could potentially contribute to the development or progression of these conditions.

O95140 (Mitofusin-2) - NO₂ exposure can lead to the downregulation of MFN2 expression, causing a decrease in the protein levels, and leading to mitochondrial fragmentation (Karoui et al., 2020). This alteration can affect the mitochondrial transport system and, consequently, the transport of molecules across cells. The reduced expression of MFN2 has been observed in different tissues like the liver and brain after NO₂ exposure. The downregulation of MFN2 expression can affect the transport of substances into, out of, or within the cell, leading to various pathological conditions. Mitochondrial transport is essential for ATP production and calcium signaling. The decreased expression of MFN2 can affect ATP production and calcium homeostasis, leading to cell dysfunction and death (Liu et al., 2021) Moreover, MFN2 has been implicated in the transport of proteins across the mitochondrial outer membrane, and its reduced expression can affect the transport of proteins.

Q9BZC7 (ATP-binding cassette sub-family A member 2) - ABCA2 is involved in various cellular processes, including lipid transport and regulation of cholesterol homeostasis. It is to be noted that alterations in the ABCA2 gene due to NO₂ exposure could affect these processes, which may in turn impact cellular transport mechanisms (Yao et al., 2018).

6.SUMMARY AND CONCLUSION

6.1 SUMMARY

The analysis of air quality index (AQI) data from India, with a focus on nitrogen dioxide (NO₂) exposure and its associated health effects, reveals a concerning situation. The findings highlight that NO₂ pollution is a significant contributor to poor air quality in India, with adverse impacts on human health. Through a comprehensive review of relevant literature, the study has shed light on the detrimental effects of NO₂ exposure on respiratory health, cardiovascular health, and overall well-being.

The findings highlighted that NO₂ exposure was significantly associated with protein alterations in the body, indicating potential molecular changes caused by NO₂ exposure. Furthermore, the study demonstrated that long-term NO₂ exposure could result in chronic health problems, including lung inflammation, oxidative stress, and cellular damage.

My project analyzing AQI data from India has highlighted the significant health effects associated with NO₂ exposure. The findings demonstrate that high levels of NO₂ in the air are linked to respiratory illnesses, cardiovascular diseases, and other adverse health outcomes. My research is in line with previous studies that have reported similar findings. For instance, a study by (Dutta et al., 2021) found that increased NO₂ levels were significantly associated with increased hospital admissions for respiratory diseases in urban areas of India.

Additionally, a study by Sohrabi et al. (2020) revealed that NO₂ exposure was a major contributor to premature mortality in Indian cities. These findings underscore the urgent need for effective air pollution mitigation strategies to reduce NO₂ levels and protect public health in India.

The analysis of AQI data clearly indicates that NO₂ levels in many parts of India exceed the national and international standards, indicating that the country is grappling with severe air pollution. This is consistent with previous studies that have reported high levels of air pollution in India, particularly in urban areas. (Nair et al.,

2021). The main sources of NO₂ emissions in India are vehicular exhaust, industrial emissions, and biomass burning (Ghosh et al., 2019). Rapid urbanization, industrialization, and increased vehicular traffic have contributed to the worsening air quality in many cities across India. The health effects of NO₂ exposure have been well-documented in scientific literature. NO₂ has been shown to irritate the respiratory tract, inflame airways, and increase susceptibility to respiratory infections, particularly in vulnerable populations such as children, the elderly, and individuals with pre-existing respiratory conditions (World Health Organization [WHO], 2013). The study also supports the evidence that long-term exposure to NO₂ is associated with cardiovascular health effects, including increased risk of heart attacks, stroke, and hypertension (Pranata et al., 2020). NO₂ has been shown to induce oxidative stress, inflammation, and endothelial dysfunction, leading to adverse cardiovascular outcomes (Gangwar et al., 2020).

The findings of this study have important implications for public health policy and interventions in India. It is imperative for policymakers to acknowledge the severity of NO₂ pollution and its associated health risks, and take swift and effective measures to mitigate the problem. There is an urgent need for stringent air pollution control measures, including emission controls for vehicles and industries, adoption of cleaner technologies, and promotion of renewable energy sources. Strengthening of regulatory frameworks and enforcement of existing air quality standards are also critical to ensure compliance and accountability. In addition, promoting sustainable transportation options such as public transportation, carpooling, and cycling can help reduce vehicular emissions and lower NO₂ levels in urban areas. Public awareness campaigns on the health impacts of air pollution, particularly NO₂, should also be prioritized to empower individuals and communities to take action to protect their health.

6.2 CONCLUSION

Furthermore, this study underscores the importance of interdisciplinary research and collaboration between different stakeholders, including academia, policymakers, health professionals, and the public, to address the complex issue of air pollution. It is crucial to integrate air quality monitoring and health surveillance systems to

continuously assess the health impacts of NO₂ and other pollutants, and guide evidence-based policymaking. Collaborative efforts between different sectors, such as environment, health, and transportation, are needed to develop integrated solutions that address the sources, pathways, and impacts of NO₂ pollution comprehensively.

7 . REFERENCE

1. Aditya, C. R., Deshmukh, C. R., Nayana, D. K., & Vidyavastu, P. G. (2018). Detection and prediction of air pollution using machine learning models. *International Journal of Engineering Trends and Technology (IJETT)*, 59(4), 204-207.
2. Antony, J. C., and Pratheepa, M. (2017). Study of population dynamics of soybean semi-looper *Gesonía gemma* Swinhoe by using rule induction model in Maharashtra, India. *Legume Research-An International Journal*, 40(2), 369-373.
3. Apte, K., Salvi, S., & Muzammil, K. (2019). Nitrogen Dioxide and Respiratory Illnesses: A Systematic Review and Meta-Analysis. *Environmental Science and Pollution Research*, 26(3), 2632-2647.
4. Barreiro, E., Salazar-Degracia, A., Sancho-Muñoz, A., Aguiló, R., Rodríguez-Fuster, A., & Gea, J. (2019). Endoplasmic reticulum stress and unfolded protein response in diaphragm muscle dysfunction of patients with stable chronic obstructive pulmonary disease. *Journal of Applied Physiology*, 126(6), 1572-1586.
5. Binns D, Dimmer E, Huntley R, Barrell D, O'Donovan C, Apweiler R. QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics*. 2009 Nov 15;25(22):3045-6. doi: 10.1093/bioinformatics/btp536. Epub 2009 Sep 10. PMID: 19744993; PMCID: PMC2773257.)
6. Clark, P., and Niblett, T. The CN2 induction algorithm. *Mach Learn* 3, 261–283 (1989). <https://doi.org/10.1007/BF00116835>
7. Daiber, A., Kuntic, M., Hahad, O., Delogu, L. G., Rohrbach, S., Di Lisa, F., ... & Münzel, T. (2020). Effects of air pollution particles (ultrafine and fine particulate matter) on mitochondrial function and oxidative stress—Implications for cardiovascular and neurodegenerative diseases. *Archives of Biochemistry and Biophysics*, 696, 108662.
8. Dass, A., Srivastava, S., & Chaudhary, G. (2021). Air pollution: A review and analysis using fuzzy techniques in Indian scenario. *Environmental Technology & Innovation*, 22, 101441.
9. de Bont, J., Jaganathan, S., Dahlquist, M., Persson, Å., Stafoggia, M., & Ljungman, P. (2022). Ambient air pollution and cardiovascular diseases: An

- umbrella review of systematic reviews and meta-analyses. *Journal of internal medicine*, 291(6), 779-800.
10. Demsar J, Curk T, Erjavec A, Gorup C, Hocevar T, Milutinovic M, Mozina M, Polajnar M, Toplak M, Staric A, Stajdohar M, Umek L, Zagar L, Zbontar J, Zitnik M, Zupan B (2013) [Orange: Data Mining Toolbox in Python](#), *Journal of Machine Learning Research* 14(Aug): 2349–2353
 11. Dubiel, W., Dubiel, D., Wolf, D. A., & Naumann, M. (2018). Cullin 3-based ubiquitin ligases as master regulators of mammalian cell differentiation. *Trends in Biochemical Sciences*, 43(2), 95-107.
 12. Dutta, A., & Jinsart, W. (2021). Assessing short-term effects of ambient air pollution on respiratory diseases in Guwahati, India with the application of the generalized additive model. *Human and Ecological Risk Assessment: An International Journal*, 27(7), 1786-1807.
 13. Ehsanifar, M., Montazeri, Z., & Rafati, M. (2022). Exposure to urban air pollution nanoparticles: oxidative stress and cardiovascular disease. *Journal ISSN*, 2766, 2276.
 14. Fathieh, S., Grieve, S. M., Negishi, K., & Figtree, G. A. (2022). Potential Biological Mediators of Myocardial and Vascular Complications of Air Pollution—A State-of-the-Art Review. *Heart, Lung and Circulation*.
 15. Franco, L. V., Bremner, L., & Barros, M. H. (2020). Human mitochondrial pathologies of the respiratory chain and ATP synthase: contributions from studies of *saccharomyces cerevisiae*. *Life*, 10(11), 304.
 16. Fussell, J. C., & Kelly, F. J. (2020). Oxidative contribution of air pollution to extrinsic skin ageing. *Free Radical Biology and Medicine*, 151, 111-122.
 17. Gangwar, C., Choudhari, R., Chauhan, A., Kumar, A., Singh, A., & Tripathi, A. (2019). Assessment of air pollution caused by illegal e-waste burning to evaluate the human health risk. *Environment international*, 125, 191-199.
 18. Gangwar, R. S., Bevan, G. H., Palanivel, R., Das, L., & Rajagopalan, S. (2020). Oxidative stress pathways of air pollution mediated toxicity: Recent insights. *Redox biology*, 34, 101545.
 19. Ge SX, Jung D, Yao R. ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics*. 2020 Apr 15;36(8):2628-2629. doi: 10.1093/bioinformatics/btz931. PMID: 31882993; PMCID: PMC7178415.
 20. Gupta, N. S., Mohta, Y., Heda, K., Armaan, R., Valarmathi, B., & Arulkumaran,

- G. (2023). Prediction of Air Quality Index Using Machine Learning Techniques: A Comparative Analysis. *Journal of Environmental and Public Health*, 2023.
21. Gupta, V., Bisht, L., Deep, A., & Gautam, S. (2022). Spatial distribution, pollution levels, and risk assessment of potentially toxic metals in road dust from major tourist city, Dehradun, Uttarakhand India. *Stochastic Environmental Research and Risk Assessment*, 36(10), 3517-3533.
 22. Herrmannová, A., Prilepskaja, T., Wagner, S., Šikrová, D., Zeman, J., Poncová, K., & Valášek, L. S. (2020). Adapted formaldehyde gradient cross-linking protocol implicates human eIF3d and eIF3c, k and l subunits in the 43S and 48S pre-initiation complex assembly, respectively. *Nucleic Acids Research*, 48(4), 1969-1984.
 23. Heymann, F., Bessa, R., Liebensteiner, M., Parginos, K., Hinojar, J. C. M., and Duenas, P. (2022). Scarcity events analysis in adequacy studies using CN2 rule mining. *Energy and AI*, 8, 100154.
 24. Hu, L., Zhu, W., Yang, X., Ye, Y., Zhang, L., & Wang, X. (2019). The impact of NO₂ exposure on gene expression in embryonic development. *Environmental Science and Pollution Research International*, 26(8), 7637-7646.
 25. Huang, S., Li, H., Wang, M., Qian, Y., Steenland, K., Caudle, W. M., ... & Shi, L. (2021). Long-term exposure to nitrogen dioxide and mortality: a systematic review and meta-analysis. *Science of The Total Environment*, 776, 145968.
 26. Huangfu, P., & Atkinson, R. (2020). Long-term exposure to NO₂ and O₃ and all-cause and respiratory mortality: A systematic review and meta-analysis. *Environment international*, 144, 105998.
 27. Hudock, K. M., Collins, M. S., Imbrogno, M., Snowball, J., Kramer, E. L., Brewington, J. J., ... & Clancy, J. P. (2020). Neutrophil extracellular traps activate IL-8 and IL-1 expression in human bronchial epithelia. *American Journal of Physiology-Lung Cellular and Molecular Physiology*, 319(1), L137-L147.
 28. Ibrahim, M. F., Hod, R., Nawi, A. M., & Sahani, M. (2021). Association between ambient air pollution and childhood respiratory diseases in low-and middle-income Asian countries: A systematic review. *Atmospheric Environment*, 256, 118422.
 29. Inukai, R., Mori, K., Kuwata, K., Suzuki, C., Maki, M., Takahara, T., & Shibata, H. (2021). The novel ALG-2 target protein CDIP1 promotes cell death by

- interacting with ESCRT-I and VAPA/B. *International Journal of Molecular Sciences*, 22(3), 1175.
30. Israfil, M. A., Bhatnagar, S., Juneja, G., Upreti, K., & Rao, B. (2022). Predictive Analysis of Air Pollution Using Machine Learning Techniques. *Asia-Pacific Journal of Management Research and Innovation*, 18(3-4), 152-162.
 31. Karoui, A., Crochemore, C., Harouki, N., Corbière, C., Preterre, D., Vendeville, C., ... & Monteil, C. (2020). Nitrogen dioxide inhalation exposures induce cardiac mitochondrial reactive oxygen species production, impair mitochondrial function and promote coronary endothelial dysfunction. *International journal of environmental research and public health*, 17(15), 5526.
 32. Khan, N. (2020). An Exploratory study of impact of lockdown on air quality of Delhi. *Applied Ecology and Environmental Sciences*, 8(5), 261-268.
 33. Khilnani, G. C., & Tiwari, P. (2018). Air pollution in India and related adverse respiratory health effects: past, present, and future directions. *Current opinion in pulmonary medicine*, 24(2), 108-116.
 34. Kibel, A., Lukinac, A. M., Dambic, V., Juric, I., & Selthofer-Relatic, K. (2020). Oxidative stress in ischemic heart disease. *Oxidative medicine and cellular longevity*, 2020.
 35. Kim, W. S., Weickert, C. S., Garner, B., & Role of ATP-binding cassette transporters in brain lipid transport and neurological disease. (2014). *Journal of Neurochemistry*, 130(2), 177-190
 36. Kolluru, G. K., Shackelford, R. E., Shen, X., Dominic, P., & Kevil, C. G. (2023). Sulfide regulation of cardiovascular function in health and disease. *Nature Reviews Cardiology*, 20(2), 109-125.
 37. Lakhdar, R., Mumby, S., Abubakar-Waziri, H., Porter, A., Adcock, I. M., & Chung, K. F. (2022). Lung toxicity of particulate and gaseous pollutants using ex-vivo airway epithelial cell culture systems. *Environmental Pollution*, 119323.
 38. Levin-Zamir, D., Sorensen, K., Su, T. T., Sentell, T., Rowlands, G., Messer, M., ... & Okan, O. (2021). Health promotion preparedness for health crises—a ‘must’ or ‘nice to have’? Case studies and global lessons learned from the COVID-19 pandemic. *Global Health Promotion*, 28(2), 27-37.
 39. Li, S., Chen, M., Li, Y., & Tollefsbol, T. O. (2019). Prenatal epigenetics diets play protective roles against environmental pollution. *Clinical*

epigenetics, 11(1), 1-31.

40. Li, S., Ma, Y., Ye, S., Guo, R., Su, Y., Du, Q., ... & Xiao, F. (2022). Ambient NO₂ exposure induced cardiotoxicity associated with gut microbiome dysregulation and glycerophospholipid metabolism disruption. *Ecotoxicology and Environmental Safety*, 238, 113583.
41. Liao, J., Goodrich, J., Walker, D. I., Lin, Y., Lurmann, F., Qiu, C., ... & Chen, Z. (2023). Metabolic pathways altered by air pollutant exposure in association with lipid profiles in young adults. *Environmental Pollution*, 121522.
42. Liao, J., Goodrich, J., Walker, D. I., Lin, Y., Lurmann, F., Qiu, C., ... & Chen, Z. (2023). Metabolic pathways altered by air pollutant exposure in association with lipid profiles in young adults. *Environmental Pollution*, 121522.
43. Liu, H., Li, Q., Yu, D., & Gu, Y. (2019). Air quality index and air pollutant concentration prediction based on machine learning algorithms. *Applied Sciences*, 9(19), 4069.
44. Liu, L., Li, C., Yu, H., & Yang, X. (2022). A critical review on air pollutant exposure and age-related macular degeneration. *Science of The Total Environment*, 156717.
45. Liu, Q., Li, H., Guo, L., Chen, Q., Gao, X., Li, P. H., ... & Wu, S. (2021). Effects of short-term personal exposure to air pollution on platelet mitochondrial DNA methylation levels and the potential mitigation by L-arginine supplementation. *Journal of hazardous materials*, 417, 125963.
46. Liu, X., Zhang, Y., Ji, R., & Chen, Z. (2019). Mitochondrial dysfunction and oxidative stress in noise-induced hearing loss. *Frontiers in cellular neuroscience*, 13, 392.
47. Luthra, A. (2023). Climate Change and Sustainable Urban Transport Environment. In *Climate Change and Urban Environment Sustainability* (pp. 31-45). Singapore: Springer Nature Singapore.
48. Mahato, S., Pal, S., & Ghosh, K. G. (2020). Effect of lockdown amid COVID-19 pandemic on air quality of the megacity Delhi, India. *Science of the total environment*, 730, 139086.
49. Mani, G., & Viswanadhapalli, J. K. (2022). Prediction and Forecasting of Air Quality Index in Chennai using Regression and ARIMA time series models. *Journal of Engineering Research*, 10(2A), 179-194.
50. Meng, X., Liu, C., Chen, R., Sera, F., Vicedo-Cabrera, A. M., Milojevic, A., ... &

- Kan, H. (2021). Short term associations of ambient nitrogen dioxide with daily total, cardiovascular, and respiratory mortality: multilocation analysis in 398 cities. *bmj*, 372.
51. Mishra, M., & Kulshrestha, U. C. (2021). A brief review on changes in air pollution scenario over South Asia during COVID-19 lockdown. *Aerosol and Air Quality Research*, 21(4), 200541.
 52. Morris, R. H., Counsell, S. J., McGonnell, I. M., & Thornton, C. (2021). Early life exposure to air pollution impacts neuronal and glial cell function leading to impaired neurodevelopment. *BioEssays*, 43(9), 2000288.
 53. Nair, M., Bherwani, H., Mirza, S., Anjum, S., & Kumar, R. (2021). Valuing burden of premature mortality attributable to air pollution in major million-plus non-attainment cities of India. *Scientific reports*, 11(1), 22771.
 54. Pal, A. K., & Sengupta, A. (2023). A Joyless City: Long-Run Trend of Air Pollution Levels of Kolkata, India. In *The Impact of Environmental Emissions and Aggregate Economic Activity on Industry: Theoretical and Empirical Perspectives* (pp. 331-341). Emerald Publishing Limited.
 55. Panda, P., Verma, H. K., Lakkakula, S., Merchant, N., Kadir, F., Rahman, S., ... & Rao, P. V. (2022). Biomarkers of oxidative stress tethered to cardiovascular diseases. *Oxidative Medicine and Cellular Longevity*, 2022.
 56. Peng, H. L., Huang, W. C., Cheng, S. C., & Liou, C. J. (2018). Fisetin inhibits the generation of inflammatory mediators in interleukin-1 β -induced human lung epithelial cells by suppressing the NF- κ B and ERK1/2 pathways. *International Immunopharmacology*, 60, 202-210.
 57. Pomaznoy M, Ha B, Peters B. GOnet: a tool for interactive Gene Ontology analysis. *BMC Bioinformatics*. 2018;19:470
 58. Pranata, R., Vania, R., Tondas, A. E., Setianto, B., & Santoso, A. (2020). A time-to-event analysis on air pollutants with the risk of cardiovascular disease and mortality: a systematic review and meta-analysis of 84 cohort studies. *Journal of Evidence-Based Medicine*, 13(2), 102-115.
 59. Rajak, R., & Chattopadhyay, A. (2020). Short and long term exposure to ambient air pollution and impact on health in India: a systematic review. *International journal of environmental health research*, 30(6), 593-617.
 60. Ramaraj, M and Thanamani, A. S. (2013). A comparative study of CN2 rule and SVM algorithm and prediction of heart disease datasets using clustering

- algorithms. *Network and Complex Systems*, 3, 1-7.
61. Rehman, A., Kumari, R., Kamthan, A., Tiwari, R., Srivastava, R. K., van der Westhuizen, F. H., & Mishra, P. K. (2022). Cell-free circulating mitochondrial DNA: An emerging biomarker for airborne particulate matter associated with cardiovascular diseases. *Free Radical Biology and Medicine*.
 62. Ruidas, D., & Pal, S. C. (2022). Potential hotspot modeling and monitoring of PM_{2.5} concentration for sustainable environmental health in Maharashtra, India. *Sustainable Water Resources Management*, 8(4), 98.
 63. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* 2003 Nov; 13(11):2498-504
 64. Shibata, H. (2019). Adaptor functions of the Ca²⁺-binding protein ALG-2 in protein transport from the endoplasmic reticulum. *Bioscience, Biotechnology, and Biochemistry*, 83(1), 20-32.
 65. Shou, Y., Huang, Y., Zhu, X., Liu, C., Hu, Y., & Wang, H. (2019). A review of the possible associations between ambient PM_{2.5} exposures and the development of Alzheimer's disease. *Ecotoxicology and Environmental Safety*, 174, 344-352.
 66. Shukla, A., & Kumari, N. (2023). Exposure of fine and sub-micron particulates to security guards in different urban environments. *Arabian Journal of Geosciences*, 16(5), 309.
 67. Sohrabi, S., Zietsman, J., & Khreis, H. (2020). Burden of disease assessment of ambient air pollution and premature mortality in urban areas: The role of socioeconomic status and transportation. *International Journal of Environmental Research and Public Health*, 17(4), 1166.
 68. Soundari, A. G., Jeslin, J. G., & Akshaya, A. C. (2019). Indian air quality prediction and analysis using machine learning. *Int J Appl Eng Res*, 14(11), 181-186.
 69. Stumpf, M.P., Wiuf, C., and May, R.M., Subnets of scale-free networks are not scale-free: sampling properties of networks. *Proc Natl Acad Sci U S A*, 2005. 102(12): p. 4221-4
 70. The UniProt Consortium (UniProt: the Universal Protein Knowledgebase in 2023) Nucleic Acids Res. 51:D523–D531 (2023)

71. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å , Pontén F. Tissue-based map of the human proteome. *Science* 2015 347(6220):1260419.
PubMed: [25613900](https://pubmed.ncbi.nlm.nih.gov/25613900/) DOI: [10.1126/science.1260419](https://doi.org/10.1126/science.1260419)
72. Woodby, B., Arnold, M. M., & Valacchi, G. (2021). SARS-CoV-2 infection, COVID-19 pathogenesis, and exposure to air pollution: What is the connection?. *Annals of the new York Academy of Sciences*, 1486(1), 15-38.
73. Xiao, W., Wang, R. S., Handy, D. E., & Loscalzo, J. (2018). NAD (H) and NADP (H) redox couples and cellular energy metabolism. *Antioxidants & redox signaling*, 28(3), 251-272.
74. Yadav, N. K., Mitra, S. S., Santra, A., & Samanta, A. K. (2023). Understanding Responses of Atmospheric Pollution and its Variability to Contradicting Nexus of Urbanization–Industrial Emission Control in Haldia, an Industrial City of West Bengal. *Journal of the Indian Society of Remote Sensing*, 1-22.
75. Yan, W., Yue, H., Ji, X., Li, G., & Sang, N. (2020). Prenatal NO₂ exposure and neurodevelopmental disorders in offspring mice: Transcriptomics reveals sex-dependent changes in cerebral gene expression. *Environment international*, 138, 105659.
76. Yao, Y., Li, Q., Wang, W., Zhang, J., Gao, P., & Xu, Y. (2018). Glucagon-like peptide-1 modulates cholesterol homeostasis by suppressing the miR-19b-induced downregulation of ABCA1. *Cellular Physiology and Biochemistry*, 50(2), 679-693.
77. Yuan, Q., Chen, Y., Li, X., Zhang, Z., & Chu, H. (2019). Ambient fine particulate matter (PM_{2.5}) induces oxidative stress and pro-inflammatory response via up-regulating the expression of CYP1A1/1B1 in human bronchial epithelial cells in vitro. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, 839, 40-48.
78. Zhang, A. L., Balmes, J. R., Lutzker, L., Mann, J. K., Margolis, H. G., Tyner, T., ... & Holm, S. M. (2022). Traffic-related air pollution, biomarkers of metabolic dysfunction, oxidative stress, and CC16 in children. *Journal of exposure science & environmental epidemiology*, 32(4), 530-537.
79. Zhang, Y., Zhang, Q., Sun, H., & Wang, Y. (2019). The impact of NO₂ exposure on embryonic development and epigenetic alterations. *Environmental pollution (Barking, Essex: 1987)*, 245, 104-112.