

Modelling of Aerosol Optical Depth Using Machine Learning Techniques



By

Ibrahim Ibn Abdul-Wajid (356875)

Muhammad Umair (345487)

Raja Taimoor (341988)

Supervisor

Dr. Erum Aamir

Co-supervisor

Ma'am Momina Ahmad

**Institute of Environmental Sciences and Engineering (IESE)
School of Civil and Environmental Engineering (SCEE)
National University of Sciences and Technology (NUST), Islamabad**

June 2024

Modelling of Aerosol Optical Depth Using Machine Learning Techniques

A thesis submitted to the National University of Sciences & Technology (NUST) in
partial fulfilment of the requirements for the degree of Bachelor of Engineering (BE) in
Environmental Engineering

By

Ibrahim Ibn Abdul-Wajid (356875)

Muhammad Umair (345487)

Raja Taimoor (341988)

Supervisor

Dr. Erum Aamir

Co-supervisor

Ma'am Momina Ahmad

**Institute of Environmental Sciences and Engineering (IESE)
School of Civil and Environmental Engineering (SCEE)
National University of Sciences and Technology (NUST), Islamabad**

June 2024

Approval Sheet

It is certified that the contents and form of the thesis entitled “Modelling of Aerosol Optical Depth Using Machine Learning Techniques” submitted by Ibrahim Ibn Abdul-Wajid, Muhammad Umair, and Raja Taimoor have been found satisfactory for the degree requirement.

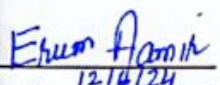

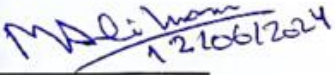

 12/14/24	Dr. Erum Aamir Assistant Professor IESE (SCEE) NUST Islamabad	 12/06/2024
Dr Erum Aamir (Project Supervisor) Assistant Professor at SCEE (IESE), NUST		Ma'am Momina Ahmad (Co-supervisor) Lecturer at SCEE (IESE), NUST
 12/06/2024	Dr. Muhammad Ali Inam Assistant Professor HoD Environmental Engineering SCEE (IESE), NUST H-12 Islamabad	
Dr Muhammad Ali Inam Head of Department at SCEE (IESE), NUST		
 12/06/2024	Prof. Dr. Imran Hashmi Associate Dean IESE (SCEE) NUST Islamabad	
Dr Imran Hashmi Associate Dean at SCEE (IESE), NUST		

Table of Contents

Acknowledgement	8
Abstract	9
Chapter 1: Introduction	11
1.1 Background	11
1.2 Problem Statement	12
1.3 Objectives	12
Chapter 2: Literature Review	13
2.1 Support Vector Regression (SVR)	13
2.2 Gradient Boosting Decision Tree (GBDT)	14
2.3 Random Forest (RF)	14
2.4 Discussion	15
Chapter 3: Methodology	16
3.1 Data Collection	17
3.1.1 Global Horizontal Irradiance (GHI)	17
3.1.2 Temperature	17
3.1.3 Relative Humidity	18
3.1.4 Wind Speed	19
3.1.5 Wind Direction	19
3.1.6 Day of the Year	20
3.1.7 Time of the Day	21
3.3 AERONET Data	26
3.3.1 Extrapolating AOD (550nm)	26
3.3.2 Averaging values	27
3.4.1 Metrics	27
3.4.1 Correlation Coefficient	27
3.4.1 Mean Absolute Error (MAE)	28
3.4.3 Mean Square Error (MSE)	28
3.4.4 Root Mean Square Error (RMSE)	29
4.1 Pre-processing data	30
4.1.1 Removing outliers and null values	30

4.1.2 Turning values into float values	31
4.1.3 Standardizing Datasets	31
4.2 Forward Feature Selection (FFS).....	31
4.3 Splitting of training and testing sets	32
4.4 Results	33
4.4 Discussion	36
Chapter 5: Improving the SVR	37
5.1 Hyper-tuning	37
5.2 Optimizers	37
5.2.1 Adam	38
5.2.2 Gray Wolf Optimizer	38
5.4 Comparison with Literature	39
5.5 Discussion	39
5.5.1 Pollution	39
5.5.2 Poor quality data	40
Chapter 6: Conclusion & Recommendations.....	41
6.1 Conclusion.....	41
6.2 Recommendations	41
6.2.1 Training the models in cleaner cities	41
6.2.2 Using Different Data.....	42
6.2.3 Using Different Input Features.....	42
Chapter 7: References	43

List of Tables

Table 1: Validation of ERA-5 datasets for Lahore.....	22
Table 2: Validation of ERA-5 datasets for Karachi.....	24
Table 3: Performances of different ML models for Lahore	33
Table 4: Performances of different ML models for Karachi	34
Table 5: The optimum hyperparameters found by our Grid Search	37
Table 6: Performance of different optimizers for Lahore's SVR model.....	38
Table 7: Performance of different optimizers for Karachi's SVR model.....	39
Table 8: Comparison of our SVR models with models from literature.....	39

List of Figures

Figure 1: Overview of project methodology	16
Figure 2: Lahore GHI	22
Figure 3: Lahore Temperature	22
Figure 4: Lahore RH	23
Figure 5: Lahore WS.....	23
Figure 6: Lahore WD.....	23
Figure 7: Karachi GHI	24
Figure 8: Karachi Temperature	24
Figure 9: Karachi RH.....	25
Figure 10: Karachi WS.....	25
Figure 11: Karachi WD.....	25
Figure 12: Overview of the pre-processing of data	30
Figure 13: A simple depiction of how FFS performs	32
Figure 14: Lahore's GBDT	33
Figure 15: Lahore's RF	34
Figure 16: Lahore's SVR.....	34
Figure 17: Karachi's GBDT	35
Figure 18: Karachi's RF	35
Figure 19: Karachi's SVR.....	36

Acknowledgement

First, we thank Allah, the Most Compassionate, the Most Merciful, and the Lord of all worlds. Second, we thank those individuals without whom this project would not have been possible. The Prophet (SAW) said: "Whoever does not thank people has not thanked Allah".

The first of such individuals is, of course, our supervisor, Dr. Erum Aamir. Without her support and patience, this project would have been a passing idea. We could not have asked for a better supervisor. The second person we must thank is Ma'am Momina Ahmad. We will forever remain indebted to the care and guidance that she provided.

We would also like to thank our families and friends for the love and support they gave us throughout this journey. A particular thank you to one Asim Ahmer of the School of Natural Sciences (SNS). Without his help, this project would not have made it to completion. We must also thank Hira Saif of the Institute of Environmental Engineering and Sciences (IESE) for all the help and guidance she lent us.

Finally, we would like to thank all the faculty and admin at IESE for providing us with an environment that allowed us to explore our potential and develop into the individuals that we are today.

Abstract

Aerosol Optical Depth (AOD) is a critical parameter in atmospheric sciences, representing the concentration of aerosols in a vertical column of the atmosphere. Accurate prediction of AOD is essential for understanding air quality, climate change, and their impacts on human health. This study explores the potential of machine learning techniques in predicting AOD levels over urban regions in Pakistan, specifically Lahore and Karachi. We employ three machine learning models: Support Vector Regression (SVR), Gradient-Boosting Decision Tree (GBDT), and Random Forest (RF), leveraging various meteorological and environmental datasets.

The datasets are pre-processed by removing outliers, handling missing values, and standardizing the data points. Key input features include temperature, relative humidity, wind speed, wind direction, and day of the year. We validate the performance of these models using metrics such as the correlation coefficient (R), Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE).

The results indicate that the SVR model, optimized using the Gray Wolf Optimizer (GWO), outperforms the other models with a correlation coefficient (R) of 0.64 for Lahore and 0.54 for Karachi. The optimized SVR model also significantly improves in MAE and RMSE, highlighting its robustness and accuracy in predicting AOD levels.

This study demonstrates the efficacy of Machine Learning (ML) techniques in environmental monitoring, providing a reliable tool for predicting AOD. The findings suggest that, by incorporating higher-quality data and a broader range of input variables, further improvements can be achieved. The successful application of these models in Pakistan could pave the way for enhanced air quality management and climate research in other regions.

List of Abbreviations:

AERONET: Aerosol Robot Network

AOD: Aerosol Optical Depth

GBDT: Gradient-Boosting Decision Tree

GHI: Global Horizontal Irradiance

KHI: Karachi

LHR: Lahore

MAE: Mean Absolute Error

MBE: Mean Bias Error

ML: Machine Learning

MSE: Mean Squared Error

RF: Random Forest

RH: Relative Humidity

RMSE: Root Mean Squared Error

SVR: Support Vector Regression

WD: Wind Direction

WS: Wind Speed

Chapter 1: Introduction

1.1 Background

Aerosols refer to particulate matter suspended in the air. They include crystal materials, inorganic materials, metals, elemental carbon, and biological substances such as pollen, spores, and animal excrement (Zaheer et al., 2023). These materials play a significant role in the radiative features of clouds and the scattering and absorption of radiation in the atmosphere. Consequently, the presence of aerosols in the Earth's atmosphere significantly alters the planet's radiative balance and, with it, the climate system (Lemmouchi et al., 2023). These effects extend to the hydrological cycle, the global surface temperature, and ecosystems (Ali et al., 2020).

Furthermore, aerosols negatively impact human health. According to the World Health Organization (WHO), about 4.2 million to 7 million people die per year because of the presence of aerosols in the air. Finer particulate matter (such as $PM_{2.5}$ and PM_{10}) enters the human lungs and causes numerous heart and respiratory-related diseases, such as cardiovascular disease, cerebrovascular disease, and asthma. Prolonged exposure to particulate matter can lead to complications related to child-bearing, such as premature births, low birth weights amongst infants, and early gestational birth ages (Ranjan et al., 2020). Sometimes, the effects that aerosol particles have on human health are indirect. For example, the acidic nature of sulfate ions encourages the availability of metals that, when exposed to humans, can lead to higher morbidity and death rates (Zaheer et al., 2023).

For these environmental and health-related issues, the study of aerosols is crucial. Studying their spatial variability and interactions with radiation in the atmosphere. South Asia, in particular, is threatened by the rising amounts of aerosols in the atmosphere due to growing populations, rapid urbanization, increased motorized traffic, changes in land use, and rising industrialization in and around urban areas. The sources of these aerosols are both natural and anthropogenic. Dust and sea salts (examples of natural aerosols) originate from arid regions in the south and the Indian Ocean. Anthropogenic aerosols are mostly from vehicular and industrial emissions (Ali et al., 2020).

The fundamental property of aerosols that quantifies their presence in atmospheric columns is aerosol optical depth (AOD). Aerosol optical depth (AOD) is an optical parameter of aerosols that is defined as the extinction coefficient of sunlight over a vertical column of aerosols in the atmosphere. Several factors influence the variation of Aerosol Optical Depth (AOD) in the

atmosphere, namely seasonality, topography, and environment. In Pakistan, AOD values peak in the summer and fall during the winter (Ali et al., 2020).

Several remote sensing techniques have been established to monitor the variation of aerosol optical depth (AOD) in the atmosphere, ranging from ground-based to satellite-based techniques. However, these monitoring systems suffer from several limitations. Ground-based monitoring stations (such as NASA's AERONET stations) can only monitor AOD levels at a certain point, hence providing no spatial resolution. Furthermore, factors such as extreme pollution events, bad weather, and equipment malfunction lead to gaps in the recorded data. Satellites offer data over a much wider spatial range, but suffer from low temporal resolution (for example, the MODIS sensors onboard NASA's Aqua and Terra satellites each provide data for AOD levels a mere two times a day).

Gapless datasets for AOD with a high temporal resolution (such as hourly) have a number of uses: For example:

- They are used to monitor diurnal patterns in aerosol levels (Lipponen et al., 2022)
- They are used to derive surface solar irradiance databases for use in the energy sector (Schroedter-Homscheidt & Oumbe, 2013)
- They are used to model PM_{2.5} levels - another important air quality parameter (Pu & Yoo, 2022)

1.2 Problem Statement

Although Pakistan possesses monitoring systems for aerosol optical depth (AOD), factors such as poor weather, equipment malfunctioning, and high pollution events lead to the inadequate monitoring of data and gaps in observations (Zaheer et al., 2023). This inhibits the country's research capabilities and ability to analyze and monitor air pollution.

1.3 Objectives

The objectives of our project are:

1. Train three Machine Learning (ML) models for Karachi and Lahore using the best input features.
2. Compare the performances of these models using appropriate metrics.
3. Attempt different optimization processes on our best-performing model

Chapter 2: Literature Review

A literature review was conducted to assess the potential Machine Learning (ML) has in providing a solution to our problem. Machine Learning, a branch of Artificial Intelligence (AI), involves the use of data to train algorithms to recognize patterns and make predictions. Thus far, the application of Machine Learning (ML) in Environmental Science and Engineering (ESE) has been limited, but progress is being made in using ML to solve complex ESE-related problems such as the modelling of wastewater treatment systems, the prediction of water availability, and to identify toxic elements in commercially-available chemical substances (Zhong et al., 2021).

The ability of ML algorithms to interpret large datasets and identify complex relationships makes them well-suited for a task such as the prediction of AOD. Already, studies have been conducted internationally to test the potential use of ML in providing AOD datasets.

One study compared the capabilities of a Support Vector Machine (SVM) and a BP Neural Network in predicting AOD values. Both showed strong non-linear fitting abilities and that the SVM was more accurate. For the dataset, approximately 2000 samples were taken over a one-year span (Jing et al., 2017). Another study used a random forest (RF) model to predict missing AOD values using meteorological and topographical parameters. The random forest proved largely successful in its predictions (Jin et al., 2022).

AOD prediction is not the only complex, air-related task that ML has been useful for. One study in China used a gradient-boosting decision tree (GBDT) to improve satellite estimations of ground PM2.5 levels. The model successfully improved the spatial resolution of the satellite data (T. Zhang et al., 2021).

After conducting a preliminary survey of the potential uses of ML in AOD prediction, we looked deeper into what models would be best suited for the task.

2.1 Support Vector Regression (SVR)

Support Vector Regression (SVR) is a widely used, supervised learning algorithm. This means that both the input data and output data are provided to the model (it is supervised) and then derives a relationship between the two sets of data. Once this “training” has been completed, the model can be used to make predictions based on new sets of data. What makes the SVR suitable for complex regression problems is its ability to approximate a relationship between the input and output datasets whilst maintaining a minimum possible error and a relative level of simplicity. This simplicity is achieved by focusing on making the model function as flat as possible and setting a

margin of tolerance (known as epsilon) within which predictions are accepted. The SVR model can be denoted as:

$$f(x) = \langle w + \phi(x) \rangle + b$$

where $f(x)$ is the general, non-linear regression function, w and b are vector and scalar weights, respectively, and $\phi(x)$ denotes the mapping of the plane within the space x (Panahi et al., 2020).

2.2 Gradient Boosting Decision Tree (GBDT)

Gradient Boosting Decision Tree (GBDT) is another form of supervised learning where the input and output data are divided into several subsets. Simple models (decision trees) are trained on their given subsets of data sequentially. Each model, once trained, provides an error the next model in the sequence attempts to lower. This sequence is continued iteratively until the approximated relationship between the input and output data is optimal (i.e. gives the lowest possible error). What makes the GBDT suitable for our task is that it is able to solve non-linear regression problems that involve multiple input features. Furthermore, the GBDT has proven to show low sensitivity to missing data in the training sets (Huan et al., 2020).

2.3 Random Forest (RF)

The Random Forest (RF) algorithm is another robust technique in Machine Learning (ML) that, like the Gradient Boosting Decision Tree (GBDT), makes use of decision trees, but not in the same manner. The Random Forest (RF) algorithm builds multiple decision trees and trains each of them on a portion of the data. The results of all the decision trees are calculated, and the mean prediction of the individual decision trees is the output. Whereas the Gradient Boosting Decision Tree (GBDT) makes use of the decision trees' ability to improve themselves sequentially, the Random Forest (RF) leverages the "collective wisdom" of the decision trees to enhance predictive accuracy and limit overfitting. The data subset provided to each decision tree is divided into an "in-bag" subset which is used for training the decision tree and an "out-of-bag" subset which is not involved in the training of the decision tree. The partitioning between the "in-bag" and "out-of-bag" sets are unique to each decision tree. This helps increase the "randomness" and improves the internal validation of the models by a significant amount. The "out-of-bag" sample for each decision tree is used to evaluate its performance. The average of all the "out-of-bag" predictions provides a metric for the relative accuracy of the Random Forest (RF) model (He et. al, 2020).

2.4 Discussion

Our review of the literature convinced us that the three most suitable Machine Learning (ML) models for our task were the Support Vector Regression (SVR), Gradient-Boosting Decision Tree (GBDT), and the Random Forest (RF).

Support Vector Regression (SVR) possesses a strong ability to approximate non-linear relationships whilst maintaining a relative level of simplicity in its model function. Given that our problem required multiple input features, we believed SVR would be most suited for the task.

Gradient-Boosting Decision Tree (GBDT) is able to derive generalizations on datasets that include missing values, which is a common issue in environmental datasets. Already the GBDT has proven capable in improving the spatial resolution of AOD datasets (T. Zhang et al., 2021) and so it became a promising candidate for our task.

Lastly, the Random Forest (RF)'s ability to include randomness in the training of each decision tree provides it with a level of robustness that is suited for diverse urban environments like Lahore and Karachi.

Given these strengths, we concluded that SVR, GBDT, and RF are well-suited for our task of predicting AOD. The next step involves developing a robust methodology for designing these models and collecting high-quality data to train them. This approach will help us address the limitations observed in previous studies and enhance the accuracy and reliability of our AOD predictions.

Chapter 3: Methodology

The three selected Machine Learning (ML) models are based on what is known as “supervised” learning. That is, both the input features and the target variables are provided to the algorithms so that a relationship between the two sets of data can be approximated. The first task we had to conduct was collecting data for both our input features and our target variable.

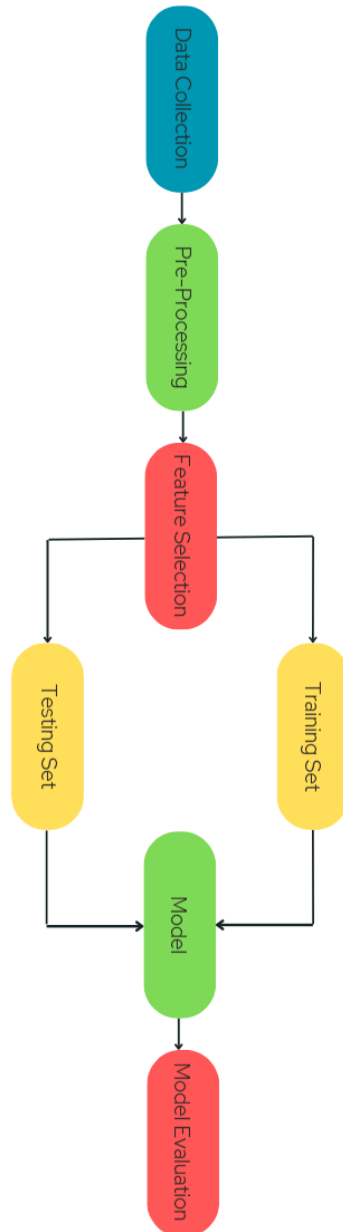


Figure 1: Overview of project methodology

3.1 Data Collection

The European Centre for Medium-Range Weather Forecasts (ECMWF) provides global atmospheric datasets spanning from 1979 to the present. The fifth and latest generation of these datasets is ERA-5. Every 3 months, these datasets are updated (Vanella et al., 2022). The benefits of ERA-5 datasets are that they offer a high spatial resolution for a wide range of climate-related variables. Furthermore, the datasets have an hourly resolution, which is precisely what is needed for the development of our models. Compared to previous generations, the ERA-5 offers an enhanced level of data accuracy. It makes use of the latest version of the ECMWF's Integrated Forecast System model (known as IFS 41r2). This version provides greater temporal output as well as increased horizontal and vertical resolutions (1 h, 0.25°, and 137 vertical levels extending from the surface, respectively). Already, many studies have been conducted that demonstrate the overall accuracy of the ERA-5 datasets compared to other reanalysis datasets and previous generations (Soares et al., 2020).

3.1.1 Global Horizontal Irradiance (GHI)

The first input feature chosen for our model was Global Horizontal Irradiance (GHI). Several studies have demonstrated the inter-relation between GHI and Aerosol Optical Depth (AOD) (Perez et al., 2020). GHI is a parameter that quantifies the amount of shortwave radiation that reaches a horizontal plane on the Earth's surface, either through the direct penetration of sunlight or through the diffusion of atmospheric radiation. AOD, as explained previously, measures the amount of sunlight that is extinguished by aerosols in the atmosphere, either by absorbing it or by scattering it. The relationship between GHI and AOD is therefore easy to understand. If GHI is the amount of sunlight being received by the Earth's horizontal surface and AOD is the amount of sunlight being extinguished in the atmosphere, it is clear that the two must be inversely proportional. A higher AOD represents a higher concentration of aerosols, which consequently entails more scattering and absorption of solar light. As AOD increases, less sunlight is able to directly reach the Earth's surface, which causes a drop in GHI levels (Gueymard et al., 2012).

3.1.2 Temperature

The second input feature selected for our models was surface temperature. The relationship between temperature and Aerosol Optical Depth (AOD) is more complex than that of GHI and AOD, mostly due to the various atmospheric processes that are involved. Typically speaking, higher temperatures relate to enhanced emissions of certain aerosols (such as organic

compounds from vegetation and particulate matter from anthropogenic sources). Therefore, in this case, higher temperatures mean a higher AOD level. Furthermore, high temperatures can accelerate chemical processes in the atmosphere that may produce secondary aerosols. These secondary aerosols contribute to the AOD levels. Such phenomenon has been observed in warm months as well as in regions experiencing heat waves (Basharat et al., 2023). It is likely that high temperatures are the main reason for why AOD levels peak in the summer time in cities across Pakistan (Nawaz et al., 2023).

One complication, however, is that in certain cases, high aerosols can correspond to lower temperatures. The reason for this is due to the impact aerosols have on the Earth's radiative balance. Aerosols that scatter sunlight can cause the Earth's surface to experience a cooling effect. On the other hand, aerosols such as black carbon absorb sunlight and consequently contribute to atmospheric heating. This two-directional relationship is responsible for various atmospheric feedback loops, in which temperature changes impact aerosols, which in turn generate temperature fluctuations (Basharat et al., 2023). It is for this reason that our project objectives include an investigation into the input features that most heavily affect AOD levels. Our study hopes to find just how much of a role temperature plays in influencing AOD levels.

3.1.3 Relative Humidity

Statistical studies have already been conducted that show a positive relationship between relative humidity and AOD in Pakistan (Tariq et al., 2021; Basharat et al., 2023; Zeb et al., 2024). The reasons for this are multifaceted.

Directly and indirectly, Relative Humidity (RH) influences AOD levels by manipulating the microphysical and optical properties of aerosols. Soot, sulphates, nitrates, organic carbon, and mineral dust are the major forms of aerosols present in the troposphere. All of these substances are relatively soluble in water. During the presence of high ambient RH, atmospheric conditions can induce an increase in water uptake by these aerosols, which increases their size as well as their residence time in the atmosphere. This uptake of water is known as aerosol hygroscopicity. The result is a significant increase in the scattering capabilities of the aerosols, which in turn increases AOD (Prasad et al., 2023).

Another way through which RH affects AOD is cloud formation. The swelling of aerosols during high RH events can cause the aerosols to perform as cloud condensation nuclei (CCN). This enhances cloud formation and influences their radiative properties (Tariq et al., 2021).

The effects that RH has on AOD is dictated by the chemical composition of the aerosols. Some aerosols exhibit greater scattering capabilities in high RH than others. For example, sea salt is an aerosol that is purely hygroscopic and hence, possesses a greater scattering coefficient than aerosols such as dust. It is therefore important to also take note of the types of aerosols present when investigating the relationship between RH and AOD (Prasad et al., 2023).

3.1.4 Wind Speed

The relationship between Wind Speed (WS) and Aerosol Optical Depth (AOD) is not straightforward and hence requires elaboration. Generally speaking, influences AOD levels by affecting the concentration, spatial variation, and the retention of aerosols in the atmosphere.

During dust storms, high wind speeds can cause the resuspending of dust particles and other particulate matter on the Earth's surface into the atmosphere. This significantly reduces the ability of sunlight to breach the Earth's lower atmosphere and hence increases AOD and reduces visibility. This phenomenon is particularly evident in regions that are arid and semi-arid (such as Iran), where soil is loose and sand can be easily lifted by the presence of high winds (Omidvar et al., 2022).

In contrast to arid regions, are coastal regions which, interestingly, also display higher AOD levels during the presence of high wind speeds. The most likely reason for this is the presence of sea salts which, as explained before, are purely hygroscopic. Strong winds can transport sea salts suspended over water to coastal regions, where the addition of humidity can cause a swelling of the sea salt particles and hence, an increase in AOD (Sun et al., 2024).

On the other hand, high wind speeds can lead to a lowering of AOD levels. This is in the event where winds are dispersing aerosols, reducing their concentration in certain regions, and hence leading to less obstruction in the path of sunlight. In such scenarios, high wind speeds help to improve the air quality, by diluting aerosols in the atmosphere and transporting pollutants away from their sources (N. Kang et al., 2020).

3.1.5 Wind Direction

As explained above, high wind speeds can both increase AOD levels as well as decrease them. A reason for this is the direction of wind speeds. The source and type of aerosols present in a region (as well as their subsequent transportation) is dictated by the direction in which the wind is blowing. For example, winds generated in urban or industrial localities can carry with them high concentrations of organic particulate matter, which can then increase AOD in places that the

winds cross (Yousefi et al., 2023). On the other hand, winds that originate in relatively cleaner regions (such as rural areas or coastal regions) may lack anthropogenic aerosols but may still carry with them natural aerosols such as sea salts and sand. These too can lead to increases in AOD levels (Prasad et al., 2023).

The role of topography must also be considered when investigating the relationship between Wind Direction (WD) and AOD. Studies have found that mountains that block the flow of winds can cause a lowering in AOD levels by preventing the spread and transportation of aerosols. On the other hand, in basins, this blocking of winds can prevent the dispersion of aerosols, which leads to their accumulation and subsequently, increases in AOD (Nakata et al., 2022).

One study in China showed that the particular matter levels in a city change by season due to the shifts in wind. When the winds began to flow from the south to the north, they brought with them anthropogenic aerosols produced from local industries, which subsequently increased AOD levels in the city (C. Zheng et al., 2017).

3.1.6 Day of the Year

Studies that investigate long-term patterns in AOD levels provide insight into the relationship between the variable we will refer to as Day of the Year (DOY) and Aerosol Optical Depth (AOD). For example, one study showed that – generally speaking – AOD levels are higher during summer days in urban cities in Pakistan than during winter days (Ahmed et al., 2020).

This general trend, however, is complicated by the monsoon season. In Pakistan, the monsoon rains occur in late summers. The presence of rainfall dilutes aerosols in the atmosphere, causing a “washing out” effect that lowers AOD levels (Khalid et al., 2022).

There is also the role that agricultural activities such as crop burning play. In Pakistan, during the winter time, stubble is burned to prepare crop land for the next season. This burning of biomass releases a wide manner of greenhouse gases and particulate matter. This in turn can lead to high AOD events (Tariq et al., 2023). Winter can also cause high AOD levels due to stagnant air conditions. The stagnant air during winters traps aerosols close to the surface, which consequently boosts AOD levels (Qayyum et al., 2022).

As can be seen, the relationship between Day of the Year (DOY) and AOD is complex. For this reason, the second objective of our project was to assess what influence DOY has on AOD predictions relative to the other input features.

3.1.7 Time of the Day

Temporal variations of Aerosol Optical Depth (AOD) must be considered when modelling AOD. This is due to the diurnal variations in aerosol levels as a consequence of changes in natural and human activities. In urban and industrial areas, the changes in AOD levels throughout the day are significant. In the early hours of the morning, humans begin their industrial operations as well as commute to their work places (more traffic). This significantly increases AOD levels. Furthermore, during the morning, there tends to be greater atmospheric stability as a result of cooler temperatures. The stagnation of winds can trap aerosols near the city's surface, elevating AOD levels (Jiang et al., 2024).

As the hours progress, temperatures increase and the planetary boundary layer rises. This can generate winds that disperse the aerosols and subsequently reduce AOD. This dilution of aerosols is aided by the presence of solar radiation that heats up the air and allows for a more vigorous mixing of air streams (Haider et al., 2017).

The difference between morning and afternoon AOD levels can be quite significant. One study conducted in Eastern India found that, due to changes in aerosol loading from anthropogenic sources and enhancement in relative humidity, the decline in AOD levels can be as great as 20% (Mukherjee & Vinoj et al., 2019). We can therefore conclude that time must be considered as an input feature in our models.

3.2 Validation

In Pakistan, at 9 locations, the World Bank has placed weather masts that have recorded ground-based data of meteorological parameters such as temperature, relative humidity, wind speed, wind direction, and GHI. The resolution of these datasets is 10 minutes (Irfan et al., 2019).

In research, ground-based observations are always considered to more accurate and “true” compared to satellite-based observations or data generated from reanalysis techniques. Therefore, we sought to validate the datasets we had acquired from ERA-5 with the datasets available from the World Bank. The World Bank datasets were not chosen as our input due to their size. The World Bank only has data available from the years 2014 to 2017. Our project demanded larger datasets (which we acquired from ERA-5). The datasets we acquired from ERA-5 spanned from 2014 to 2022. The years of data that we had available from the World Bank were used to validate the accuracy of the corresponding years of ERA-5 data.

3.2.1 Results of validation

The table provides metrics for the validation of Lahore's ERA-5 datasets with the World Bank datasets.

Table 1: Validation of ERA-5 datasets for Lahore

Lahore (LHR)	GHI	Temperature	RH%	Wind Speed	Wind Direction
R	0.98	0.95	0.56	0.39	0.36
Mean Bias Error (MBE)	-14.86	-1.68	28.27	-0.05	-63.49
Root Mean Square Error (RMSE)	59.51	3.58	35.09	1.54	120.41

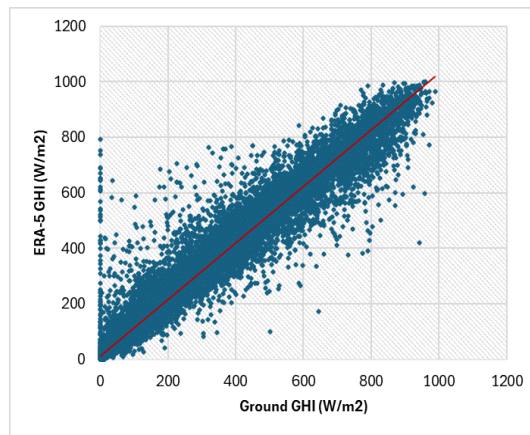


Figure 2: Lahore GHI

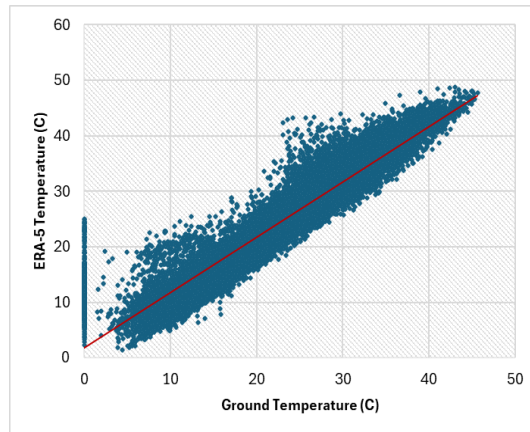


Figure 3: Lahore Temperature

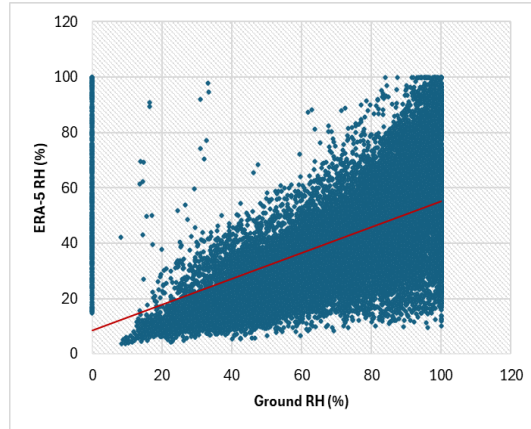


Figure 4: Lahore RH

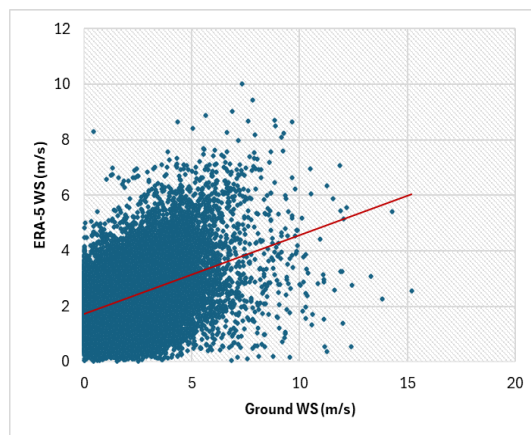


Figure 5: Lahore WS

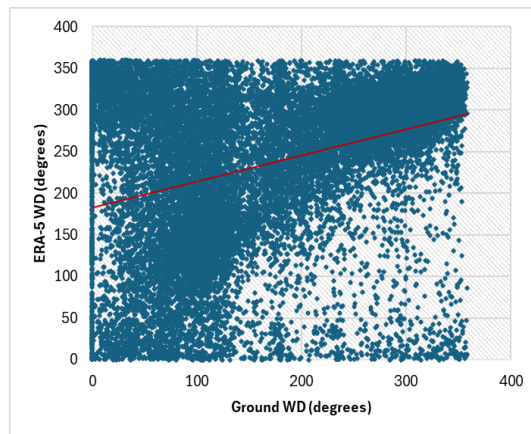


Figure 6: Lahore WD

The same validation was performed for the datasets in Karachi and their metrics are provided below.

Table 2: Validation of ERA-5 datasets for Karachi

Karachi (KHI)	GHI	Temperature	RH%	Wind Speed	Wind Direction
R	0.52	0.92	0.83	0.79	0.54
Mean Bias Error (MBE)	-10.34	0.52	-1.08	-0.77	-23.64
Root Mean Square Error (RMSE)	130.57	1.97	12.60	1.43	79.29

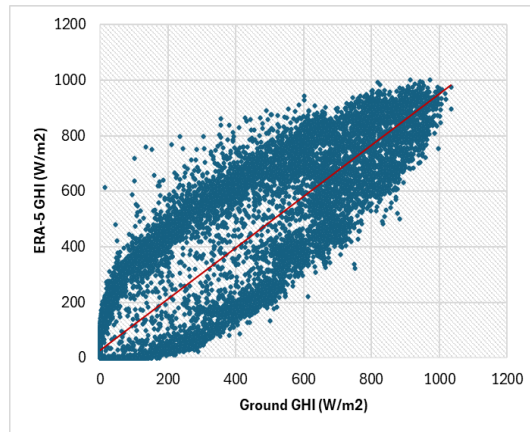


Figure 7: Karachi GHI

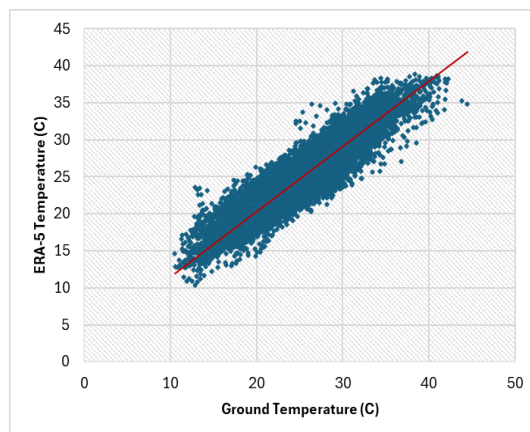


Figure 8: Karachi Temperature

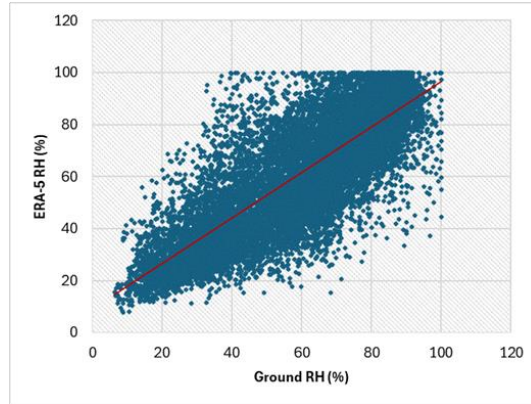


Figure 9: Karachi RH

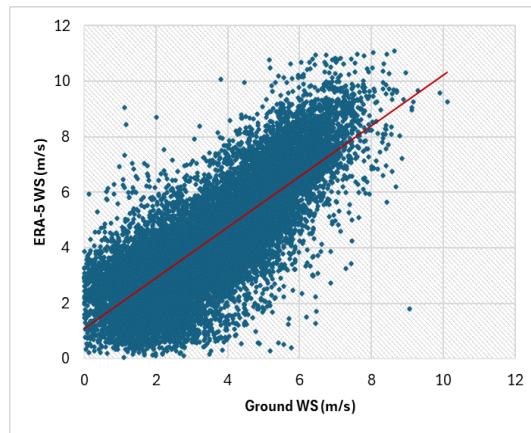


Figure 10: Karachi WS

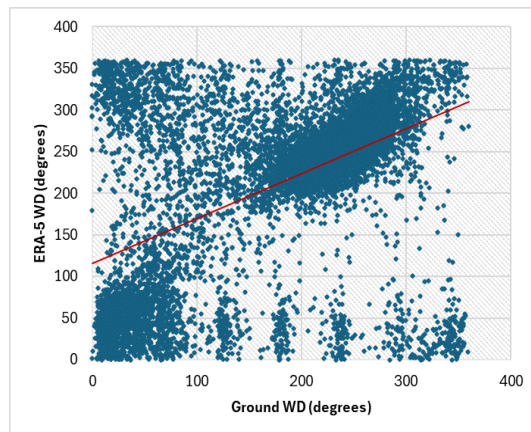


Figure 11: Karachi WD

3.2.1 Discussion

The results show that certain parameters possess a higher correlation between ERA-5 data and World Bank data than others. These differences are not consistent between cities. For example, GHI in Lahore had a much greater R value (0.98) than GHI in Karachi (0.52). Some parameters, such as WS and WD in Lahore, had R values much lower than expected (0.39 and 0.36 respectively). However, studies have demonstrated that even in cases where ERA-5 data displays a low correlation to ground data, it still can be used in the development of ML models for solving complex, air-related problems (Wang et al., 2023). As such, we proceeded with the data that we had available in the development of our models.

3.3 AERONET Data

For our target variable (AOD), we collected data from the AERONET stations present in Lahore and Karachi. These are the only two available AERONET stations in Pakistan. Their respective coordinates are (74.324, 31.542) and (67.030, 24.870). It is important to note that when collecting the data for our input features from the ERA-5 website, data extracted within a 40-km radius of the AERONET stations was taken. This is in line with the methodology developed by Lanzaco et al. (2017). Were the input features taken from a location further than 40-km away from the AERONET stations, then there could be no claim to spatial correspondence between our input features and target variable.

The highest quality dataset available from the AERONET website was collected. This was the Version 3 Level 2.0 (Solar) measurements. These datasets include AOD values taken every 15 minutes during the day time and are processed to remove influence from cloud coverage and other atmospheric interferences.

3.3.1 Extrapolating AOD (550nm)

Aerosols have distinct properties when it comes to the scattering and absorbing of sunlight. It is for this reason that scientists measure AOD at different wavelengths. Each wavelength provides unique insights into the different types of aerosols present in a region. The common reference wavelength for studies is AOD measured at 550nm (Kumar et al., 2022). It is for this reason that we chose AOD (550nm) as our target variable.

However, AOD (550nm) is not provided by the AERONET stations. Instead, an extrapolation method using available AOD values had to be used to derive our own dataset for AOD (550nm).

The extrapolation method used was the same one Cesnulyte et al. (2014) used in their validation of ERA-5's AOD datasets against AERONET datasets. The formula given is:

$$AOD_{550} = AOD_{500} \left(\frac{550}{500} \right)^{-\alpha}$$

where AOD_{500} is the AOD values measured at 500nm (provided by AERONET) and α is Ångström exponent measured between the wavelengths 440-870nm.

3.3.2 Averaging values

AERONET provides AOD values that are measured at intervals of 15 minutes. The objective of our project was to develop models that predict AOD values at an hourly resolution. The input data we took from ERA-5 was already available in intervals of 1 hour. To ensure consistency in time, we took the hourly averages of our extrapolated AOD values.

3.4.1 Metrics

In Machine Learning (ML), models are evaluated on the basis of certain metrics. It is these metrics that are interpreted and used to assess the performance and accuracy of the models. Due to the nature of individual metrics, no single metric can be used to assess the accuracy of a model. Instead, several must be used. The combination of their unique insights helps to evaluate the model's predictive abilities. We chose to use the same metrics that Zaheer et al. (2023) used in assessing the performance of the model they had developed for AOD prediction. These were: the coefficient of determination (R), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE).

3.4.1 Correlation Coefficient

The correlation coefficient, denoted as R, measures the strength and direction between the observed values of a dataset and the predicted values provided by the model. In Machine Learning (ML), the correlation coefficient indicates how well the input features predict the target variable. A high R value (close to 1) indicates strong efficiency whereas a low value (close to 0) indicates low predictive power.

3.4.1 Mean Absolute Error (MAE)

MAE is the average of the absolute errors between the actual values and the predicted values. It is calculated as:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y(i) - y'(i)|$$

n = number of observations

y = true value

y' = predicted value

Unlike MSE and RMSE, MAE does not square the errors, so it is less sensitive to outliers. This makes it a more robust measure in the presence of outliers. MAE provides a clear and direct interpretation of the average error magnitude, making it easy to understand. Using MAE alongside MSE and RMSE can provide complementary insights, as it balances the effects of outliers differently.

3.4.3 Mean Square Error (MSE)

MSE is the average of the squares of the errors, where the error is the difference between the actual value and the predicted value. It is calculated as:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y(i) - y'(i))^2$$

n = number of observations

y = true value

y' = predicted value

By squaring the errors, MSE penalizes larger errors more than smaller ones, which helps in identifying models that make significant mistakes. Many ML algorithms, including linear regression, use MSE as a loss function to minimize during training, ensuring that the model finds the best fit for the data.

3.4.4 Root Mean Square Error (RMSE)

RMSE is the square root of the mean squared error. It provides the standard deviation of the prediction errors, offering a measure of how spread out these errors are.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y(i) - y'(i))^2}$$

n = number of observations

y = true value

y' = predicted value

RMSE is expressed in the same units as the target variable, making it more interpretable in the context of the problem. Like MSE, RMSE penalizes larger errors more significantly, making it useful for applications where larger errors are particularly undesirable. RMSE provides a straightforward measure of average prediction error magnitude, which can be easier to interpret and communicate.

Chapter 4: Model development

4.1 Pre-processing data

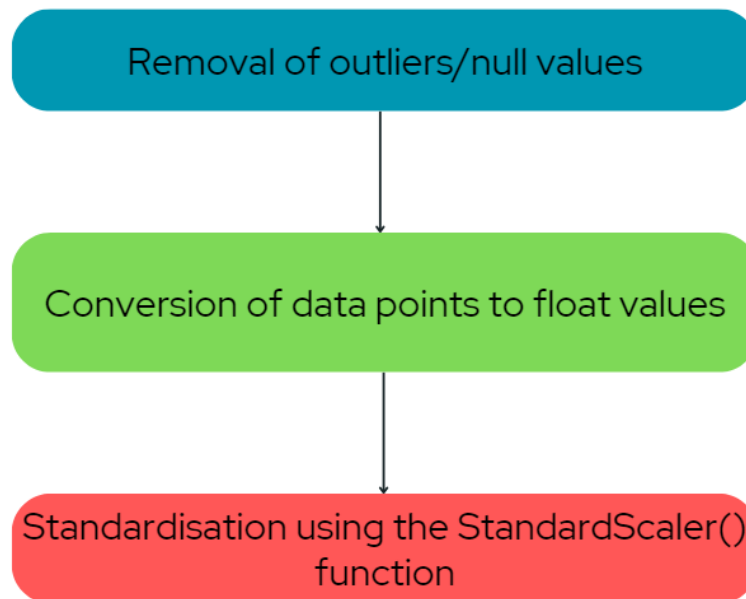


Figure 12: Overview of the pre-processing of data

4.1.1 Removing outliers and null values

The first step in pre-processing our data was removing outliers. Outliers are those values in a dataset that significantly differ from the other data points. The existence of outliers in weather and climate-related datasets can be for a number of reasons. Extreme weather events, equipment malfunctioning, and errors in data processing are just a few examples. Outliers, if fed into a model, can skew the results and negatively impact the model's ability to approximate a relationship between the input features and target variables (Boukerche et al., 2020).

One effective method of removing outliers is the Interquartile Range (IQR) method. In this method, the dataset is divided into percentiles. The difference is found between the 75th percentile (known as Q3) and the 25th percentile (known as Q1). This difference is the interquartile range (QR). Data points that fall below $Q1 - 1.5/QR$ or above $Q3 + 1.5/QR$ are identified as outliers and are subsequently removed. By removing these extreme values, we improve the accuracy of our models' predictions (Frery et al., 2023).

The other data that must be removed is null values. Null values are those values that are either missing or are undefined. Typically, they are a result of problems faced during the time of

recording or during the processing of the data. To ensure the integrity of the dataset, null values must be removed (Mijwil et al., 2023).

4.1.2 Turning values into float values

The common practice in ML is to convert all data points into float values. By floating our data, we provide our models with the necessary granularity that is required for the computation of numerical values. Without first floating our data, other pre-processing techniques such as standardization are not possible. Another benefit of floating data is that float values were optimized for computational processes, especially large datasets. This allows ML models to perform complex computations in less time and with less memory usage. Overall, use float values boosts the performance as well as the accuracy of ML models, giving us better predictions and better results (Joshi 2024).

4.1.3 Standardizing Datasets

The next critical step in pre-processing data in ML is standardizing the datasets. In standardization, the datasets are scaled down so that mean value is 0 and the standard deviation is 1. This practice is based on the understanding that the variables of different datasets do not share the same numerical range. As a result, they will not make equal (or fair) contributions to the model's training function. In some cases, not standardizing the datasets can cause the model's predictions to be bias towards a certain input feature (Testas et al., 2023). For example, before standardizing our datasets, it was observed that our GHI input feature was making an unfairly large contribution to the model's performances (on account of its large scale of values). After standardizing the datasets, this contribution was curtailed.

4.2 Forward Feature Selection (FFS)

The initial selection of our input features was based on a literature review and our understanding of what we believed would be the most important input features for our model. However, what we wished to find were the features that most significantly contribute to the predictions of our target variable.

For this, Forward Feature Selection (FFS) was selected. FFS is an ML method that helps a model identify the input features that are most important in making predictions. It is an iterative process that begins with no features and then one by one, adds the features into the model. Each time, it assesses the contribution that the feature is making to the model's performance. This is based on how much each feature helps to lower the model's errors. The metric we used to assess the errors

was Mean Squared Error (MSE). The process continues until the added features stop making a significant improvement to the predictions. The benefit of FFS is that it aids in simplifying models and reduces the overfitting of data. By eliminating irrelevant features, it also boosts computational efficiency (Zaheer et al., 2023).

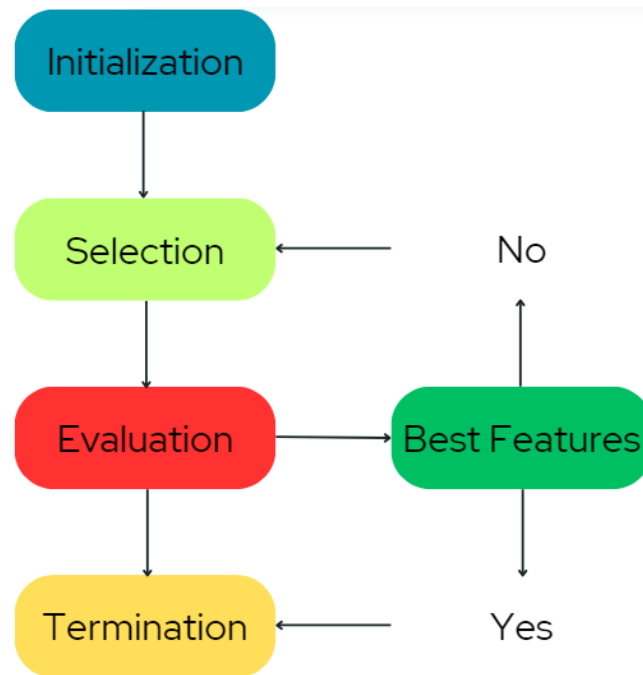


Figure 13: A simple depiction of how FFS performs

In our case, the five best input features were found to be: temperature, relative humidity, wind speed, wind direction, and Day of the Year (DOY).

4.3 Splitting of training and testing sets

For every supervised learning technique, the dataset must be divided into a training and a testing set. The training set is used to train the model and allow it to approximate a relationship between the input features and the target variable. The testing set is used to test the capabilities of the model in making new predictions. Standard practice is to split the data 70-80% and 20-30% between the training and testing sets. Such a split ensures that the models are able to generalize data and are robust in their predictions (N. Chen et al., 2021).

4.4 Results

The results of our models based on the selected metrics are given in the tables below.

Table 3: Performances of different ML models for Lahore

Lahore (LHR)	SVR	RF	GBDT
R	0.5244	0.4929	0.5157
Mean Square Error (MSE)	0.055	0.060	0.059
Mean Absolute Error (MAE)	0.180	0.191	0.189
Root Mean Square Error (RMSE)	0.235	0.243	0.242

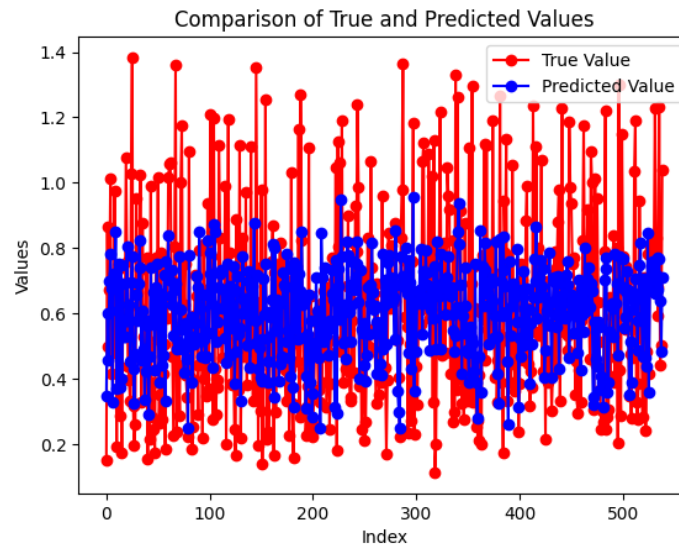


Figure 14: Lahore's GBDT

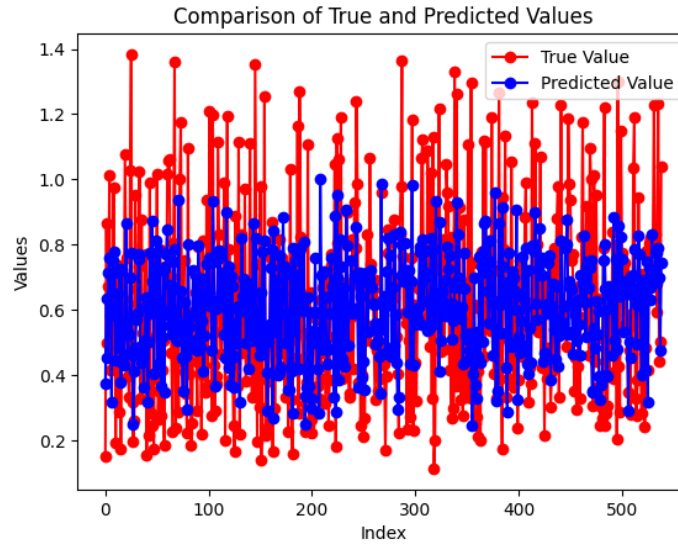


Figure 15: Lahore's RF

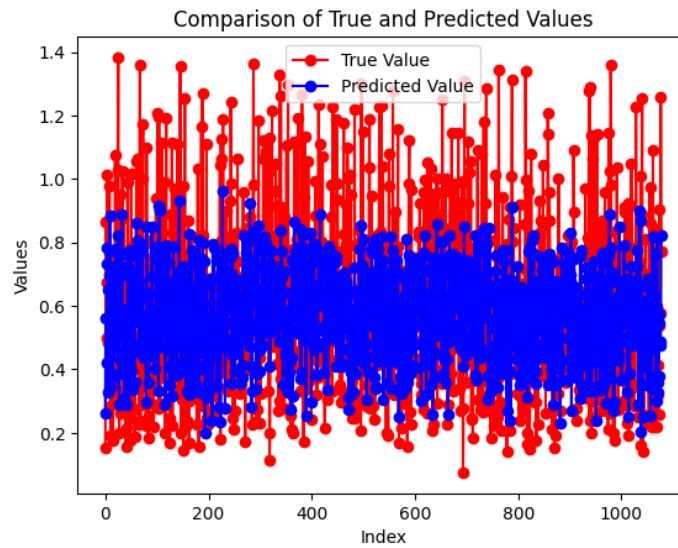


Figure 16: Lahore's SVR

Table 4: Performances of different ML models for Karachi

Karachi (KHI)	SVR	RF	GBDT
R	0.4066	0.4074	0.3701
Mean Square Error (MSE)	0.018	0.017	0.017
Mean Absolute Error (MAE)	0.110	0.104	0.107
Root Mean Square Error (RMSE)	0.136	0.130	0.133

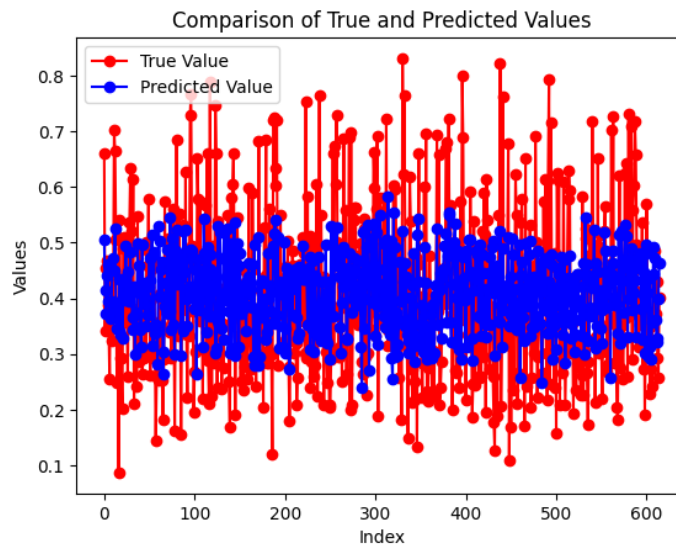


Figure 17: Karachi's GBDT

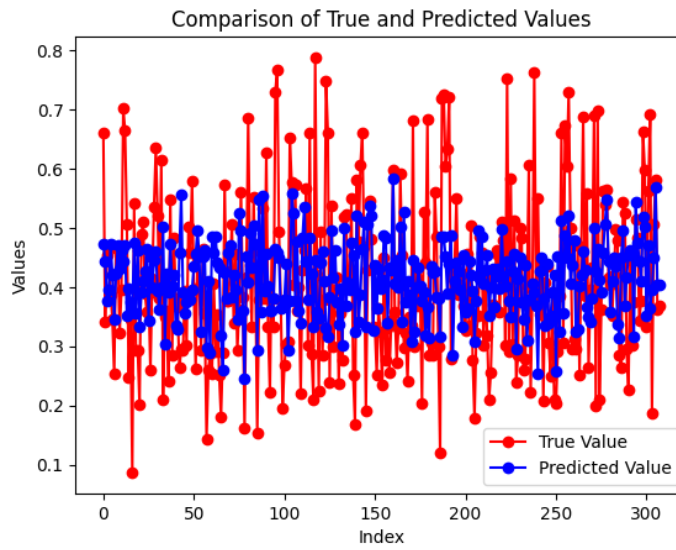


Figure 18: Karachi's RF

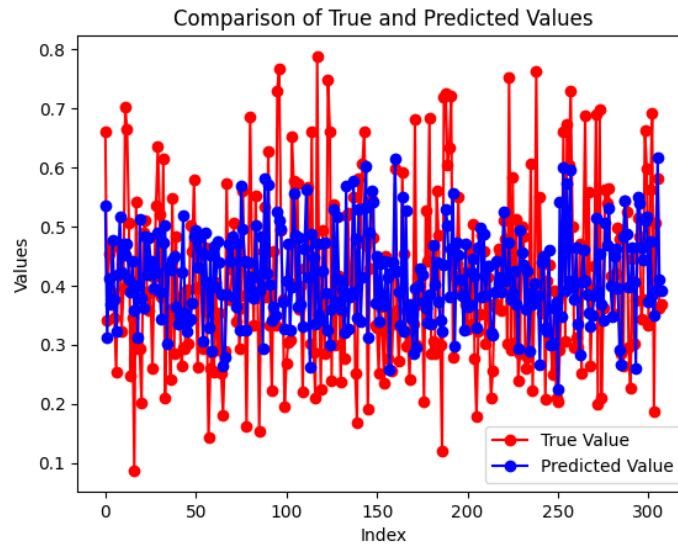


Figure 19: Karachi's SVR

4.4 Discussion

The results shown above display our first attempt at developing models for Lahore and Karachi. The performances of all our models, when compared to models available in the literature, were poor. For example, the model developed by Zaheer et al. (2023) had the following metrics: $R = 0.77$, $MSE = 0.0049$, $MAE = 0.06$, and $RMSE = 0.07$.

In the case of our models, the SVR performed best in Lahore whereas the SVR and RF in Karachi performed almost equally.

Chapter 5: Improving the SVR

After the initial attempts made at modelling AOD, we decided to focus our efforts on improving the performance of our SVR models. We conducted a deeper literature review (Lanzaco et al., 2017; N. Chen et al., 2021; Zaheer et al., 2023) into the capabilities of SVR and arrived at the conclusion that the SVR was the best model for us to focus our efforts on.

5.1 Hyper-tuning

In an SVR model, there are two main hyperparameters that influence the performance of the model. The first is epsilon (ϵ) which controls the margin of tolerance. If the epsilon is low, then there is a lower penalty for misclassified data points but the model's ability to predict target variables is decreased. If epsilon is high, there is a larger penalty for errors but the model may not be able to generalize as well when given new data. The second is C which controls the trade-off between the training error and the margin that divides our datasets. If C is too low, the margin will be too great and data points will be misclassified. If C is too high, the SVR's plane will perform overfitting, in which it will be too specific in its classification of the training dataset and will therefore perform poorly on new, unseen data (Y. Wang et al., 2023).

Due to the significance of these hyperparameters, we attempted to find the optimal values for our models. For this, we made use of the Grid Search function. This function tests different combinations of hyperparameters within a provided range and identifies which combination performs the best.

Table 5: The optimum hyperparameters found by our Grid Search

Epsilon (ϵ)	C
0.1	1

5.2 Optimizers

An optimizer is an algorithm used in ML to adjust the weights and biases of the models with the goal of improving the predictive abilities of the model. The library we were using to develop our models, Scikit-learn, provides the most common optimizer, the gradient descent. The gradient descent is an optimization algorithm that performs iterative steps to minimize the error function. We decided to make use of other optimizers in order to improve our SVR's performance.

5.2.1 Adam

Adam (short for Adaptive Moment Estimation) is an advanced optimization algorithm. It is widely used in training deep learning models due to its efficiency and ability to handle sparse gradients. Adam adjusts the learning rate for each parameter dynamically, making it suitable for problems involving multiple input parameters and large datasets. Adam includes mechanisms to correct biases in the estimates of the first and second moments, ensuring more reliable and accurate parameter updates (Ghimire et al., 2022). The library used for the Adam optimizer is Pytorch.

5.2.2 Gray Wolf Optimizer

The Gray Wolf Optimizer (GWO) is a nature-inspired algorithm based on the social hierarchy and hunting behavior of gray wolves in the wild. It was proposed by Seyedali Mirjalili in 2014 and has been applied successfully to various optimization problems. GWO simulates the leadership hierarchy of gray wolves, with four types of wolves: alpha, beta, delta, and omega. The alpha wolves lead the pack, followed by beta and delta wolves, with omegas being the lowest-ranking members. The algorithm mimics the cooperative hunting process of gray wolves, including tracking, encircling, and attacking prey. This process is modeled mathematically to perform optimization. GWO balances exploration (searching for new solutions) and exploitation (refining existing solutions) effectively, making it a robust optimizer (R. Liu et al., 2021).

5.3 Results

The tables below give the performances of SVR for both cities using the different optimizers.

Table 6: Performance of different optimizers for Lahore's SVR model

Lahore (LHR)	SVR	SVR with Adam Optimizer	SVR with GWO Optimizer
R	0.5244	0.545	0.64
Mean Square Error (MSE)	0.055	0.035	0.009
Mean Absolute Error (MAE)	0.180	0.11	0.04
Root Mean Square Error (RMSE)	0.235	0.187	0.094

Table 7: Performance of different optimizers for Karachi's SVR model

Karachi (KHI)	SVR	SVR with Adam Optimizer	SVR with GWO Optimizer
R	0.4066	0.454	0.54
Mean Square Error (MSE)	0.018	0.016	0.014
Mean Absolute Error (MAE)	0.110	0.104	0.08
Root Mean Square Error (RMSE)	0.136	0.126	0.118

5.4 Comparison with Literature

Table 8: Comparison of our SVR models with models from literature

	(Jing et al., 2020)	(Zaheer et al., 2023)	Lahore Model	Karachi Model
R	0.87	0.77	0.64	0.54
Mean Square Error (MSE)	0.0009	0.0049	0.09	0.014
Mean Absolute Error (MAE)	-	0.06	0.04	0.08
Root Mean Square Error (RMSE)	0.03	0.07	0.094	0.118

5.5 Discussion

As can be seen from the tables above, the optimizers only marginally improved the performances of our SVR. Even after improvement, our models did not perform as well as models provided by the literature. The possible reasons for the performance of our models are discussed below.

5.5.1 Pollution

Both Karachi and Lahore are placed within the top 5 cities in Pakistan with poorest air quality. The roles of various air pollutants in influencing AOD levels is still being studied (Bilal et al., 2021)

but it is likely that our models' relatively low predictive abilities are due to the influence of air pollutants on AOD levels.

5.5.2 Poor quality data

Our validation test of the ERA-5 datasets did take note of the low accuracy of certain parameters (such as WD and WS in Lahore). It is possible that our models' accuracy is being affected by the quality of the data being used to train them.

Chapter 6: Conclusion & Recommendations

6.1 Conclusion

The findings of our project can be summarized below:

- The forward feature selection (FFS) found the following 5 input features to be the best: temperature, relative humidity, wind speed, wind direction, and day of the year (DOY).
- The best-performing model after optimization and hyper-tuning was the Support Vector Regression (SVR).
- The best performing optimizer was the Gray Wolf Optimizer (GWO).

Although our models did not achieve a desired level of accuracy, they still demonstrated the potential ML has in solving complex, non-linear problems in ESE (Environmental Science and Engineering).

Furthermore, it must be noted that an important aspect of modelling is defining the purpose of the model i.e. for what application it is intended to serve. Although, when compared to the literature, our models do not perform as highly, that does not mean that our models are redundant.

Schroedter-Homscheidt & Oumbe (2013) did a study that demonstrated that AOD models with as low a correlation as 0.45 and an RMSE as high as 0.28 could be used in the calculation of surface solar irradiance for use in the energy sector. Our models, in comparison, perform much better than these. This indicates a potential application of the models we developed.

6.2 Recommendations

There are two main ways that our models can be further improved.

6.2.1 Training the models in cleaner cities

As discussed previously, it is possible that the pollution in Karachi and Lahore hamper the ability of the models to approximate a relationship between the provided input features and the target variable. It is possible that a model trained in a cleaner city will do better at defining the relationship between the input features we provided and our target variable.

6.2.2 Using Different Data

The datasets used for the input features were ERA-5. We suggest that further studies make use of other available datasets, particularly datasets that display a higher correlation with ground values. It is likely that this will significantly improve the performance of models in predicting AOD.

6.2.3 Using Different Input Features

Several studies have demonstrated the relationship between AOD and other meteorological parameters, such as rainfall (Gautum et al., 2022), as well as air pollutants such as nitrous oxides (Nichol et al., 2020) and carbon monoxide (Buchholz et al., 2021). We suggest that future studies make use of these variables as input features and subsequently investigate their potential in predicting AOD.

Chapter 7: References

1. Zaheer, K., Saeed, S., & Tariq, S. (2023). Prediction of aerosol optical depth over Pakistan using novel hybrid machine learning model. *Acta Geophysica*, 71(4), 2009–2029. <https://doi.org/10.1007/s11600-023-01072-x>
2. Lemmouchi, F., Cuesta, J., Lachatré, M., Brajard, J., Coman, A., Beekmann, M., & Derognat, C. (2023). Machine Learning-Based Improvement of Aerosol Optical Depth from CHIMERE Simulations Using MODIS Satellite Observations. *Remote Sensing*, 15(6), 1510. <https://doi.org/10.3390/rs15061510>
3. Ranjan, A. K., Patra, A. K., & Gorai, A. K. (2020). A Review on Estimation of Particulate Matter from Satellite-Based Aerosol Optical Depth: Data, Methods, and Challenges. *Asia-Pacific Journal of Atmospheric Sciences*, 57(3), 679–699. <https://doi.org/10.1007/s13143-020-00215-0>
4. Zhong, S., Zhang, K., Bagheri, M., Burken, J. G., Gu, A., Li, B., Ma, X., Marrone, B. L., Ren, Z. J., Schrier, J., Shi, W., Tan, H., Wang, T., Wang, X., Wong, B. M., Xiao, X., Yu, X., Zhu, J. J., & Zhang, H. (2021). Machine Learning: New ideas and tools in environmental science and engineering. *Environmental Science & Technology*. <https://doi.org/10.1021/acs.est.1c01339>
5. Panahi, M., Gayen, A., Pourghasemi, H. R., Rezaie, F., & Lee, S. (2020). Spatial prediction of landslide susceptibility using hybrid support vector regression (SVR) and the adaptive neuro-fuzzy inference system (ANFIS) with various metaheuristic algorithms. *Science of the Total Environment*, 741, 139937. <https://doi.org/10.1016/j.scitotenv.2020.139937>
6. Huan, J., Li, H., Li, M., & Chen, B. (2020). Prediction of dissolved oxygen in aquaculture based on gradient boosting decision tree and long short-term memory network: A study of Chang Zhou fishery demonstration base, China. *Computers and Electronics in Agriculture*, 175, 105530. <https://doi.org/10.1016/j.compag.2020.105530>
7. He, S., Wu, J., Wang, D., & He, X. (2022). Predictive modeling of groundwater nitrate pollution and evaluating its main impact factors using random forest. *Chemosphere*, 290, 133388. <https://doi.org/10.1016/j.chemosphere.2021.133388>
8. Soares, P. M. M., Lima, D. C. A., & Nogueira, M. (2020). Global offshore wind energy resources using the new ERA-5 reanalysis. *Environmental Research Letters*, 15(10), 1040a2. <https://doi.org/10.1088/1748-9326/abb10d>
9. Vanella, D., Longo-Minnolo, G., Belfiore, O. R., Ramírez-Cuesta, J. M., Pappalardo, S., Consoli, S., D'Urso, G., Chirico, G. B., Coppola, A., Comegna, A., Toscano, A., Quarta, R., Provenzano, G., Ippolito, M., Castagna, A., & Gandolfi, C. (2022). Comparing the use of ERA5 reanalysis dataset and ground-based agrometeorological data under different climates and topography in Italy. *Journal of Hydrology. Regional Studies*, 42, 101182. <https://doi.org/10.1016/j.ejrh.2022.101182>
10. Basharat, U., Tariq, S., Chaudhry, M. N., Khan, M., Agyekum, E. B., Mbasso, W. F., & Kamel, S. (2023). Seasonal correlation of aerosols with soil moisture, evapotranspiration, and vegetation over Pakistan using remote sensing. *Heliyon*, 9(10), e20635. <https://doi.org/10.1016/j.heliyon.2023.e20635>

11. Gueymard, C. A. (2012). Temporal variability in direct and global irradiance at various time scales as affected by aerosols. *Solar Energy*, 86(12), 3544–3553. <https://doi.org/10.1016/j.solener.2012.01.013>
12. Nawaz, M., Tariq, S., Mariam, A., & Ul-Haq, Z. (2023). Investigating the variability in aerosol optical depth and associated population exposure risk in Pakistan using MAIAC data. *Water, Air and Soil Pollution/Water, Air & Soil Pollution*, 234(10). <https://doi.org/10.1007/s11270-023-06661-6>
13. Ali, G., Bao, Y., Ullah, W., Ullah, S., Guan, Q., Liu, X., Li, L., Lei, Y., Li, G., & Ma, J. (2020). Spatiotemporal Trends of Aerosols over Urban Regions in Pakistan and Their Possible Links to Meteorological Parameters. *Atmosphere*, 11(3), 306. <https://doi.org/10.3390/atmos11030306>
14. Jing, L., Feng, X., Jintuo, L., Rui, M., Weiliang, L., & Yongjun, L. (2017). Contrastive research of SVM and BP neural network in AOD prediction. *Chinese Control Conference (CCC)*. <https://doi.org/10.23919/chicc.2017.8027996>
15. Jin, Z., Wang, S., Ma, J., & Zhou, J. (2022). Estimating Hourly Full-Coverage Himawari-8 AHI AOD with Spatiotemporal Random Forest Model. *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*. <https://doi.org/10.1109/igarss46834.2022.9883712>
16. Zhang, T., He, W., Zheng, H., Cui, Y., Song, H., & Fu, S. (2021). Satellite-based ground PM_{2.5} estimation using a gradient boosting decision tree. *Chemosphere*, 268, 128801. <https://doi.org/10.1016/j.chemosphere.2020.128801>
17. Schroedter-Homscheidt, M., & Oumbe, A. (2013). Validation of an hourly resolved global aerosol model in answer to solar electricity generation information needs. *Atmospheric Chemistry and Physics*, 13(7), 3777–3791. <https://doi.org/10.5194/acp-13-3777-2013>
18. Perez, M. J. R., Keelin, P., Perez, R. R., Kubinieć, A., & Stackhouse, P. (2020). Observed Recent Trends in the Solar Resource across North America: Changing Cloud-cover, AOD and the Implications for PV Yield. *47th IEEE Photovoltaic Specialists Conference (PVSC)*. <https://doi.org/10.1109/pvsc45281.2020.9300861>
19. Tariq, S., Nawaz, H., Ul-Haq, Z., & Mehmood, U. (2021). Investigating the relationship of aerosols with enhanced vegetation index and meteorological parameters over Pakistan. *Atmospheric Pollution Research*, 12(6), 101080. <https://doi.org/10.1016/j.apr.2021.101080>
20. Zeb, B., Alam, K., Khan, R., Ditta, A., Iqbal, R., Elsadek, M. F., Raza, A., & Elshikh, M. S. (2024). Characteristics and optical properties of atmospheric aerosols based on long-term AERONET investigations in an urban environment of Pakistan. *Scientific Reports*, 14(1). <https://doi.org/10.1038/s41598-024-58981-0>
21. Prasad, P., Basha, G., & Ratnam, M. V. (2023). Impact of Relative Humidity on the vertical distribution of aerosols over India. *Atmospheric Research*, 281, 106468. <https://doi.org/10.1016/j.atmosres.2022.106468>
22. Omidvar, K., Dehghan, M., & Khosravi, Y. (2022). Assessment of relationship between aerosol optical depth (AOD) index, wind speed, and visibility in dust storms using genetic algorithm in central Iran (case study: Yazd Province). *Air Quality, Atmosphere & Health*, 15(10), 1745–1753. <https://doi.org/10.1007/s11869-022-01214-y>

23. Sun, K., Dai, G., Wu, S., Reitebuch, O., Baars, H., Liu, J., & Zhang, S. (2024). Effect of wind speed on marine aerosol optical properties over remote oceans with use of spaceborne lidar observations. *Atmospheric Chemistry and Physics*, 24(7), 4389–4409. <https://doi.org/10.5194/acp-24-4389-2024>
24. Kang, N., Deng, F., Khan, R., Kumar, K. R., Hu, K., Yu, X., Wang, X., & Devi, N. L. (2020). Temporal variations of PM concentrations, and its association with AOD and meteorology observed in Nanjing during the autumn and winter seasons of 2014–2017. *Journal of Atmospheric and Solar-terrestrial Physics*, 203, 105273. <https://doi.org/10.1016/j.jastp.2020.105273>
25. Zheng, C., Zhao, C., Zhu, Y., Wang, Y., Shi, X., Wu, X., Chen, T., Wu, F., & Qiu, Y. (2017). Analysis of influential factors for the relationship between PM_{2.5} and AOD in Beijing. *Atmospheric Chemistry and Physics*, 17(22), 13473–13489. <https://doi.org/10.5194/acp-17-13473-2017>
26. Yousefi, R., Wang, F., Ge, Q., Shaheen, A., & Kaskaoutis, D. G. (2023). Analysis of the Winter AOD Trends over Iran from 2000 to 2020 and Associated Meteorological Effects. *Remote Sensing*, 15(4), 905. <https://doi.org/10.3390/rs15040905>
27. Nakata, M., Kajino, M., & Sato, Y. (2021). Effects of mountains on aerosols determined by AERONET/DRAGON/J-ALPS measurements and regional model simulations. *Earth and Space Science*, 8(12). <https://doi.org/10.1029/2021ea00197>
28. Ahmad, M., Tariq, S., Alam, K., Anwar, S., & Ikram, M. (2020b). Long-term variation in aerosol optical properties and their climatic implications over major cities of Pakistan. *Journal of Atmospheric and Solar-terrestrial Physics*, 210, 105419. <https://doi.org/10.1016/j.jastp.2020.105419>
29. Khalid, B., Khalid, A., Muslim, S., Habib, A., Khan, K., Alvim, D. S., Shakoore, S., Mustafa, S., Zaheer, S., Zoon, M., Khan, A. H., Ilyas, S., & Chen, B. (2021b). Estimation of aerosol optical depth in relation to meteorological parameters over eastern and western routes of China Pakistan economic corridor. *Journal of Environmental Sciences/Journal of Environmental Sciences*, 99, 28–39. <https://doi.org/10.1016/j.jes.2020.04.045>
30. Qayyum, F., Tariq, S., Ul-Haq, Z., Mehmood, U., & Zeydan, Ö. (2022). Air pollution trends measured from MODIS and TROPOMI: AOD and CO over Pakistan. *Journal of Atmospheric Chemistry*, 79(3), 199–217. <https://doi.org/10.1007/s10874-022-09436-1>
31. Jiang, X., Wang, Y., Wang, L., Tao, M., Wang, J., Zhou, M., Bai, X., & Gui, L. (2024). Characteristics of Daytime-And-Nighttime AOD differences over China: a perspective from CALIOP satellite observations and GEOS-Chem model simulations. *Journal of Geophysical Research. Atmospheres*, 129(8). <https://doi.org/10.1029/2023jd039158>
32. Haider, R., Yasar, A., & Tabinda, A. B. (2017). Urban emission patterns at a Semi-Arid site in Lahore, Pakistan. *Polish Journal of Environmental Studies*, 26(1), 59–68. <https://doi.org/10.15244/pjoes/64284>
33. Irfan, M., Zhao, Z., Ahmad, M., & Mukeshimana, M. C. (2019). Solar energy development in Pakistan: barriers and policy recommendations. *Sustainability*, 11(4), 1206. <https://doi.org/10.3390/su11041206>
34. Wang, Z., Chen, P., Wang, R., An, Z., & Qiu, L. (2023). Estimation of PM_{2.5} concentrations with high spatiotemporal resolution in Beijing using the ERA5 dataset and

- machine learning models. *Advances in Space Research*, 71(8), 3150–3165. <https://doi.org/10.1016/j.asr.2022.12.016>
35. Lanzaco, B. L., Olcese, L. E., Palancar, G. G., & Toselli, B. M. (2017). An improved aerosol optical depth map based on Machine-Learning and MODIS data: development and application in South America. *Aerosol and Air Quality Research*, 17(6), 1623–1636. <https://doi.org/10.4209/aaqr.2016.11.0484>
36. He, L., Wang, L., Lin, A., Zhang, M., Bilal, M., & Tao, M. (2017). Aerosol Optical Properties and Associated Direct Radiative Forcing over the Yangtze River Basin during 2001–2015. *Remote Sensing*, 9(7), 746. <https://doi.org/10.3390/rs9070746>
37. Kumar, A. (2021). Significant characteristics of aerosol optical depth and cloud cover fraction over the South West region of India. *Indian Journal of Physics/Indian Journal of Physics/Indian Journal of Physics and Proceedings of the Indian Association for the Cultivation of Science*, 96(1), 1–12. <https://doi.org/10.1007/s12648-021-02091-4>
38. Cesnulyte, V., Lindfors, A. V., Pitkänen, M. R. A., Lehtinen, K. E. J., Morcrette, J., & Arola, A. (2014b). Comparing ECMWF AOD with AERONET observations at visible and UV wavelengths. *Atmospheric Chemistry and Physics*, 14(2), 593–608. <https://doi.org/10.5194/acp-14-593-2014>
39. Boukerche, A., Zheng, L., & Alfandi, O. (2020). Outlier detection. *ACM Computing Surveys*, 53(3), 1–37. <https://doi.org/10.1145/3381028>
40. Frery, A. C. (2023). Interquartile range. In *Encyclopedia of earth sciences series/Encyclopedia of earth sciences* (pp. 664–666). https://doi.org/10.1007/978-3-030-85040-1_165
41. Mijwil, M. M., Abdulqader, A. W., Ali, S. M., & Sadiq, A. T. (2023). Null-values imputation using different modification random forest algorithm. *IAES International Journal of Artificial Intelligence*, 12(1), 374. <https://doi.org/10.11591/ijai.v12.i1.pp374-383>
42. Joshi, K. (2024, January 11). Floating points in deep learning: Understanding the basics. *Medium*. <https://medium.com/@krinaljoshi/floating-points-in-deep-learning-understanding-the-basics-93459f77a266>
43. Testas, A. (2023). Random Forest Classification with Scikit-Learn and PySpark. In *Apress eBooks* (pp. 243–258). https://doi.org/10.1007/978-1-4842-9751-3_9
44. Chen, N., Yang, M., Du, W., & Huang, M. (2021). PM2.5 Estimation and Spatial-Temporal Pattern Analysis Based on the Modified Support Vector Regression Model and the 1 km Resolution MAIAC AOD in Hubei, China. *ISPRS International Journal of Geo-information*, 10(1), 31. <https://doi.org/10.3390/ijgi10010031>
45. Wang, Y., Wu, J., Hu, Z., & McLachlan, G. J. (2023). A new algorithm for support vector regression with automatic selection of hyperparameters. *Pattern Recognition*, 133, 108989. <https://doi.org/10.1016/j.patcog.2022.108989>
46. Ghimire, S., Bhandari, B., Casillas-Pérez, D., Deo, R. C., & Salcedo-Sanz, S. (2022). Hybrid deep CNN-SVR algorithm for solar radiation prediction problems in Queensland, Australia. *Engineering Applications of Artificial Intelligence*, 112, 104860. <https://doi.org/10.1016/j.engappai.2022.104860>
47. Liu, R., Peng, J., Leng, Y., Lee, S., Panahi, M., Chen, W., & Zhao, X. (2021). Hybrids of Support Vector Regression with Grey Wolf Optimizer and Firefly Algorithm for Spatial

- Prediction of Landslide Susceptibility. *Remote Sensing*, 13(24), 4966. <https://doi.org/10.3390/rs13244966>
48. Bilal, M., Mhawish, A., Nichol, J. E., Qiu, Z., Nazeer, M., Ali, M. A., De Leeuw, G., Levy, R. C., Wang, Y., Chen, Y., Wang, L., Shi, Y., Bleiweiss, M. P., Mazhar, U., Atique, L., & Ke, S. (2021b). Air pollution scenario over Pakistan: Characterization and ranking of extremely polluted cities using long-term concentrations of aerosols and trace gases. *Remote Sensing of Environment*, 264, 112617. <https://doi.org/10.1016/j.rse.2021.112617>
49. Gautam, S., Elizabeth, J., Gautam, A. S., Singh, K., & Abhilash, P. (2022). Impact assessment of aerosol optical depth on rainfall in Indian rural areas. *Aerosol Science and Engineering*, 6(2), 186–196. <https://doi.org/10.1007/s41810-022-00134-9>
50. Nichol, J. E., Bilal, M., Ali, M. A., & Qiu, Z. (2020). Air Pollution Scenario over China during COVID-19. *Remote Sensing*, 12(13), 2100. <https://doi.org/10.3390/rs12132100>
51. Buchholz, R. R., Worden, H. M., Park, M., Francis, G., Deeter, M. N., Edwards, D. P., Emmons, L. K., Gaubert, B., Gille, J., Martínez-Alonso, S., Tang, W., Kumar, R., Drummond, J. R., Clerbaux, C., George, M., Coheur, P., Hurtmans, D., Bowman, K. W., Luo, M., . . . Kulawik, S. S. (2021). Air pollution trends measured from Terra: CO and AOD over industrial, fire-prone, and background regions. *Remote Sensing of Environment*, 256, 112275. <https://doi.org/10.1016/j.rse.2020.112275>
52. Mukherjee, T., & Vinoj, V. (2019). Atmospheric aerosol optical depth and its variability over an urban location in Eastern India. *Natural Hazards*, 102(2), 591–605. <https://doi.org/10.1007/s11069-019-03636-x>

FYDP Report.pdf

ORIGINALITY REPORT

16%

SIMILARITY INDEX

12%

INTERNET SOURCES

11%

PUBLICATIONS

6%

STUDENT PAPERS

PRIMARY SOURCES

1	link.springer.com Internet Source	1%
2	Submitted to Higher Education Commission Pakistan Student Paper	1%
3	www.mdpi.com Internet Source	1%
4	Submitted to ctu Student Paper	1%
5	Meysam Alizamir, Kaywan Othman Ahmed, Sungwon Kim, Salim Heddami, AliReza Docheshmeh Gorgij, Sun Woo Chang. "Development of a robust daily soil temperature estimation in semi-arid continental climate using meteorological predictors based on computational intelligent paradigms", PLOS ONE, 2023 Publication	<1%
6	Muhammad Haseeb, Zainab Tahir, Syed Amer Mahmood, Saira Batool, Aqil Tariq, Linlin Lu,	<1%

Walid Soufan. "Spatio-temporal assessment of aerosol and cloud properties using MODIS satellite data and a HYSPLIT model: Implications for climate and agricultural systems", Atmospheric Environment: X, 2024
Publication

7 Submitted to University of Birmingham <1 %
Student Paper

8 Submitted to Universidad de Jaén <1 %
Student Paper

9 www.researchgate.net <1 %
Internet Source

10 iopscience.iop.org <1 %
Internet Source

11 researchoutput.csu.edu.au <1 %
Internet Source

12 pubmed.ncbi.nlm.nih.gov <1 %
Internet Source

13 Submitted to Tias Business School <1 %
Student Paper

14 Submitted to University of East London <1 %
Student Paper

15 docslib.org <1 %
Internet Source

16 Submitted to National Economics University

<1 %

17

Song He, Jianhua Wu, Dan Wang, Xiaodong He. "Predictive modeling of groundwater nitrate pollution and evaluating its main impact factors using random forest", *Chemosphere*, 2021

Publication

<1 %

18

Submitted to University of Newcastle

Student Paper

<1 %

19

iwaponline.com

Internet Source

<1 %

20

repository.ums.ac.id

Internet Source

<1 %

21

sciencepg.com

Internet Source

<1 %

22

mobt3ath.com

Internet Source

<1 %

23

Submitted to University of Zululand

Student Paper

<1 %

24

dspace.daffodilvarsity.edu.bd:8080

Internet Source

<1 %

25

www.frontiersin.org

Internet Source

<1 %

Submitted to Associatie K.U.Leuven

26

Student Paper

<1 %

27

Submitted to Australian National University

Student Paper

<1 %

28

www.nature.com

Internet Source

<1 %

29

apps.dtic.mil

Internet Source

<1 %

30

downloads.cambridge.edu.au

Internet Source

<1 %

31

downloads.hindawi.com

Internet Source

<1 %

32

hdl.handle.net

Internet Source

<1 %

33

www.coursehero.com

Internet Source

<1 %

34

d.docksci.com

Internet Source

<1 %

35

Submitted to BITS, Pilani-Dubai

Student Paper

<1 %

36

Submitted to Marquette University

Student Paper

<1 %

37

Ozyurt, D.B.. "Theory and practice of simultaneous data reconciliation and gross

<1 %

error detection for chemical processes",
Computers and Chemical Engineering,
20040315

Publication

38

drpress.org

Internet Source

<1 %

39

dspace.mit.edu

Internet Source

<1 %

40

liu.diva-portal.org

Internet Source

<1 %

41

Hamza Turabieh. "A Hybrid ANN-GWO Algorithm for Prediction of Heart Disease", American Journal of Operations Research, 2016

Publication

<1 %

42

Theogan Logan Pillay, Akshay Kumar Saha. "A Review of Metaheuristic Optimization Techniques for Effective Energy Conservation in Buildings", Energies, 2024

Publication

<1 %

43

Submitted to University of Leicester

Student Paper

<1 %

44

ai.usz.edu.pl

Internet Source

<1 %

45

utpedia.utp.edu.my

Internet Source

<1 %

46	Submitted to University of Stirling Student Paper	<1 %
47	Wen Ma, Jianli Ding, Rui Wang, Jinlong Wang. "Drivers of PM2.5 in the urban agglomeration on the northern slope of the Tianshan Mountains, China", Environmental Pollution, 2022 Publication	<1 %
48	aaqr.org Internet Source	<1 %
49	acikbilim.yok.gov.tr Internet Source	<1 %
50	c.coek.info Internet Source	<1 %
51	hh.diva-portal.org Internet Source	<1 %
52	ikee.lib.auth.gr Internet Source	<1 %
53	www.tandfonline.com Internet Source	<1 %
54	Ahmad Qadeib Alban, Ammar Abulibdeh, Lanouar Charfeddine, Rawan Abulibdeh, Abdelgadir Abuelgasim. "A Comprehensive Machine and Deep Learning Approach for Aerosol Optical Depth Forecasting: New	<1 %

Evidence from the Arabian Peninsula", Earth Systems and Environment, 2024

Publication

55

Navneet Kumar, Anirban Middey. "Interaction of aerosol with meteorological parameters and its effect on the cash crop in the Vidarbha region of Maharashtra, India", International Journal of Biometeorology, 2022

Publication

56

Sami Ghordoyee Milan, Abbas Roozbahani, Naser Arya Azar, Saman Javadi. "Development of adaptive neuro fuzzy inference system – Evolutionary algorithms hybrid models (ANFIS-EA) for prediction of optimal groundwater exploitation", Journal of Hydrology, 2021

Publication

57

Yunxiao Chen, Mingliang Bai, Yilan Zhang, Jinfu Liu, Daren Yu. "Proactively selection of input variables based on information gain factors for deep learning models in short-term solar irradiance forecasting", Energy, 2023

Publication

58

earthzine.org

Internet Source

59

etd.ummy.ac.id

Internet Source

<1 %

<1 %

<1 %

<1 %

<1 %

60	journal.esj.edu.iq Internet Source	<1 %
61	nemertes.library.upatras.gr Internet Source	<1 %
62	pr.hec.gov.pk Internet Source	<1 %
63	publishup.uni-potsdam.de Internet Source	<1 %
64	"Climate Change and Human Adaptation in India", Springer Science and Business Media LLC, 2024 Publication	<1 %
65	Andrea Giudici, Mikolaj Zbigniew Jankowski, Rain Männikus, Fatemeh Najafzadeh, Ülo Suursaar, Tarmo Soomere. "A comparison of Baltic Sea wave properties simulated using two modelled wind data sets", Estuarine, Coastal and Shelf Science, 2023 Publication	<1 %
66	Düzgün, Cansu. "Quantifying Simulated Deep Convective Transport and Its Sensitivity to Lightning Data Assimilation in a Supercell and Mesoscale Convective System", The Florida State University, 2023 Publication	<1 %

67

Emadaldin Mohammadi Golafshani, Mehrdad Arashpour, Alireza Kashani. "Green mix design of rubbercrete using machine learning-based ensemble model and constrained multi-objective optimization", *Journal of Cleaner Production*, 2021

Publication

<1 %

68

Gohar Ali, Yansong Bao, Waheed Ullah, Safi Ullah, Qin Guan, Xulin Liu, Lin Li, Yuhong Lei, Guangwen Li, Jun Ma. "Spatiotemporal Trends of Aerosols over Urban Regions in Pakistan and Their Possible Links to Meteorological Parameters", *Atmosphere*, 2020

Publication

<1 %

69

Kadhirvel Boopathi, B Krishnan, J Bastin, Suchit Hoti, Reddy Prasad D.M. "Feasibility study of offshore wind energy on the coast of Sri Lanka", *International Journal of Green Energy*, 2023

Publication

<1 %

70

Lin, Chin-An. "The Impact of Wildfire Activities on Air Quality and Solar Energy in New York and California", *State University of New York at Albany*, 2024

Publication

<1 %

71

Pakize Erdogmus, Fatih Kayaalp. "Chapter 1 Introductory Chapter: Clustering with Nature-

<1 %

Inspired Optimization Algorithms", IntechOpen, 2020

Publication

72

Ruiz-Arias, José A., Jimy Dudhia, Francisco J. Santos-Alamillos, and David Pozo-Vázquez. "Surface clear-sky shortwave radiative closure intercomparisons in the Weather Research and Forecasting model : SW CLOSURE INTERCOMPARISONS IN WRF", Journal of Geophysical Research Atmospheres, 2013.

Publication

<1 %

73

Wahyu Luqmanul Hakim, Fatemeh Rezaie, Arip Syaripudin Nur, Mahdi Panahi, Khabat Khosravi, Chang-Wook Lee, Saro Lee. "Convolutional neural network (CNN) with metaheuristic optimization algorithms for landslide susceptibility mapping in Icheon, South Korea", Journal of Environmental Management, 2022

Publication

<1 %

74

aaltodoc2.org.aalto.fi

Internet Source

<1 %

75

acp.copernicus.org

Internet Source

<1 %

76

arxiv.org

Internet Source

<1 %

77

assets.researchsquare.com

Internet Source

<1 %

78

cdr.lib.unc.edu

Internet Source

<1 %

79

d197for5662m48.cloudfront.net

Internet Source

<1 %

80

de.slideshare.net

Internet Source

<1 %

81

delibra.bg.polsl.pl

Internet Source

<1 %

82

eprints.usq.edu.au

Internet Source

<1 %

83

gmd.copernicus.org

Internet Source

<1 %

84

hal-mines-paristech.archives-ouvertes.fr

Internet Source

<1 %

85

klme.nuist.edu.cn

Internet Source

<1 %

86

mdpi-res.com

Internet Source

<1 %

87

par.nsf.gov

Internet Source

<1 %

88

repositorio.aemet.es

Internet Source

<1 %

89

www.atmos-chem-phys.net

Internet Source

<1 %

90

www.scilit.net

Internet Source

<1 %

91

Rehana Khan, Kanike Raghavendra Kumar, Tianliang Zhao, Waheed Ullah, Gerrit de Leeuw. "Interdecadal Changes in Aerosol Optical Depth over Pakistan Based on the MERRA-2 Reanalysis Data during 1980–2018", *Remote Sensing*, 2021

Publication

<1 %

92

Pedro M M Soares, Daniela C A Lima, Miguel Nogueira. "Global offshore wind energy resources using the new ERA-5 reanalysis", *Environmental Research Letters*, 2020

Publication

<1 %

93

Yanhong Zou, Yuting Chen, Hao Deng. "Gradient Boosting Decision Tree for Lithology Identification with Well Logs: A Case Study of Zhaoxian Gold Deposit, Shandong Peninsula, China", *Natural Resources Research*, 2021

Publication

<1 %

94

You-Gan Wang, Jinran Wu, Zhi-Hua Hu, Geoffrey J. McLachlan. "A new algorithm for support vector regression with automatic

<1 %

selection of hyperparameters", Pattern Recognition, 2023

Publication

Exclude quotes On

Exclude matches Off

Exclude bibliography On