International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
IJISAE
ISSN:2147-6799 www.ijisae.org Original Research Paper

Forecasting Ambient Air Pollution of Ludhiana, Punjab Based on Mamdani Inference System

Bachandeep Singh Bhathal[1], Dr. Gaurav Gupta[2], Dr. Brahmaleen K. Sidhu[2]

Submitted:12/03/2024 **Revised**: 27/04/2024 **Accepted**: 04/05/2024

Abstract: Air is a necessity for living things to survive, and a drastic change has taken place in the air pollution levels due to pandemics. The prediction of air pollution in ambient air had become a requirement to save mankind and other living things. This paper presents a system for prediction of ambient air quality. The proposed system is based on the Mamdani Fuzzy Inference (MFI) system. The required input data is collected from the specific area of Ludhiana, Punjab (India). The pollutants covered are Ammonia, $PM_{2.5}$ µm, PM_{10} µm, Carbon Monoxide, and Sulfur Dioxide. Around fifty rules were framed in this model for the day-to-day prediction process. The results were obtained and compared by correlation, Index of Agreement (IOA), Mean Absolute (MA) Percentage error, Mean Absolute (MA) error, and Root Mean Square (RMS) error, where the correlation of min-max was 0.9268 depicts the positive results. The results were found to be approximately 93% accurate to the real values.

Keywords: *Mamdani, Fuzzy Inference System, Air Quality Index, Air Pollution, Ambient Air*

1. INTRODUCTION

For a long time, people have benefited from weather forecasting systems. These systems help us predict the weather by processing a vast amount of data. Similarly, there is a need for a system that can predict daily ambient air quality. Just like weather forecasts, air quality predictions require a lot of data, which is considered one of the most valuable resources in the world. This data comes from many sources. In the context of air quality prediction, data can be categorized based on its origin. For example, some data comes from contaminants in water, air, or soil caused by living organisms like bacteria and fungi. Other types of data involve exposure to chemicals, which can be in the form of solid particles, droplets, or gases. Dealing with the increasing air pollution resulting from technological advancements is one of the most significant challenges the world faces today. Pollution has increased considerably in recent years, posing a threat to our health and the environment. This rise in pollution is a complex issue that affects ecosystems and human well-being. It is crucial to recognize the various sources and types of pollution to address this global challenge effectively. To tackle air pollution and protect environmental health, it is vital to have precise exposure estimations. Accurate data collection and analysis are essential for understanding how pollutants affect different regions and populations. With accurate predictions, policymakers and scientists can develop strategies to mitigate pollution and its harmful effects. This requires collaboration among researchers, governments, and communities to ensure a healthier future for everyone.

To predict air quality, a variety of methods and algorithms are available to process data effectively. One innovative approach introduced by [1] is the use of a fuzzy system. Fuzzy systems are mathematical frameworks that interpret data using linguistic terms, making them well-suited for handling systems with uncertainties or ambiguities in data interpretation. This means that fuzzy systems can deal with data that is not always clear-cut or straightforward, making them a powerful tool for air quality prediction. Fuzzy systems offer several advantages over traditional regression systems. One significant benefit is their ability to handle large amounts of computational power and manage a vast number of variables, even when these variables have little correlation with each other. This is important because air quality data often involves many different factors that can influence the results. For instance, weather conditions, pollution sources, and geographical features can all impact air quality, and these factors may not always be directly related. Traditional regression models can struggle to scale effectively when dealing with such a large number of variables, especially if there are weak or complex relationships among them. Fuzzy systems, on the other hand, are designed to manage these complexities by using a more flexible approach to data interpretation. This flexibility allows fuzzy systems to provide more accurate predictions even when dealing with uncertain or incomplete data.

To demonstrate the effectiveness of the fuzzy system, an air quality index (AQI) model has been used as a case study. The AQI model provides a comprehensive initial database of inputs and exhibits significant variations in pollution

[1]*Assistant Professor, University Institute of Computing, Chandigarh University, Gharuan*
[2]*Assistant Professor, Department of CSE, Punjabi University, Patiala*

concentrations, which are crucial for a fuzzy inference system (FIS) to showcase its capabilities. This variation in data is necessary because it helps the model learn and adapt to different scenarios, improving its overall performance. Even when the relationships between the variables used in the model are not well understood, the fuzzy system can still function effectively as a pioneering tool for predicting complex systems. This means that even if we don't fully know how all the factors are interconnected, the fuzzy system can still make accurate predictions. This is particularly useful in environmental studies where many variables can influence outcomes. The focus on air pollution in this model is due to the significant impact of ambient air pollution on ecosystems and public health. Air pollution poses a threat to both human health and the environment, making it essential to have reliable prediction models that can help mitigate its effects. By using the fuzzy system in the AQI model, researchers can better understand and predict pollution levels, leading to more effective strategies for reducing pollution. Calculating the ambient air quality index is also important for monitoring a country's economic progress and its ecological footprint. By understanding how pollution levels change over time, governments and organizations can implement smarter practices to minimize environmental impact while promoting sustainable growth. The AQI serves as a vital tool for assessing the overall health of the environment and guiding policy decisions. Currently, measuring the AQI involves determining how each pollutant compares to acceptable standards for different pollution categories. This approach relies on predefined rules for each pollutant, which can limit the accuracy of predictions. To enhance this process and achieve better results, the Mamdani fuzzy system is employed. The Mamdani fuzzy system is known for its ability to handle complex and uncertain data, making it ideal for improving AQI predictions. [2].

The Mamdani fuzzy inference system has been utilized to create a non-linear regression model for predicting air quality. This type of system is especially useful for dealing with complex relationships between variables, allowing for more accurate predictions. Once the model was developed, it was tested using several performance parameters, including the Index of Agreement (IOA), root mean square error (RMSE), correlation, mean absolute error (MAE), and mean absolute percentage error (MAPE). These parameters help evaluate how well the model predicts air quality by comparing its predictions to actual observations. The data used to test this model was collected from Ludhiana, Punjab. This data set includes a variety of pollutants, ranging from ammonia to particulate matter with a diameter of 10 micrometers (PM10), as well as carbon monoxide (CO) and sulfur dioxide (SO2). These pollutants are significant contributors to air pollution and can have serious health and environmental impacts. By including a wide range of pollutants, the model can better capture the complexity of air quality in the region. Once the model was applied to this data, the results were calculated and measured according to the air quality index (AQI). The AQI is a standard tool used to communicate how polluted the air currently is or how polluted it is forecast to become. It takes into account the concentration levels of different pollutants and translates these into a single number or category that indicates the overall air quality. The use of the Mamdani fuzzy system in this context allows for a more nuanced understanding of air quality, as it can effectively manage the uncertainties and non-linear relationships inherent in environmental data. By using the Mamdani fuzzy inference system, the model can provide more accurate predictions of air quality, which are crucial for both public health and environmental management. Policymakers and environmental agencies can use these predictions to implement measures that reduce pollution levels and protect communities from harmful exposure. Moreover, this approach can serve as a model for other regions facing similar air quality challenges, helping to improve air quality management practices worldwide.

2. FORECASTING RELATED WORK

[3] has designed a collaborative recommender system specifically to help farmers make informed decisions about their crops. This system predicts or recommends which crops are most suitable for a farmer's location, taking into account various factors such as historical weather conditions. By analyzing past weather patterns, the system can offer insights into which crops are likely to thrive in a particular area. In addition to crop recommendations, the system also suggests other essential agricultural inputs like seeds, fertilizers, pesticides, and tools that are useful for harvesting. These recommendations aim to optimize crop yield and efficiency, ensuring that farmers have the resources they need to succeed. The system uses the Mamdani fuzzy inference system to enhance its predictions. This system provides farmers with a preliminary idea of crop yield and sowing by forecasting potential outcomes. The fuzzy system can handle the uncertainties and variations in agricultural data, offering more accurate and reliable predictions for farmers. By utilizing this collaborative recommender system, farmers can make better decisions about which crops to plant and how to manage them effectively. The system's recommendations are tailored to each farmer's specific location and conditions, making it a valuable tool for increasing productivity and sustainability in agriculture. This approach is particularly beneficial for farmers who may not have access to extensive agricultural research or expertise. By leveraging advanced technology and data analysis, the system empowers farmers to make informed choices that can lead to improved crop yields and economic outcomes.

[4] developed a Mamdani Fuzzy Inference System (MFIS) to evaluate the green performance of companies. This system is designed to help businesses assess how environmentally friendly their operations are by analyzing

various factors. The MFIS uses several fuzzy inference systems (FISs) to process different dimensions of green supply chain management (GSCM). GSCM is a concept that integrates environmental thinking into supply chain management, covering aspects such as product design, material sourcing, and manufacturing processes. In this approach, each dimension of GSCM is evaluated individually, and the results are then used as inputs for the main FIS. By combining these inputs, the MFIS can provide a comprehensive assessment of a company's green performance. This multi-step process allows the system to capture a wide range of environmental factors and provide more accurate evaluations. The results generated by the MFIS offer valuable insights into various aspects of a company's operations. For example, it provides information about the amount of solid or liquid waste produced, how energy is used, and the utilization of other resources. These insights are crucial for companies looking to improve their environmental impact and make more sustainable decisions. By evaluating these factors, companies can identify areas where they need to reduce waste or improve resource efficiency. This information helps businesses make informed decisions about how to enhance their environmental performance and align with sustainable practices. As a result, the MFIS supports companies in their efforts to reduce their ecological footprint and contribute to environmental protection. The use of the MFIS for evaluating green performance is particularly important in today's business landscape, where sustainability is increasingly becoming a priority for companies and consumers alike. By providing detailed assessments and actionable recommendations, the MFIS helps businesses not only comply with environmental regulations but also gain a competitive advantage by demonstrating their commitment to sustainability.

[5] employed the Mamdani Fuzzy Inference System (MFIS) to predict and enhance intelligent transportation systems in smart cities. This initiative aims to improve the quality of life for city residents by making urban transportation more efficient and responsive to their needs. In the rapidly evolving environment of smart cities, effective transportation management is essential to ensure smooth mobility and reduce congestion. The MFIS model was specifically applied to evaluate traffic congestion control in smart cities. Traffic congestion is a significant issue in urban areas, causing delays, increasing pollution, and reducing the overall quality of life. By using the Mamdani Fuzzy Inference System, Bachandeep's model can analyze various factors that contribute to traffic congestion and predict future traffic patterns. This predictive capability allows city planners and transportation authorities to develop strategies to alleviate congestion and improve traffic flow. The model has proven to be beneficial in predicting future traffic conditions, offering insights that help in planning and managing urban transportation systems more effectively. By understanding potential congestion points and traffic patterns, cities can implement measures such as adjusting traffic signal timings, optimizing public transportation routes, and promoting alternative transportation modes to minimize congestion. In addition to traffic management, the model suggests that considering other factors like environmental conditions, healthcare, and security surveillance can enhance predictions and decision-making. Environmental conditions, such as air quality and weather, can significantly impact transportation systems. For instance, poor air quality may lead to restrictions on vehicle usage, while adverse weather conditions can cause disruptions in traffic flow. Healthcare considerations are also important, as efficient transportation systems can improve access to medical facilities and reduce response times for emergency services. Security surveillance, on the other hand, can play a crucial role in ensuring the safety of transportation systems by monitoring for potential threats and coordinating emergency responses. By integrating these additional factors into the transportation model, cities can develop more comprehensive and adaptive solutions to urban mobility challenges. This holistic approach not only addresses traffic congestion but also supports broader city goals, such as sustainability, public health, and safety.

[6] has developed a rain detection system that utilizes various inputs to accurately determine weather conditions. This system is designed to detect rain and provide information about rainy weather levels, helping individuals and organizations make better decisions based on weather conditions. The system gathers inputs using an Arduino, a versatile microcontroller that can collect data from different sensors. These sensors include humidity, rain, and temperature sensors, which provide essential information about the current weather. The Arduino collects this data and sends it sequentially to the Fuzzy Inference System (FIS) using a specific communication protocol. Once the data reaches the FIS, it is processed using a specialized algorithm. This algorithm analyzes the input data from the sensors and compares the results to assess the likelihood and intensity of rain. The FIS is particularly useful in this context because it can handle the inherent uncertainties and variations in weather data, leading to more accurate and reliable predictions. By combining data from multiple sensors, the rain detection system can offer a comprehensive assessment of weather conditions. It considers various factors, such as humidity levels, temperature, and rain intensity, to provide a detailed understanding of the current weather. This information is valuable for planning outdoor activities, managing agricultural operations, and ensuring the safety of transportation systems during rainy conditions. The results of the system have proven to be effective and responsive, providing timely and accurate information about rainy weather levels. This responsiveness is crucial in situations where quick decisions are needed, such as when deciding whether to proceed with outdoor events or take

precautionary measures to protect crops from potential damage.

3. FORECASTING METHODOLOGY FOR AMBIENT AIR POLLUTION

In our system, we use fuzzy inference, where the outcomes of the fuzzy rules are expressed as crisp numbers. A crisp number is a clear and precise value, as opposed to a range or degree of possibility often used in fuzzy logic. This approach simplifies the calculation process, making it one of the main advantages of using fuzzy inference. One of the significant benefits of this method is that it allows for straightforward integration of new rules into the system. In traditional systems, adding new rules or adjusting existing ones can be complex and time-consuming. However, with fuzzy systems, you can easily incorporate additional rules as needed, without having to overhaul the entire system. This flexibility makes it easier to adapt and improve the system over time. Furthermore, fuzzy inference systems allow for the easy addition of new membership functions. Membership functions are used in fuzzy logic to define how each input variable relates to different categories or sets. Adding new membership functions enables the system to handle a broader range of inputs and scenarios, enhancing its overall accuracy and responsiveness. This adaptability is an attractive feature of fuzzy systems, as it allows them to evolve and improve continuously. As new data becomes available or as the requirements of the system change, it can be updated to incorporate this information, ensuring it remains effective and relevant.

The data for each pollutant has been centralized at a single location, meaning all the information for each type of pollutant is collected and stored together. This centralized approach helps in organizing and managing the data more effectively. To analyze this data, several mathematical calculations were performed, including finding the mean, median, minimum, and maximum values. These calculations provide a summary of the data and help in understanding its general characteristics. The mean gives the average value, the median shows the middle value when the data is ordered, while the minimum and maximum values indicate the range of the data. This processed data was then used in the Mamdani Fuzzy Inference System (MFIS). In the MFIS, various membership rules are applied to interpret the data and make predictions. Membership rules define how different data values fit into different categories or fuzzy sets, helping the system to evaluate the data more effectively. Once the MFIS has processed the data using these rules, the results are compared with several performance parameters to assess the accuracy and effectiveness of the model. These parameters include:

- **Index of Agreement (IOA):** Measures how well the model's predictions match the observed data.
- **Root Mean Square Error (RMSE):** Calculates the square root of the average squared differences between predicted and observed values, providing an indication of prediction accuracy.
- **Correlation:** Assesses the strength and direction of the relationship between predicted and actual values.
- **Mean Absolute Error (MAE):** Computes the average absolute differences between predicted and observed values, giving a straightforward measure of prediction accuracy.
- **Mean Absolute Percentage Error (MAPE):** Measures the average percentage error between predictions and actual values, providing a relative measure of accuracy.

By comparing the calculated values with these parameters, we can evaluate how well the Mamdani Fuzzy Inference System performs in predicting pollution levels. This comparison helps in fine-tuning the model and ensuring that it provides accurate and reliable predictions for environmental monitoring and decision-making.

3.1 Fuzzy Inference System

The concept of set membership is crucial in describing how fuzzy systems work. In fuzzy logic, set membership helps define how objects or elements fit into a particular set or category within a broader universe of discourse. In traditional (or classic) set theory, an object either belongs to a set or it does not. This means the membership is binary: an object is either a member of the set (1) or not a member (0). However, fuzzy logic introduces a more flexible approach through fuzzy sets, where membership is not simply binary but can vary continuously. In a fuzzy set, the membership of an object is described in terms of a degree of membership. This degree is represented by a membership function, which assigns a value between 0 and 1 to each object. Here's what these values mean:

- **Membership Function Value of 0:** Indicates that the object does not belong to the fuzzy set at all.
- **Membership Function Value of 1:** Indicates that the object fully belongs to the fuzzy set.
- **Membership Function Value Between 0 and 1:** Represents partial membership, meaning the object partially belongs to the fuzzy set. The exact value reflects the degree or extent to which the object fits into the set.

This concept allows for a more nuanced and imprecise representation of set membership compared to classic set theory. For example, instead of categorizing an object as simply "hot" or "cold," a fuzzy set might categorize it as "somewhat hot" or "fairly cold," with membership values reflecting degrees of heat or coldness.

Here now if an element "x" in the nature "X" of a crisp sets can be the member of any crisp set "A" or it cannot be the member.

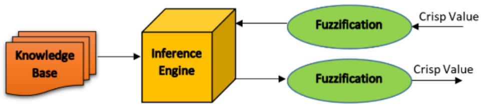

Fig 1: Simplified Fuzzy Inference System

The mathematical representation of this this binary issue of membership with the indicator function is as:

$$X_A(x) = \begin{cases} 1, x\epsilon A \\ 0, x\epsilon A \end{cases}$$

Let us take an example to explain this concept, consider two colors red and blue in which the red can be used as stating hot and the blue can be used as stating cold situations then in this the red will be taken as a 1 which will denote hot, whereas the blue will be considered as 0 which is for cold. The output as a membership of the function will be as 0 for blue which is cold and 1 is for red which is hot. [7]

The fuzzy definition of a set helps the system to get introduced with the indefinite elements which are introduced in the processing of data and the same can be represented as:

$$\mu_A(x)\epsilon[0,1]$$

Let us take an element pink which is a color and is defined in the set mentioned above will also occupies a membership value between 0 and 1. Therefore the set "A" can be manipulated with the color set "A" which can then defined as membership function too.

According to the following illustration the structure of the fuzzy inference system can be represented,

Fig 2: Fuzzy Inference System

The FIS process involves three stages which are as Fuzzification, Rule Evaluation, and De-Fuzzification. [8]

- **Fuzzification:** This stage is responsible for translating precise input values into fuzzy sets. The process begins with crisp data sets, which are exact and well-defined measurements. Fuzzification converts these crisp values into fuzzy sets, representing degrees of membership in various categories. This transformation involves assigning truth values to the inputs based on predefined membership functions. The result is a set of fuzzy values that reflect the degree to which each input belongs to different fuzzy sets. These fuzzy values are then used in the next stage of the system for further processing.
- **Inference Engine/Rule Evaluation:** In this stage, the fuzzy values obtained from the fuzzification process are used for further computation. The inference engine applies a set of rules from the knowledge database to these fuzzy values. These rules are typically formulated in an IF-THEN format and are used to derive conclusions or make decisions based on the input data. The inference engine combines the fuzzy values according to these rules to generate an intermediate fuzzy output. This output remains in fuzzy form and reflects the system's decision-making process based on the provided inputs and the rules applied.
- **De-Fuzzification:** The final stage involves converting the fuzzy output into a crisp value. De-fuzzification takes the fuzzy results produced by the inference engine and translates them back into precise, well-defined values. This process involves determining the best representative crisp value from the fuzzy output, which may include calculating a weighted average or another method to consolidate the fuzzy results. The goal of de-fuzzification is to provide a clear and actionable output, reflecting the degree of membership in a specific category or decision outcome, which can then be used for practical applications or further analysis.

Among functioning of all the stages, crucial one is knowledge base. [9]

3.2 Mamdani fuzzy inference system

The foundation of FIS is relaying on the rule of IF-THEN property.

"IF 'x' is 'A' then 'y' is 'B'"

This is then translated into the fuzzy inference engine. But MFIS generates the graphical inference technique which are used to uncover the results for the corresponding past history inputs. The membership functions and the number of inputs are directly affecting the rules used in the system. Various assumption and combination techniques are followed by the membership functions. [10]

The well-known form of the rules in the MIS can be described as following equation:

If x_1 is Z_1^k and x_2 is Z_2^k then y^k is B^k for k-1,2,3….r.

Where Z_1^k and Z_2^k are fuzzy sets signifying the kth antecedent pair and Bk is the kth resultant.

In this the accumulation can be taken as sum of model; the repercussion can be defined as min, or product.

Here for the system now we can consider the input (i) and input (j) be the previous circumstances; and for these parameters the repercussion function can be as follow: [11]

$$\mu_{c_k}(y) = \max[\min[\mu_{Z_1^k}(input(i)), \mu_{Z_2^k}(input(j))]]$$

For k = 1,2,3,…,r

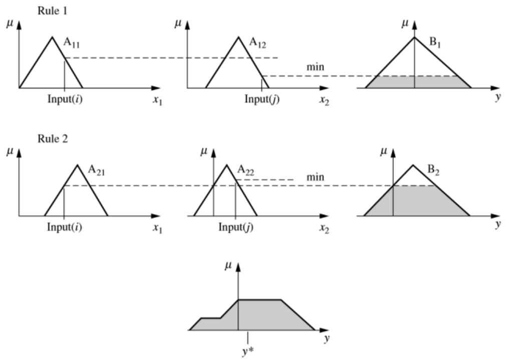

Fig 3: Max-Min Mamdani System *[11]*

For the implementation of max-product, the products of the forerunner membership functions are the subsequent membership functions whereas the huge number of portion is taken as a final resultant. This can be defined as,

$$\mu_{c_k}(y) = \max\left[\mu_{Z_1^k}(input(i))_k * \mu_{Z_2^k}(input(j))\right]$$

For k=1,2,3,….,r

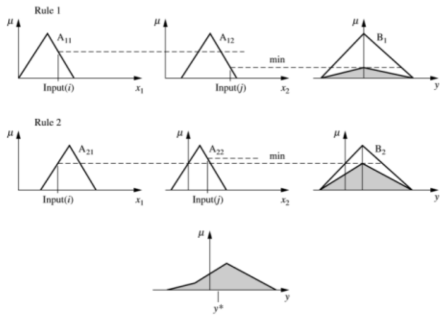

Fig 4: Max-Product Mamdani System *[11]*

In this research, a range of hazardous pollutants has been taken into consideration, including Carbon Monoxide (CO), Sulfur Dioxide (SO_2), Ammonia (NH_3), Particulate Matter (PM_{10}), and Particulate Matter ($PM_{2.5}$). These pollutants are known for their adverse effects on both human health and the environment. To analyze and predict air quality, the study relies on historical data of these pollutants. This data is recorded hourly and collected at a centralized location. By aggregating and centralizing this historical data, the research can effectively use it in the application model for making predictions. The main parameter that has been considered for this research is the Air Quality Index (AQI). The AQI is a composite measure that reflects the overall level of air pollution based on the concentrations of the pollutants mentioned. By using historical records of CO, SO_2, NH_3, PM_{10}, and $PM_{2.5}$, the application model is able to estimate the current air quality and make predictions about future pollution levels. The centralized historical data serves as the foundation for these predictions, allowing for a comprehensive assessment of air quality through the AQI. [12]

The rules employed in this model are centered on the Air Quality Index (AQI). Specifically, the model operates on the principle that if any pollutant has a high concentration, the resulting AQI will be high, indicating poor air quality. Conversely, if the concentration of pollutants is low, the AQI will also be low, reflecting better air quality. The model uses these rules to evaluate and simulate air quality based on historical data of various pollutants. By analyzing the concentration levels of pollutants such as CO, SO_2, NH_3, PM_{10}, and $PM_{2.5}$, the system determines the AQI value. High pollutant concentrations translate to higher AQI levels, signaling increased pollution, while lower pollutant levels result in lower AQI values, indicating cleaner air. After applying these rules and simulating different techniques and assumptions, the model generates results that reflect the predicted air quality. These results help in understanding and forecasting the air quality based on the pollutant concentrations and their impact on the AQI. [13]

4. RESULTS

The MFIS simulation are done using MatLab after following steps and results different parameter values are tabulated in the table -1.

Steps to simulate system and obtain results

- The data used to assess the system's capabilities is sourced from a combination of online platforms and offline methods. For this research, Ludhiana has been chosen as the focal point due to its significance as an industrial hub in Punjab. Specifically, measurements are taken at Samrala Chowk, a prominent location in Ludhiana, Punjab, India. This strategic choice of location ensures that the data reflects the diverse industrial activities and environmental conditions characteristic of this region.
- We have collected a total of 6,839 samples, which are used as input data for the system. These samples include measurements of five major ambient air pollutants: Carbon Monoxide (CO), Sulfur Dioxide (SO_2), Ammonia (NH_3), Particulate Matter (PM_{10}), and Particulate Matter ($PM_{2.5}$).
- The measurement level for all the pollutant is in the unit of ($\mu g/m^3$).
- Air Quality Index has its own measures which are divided into various five semantic metrics named as good, moderate, poor, very poor and severe
- The pollution levels divided into semantic metrics as low, moderate and poor.
- All of the samples are taken from historical 2018 and 2019 years.

- No modification or manipulation in acquired data has been done such as scaling or normalization before input.

Table 1: Simulation Results

Measurement Factors	Max - Min	Sum - Product	Max - Product
Correlation	0.9268	0.9713	0.922
IOA	0.9279	0.9358	0.9386
MA Percentage Error	0.1989	0.2358	0.1671
MA Error	33.3386	32.8495	29.7418
RMS Error	40.5516	38.3097	40.507

The values are compared across different metrics and scales as detailed in the table above. Specifically, the correlation coefficients for the Max-Min, Sum-Product, and Max-Product methods are 0.93, 0.97, and 0.92, respectively. These values are very close to 1, indicating a high level of accuracy, approximately 93%. This suggests that the methods provide reliable results in terms of correlation with the actual data. In addition to these correlation metrics, the Max-Min, Sum-Product, and Max-Product values have been assessed on various scales to evaluate the robustness of the underlying model. The model's reliability has been further scrutinized through tests on different scales, including the Index of Agreement (IOA), Mean Absolute (MA) Error, and Root Mean Square (RMS) Error. These additional evaluations help in understanding how well the model performs across different scenarios and measurement criteria. The results of these tests are also illustrated in the figures below, which provide visual representations of the model's performance. Additionally, the membership functions for the model are categorized into low, moderate, and high levels. Figure 11 displays the rule set definition for the Air Quality Index, showing how different levels of air quality are interpreted and applied within the model. This comprehensive evaluation ensures that the model not only fits the historical data well but also maintains accuracy and reliability across various metrics and scales. [14]

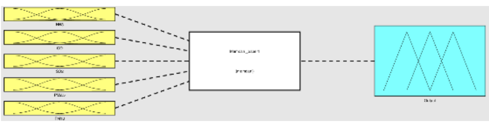

Fig 5: Fuzzy input/output combination

The fuzzy inputs used are:

- Ammonia (NH_3)
- Carbon monoxide (CO)
- Sulfur Dioxide (SO_2)
- Particulate Matter 2.5 micrometers diameter ($PM_{2.5}$)
- Particulate Matter 10 micrometers diameter (PM_{10})

Three steps which are used in the implementation of fuzzy system as in the classical way are shown below: [11]

Linguistic variable and fuzzy set determination

The combination of fuzzy input/output can be seen in the Figure 5 where each of the input and the output is divided into the numerous of fuzzy sets and the same can be seen from Figure 6 to Figure 10. [15]

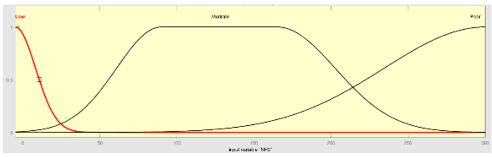

Fig 6: NH3 Membership Function

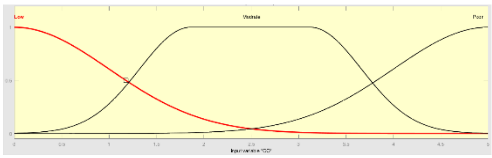

Fig 7: CO Membership Function

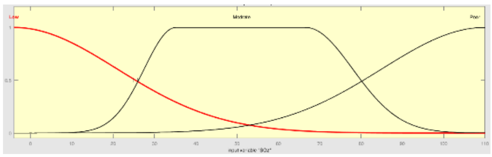

Fig 8: SO_2 Membership Function

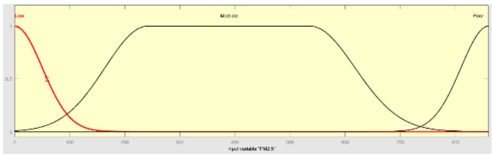

Fig 9: $PM_{2.5}$ Membership Function

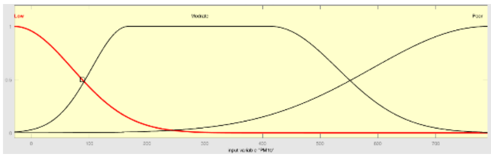

Fig 10: PM_{10} Membership Function

The output is divided into five fuzzy sets which describes the air quality index. The output fuzzy sets are as follow:

- Good
- Moderate
- Poor
- Very Poor
- Severe

The Gaussian set is used to divide the output consumption intervals and the same can be seen in Figure 11. The output of air quality index is divided from the range of 0 to 500 in Figure 11.

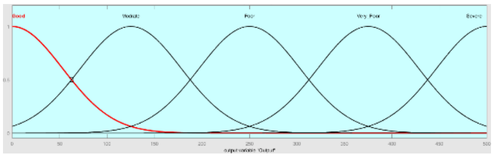

Fig 11: AQI Membership Functions

Constructing fuzzy rules

In this study, we have developed and used 50 fuzzy rules to predict the ambient AQI value and are shown in figure 12. [16] [17] The figure describes some of the rules which are used in the mamdani fuzzy system. These can be set with the "AND" or "OR" connection and can be seen in various languages as per the requirement of the user. One of the rule which has been implemented in the mamdani fuzzy inference system is as:

(NH_3==Low) & (CO==Moderate) & (SO_2==Low) & ($PM_{2.5}$==Moderate) & (PM_{10}==Moderate) => (Output=Poor) (1)

In the above rule we can see that if the concentration of Ammonia (NH_3) is low along with the composition of other concentrations as Carbon Monoxide (CO) is Moderate, Sulphur Dioxide SO2 if Low, Particulate Matter with the micrometer diameter 2.5 is Moderate and Particulate matter with the micrometer diameter 10 is Moderate then the Output is Poor with the weight of 1, same like this the total number of 50 rules were defined to get the output. [18]

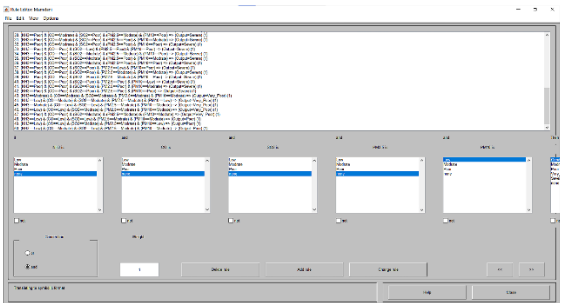

Fig 12: Mamdani Rules

Performing fuzzy inference into system

This is a process to articulate the mapping using fuzzy logic from the given input to output. The decision can be made on after providing basis. The process of fuzzy inference include three things which are membership functions, fuzzy logic operators and if-than rules.To compute the mapping form input values to output values this process is used along with this it contain fuzzification aggregation and defuzzification as a sub-processes. [19]

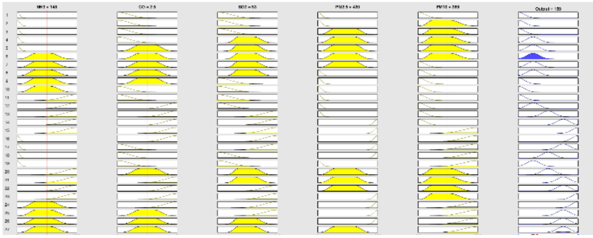

Fig 13: Rule Set Defining AQI

The "and" operator is used for the rules which are defined as per the air quality standards. If the pollutant values are high then the simulated forecast in the form of result is shown high and if it is low then it is set to low.

5. CONCLUSION AND FUTURE SCOPE

A Based on the results from the fuzzy inference system, there is a strong correlation between the target values and the output of the model simulations, as well as a favorable Index of Agreement (IOA). This indicates that the model performs well in aligning its predictions with the actual data. The model itself is a well-designed, straightforward mathematical regression model that utilizes the concept of membership functions and set theory. By integrating these elements, the model provides robust solutions for air quality prediction. It effectively translates the membership function concepts into practical outputs, demonstrating its reliability in evaluating and interpreting data. The Air Quality Index (AQI) is determined using a set of IF-THEN rules, which form the basis of the model's decision-making process. These rules are applied to assess and classify air quality levels. Despite the complexity of working with fuzzy logic and various input parameters, the application of powerful linguistic rules has made it possible to address and manage any complications that arise during the process. In summary, the fuzzy inference system shows a good match between target and predicted values, backed by a solid IOA. The use of a simple mathematical regression model, along with the membership function concept and IF-THEN rules, ensures that the model delivers accurate and effective results. The ability to overcome challenges using well-defined linguistic rules further enhances the model's efficacy in air quality assessment. [20] The noticeable weakness of this model can arise when the membership functions have a high degree of overlap. When there is significant overlap between membership functions, it can create uncertainty and lead to inaccuracies in the model's predictions. To address this issue, it is important to optimize the membership functions to minimize overlap and reduce extremes in error for future models. By refining the membership functions, the model can better distinguish between different fuzzy sets and improve the accuracy of its predictions.

References

[1] L. Zadeh, "Fuzzy logic = computing with words," IEEE Transactions on Fuzzy Systems, pp. 103-111, 1996.

[2] E. H. Mamdani, “Application of Fuzzy Logic to Approximate Reasoning Using Linguistic Synthesis,” IEEE Transactions on Computers, pp. 1182-1191, 1977.
[3] M. Kuanr, B. Kesari Rath and S. Nandan Mohanty, “Crop Recommender System for the Farmers using Mamdani Fuzzy Inference Model,” International Journal of Engineering & Technology, pp. 277-280, 2018.
[4] E. Pourjavad and A. Shahin, “The Application of Mamdani Fuzzy Inference System in Evaluating Green Supply Chain Management Performance,” International Journal of Fuzzy Systems, Springer, p. 901–912, 2018.
[5] K. Iqbal, M. A. Khan and A. Fatima, “Intelligent Transportation System (ITS) for Smart-Cities using Mamdani Fuzzy Inference System,” International Journal of Advanced Computer Science and Applications, pp. 94-105, 2018.
[6] Y. Ardiansyah, R. Sarno and O. Giandi, “Rain Detection System for Estimate Weather Level Using Mamdani Fuzzy Inference System,” in International Conference on Information and Communications Technology, 2018.
[7] Yeganeh, M. G. Hewson, S. Clifford, L. D. Knibbs and L. Morawska, “A satellite-based model for estimating PM2.5 concentration in a sparsely populated environment using soft computing techniques,” Environmental Modelling and Software, p. 84–92, 2017.
[8] S. H. Chen, A. J. Jakeman and J. P. Norton, “Artificial Intelligence techniques: An introduction to their use for modelling environmental systems,” Mathematics and Computers in Simulation, p. 379–400, 2008.
[9] W. Ding, J. Zhang and Y. Leung, “Prediction of air pollutant concentration based on sparse response back-propagation training feedforward neural networks,” Environmental Science and Pollution Research, vol. 23, no. 19, pp. 19481-19494, 2016.
[10] K. Gorai, P. Goyal and K. Kanchan, “A Review on Air Quality Indexing System,” Asian Journal of Atmospheric Environment, pp. 101-113, 2015.
[11] T. Ross, Fuzzy logic with engineering applications, New Mexico, USA: John Wiley, 2004.
[12] X. Zhu, Z. Huang, S. Yang and G. Shen, “Fuzzy Implication Methods in Fuzzy Logic,” IEEE Computer Society, p. 154–158, 2007.
[13] N. Güler Dincer and Ö. Akkuş, “A new fuzzy time series model based on robust clustering for forecasting of air pollution,” Ecological Informatics, vol. 43, pp. 157-164, 2018.
[14] Bougoudis, K. Demertzis, L. Iliadis, V. D. Anezakis and A. Papaleonidas, “FuSSFFra, a fuzzy semi-supervised forecasting framework: the case of the air pollution in Athens,” Neural Computing and Applications, vol. 29, no. 7, pp. 375-388, 2018.
[15] Z. Yuan, X. Zhou, T. Yang, J. Tamerius and R. Mantilla, “Predicting traffic accidents through heterogeneous urban data: A case study,” Proceedings of the 6th International Workshop on Urban Computing (UrbComp 2017), Halifax, NS, Canada., 2017.
[16] CPCB, 18 November 2009. [Online]. Available: https://cpcb.nic.in/displaypdf.php?id=aG9tZS9haXItcG9sbHV0aW9uL1JlY3ZlZC1OYXRpb25hbC5wZGY=. [Accessed May 2021].
[17] H.-F. Wang and R.-C. Tsaur, “Insight of a fuzzy regression model,” Fuzzy Sets and Systems, pp. 355-369, 2000.
[18] R. Mamlook, O. Badran and E. Abdulhadi, “A fuzzy inference model for short-term load forecasting,” Energy Policy, pp. 1239-1248, 2009.
[19] N. Gouveia and T. Fletcher, “Time series analysis of air pollution and mortality: effects by cause, age and socioeconomic status,” Journal of Epidemiology and Community Health, pp. 750-755, 2000.
[20] Rojas, O. Valenzuela, M. Anguita and A. Priet, “Analysis of the operators involved in the definition of the implication functions and in the fuzzy inference process,” International Journal of Approximate Reasoning, pp. 367-389, 1998.