Proceedings of Science and Mathematics
Faculty of Science, Universiti Teknologi Malaysia
Vol. 23, 2024, page 133-143

Fuzzy Inference System Model for Air Quality Index Prediction

Zulfazleen Natasha Zulkiflee, Amidora Idris
Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia
Corresponding author: amidora@utm.my

Abstract
Air pollution, a critical environmental issue, has serious consequences for human health, ecosystems, and the climate. This phenomenon is defined by the presence of harmful substances in the atmosphere, such as particulate matter (PM), nitrogen oxides (NO_x), sulphur dioxide (SO_2), carbon monoxide (CO), and ozone (O_3). In Malaysia, major sources of air pollution include industrial activities, vehicle emissions, agricultural practices, and transboundary haze caused by forest fires in neighboring countries. The health consequences range from respiratory and cardiovascular disease to premature death, with children and the elderly being especially vulnerable. Additionally, air pollution in Malaysia contributes to environmental degradation, as evidenced by acid rain, eutrophication, and biodiversity loss. Addressing air pollution in Malaysia requires broad strategies that include regulatory policies, technological advancements, and public awareness campaigns. Effective mitigation can result in better public health outcomes, higher environmental quality, and progress toward climate stability. Fuzzy inference system (FIS) is a widely used and effective application of fuzzy logic in various fields. This method is ideal for dealing with well-known environmental factors and making informed decisions under certainty. Therefore, this study proposes a fuzzy logic inference model to investigate a prediction model for the air quality index (AQI) in Pasir Gudang city based on pollutant concentrations (particulate matter 10 (PM_{10}), carbon monoxide (CO), and nitrogen dioxide (NO_2). There are five levels of air quality index (AQI) which are good, moderate, unhealthy, very unhealthy, and hazardous. This measurement is based on a classification by the Department of Environment (DOE) under the Ministry of Science, Technology, and Innovation. Results from actual data are compared to those from the proposed model. As a result, the air quality index prediction model's accuracy rate of 89.17% indicates that it meets the standard for forecasting AQI values.
Keywords: Air pollution; Air Quality Index; Fuzzy Inference System

1. Introduction

Air quality has become a major concern as cities and industry modernize. Deteriorating air quality, typified by increased pollutant concentrations, causes major health and environmental problems around the world, including Malaysia. Air pollution in Malaysia, like in many other places, is produced by both anthropogenic (man-made) and natural factors. Some common sources include the combustion of fossil fuels, vehicle emissions, deforestation, burn the biomass and garbage disposal. Air pollution has far-reaching consequences for ecosystems and wildlife in addition to humans. It contributes to climate change by raising the level of greenhouse gases in the atmosphere. Furthermore, it destroys ecosystems through acid rain, which affects soil and water bodies, as well as altering animal behavior and physiology, potentially upsetting entire food chains.

Given the serious health and environmental consequences of air pollution, accurately predicting the AQI is critical. The AQI is a numerical scale that indicates how filthy the air is currently or is expected to become. Researchers have used a variety of methodologies to forecast the AQI, allowing authorities to take more prompt action. Some method to predict air quality are by using a machine learning approach [1] and by using an effective hybrid deep learning model [2].

The most well-known application of fuzzy logic is fuzzy inference system (FIS), where membership functions are generally set manually through trial and error [3]. The FIS technique has an expressive output strength that allows users to easily understand and change the results [4]. The main parts in FIS include fuzzification, fuzzy decision, fuzzy rule-based, and defuzzification [5]. There are two types of FIS in terms of inference process which are Mamdani FIS [6] and Takagi-Sugeno-Kang (TSK) FIS [7]. The Mamdani technique is frequently preferred when transparency and ease of understanding are critical, because it allows for the use of normal language to define rules.

In this study, Mamdani FIS) are used to estimate the value of AQI. The application of genetic algorithms in the Mamdani FIS provides for a more streamlined and cost-effective approach to parameter optimization [8]. This means that fewer parameters must be optimized when compared to the Sugeno FIS, resulting in a more cost-effective modeling procedure. Mamdani FIS is a good alternative for forecasting AQI because of its capacity to include human expertise and language interpretability.

The study aims to understand the concept of FIS and AQI, to investigate a prediction model for the AQI in Pasir Gudang using a Mamdani FIS, to apply the triangular and trapezoidal membership functions in a fuzzy decision-making process, to analyse the data of air quality index based on concentrations of air pollutants such as particulate matter 10 (PM_{10}), carbon monoxide (CO), and nitrogen dioxide (NO_2).

2. Literature Review

2.1. Fuzzy Set

A fuzzy set in a universe of discourse is characterized by a membership function $\mu_A(x)$ that takes values in the interval [0,1]. That membership function of a classical set can only take two-values zero or one. Whereas the membership function of a fuzzy set is a continuous function with range [0,1]. There are various kinds of fuzzy numbers, but triangular and trapezoidal fuzzy numbers are the most widely used in real-world problems. Triangular fuzzy number is defined as $A = (a, b, c)$ where $a, b,$ and c are all real numbers, and its membership function is shown in Equation 1 below

$$\mu_A(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \le x \le b \\ \frac{c-x}{c-b}, & b \le x \le c \\ 0, & x > a \end{cases} \tag{1}$$

Triangular function in Equation 1 is defined by a lower limit a, an upper limit c, and a value b, where $a < b < c$. The triangular membership function is used to transform the linguistic values into a range of 0 - 1. The graph of the triangle membership function is shown in Figure 1.

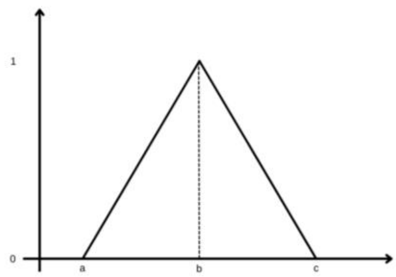

Figure 1 The Graph of the Triangle Membership Function

A fuzzy set $A = (a, b, c, d)$ is said to be trapezoidal fuzzy number function is given by where $a \le b \le c \le d$ and its membership function is shown in Equation 2 below

$$\mu_A(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a < x \leq b \\ 1, & b \leq x \leq c \\ \frac{c-x}{d-c}, & c < x \leq d \\ 0, & x \geq d \end{cases} \tag{2}$$

In Equation 2, the trapezoidal function is defined by a lower limit a, an upper limit d, a lower support limit b, and an upper support limit c [9]. The graph of the trapezoidal membership function is shown in Figure 2.

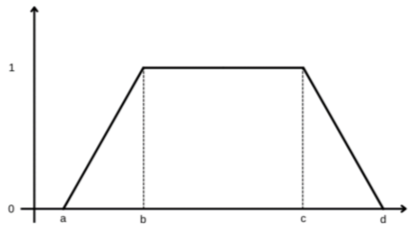

Figure 2 The Graph of the Trapezoidal Membership Function

2.2. Fuzzy Logic

Fuzzy logic, an extension of classical logic, presents a mathematical framework for dealing with imprecise or uncertain data. It is derived from fuzzy set theory and differs from traditional binary logic in that it allows for varying degrees of truth between the conventional true and false divisions. In 1965,the concept of infinite values has since evolved rapidly. Its foundation is the fuzzy set concept [10].

Nowadays, many researchers use fuzzy logic methods, and it has become increasingly popular, especially in fields that require dealing with uncertainty and imprecise information. For example, fuzzy logic can be apply in game AI for highlighting its simplicity and expressive power, essential for effective game development [11]. Other than that, the potential of fuzzy logic in finance also have been critically examine and emphasizing its underutilization in addressing banking crises [12]. Finally, fuzzy logic controller also can be used to adapt the changing weather condition in air conditioning systems [13].

2.3. Fuzzy Inference System (FIS)

The FIS incorporates an expert's knowledge and expertise in the design of a system that manages a process whose input-output relations are defined by a set of fuzzy control rules, such as IF-THEN rules [14].

The FIS is made up of four major parts. First, the process of fuzzification is the act of calculating the value of a crisp or inserting a membership degree value. The fuzzification process seeks to convert a definite value input into a fuzzy input. During the fuzzification process, the system converts the input value to linguistics and compares it to the rule base. The second is the fuzzy decision process, which is referred to as the selection of membership functions for input and output variables. Linguistic values can be expressed using the fuzzy set method. The membership functions of a fuzzy set are defined. To standardize the crisp for its simplicity and computing performance, triangular and trapezoidal membership functions are commonly used. The next step is the fuzzy rule base. In the fuzzy inference process, the rules that determine the membership functions of input and output are used. The first rules are linguistic in nature and are also known as "IF- THEN" rules.Finally, the defuzzification process, where the value of the output will be found in the value of crisp, that attempts to transform the results received from the rule base so that it becomes one degree of membership with the premises "crisp," is the final phase of this fuzzy [15].

There are a variety of types of FIS, such as adaptive network-based FIS, ellipsoidal FIS, Tsukamoto, generalized modulus ponens (GMP) FIS, and hierarchical FIS. However Mamdani and Sugeno are the most frequently used. Numerous researchers have conducted research on the FIS. For instance, the FIS is used to forecast hotel demand and determines best pricing methods for online hotel searching using a fuzzy inference system [16]. Other than that, Mamdani was one of the first fuzzy systems to regulate a steam engine and boiler combination using a set of fuzzy rules supplied by experienced human operators [17]. This method has been applied successfully in a range of industrial problems. In the previous study, many researchers have been used Mamdani FIS. Additionally, the Mamdani FIS was discovered to be effective and responsive in delivering information about the level of rainy weather in this rain detecting system [15].

2.4. Air Quality Index (AQI)

The AQI is a numerical scale that conveys the quality of air in a specific region, serves as a key instrument for public health and environmental awareness. It alerts society about how clean or filthy the air is and what health impacts may be a worry for the general population. The AQI is extensively used by environmental agencies and health organizations to communicate to the public the possible concerns connected with air pollution. Particulate matter (PM_{10} and $PM_{2.5}$), ground-level ozone (O_3), sulphur dioxide (SO_2), nitrogen dioxide (NO_2), and carbon monoxide (CO) concentrations are used to calculate the AQI. Each of these pollutants has a distinct health impact, and their concentrations are monitored in relation to set air quality guidelines.

The AQI readings are categorized as good, moderate, unhealthy, very unhealthy, and hazardous. These categories help to explain the amount of health concern associated with certain pollutant concentration levels. Each signifying a distinct level of health concern as shown in Table 1.

Table 1 The Air Quality Index Level

Indicator	AQI value	Explanation
Good	0 – 50	Air quality is considered satisfactory, and air pollution poses little or no risk.
Moderate	50 - 100	Air quality is acceptable; however, for some pollutants there may be a moderate health concern for a very small number of people.
Unhealthy	100 - 150	Everyone may begin to experience health effect.
Very Unhealthy	150 - 200	Health warnings of emergency conditions.
Hazardous	> 200	Health alert: everyone may experience more serious health effects.

In this study only three variables of air quality index (AQI) have been used which is particulate matter 10 (PM_{10}), carbon monoxide (CO) and nitrogen dioxide (NO_2). This study focuses on CO, PM_{10}, NO_2 because they are important components of the AQI, which measures major health impacts, environmental significance and regulatory importance. Particulate matter also known as PM where Some particles are large or dark enough to be seen with the human eye, such as dust, dirt, soot, or smoke. Others are so minuscule that they can only be seen with an electron microscope. Aside from that, nitrogen dioxide (NO_2) is one of a class of extremely reactive gases known as nitrogen oxides (NO_x). The primary source of NO_2 in the atmosphere is the combustion of fuel. NO_2 is produced by emissions from cars, lorries, and buses, as well as power plants and off-road equipment. Last but not least, carbon monoxide (CO) is a toxic, colourless, and odorless gas produced by incomplete hydrocarbon combustion of fossil fuels, leads to headaches, nausea, and dizziness, as well as increasing the risk of heart disease, and high quantities can be fatal.

Other than the Mamdani FIS, there have been several studies on forecasting air pollution in the previous year. By using an inverse distance weighted (IDW) geostatistical approach analysis, the AQI can be predicted [18]. The other method are by using unique optimal-combined model for AQI forecasting based on complementary ensemble empirical mode decomposition (CEEMD), particle swarm optimization and gravitational search algorithm (PSOGSA), and particle swarm optimization (PSO) [19].

3. Methodology

3.1. The Fuzzy Inference System Model

The method of fuzzy inference system used to predict the AQI consists of 5 steps as stated below:

Step 1 : Data Collection

Data of air quality index in Pasir Gudang, Johor was collected from 1^{st} February 2022 until 31^{st} May 2022 and used in the FIS process by using MATLAB R2023b.

Step 2 : Fuzzification Process (inputs and outputs variables)

As observed in Figure 3, the design of the fuzzy based air quality index prediction model is designed using the Fuzzy Toolbox and the Mamdani FIS, integrated within the MATLAB software environment. It illustrates how air pollutants affect the AQI.

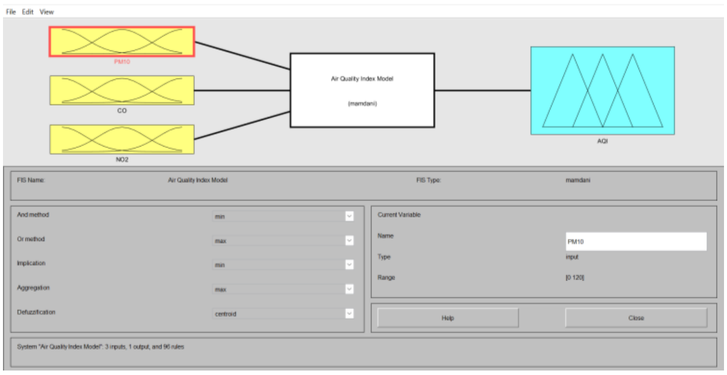

Figure 3 Design of The Air Quality Index Prediction Model for PM_{10} , CO, and NO_2 vs AQI.

Step 3: Fuzzy Decision (selection of membership functions for Inputs and output variables)

In this study, the triangular membership function, known as trimf in MATLAB is applied in the design of the proposed fuzzy based prediction model. The triangular membership function is computationally efficient and is used to normalize crisp inputs.

The Mamdani FIS is suitable for designing AQI predictions model as both the inputs and output of the FIS are represented by the values of linguistic variables. In order to transform crisp input values into fuzzy values, the membership function of each input is determined. Table 2 illustrates the linguistic variables and range for PM_{10} , CO and NO_2 and AQI.

Table 2: The Linguistic Variables and Range for Inputs and Output

Functions	**Criterion**	**Linguistic Variable**	**Range**
Inputs	Particulate Matter 10 (PM_{10})	Excellent	< 35
		Good	35 – 52.5
		Moderate	52.5 – 70
		Poor	70 – 87.5
		Very Poor	87.5 – 105
		Severe	> 105
	Carbon Monoxide (CO)	Excellent	< 10
		Good	10 – 15
		Moderate	15 – 20
		Poor	> 20
	Nitrogen Dioxide (NO_2)	Excellent	< 25
		Good	25 – 37.5
		Moderate	37.5 – 50
		Poor	> 50
Outputs	Air Quality Index (AQI)	Good	< 100
		Moderate	100 – 150
		Unhealthy	150 – 200
		Very Unhealthy	200 – 250
		Hazardous	> 300

The membership function editor for inputs and output are illustrated in Figure 4 until Figure 7. The membership function input category was obtained by setting up the range of each input as shown in Table 2.

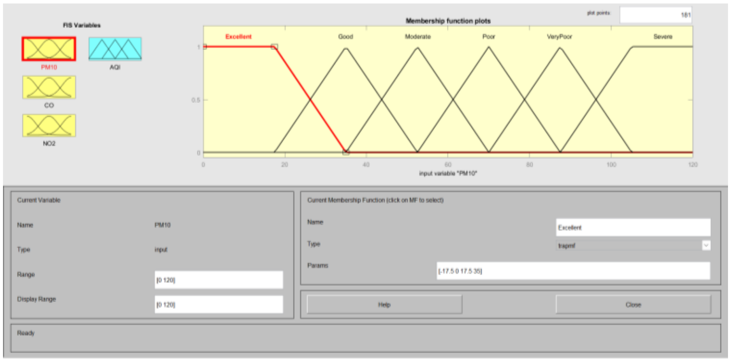

Figure 4 The Membership Function of Particulate Matter 10 (PM_{10}).

Figure 5 The Membership Function of Carbon Monoxide (CO)

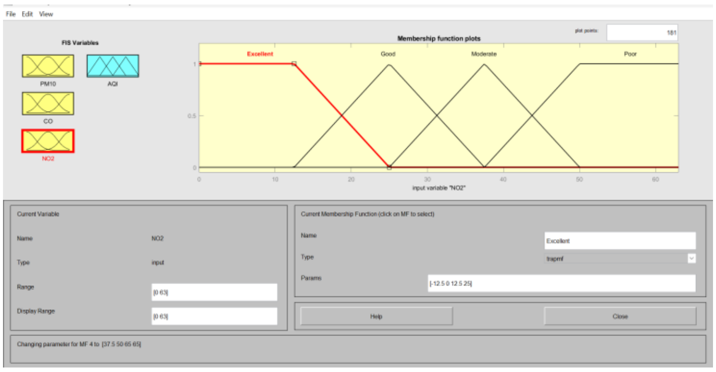

Figure 6 The Membership Function of Nitrogen Dioxide (NO_2).

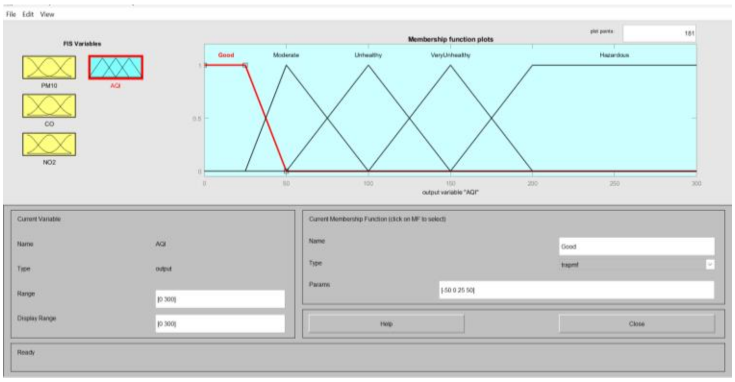

Figure 7 The Membership Function of Air Quality Index (AQI).

Step 4 : Fuzzy Rule Base (determination of rule base application)

The rules decided the membership functions of input and output are utilized in fuzzy inference process. The first rules are linguistic and also known as “IF-THEN” regulations. The rules below show an example of IF-THEN rules of AQI generated from this process.

IF (PM_{10} is Poor) and (CO is Moderate) and (NO_2 is Excellent) THEN (AQI is Unhealthy)
IF (PM_{10} is Excellent) and (CO is Excellent) and ($NO_2$is Excellent) THEN (AQI is Good)
IF (PM_{10} is Good) and (CO is Moderate) and ($NO_2$is Moderate) THEN (AQI is Moderate)

Figure 8 shows the rule editor that is set up in the MATLAB software. There are 96 rules used to evaluate the AQI by using Mamdani IF-THEN rules to determine whether it is good, moderate, sensitive, unhealthy, very unhealthy, and hazardous.

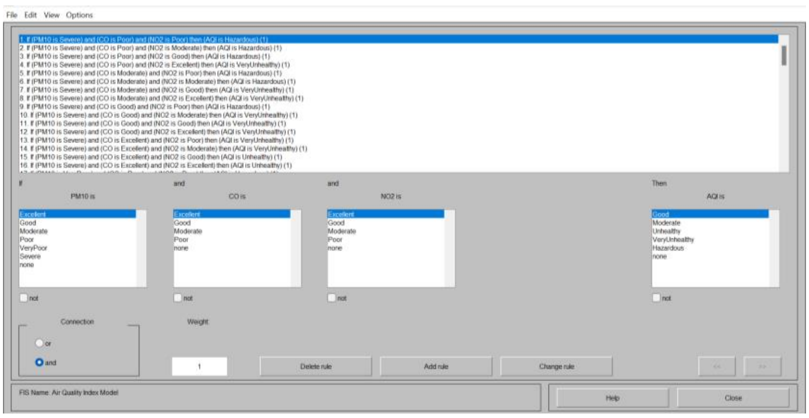

Figure 8 Rule Editor

Step 5 : Defuzzification Process

The defuzzification process transforms the fuzzy value into a crisp value. After the rules are set, the data has been inserted in the rule viewer and the result is generated as shown in the right-end column of Figure 9. The rule viewer is used in a fuzzy inference diagram. This diagnostic tool identifies active rules and the impact of each membership functions on results.

Figure 9 illustrates how air pollutants PM_{10}, CO, and NO_2 concentration levels significantly impact the Pasir Gudang Air Quality Index from a rule perspective. For example, if PM_{10} , CO, and NO_2 concentrations are 21 $\mu g/m^3$, 3 ppm, and 14 ppb, the predicted AQI is 47.

Figure 9 Rule Viewer (Rules 1-96)

Figure 10 shows how the surface viewer generates and plots an output surface map for the model, displaying the dependency of one or more outputs on any one or two inputs. In this case, the surface view shows how the output AQI varies with the concentrations of the air pollutants PM_{10}, CO, and NO_2. Any changes in air pollutant concentrations affects the AQI. The plot illustrates the relationship between air pollutants and the AQI. Any increase in air pollutant input concentrations leads to an increase in the AQI, regardless of other air pollutant values.

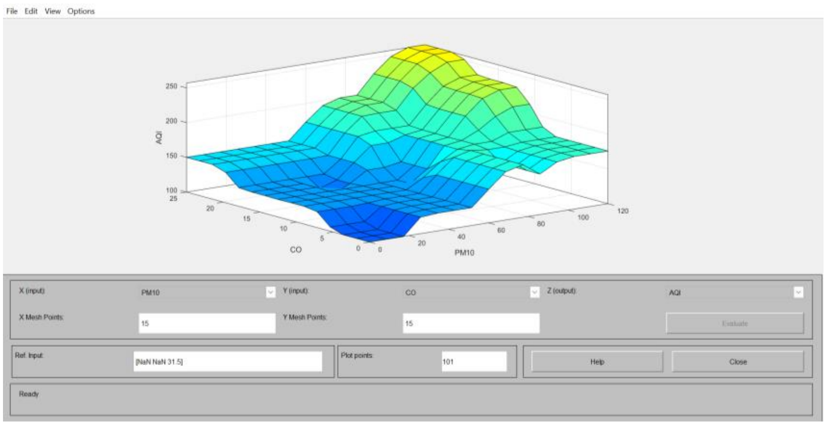

Figure 10 Surface Viewer

4. Results and Discussion

4.1. The Comparison between Actual Data and Mamdani AQI Model

The results of actual data output are shown in Table 3. The value is obtained based on the 96 IF-Then rules that are mentioned in defuzzification process. For example, IF (PM_{10} is Excellent) and (CO is

Excellent) and(NO_2 is Excellent) THEN (AQI is Good) with the input value PM_{10} is 22 $\mu g/m^3$, CO is 2 ppm and NO_2 is 16 ppb.

Table 3 :Result output of AQI from MATLAB

Inputs			**MATLAB (Mamdani Status)**	
PM_{10}	CO	NO_2	Output AQI	Linguistic Value
22	2	16	56	Moderate
25	3	14	51	Moderate
10	2	10	38	Good
25	4	17	62	Moderate
45	2	10	72	Moderate

Table 4 shows the comparison statements of the results from actual data output and AQI prediction from MATLAB. The statements is “TRUE” when the linguistic variable for actual data is “Good” and AQI Prediction from MATLAB is also “Good”. However, if they both are with different linguistic variable then the statement is “FALSE”. The comparison statements in Table 4 are part of 120 statements overall.

Table 4:The Result of Comparison Statements between The Actual Data Output and Air Quality Index Model

Date	**Actual Data**		**AQI Prediction (MATLAB)**		**Statements**
	AQI	Linguistic Variable	AQI	Linguistic Variable	
27032022	55	Moderate	45	Good	FALSE
28032022	58	Moderate	55	Moderate	TRUE
29032022	47	Good	42	Good	TRUE
30032022	50	Good	59	Good	TRUE
31032022	53	Moderate	57	Moderate	TRUE
01042022	63	Moderate	49	Good	FALSE

Figure 11 illustrates the results of the overall comparison statements between the actual air quality index data and the AQI predictions made by the model over a period of 120 days. The comparison reveals that out of these 120 days, the AQI model made correct predictions (TRUE statements) on 107 days and incorrect predictions (FALSE statements) on 13 days.

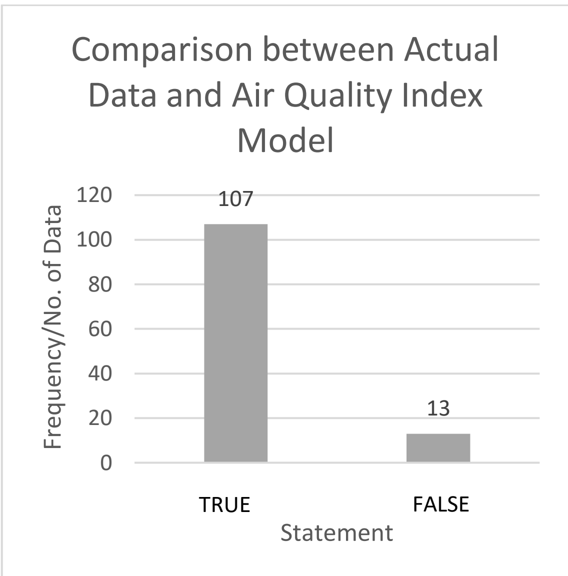

Figure 11 The Graph of Comparison between Actual Data and AQI model.

The level of accuracy rate is calculated as follows:

$$\text{Accuracy Rate } = \frac{107}{120} \times 100\%$$

$$= 89.1667\%$$

This calculation shows that the AQI model has an accuracy rate of approximately 89.17%, indicating that the model's predictions match the actual data of AQI 89.17% of the time.

Conclusion

In this paper, the air quality index prediction model based on the fuzzy logic inference system was investigated to predict the AQI in Pasir Gudang city, according to the air pollutant data concentrations. It was observed that fuzzy logic algorithms are capable in determining the AQI and therefore, can be used to predict and estimate the air quality index in real time, based on the given air pollutant concentrations. With the accuracy rate 89.17%, indicates that it meets the standard for forecasting the AQI values. Hence, this can reduce the effects of air pollution on both human and the environment.
.

Acknowledgement

I wish to express my sincere gratitude to all who have contributed throughout the course of this work. The completion of this work would not have been feasible without their participation and assistance

References

[1] Liu, H., Li, Q., Yu, D., & Gu, Y. (2019). Air quality index and air pollutant concentration prediction based on machine learning algorithms. *Applied Sciences (Switzerland)*, *9*(19). https://doi.org/10.3390/app9194069

[2] Sarkar, N., Gupta, R., Keserwani, P. K., & Govil, M. C. (2022). Air Quality Index prediction using an effective hybrid deep learning model. *Environmental Pollution*, *315*. https://doi.org/10.1016/j.envpol.2022.120404

[3] Samira, K., & Ahmad, J. (2016). A New Artificial Intelligence Method for Prediction of Diabetes Type2. In *Bulletin de la Société Royale des Sciences de Liège* (Vol. 85).

[4] Swati R, Chaudhari, & Manoj E. Patil. (2015). *Comparative Analysis of Fuzzy Inference System for Air Conditioner*.

[5] Kumaravel R, & Vallinayagam V. (n.d.). 484 A FUZZY INFERENCE SYSTEM FOR AIR QUALITY IN USING MATLAB, CHENNAI, INDIA. In *J. Environ. Res. Develop. Journal of Environmental Research And Development* (Vol. 7, Issue 1).

[6] Pourjavad, E., & Shahin, A. (2018). The Application of Mamdani Fuzzy Inference System in Evaluating Green Supply Chain Management Performance. *International Journal of Fuzzy Systems*, *20*(3), 901–912. https://doi.org/10.1007/s40815-017-0378-y

[7] Cavallaro, F. (2015). A Takagi-Sugeno fuzzy inference system for developing a sustainability index of biomass. *Sustainability (Switzerland)*, *7*(9), 12359–12371. https://doi.org/10.3390/su70912359

[8] Özger, M. (2009). Comparison of fuzzy inference systems for streamflow prediction. *Hydrological Sciences Journal*, *54*(2), 261–273. https://doi.org/10.1623/hysj.54.2.261

[9] Dison Ebinesar, M., & Dison, M. E. (2015). Different Types of Fuzzy Numbers and Certain Properties. *Journal of Computer and Mathematical Sciences |www.Compmath-Journal.Org*, *6*(11), 631–651. www.compmath-journal.org

[10] Zadeh Lotfi A. (1965). *Fuzzy sets. Inform and Control* (Vol. 8).

[11] Pirovano, M. (2012). *The use of Fuzzy Logic for Artificial Intelligence in Games*. http://www.maxis.com/

[12] Sanchez-Roger, M., Oliver-Alfonso, M. D., & Sanchís-Pedregosa, C. (2019). Fuzzy logic and its uses in finance: A systematic review exploring its potential to deal with banking crises. In *Mathematics* (Vol. 7, Issue 11). MDPI AG. https://doi.org/10.3390/math7111091

[13] Riyadh Waheed, S., Mundher Adnan, M., Mohd Suaib, N., & Shafry Mohd Rahim, M. (2020). Fuzzy Logic Controller for Classroom Air Conditioner. *Journal of Physics: Conference Series*, *1484*(1). https://doi.org/10.1088/1742-6596/1484/1/012018

[14] Francesco Camastra, Angelo Ciaramella, Valeria Giovannelli, Matteo Lener, Valentina Rastelli, Antonino Staiano, Giovanni Staiano, & Alfredo Starace. (2014). *A fuzzy decision system for genetically modified plant environmental risk assessment using Mamdani inference*.

[15] Ahmad Yusuf Ardiansyah, & Riyanarto Sarno. (2018). *Rain Detection System for Estimate Weather Level Using Mamdani Fuzzy Inference System*.

[16] Ahmadi, S. A., & Ghasemi, P. (2023). Pricing strategies for online hotel searching: a fuzzy inference system procedure. *Kybernetes*, *52*(11), 4913–4936. https://doi.org/10.1108/K-03-2022-0427

[17] Kothamasu, R., & Huang, S. H. (2007). Adaptive Mamdani fuzzy model for condition-based maintenance. *Fuzzy Sets and Systems*, *158*(24), 2715–2733. https://doi.org/10.1016/j.fss.2007.07.004

[18] Jumaah, H. J., Ameen, M. H., Kalantar, B., Rizeei, H. M., & Jumaah, S. J. (2019). Air quality index prediction using IDW geostatistical technique and OLS-based GIS technique in Kuala Lumpur, Malaysia. *Geomatics, Natural Hazards and Risk*, *10*(1), 2185–2199. https://doi.org/10.1080/19475705.2019.1683084

[19] Zhu, S., Yang, L., Wang, W., Liu, X., Lu, M., & Shen, X. (2018). Optimal-combined model for air quality index forecasting: 5 cities in North China. *Environmental Pollution*, *243*, 842–850. https://doi.org/10.1016/j.envpol.2018.09.025