Application of Fuzzy Inference System in the Prediction of Air Quality Index

Nor Hayati Shafii[1]*, Nur Aini Mohd Ramle[2], Rohana Alias[3], Diana Sirmayunie Md Nasir[4], Nur Fatihah Fauzi[5]

[1,2,3,4,5] *Faculty of Computer and Mathematical Sciences,*
Universiti Technologi MARA, Perlis Branch, Arau Campus, 02600 Arau, Perlis, Malaysia

Corresponding author: * norhayatishafii@uitm.edu.my

HIGHLIGHTS

- Fuzzy Inference System (FIS) is used to develop a model in predicting Air Quality Index (AQI).
- A model is developed to evaluate the AQI of air pollution by considering the value of three major inputs, namely Particulate Matter (PM10), Carbon Dioxide (CO) and Nitrogen Dioxide (NO2).
- The evaluation system is based on Mamdani fuzzy inference system through four major steps namely Fuzzification, Fuzzy decision, Fuzzy rule based and Defuzzification.
- The level of accuracy rate is calculated to measure the efficiency of the proposed model.

ABSTRACT

Air pollution is the presence of substances in the atmosphere that are harmful to the health of humans and other living beings. It is caused by solid and liquid particles and certain gases that are suspended in the air. The air pollution index (API) or also known as air quality index (AQI) is an indicator for the air quality status at any area. It is commonly used to report the level of severity of air pollution to public and to identify the poor air quality zone. The AQI value is calculated based on average concentration of air pollutants such as Particulate Matter 10 (PM10), Ozone (O3), Carbon Dioxide (CO2), Sulfur Dioxide (SO2) and Nitrogen Dioxide (NO2). Predicting the value of AQI accurately is crucial to minimize the impact of air pollution on environment and human health. The work presented here proposes a model to predict the AQI value using fuzzy inference system (FIS). FIS is the most well-known application of fuzzy logic and has been successfully applied in many fields. This method is proposed as the perfect technique for dealing with environmental well known and tackling the choice made below uncertainty. There are five levels or indicators of AQI, namely good, moderate, unhealthy, very unhealthy, and hazardous. This measurement is based on classification made from the Department of Environment (DOE) under the Ministry of Science, Technology, and Innovation (MOSTI). The results obtained from the actual data are compared with the results from the proposed model. With the accuracy rate of 93%, it shows that the proposed model is meeting the highest standard of accuracy in forecasting the AQI value.

Keywords: air pollution, Air Quality Index, Fuzzy Inference System, accuracy

INTRODUCTION

In the era of globalization and the age of industrialization 4.0, environmental pollution is often one of the topics discussed in the media. Most countries in the world, including Malaysia, have the same problem. Many regions in Malaysia are facing extreme air quality issues. In Malaysia, state, local and federal associations are involved in identifying where air pollution occurs, causes of air pollution, and how to control air pollution. Air pollution is caused by many reasons such as vehicle exhaust fumes, fossil fuel-based power plants, exhaust from industrial factories and plants**,** agricultural and construction activities etc. It may cause diseases, allergies and even death to humans. It may also cause harm to other living organisms such as animals and food crops and may damage the natural environment. Observing AQI accurately is essential to detect pollution peaks and make early preventive measures that can minimize the impact of air pollution on environment and human health.

Previous studies have used various approaches to predict the air quality index (AQI). In this study, we focus on fuzzy inference system (FIS) to predict the AQI value. According to Chaudhari and Patil (2014), FIS approach has an expressive output strength without difficulty to understand the results and manipulate the goal. FIS can capture converting environment as a professional information and without problems. The very first application of this technique was to control a steam engine and the result obtained was as good as when controlled by experienced human operators. According to Sowlat et al. (2011) FIS has been effectively used in many fields including programmed control, data arrangement, decision analysis, skillful systems, and computer vision. For this reason, fuzzy inference systems are referred to by various names, such as fuzzy expert systems, fuzzy rule-based systems, fuzzy skilled systems, fuzzy modeling, fuzzy associative memory, fuzzy logic controllers and ambiguously fuzzy systems etc. FIS are broadly applicable in finance, scientific, economic, and engineering etc. due to the intuitive nature of the system and its ability to investigate decisions of a human. Samira and Ahmad (2016) found that FIS is the most well-known application of fuzzy logic where membership functions should be generally adjusted manually through trial and error. According to Cavallaro (2015), fuzzy logic reasoning consists of two kinds of information. The first concerns with the labels and membership functions allocated to the input and output variables. The correct desire of those symbolizes is one of the greatest serious degrees inside the design model. Another kind of information is associated with rule base which is a technique of converting the fuzzy values of the inputs to fuzzy values of the outputs. Kumaravel (2012) concluded that the important elements in FIS are fuzzification, fuzzy decision, fuzzy rule-based and defuzzification.

According to Tiwari (2015) FIS has assumed that each rule–is activated at each cycle and contributes collectively to the solution. It is a parallel single inference. However, the inference process can be preserved as the new inferred results that can be fed once more as inputs. Methods based on the fuzzy set theory must be carried out within the context of environmental numbers. The limits among the right and an unacceptable concentration is not to be considered as sharp, but as fuzzy, with implications for subsequent movement plans. There the usage of fuzzy numbers is proposed as the perfect technique for dealing with environmental well known and tackling the choice made below uncertainty.

METHODOLOGY

Method of Data Collection

The hourly data of AQI used in this study are secondary data obtained from Department of Environment under the Ministry of Science, Technology, and Innovation (MOSTI) for Kelantan state. The are 122 data sets collected from 1st June to 31st July 2016. It was recorded from Continuous Air Quality Monitoring Stations (CAQM) located in both industrial and urban areas in Kelantan.

Rating scales for Air Quality Index (AQI)

The rating scale of the AQI status indicator consists of Good, Moderate, Unhealthy, Very Unhealthy, and Hazardous as shown in Table 1. This measurement is based on classification made from the Department of Environment (DOE) under the Ministry of Science, Technology, and Innovation (MOSTI).

Table 1: Indicators for output, AQI

API	Indicator	Explanation
3 - 5.75	Good	Low pollution without any bad effect on health
5.76 - 8.5	Moderate	Moderate pollution that does not pose any bad effect on health
8.51 - 11.25	Unhealthy	Worsen the health condition and low tolerance of physical exercises to people with heart and lung complications
11.25-14.0	Very Unhealthy	Worsen the health condition and low tolerance of physical exercises to people with heart and lung complications. Affect public health
>14.0	Hazardous	Hazardous to high-risk people and public health

There are five major pollutants that contribute to air pollution, namely Sulfur Dioxide (SO_2), Nitrogen Dioxide (NO_2), Carbon Dioxide (CO), Ozone (O_3) and Particulate Matter (PM10). The most significant input variables used in this study are PM10, CO and NO_2.

Data Analysis

Data analysis is done by using MATLAB. It is a completely effective software that has numerous built-in tools for solving problems and for graphical illustrations. The method of data analysis used to predict the AQI using FIS consists of four steps; 1) Fuzzification of input and output variables 2) Fuzzy decision (selection of membership functions for input and output variables), 3) Fuzzy rule-based (determination of rule base application) and 4) Defuzzification. The structure of a fuzzy system is shown in Figure 1.
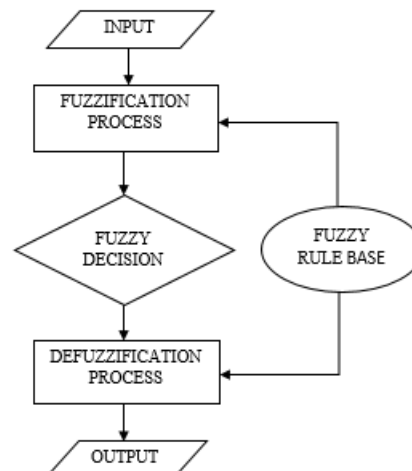

Figure 1: The Process of Fuzzy Inference System

I) Fuzzification of Input and Output Variables

Fuzzification is a process of transforming or changing the real crisp value into the fuzzy value. The operation interprets real crisp input or else measured values into linguistic ideas using appropriate membership functions. Fuzzy logic implements human experiences and preferences via membership functions and fuzzy rules. In fuzzy logic, each linguistic variable is associated with confidence values such that each term has its own confidence value. Appropriate fuzzy linguistic values are allocated for each fuzzy variable. The input and output are converted in the form of linguistic layout. As a result, the input value is used to generate output in the rule viewer as given in Equation (1).

$$\mu_c(y) = \max_k [\min[\mu_A(input(i)), \mu_B(input(j))]], k = 1, 2, 3, ..., r \tag{1}$$

Figure 2 shows the fuzzy inference system that consists of three inputs (PM10, CO and NO2) and one output which is air quality index (AQI). By using FIS editor and membership function editor from MATLAB software all the membership functions are developed.
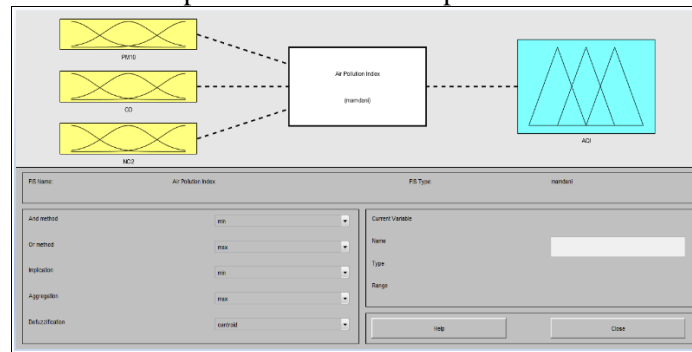

Figure 2: FIS Editor

II) Fuzzy Decision - Selection of Membership Functions for Input and Output Variables

Linguistic values are expressed inside the method of fuzzy sets. A fuzzy set is defined by its membership functions. Generally, triangular and trapezoidal membership functions are utilized to standardize the crisp for its simplicity and computational performance. It is defined mathematically in the following equations;

Triangular function in Equation 2 is defined by a lower limit a, an upper limit b, and a value m, where a < m < b. The Triangular membership function is used to transform the linguistic values into a range of 0 - 1.

$$\mu_A(x) = \begin{cases} 0, x \leq a \\ \dfrac{x-a}{m-a}, a \leq x \leq m \\ \dfrac{b-x}{b-m}, m \leq x \leq b \\ 0, x \geq b \end{cases} \tag{2}$$
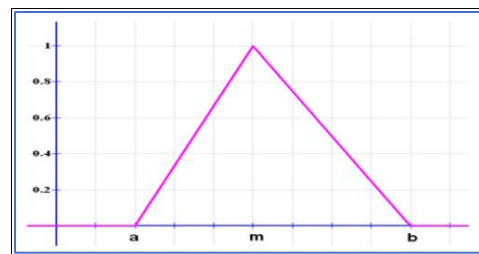

Figure 3: Triangular Membership Function

Trapezoidal function in Equation 3 is defined by a lower limit a, an upper limit d, a lower support limit b, and an upper support limit c, where a < b < c < d.

$$\mu_A(x) = \begin{cases} 0, x < a \\ \dfrac{x-a}{b-a}, a \leq x \leq b \\ 1, b \leq x \leq c \end{cases} \qquad (3)$$
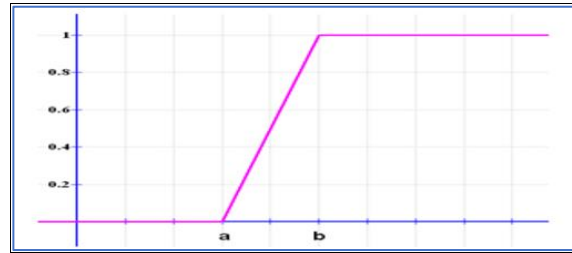

Figure 4: Trapezoidal membership function

Table 2 below shows the fuzzy set input indicators and the fuzzy set of output indicators. Table 3 shows the range of each input and Table 4 shows the parameters used in the membership function editor for each input.

Table 2: Indicators for output and input

Inputs (Fuzzy Variables)			Output
PM10	CO	NO2	AQI
Excellent	Excellent	Excellent	Good
Good	Good	Good	Moderate
Moderate	Moderate	Moderate	Unhealthy
Poor	Poor	Poor	Very Unhealthy
Very Poor			Hazardous
Severe			

Table 3: Range for each input

Inputs			Indicators
PM10	CO	NO2	
< 35	< 10	< 25	Excellent
35 - 52.5	10 - 15	25 - 37.5	Good
52.5 -70	15 - 20	37.5 - 50	Moderate
70 - 87.5	> 20	> 50	Poor
87.5 - 105			Very Poor
> 105			Severe

Table 4: Parameters for membership function of the inputs

Parameters			Indicators
PM10	CO	NO2	
[-17.5 0 17.5 35]	[-5 0 5 10]	[-12.5 0 12.5 25]	Excellent
[17.5 35 52.5]	[5 10 15]	[12.5 25 37.5]	Good
[35 52.5 70]	[10 15 20]	[25 37.5 50]	Moderate
[52.5 70 87.5]	[15 20 25 25]	[37.5 50 62.5 62.5]	Poor
[70 87.5 105]			Very Poor
[87.5 105 122.5 122.5]			Severe

Figures 5, 6 and 7 show the membership function editor for three inputs PM10, NO2, CO and Figure 8 shows the membership function editor for output AQI. The membership function input category was obtained by setting up the range of each input variable as shown in Table 4. This is then, followed by setting up the parameter for each input as shown in Table 5
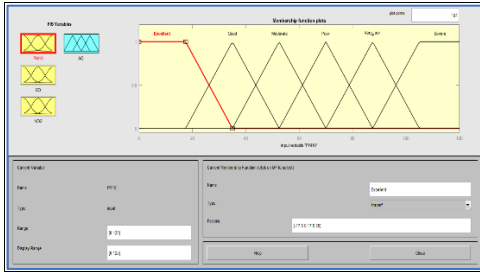
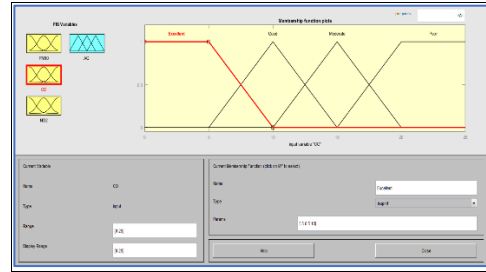**Figure 5:** Membership function editor of PM10 (input 1)

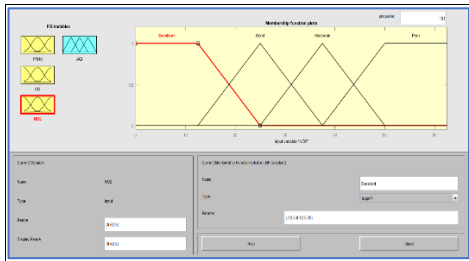**Figure 6:** Membership function editor of CO (input 2)

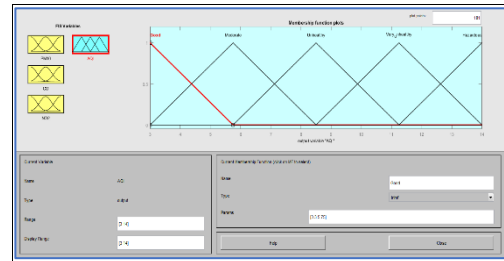**Figure 7:** Membership function editor of NO2 (input 3)

Figure 8: Membership function editor of AQI (output)

The membership functions of input 1, 2 and 3 and also output are shown in details in Table 5. The functions are formulated based on equation (2) and (3).

Table 5: Membership function of input 1, 2, 3 and output

Input 1: PM10	Input 2: CO	Input 3: NO2	Output: AQI
$\mu_{Excellent}(x) = \begin{cases} 1, 0 \le x \le 17.5 \\ \dfrac{35-x}{17.5}, 17.5 \le x \le 35 \\ 0, otherwise \end{cases}$	$\mu_{Good}(x) = \begin{cases} \dfrac{x-5}{5}, 5 \le x \le 10 \\ \dfrac{15-x}{5}, 10 \le x \le 15 \\ 0, otherwise \end{cases}$	$\mu_{Good}(x) = \begin{cases} \dfrac{x-12.5}{12.5}, 12.5 \le x \le 25 \\ \dfrac{37.5-x}{12.5}, 25 \le x \le 37.5 \\ 0, otherwise \end{cases}$	$\mu_{Moderate}(x) = \begin{cases} \dfrac{x-3}{2.75}, 3 \le x \le 5.75 \\ \dfrac{8.5-x}{2.75}, 5.75 \le x \le 8.5 \\ 0, otherwise \end{cases}$
$\mu_{Good}(x) = \begin{cases} \dfrac{x-17.5}{17.5}, 17.5 \le x \le 35 \\ \dfrac{52.5-x}{17.5}, 35 \le x \le 52.5 \\ 0, otherwise \end{cases}$	$\mu_{Moderate}(x) = \begin{cases} \dfrac{x-10}{5}, 10 \le x \le 15 \\ \dfrac{20-x}{5}, 15 \le x \le 20 \\ 0, otherwise \end{cases}$	$\mu_{Moderate}(x) = \begin{cases} \dfrac{x-25}{12.5}, 25 \le x \le 37.5 \\ \dfrac{50-x}{12.5}, 37.5 \le x \le 50 \\ 0, otherwise \end{cases}$	$\mu_{Unhealthy}(x) = \begin{cases} \dfrac{x-5.75}{2.75}, 5.75 \le x \le 8.5 \\ \dfrac{11.25-x}{2.75}, 8.5 \le x \le 11.25 \\ 0, otherwise \end{cases}$
$\mu_{Moderate}(x) = \begin{cases} \dfrac{x-35}{17.5}, 35 \le x \le 52.5 \\ \dfrac{70-x}{17.5}, 52.5 \le x \le 70 \\ 0, otherwise \end{cases}$	$\mu_{Poor}(x) = \begin{cases} 0, x \le 15 \\ \dfrac{x-15}{5}, 15 \le x \le 20 \\ 0, otherwise \end{cases}$	$\mu_{Poor}(x) = \begin{cases} 0, x \le 37.5 \\ \dfrac{x-37.5}{12.5}, 37.5 \le x \le 50 \\ 1, 50 \le x \le 62.5 \\ 0, otherwise \end{cases}$	$\mu_{VeryUnhealthy}(x) = \begin{cases} \dfrac{x-8.5}{2.75}, 8.5 \le x \le 11.25 \\ \dfrac{14-x}{2.75}, 11.25 \le x \le 14 \\ 0, otherwise \end{cases}$
$\mu_{Poor}(x) = \begin{cases} \dfrac{x-52.5}{17.5}, 52.5 \le x \le 70 \\ \dfrac{87.5-x}{17.5}, 70 \le x \le 87.5 \\ 0, otherwise \end{cases}$			$\mu_{Hazardous} = \begin{cases} \dfrac{x-11.25}{2.75}, 11.25 \le x \le 14 \\ 0, otherwise \end{cases}$
$\mu_{VeryPoor}(x) = \begin{cases} \dfrac{x-70}{17.5}, 70 \le x \le 87.5 \\ \dfrac{105-x}{17.5}, 87.5 \le x \le 105 \\ 0, otherwise \end{cases}$			

$\mu_{Severe}(x) = \begin{cases} 0, x \le 87.5 \\ \dfrac{x-87.5}{17.5}, 87.5 \le x \le 105 \\ 1, 105 \le x \le 122.5 \end{cases}$			

The range of output is between 3 to 14. Table 6 shows the range of output and Table 7 shows the parameters that are used in the membership function editor for output data.

Table 6: Parameters for membership function of output

Parameters	Indicators
[3 3 5.75]	Good
[3 5.75 8.5]	Moderate
[5.75 8.5 11.25]	Unhealthy
[8.5 11.25 14]	Very Unhealthy
[11.25 14 14]	Hazardous

Table 7: Range of output

Indicators of output (AQI)	Range
Good	3 – 5.75
Moderate	5.76 – 8.5
Unhealthy	8.51 – 11.25
Very Unhealthy	11.26 - 14
Hazardous	>14

III) 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 suggested fuzzy model is based on the Mamdani fuzzy model structure as given Equation 4.

$$R_i: \text{IF } X_1 \text{ is } L_1 \wedge X_2 \text{ is } L_2 \wedge X_3, \ldots \text{is } L_{M-1} \wedge X_M \text{ is } \overset{i}{L}M \text{ THEN } \mu_R^i \rightarrow Y \tag{4}$$

Where i: 1,…n
$\overset{i}{L} \in \{L_{j,k}\}$ j: 1,…, m; k: 1,...,n; n: the number of linguistic labels defined for the j input variables

Table 8 shows the arrangement of fuzzy number for each input variable. It is used to produce IF-THEN rules. The calculation below shows the steps of how to find the minimum and maximum value for the input variables that are going to be used to find the arrangement of fuzzy number for output variables.

Minimum:	Maximum:	Difference of value:	Example of calculation:
1 + 1 + 1 = 3	6 + 4 + 4 = 14	14 – 3 = 11	Good (3-5.2)
		11 / 5 = 2.2	3 + 2.2 = 5.2

Table 8: Arrangement for three input variables and output variables

Input Variables	Fuzzy Number	Output variable	Range of Fuzzy Number
PM10	Severe (1)	AQI	Good (3-5.2)
	Very Poor (2)		Moderate (5.21-7.4)
	Poor (3)		Unhealthy (7.41-9.6)
	Moderate (4)		Very Unhealthy (9.61-11.8)
	Good (5)		Good (11.81-14)
	Excellent (6)		
CO & NO2	Poor (1)		
	Moderate (2)		
	Good (3)		
	Excellent (4)		

Based on these arrangements, 96 best rules are generated. The rules below show an example of If-Then rules of AQI generated from this process.

IF (PM10 is severe) and (CO is poor) and (NO2 is poor) THEN (AQI is hazardous)
IF (PM10 is severe) and (CO is poor) and (NO2 is moderate) THEN (AQI is hazardous)
IF (PM10 is severe) and (CO is poor) and (NO2 is excellent) THEN (AQI is very unhealthy)

Figure 9 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, unhealthy, very unhealthy, and hazardous.
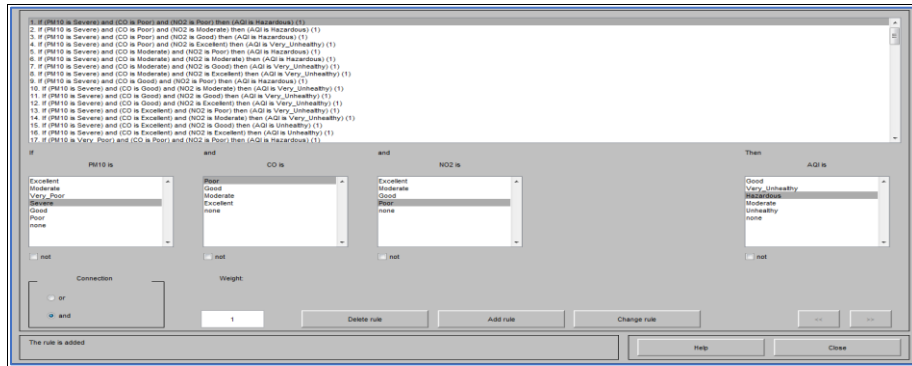

Figure 9: Rule editor

IV) Defuzzification Process

The defuzzification transforms the fuzzy value into a crisp value. In this study, a centroid method is applied, which is one of the greatest not unusual methods for defuzzification process. Fuzzy number is geometrically determined by using this approach.

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 10.
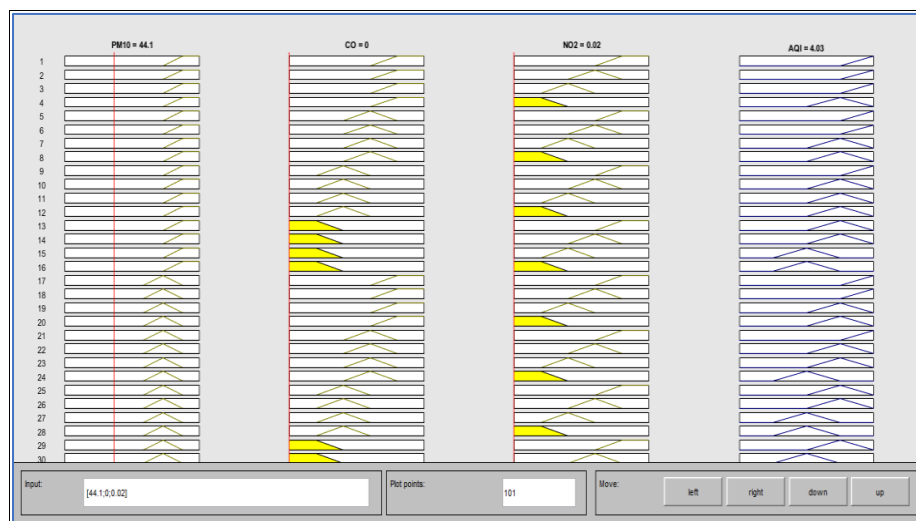

Figure 10: Rule Viewer (Rules 1- 96)

FINDINGS AND DISCUSSIONS

The results of AQI by using Mamdani Fuzzy Inference System are shown in Table 9. The values in Table 9 are part of 122 outputs in overall. The status in the linguistic value is all "Good".

Table 9: Result output of AQI from MATLAB (Mamdani status)

LOCATION	INPUTS			MATLAB (Mamdani Status)	
				OUTPUT	LINGUISTIC VALUE
	PM10	CO	NO2	AQI	
Kota Bharu	26.404	0.4325	0.008	4.04	Good
	37.94	0.625	0.01	3.91	Good
	35.612	0.45	0.009	3.88	Good
	40.104	0.4875	0.015	3.95	Good
	49.595	0.6325	0.01	3.91	Good
	48.009	0.6512	0.01	3.93	Good

The results of actual data output are shown in Table 10. The value is obtained based on the 96 If-Then rules that are mentioned in the early subtopic of data analysis. For example, IF (PM10 is Excellent) and (CO is Excellent) and (NO2 is Excellent) THEN (AQI is Good) with the input value PM10 is 26.404, CO is 0.4325 and NO2 is 0.008.

Table 10: The result of actual data output for AQI

LOCATION	INPUTS						ACTUAL DATA
	PM10	LINGUISTIC VALUE	CO	LINGUISTIC VALUE	NO2	LINGUISTIC VALUE	LINGUISTIC VALUE
Kota Bharu	26.404	Excellent	0.4325	Excellent	0.008	Excellent	Good
	37.94	Good	0.625	Excellent	0.01	Excellent	Good
	35.612	Good	0.45	Excellent	0.009	Excellent	Good
	40.104	Good	0.4875	Excellent	0.015	Excellent	Good
	49.595	Good	0.6325	Excellent	0.01	Excellent	Good
	48.009	Good	0.6512	Excellent	0.01	Excellent	Good

Table 11 shows the comparison statements of the result from actual data output and Mamdani fuzzy inference status for AQI reading. The statement is "TRUE" when the linguistic value for actual data is "Good" and Mamdani status from MATLAB is also "Good".This means that when linguistic value for actual data and Mamdani status are same, the statement is TRUE. However, if they both are with different linguistic value then the statement is "FALSE". The comparison statements in Table 11 are part of 122 statements in overall.

Table 11: The result of comparison statements between the actual data output and Mamdani status

DATE	LINGUISTIC VALUES			STATEMENTS
	ACTUAL DATA	Mamdani Status		
		AQI	MATLAB	
20160601	Good	4.04	Good	TRUE
20160602	Good	3.91	Good	TRUE
20160603	Good	3.88	Good	TRUE

20160704	Moderate	5.75	Good	FALSE
20160705	Moderate	5.75	Good	FALSE
20160706	Moderate	5.75	Good	FALSE
20160707	Good	3.93	Good	TRUE
20160708	Good	3.91	Good	TRUE

The results of comparison statements in overall are represented in Figure 11. From the total data collected within 122 days, there are 113 TRUE statements and 9 FALSE statements. That means there are 113 correct responses out of 122 data.
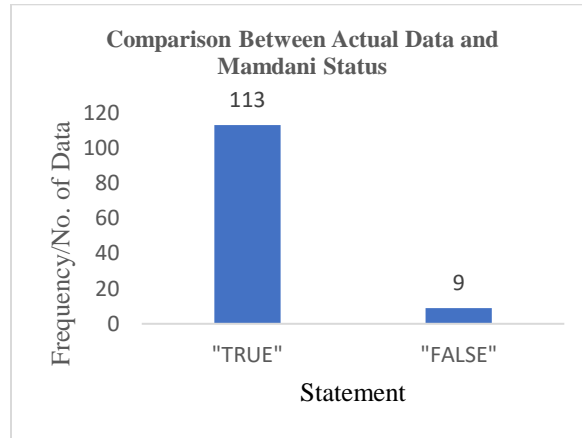

Figure 11: Number of TRUE and FALSE statements

The level of accuracy rate is calculated as follows.

$$\text{Accuracy Rate} = \frac{113}{122} \times 100$$
$$= 0.9262 \ @ \ 93\%$$

CONCLUSION AND RECOMMENDATIONS

The main objective of this study is to predict the value AQI accurately using a proposed model of Fuzzy Inference System. A model of fuzzy inference system (FIS) is developed to evaluate the air quality index (AQI) of air pollution by considering the value of three major inputs, namely PM10, CO and NO2. The evaluation system is based on Mamdani fuzzy inference system through four major steps: Fuzzification process, Fuzzy decision, Fuzzy rule-based and Defuzzification. The higher value of accuracy rate (93%) has shown that the proposed model of fuzzy inference system using Mamdani approach works relatively well and efficient in predicting the Air Quality Index (AQI). Thus, the model used is perfect and ideal to access the air quality index (AQI) accurately.

This study has applied Mamdani Rule-Based system to analyse AQI data. In future it might be possible to analyze the data by using other fuzzy inference system approaches, such as Sugeno Fuzzy Inference System.

REFERENCES

Chaudhari, S. R., & Patil, M. E. (2014). Comparative Analysis of Fuzzy Inference Systems for Air Conditioner. *International Journal of Advanced Computer Research*, 4(4), 922–927.

Sowlat, M.H., Gharibi, H., Yunesian, M., Mahmoudi, T. M., & Lotfi, S. (2011). A Novel, Fuzzy-Based Air Quality Index (FAQI) For Air Quality Assessment. *Atmospheric Environment Journal*, 45(12), 2050– 2059.

Samira, K., & Ahmad, J. (2016). A New Artificial Intelligence Method for Prediction of Diabetes Type2. *Bulletin de La Société Royale des Sciences de Liège*, 85, 376–391.

Cavallaro, F. (2015). A Takagi-Sugeno Fuzzy Inference System For Developing A Sustainability Index Of Biomass. *Sustainability Journal*, 7(9), 12359–12371.

Kumaravel, R., Vallinayagam, V., Kumaravel, R., & Vallinayagam, V. (2012). A Fuzzy Inference System for Air Quality in Using Matlab. *Journal of Environmental Research and Development*, 7(1), 484–495.

Department of Environment Malaysia (DOE) website (2000). "A guide to air pollutant index (API) in Malaysia". Retrieved Julai 01, 2020, from https://www.doe.gov.my/portalv1/en/info-umum/english-air-pollutant-index-api/100

Tiwari, P. (2015). Computational Methods of Air Quality Indices: A Literature Review. *Journal of Environmental Science, Toxicology and Food Technology*. 1(5), 46-49.