International Journal on Advanced Science Engineering Information Technology

Vol.14 (2024) No. 4
ISSN: 2088-5334

Central Tendency Data Real-Time Acid Rain Measurement to Evaluate Tool's Performance Using Statistical Analysis

Ardiansyah Ramadhan [a,*], Indra Chandra [b], Wiwiek Setyawati [c], Dyah Aries Tanti [c], Asri Indrawati [c], Achmad Faiz Alawi [b], Brety Fetrecia Br Karo [b], Viny Aulia Sabilla [b], Anandha Putri Ayu Prihatini [b]

[a] *Electrical Engineering (Graduate Program), School of Electrical Engineering, Telkom University, Bandung, Indonesia*
[b] *Engineering Physics, School of Electrical Engineering, Telkom University, Bandung, Indonesia*
[c] *The Research Center for Climate and Atmosphere, National Research and Innovation Agency, Bandung, Indonesia*
Corresponding author: [*]ardiansyah1312@gmail.com

Abstract*—Increased population growth has implications for increased industrial and transportation activities. This activity increased gas emissions in the form of Sulfur Dioxide (SO_2) and Nitrogen Dioxide (NO_2), resulting in acid rain. The measuring instrument, a rain gauge, is used to measure acid rain in Greater Bandung at Telkom University, Indonesia, but it is currently unable to measure in real-time. Therefore, this study aims to measure acid rain in real-time by measuring parameters such as pH, temperature, conductivity, and precipitation to test the acidity contained in rainwater and evaluate the tool's performance with data for March 2022. The analysis uses various statistical methods: anomaly detection, outlier detection, central tendency, person correlation analysis, Mean Absolute Percentage Error, T-test, and Analysis of Variance (ANOVA). Each parameter can be worked out in real-time based on the central tendency and rain dispersion results. If there is only daily rain, then the correlation for each parameter is the highest, namely pH if there is acid rain. The real-time average pH is 6.45; in the laboratory, it is 6.56, so the MAPE value is 5.37 (good category). Even though this tool can work well, it needs to improve the quality of the temperature control in it. Since temperature significantly affects pH, the results show a negative correlation of -0.80 between pH and temperature. In the ANOVA test, the resulting p-value, when compared with data in the laboratory, is >0.05, meaning that the average daily pH does not have a significant difference from the average laboratory test.

***Keywords*—Acid rain; real-time; statistical analysis; central tendency; evaluate performance.**

Manuscript received 12 Jul. 2023; revised 1 Jun. 2024; accepted 9 Jul. 2024. Date of publication 31 Aug. 2024.

I. INTRODUCTION

In the Positive Matrix Factorization (PMF) model, local and distant emission sources can affect the acid deposition characteristics of urban and suburban areas [1]. Gas emissions in the form of Sulfur Dioxide (SO_2) and Nitrogen Dioxide (NO_2) are the causes of acid deposition originating from industrial and transportation activities [1], [2]. Traffic activity, industrial and biomass burning, and sea salt from long-distance transport emission sources are examples of local emissions [1]. Acid rain is the acidic component included with all forms of precipitation, as sulfuric or nitric acids that fall to the ground from the atmosphere in the dry or wet form are released into the atmosphere and transported by wind and air currents [2]. Rain is naturally acidic with a pH <6, but rain with a pH below 5.6 is defined as acid rain [2]. Measurement of acid rain through wet deposition and dry deposition has been carried out using manual sampling in both urban and suburban areas of Indonesia. The year 2019 - 2022 using wet and dry deposition data [1]. The pH value at the Telkom University measurement location tends to be more acidic than the Cipedes and Tanjung Sari measurement locations because the anion concentration reaches 52% of the total concentration at Telkom University [1]. Table 1 shows 2015-2019 and other information about sampling locations for manual acid rain measurements [3].

The potential cause of rainwater acidity can be seen through NO_3- and $nssSO_4^{2-}$ ions, while the neutralizing compounds are NH_4+ and $nssCa_2$+. The nitrate equivalent fraction showed that HNO_3 was more influential in acid deposition in Jakarta, Serpong, and Bandung, but in Kototabang and Maros, it was more due to H_2SO_4 [3]. Several studies on air pollution in cities in Indonesia indicate health, economic, and climate change issues [4]-[8]. These studies correlate with increased acid rain from actual emissions [9],

[10]. Assessment and monitoring of acid rain in the Bandung area have identified several processes and measurement parameters that have an adverse impact as an initial stage but are not real-time yet [11], [12]. The impact of acid rain results in contamination of the soil [13]-[17], composition [13], [14], [18], damage to materials or buildings [19], the emergence of diseases that affect human or animal health [2], damage to asphalt [20]-[22] and plant damage [23]-[26].

TABLE I
ACID DEPOSITION MONITORING LOCATIONS IN INDONESIA WHICH ARE MEMBERS OF EANET [3]

Location	Class	Latitude (LS)	Longitude (BT)	Altitude (m)
Serpong	rural	$06^0$21'01.9"	$106^0$40'04.7"	64
Bandung	urban	$06^0$53'41.63"	$107^0$35'11.1"	753
Jakarta	urban	$06^0$09'21.51"	$106^0$50'32.67"	7
Kota Tabang	isolate	$0^0$12'8.70"	$100^0$19'4.50"	845
Maros	rural	04^0 59' 50.29'	$119^0$34'17.73"	1

Water acidity has been assessed and monitored in several countries [27]-[30]. In Indonesia, acid rain monitoring is manual [3], and a calibrated real-time rainwater monitoring tool has been carried out [31]. Manual rainwater quality measurement tools [3] still use a rain gauge that collects rainwater-taking rainwater manually-storage samples in the freezer-measurement each parameter in the laboratory [11]. In contrast, real-time rainwater quality measurement tools [31] can measure direct data when it rains with all the same parameters measured in the laboratory. The data was validated using the manual tool, but it required a very long process because it used a tool that could not directly measure all parameters. The real-time tool [31] has been calibrated but requires data validation to compare the resulting data with manual tools [3]. Therefore, in this study, a comparative test of the results of the two processes and statistical analysis was carried out between the results of manual measurements carried out in the laboratory and the results of real-time data that have been processed using machine learning as an evaluation of the performance of tools for initial studies taken from one of the locations. The measurement location is at Telkom University, Bandung, Indonesia, as the data is processed.

II. MATERIALS AND METHOD

The data taken are samples from manual tools [3] and real-time data tools [31] based on measurement results at Telkom University in March 2022. The method used is anomaly detection, and outliers in sliding windows are carried out on real-time data. Fig. 1 shows a statistical analysis of a real-time acid rain data scheme to evaluate tools. Comparisons were made on laboratory data using samples [3] and [31] against the validated real-time daily average data during the rain.

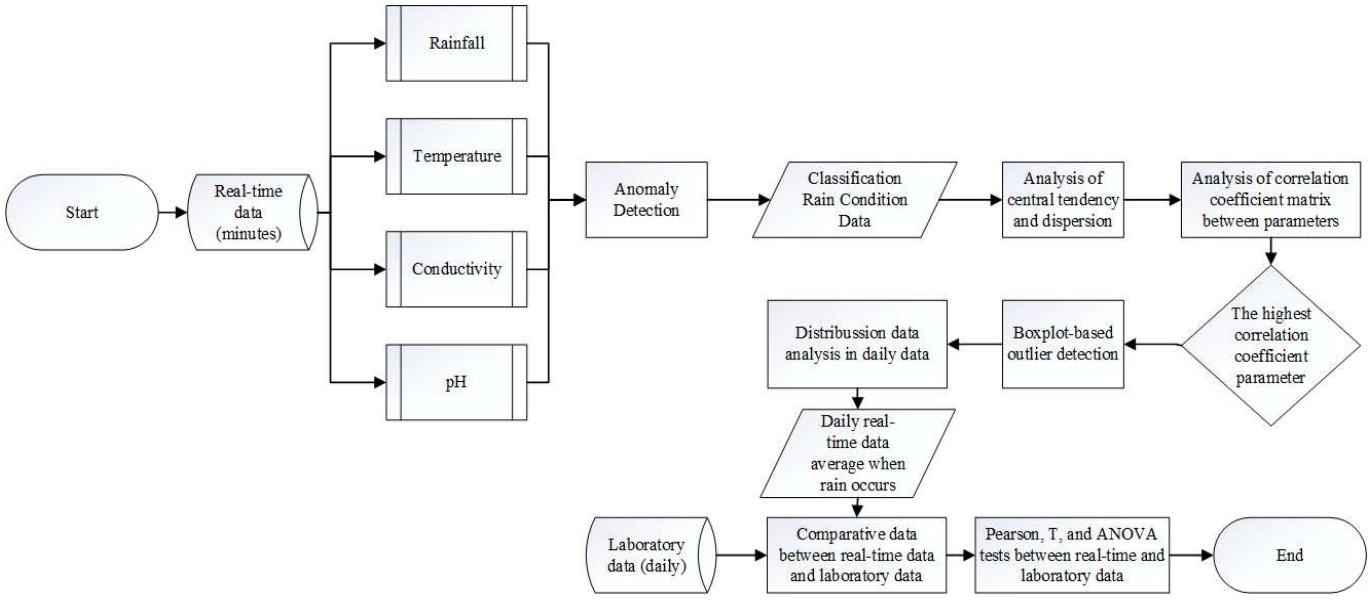

Fig. 1 Statistical Analysis of Real-Time Acid Rain Measurement Data Scheme to Evaluate Tools

A. Anomaly Detection

One increasingly effective technique is Machine Learning (ML), which is essential in this field [32]. One technique for identifying unexpected items or events in a data set that differ from the norm is anomaly detection [33]. At this stage, anomaly detection is carried out on the characteristics of each sensor used, namely pH [34], conductivity [35], and temperature [36].

B. Boxplot-based Outlier Detection

Outliers are carried out in stages, starting from raw data and data that will be used for comparison. Tukey (1977) introduces rules for constructing boxplots for filtering outliers [37]. This method drains the quartiles, not the mean and standard deviation [37]. Symmetrical and skewed data can use the Turkey method [37]. The Tukey method has the following rules:

$$Q_1 - 1.5IQR, Q_3 + 1.5IQR \quad (1)$$

$$Q_1 - 3IQR, Q_3 + 3IQR \quad (2)$$

Q_1 is the first and Q_3 is the third quartile [37]-[38]. The difference between the inside and outside fences is the IQR (interquartile range) [37]-[38]. The interval with 1.5 IQR is called the inner fence located below Q_1 and above Q_3 at a distance of 1.5 IQR, and the interval with 3 IQR is called the outer fence located below Q_1 and above Q_3 at a distance of 3IQR [37]-[38]. Observations between the inner fence and the outer fence are potential outliers, and observations as possible outliers outside the outer fence [37]-[38]. With this, it can detect more observations as outliers as the skewness of the data increases [37]-[38].

C. Central Tendency

Characteristic functions that provide information on a virgin, such as the average, median, and mode, refer to the location of the distribution. We will measure this for the population (that is, the set of all the elements we describe) and for a sample drawn from the population, as well as for grouped and ungrouped data [39], [40].

1) The population mean: denoted by (Greek letter mu), and for the sample, by X (pronounced "X bar") [40]. For ungrouped data, X is calculated using the following formula [40]:

$$\mu = \frac{\Sigma X}{N} \quad and \quad \bar{x} = \frac{\Sigma X}{n} \quad (3)$$

ΣX refers to the sum of all observations, for N and n refers to the number of observations in the population and sample. For grouped data, X is calculated by [40]

$$\mu = \frac{\Sigma fX}{N} \quad and \quad \bar{x} = \frac{\Sigma fX}{n} \quad (4)$$

2) Median: The median for ungrouped data is the middle value when all are arranged in ascending or descending order of value [40]:

$$median = the\left(\frac{N+1}{2}\right) th\ item\ in\ the\ data\ array \quad (5)$$

N refers to the number of items in the population (n for a sample). The median for grouped data is given by the formula [35]:

$$median = L + \frac{\frac{n}{2} - F}{f_m} \quad (6)$$

where: L = lower limit of the median class, n = the number of observations in the data set, F = sum of the frequencies up to but not including the median class, f_m = frequency of the median class, c = width of the class interval [40].

3) Mode: The value that appears most often in the data set is the mode. For grouped data, we obtain [40]:

$$mode = L + \frac{d_1}{d_1 + d_2} c \quad (7)$$

where: L = lower limit of the modal class, d_1= frequency of the modal class minus the frequency of the previous class, d_2= frequency of the modal class minus the frequency of the following class and c = width of the class interval [40]. Dispersion refers to the variability or spread of the data. The most important measures of dispersion are (1) the average deviation, (2) the variance, and (3) the standard deviation. [39]-[40].

D. Pearson's Correlation

In 2010 [41], Pearson correlation analysis on acid rain was carried out, but this study used manual data using a rain gauge [3].

$$r = \frac{n\sum_{i=1}^{n} X_i Y_i - \sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{\sqrt{n\sum_{i=1}^{n} X_i{}^2 - \left(\sum_{i=1}^{n} X_i{}^2\right)^2}\ \sqrt{\sum_{i=1}^{n} Y_i{}^2 - \left(\sum_{i=1}^{n} Y_i{}^2\right)^2}} \quad (8)$$

A probability value (p-value) is used to test the influence between variables, or in SPSS; it is written GIS (significance). If the p-value is lower than α (alpha), the influence between variables is significant. In contrast, if the p-value is higher than α, the influence between variables is insignificant. The value of α is the value of the confidence coefficient, which in this calculation is used α =0.05 or a 95% confidence level [41].

E. Mean Absolute Percentage Error (MAPE)

The statistical tool used to measure the accuracy of statistical models in making predictions or forecasting is the Mean Absolute Percentage Error (MAPE) [42]. In this real-time measurement of acid rain, MAPE is used to see the tool's performance from the values obtained to be compared with results in the laboratory.

$$MAPE = \frac{1}{n}\sum \left|\frac{A_i - F_i}{A_i}\right| x\ 100\% \quad (9)$$

where: n = sample size, A_i = actual data values, F_i =forecasting data values.

F. T and Analysis of Variance (ANOVA) Test

In statistics, the T-test and ANOVA tests are widely used in analyzing data [43]-[45], focusing on independent two-sample t-tests and one-way ANOVA. The t-test function tests the difference between two independent population means. In real-time, measurements of laboratory testing values are carried out by looking at the T-test results' characteristics to evaluate the tool's performance from the data obtained. Welch's t statistic can be calculated using the following formula:

$$t = \frac{M_2 - M_1}{\sqrt{\frac{s_1^2}{n_1}}} \quad (10)$$

where: t = t-test statistic, M_2= mean of the second group, M_1= mean of the second group, s_1^2 = standard deviation of the first group, and n_1= sample size of the first group. In one-way ANOVA, test the mean difference between two or more independent groups [46]. As with Welch's t-test, the F* Brown-Forsythe statistic [47], which is robust to violations of the HOV assumption, and can be calculated using individual group variances as follows:

$$F^* = \frac{\Sigma_{j=1}^{k} n_j (M_j - M)^2}{\Sigma_{j=1}^{k} (1 - \frac{n_j}{N}) s_j^2} \quad (11)$$

III. Results and Discussion

Fig. 2 shows the result of real-time data on measurements of acid rain in March 2022. This measurement shows changes

in the values of various parameters to measure the level of acid rain. The parameters measured were pH, temperature, rainfall, and conductivity. Each parameter is measured daily for 24 hours with a validated data retrieval resolution in minutes. The data displayed results from daily measurements of tools that work in total before being classified into rainy conditions. The tool's performance on the graph can work and is responsive according to rainy or weather conditions. However, this figure has not focused on classifying when it rains.

When rainwater falls, it is read by the rain sensor. Furthermore, rainwater will fall into the chamber to measure its quality. Each parameter will continue to be measured until rainwater is removed manually. Furthermore, from the real-time data results, the overall identification of centroid tendency and the analysis that will be carried out only use data on a rainy day. In March 2022, Fig.3 shows five rainy days: the 7th, 8th, 10th, 11th, and 13th.

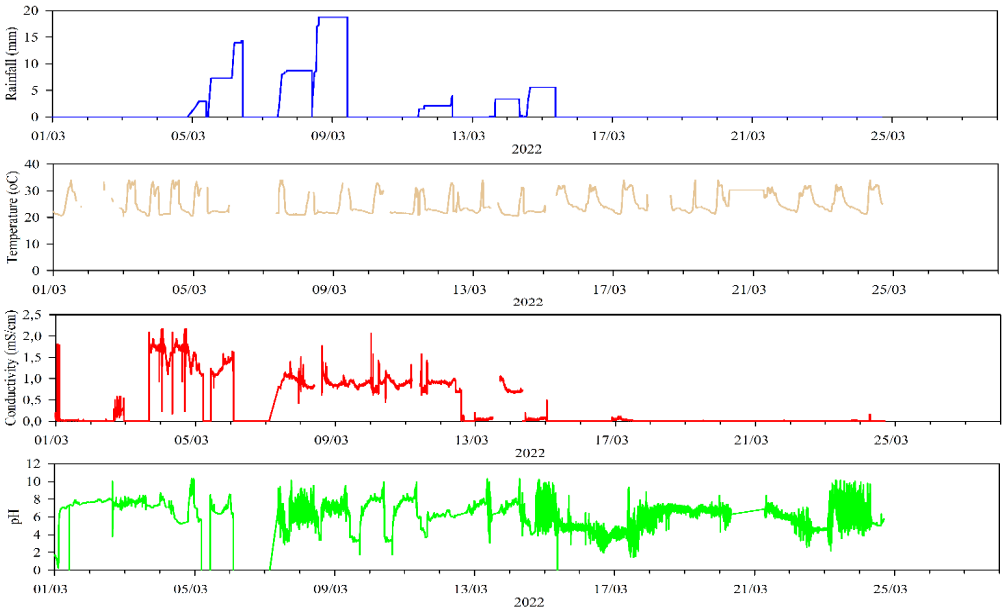

Fig. 2 Real-time measurement of acid rain March 2022

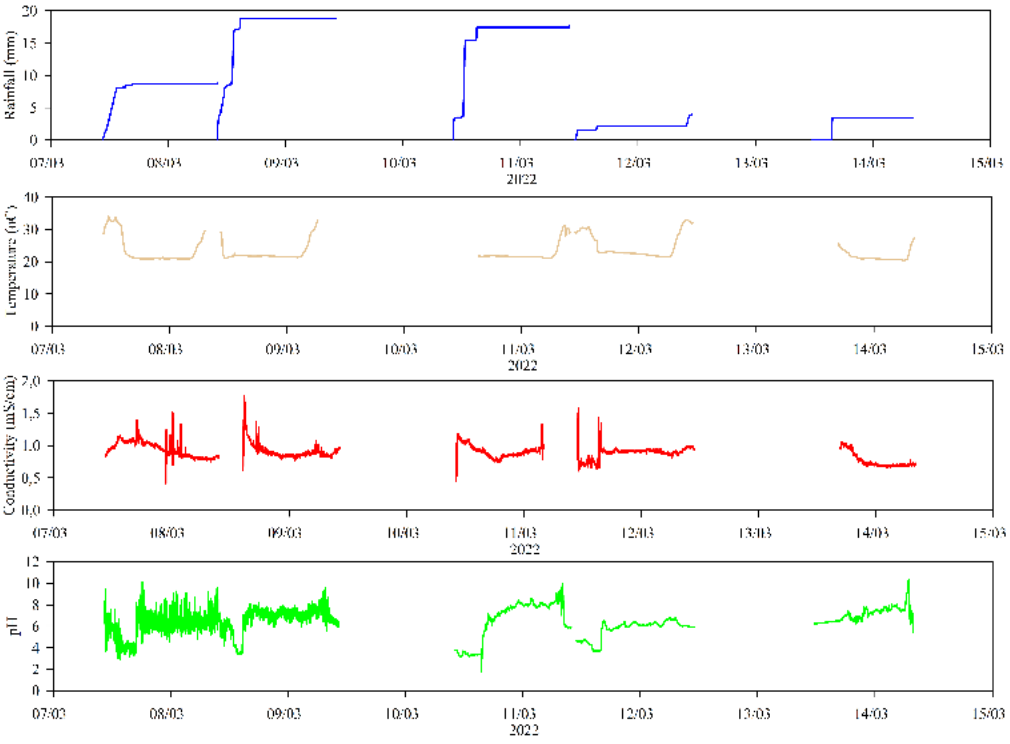

Fig. 3 Acid rain measurement data when conditions occurred rain in March 2022

Each of these data will be used as analysis data per day. Each day will identify each parameter; the results on real-time rain data have been classified, but this analysis requires further study. In Facilitating the analysis of daily rainfall that occurs every day, an initial classification of rain is carried out according to the time of occurrence of rain in Table 2.

TABLE II
INITIALIZATION OCCURS IN RAINY CONDITIONS

No	Date	Condition	Classification
1.	March 7th 2022	Rain	D_1
2.	March 8th 2022	Rain	D_2
3.	March 10th 2022	Rain	D_3
4.	March 11th 2022	Rain	D_4
5.	March 13th 2022	Rain	D_5

Each parameter is analyzed by central tendency and daily dispersion by looking at the distribution of real-time measurement data that occurs [39]. Preliminary analysis: mean, median, mode, standard deviation, and precision. The purpose of this measurement is the same as previous studies, namely, to see the work of time intervals [48], describing the data model [39], and investigation of a pattern [49]. Real-time acid rain shows central tendency and dispersion data, as shown in Table 3, which has been classified based on daily rain, which starts when it rains until it is taken manually. The aim is to collect rainwater manually even though there is real-time data because it can be compared with results in the laboratory [1], [3], [11], [41].

Based on the data presented, the central tendency data for real-time rain measurements from D1-D5 can work well despite the relatively high measured dispersion. Temperature is one with reasonably high dispersion [50]-[51]. The high data distribution on this parameter is caused by the absence of a temperature controller that can withstand the temperature conditions in the measurement medium. In addition, the daily cycle from morning to afternoon to evening to night produces temperature data that will continue to vary. Even though the temperature parameter has a high standard deviation, this parameter is quite precise compared to other data [46]. Temperature is closely related to pH [46]. The pH value is low when the temperature is high, and the pH increases when the temperature is low [46].

TABLE III
CENTRAL TENDENCY AND DISPERSION OCCUR IT IS RAINING

Class	Parameters	Rainfall	Temp	EC	pH
D1	n	1410	1248	1389	1408
	Min	0.00	20.86	0.42	2.92
	Max	8.79	34.00	1.51	10.10
	Mean	8.01	23.80	0.94	6.06
	Mode	8.68	20.88	0.80	6.06
	Median	8.68	21.06	0.94	6.22
	Std	1.83	4.39	0.12	1.12
	Precision	5.50	13.16	0.37	3.36
D2	n	1461	1192	1192	1461
	Min	0.00	21.13	0.62	3.39
	Max	18.74	32.77	1.77	9.55
	Mean	17.08	22.74	0.91	6.39
	Mode	18.74	21.63	0.85	7.08
	Median	18.74	21.75	0.88	6.98
	Std	4.07	2.59	0.12	1.07
	Precision	12.22	7.76	0.37	3.22
D3	n	1428	1128	1082	1428
	Min	0.00	21.19	0.45	1.73
	Max	17.72	31.13	1.33	9.97
	Mean	15.83	22.65	0.91	6.56
	Mode	17.43	21.56	0.94	3.36
	Median	17.43	21.63	0.91	7.45
	Std	4.16	2.83	0.10	1.92
	Precision	12.48	8.48	0.31	5.75
D4	n	1404	1404	1404	1404
	Min	0.00	21.50	0.61	3.71
	Max	4.03	32.94	1.58	6.79
	Mean	2.09	24.93	0.89	5.79
	Mode	2.15	22.69	0.92	6.05
	Median	2.15	22.85	0.91	6.05
	Std	0.43	3.76	0.09	0.80
	Precision	1.28	11.28	0.26	2.39
D5	n	1237	926	926	1207
	Min	0.00	20.44	0.66	5.48
	Max	3.38	27.27	1.06	10.30
	Mean	2.73	21.66	0.76	7.09
	Mode	3.38	21.06	0.71	6.42
	Median	3.38	21.00	0.71	7.12
	Std	1.32	1.46	0.11	0.69
	Precision	3.96	4.39	0.33	2.08

Furthermore, this parameter's high standard deviation is due to the addition of data that continues to occur when it rains, so this sensor has quite varied data. Following the principle of the rain sensor, each end of the water will produce an increase in data worth 0.000236667 mm. To identify a correlation matrix between parameters, a real-time comparison against the average values listed in Table 4 is carried out.

TABLE IV
CORRELATION COEFFICIENT MATRIX BETWEEN PARAMETERS REAL-TIME RAIN MEASUREMENT

	Rainfall	pH	EC	Temp
Rainfall		0.05	0.58	-0.19
pH	0.05		-0.73	-0.80
EC	0.58	-0.73		0.60
Temp	-0.19	-0.80	0.60	

Rainfall has a relatively small correlation with -0.19 (to temperature), so the correlation between these two variables is not strong [52]-[53]. However, pH and conductivity are positively correlated for rainfall, so the results show the same as research done manually in the laboratory [41]. In addition, pH and temperature are negatively correlated, which overall, the negative correlation between pH and temperature can be caused by a combination of chemical, physical, and biological factors that affect the water solution in the environment [54]-[56]. Furthermore, pH has the strongest correlation to all parameters, so in the case of acid rain, it needs further analysis, as shown in Figure 4, where the real-time results are carried out for outlier detection to eliminate installment data.

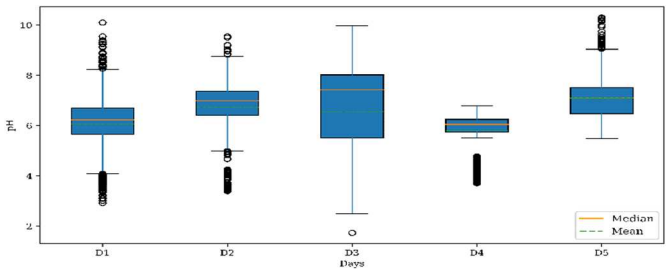

Fig. 4 Boxplot-based outlier detection of real-time pH parameters when it rains

There is one outlier in the D3 figure above, but the median value to the average value is quite far. Therefore, it is necessary to study the distribution of data related to all data when it rains because there are still outlier data that need to be removed to ensure that D1-D5 data has been validated before taking the average value. Statistically, the data distribution regarding the mean, median, and mode can analyze the data distribution, whether positively skewed, negatively skewed, or symmetrical. From the analysis, referring to Fig. 5, D1, D3, and D5 skewed positively because mean>median>mode and D2 and D3 doubled negatively because mode>median>median. This rule refers to the skew of the frequency distribution [40]. The assumption included in the data distribution is a normal distribution. This real-time validation aims to allow data to be compared head-to-head with results in the laboratory and minimize error.

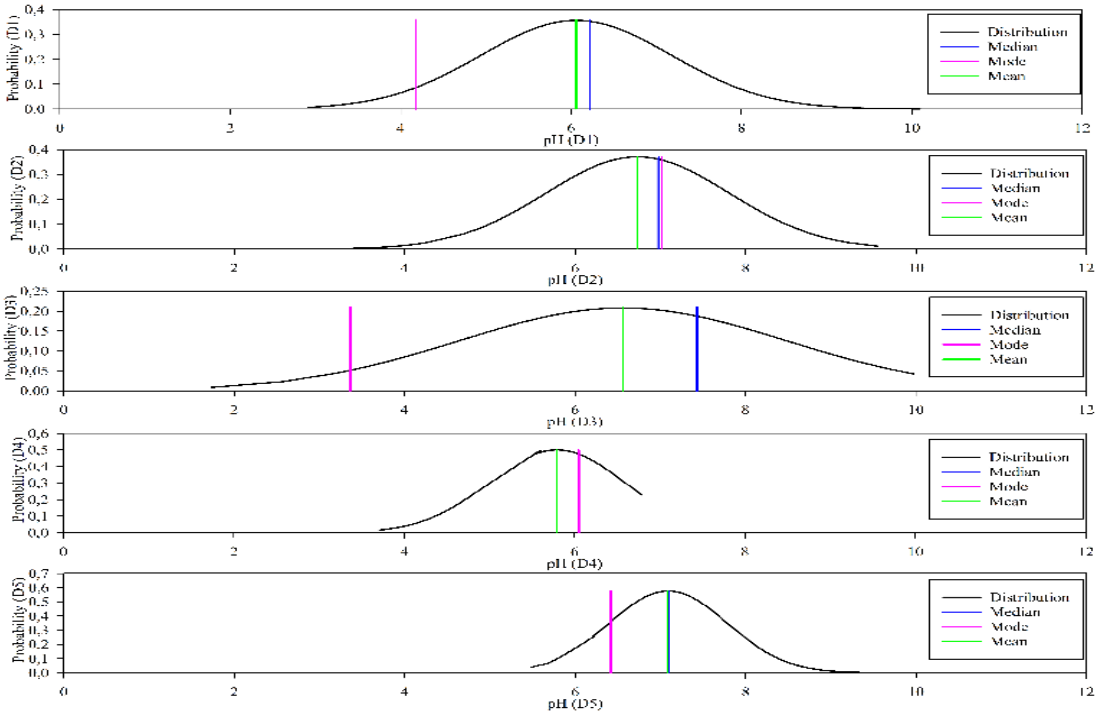

Fig. 5 Data distribution of each real-time pH parameter

After removing the outliers [57] and analyzing the real-time distribution of the pH parameter data, the validation data will be compared [58]. Validation takes daily average data for real-time and laboratory average results. Figure 6 shows the comparative data between the two results.

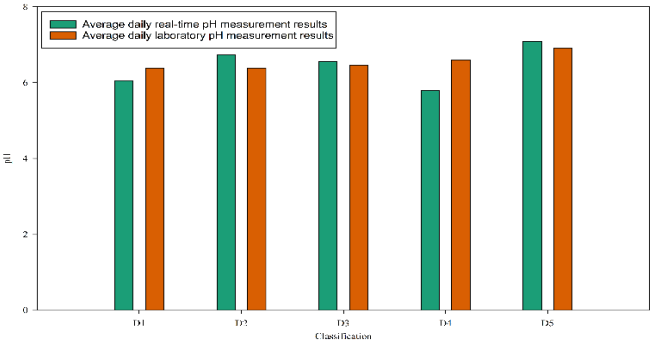

Fig. 6 Comparison of real-time daily average pH data and laboratory averages

During D1-D5 real-time measurements, the average value is 6.45, and the laboratory is 6.56. Moreover, it obtained a MAPE value of 5.37, which is categorized as good [59]. The data to compare has a low error rate, so daily real-time results can approximate laboratory results. This data must go through anomaly detection and outlier detection so that real-time measurement results can be validated. After the two averages are compared, a t-test can be carried out, assuming the variance of the data is the same, as shown in Table 5.

TABLE V
T-TEST: TWO-SAMPLE ASSUMING EQUAL VARIANCES AVERAGE DAILY LABORATORY AND REAL-TIME

	Average daily laboratory pH measurement results (Ai)	**Average daily real-time pH measurement results (Fi)**
Mean	6.556	6.448503669
Variance	0.04948	0.273129212
Observations	5	5
Pooled Variance	0.161304606	
Hypothesized Mean Difference	0	
df	8	
t Stat	0.42319474	
P(T<=t) one-tail	0.341650178	
t Critical one-tail	1.859548038	
P(T<=t) two-tail	0.683300355	
t Critical two-tail	2.306004135	

With the results of hypothesis 0, both the average daily yield data in March, there is no significant difference with a p-value level> 0.05, which means that real-time data and data have minimal data differences. However, for further statistical analysis, namely by testing the comparison of data groups with the ANOVA test. The summary of ANOVA in Table 6 has the same variance and average as Table 5.

The p-value generated from the ANOVA test in Table 7 is > 0.05, meaning that there is no difference in both laboratory data and actual data so that the daily average value can be used as an initial analysis of acid rain without having to decide which is better and faster than the previous process. Variability between groups (Between Groups) has a Sum of Squares (SS) value of 0.029, indicating a difference in variability between groups or treatments in the study.

TABLE VI
SUMMARY ANOVA

Groups	Count	Sum	Average	Variance
Ai	5	32.78	6.556	0.04948
Fi	5	32.24251834	6.448504	0.273129

Meanwhile, within groups (Within Groups) has an SS value of 1.3, which describes the variation in the same group or similar treatment. The analysis results show no significant difference between groups in terms of average because the p-value (0.7) is more significant than the significance level usually used (usually 0.05). The p-value indicates the statistical significance of the F-ratio. F-ratio is the ratio between the Mean Squared (MS) between groups and MS within groups. The unachieved F critical value (F crit = 5.31) also confirms that the F Critical obtained is the limit value compared to the F-ratio value. There is no significant difference between groups.

Thus, there is insufficient evidence to reject the null hypothesis, which states no significant difference between groups regarding average. Compared to previous studies [41], it still uses laboratory data without comparison with real-time data. However, to obtain consistent daily pH data results between laboratory and real-time data, this study found that IoT-based rain quality measuring instruments that use real-time data can meet the requirements of statistical testing [60], It is necessary to go through various statistical test processes, such as anomaly and outlier detection, to compare daily results with laboratory results [61]-[63]. The tool's data can be evaluated through the stages carried out to improve the best results. Overall, various statistical processes and tests show ANOVA values not far from the real-time and laboratory average data.

TABLE VII
ANOVA VARIANCES AVERAGE DAILY LABORATORY AND REAL-TIME

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	0.029	1	0.029	0.18	0.7	5.31
Within Groups	1.3	8	0.16			
Total	1.32	9				

IV. CONCLUSION

Based on central tendency and dispersion data during rainfall by measuring min, max, mean, median mode, standard deviation, and precision, rainwater quality measurement tools can work on each parameter in real time. These parameters are pH, rainfall, temperature, and conductivity. Real-time data, data is stored and can be processed statistically. Real-time data is processed when it rains or cannot be used as an alternative data information system other than manual measurements. It is evident from the normal distribution of data during measurements and the daily data results, which, after going through the anomaly detection and outlier detection processes, can be compared with the results in the laboratory. The average real-time pH is 6.45, and the laboratory is 6.56, so the MAPE value of 5.37 is categorized as good in March 2023 so that this tool can identify rain quality conditions. Even though this tool can work well, its temperature control quality needs to be improved. Because pH greatly influences temperature, this can be seen from the results, which show a robust negative correlation, namely -0.80, between pH and temperature. pH - temperature can be caused by a combination of chemical, physical, and biological factors that affect the water solution in the environment.

Even though pH is significant for identifying acid rain and does not significantly correlate with rainfall and conductivity, other supporting parameters, such as rainfall and conductivity, are still needed to determine whether or not rain conditions and further analysis when different conditions exist. In the ANOVA test, the p-value generated after comparison with data in the laboratory is >0.05, meaning that the daily average pH does not have a large significant difference with the average laboratory test. The implications for the future will be even better when the data has been validated in real-time so that it can be used for information systems and decision support systems or early warning systems because based on performance, the results of performance evaluated through statistical data are currently suitable for producing information on acid rain conditions—or not acid rain in real-time.

ACKNOWLEDGMENT

This research was financially supported by Bandung Techno Park Telkom University (246/BTP/PKS.01/0.0/2022) and DIKTI through the Directorate of Research and Community Service Telkom University (983/PNLT3/PPM/2022). We thank the Central Laboratory for Climate and Atmospheric Research—National Research and Innovation Agency (BRIN) for collaborating and providing access to laboratory facilities and tools (006/SAM4/TE-DEK/2019).

REFERENCES

[1] A. Indrawati *et al.*, "Characteristics of Acid Deposition in Urban and Sub-Urban Area," in *Springer Proceedings in Physics*, Springer Science and Business Media Deutschland GmbH, 2022, pp. 171–182. doi: 10.1007/978-981-19-0308-3_13.

[2] U.S. EPA, "Acid Rain," U.S. Environmental Protection Agency. Accessed: Jul. 15, 2023. [Online]. Available: https://www.epa.gov/acidrain/what-acid-rain

[3] R. Puji Lestari *et al.*, "Chemical Composition of Wet Deposition in Jakarta, Serpong, Bandung, Kototabang, and Maros During 2015-2019," *ECOLAB*, vol. 15, no. 2, pp. 89–100, 2021, doi:10.20886/jklh.2021.15.2.89-100.

[4] Y. Kim and V. Radoias, "Severe Air Pollution Exposure and Long-Term Health Outcomes," *Int J Environ Res Public Health*, vol. 19, no. 21, Nov. 2022, doi: 10.3390/ijerph192114019.

[5] G. Syuhada *et al.*, "Impacts of Air Pollution on Health and Cost of Illness in Jakarta, Indonesia," *Int J Environ Res Public Health*, vol. 20, no. 4, Feb. 2023, doi: 10.3390/ijerph20042916.

[6] H. Sani, T. Kubota, J. Sumi, and U. Surahman, "Impacts of Air Pollution and Dampness on Occupant Respiratory Health in Unplanned Houses: A Case Study of Bandung, Indonesia," *Atmosphere (Basel)*, vol. 13, no. 8, Aug. 2022, doi:10.3390/atmos13081272.

[7] L. K. Widya *et al.*, "Comparison of spatial modelling approaches on PM10 and No2 concentration variations: A case study in Surabaya city, Indonesia," *Int J Environ Res Public Health*, vol. 17, no. 23, pp. 1–15, Dec. 2020, doi: 10.3390/ijerph17238883.

[8] I. Chandra, K. Nisa, and E. Rosdiana, "Preliminary study: Health risk analysis of PM2.5 and PM10 mass concentrations in Bandung Metropolitan," in *IOP Conference Series: Earth and Environmental Science*, IOP Publishing Ltd, Jul. 2021. doi: 10.1088/1755-1315/824/1/012049.

[9] X. Du, L. Zhang, and H. Ijang, "Variation characteristics and correlation analysis of air quality index and acid rain in Dongying City," Mar. 2021.

[10] X. Chen *et al.*, "Analysis of the Spatio-temporal Changes in Acid Rain and Their Causes in China (1998-2018)," *J Resour Ecol*, vol. 12, no. 5, pp. 593–599, Sep. 2021, doi: 10.5814/j.issn.1674-764x.2021.05.002.

[11] D. Rievaldo *et al.*, "Effect of Rainfall on The Chemical Composition of Rainwater in Monitoring Acid Deposition In Greater Bandung," *Indonesian Journal of Urban And Environmental Technology*, pp. 49–62, Feb. 2023, doi: 10.25105/urbanenvirotech.v6i1.14051.

[12] N. Y. Hasan, Driejana, A. Sulaeman, and H. D. Ariesyady, "Water quality indices for rainwater quality assessment in Bandung urban region," in *IOP Conference Series: Materials Science and Engineering*, Institute of Physics Publishing, Nov. 2019. doi:10.1088/1757-899X/669/1/012044.

[13] W. Zhang, K. Zhao, B. Wan, Z. Liang, W. Xu, and J. Li, "Chromium Transport and Fate in Vadose Zone: Effects of Simulated Acid Rain and Colloidal Types," *Int J Environ Res Public Health*, vol. 19, no. 24, Dec. 2022, doi: 10.3390/ijerph192416414.

[14] Z. Liu *et al.*, "Effect of simulated acid rain on soil CO2, CH4 and N2O emissions and microbial communities in an agricultural soil," *Geoderma*, vol. 366, May 2020, doi:10.1016/j.geoderma.2020.114222.

[15] M. Zhou, H. Hu, J. Wang, Z. Zhu, and Y. Feng, "Nitric Acid Rain Increased Bacterial Community Diversity in North Subtropical Forest Soil," *Forests*, vol. 13, no. 9, Sep. 2022, doi: 10.3390/f13091349.

[16] Y. Zheng, Y. Wang, Y. Zheng, and Y. Li, "Effects of Simulated Acid Rain on Soil Enzyme Activity and Related Chemical Indexes in Woodlands," *Forests*, vol. 13, no. 6, Jun. 2022, doi:10.3390/f13060860.

[17] J. Long *et al.*, "The leaching of antimony and arsenic by simulated acid rain in three soil types from the world's largest antimony mine area," *Environ Geochem Health*, vol. 44, no. 12, pp. 4253–4268, Dec. 2022, doi: 10.1007/s10653-021-01188-3.

[18] Z. Liu *et al.*, "Higher sensitivity of microbial network than community structure under acid rain," *Microorganisms*, vol. 9, no. 1, pp. 1–18, Jan. 2021, doi: 10.3390/microorganisms9010118.

[19] I. Rodríguez, A. Ortiz, P. Caldevilla, S. Giganto, G. Búrdalo, and M. Fernández-Raga, "Comparison between the Effects of Normal Rain and Acid Rain on Calcareous Stones under Laboratory Simulation," *Hydrology*, vol. 10, no. 4, Apr. 2023, doi:10.3390/hydrology10040079.

[20] X. Zhang, I. Hoff, and R. G. Saba, "Response and deterioration mechanism of bitumen under acid rain erosion," *Materials*, vol. 14, no. 17, Sep. 2021, doi: 10.3390/ma14174911.

[21] J. Wei, Q. Chen, J. Du, K. Liu, and K. Jiang, "Study on the Durability of Acid Rain Erosion-Resistant Asphalt Mixtures," *Materials*, vol. 15, no. 5, Mar. 2022, doi: 10.3390/ma15051849.

[22] J. Wei, Q. Li, K. Liu, and K. Jiang, "Study on Acid Rain Corrosion Resistance and Action Mechanism of Graphene Oxide Modified Epoxy Asphalt," *International Journal of Pavement Research and Technology*, 2023, doi: 10.1007/s42947-023-00292-0.

[23] A. M. Hamed and B. I. Wahab, "Measuring the Acid Rain in Heet City of Iraq," in *IOP Conference Series: Earth and Environmental Science*, Institute of Physics, Aug. 2022. doi: 10.1088/1755-1315/1060/1/012022.

[24] V. M. Rodríguez-Sánchez, U. Rosas, G. Calva-Vásquez, and E. Sandoval-Zapotitla, "Does acid rain alter the leaf anatomy and photosynthetic pigments in urban trees?," *Plants*, vol. 9, no. 7, pp. 1–16, Jul. 2020, doi: 10.3390/plants9070862.

[25] Z. Shi, J. Zhang, Z. Xiao, T. Lu, X. Ren, and H. Wei, "Effects of acid rain on plant growth: A meta-analysis," *Journal of Environmental Management*, vol. 297. Academic Press, Nov. 01, 2021. doi:10.1016/j.jenvman.2021.113213.

[26] H. T. T. Pham, A. T. Nguyen, T. T. H. Nguyen, and L. Hens, "Stakeholder Delphi-perception analysis on impacts and responses of acid rain on agricultural ecosystems in the Vietnamese upland," *Environ Dev Sustain*, vol. 22, no. 5, pp. 4467–4493, Jun. 2020, doi:10.1007/s10668-019-00393-6.

[27] M. O. Eyankware and B. E. Ephraim, "A Comprehensive Review of Water Quality Monitoring and Assessment in Delta State, Southern Part of Nigeria," *Journal of Environmental and Earth Sciences*, vol. 3, no. 1. Bilingual Publishing Co., pp. 16–28, Apr. 01, 2021. doi:10.30564/jees.v3i1.2900.

[28] B. O. Rosseland, "The legacy from the 50 years of acid rain research, forming present and future research and monitoring of ecosystem impact: This article belongs to Ambio's 50th Anniversary Collection. Theme: Acidification," *Ambio*, vol. 50, no. 2, pp. 273–277, Feb. 2021, doi: 10.1007/s13280-020-01408-7.

[29] H. T. T. Pham, A. T. Nguyen, A. T. N. Do, and L. Hens, "Impacts of simulated acid rain on the growth and the yield of soybean (Glycine max (l.) merr.) in the mountains of northern vietnam," *Sustainability (Switzerland)*, vol. 13, no. 9, May 2021, doi: 10.3390/su13094980.

[30] T. Navrátil *et al.*, "Mercury cycling during acid rain recovery at the forested Lesní potok catchment, Czech Republic," *Hydrol Process*, vol. 35, no. 6, Jun. 2021, doi: 10.1002/hyp.14255.

[31] I. Chandra *et al.*, "Peralatan Pemantauan Air Hujan," P00202208927, Aug. 21, 2022

[32] A. B. Nassif, M. A. Talib, Q. Nasir, and F. M. Dakalbab, "Machine Learning for Anomaly Detection: A Systematic Review," *IEEE Access*, vol. 9. Institute of Electrical and Electronics Engineers Inc., pp. 78658–78700, 2021. doi: 10.1109/access.2021.3083060.

[33] H. Su, Z. Wu, H. Zhang, and Q. Du, "Hyperspectral Anomaly Detection: A survey," *IEEE Geosci Remote Sens Mag*, vol. 10, no. 1, pp. 64–90, Mar. 2022, doi: 10.1109/mgrs.2021.3105440.

[34] DFROBOT, "SKU: SEN0161." Accessed: Aug. 30, 2023. [Online]. Available: https://wiki.dfrobot.com/PH_meter_SKU__SEN0161_

[35] DFROBOT, "SKU: DFR0300." Accessed: Aug. 30, 2023. [Online]. Available: https://wiki.dfrobot.com/Analog_EC_Meter_SKU_DFR0300

[36] DFROBOT, "SKU: DFR0198." Accessed: Aug. 30, 2023. [Online]. Available: https://wiki.dfrobot.com/Waterproof_DS18B20_Digital_Temperature_Sensor__SKU_DFR0198_

[37] A. Boukerche, L. Zheng, and O. Alfandi, "Outlier Detection: Methods, Models, and Classification," *ACM Comput Surv*, vol. 53, no. 3, Jun. 2020, doi: 10.1145/3381028.

[38] S. Seo, "A Review and Comparison of Methods for Detecting Outliers in Univariate Data Sets," University of Pittsburgh, 2006.

[39] D. Chakrabarty, "Model Describing Central Tendency of Data," *Int J Adv Res Sci Eng Technol*, vol. 8, no. 9, 2021, [Online]. Available: www.ijarset.com

[40] D. Salvatore, D. Reagle, and M.-H. New York Chicago San Francisco Lisbon London Madrid Mexico City Milan New Delhi San Juan Seoul Singapore Sydney Toronto, "Theory and Problems of Statistics and Econometrics Second Edition Schaum's Outline Series," 2002, doi:10.1036/0071395687.

[41] T. Budiwati, A. Budiyono, W. Setyawati, and A. Indrawati, "Analisis Korelasi Pearson Untuk Unsur-Unsur Kimia Air Hujan Di Bandung," *Jurnal Sains Dirgantara*, vol. 7, 2010.

[42] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Comput Sci*, vol. 7, pp. 1–24, 2021, doi: 10.7717/PEERJ-CS.623.

[43] Z. Yu, M. Guindani, S. F. Grieco, L. Chen, T. C. Holmes, and X. Xu, "Beyond t test and ANOVA: applications of mixed-effects models for more rigorous statistical analysis in neuroscience research," *Neuron*, vol. 110, no. 1. Cell Press, pp. 21–35, Jan. 05, 2022. doi:10.1016/j.neuron.2021.10.030.

[44] Q. Liu and L. Wang, "t-Test and ANOVA for data with ceiling and/or floor effects," *Behav Res Methods*, vol. 53, no. 1, pp. 264–277, Feb. 2021, doi: 10.3758/s13428-020-01407-2.

[45] D. van den Bergh, E.-J. Wagenmakers, and F. Aust, "Bayesian Repeated-Measures ANOVA: An Updated Methodology Implemented in JASP," Jun. 2022, doi: 10.31234/osf.io/fb8zn.

[46] S. E. Maxwell and H. D. Delaney, *Designing experiments and analyzing data : a model comparison perspective*. Lawrence Erlbaum Associates, 2004.
[47] T.-S. Lim and W.-Y. Loh, "A comparison of tests of equality of variances 1," 1996.
[48] G. Gümüş and F. Balcı, "Working memory for time intervals: Another manifestation of the central tendency effect," *Psychon Bull Rev*, Jun. 2023, doi: 10.3758/s13423-023-02324-z.
[49] T. N. Cason, D. Friedman, and E. Hopkins, "An experimental investigation of price dispersion and cycles," *Journal of Political Economy*, vol. 129, no. 3, pp. 789–841, Mar. 2021, doi:10.1086/712445.
[50] A. Qamar, Z. Anwar, H. Ali, S. Imran, R. Shaukat, and M. Mujtaba Abbas, "Experimental investigation of dispersion stability and thermophysical properties of ZnO/DIW nanofluids for heat transfer applications," *Alexandria Engineering Journal*, vol. 61, no. 5, pp. 4011–4026, May 2022, doi: 10.1016/j.aej.2021.09.028.
[51] A. M. Kozlowski *et al.*, "Heparin in Acid and Alkaline Environments—A Study of the Correlations between Hydrodynamic Properties and Desulphation," *Polysaccharides*, vol. 4, no. 2, pp. 88–98, Mar. 2023, doi: 10.3390/polysaccharides4020007.
[52] C. Qiang and Y. Deng, "A new correlation coefficient of mass function in evidence theory and its application in fault diagnosis," *Applied Intelligence*, vol. 52, no. 7, pp. 7832–7842, May 2022, doi:10.1007/s10489-021-02797-2.
[53] S. Li and C. Wei, "Hesitant fuzzy linguistic correlation coefficient and its applications in group decision making," *International Journal of Fuzzy Systems*, vol. 22, no. 6, pp. 1748–1759, Sep. 2020, doi:10.1007/s40815-020-00876-z.
[54] X. Zhang, Z. Qu, Z. Tang, and M. Iqbal, "pH-temperature coupled regulation for promoted nanofluidic osmotic energy conversion," *Desalination*, vol. 572, Mar. 2024, doi: 10.1016/j.desal.2023.117131.
[55] D. Liu *et al.*, "Self-Assembly Study of Block Copolypeptoids in Response to pH and Temperature Stimulation," *Polymers (Basel)*, vol. 16, no. 8, Apr. 2024, doi: 10.3390/polym16081082.
[56] A. S. Al-Fahal, A. S. Ahmed, A. K. Mohammed, and W. S. Mohammed-Ali, "Evaluation of Rainwater Harvesting Systems for Drinking Water Quality in Iraq," *International Journal of Design and Nature and Ecodynamics*, vol. 19, no. 1, pp. 111–117, Feb. 2024, doi:10.18280/ijdne.190113.
[57] J. Yang, S. Rahardja, and P. Fränti, "Mean-shift outlier detection and filtering," *Pattern Recognit*, vol. 115, Jul. 2021, doi:10.1016/j.patcog.2021.107874.
[58] H. Schwarzmeier *et al.*, "Theranostic markers for personalized therapy of spider phobia: Methods of a bicentric external cross-validation machine learning approach," *Int J Methods Psychiatr Res*, vol. 29, no. 2, Jun. 2020, doi: 10.1002/mpr.1812.
[59] M. J. Hamayel and A. Y. Owda, "A Novel Cryptocurrency Price Prediction Model Using GRU, LSTM and bi-LSTM Machine Learning Algorithms," *AI*, vol. 2, no. 4, pp. 477–496, Oct. 2021, doi:10.3390/ai2040030.
[60] A. Ramadhan *et al.*, "Real-Time Measurement Analysis of Rainwater Quality Based on the Internet of Things," in *2023 29th International Conference on Telecommunications (ICT)*, 2023, pp. 1–6. doi:10.1109/ICT60153.2023.10374038.
[61] H. Li, Z. Wan, and H. He, "Real-Time Residential Demand Response," *IEEE Trans Smart Grid*, vol. 11, no. 5, pp. 4144–4154, 2020, doi: 10.1109/TSG.2020.2978061.
[62] W. D. Xu, M. J. Burns, F. Cherqui, and T. D. Fletcher, "Enhancing stormwater control measures using real-time control technology: a review," *Urban Water J*, vol. 18, no. 2, pp. 101–114, Feb. 2021, doi:10.1080/1573062X.2020.1857797.
[63] M. Sarrab, S. Pulparambil, and M. Awadalla, "Development of an IoT based real-time traffic monitoring system for city governance," *Glob Transit*, vol. 2, pp. 230–245, Jan. 2020, doi:10.1016/j.glt.2020.09.004.