# IJASEIT Artikel

*by* Sri Listia Rosa

International Journal on Advanced Science Engineering Information Technology

Vol.12 (2022) No. 5
ISSN: 2088-5334

# Identifying Water Pollution Sources Using Real-Time Monitoring and IoT

Sri Listia Rosa [a], Evizal Abdul Kadir [a,*], Apri Siswanto [a], Mahmod Othmand [b], Hanita Daud [b]

[a] *Department of Informatics Engineering, Universitas Islam Riau, Marpoyan, Pekanbaru, Riau, 28284, Indonesia*
[b] *Department of Fundamental and Applied Science, Universiti Teknologi Petronas, Seri Iskandar, Perak, 32610, Malaysia*

*Corresponding author: *evizal@eng.uir.ac.id*

***Abstract*— Water is a natural resource essential for basic human life; however, water pollution is deteriorating in major water sources, such as rivers, seas, and lakes. This study evaluated and identified specific pollution sources owing to numerous industrial and other potential sources of pollution along the enormous length of a Siak river. A water pollution detection system was installed and deployed at river measurement stations having the potential to pollute, particularly near industrial sites releasing chemicals and wastewater. Data obtained from the system was analyzed using an algorithm to detect and assess any abnormal behavior change in the data over time. Six detection systems were deployed around the river, primarily in residential and industrial areas. As the studied river is one of the deepest in Indonesia, this research focused only on analyzing and identifying the sources of polluted water around Pekanbaru where many people, including water supply companies, utilize the river water. Water pollution sources were identified at sensor nodes two and four, which indicated through abnormal data that various types of material were present in the river and were detected using the sensors system. Several processes are required to improve the location data accuracy, e.g., improving the algorithm using training data, performing several iterations, increasing data from the sensor, and repeating the process many times to establish the source coordinate, as shown in the results.**

***Keywords*— Water pollution; source identification; sensor; algorithm; smart system; Siak river.**

*Manuscript received 29 Sep. 2021; revised 18 Apr. 2022; accepted 6 Jun. 2022. Date of publication 31 Oct. 2022.*


## I. Introduction

Water is a vital necessity for all living things in the world, notably human life. Water also serves as the primary environmental preservation and sustainability of the earth. Water is a natural resource found in the sea, river, lake, or underground. In recent decades, due to human development and exploration, many water resources got polluted and contaminated by chemical or environmental damage that then impacted the water resources. Many attempts have been made to protect environmental and water resources. However, certain areas or regions have difficulties due to authority or local government rule and regulation.

Furthermore, many people are not accountable for environmental damage caused by cutting trees and illegal exploration, resulting in environmental pollution. Indonesia is one of the tropical countries in the Southeast Asia Region that is rich in tropical trees and high-intensity of rainfall. As a tropical country, Indonesia boasts a plethora of rivers and lakes, especially in Riau Province, which has five big rivers and some other small or sub-rivers scattered around the region. River water is one of the main sources of water supply in most houses and industries in Indonesia, especially the community that lives surrounding the river. The river is a part of their daily life, and most of their activities use the river, such as cleaning and washing, showering, fishing, transportation, etc. Additionally, most of the activities on the river then water become potentially contaminated and polluted by many sources of pollution, including household waste and a dirty environment.

The approach for water source contamination and identification of water distribution system in a model of linear trees with a programming algorithm [1]. Data training and validation are done by an algorithm linear programming formulation, the rule, and structure clarification to discover the source of pollutant water contamination and then identify the location or coordinate. Introduces a new sensor based on a combination of interdigital electromagnetic and planar meander for water quality monitoring and contamination level in a water source. A Series of experiments were conducted to determine the features of a sensor. The two nitrates form, such as ammonium and sodium, were mixed in numerous times with a calculated ratio to examine the sensor's response in the detection of contamination.

Furthermore, water samples were obtained from various sources to evaluate and ensure that the sample was suitable for analysis in a different kind of water. The final results were compared and analyzed using the Nuclear Magnetic Resonance technique [2]. As discussed, historically accurate detection data of groundwater contamination and pollution is required to monitor and detect the impact of considerable environmental damage. To assemble such appropriate data and method to analyze with the maximum spatial distribution and segmented elaboration of the historical water pollution data. The configuration of the water monitoring system and some installation points and samplings are required and should be carefully considered as the fundamental parameters in optimization. In the New Moscow region, metaheuristic test approaches for optimization of monitoring procedure and multi factors assessment of water quality detection were used [3].

Using discriminant analysis, factor analysis, principal component analysis, and cluster analysis (CA) in multivariate statistical techniques were employed to evaluate the spatial and temporal of river basins and then interpret large datasheets to identify water content and supply [4]. The hierarchy of CA splits and breaks the data into several months to achieve high accuracy data in analysis and the evaluation of complex data sets from source information to monitor the water quality in the river and manage good water sources. In recent years, water pollution caused by soil and land damage has become increasingly significant, owing to industrial development and a massive urban population [5]. Heavy metal pollution in the soil has become widespread because of its strong chemical toxicity. High chemical concentration content in the soil directly affects nutrients and soil function and reduces the biological activities in the soil. A method of Long Short-Term Memory (LSTM) network model applied to inverting four major water parameters, including dissolved oxygen (DO), Pounds Hydrogenic (PH), Ammonia–Nitrogen (NH3–H), and Chemical Oxygen Demand (CODMn). In long-term dynamic monitoring of freshwater resources, LSTM was employed to analyze the quality based on the four major parameters mentioned [6]. The model has exhibited good performance in assessing water quality with the Mean Relative Error and coefficient of determination with relative Root Mean Square Error (rRMSE). Water quality is one of the most important issues and critical indicators in the environment and living things; the pollution makes contamination accidentally or by human activities. The discussion on the assessment and criticism of the water pollution inapplicability of various technologies applied in real-time monitoring in water pollution and addressing those issues have high attention and have been reported in a test bed in real-life scenarios. in addition, the sensor's performance in detection sensing range, limit of detection, and accuracy are given main attention and priority to verify the detection results [7], [8].

Natural water is one of the basic needs for human life and other living things on earth. In recent decades, water pollution has critically damaged the source due to urbanization and other aspects of human activities. A water quality monitoring system is needed at water sources such as a river, lake, or other sources to know the water quality status. In addition, a real-time monitoring system is required more urgently by placing sensors on the source of water supply to achieve and update information. A method Poisson distribution algorithm model was used to simulate the distribution and change of water demand in residential areas than to find out the location of the source of water pollution [9]. Using a Water Distribution Network (WDN) in the identification and analysis of water contamination included two types of WDN in the analysis used in problem analysis: transient state conditions and steady. The use of WDN is to achieve continuous water contamination, as presented. An approach was applied to estimate the magnitude of concentration and the sources of water contamination. A set of simulations and measurements by applying methods in superposition as a model to embed and relate the contamination distribution in polluted water. Environmental pollution caused by industrial operations especially waste material emitted from processing contains chemicals and toxins that are dangerous to human life and the environment [10]. The use of a multifunctional miniaturized to monitor water quality monitoring system (WQMS) consists of continuous water quality monitoring and simultaneous use of a wireless communication system. The electrodes sensor integrated with polydimethylsiloxane flow channels was applied in the system as a basic compound sensor. The multi-sensor was employed be able to identify water concentration parameters concurrently in the measurement of water temperature, pH, copper ion concentration, electrical conductivity, and other sensitive material detected. A Long Range (LoRa) implement in this monitoring system to achieve continuous monitoring of detected data, the advantages of using WQMS as multifunctional monitoring, such as small in size, low cost, continuous sampling, easy maintenance, and long-term monitoring up to many days [11].

A machine learning algorithm to identify water supply pollution sources [12], using a parallel processing system. The algorithm combining of Artificial Neural Networks (ANN) to classify the water pollution sources by using analysis of Random Forests regression to determine the significance of variables in a water contamination event, for example, start and end time as well as the concentration of the chemical contaminant. Performing Monte Carlo hydraulic simulations and water quality in parallel processing, recording all data sensed by all the sensors placed in the water source within water supply network them selecting the most potential polluted source based on the selection of nodes in the sensor network by using machine learning. Development of a new inversion of the Colored Dissolved Organic Matter (CDOM) algorithm by using advantages of the algorithm of sparse learning as representative and also well-known as Least Absolute Shrinkage and Selection Operator (LASSO), were implemented in order to classify the optimal bands of arithmetic for the CDOM inversion and estimation the parameters model in a globally as well as robustness manner then method of statistics-based.

Furthermore, two stages of the framework in inversion are discussed for improving the stability of LASSO to address inadequate situ samples circumstance. With this framework, the two different schemes in the application cover the band of arithmetic in terms of information propagation between the two stages as the application of the two algorithms [13]. A method of coupling machine learning applied in solving water contamination problems in water distribution network supply includes the determination of the exact source of water contamination start and end times of an event for every node individually. Two algorithmic in different frameworks were constructed. Both algorithms maximize the Random Forest algorithm for classification of the main source of water contamination in a node candidate, while another one of the frameworks used the stochastic fireworks optimization to determine the water contamination start and end time [14], [15].

Method to identify water supply contamination source in a network based on an algorithm of random forest classifying used in this discussion [16]. The algorithm was tested in two kinds of different water benchmark distribution networks with different sensor placement locations to achieve accurate data collection. In each network, consider the number of contaminated amounts in random scenarios and random simulation parameters of water quality in time series. The pollution source location data is recorded in many types of data such as pollution stand and end time, minimum and maximum chemical contamination quantity, and how big the area of the contaminated water source. The installation of the sensor was designed and investigated accordingly to obtain robustness in the data collection and verification of every single data to find the exact location of the water pollution source. Identification of water source contamination used Genetic Algorithm (GA) as discussed, the first method to obtain the rough source of water contamination by release analyzing time to the double peak typical phenomenon in region and then presenting the rough position of water source by using Lagrange tracing scheme [17]. Calculation and measurement results show the correlation of the concentration at water contamination measurement point, high resolution and mass optimization model apply to improve the results and repeat several times alternatively. Another discussion on the use of GA to identify water pollutant sources in river water supply by implementing intelligent and optimization methods to overcome the weakness of accuracy in location identification. Further, the Basic Genetic Algorithm (BGA) initiative will be established to continue and assist GA in identifying and finding a source of contamination. Data processing and optimization tool with a one-dimensional formula applied as a solution in analysis of unsteady water quality equation to achieve the results [18], [19].

Pollution management and surface water incident are emergencies to protect the environment. The method to identify and localize the source of water pollution is a crucial and urgent matter to do in quick response to protect the environment and water supply. A multi-sensor water quality monitoring and management system for water pollution sources identification and localization based on Wireless Sensor Networks (WSNs) system. The multi-sensor system is embedded into two types of sensor nodes to send the information to WSNs sink (gateway). An algorithm called utilizing Radial Basis Function Network (RBFN) applies in this discussion to check the factor of the pollution [20]–[22]. A coarse method of the location based on the land contour is used for advanced water pollution identification and localization for better diffusion models and detailed analysis of source contaminated water. The simulation and experiment model results compare to check the accuracy and determination of the final contaminated location for further action. The detection of water source pollution using a sensor network is challenging on this issue, where the sensor node has to place carefully in a correct location to obtain optimum data detection [23]–[25]. While both dynamic and static pollution sources are related to the binary hypotheses, specific inspection and test statistics are given high attention and applied in this case. Experiment and simulation results of detection give test feedback to illustrate the source of water pollution, contamination, and large contaminated areas. The number of nodes and sample data is another high attention for this experiment and simulation to observe several times to achieve high accuracy detection results [26]–[28].

## II. Method and Data Collection

Siak River is located in the capital of Riau Province, which is in the city of Pekanbaru. The river plays a significant role for the community and the Riau economy since the Siak River is a transportation route connecting Pekanbaru to other cities besides land transportation.

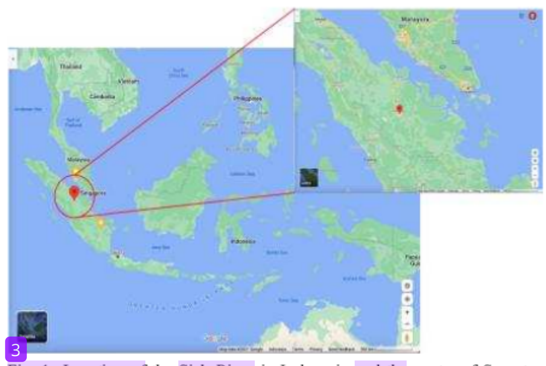


Fig. 1 Location of the Siak River in Indonesia and the center of Sumatra Island

Compared to other modes of transportation, using the river as the primary means of transportation is more efficient. The total length of rivers is more than 200 km from the upstream to the downstream, which is the end of the sea. Along the river, which is quite long, there are many obstacles and potential contamination of the river's water. Thus detecting and monitoring river water is an essential and urgent task because the river water has been badly contaminated. Furthermore, the operation of several industries and factories surrounding and along the river contributes to the worsening water pollution, besides waste from the community and other trash around the river. Figure 1 shows the map of Indonesia and Riau Province in the center of Sumatra Island, which is the place of the Siak River for this study and research location.

A preliminary survey indicates that the current condition of the river is quite bad and highly contaminated by various materials due to several industrial operations along the river and local communities and developing residential areas. The river connects to many sub rivers or branches, each contributing to water pollution. Some sub-rivers are ended upstream where villages and communities live without care for the environment. Furthermore, the material that causes water pollution comes in many kinds and from various sources. Thus, a smart system with multi-parameters of a sensor for detecting material and chemical content is required to achieve valid and accurate data. Several sample data collected will affect the accuracy of data analysis as much as the possible number of sample data required to accomplish an accurate final decision of the water pollution source's location.

Figure 2 illustrates the Siak River in the heart of Pekanbaru City with specific river roots that have numerous branches crossing the city of Pekanbaru. Siak River divides the city of Pekanbaru between the north and south parts, as the river crossing in the center of the city greatly impacts the city's daily activities and communities. Thus, potential river water gets contaminated from the pollution is very high the busy river.

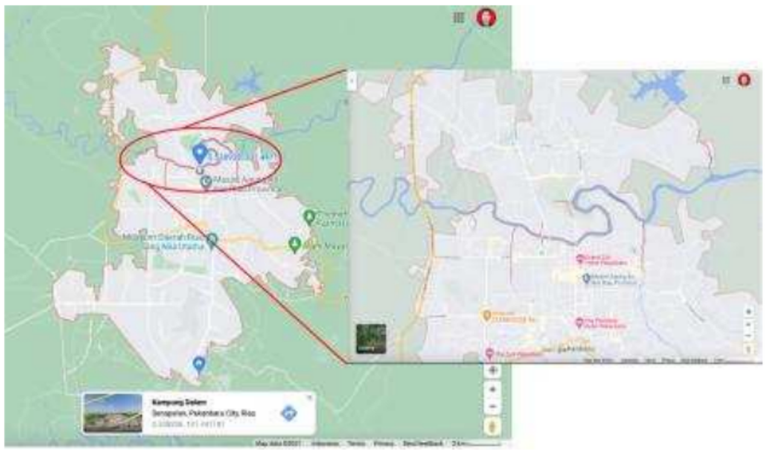


Fig. 2 Siak River in Riau Province divides the Pekanbaru City

### A. *Water pollution sensing system*

In normal conditions, water has many parameters that can be measured by observation or using measurement tools such as sensors or probes. While the water gets contaminated, the value of water parameters varies according to the quantity of material or chemical polluting the water, how much the substance can be measured, and how the changes are checked. Several water values can be measured in either normal conditions or when contaminated. Table 1 displays the common parameters in most of the water. The most common parameters in the water are temperature and pH. These two parameters, as the basic, have to measure and are very closely connected to the quality of water. Additional parameters to support and measure accuracies are water dissolved oxygen (DO) and the water's electrical conductivity (EC) content. All of these parameters have a strong justification to determine how water gets polluted and the type of chemical-based or intensity measured. Multi sensing system proposed and designed in this study and experiment to collect water parameters data, sensor sensitivity, and accuracy are taking attention in determining the type of sensors used. Water temperature, pH, Dissolve Oxygen (DO), and EC are the four fundamental parameters employed to determine water quality, as shown in Table 1. There are relationships between those sensor parameters, and considering all these data based on sensors detection will improve decision-making in a final determination of source location identification.

TABLE I
RIVER WATER POLLUTION SENSING PARAMETERS

| Parameter | Range | Accuracy | Sensor Type |
|---|---|---|---|
| Temperature | 0 to 60 ° C | ± 0.5 ° C | Thermistor |
| pH | 0 to 14 | ± 0.1 | Glass Electrode |
| DO | 0 to 20 mg/L | ± 0.5 mg/L | Polarography |
| Electrical Conductivity | 0 to 50 | ± 0.5 | Conductivity Measurement |

In this research, a set of sensing systems was designed and developed to detect river water parameters; many sensors have been used according to the water parameters. A designed block diagram of a river water sensing system, as illustrated in figure 3, where temperature, pH, DO, and EC sensors were installed. The signal from the sensor-based detection commonly has a noise; as a result, a signal conditioning procedure is required before going to the Main Control Unit (MCU), while Radio Frequency and antenna system are utilized to transmit the data to the Wireless Sensor Network (WSN) node station.

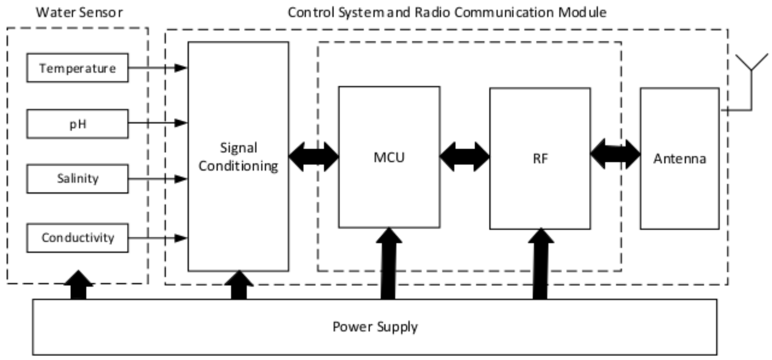



Fig. 3 Block diagram of the sensor system to detect water pollution parameters

The proposed sensor system is designed with real-time monitoring for all the data detected by the water sensor. Multiple nodes will be able to gather many data from multiple locations as sensor nodes deploy at the measurement points. To enable real-time monitoring, a high-speed data transfer system was built, as indicated in Figure 4, the block diagram of the communication system to collect data from the nodes and transfer it to a backend system. While 4G cellular communication technology was utilized as media to transfer data, most of the nodes sensing locations are covered by cellular signals. Data from sensor nodes is collected by the WSN gateway system, which is subsequently sent to the data center (backend system) for analysis.

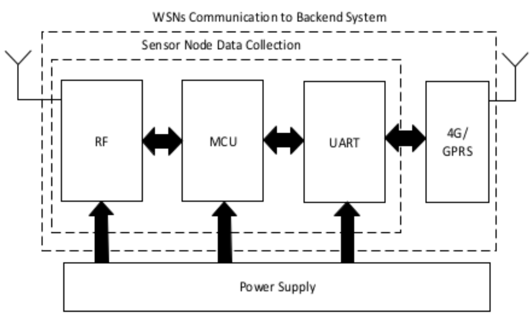



Fig. 4 Communication system block diagram of the sensor network

## B. *Real-time detection and monitoring system*

Due to the river's length to monitor, many sensor nodes are installed at the potential areas where water gets contaminated to get precise data and location. Figure 5 depicts a sensor network for detecting river water pollution that connects several sensor nodes deploy on-site. Detected water parameters data is temporarily stored in the buffer memory in the WSN sink node before forwarding it to the data center utilizing cellular data communication or Global System for Mobile Communication (GSM) modem. All the data are analyzed in a central unit at the monitoring system to assess pollution. Real-time communication between sensor nodes and the server at the backend system is vital to confirm all the data transfers as fast as possible and avoid losses. Sensor nodes place at the actual site, which is surrounding the river that may have much interference with the potential to disrupt the communication system; a sink node is a node cluster for many sensor nodes implemented and configured in the system to reduce the data loss and interrupted communication system. According to an early survey, the distance of the river to be monitored in the city of Pekanbaru is about 30 km. The current setup deployed six sensors node to detect the abnormalities in the river water. Among the six nodes, one node serves as a sink or gateway to communicate to the backend system, while the other five send the data to the sink node.

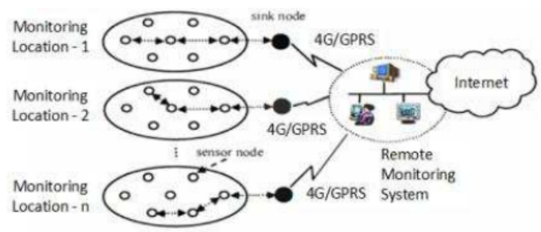



Fig. 5 Communication link of a sensor node to the data gateway system

The designed water pollution sensing system has been fabricated, and initial testing has been conducted in a laboratory to gather preliminary data and calibrate all the sensors to the optimum sense value. Figure 6 illustrates a set of fabricated sensing systems set up in a container consisting of water to check the sensor's functionality. Water contamination is simulated to various values, for example, temperature sensor by heating the water then back to the normal temperature and cooling down the water by putting ice or other material. Similar concepts have been done for other parameters to test and ensure the sensor is working and detection values are changing.

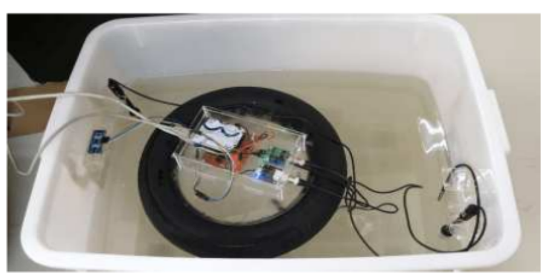


Fig. 6 Initial testing of water sensor for calibration

As mentioned earlier, several sensors are employed to detect actual water values in the river in a prototype sensor system constructed based on actual conditions in river water. An actual prototype of the sensor system is shown in figure 7, while figure 7(a) displays testing conducted in a river with a floating tube, and figure 7(b) shows a solar panel installed in the system as a power supply for sensors and other components. All the detected data sending to the data center through cellular communication as a GSM antenna is installed at the side of the box.

Fig. 7 Prototype of the water pollution sensing system; (a) front side, which is the control unit, and (b) solar panel as a power supply to the sensor system

*C. Data collection and analysis*

The Siak River is located in the center of Pekanbaru City in Riau Province, at the coordinate 0°32'27.9"N 101°26'15.8"E. The river divides the city into two parts, north and south. To collect water quality data, evaluate the pollution concentration in the water, and find the location, the scenarios that sketch the river as displayed in figure 8, the river map refers to the early map as shown in figure 2, which shows how actual river with several tributaries. The river consists of six tributaries or sub-rivers, each of which has the potential to contribute and contaminate river water since the branches originate from residential and industrial areas. While the main body of the river with wide by about 100-150 meters and has a depth starting from 20 meters to more than 100 meters, it is one of the deepest rivers in Indonesia. Deployment of sensor nodes must be carefully considered because the river is deep and has fast river flow, especially during rain and flooding.

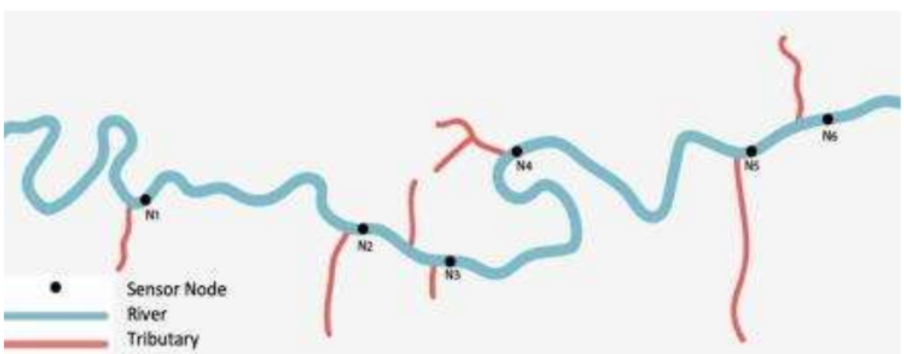



Fig. 8 River with sensor nodes installed at the measurement point

Water pollution data were obtained from the sensor nodes deployed to the field. There are six sensor nodes as monitoring points installed, with each node positioned at a specific location based on a preliminary survey that indicated a high potential for contamination and polluted river water. Table 2 presents the detailed coordinate of those sensor node locations, with all the nodes installed after the river branch or tributary to obtain accurate water pollution data, as mentioned in the early placement of sensors after the branch, then will get detailed water pollution data. All the detected data from sensor nodes is sent to the backend system through the WSN sink gateway to continue the next procedure, filtering and analyzing the data utilizing an artificial intelligence algorithm to accomplish and identify the exact location of sources of water pollution.

TABLE III
SENSOR NODES DEPLOYMENT LOCATION

| Sensor Node | Deployment Coordinate |
|---|---|
| 1 | 0°32'57.1"N 101°24'59.4"E |
| 2 | 0°32'40.8"N 101°25'50.9"E |
| 3 | 0°32'27.9"N 101°26'15.8"E |
| 4 | 0°33'09.5"N 101°26'50.5"E |
| 5 | 0°32'48.2"N 101°28'09.7"E |
| 6 | 0°33'34.9"N 101°29'25.3"E |

Raw data is obtained from measuring river water parameters by a sensor deployed at the monitoring point and filtering method done at the node station. The identity of source's water pollution location began with the initialization of all the sensor nodes, followed by the measurement of all the detection data stored in the internal memory of the microcontroller unit. All the valid data after the filtering process is sent to the data center at the backend system to begin the next step, which is running analysis, by utilizing an advanced system to conduct analysis using artificial intelligence algorithm for measurement data and data training. Once the data has been analyzed, then check the error; if the value is more than the threshold, return to the analysis for optimization; if it is less than the threshold, find the source location of water pollution and close by the end of execution. Figure 9 indicates a complete flow of the method in identifying pollution source locations.

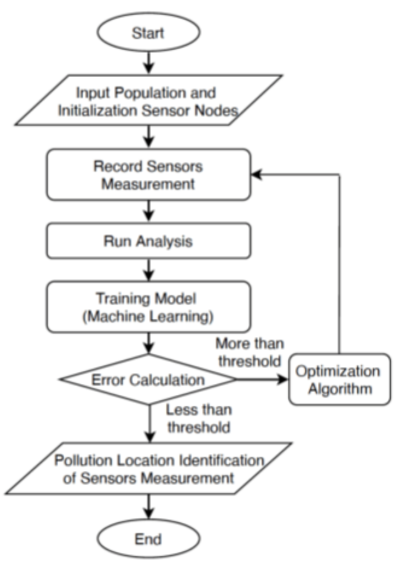



Fig. 9 Flowchart of measurement and optimization method

In a dynamic river water scenario, the contamination follows the water flow; for example, in a river, water flow from water upstream to downstream, the flow of water is along the *y* direction with the flow rate is *u, and* there is an instantaneous source at $(\zeta, \eta)$, the mass of the pollutant is *M*, and the diffusion coefficient is *D*. The concentration at $(x_i, y_i)$ is as equation (1) [29].

$$C(x_i, y_i, t) = \frac{M}{4\pi D} \exp\left\{\frac{1}{4t\,D}\left[(y_i - \zeta - ut)^2 + (x_i - \eta)^2\right]\right\} \quad (1)$$

In this situation, the diffusion is affected by the flow, the pollution on one side of the sources, and the concentration field, as shown in figure 10.

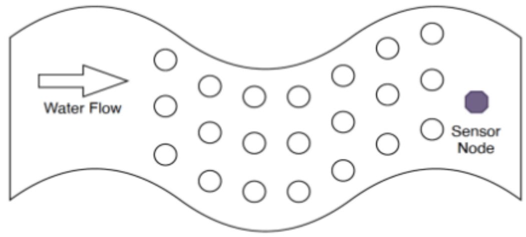



Fig. 10 River water pollution source at the monitoring location

The current detection approach uses hypothesis and testing [23] and assumes that the analysis defines the distribution of noise *e* is. However, in the actual measurement environment, $\delta$ is normally undefined. Thus, the analysis of this research work utilized the specific technique as follows.

Investigate the mean value $\mu_k$ of the monitoring data:

$$\tilde{C}(x_i, y_i, t_{e+k}) - \tilde{C}(x_i, y_i, t_e), k = 1,2, \dots, L.$$

The binary hypotheses are given by,

$$H_0^{(1)}: \ \mu_k = 0,$$

$$H_1^{(1)}: \ \mu_k \neq 0.$$

The test statistic is,

$$G_1 = \frac{\overline{C_k}}{S_k\sqrt{L}}, \quad (2)$$

Where $\overline{C_k} = (1/L)\sum_{K=1}^{L}(\tilde{C}(x_i, y_i, t_{l+k}) - \tilde{C}(x_i, y_i, t_l))$ and

$$S_k^2 = \left(\frac{1}{L-1}\right)\left(\sum_{k=1}^{L}\tilde{C}(x_i, y_i, t_{l+k}) - \tilde{C}(x_i, y_i, t_l) - \overline{C_k}\right)^2 (3)$$

Let, the number of concentrations in variation satisfy by,

$$|G_1| \geq t_{\frac{\alpha}{2}}(L-1), \quad (4)$$

That is, when $P\left(|G_1| \geq t_{\frac{\alpha}{2}}(L-1)\right) = \alpha$, reject $H_0^{(1)}$ and its deduction that sensor nodes have detected the source of water pollution. Please take note that $\alpha$ is given by a significance level and $t_{\frac{\alpha}{2}}$ is the $\frac{\alpha}{2}$ quantile of *t*-distribution.

In the analysis and measurement, some basic values of the significant level [30],[31]. The detection results are based on sensing node data to analyze the pollution location. Localization of water pollution identification based on algorithms applied in the data analysis. The mathematical model of localization algorithms refers to the diffusion models, as follows [21]. If $C(x', y', t_l, \zeta, \eta)$ is concentration refer to the theoretical at the node $(x_i, y_i)$ provide by the diffusion model, $\tilde{C}\ (x', y', t_l) = C(x', y', t_l, \zeta, \eta) + \theta$ is the value of corresponding monitoring with the noise $\theta$, and $f_j(\zeta, \eta), j = 1,2,3, \dots$, represent the related constraint of $(\zeta, \eta)$, by the assumptions of whether then the distributions of measured noise determine or not and whether distribution is normal distributions or not, there are many ways to estimate the method that can be provided. Least Squares is the common use method because of the advantage that methods simple and can be applied in the practical applications, while the distribution of $\theta$ is undetermined as follows:

$$\min \sum_{i=1}^{n}\left[\tilde{C}_i - C\left(x_i, y_{i,\zeta,\eta}\right)\right]^2 \quad (5)$$
$$\text{s.t } f_j\,(\zeta,\eta), j = 1,2,3,\ldots$$

Refer to the figure 10, a system with rectangular coordinate with the water direction along the river flow is *y*, rectangular coordinate of the system, the symmetry axis $y = \eta$. The source location of the diffusion by $(\zeta, \eta)$ is on the axis of symmetry. While, there are two nodes $(x_1, y_1)$ and $(x_1, y_2)$ with the same value of x-coordinate on the same contour, it can be calculated as:

$$\eta = \frac{y_1+y_2}{2} \quad (6)$$

While there is superimposed effect in the concentration, the point is far from the water flow on the contour are still on a circle. By choosing any *n* points marked as $(x_1, y_1)$, $(x_2, y_2)$, $(x_3, y_3)$, …, $(x_n, y_n)$ to determine and locate the source of pollution; then can determine as.

$$\begin{gathered}(x_1 - \zeta)^2 + (y_1 - \eta)^2 = r^2,\\ (x_2 - \zeta)^2 + (y_2 - \eta)^2 = r^2,\\ \vdots \\ (x_n - \zeta)^2 + (y_n - \eta)^2 = r^2,\end{gathered} \quad (7)$$

Where *r* is the circle radius. And it can be written as,

$$\begin{gathered}x_1^2 - x_n^2 + y_1^2 - y_n^2 - 2(x_1 - x_n)\hat{\zeta} - 2(y_1 - y_n)\eta = 0,\\ x_2^2 - x_n^2 + y_2^2 - y_n^2 - 2(x_2 - x_n)\hat{\zeta} - 2(y_2 - y_n)\eta = 0,\\ \vdots \\ x_{n-1}^2 - x_n^2 + y_{n-1}^2 - y_n^2 - 2(x_{n-1} - x_n)\hat{\zeta} - 2(y_{n-1} - y_n) = 0\end{gathered} \quad (8)$$

Then, it can obtain by,

$$\hat{\zeta} = \frac{1}{n-1}\sum_{i=1}^{n-1}\frac{(x_i^2 - x_n^2 + y_i^2 - y_n^2 - 2(y_{i} - y_n)\eta)}{2(x_{i} - x_n)} \quad (9)$$

and the amount of residual $\Lambda$ write by,

$$\Lambda = \sum_{i=1}^{n-1}\left|x_i^2 - x_n^2 + y_i^2 - y_n^2 - 2(x_i - x_n)\hat{\zeta} - 2(y_i - y_n)\eta\right| \quad (10)$$

Figure 10 shows the coordinate of rectangular system, if their movement of water flow along the river with direction *y*. There are several sensor nodes with coordinate value at the same unidirectional direction. Based on equation (7) ~ (10), the whole localization algorithm is as follows.

*1) Step 1.* Give a threshold < and let the counting marks be $l = 1$ and $j = 1$.

*2) Step 2.* At sampling time $t_l$, connect any two points $(x, y)$ and $(x'y')$ when the concentrations $C(x', y', t_l)$ and $C(x, y, t_l)$ satisfy $|C(x', y', t_l) - C(x, y, t_l)| \leq \beta$

*3) Step 3.* If the number of nodes connected more than 4, it could be deduced that the number of nodes in the same contour goes to step 4. Else, *l* adjusted to $l = l+1$ and go to step 2.

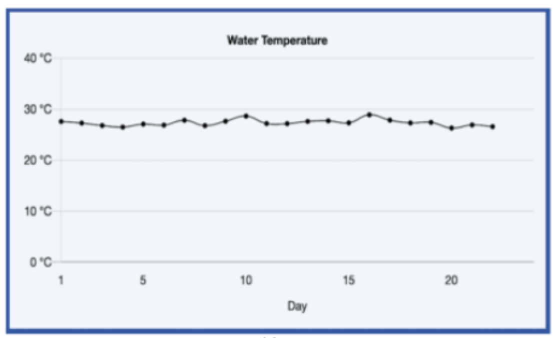
*4) Step 4.* If *N* is the number of sensor nodes on the same contour and obtained in step 3. The *N* node, if there are two nodes $(x_1,y_1)$ and $(x_1,y_2)$ by the same position of *x*, then estimation of $\eta$ able to calculate.

*5) Step 5.* By using the SL-n algorithm to obtain estimation; in the *N* point which on the similar contour, by choosing any points of *n* to obtain estimation as in (12). The $C_N^n$ result $\left(\hat{\zeta}^{(j)}, \Lambda^{(j)}\right), j = 1,2,3, \ldots, C_N^n$ , will be achieved.

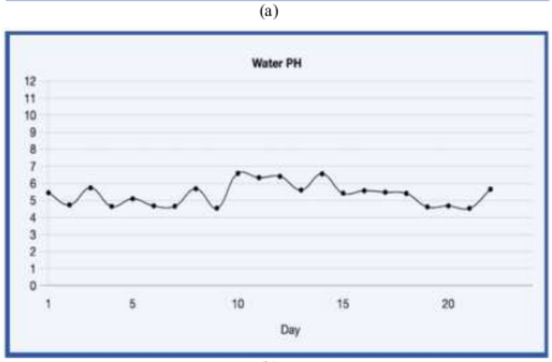
*6) Step 6.* Find the minimum number of values in $\{\Lambda^{(j)}\}$ and let the estimation of corresponding pollution location by of the minimum value of the estimation in ultimate of $\zeta$.

## III. Results and Discussion

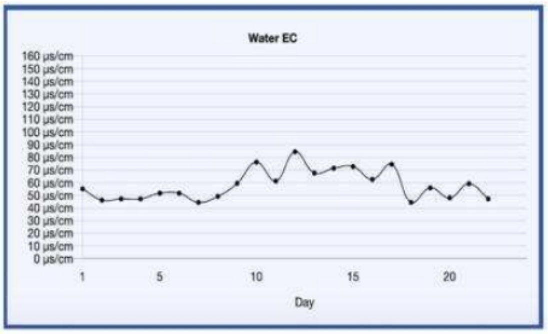
River water data detection and measurement results are evaluated at the central processing in the MCU then display the reading results in the graph to examine how the data behaves, as well as recorded to check the abnormality in the long-time study. Figure 11 shows the graph of measurement results, while figure 11(a) being a measurement graph of water temperature, figure 11(b) is measurement results of water pH, while figure 11(c) is a measurement of EC of water and figure 11(d) being measurement results of dissolved oxygen (DO) of water.



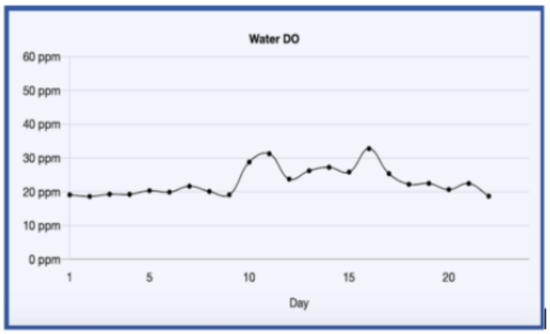

(a)




(b)




(c)




(d)

Fig. 11 Results of water pollution measurement; (a) water temperature, (b) water pH, (c) electrical conductivity (EC), and (d) water dissolve oxygen (DO)

To verify measurement data from the measurement, the displaying data range can be modified based on daily, weekly, monthly, and yearly scales. Furthermore, any abnormality in data displayed and recorded in the system would generate an alert to remind you of anything happening at the monitoring point. The upper and lower limit unit values of every measurement was calibrated based on standard values of river water properties. This calibration has been in preliminary testing in the laboratory as initial testing to calibrate all the sensors and to achieve high accuracy data reading, method of calibration by combining several techniques including manual and conventional measurement models.

The results of measurement and analysis to find and identify the cause of water pollution contamination based on raw data detected by sensors at the monitoring point is vary from one sensor node to the next; every sensor node has a coverage area, and water parameter reading may differ. The last node at the downstream location may have a high possibility of detecting contaminated data compared to 1the upstream of the river. Figure 12 depicts the pollution source findings based on data analysis from the sensor node, particularly nodes two and four, which indicate high contamination on all parameters and features of the detected and recorded data. The study's findings are compiled in a unit of time to conclude and determine whether or not the water is contaminated. In comparison, a real-time monitoring system displays sensors' most recent data reading for visual analysis and live monitoring.

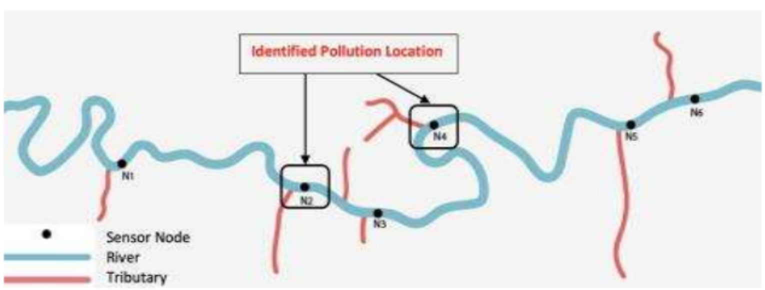



Fig. 12 Results of water pollution source location at the sensor node two and four.

According to the data training and detected from sensor nodes two and four in the identified measurement of water pollution, contamination is not permanent. It may take some time before water quality recovers when pollution sources are removed, for example, chemicals ejected from industries going into the river, then the material follows river flow then neutralized, or going into the last miles which end up in the sea. Many other situations, including data analysis and noise data sent by the sensor node, have been developed based on trial and testing.

## IV. Conclusion

A system for detecting and monitoring river water pollution was developed to obtain an automatic and intelligent system for discovering and identifying the sources of river water pollution. Six sensor nodes were deployed at the measurement points to capture live data from the river, and the sensors installed refer to the scenario of the river with the tributaries or branches that emerge naturally. Sensor nodes may detect and transfer the data to a system inside the node for temporary memory, and the backend system validates the data after filtering to record and show it onto the display for analysis. Data training and analysis with intelligence algorithms were conducted to accurately and correctly identify pollution sources. Two source locations were detected according to the sensor training data: nodes two and four, with record-high water contamination. Finally, the process was concluded by locking the system and location for further evaluation.

## Acknowledgment

The authors thank the Ministry of Education, Culture, Research, and Technology (KEMENBUDRISTEK) of Indonesia for funding this research project with contract number is 06/Kontrak/LPPM/4-2021, as well as Universitas Islam Riau, Indonesia as well Universiti Teknologi Petronas, Malaysia to support the facilities."


## References


[1] P. Ami and O. Avi, "Contamination Source Identification in Water Systems: A Hybrid Model Trees–Linear Programming Scheme," J. Water Resour. Plan. Manag., vol. 132, no. 4, pp. 263–273, Jul. 2006, doi: 10.1061/(ASCE)0733-9496(2006)132:4(263).

[2] M. A. M. Yunus and S. C. Mukhopadhyay, "Novel Planar Electromagnetic Sensors for Detection of Nitrates and Contamination in Natural Water Sources," IEEE Sens. J., vol. 11, no. 6, pp. 1440–1447, 2011, doi: 10.1109/JSEN.2010.2091953.

[3] E. Yudina et al., "Optimization of Water Quality Monitoring Networks Using Metaheuristic Approaches: Moscow Region Use Case," water, vol. 13, no. 7. 2021, doi: 10.3390/w13070888.

[4] J. Liu et al., "Water quality assessment and source identification of the Shuangji River (China) using multivariate statistical methods," PLoS One, vol. 16, no. 1, p. e0245525, Jan. 2021.

[5] H.-H. Jiang, L.-M. Cai, H.-H. Wen, and J. Luo, "Characterizing pollution and source identification of heavy metals in soils using geochemical baseline and PMF approach," Sci. Rep., vol. 10, no. 1, p. 6460, 2020, doi: 10.1038/s41598-020-63604-5.

[6] C. Qi, S. Huang, and X. Wang, "Monitoring Water Quality Parameters of Taihu Lake Based on Remote Sensing Images and LSTM-RNN," IEEE Access, vol. 8, pp. 188068–188081, 2020, doi: 10.1109/ACCESS.2020.3030878.

[7] I. Yaroshenko et al., "Real-Time Water Quality Monitoring with Chemical Sensors," Sensors, vol. 20, no. 12. 2020, doi: 10.3390/s20123432.

[8] N. S. Salikova, J. Rodrigo-Ilarri, K. K. Alimova, and M.-E. Rodrigo-Clavero, "Analysis of the Water Quality of the Ishim River within the Akmola Region (Kazakhstan) Using Hydrochemical Indicators," Water, vol. 13, no. 9. 2021, doi: 10.3390/w13091243.

[9] X. Yan, J. Gong, and Q. Wu, "Pollution source intelligent location algorithm in water quality sensor networks," Neural Comput. Appl., vol. 33, no. 1, pp. 209–222, 2021, doi: 10.1007/s00521-020-05000-8.

[10] O. S. Adedoja, Y. Hamam, B. Khalaf, and E. R. Sadiku, "Development of an Algorithm for the Estimation of Contamination Sources in a Water Distribution Network," IEEE Access, vol. 8, pp. 200412–200419, 2020, doi: 10.1109/ACCESS.2020.3035573.

[11] H.-C. Yu et al., "Development of Miniaturized Water Quality Monitoring System Using Wireless Communication," Sensors, vol. 19, no. 17. 2019, doi: 10.3390/s19173758.

[12] L. Grbčić, I. Lučin, L. Kranjčević, and S. Družeta, "A Machine Learning-based Algorithm for Water Network Contamination Source Localization," Sensors, vol. 20, no. 9. 2020, doi: 10.3390/s20092613.

[13] R. Zhang et al., "Developing New Colored Dissolved Organic Matter Retrieval Algorithms Based on Sparse Learning," IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., vol. 13, pp. 3478–3492, 2020, doi: 10.1109/JSTARS.2020.3003593.

[14] L. Grbčić, L. Kranjčević, and S. Družeta, "Machine Learning and Simulation-Optimization Coupling for Water Distribution Network Contamination Source Detection," Sensors, vol. 21, no. 4. 2021, doi: 10.3390/s21041157.

[15] Lukmana, H. Hidayata, L. Subehi, R. Dina, N. Mayasari, I. Melati, and Y. Sudriani, "Loads of Pollution to Lake Toba and Their Impacts", International Journal on Advanced Science Engineering Information Technology, vol. 11, no. 3, 2021, pp. 930-936.

[16] L. Grbčić, I. Lučin, L. Kranjčević, and S. Družeta, "Water supply network pollution source identification by random forest algorithm," J. Hydroinformatics, vol. 22, no. 6, pp. 1521–1535, Sep. 2020, doi: 10.2166/hydro.2020.042.

[17] S. Zhang and X. Xin, "Pollutant source identification model for water pollution incidents in small straight rivers based on genetic algorithm," Appl. Water Sci., vol. 7, no. 4, pp. 1955–1963, 2017, doi: 10.1007/s13201-015-0374-z.

[18] Y. Q. Long, C. Y. Wu, and J. P. Wang, "The Influence of Estimated Pollution Range on the Groundwater Pollution Source Identification Method Based on the Simple Genetic Algorithm," Appl. Mech. Mater., vol. 587–589, pp. 836–841, 2014, doi: 10.4028/www.scientific.net/AMM.587-589.836.

[19] I. G. A. P. Eryani and Nurhamidah, "Sedimentation Management Strategy in River Estuary for Control the Water Damage in Downstream of Ayung River", International Journal on Advanced Science Engineering Information Technology, vol. 10, no. 2, 2020, pp. 743-748.

[20] J. Ma, F. Meng, Y. Zhou, Y. Wang, and P. Shi, "Intelligent Water Pollution Source Identification and Localization in Wireless Sensor Networks," in 2018 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), 2018, pp. 1300–1305, doi: 10.1109/IMCEC.2018.8469723.

[21] E. A. Kadir, A. Siswanto, S. L. Rosa, A. Syukur, H. Irie, and M. Othman, "Smart Sensor Node of WSNs for River Water Pollution Monitoring System," in 2019 International Conference on Advanced Communication Technologies and Networking (CommNet), 2019, pp. 1–5, doi: 10.1109/COMMNET.2019.8742371.

[22] E. A. Kadir, A. Syukur, M. Othman, and B. Saad, "Smart Sensor Node of Wireless Sensor Networks (WSNs) for Remote River Water Pollution Monitoring System BT - Optimization Based Model Using Fuzzy and Other Statistical Techniques Towards Environmental Sustainability," S. A. A. Karim, E. A. Kadir, and A. H. Nasution, Eds. Singapore: Springer Singapore, 2020, pp. 29–40.

[23] X. Luo, J. Yang, and L. Chai, "Water pollution source detection in wireless sensor networks," in 2015 IEEE International Conference on Information and Automation, 2015, pp. 2311–2315, doi: 10.1109/ICInfA.2015.7279671.

[24] A. Delgadoa, F. Gila, J. Chullunquía, T. Valdivia, and C. Carbajal, "Water Quality Analysis in Mantaro River, Peru, Before and After the Tailing's Accident Using the Grey Clustering Method", International Journal on Advanced Science Engineering Information Technology, vol. 11, no. 3, 2021, pp. 917-922.

[25] E. A. Kadir, H. Irie, and S. L. Rosa, "River Water Pollution Monitoring using Multiple Sensor System of WSNs (Case: Siak River, Indonesia)," in 2019 6th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI), 2019, pp. 75–79, doi: 10.23919/EECSI48112.2019.8976991.

[26] O. Khutsoane, B. Isong, N. Gasela, and A. M. Abu-Mahfouz, "WaterGrid-Sense: A LoRa-Based Sensor Node for Industrial IoT Applications," IEEE Sens. J., vol. 20, no. 5, pp. 2722–2729, 2020, doi: 10.1109/JSEN.2019.2951345.

[27] S. L. Ullo and G. R. Sinha, "Advances in Smart Environment Monitoring Systems Using IoT and Sensors,"Sensors, vol. 20, no. 11. 2020, doi: 10.3390/s20113113.

[28] E. A. Kadir, H. Irie, S. L. Rosa, B. Saad, S. K. A. Rahim, and M. Othman, "Remote Monitoring of River Water Pollution Using Multiple Sensor System of WSNs and IoT BT - Sensor Networks and Signal Processing," 2021, pp. 99–113.

[29] Z. Z. Peng, T.X.Yang, X. J. Liang, and Z. S. Gu, Mathematical Model of Water Environment and Its Application. Beijing: Chemical Industry Press, 2007.

[30] A. R. John, Mathematical Statistics and Data Analysis, Third. New York, USA: Duxbury Press, 2006.

[31] C. E. R. Calderón, O. J. S. Parra, and S. C. V. Ayala, "Water Quality Prediction and Detection of the Vibrio Cholerae Bacteria", International Journal on Advanced Science Engineering Information Technology, vol. 11, no. 6, 2021, pp. 2369-2374.

IJASEIT Artikel

ORIGINALITY REPORT

| 8% | 7% | 3% | 0% |
|---|---|---|---|
| SIMILARITY INDEX | INTERNET SOURCES | PUBLICATIONS | STUDENT PAPERS |

PRIMARY SOURCES

| | | |
|---|---|---|
| 1 | ijaseit.insightsociety.org<br>Internet Source | 4% |
| 2 | www.hindawi.com<br>Internet Source | 1% |
| 3 | repository.uir.ac.id<br>Internet Source | 1% |
| 4 | Ruihao Zhang, Ruru Deng, Yingfei Liu, Yeheng Liang, Longhai Xiong, Bin Cao, Wenzhi Zhang. "Developing New Colored Dissolved Organic Matter Retrieval Algorithms Based on Sparse Learning", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2020<br>Publication | 1% |
| 5 | pubmed.ncbi.nlm.nih.gov<br>Internet Source | 1% |
| 6 | repository.ubaya.ac.id<br>Internet Source | 1% |

| | | | |
|---|---|---|---|
| Exclude quotes | On | Exclude matches | < 1% |
| Exclude bibliography | On | | |