Enhancing Water Quality Assessment in Indonesia Through Digital Image Processing and Machine Learning

Athiya Iffaty¹, Adinda Salsabila², Adis Aufa Rafiqhi³, Rivansyah Suhendra⁴*, Muhammad Yusuf⁵ and Novi Reandy Sasmita⁶

¹Department of Urban and Regional Planning, Universitas Diponegoro, Semarang, Indonesia; ²Department of Computer Science, Universitas Padjadjaran, Bandung 45363, Indonesia; ³Department of Civil Engineering, Faculty of Engineering, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia; ⁴Department of Information Technology, Faculty of Engineering, Universitas Teuku Umar, Aceh Barat 23681, Indonesia; ⁵Department of Pharmacy, STIKes Assyifa Aceh, Banda Aceh 23242, Indonesia; ⁶Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia

* Correspondence: rivansyahsuhendra@utu.ac.id

Abstract

Indonesia's diverse climate types, influenced by its unique geographical features, pose significant environmental challenges, including water quality issues related to turbidity and Total Dissolved Solids (TDS). Many Indonesians lack awareness of water quality, particularly turbidity, which can harbor harmful microorganisms. To address these challenges, this study employs digital image processing and machine learning, specifically Support Vector Machine (SVM) algorithms, for water quality assessment. A dataset of 80 water images, categorized into seven turbidity classes, is used to train and test the model. Results show a clear correlation between turbidity levels and TDS concentrations and pH values. The system accurately assesses water suitability for different sources, offering a user-friendly and cost-effective solution for water quality monitoring in dynamic environmental conditions. However, limitations include the dataset size and the narrow focus on turbidity. Future research could expand to encompass a broader range of water quality factors. This approach holds promise for enhancing water quality management in Indonesia and similar regions.

Introduction

Indonesia's unique geographic location near the equator gives rise to three distinct climate types: monsoonal, tropical, and maritime. These climates are significantly influenced by the country's geographic, topographic, and orographic features [1–3]. These climatic conditions have significant environmental consequences, including shifts in rainfall patterns and tidal sea-level fluctuations [4]. Variations in rainfall can lead to adverse effects, such as flooding and drought, with far-reaching impacts on Indonesia's agriculture, fisheries, and public health [5]. Prolonged dry seasons in different regions can result in water shortages, creating conditions favorable for the proliferation of bacteria, viruses, fungi, and parasites due to higher humidity levels [6–8]. Water contamination is a concern, especially during the rainy season when it mixes with upstream sources.

Water is a vital resource with far-reaching impacts on Earth's natural processes and human existence. However, not all water sources are safe for human consumption, as seemingly clear water may contain hidden impurities [9,10]. According to the Regulation of the Minister of Health of the Republic of Indonesia, water quality encompasses a range of fundamental physical attributes, such as odor, total dissolved solids (TDS), turbidity, taste, and color [11]. These attributes significantly affect water quality.
Regrettably, many Indonesians remain unaware of water quality, particularly regarding turbidity [12]. Turbid water often harbors harmful bacteria and microorganisms. TDS, or Total Dissolved Solids, is a critical parameter for assessing water quality. High TDS levels indicate potentially hazardous contamination, including sulfate, bicarbonate, and other particulates. Excessive TDS in public water sources can lead to skin diseases and harm to internal organs.

In 2018, data from the Environmental and Forestry Agency of Aceh Province showed that TDS levels in Krueng Aceh were excessively high during the rainy season but within normal ranges during the dry season. Rainwater from upstream sources entering Krueng Aceh made the water cloudier during the rainy season due to the increased presence of harmful microorganisms and bacteria, demonstrating the influence of climate factors [13]. This condition demonstrates that elevated TDS levels and turbidity can also be influenced by climate factors.

While water testing tools are indeed accessible to the public, their potential remains largely untapped, primarily because they are often perceived as challenging to use, expensive, and not readily available. These barriers can deter individuals from actively engaging in water quality monitoring and safeguarding their health and the environment. Despite the wealth of information and technology at our disposal, raising awareness about the importance of water testing and improving the accessibility of user-friendly, affordable tools is essential to encourage more people to take proactive steps in ensuring clean and safe water for all.

One promising solution for addressing the challenges associated with water testing tools is harnessing machine learning [14–16]. Machine learning is a subset of artificial intelligence (AI) that involves the use of algorithms and statistical models to enable computers to learn and make predictions or decisions without explicit programming [17–19]. In essence, it allows machines to analyze and identify patterns in data, learn from those patterns, and use that knowledge to perform tasks or make decisions [20,21]. Machine learning is widely used in predictive modeling for tasks like forecasting [22], classification [23–25], regression [26–28] and clustering [29,30]. These algorithms have the potential to automate the analysis of water samples, making the process more accessible, affordable, and efficient. In doing so, they empower individuals and communities to take proactive steps in safeguarding clean and safe water for all while also ensuring a healthier and more sustainable environment.

In this study, we aim to utilize digital image processing and machine learning algorithms for water quality assessment. This approach holds promise as a user-friendly and cost-effective solution to significantly enhance water quality monitoring practices with dynamic environmental conditions, like Indonesia, ultimately contributing to public health and bolstering environmental sustainability.

**Materials and Methods**

**Water Sample Preparation**

Turbid water samples for simulation are created by mixing clay from Krueng Aceh River, Indonesia, with concentrations measured in percentages (%), resulting in 500 mL for each concentration. Turbidity levels are established in a series of concentrations: 0%, 10%, 20%, 40%, 60%, 80%, and 100%, with 0% representing clear water. Measurements of TDS and pH are taken for each simulated sample to obtain data relevant to image processing.

**Dataset**

Each turbid water sample was photographed from an optimal distance to obtain the requisite image data for our study. The water image capture process involved using glass containers containing water samples with varying levels of turbidity, Total Dissolved Solids (TDS), and pH. Consequently, the dataset comprises images saved with the "jpg" extension.
The dataset, comprising 80 water images, is categorized into seven distinct classes based on turbidity levels. These classes include Class 1 (clear, 0% turbidity), Class 2 (10% turbidity), Class 3 (20% turbidity), Class 4 (40% turbidity), Class 5 (60% turbidity), Class 6 (80% turbidity), and Class 7 (100% turbidity). The visualization of each sample of the water images can be found in Figure 1. Each image is represented as an RGB image with a region of interest (ROI) size of 512 x 512 pixels. The dataset is further divided into two subsets: the training set, comprising 60 images used for model development, and the testing set, consisting of 20 images used for model evaluation.

Figure 1. The visualization of each water turbidity level.

Image Classification Methods

The image water classification method encompasses four crucial stages, each contributing to the overall process. These stages are data input, preprocessing, feature extraction, and Model Generation and Classification, as illustrated in Figure 1.

Figure 2. Flowchart of this study.

In the first step, we utilized the training images as the input dataset for model training, employing the Support Vector Machine (SVM) algorithms. SVM is a powerful and widely recognized machine learning technique known for its effectiveness in a diverse array of tasks. This choice was made due to their well-established effectiveness across a range of tasks [31–34].

To enhance the robustness of our approach, we incorporated the preprocessing step, which involves the transformation of the original RGB (Red, Green, Blue) images into grayscale. Grayscale transformation simplifies the image data by converting it into shades of gray, thus reducing the complexity of the images [35,36]. This conversion is essential to standardize the input data for further analysis and helps eliminate the influence of color variations, ensuring that the classification process is primarily based on the intrinsic characteristics of water bodies and their surroundings.

Moving on to the feature extraction stage, this step involves identifying and isolating specific characteristics or attributes from the preprocessed grayscale images that can be used for
classification purposes. In this study, the chosen feature for extraction is the color feature. This means that the feature extraction process aims to quantify and characterize the color information present in the images [37]. By focusing on color as the primary feature, the method is tailored to differentiate various water types, taking into account variations in color that may exist in different water bodies. Extracting color-related information is vital as it offers insights into water quality, contamination, or other factors that influence the color of the water in the images.

After completing the feature extraction stage, the next step involves using the trained model to classify the water turbidity in a testing dataset. This process aims to leverage the information gathered during training to make predictions about the turbidity levels of water bodies in new, unseen images. The results of the classification process provide insights into the turbidity levels of the water bodies depicted in the testing images.

**Results and Discussion**

The objective of this study is to employ digital image processing and machine learning algorithms for the assessment of water quality. To achieve this, an SVM model was trained to categorize water turbidity levels into seven distinct classes. These classes allow for the estimation of Total Dissolved Solids (TDS) and pH values. In Table 1, we present the TDS and pH values that were measured for each class of water turbidity levels. This dataset was derived from in-situ measurements conducted on simulated samples using a portable measurement device. The data collected from these measurements serves as a valuable reference for predicting the turbidity levels, TDS, and pH values of objects within the testing dataset.

<table>
<thead>
<tr>
<th>Water Turbidity Level</th>
<th>Parameter</th>
<th>TDS (mg/L)</th>
<th>pH</th>
</tr>
</thead>
<tbody>
<tr>
<td>0% ± SD</td>
<td>TDS</td>
<td>9.1 ± 0.6</td>
<td>6.90 ± 0.20</td>
</tr>
<tr>
<td>10% ± SD</td>
<td>TDS</td>
<td>327.1 ± 21.5</td>
<td>7.27 ± 0.05</td>
</tr>
<tr>
<td>20% ± SD</td>
<td>TDS</td>
<td>371.6 ± 13.5</td>
<td>7.37 ± 0.07</td>
</tr>
<tr>
<td>40% ± SD</td>
<td>TDS</td>
<td>598.1 ± 32.7</td>
<td>7.47 ± 0.06</td>
</tr>
<tr>
<td>60% ± SD</td>
<td>TDS</td>
<td>869.1 ± 11.7</td>
<td>7.59 ± 0.03</td>
</tr>
<tr>
<td>80% ± SD</td>
<td>TDS</td>
<td>940.3 ± 8.5</td>
<td>7.85 ± 0.12</td>
</tr>
<tr>
<td>100% ± SD</td>
<td>TDS</td>
<td>1490.3 ± 22.6</td>
<td>8.85 ± 0.11</td>
</tr>
</tbody>
</table>

**Table 1.** Classification for each object based on class and estimated TDS and pH water values.

The Table 1. provides valuable insights into the relationship between water turbidity levels, TDS concentrations, and pH values. At the lowest turbidity level of 0%, the TDS concentration in the water is measured at approximately 9.1 mg/L, with a relatively low standard deviation of 0.6, indicating a degree of consistency in this measurement. However, as turbidity levels increase, there is a noticeable trend of rising TDS concentrations, indicating a potential correlation between water turbidity and TDS levels. For instance, at 100% turbidity, the TDS concentration significantly increases to 1490.3 mg/L, with a standard deviation of 22.6.

Regarding pH values, the data also reveals interesting trends. At 0% turbidity, the pH of the water is recorded at 6.90, with a standard deviation of 0.20. As turbidity levels rise, there is a gradual increase in pH, with higher turbidity levels associated with higher pH values. For instance, at 100% turbidity, the pH level is measured at 8.85, with a standard deviation of 0.11. These findings suggest that both TDS concentrations and pH values may vary systematically with changes in water turbidity, and this information is crucial for understanding water quality and environmental conditions.

Table 2 represents the classification of water suitability levels assigned to newly acquired images, ranging from “Very good” with 0% turbidity to “Dangerous” with 100% turbidity. After testing these new images, the system will output the percentage of water turbidity, allowing
for a precise characterization of the water quality. The information gleaned from this turbidity level assessment is instrumental in estimating key parameters for water quality. This comprehensive evaluation based on newly acquired image data is vital for understanding and managing water quality in various contexts.

Table 2. Water suitability assessment.

<table>
<thead>
<tr>
<th>Water Suitability Level</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very good</td>
<td>0%</td>
</tr>
<tr>
<td>Very good</td>
<td>10%</td>
</tr>
<tr>
<td>Good</td>
<td>20%</td>
</tr>
<tr>
<td>Good</td>
<td>40%</td>
</tr>
<tr>
<td>Can be used</td>
<td>60%</td>
</tr>
<tr>
<td>Bad</td>
<td>80%</td>
</tr>
<tr>
<td>Dangerous</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 3 presents the outcomes of the testing data, showcasing the system's predictions regarding the water quality. This testing process involved using images of mineral water from beverage products and river water from the Banda Aceh region. The image data was input into the system and processed according to the methodology described earlier. The system's results reveal distinct categories for the tested water samples. Mineral water is classified under Group 1, denoting clear water with 0% turbidity. The system estimates the Total Dissolved Solids (TDS) content for this group to be $9.1 \pm 0.6$ mg/L, with a pH of $6.90 \pm 0.20$, indicating excellent suitability for consumption. Conversely, river water falls into Group 3, which corresponds to a turbidity level of 20%. The system estimates the TDS content for this group to be $371.6 \pm 13.5$ mg/L, with a pH of $7.37 \pm 0.07$, signifying good suitability. These findings align with in-situ measurements taken using a portable measurement device on sample water, confirming the system's accuracy in assessing water quality and turbidity levels for different water sources.

Table 3. Image processing of water samples.

<table>
<thead>
<tr>
<th>No</th>
<th>Water Image</th>
<th>Type of Water</th>
<th>Class</th>
<th>Suitability</th>
<th>TDS (mg/l)</th>
<th>pH</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Mineral</td>
<td>Class 1</td>
<td>Very good</td>
<td>$9.1 \pm 0.6$</td>
<td>$6.90 \pm 0.20$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Water</td>
<td>(clear, 0% turbidity)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Mineral</td>
<td>Class 1</td>
<td>Very good</td>
<td>$9.1 \pm 0.6$</td>
<td>$6.90 \pm 0.20$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Water</td>
<td>(clear, 0% turbidity)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>River</td>
<td>Class 3</td>
<td>Good</td>
<td>$371.6 \pm 13.5$</td>
<td>$7.37 \pm 0.07$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Water</td>
<td>(20% Turbidity)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

These findings demonstrate the system's capacity to accurately categorize water samples, as exemplified by the classification of mineral water and river water. These results highlight the practical application of this technology in the real world, with implications for water quality management in various sectors.

Conclusions

In conclusion, this research demonstrates a promising and innovative approach to assess water quality in regions prone to dynamic environmental changes, such as Indonesia. The integration of digital image processing and machine learning, specifically SVM allows for efficient and automated classification of water samples based on turbidity levels, enabling quick and
accessible water quality assessment. The estimation of TDS and pH values from image data provides a valuable tool for understanding water quality. The findings indicate the system’s potential to accurately evaluate water suitability for various applications, from very good to dangerous, offering a user-friendly and cost-effective solution for water quality monitoring and enhancing public health and environmental sustainability in Indonesia and similar regions.

While the results of this study offer promising insights into the application of digital image processing and machine learning for water quality assessment, several limitations should be acknowledged. One significant constraint is the size of the dataset employed for this research. The dataset encompasses a relatively limited number of water samples and a specific range of turbidity levels. A more extensive and diverse dataset would undoubtedly enhance the model’s accuracy and the generalizability of its findings to a broader spectrum of water sources and conditions. A larger and more varied dataset would better capture the intricacies of water quality in the real world. Additionally, while this research primarily focuses on the influence of turbidity levels on water quality, it’s essential to acknowledge that water quality is a multidimensional parameter. It is affected by a complex interplay of factors, including chemical composition, the presence of microorganisms, and contaminants. The relatively narrow scope of this study might not fully encompass the complexity of real-world water quality scenarios. Future research could extend its purview to encompass these additional variables, providing a more holistic understanding of water quality.

**Funding:** This study does not receive external funding.

**Ethical Clearance:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data is available upon reasonable request to the corresponding author.

**Acknowledgments:** We extend our sincere thanks to Mr. Muhibbul Khibri, Mrs. Cut Anizar, and all the teachers at SMA Negeri 10 Fajar Harapan Banda Aceh for their valuable support in this research.

**Conflicts of Interest:** All the authors declare that there are no conflicts of interest.

**References**


