Spectral Mixture Analysis-based WQI with Convolutional Long Short-Term Memory Techniques

Ika Oktavianti, Yusuf Hartono, Sukemi
Department of Informatics Engineering, Sriwijaya University, Indonesia

***Abstract*—Surface water, including river water, is an important natural resource for human life. However, river water quality in Indonesia often declines due to various factors, such as excessive water consumption, waste pollution, and natural disasters. This study aims to predict the Water Quality Index (WQI) of rivers using Spectral Mixture Analysis with deep learning architecture. The methods used in this study are Spectral Mixture Analysis (SMA) and Convolutional Long Short-Term Memory (ConvLSTM). SMAs are used to decompose the spectral signatures of water quality components and provide insight into the composition of water bodies. ConvLSTM, a deep learning architecture, is used to capture temporal dependencies and spatial patterns in water quality data. The results showed that the percentage of WQI prediction accuracy for 345-band model was better than 234-band model, reaching 34.78%. The visible color spectrum that represents the Meets (M) and Light (R) Pollution Index is Blue (0, 0, 255) and wavelengths ranging from 0.53 µm to 0.88 µm. The test results of the ConvLSTM hybrid model on 8 mandatory parameters of River WQI measurements at 30 watershed monitoring points of North Musi Rawas Regency from 2021 to 2023, the accuracy value reaches 96% or it is considered that the performance of this model is acceptable. This research proves that Spectral Mixture Analysis with hybrid model Convolutional Long Short-Term Memory techniques is effectively capable of predicting and monitoring the WQI of rivers and these results can be used to take appropriate steps in determining policies.**

Keywords*—*Water quality index; Spectral Mixture Analysis; remote sensing; deep learning; convolutional long short-term memory

I. Introduction

Surface water is water that collects above ground or in springs, rivers, lakes, wetlands, reservoirs or other puddles that do not experience infiltration underground. Surface water classified as river water is widely used for various purposes including drinking, household needs, irrigation, power generation, and industry, as well as supporting all forms of life and affecting health, lifestyle, and human economic well-being [1]. In its utilization, river water quality is influenced by the environment around the river flow and in general the water quality in the upstream part is higher than the downstream. This is caused by industrial, and household waste and all human daily activities that go directly to the river without going through the processing or purification process first.

Currently, the water quality in Indonesia is still relatively low. The low quality of water is due to the influence of contamination of domestic waste, agricultural and livestock industries. The problem of poor water quality is also influenced by excessive water consumption, limited sources of clean water, household waste pollution, and industrial activities. On the other hand, the growing population and the growing industrial expansion require a lot of water supply. The water needed is not only used for household purposes but also industry [2]. South Sumatra as one of the provinces in Indonesia which is fed by nine tributaries of the Musi River has a relatively low level of water pollution compared to the national level. The Water Quality Index (WQI) value for South Sumatra Province in 2017 was 63.81 higher than the National WQI of 53.20. However, in 2022 the WQI value in South Sumatra Province has decreased to 59.85 but the pollution status category is still quite high. Based on observations, it is known that the value of WQI in South Sumatra Province fluctuates from year to year.

Fluctuations in the value of the WQI in South Sumatra are also influenced by natural disasters. Flood natural disasters can have an impact on poor water quality and damage river ecosystems. Floods that deliver poor-quality water to settlements can also adversely affect health. North Musi Rawas Regency is one of the areas in South Sumatra that is often affected by floods. This is because many villages in this district are in the Watershed Area (DAS). In monitoring and controlling watersheds in an area, the availability of comprehensive and accurate data is needed, while the current condition of data related to watershed conditions is still very limited to access. Therefore, identification and inventory of watershed conditions in North Musi Rawas Regency is very necessary. The need for various kinds of data to support research and policy-making carried out by government agencies is very important to allocate budgets effectively and efficiently.

The use of technology in data and information processing can help decision-making and target achievement become more effective and efficient. Based on the phenomena that occur today, data and information related to the river WQI become indispensable in identifying polluted river water. The use of the WQI can facilitate the determination of river water quality and facilitate the provision of information to those in need [3]. Regular monitoring of water quality is an effort to ensure the water used is safe and healthy for humans and the environment.

One of the methods used to determine the value of the WQI is based on Spectral Mixture Analysis with remote sensing. Spectral Mixture Analysis is an analysis method that uses a combination of certain algorithms using values from end members in the spectral library which is usually done to identify an object that is indicated to have mixed pixels. The advantage of using the SMA method can provide detailed information up to the subpixel level quantitatively from land cover [4]. Furthermore, remote sensing is generally used to examine physical parameters of water quality that have visual characteristics such as water surface temperature, water turbidity, and dissolved solids [5].

Various approaches are used to investigate water quality indices by remote sensing. General methodologies for evaluating concentrations of different variables have evolved from simple linear regression methods and nonlinear multiple regression to principal component analysis (PCA) and neural networks [6–9]. A number of these investigations used tape mathematical algorithms to select correlated single bands and band ratios to map the distribution of spatial indicators [10–11]. However, some conventional regression models may no longer be optimal, especially when there are complex nonlinear relationships between water system behavior and environmental factors.

In recent years, by not changing the classical approach, the use of big data tools and technologies in the water quality sector has become a consensus, although several machine learning-based studies have shown promising results in overcoming low accuracy in time series using simple empirical models [12-13]. However, thoroughly understanding the complex two-way interaction in temporal and spatial contexts is still a challenge [14–16]. Deep learning techniques provide an opportunity to study the characteristics of spatial or temporal correlations [17-20]. Previous studies found that the new artificial intelligence (AI) approach of the Convolutional Long Short-Term Memory (ConvLSTM) model dramatically outperformed classical sequence modeling methods in capturing spatiotemporal correlation data from satellite imagery input. Previous research had been carried out on Lake Small Prespa in Greece, namely predicting water quality variables, namely DO and Chlorophyll-a. The research results show that the Hybrid CNN-LSTM model succeeded in capturing low- and high-level water quality variables, especially for DO concentration. The data used in this research are time series data for water quality parameters such as pH, temperature, DO, Chl-a. The disadvantage of this research is that it uses 3 water measurement parameters and must collect data in the field periodically, so it requires time and money, and is less effective if used as a baseline for predicting water quality in other lakes [26]. However, attempts to apply the ConvLSTM algorithm to water quality extraction are still rare, so special attention is needed in applying WQI modeling.

In this study, we recommend using ConvLSTM quickly by taking data from the results of Landsat 8 OLI/ TIRS spectral standardization at the location of the monitoring point. The initial method used was Spectral Mixture Analysis which was later combined with the Convolutional technique of Long Short-Term Memory to study short-term spatial and temporal characteristics. The Spectral Mixture Analysis method with the Convolutional Long Short-Term Memory technique conducted in this study is believed to make a good combination in measuring WQI in Indonesia in general and in South Sumatra Province in particular. This paper is structured as follows: Section I describes the background and literature review of the various publications related to this study. Data sets and research methods are discussed in Section II. Results and discussion are explained in Section III and Section IV continued with conclusions.

II. MATERIAL AND METHOD

SMAs are used to decompose the spectral signatures of water quality components, providing valuable insight into the composition of water bodies. ConvLSTM, a deep learning [21] architecture, is used to capture temporal dependencies and spatial patterns in water quality data. By combining spectral information from SMA with the sequential analysis capabilities of ConvLSTM, our proposed method, called the Spectral Mixture Analysis WQI (SMA-WQI), offers a comprehensive framework for assessing water quality conditions.

The effectiveness of the SMA-WQI model was evaluated using performance metrics such as Mean Absolute Error (MAE) and Correlation Coefficient (r) because previous experimental results showed that the SMA-WQI model outperformed the base model, demonstrating its superiority in water quality assessment. In addition, sensitivity analysis was performed to test the robustness of the SMA-WQI model against variations in input parameters and model configuration. The proposed approach promises to improve water quality monitoring and management practices, providing valuable insights for environmental decision-makers and policymakers.

A. *Spectral Mixture Analysis*

The data used in this study consisted of in-situ measurements obtained from 30 monitoring points spread across seven sub-districts in North Musi Rawas Regency, representing river conditions in the administrative area of North Musi Rawas Regency. Furthermore, this data will be standardized with spectral data. In-situ data was collected manually, and samples were then taken to the Environmental Laboratory of BLUD UPT Musi Rawas Regency to test the parameters of Chemical Oxygen Demand (COD) and Total Suspended Solids (TSS) according to certified test procedures (Fig. 1). The Environmental Laboratory of BLUD UPT Musi Rawas Regency will then issue a Test Results Report (LHU). Eight parameters are included in the test results, namely pH, Dissolved Oxygen (DO), TSS, COD, Biochemical Oxygen Demand (BOD), Nitrate, Fecal Colliformes (FC), and Phosphate. The study was conducted from 2021 to 2023.

Fig. 1. River water sampling process.

The spectral data used came from the Landsat 8 OLI/TIRS Satellite equipped with 11 bands, which can represent the spectrum of each river monitoring point. Satellite data collection is carried out by the process of spectral down sampling by determining the location code of each river monitoring point through USGS Earth Explorer, resulting in path code 125 and row 062. Based on this location code, satellite data is taken every semester from 2021 to 2023.

Landsat 8 OLI/TIRS data is then processed using the Software Integrated Land and Water Information System (ILWIS) application. ILWIS is a PC-based Geographic Information System (GIS) and Remote Sensing software developed in 2005. ILWIS provides a wide range of figure processing, spatial analysis, and digital mapping functions. Reasons for using ILWIS include its open-source nature, consistency in the use of georeferenced systems on earth maps, good raster processing capabilities, adherence to topological principles during the editing process, attractive layout display, and the ability to integrate spatial data from various formats as well as tabular data.

The satellite data and in-situ data were then adjusted to the standard using 10 color clusters and analyzed using the Spectral Mixture Analysis method by combining Band 2, Band 3, and Band 4 for the first model, as well as a combination of Band 3, Band 4, and Band 5 for the second model, according to their respective functions. Clustering using 10 colors is done to produce more complete and consistent color differences. Furthermore, the color pixel value is nominated into RGB standards with a color range of 0-255 through a stretching process, resulting in a color combination that matches the RGB combination. The standardized data is then processed using the SMAs method (see Fig. 2).

B. Convolutional Long Short-Term Memory (ConvLSTM)

In this study, using the Convolutional Long Short-Term Memory approach to study long-term spatial and temporal characteristics. ConvLSTM was developed specifically in assisting problems of predicting spatial-temporal sequences and according to previous research, is more effective in extracting spatial and temporal characteristics from feature graphs [22]. This allows ConvLSTM to analyze and predict events in time series, to combine spatial data from a single feature map [23]. To create the ConvLSTM equation [24], the equation is used:

$$it = \sigma(Wpi * Xt + Whi * K(t-1) + Wci \circ C(t-1) + yi) \quad (1)$$

$$ft = \sigma(Wpf * Xt + Whf * K(t-1) + Wcf \circ C(t-1) + yi) \quad (2)$$

$$Ct = ft \circ C(t-1) + it \circ tanh(Whc * K(t-1) + Wxc * Pt + yc) \quad (3)$$

$$Ot = \sigma(Wpo * Pt + Who * K(t-1) + Wco \circ Ct + yo) \quad (4)$$

$$ft = Ot \circ tanh(Ct) \quad (5)$$

The major innovation of LSTM is its memory cell Ct which essentially acts as an accumulator of the state information. The cell is accessed, written and cleared by several self-parameterized controlling gates. Every time a new input comes, its information will be accumulated to the cell if the input gate it is activated (1). Also, the past cell status $C(t-1)$ could be "forgotten" in this process if the forget gate ft is on (2) (3). Whether the latest cell output Ct will be propagated to the final state ht is further controlled by the output gate Ot (4) (5). The it, ft, and Ot gates each represent the 3D tensor of ConvLSTM. The last 2D that is spatial is rows and columns. The operators "$*$" and "$\circ$", respectively, are convolution operators and "Hadamard products". In this case, ConvLSTM is equipped with a batch normalization layer and a dropout layer.

The resulting application of this study is a combination of Statistical Analysis of Temporal Dynamics and Spatial Pattern Identification, namely WQI Spectral Database and Water Period Differences, which can be useful for monitoring water quality in different river areas (see Fig. 3).

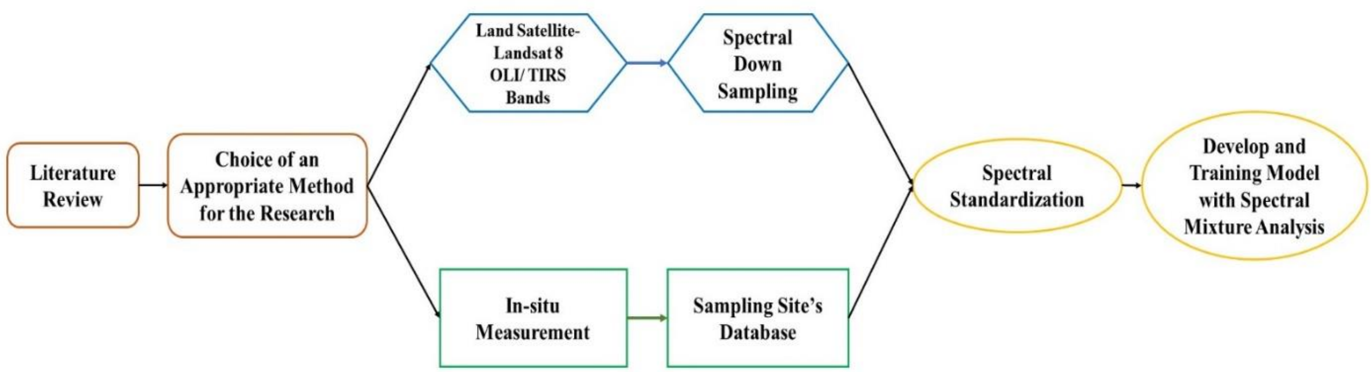


Fig. 2. Research method with SMA.

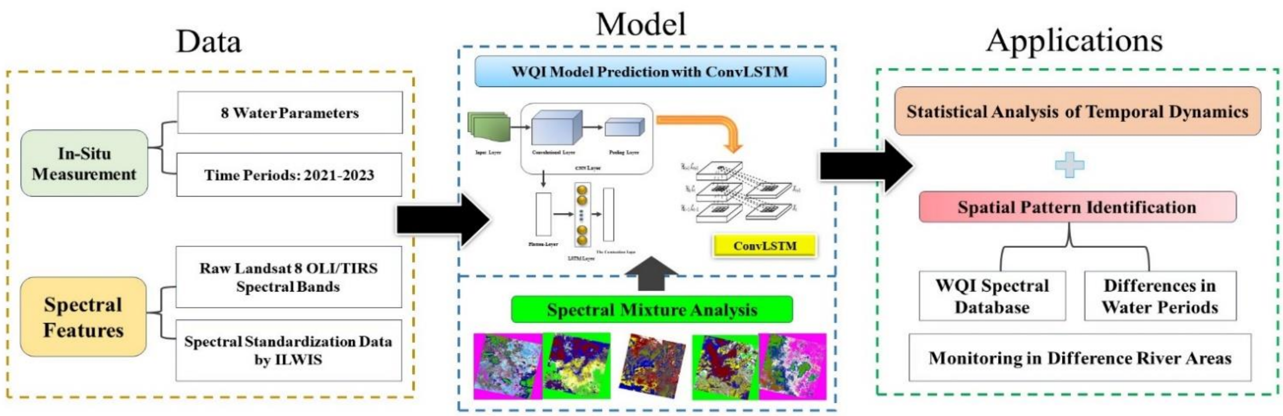


Fig. 3. A Framework of WQI based on SMA and ConvLSTM techniques.

III. Experimental Results and Discussion

A. Experimental Results

The results of this study consist of Pollution Index Score, WQI Score, results of standardization of in-situ data and Landsat 8 OLI/ TIRS data, as well as the results of Spectral Mixture Analysis with a Convolutional Long Short-Term Memory techniques approach.

1) Pollution index and river WQI: The results of monitoring at river monitoring points are separated per category of sample points that meet water quality, so that the Pollution Index (IP) and Index Value per Water Quality are obtained as the basis for calculating the WQI score. Data on the Pollution Index (IP) per semester from 2021 to 2023 can be seen in Fig. 4.

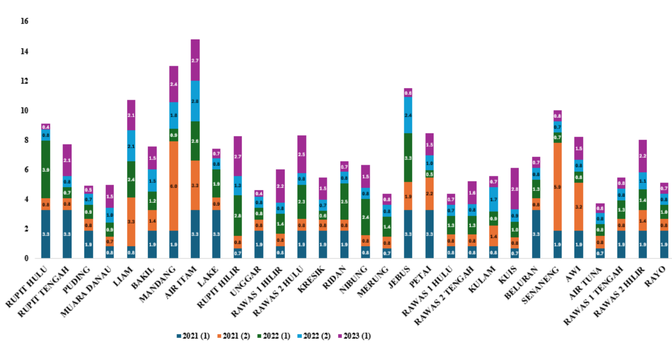


Fig. 4. IP diagram at river monitoring point for 2021 – 2023.

The results of the calculation of the Index value per Water Quality obtained by the WQI Score in 2021 are 59.7; Year 2022 is 61.3 and Year 2023 is 58.7 (Table I). Based on the Canadian Council of Ministers of the Environment (CCME), the higher the WQI River, the better the water quality, it can be concluded that the status of WQI Sungai in Muratara Regency from 2021 to 2023 is monitored poorly. This means water quality is often threatened and compromised, conditions often deviating from naturally desired levels [25].

2) In-Situ data standardization and satellite data of landsat 8 OLI/TIRS: The spectral standardization process produces 8-9 color spectrums with wavelengths from 0.53 µm to 0.88 µm. In the 234-band model, the colors that are not visible from the standardization process are red (255, 0, 0) and magenta (249, 0, 255) with the standardization chart can be seen in Fig. 6 with the highest number of color spectrum, namely blue (0, 0, 255) as many as 39 points and then navy (0, 0, 128) as many as 36 points (Fig. 5). While in the 345-model band, the color that is not visible is aqua color (0, 255, 255) with the standardization chart can be seen in Fig. 6 with the highest number of color spectrum, namely blue (0, 0, 255) as many as 49 points and then navy colors (0, 0, 128) as much as 29 points.

TABLE I. River WQI in North Musi Rawas Region for 2021-2023

(a)

Water Quality	Index Value Weight	Number of Points for Water Quality		
		2021	2022	2023
Good	70	31	34	26
Light	50	27	26	34
Moderate	30	2	0	0
Heavy	10	0	0	0

(b)

Water Quality	Index Value Weight	Percentage of Fulfillment of Quality Standards		
		2021	2022	2023
Good	70	52%	57%	43%
Light	50	45%	43%	57%
Moderate	30	3%	0%	0%
Heavy	10	0%	0%	0%

(c)

Water Quality	Index Value Weight	Index Value for Water Quality		
		2021	2022	2023
Good	70	36,2	39,7	30,3
Light	50	22,5	21,7	28,3
Moderate	30	1	0	0
Heavy	10	0	0	0
WQI = Not Good		59,7	61,3	58,7

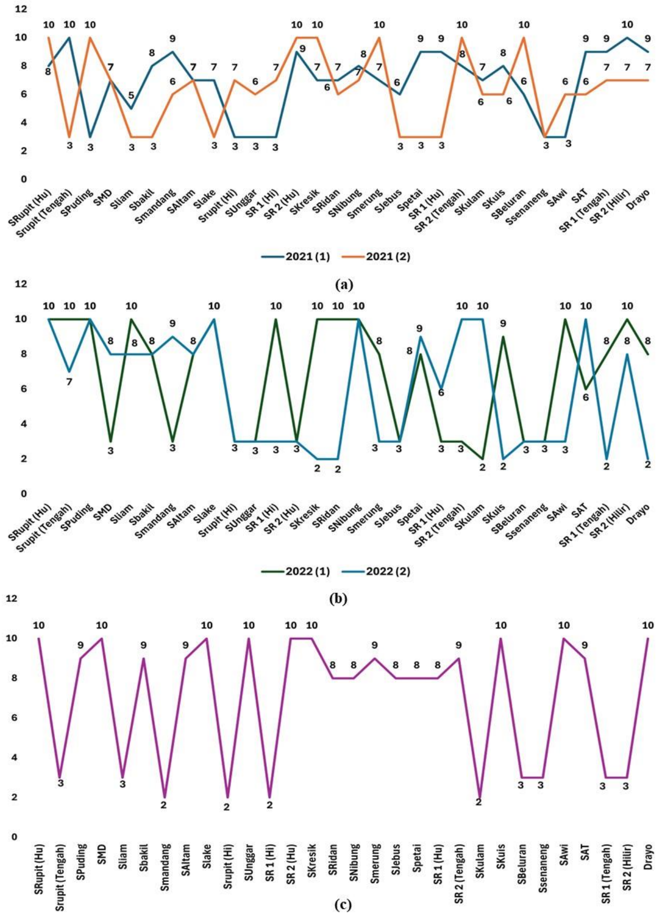
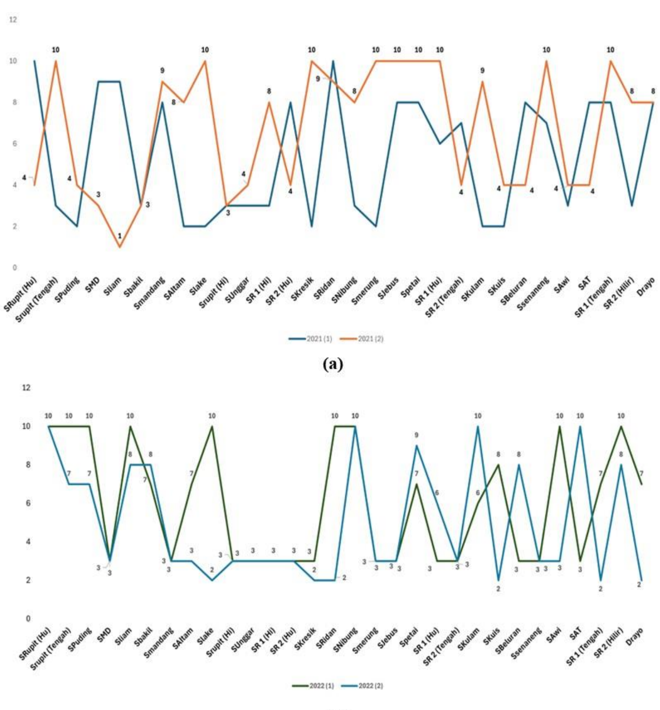


Fig. 5. 234-Band model standardization chart (a) 2021 Semester 1 and 2; (b) 2022 semester 1 and 2; (c) 2023 semester 1.



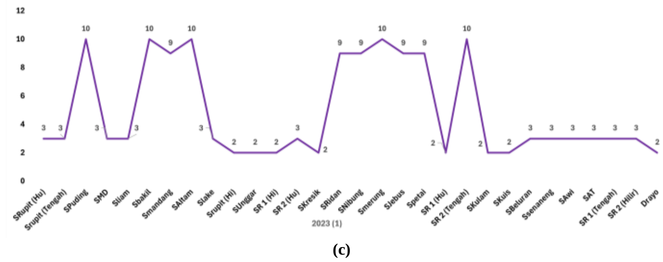


Fig. 6. 345-Band model standardization chart (a) 2021 semester 1 and 2; (b) 2022 semester 1 and 2; (c) 2023 semester 1.

3) Spectral mixture analysis process: The Landsat 8 OLI satellite data process using the ILWIS application remote sensing system produces mixed spectral maps per semester from 2021 - 2023 (Fig. 7).

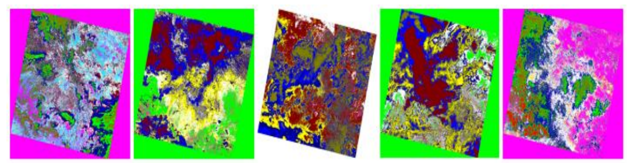


Fig. 7. Mixture spectral map.

The results of standardization produce color combinations that affect two categories of pollution index, namely pollution Meets (M) and Light pollution (R) (Table II).

TABLE II. LEVEL OF ACCURACY WITH SPECTRAL MIXTURE ANALYSIS FOR PREDICTION OF WQI

Band Model	Accuracy Percentage (%)			
	Good Pollution Index (M)	Color Spectrum	Light Pollution Index (R)	Color Spectrum
234	26,58	Blue	27,54	Navy
345	31,65	Blue	34,78	Blue

Based on the observations of Spectral Mixture Analysis, the accuracy rate of the 345-band model is better, reaching 34.78%, compared to the 234-band model which is only 26.58%. Although the accuracy level of the 345-band model is better, the level of accuracy in predicting the WQI is still considered too small. This can be caused by several things, including because: 1) the comparison of wavelengths with the condition of 8 mandatory parameters being measured; and 2) Landsat 8 OLI/TIRS satellite data at two monitoring point locations where cloud cover reached 84%, thus affecting the accuracy of the analysis.

4) Convolutional long short-term memory analysis

a) Selection of eigenvalues: From the number of missing values in each indicator, the dissolved oxygen index value lost in this area is the smallest, and prediction of index using this model has higher accuracy and reliability. Therefore, dissolved oxygen will be the main research to evaluate water quality in this region. The causes of value loss are broadly divided into human factors and natural factors. The natural factor is that in the data collection process, due to machine factors, there are some data collectors that fail, or the collected data cannot be stored, resulting in some data that cannot be saved. The human factor is caused by human error in the collection process resulting in the loss of some data. If there are enough samples in the data set, the ratio of missing values is relatively small. This small amount of missing value has less impact on the overall situation and can be eliminated directly. Therefore, the missing data value is immediately deleted in this experiment.

b) Handling outliers: In the process of collecting data, there will be abnormal objects due to different types of data sources, data measurements, and collection errors. Abnormal objects are often called outliers. Outlier detection, also known as deviation detection and exclusion mining, is often used as an important part of data mining. The task is to find objects that differ significantly from most data. Therefore, most data mining methods treat this difference in information as noise. Boxplots use the distance between interquartile values as the basis for assessment, so they have objectivity and superiority in identifying outliers. It can be seen from the data set that there is an excessive data between the maximum and minimum dissolved oxygen values, which may be due to the length of the working time of the instrument and the aging of the instrument, resulting in errors in the data at the same time. some point in time. Therefore, such data can be regarded as abnormal values during data analysis.

c) Normalization: In order for the ConvLSTM model to converge faster and have higher stability in the training process, dissolved oxygen data is normalized. To prevent the model from performing well in training sets but generally in test sets, its generalizability is weak. Therefore, this paper takes the resample data of dissolved oxygen concentration data in chronological order by day and divides the data into training sets and verification sets with a proportion of 8:2 in the training process. That is, 732 sample data was used to verify the performance of the model. The results of abnormal data that have not been normalized from the 8 mandatory parameters of WQI measurement are shown in Fig. 8.

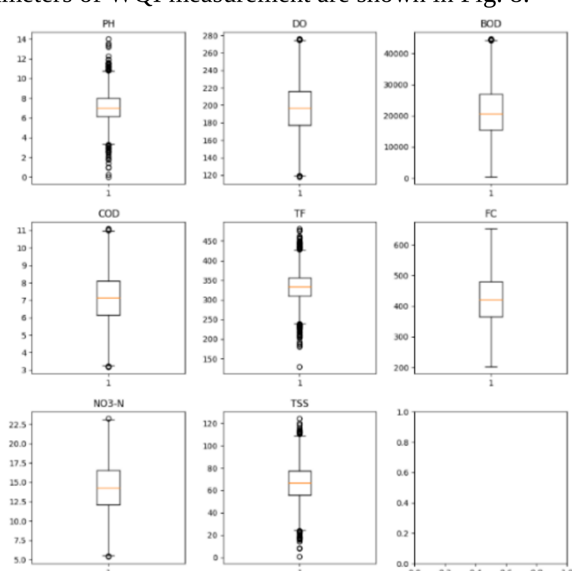


Fig. 8. Abnormal data from 8 parameters with interpolation method.

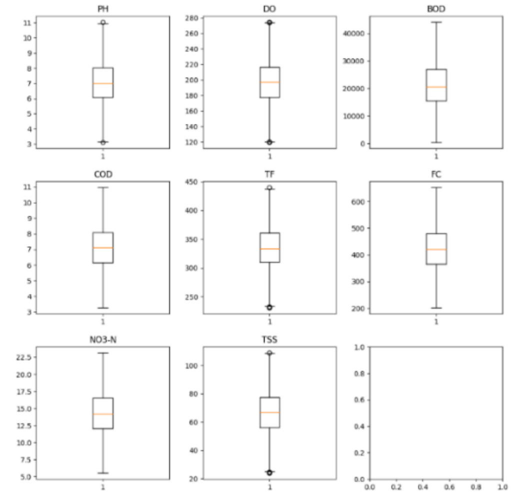


Fig. 9. Normalized data with interpolation method.

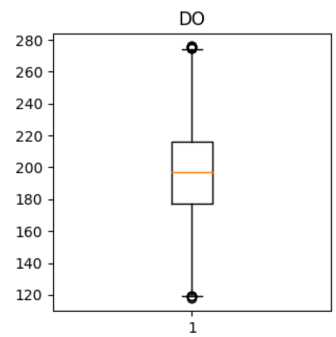


Fig. 10. Data cleaning results from Dissolved Oxygen parameter.

The effect of removing abnormal values in the diagram (Data Cleaning) of the 8 Mandatory Parameters using the interpolation method is shown in Fig. 9, and the comparison diagram after handling abnormal values specific to the dissolved oxygen parameter (Fig. 10). Fig. 11 shows WQI from dissolved oxygen parameter.

d) Data recovery: When evaluating the model after training, to eliminate the impact of normalization on the prediction results, the prediction data needs to be recovered to evaluate the error of the model's prediction values.

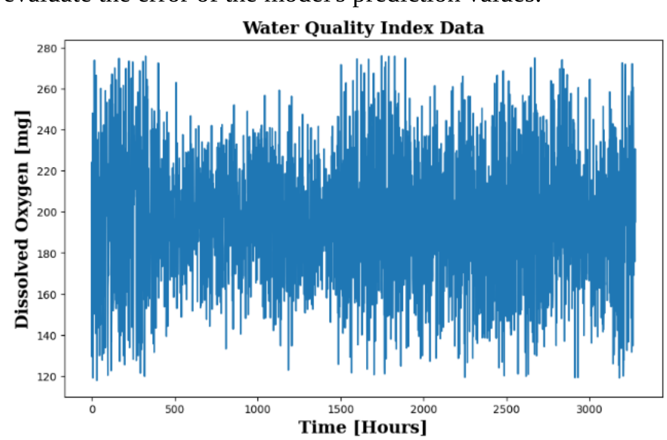


Fig. 11. WQI from dissolved oxygen parameter.

To influence the model's prediction against actual data, this experiment uses the ConvLSTM model to predict the dissolved oxygen concentration in the test sample. At the same time, in order to compare the predictive effects of the two models more clearly, we performed a visual comparative analysis of the data of the first sample, as shown in Fig. 8 and 9. It can be seen that although the ConvLSTM hybrid model can better predict periodic changes in dissolved oxygen, the fit between more significant values and smaller values is relatively poor, resulting in relative deviations. It can be shown that ConvLSTM has stronger prediction performance.

e) Result test of ConvLSTM model: To test the application of the ConvLSTM model in predicting the WQI, commonly used evaluation metrics such as Mean Absolute Error (MAE) and Correlation Coefficient (*r*) were used. Fig. 12 shows the results of processing MAE and r values data from the four ConvLSTM models tested, while Fig. 13 shows the MAE distribution in histogram form.

	PH	DO	BOD	COD	TF	FC \
0	NaN	204.890455	20791.318981	7.300212	368.516441	564.308654
1	3.716080	129.422921	18630.057858	6.635246	NaN	592.885359
2	8.099124	224.236259	19909.541732	9.275884	NaN	418.606213
3	8.316766	214.373394	22018.417441	8.059332	356.886136	363.266516
4	9.092223	181.101509	17978.986339	6.546600	310.135738	398.410813
...	...	...	...	...	...	...
3271	4.668102	193.681735	47580.991603	7.166639	359.948574	526.424171
3272	7.808856	193.553212	17329.802160	8.061362	NaN	392.449580
3273	9.419510	175.762646	33155.578218	7.350233	NaN	432.044783
3274	5.126763	230.603758	11983.869376	6.303357	NaN	402.883113
3275	7.874671	195.102299	17404.177061	7.509306	NaN	327.459760

	NO3-N	TSS	Potability
0	10.379783	86.990970	0
1	15.180013	56.329076	0
2	16.868637	66.420093	0
3	18.436524	100.341674	0
4	11.558279	31.997993	0
...	...	...	...
3271	13.894419	66.687695	1
3272	19.903225	NaN	1
3273	11.039070	69.845400	1
3274	11.168946	77.488213	1
3275	16.140368	78.698446	1

Fig. 12. MAE distribution from 8 parameters of water quality.

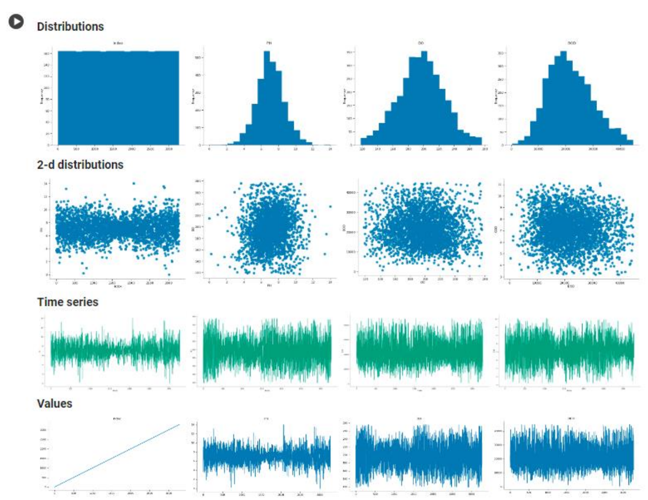


Fig. 13. MAE distribution in histogram.

The test results of the application of the ConvLSTM model with data on 8 mandatory parameters from 2021 to 2023 show an *r* score of 0.96575094 and an RMSE accuracy value of 96% or it is considered that the performance of this model is acceptable or better than the accuracy of applying the model CNN and LSTM separately (range ± 92% - 93%).

B. Analysis Discussion

River WQI monitoring in North Musi Rawas Regency carried out in this research was based on IP values with 4 categories, namely Meet (M), Light (R), Medium (S) and Heavy (B). From the results of in-situ data processing, it was obtained that the number of Sample Meet (M) points was 91, Light (R) was 87 and Medium (S) was 2 data, and there was no heavy pollution category (B). The amount of pollution M and R is not significantly different and is still much greater than the amount of pollution status S and no pollution status B was found. This can mean that in general the condition of the river in the Muratara watershed is still relatively good even though the WQI value is in the Poor category.

In terms of standardization of in-situ data and Landsat 8 OLI/TIRS satellite data, it produces a spectral library which can then be used in the Spectral Mixture Analysis process. In the Landsat 8 OLI spectral image library, endmember values are obtained from the region of interest (roi). Pixels that have been selected using the ROI method can be seen as endmember values using the pixel purity index method. Standardization can be used as a basis for conducting Mixed Spectral Analysis in this research. So that the analysis results are more accurate and the resulting color spectrum to represent pollution status is better.

The Spectral Mixture Analysis carried out in this research first produces a mixed spectral map. The level of accuracy on mixed spectral maps is influenced by the percentage of cloud cover at the location, and the accuracy of the coordinate points and includes 8 mandatory parameters for measuring water quality. The color resulting from the Spectral Mixture Analysis of the 345-band model for the Pollution Index categories M and R is Blue. This means that if the river water is in good condition or there is light pollution, the reflected color that will appear on the 345-band model is blue. The reason is the characteristics of band 5 which measures near infrared, or NIR. This part of the spectrum is critical to ecology because healthy plants reflect water on their leaves, scattering the wavelengths back into the sky. The Pollution Index Category S is not discussed because the number of samples is too small and can bias the analysis results.

IV. Conclusion

This research produces a river WQI monitoring model using the SMA method, namely a 345-band model with a visible color spectrum which represents the Met (M) and Light (R) Pollution Index which is Blue and wavelengths (spectrum) ranging from 0.53 µm up to 0.88 µm. The results of testing the application of the hybrid ConvLSTM model with data on 8 mandatory parameters for River WQI measurements at 30 watershed monitoring points in Muratara Regency from 2021 to 2023, produced an accuracy value of 96% or it is considered that the performance of this model is acceptable or better than the accuracy of applying the model CNN and LSTM separately (range ± 92% - 93%). The results of these predictions can be used as a baseline for river monitoring, not only in South

Sumatra Province, but also throughout Indonesia. This research also shows that the results of monitoring with the SMA Model with the ConvLSTM hybrid modeling technique can be used effectively to predict and monitor river WQI, saving time and costs and these results can be used to take appropriate steps in determining policies.

This research can still be developed in the future by continuing to use the SMA model with the ConvLSTM approach, but considering environmental factors, such as rainfall (Rain-ConvLSTM), distance between pollution sources (Distance-ConvLSTM), and connectivity between measurement points (Connectivity- ConvLSTM) is estimated to have better performance in predicting WQI. Accurate rainfall forecasts are very important because they have a big impact on people's social and economic activities. Rainfall data processed using the ConvLSTM model is expected to reduce the RMSE value by up to 23% so that the WQI prediction value is more accurate. This also applies to other environmental considerations.

REFERENCES

[1] P.U. Igwe, C.C. Chukwudi, F.C. Ifenatuorah, I.F. Fagbeja, and C.A. Okeke. "A review of environmental effects of surface water pollution", *International Journal of Advanced Engineering Research and Science*, *4*(12), 237340, 2017.

[2] A.V. Saraswati. March 24, 2023. "World Water Day: How Good Is World Water Quality?" Green Info, https://greeneration.org/publication/green-info/hari-air-sedunia-kualitas-air-indonesia/. Retrieved June 18, 2023.

[3] E. K. Sari and O. E. Wijaya. "Determination of Water Quality Status by Pollution Index Method and Ogan River Pollution Control Strategy Ogan Komering Ulu Regency", *Journal of Environmental Sciences*, volume 17, issue 3: 486-491. 2019.

[4] G. Indrajid and D.K. Sari. "Utilization of Landsat 8 OLI Imagery for Identification of B3 Lime Polluted Agricultural Land with Spectral Mixture Analysis Method", *Indonesian Journal of Geospatial*, *5* (2). 2018.

[5] A. Ramadianto and T. Gunawan. "Utilization of Remote Sensing Imagery for Water Quality Mapping of Jatiluhur Reservoir, Purwakarta Regency, West Java Province", 2014.

[6] X. Sun, Y. Zhang, K. Shi, N. Li, W. Wang, X. Huang, B. Qin. "Monitoring Water Quality Using Proximal Remote Sensing Technology", *Sci. Total Environ*. 803, 149805, 2022.

[7] P. Brezonik, K.D. Menken, M. Bauer. "Landsat-based Remote Sensing of LakeWater Quality Characteristics, Including Chlorophyll and Colored Dissolved Organic Matter (CDOM)". *Lake Reserv. Manag*. 21, 373–382, 2005.

[8] M. Bonansea, M.C. Rodriguez, L. Pinotti, S. Ferrero. "Using multi-temporal Landsat imagery and linear mixed models for assessing water quality parameters in Río Tercero reservoir (Argentina)". *Remote Sens. Environ*. 158, 28–41, 2015.

[9] T. Rajaee, S. Khani, M. Ravansalar. "Artificial intelligence-based single and hybrid models for prediction of water quality in rivers: A review". *Chemom. Intell. Lab. Syst*. 200, 103978, 2020.

[10] K. Sudheer, I. Chaubey, V. Garg. "Lake Water Quality Assessment from Landsat Thematic Mapper Data Using Neural Network: An Approach to Optimal Band Combination Selection", *J. Am. Water Resour. Assoc*. 42, 1683–1695. 2006.

[11] H. Lou, Y. Zhang, S. Yang, X. Wang, Z. Pan, Y. Luo. A New Method for Long-Term River Discharge Estimation of Small- and Medium-Scale Rivers by Using Multisource Remote Sensing and RSHS: Application and Validation". *Remote Sens*. 14, 1798. 2022.

[12] S. Wang, J. Li, B. Zhang, Z. Lee, E. Spyrakos, L. Feng, C. Liu, H. Zhao, Y. Wu, L. Zhu, et al. "Changes of Water Clarity in Large Lakes and Reservoirs Across China Observed from Long-Term MODIS", *Remote Sens. Environ*. 247, 111949, 2020.

[13] A. Najah Ahmed, F. Bint Othman, H. Abdulmohsin Afan, R. Khaleel Ibrahim, C. Ming Fai, M. Shabbir Hossain, M. Ehteram, A. Elshafie. "Machine Learning Methods for Better Water Quality Prediction", *J. Hydrol*. 578, 124084, 2019.

[14] A. Parsaie, A.H. Nasrolahi, A.H. Haghiabi. "Water Quality Prediction Using Machine Learning Methods", *Water Qual. Res. J.* 53, 3–13, 2018.

[15] U.S. Abobakr Yahya, A.N. Ahmed, F. Bint Othman, R.K. Ibrahim, H.A. Afan, A. El-Shafie, C.M. Fai, M.S. Hossain, M. Ehteram, A, Elshafie, A. "Water Quality Prediction Model Based Support Vector Machine Model for Ungauged River Catchment under Dual Scenarios". *Water*, 11, 1231, 2019.

[16] M. Gad, L. Hou, M. Cao, B. Adyari, L. Zhang, D. Qin, C.P. Yu, Q. Sun, A. Hu. "Tracking Microeukaryotic Footprint in A Peri-Urban Watershed, China Through Machine-Learning Approaches". *Sci. Total Environ*. 806, 150401, 2022.

[17] V. Sagan, K.T. Peterson, M. Maimaitijiang, P. Sidike, J. Sloan, B.A. Greeling, S. Maalouf, C. Adams. "Monitoring Inland Water Quality Using Remote Sensing: Potential and Limitations of Spectral Indices, Bio-Optical Simulations, Machine Learning, And Cloud Computing". *Earth-Sci. Rev.* 205, 103187, 2020.

[18] X. Shi, Z. Chen, H. Wang, D.Y. Yeung, W.K. Wong, W.C. Woo. "Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting", In *Proceedings of the 28th International Conference on Neural Information Processing Systems (NIPS'15), Montreal, QC, Canada*, 7–12 December 2015.

[19] M. Claverie, J. Ju, J.G. Masek, J.L. Dungan, E.F. Vermote, J.C. Roger, S.V. Skakun, C. Justice. "The Harmonized Landsat and Sentinel-2 surface reflectance data set", *Remote Sens. Environ*. 219, 145–161, 2018.

[20] F. Pu, C. Ding, Z. Chao, Y. Yu, X. Xu. "Water-Quality Classification of Inland Lakes Using Landsat8 Images by Convolutional Neural Networks", *Remote Sens*. 11, 1674. 2019.

[21] University of Twente. March 19, 2018. "ILWIS – Remote Sensing and GIS Software". https://www.itc.nl/ilwis/. Retrieved December 04, 2023.

[22] E. Elsayed and D. Fathy, "Semantic Deep Learning to Translate Dynamic Sign Language", *International Journal of Intelligent Engineering and Systems*, vol. 14, no. 1, pp. 316–325, Feb. 2021, doi: 10.22266/ijies2021.0228.30.

[23] E. A. Mahareek., et al. "Detecting Anomalies in Security Cameras with 3D-Convolutional Neural Network and Convolutional Long Short-Term Memory", *International Journal of Electrical and Computer Engineering (IJECE)*. Vol. 14, No. 1, February 2024, pp. 993-1004. ISSN: 2088-8708, DOI: 10.11591/ijece. v14i1.

[24] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W. Wong, and W. Woo, "Convolutional LSTM network: a machine learning approach for precipitation nowcasting," *Advances in Neural Information Processing Systems*, pp. 802–810, Jun. 2015.

[25] Y. Romdania et al., "Study of the Use of IP, Storet, and CCME WQI Methods in Determining Water Quality Status", *Spatial Journal: Geographic Communication and Information Platform*, 05 July 2018.

[26] Barzegar, R., Aalami, M.T. & Adamowski, J. (2020). Short-term water quality variable prediction using a hybrid CNN–LSTM deep learning model. *Stoch Environ Res Risk Assess* 34, 415–433. https://doi.org/10.1007/s00477-020-01776-2.