Predicting the Extent of Sidoarjo Mud Flow Using Remote Sensing

Wishnumurti Wicaksono* & Sani M. Isa

Computer Science Department, Binus Graduate Program, Bina Nusantara University, 11530, Indonesia
*E-mail: wishnumurti.wicaksono@binus.ac.id

Abstract. The Sidoarjo mud flow in East Java is the result of a natural phenomenon in which hot mudflow occurs due to volcanic activity. The Sidoarjo mud flow resulted in a considerable ecological disaster in the area. In this study, by using the Modification of Normalized Difference Water Index (MNDWI) technique we measured the extension of the mudflow area from 2013 to 2020 using Landsat 8 satellite data imagery. This study is meant to predict the extension of the mud flow area in the research site by comparing regression and neural network techniques in order to find the best approach. The RPROP MLP neural network technique was used to predict the Sidoarjo mud-flowing area in 2021 to 2025. Surprisingly the results of these calculations showed that the RPROP MLP neural network with three hidden layers and 20 neurons performed the best, with an R square value for training of 0.77915565 and for testing of 0.78321550.

Keywords: *MNDWI; neural network; regression; remote sensing; Sidoarjo mudflow.*

1 Introduction

Since May 2006 the phenomenon of overflowing hot mudflows has been occurring in Porong district, Sidoarjo, East Java, Indonesia [1]. As a result, an area of mud mixed with water was formulated however it has been enclosed by an artificial embankment. To prevent the mudflow from passing through the embankment, excess mud was drained into the Porong river [1]. As The mud comprised heavy metals, it could risk the surrounding environment being profaned, contaminating aquatic ecosystems and threatening the survival of aquatic organisms (biosecurity) as well as tainting the fishery products in the surrounding ecosystem [2].

Hot mudflow or sedimentary volcanism is a process that encourages extrusion to the topographical surface of material originating from deeply buried sediments, such as asphalt, rock fragments or blocks, water and gas, which can be divided into several dominant fractions. This mechanism is usually in line with traps of hydrocarbons in the bowels of the earth [3]. Koesoemadinata, an expert in this field, describes mud volcanoes as an extrusion on the surface of the earth in the

Received July 6th, 2021, 1st Revision December 30th, 2021, 2nd Revision March 4th, 2021, Accepted for publication March 30th, 2022.

form of clay or mud by which the morphology forms a cone, augmented by drainage and driven by a strong gas flow. Often the excretory gaseous material is followed by an explosion and burning, making the extrusion looks very much like a magmatic volcano [4].

The aim of this study was to use up-to-date remote sensing technology available as an aid to distinguish the mud-flowing area and the water surface area within the Sidoarjo mudflow embankment. In addition, the surface areas of mud and water in the Sidoarjo Mud embankment were calculated. Hence, multiple regression and neural network techniques were compared in predicting the surface area of mud and water in the Sidoarjo Mud embankment. The best prediction method was then selected to make predictions for the next five-year period.

According to Xu [5], it is possible to calculate a large area of water body using remote sensing techniques by utilizing satellite imaging data and the Modification of Normalized Difference Water Index (MNDWI) algorithm. MNDWI is an algorithm developed from the Normalized Difference Water Index (NDWI) [5], which is an algorithm exerting the middle infra-red (MIR) and green band in satellite image data to calculate an area of water [5]. In this research of distinguishing the water area and the mud area, the MNDWI technique was applied on remote sensing data. To predict the surface areas of water and mud in 2021 to 2025, a neural network technique was used.

2 Literature Review

Remote sensing can be broadly defined as gathering and interpreting information about a certain object in a certain area. It is a technique for attaining information about a geophysical object or phenomenon without making physical contact with the object, generally known as scanning the earth's surface by satellites or high-flying aircraft. Hence the data obtained are processed and interpreted [6-9].

According to Xu [5], the application of NDWI on water body areas containing built-up land faces obstacles since the water index obtained is mixed with built-up land noise. As a matter of fact MNDWI has the following formula expression shown below:

$$\text{MNDWI} = (\textit{Green} - \text{MIR}) / (\textit{Green} + \text{MIR}) \quad (1)$$

MIR is the middle infra-red region, lies in band 6 (SWIR 1), while Green is situated in band 3 of the Landsat 8 satellite images.

The application of MNDWI to extract water data from satellite images has been implemented in several papers. Du [10] emphasized water bodies mapping from

Sentinel-2 Imagery with MNDWI at a 10-m spatial resolution. The research area was located in the coastal area of Venice, Italy, covering 550 km^2 consisting of shallow waters, coastal swamps, islands, and canals, all of which are sensitive to changes in water level. Over a hundred-year period there has been a replenishment of 23 cm in water level, resulting in an expansion of the water surface area. The research aimed to produce 10-m MNDWI from Sentinel-2 images using the images that had been enhanced with image sharpening in the Short Wave Infra-Red (SWIR) band. Additionally, the 10-m performance produced by MNDWI was examined by mapping the water bodies and comparing it with 10-m NDWI and 20-m MNDWI of existing water maps [10].

Interestingly, the research conducted by Liang and Yan [11] was held at Hongjian Lake, Shaanxi Province, China. Home of the relict gull bird, a protected animal. During that period, Hongjian Lake is threatened by the continuous shrinkage. The research utilized the MNDWI approach with 336 Landsat remote sensing images taken from 1988-2015 to extract the area of water. The extraction outcomes were then correlated with various factors that could affect the size of the water in the lake.

The MNDWI method was also used by Li, *et al.* in [12]. This research examined water bodies cited in a Mediterranean coastal lagoon and its surroundings. It is a crucial location due to an Important collection of habitats for several species of wild animals facing significant threats from agricultural human activities. This study also monitored the Spatio-temporal dynamics of surface water patterns in the Narbonne Regional Nature Park based on Landsat data (2002-2016) using NDVI, NDWI and MNDWI. This type of research is valid or coastal resort management at the surface water patch level. It can also be easily applied to similar locations elsewhere with basic information for environmental management and protection provided for such locations.

3 Research Method

3.1 Study Area

Geographically, the Sidoarjo mudflow is located 112° 42' east longitude and 07° 31' south latitude, in Porong district, Porong sub-district in the southern part of Sidoarjo Regency, about 12 km south of Sidoarjo city. This sub-district is bordered by Gempol district (Pasuruan regency) to the south (Figure 1).

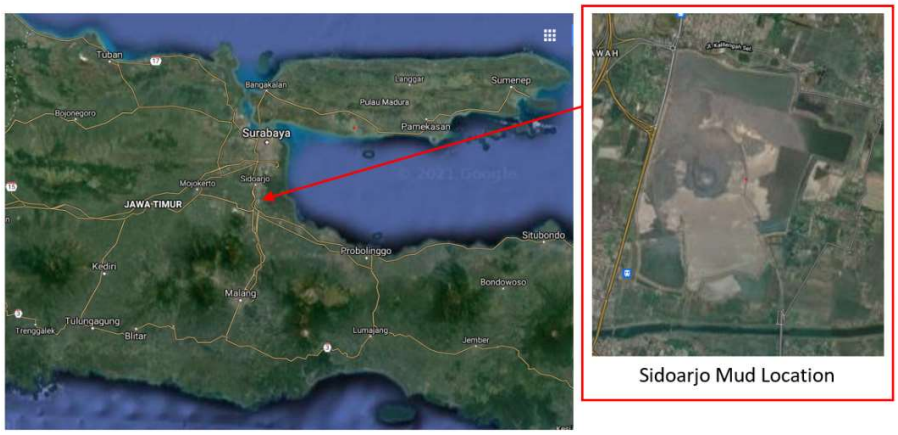

Figure 1 Sidoarjo mud flow location [13].

3.2 Research Method

The research methods applied were data collection, satellite image data pre-processing, determination the area of interest, data classification image, prediction and area calculation, result analysis and evaluation (Figure 2).

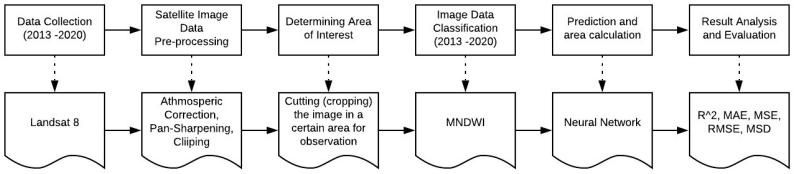

Figure 2 Research methodology.

3.3 Data Collection

The data used in this study were taken from the Landsat 8 satellite originating from https://earthexplorer.usgs.gov/. Furthermore, the data that were used to measure the area of mud distribution covered the period of 2013 to 2020. The reason for obtaining the data in 2013 was because the Landsat 8 satellite started operating on February 11, 2013, therefore the available data existed since 2013 [14]. Another reason is that the data of the Landsat 7 satellite cannot be processed due to the large distortion of the study area (Sidoarjo mudflow). Furthermore, the data that was used in this study is shown in Table 1.

Table 1 The Landsat 8 data used for this study.

Landsat Product Identifier	Acquisition Date
LC08_L1TP_118065_20131101_20170429_01_T1	2013-11-01
LC08_L1TP_118065_20140917_20170419_01_T1	2014-09-17
LC08_L1TP_118065_20151123_20170401_01_T1	2015-11-23
LC08_L1TP_118065_20160906_20170321_01_T1	2016-09-06
LC08_L1TP_118065_20170909_20170927_01_T1	2017-09-09
LC08_L1TP_118065_20181030_20181115_01_T1	2018-10-30
LC08_L1TP_118065_20191001_20191018_01_T1	2019-10-01
LC08_L1TP_118065_20201003_20201015_01_T1	2020-10-03

3.4 Satellite Image Data Pre-processing

Satellite data recorded by sensors is strongly influenced by atmospheric conditions, the angle of data recorded from the sensor, and the time of data recorded. This causes data of the satellite image to have biased information values that must be rectified. Thus, a rectification is needed to eliminate this bias [7-9]. Atmospheric correction was conducted using the Dark Object Subtraction (DOS) technique, which searches each band for the darkest pixel value. Assuming that dark matter does not reflect light, any value greater than zero must have resulted from the scattering of the atmosphere. Scattering is removed by subtracting this value from each pixel in the image [15].

Based on CCRS (2014) [16], any remote sensing image, regardless of whether the image is obtained using a satellite active multispectral scanner, an in-aircraft photography system, or another platform/sensor combination, will have various geometric distortions. Steps to rectify these issues are carried out by preprocessing the satellite data using the metadata (MTL) file [17]. Thus, Landsat MTL files contain beneficial information about data processing, while values important for enhancing Landsat data (such as conversion to reflectance and radiance) are also included in this file type.

Upon the completion of atmospheric rectification, the image is sharpened with the pan-sharpening technique. As a matter of fact, pan-sharpening is a combination of spectral information from the multispectral band (MS), which has a lower spatial resolution with the panchromatic band, which has a higher resolution. The main goal is to produce multispectral images with the spatial resolution from a sharper panchromatic band [18]. The figure shown in 3(a) is the outcome of preprocessing satellite data of the Sidoarjo mud area.

3.5 Determining Area of Interest

Determining the area of interest is held by cutting (cropping) the image in a certain area to be observed, making it easier to analyze the image according to the scope taken in this study (Sidoarjo mudflow area). In the image of data processing, the whole image is not entirely used. The area of interest was determined based on the area inside the Sidoarjo mudflow embankment (Figure 3(b)).

3.6 Image Data Classification

Image classification was conducted by using the Modification of Normalized Difference Water Index method, an algorithm developed based on the Normalized Difference Water Index [5]. The results of classification using the QGIS application are shown in Figure 3(c). According to Xu (2006) [5], the index 0 obtained from the results of classification using the MNDWI method can be used as the threshold value for determining water areas. For water surfaces, the index value is higher (above 0). For index values below 0, the water surface in the area decreases or the water surface is mixed with solid objects (Figure 3(d)).

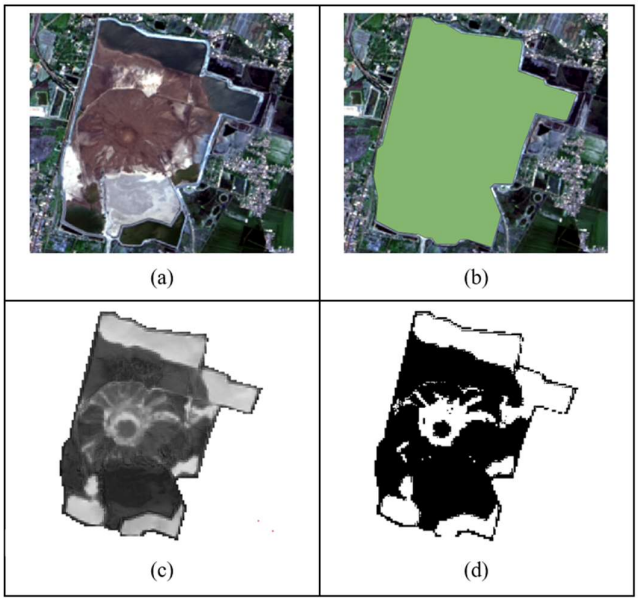

Figure 3 Satellite image processing: (a) Satellite data preprocessing, (b) area of interest, (c) MNDWI Sidoarjo, (d) mud (black) and water (white) areas.

3.7 Prediction Using Multiple Regression

Multiple regression is an algorithm used to find the relationship between one dependent variable and more than one independent variable. To make predictions in data mining, the multiple regression algorithm can be utilized. The general form of the multiple regression equation is described as follows [19]:

$$Y_{ij} = \beta_0 + \beta_1 X_{ij1} + \beta_2 X_{ij2} + \cdots + \beta_{p-1} X_{ij,p-1} + \varepsilon_{ij} \quad (2)$$

where Y_{ij} is the dependent variable from i to j observations. The independent variable is denoted by $X_{ij1}, X_{ij}, \ldots, X_{ij,p-1}$. $\beta_0, \beta_1, \beta_2, \ldots, \beta_{p-1}$ are parameters, while ε_{ij} are errors.

3.8 Prediction Using Neural Network

The prediction on the area of the Sidoarjo mudflow sediment distribution over a period of several years can be calculated using the RPROP MLP (Resilient Propagation Multi Layer Perceptron) neural network method based on historical data analysis. RPROP MLP is a scheme that performs a direct adaptation of the weighting stages based on local gradient, local information [20]. Analyzing a time series means sorting out past data, turning it into components and then projecting them into the future.

3.9 Numeric Scoring

To evaluate the prediction outcomes in the study, the values of R square and mean square error in the index of each pixel were used. R square was used to interpret the linear relationship between predictor variables and responses in the form of proportions [21].

$$R^2 = \frac{SSR}{SST} \quad (3)$$

SSR (sum of squares regression) is the total square of the difference between the value of the predictive variable and the average value of the target formulated as follows:

$$SSR = \sum_{i=1}^{n} (y_i - \underline{y})^2 \quad (4)$$

By which y_i is the predictive variable and $\underline{y}$ is the average value of the target.

SST (sum of squares total) is the total square of the difference between the value of the target variable and the average value of the target, formulated as follows:

$$SST = \sum_{i=1}^{n} (y - \underline{y})^2 \quad (5)$$

The mean square error (MSE) is the average value of the error squared [21]:

$$MSE = \frac{1}{n}\sum_{i=1}^{n} \quad (Y_i - Y)^2 \tag{6}$$

The variable Y_i is the predicted value, while Y is the actual value, and n is the figure of predicted data.

4 Results and Discussion

In this study, we used the QGIS 3.8.0 Zanzibar application and the SCP 7.3.2 Matera plugin to preprocess the Landsat 8 data from 2013 to 2020. All Landsat 8 data from 2013 to 2020 were then processed using Xu's [5] MNDWI formula, the outcomes of which were then converted to the pixel dataset.

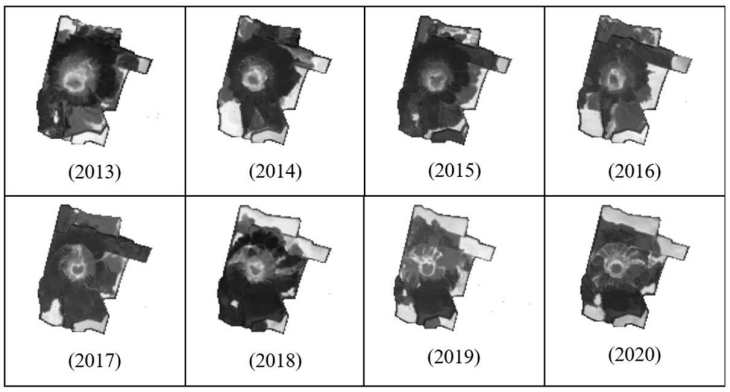

Figure 4 MNDWI calculation of Sidoarjo mudflow area from 2013 to 2020.

The MNDWI pixel index dataset from 2013 to 2020 consisted of 27,692 records and was randomly divided into 80 percent for training and 20 percent for testing. The training and testing data were then used as input for modeling using multiple regression and neural network (Figure 5).

Table "default" - Rows: 27692 | Spec - Columns: 10 | Properties | Flow Variables

Row ID	pixel_x	pixel_y	index_2013	index_2014	index_2015	index_2016	index_2017	index_2018	index_2019	index_2020
Row0	1	41	-0.398	-0.494	-0.439	-0.424	-0.362	-0.289	-0.367	-0.273
Row1	1	42	-0.424	-0.471	-0.419	-0.452	-0.415	-0.306	-0.376	-0.297
Row2	1	43	-0.409	-0.405	-0.406	-0.417	-0.44	-0.261	-0.339	-0.346
Row3	1	44	-0.393	-0.436	-0.341	-0.452	-0.483	-0.261	-0.398	-0.378
Row4	1	45	-0.387	-0.493	-0.41	-0.44	-0.466	-0.288	-0.377	-0.32
Row5	1	46	-0.408	-0.513	-0.489	-0.444	-0.438	-0.261	-0.367	-0.307
Row6	1	47	-0.363	-0.475	-0.47	-0.406	-0.418	-0.188	-0.32	-0.292
Row7	1	48	-0.341	-0.463	-0.419	-0.361	-0.434	-0.183	-0.313	-0.265
Row8	1	49	-0.327	-0.47	-0.393	-0.38	-0.435	-0.19	-0.267	-0.173
Row9	1	50	-0.359	-0.437	-0.418	-0.355	-0.419	-0.206	-0.266	-0.191
Row10	1	51	-0.375	-0.468	-0.41	-0.364	-0.435	-0.179	-0.312	-0.239

Figure 5 Sample input MNDWI pixel index dataset for multiple regression and neural network modeling.

According to Ye [22], regression and neural network modeling techniques can be used to make predictions, in which both models can produce predictive output values based on input from more than one independent variable.

In this research, the MNDWI pixel index dataset was processed using the Multiple Linear Regression to set up a prediction model. In terms of calculating the predictions, Knime 3.7.2 can be used (Figure 6). The input variable was the MNDWI pixel index data taken in between 2013 to 2019, used to predict the MNDWI pixel index values for 2020.

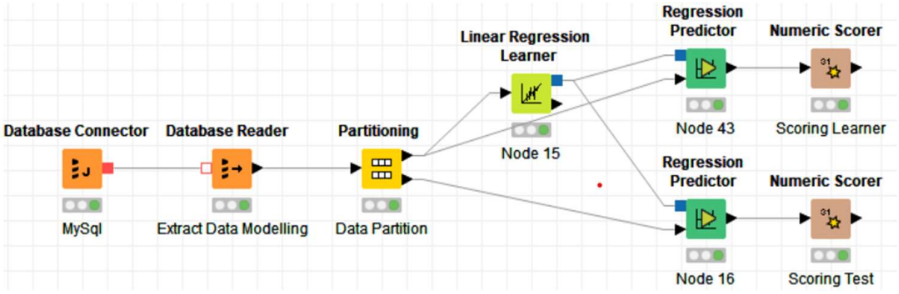

Figure 6 Modeling calculation using Linear Regression Learner.

The MNDWI index pixel dataset was also processed using the RPROP MLP neural network to set up a prediction model (Figure 7).

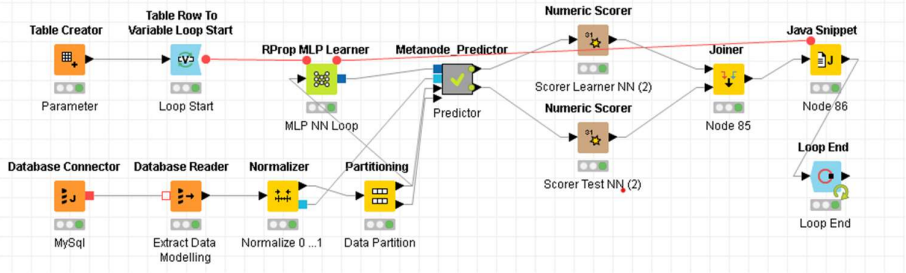

Figure 7 Modeling calculation using RPROP MLP neural network.

To get accurate predictions from the calculations of the multiple regression and neural network modeling, R square and mean square error were used (Table 2). From the neural network and multiple regression models run, the highest R square value was produced by the neural network model with 3 hidden layers and 20 neurons: 0.779156 for training and 0.783216 for testing. The lowest value of MSE in the model with 3 hidden layers and 20 neurons was 0.018508 for training

and 0.018052 for testing. Hence, to make predictions, the neural network with 3 hidden layers and 20 neutrons was used.

Table 2 R square and MSE.

		R^2		MSE	
Layer	**Neurons**	**Learner**	**Test**	**Learner**	**Test**
1	10	0.72704	0.734819	0.022875	0.022082
2	10	0.746981	0.752894	0.021204	0.020576
3	10	0.751999	0.755324	0.020783	0.020374
1	20	0.732128	0.738529	0.022449	0.021773
2	20	0.77059	0.774455	0.019226	0.018781
3	**20**	**0.779156**	**0.783216**	**0.018508**	**0.018052**
1	30	0.736458	0.742173	0.022086	0.021469
2	30	0.776663	0.781042	0.018717	0.018233
3	30	0.759720	0.761166	0.020136	0.019888
Multiple regression		0.720414	0.723490	0.023371	0.023260

Figure 8 shows the neural network using input data of MNDWI pixel index values from 2013 to 2019 (yellow nodes) and three hidden layers (green nodes) to produce a predicted output target of MNDWI pixel index values for 2020.

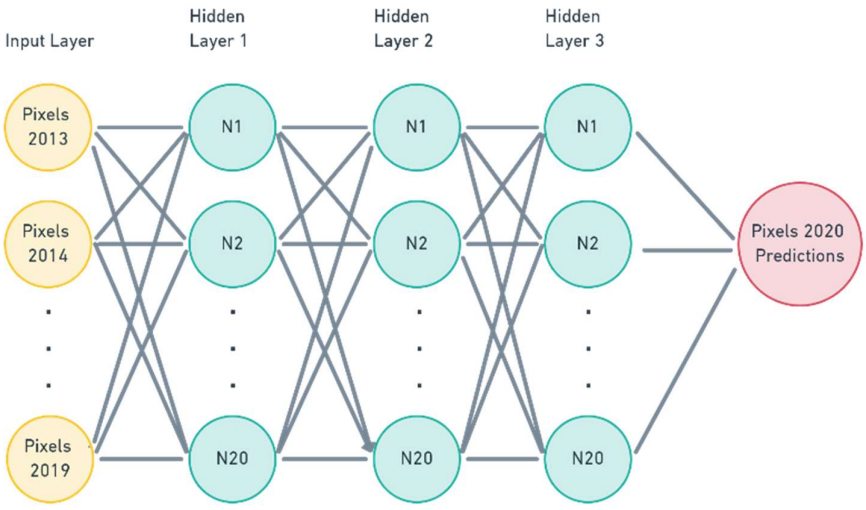

Figure 8 Neural network diagram with three hidden layers and twenty neurons.

Table 3 depicts the evaluation outcomes obtained for the neural network model with 3 hidden layers and 20 neurons using the RPROP MLP algorithm.

Table 3 Neural network R square and MSE.

	Learner	Test
R^2	0.77916	0.78322
mean absolute error	0.09486	0.09388
mean squared error	0.01851	0.01805
root mean squared deviation	0.13604	0.13436
mean signed difference	0.00193	0.00215

This research paper used the neural network model with the highest R square tested and training values and the lowest MSE value, a model with 3 hidden layers and 20 neurons, equipped with the RPROP MLP algorithm to predict the values of the MNDWI index for the next five years.

By using the RPROP MLP neural network model with 3 hidden layers and 20 neurons, the MNDWI index from 2021 to 2025 was obtained. The index was categorized based on the following threshold: if an index value was below 0 it belonged to the mud category and if it was above 0, it was in the water category. Using the outcome of this threshold, the calculation of the mudflow area and the water surface was calculated for the period of 2021 to 2025 (Table 4).

Table 4 Sidoarjo mud and water area.

Year	Mud (m^2)	Water (m^2)
2013	5,242,725	987,975
2014	4,745,250	1,485,450
2015	5,411,925	818,775
2016	4,588,200	1,642,500
2017	5,456,475	774,225
2018	4,532,850	1,697,850
2019	4,385,025	1,845,675
2020	4,151,700	2,079,000
Prediction 2021	4,236,750	1,993,950
Prediction 2022	4,482,675	1,748,025
Prediction 2023	4,366,800	1,863,900
Prediction 2024	4,577,625	1,653,075
Prediction 2025	4,538,700	1,692,000

Figure 9 is the result of the movement of the Sidoarjo mudflow area based on the results of calculations using MNDWI and the neural network method. The mudflow area is highlighted in black while the water area is presented in white.

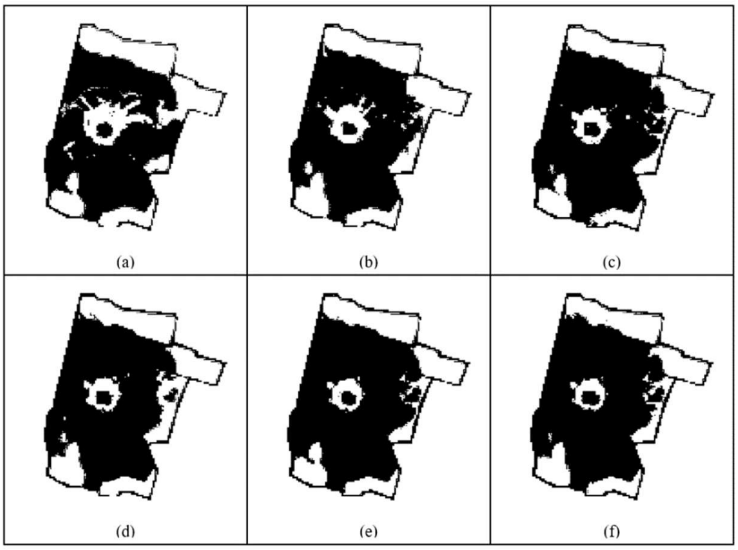

Figure 9 Prediction of Sidoarjo mudflow area: (a) mud area in 2020, (b) mud area prediction in 2021, (c) mud area prediction in 2022, (d) mud area prediction in 2023, (e) mud area prediction in 2024, (f) mud area prediction in 2025.

5 Conclusion

Based on our research on the surface area of the Sidoarjo mudflow using remote sensing imagery in the form of Landsat 8 imagery in the years 2013 to 2020, the classification of mud area and water surface area was conducted. The classified outcome was based on the calculation of the MNDWI value, which resulted in the areas of mud deposits and water surfaces. This study developed a method for predicting the area of mud surface sediment by using the RPROP MLP neural network technique for forecasting.

In future works, research is applied using satellite imagery that has a higher resolution, making the detection of the area of mud and water surface more accurate. Additionally, follow-up research will use other predictive techniques, hence the results obtained can be used as a comparison to gain better prediction techniques.

References

[1] Walhi, *Breaking Corporate Impunity*, 2019. (Text in Indonesian)

[2] Purnomo, T., *Plants Species in Situ Has Resistance to the Lapindo Hot Mud Contaminants*, Paper National Conference, 2009.

[3] Higgins, G.E. & Saunders, J.B., *Mud Volcanoes: Their Nature and Origin*, Basel, Switzerland,1974.

[4] Koesoemadinata, R.P., *Geology Oil and Gas,* 2nd ed. Bandung: Penerbit ITB, 1980. (Text in Indonesian)

[5] Xu, H., *Modification of Normalised Difference Water Index (NDWI) To Enhance Open Water Features in Remotely Sensed Imagery*, International Journal of Remote Sensing, **27**(14), pp. 3025-3033, 2006. DOI: 10.1080/01431160600589179.

[6] Schumann, G.J.P., *Remote Sensing of Floods*, Oxford Research Encyclopedia of Natural Hazard Science, no. March, pp. 1-30, 2017. DOI: 10.1093/acrefore/9780199389407.013.265.

[7] James, R.H.W. & Campbell, B., *Introduction to Remote Sensing*, 717, 2011.

[8] Sabins, F.F. & Ellis, J.M., *Remote Sensing: Principles, Interpretation and Applications*, Long Grove, Illinois: Waveland Press, Inc., 2020.

[9] Sanderson, R., *Introduction to Remote Sensing*, New Mexico State University, 2006, https://faculty.kfupm.edu.sa/crp/bramadan/crp514/readings/7%20-%20Introd_Remote_Sensing_Dr_Sanderson_New_Mexico_State_Univ_38Pages.pdf, (December 2020).

[10] Du, Y., Zhang, Y., Ling, F., Wang, Q., Li, W. & Li, X., *Water Bodies' Mapping from Sentinel-2 Imagery with Modified Normalized Difference Water Index at 10-M Spatial Resolution Produced by Sharpening the SWIR Band*, Remote Sensing, **8**(4), pp. 1-19, 2016. DOI: 10.3390/rs8040354.

[11] Liang, K. & Yan, G., *Application of Landsat Imagery to Investigate Lake Area Variations and Relict Gull Habitat in Hongjian Lake*, Ordos Plateau, China, Remote Sensing, **9**(10), pp. 1-23, 2017, DOI: 10.3390/rs9101019.

[12] Li, Z., Feng, Y., Dessay, N., Delaitre, E., Gurgel, H. & Gong, P., *Continuous Monitoring of the Spatio-Temporal Patterns of Surface Water in Response to Land Use and Land Cover Types in A Mediterranean Lagoon Complex*, Remote Sensing, **11**(12), pp. 1-19, 2019. DOI: 10.3390/rs11121425.

[13] Google Inc, *Google Maps: Sidoarjo*, 2020. https://www.google.com/maps

[14] Masek, J.G., *Landsat 8 Mission Details,* Available at: https://landsat.gsfc.nasa.gov/landsat-8/mission-details, (December 2020).

[15] Chavez, P.S., *Image-Based Atmospheric Corrections – Revisited and Improved*, Photogrammetric Engineering and Remote Sensing, **62**(9), pp. 1025-1036, 1996.

[16] CCRS, Fundamental of Remote Sensing, CCRS, Canada, 2014.

[17] Luca, C., *Semi-Automatic Classification Plugin Documentation Release 5.3.2.1 Luca Congedo*, pp. 54-55, July 2014, 2016, DOI: 10.13140/RG.2.1.1219.3524.

[18] Johnson, B.A., Tateishi, R. & Hoan, N.T., *Satellite Image Pansharpening Using a Hybrid Approach for Object-based Image Analysis*, ISPRS International Journal of Geo-Information, **1**(3), pp. 228-241, 2012. DOI: 10.3390/ijgi1030228.

[19] Neter, J., Kutner, M.H., Nachtsheim, C.J. & Wasserman, W., *Applied Linear Statistical Models*, Chicago, Irwin, 1996.

[20] Riedmiller, M. & Braun, H., *Direct Adaptive Method for Faster Backpropagation Learning: The RPROP Algorithm*, IEEE International Conference on Neural Networks, pp. 586-591, 1993. DOI: 10.1109/icnn.1993.298623.

[21] Larose, D.T., *Data Mining Methods and Models*, Wiley-Interscience, 2015.

[22] Ye, N. (n.d.), *Data Mining: Theories, Algorithms, and Examples*, Boca Raton: CRC Press, 2017.