JOURNAL OF APPLIED GEOSPATIAL INFORMATION

Vol 7 No 2 2023

http://jurnal.polibatam.ac.id/index.php/JAGI
ISSN Online: 2579-3608

Combining Two Classification Methods to Predict Jakarta Bay Seabed Type Using Multibeam Echosounder Data

Steven Solikin[1*], Angga Dwinovantyo[2], Henry M. Manik[1], Sri Pujiyati[1], Susilohadi Susilohadi[3]

[1]Department of Marine Science and Technology, Faculty of Fisheries and Marine Sciences, IPB University, Dramaga, Bogor 16680, West Java, Indonesia

[2]Research Center for Oceanography, The National Research and Innovation Agency (BRIN). Jl. Pasir Putih Raya No.1, Jakarta 14430, Indonesia

[3]Research Center for Geological Resources, The National Research and Innovation Agency (BRIN). Jl. M.H. Thamrin No. 8, Jakarta 10340, Indonesia

*Corresponding author e-mail: steven-so@apps.ipb.ac.id

Received: August 22, 2023
Accepted: October 04, 2023
Published: October 04, 2023

Abstract

Classification of seabed types from multibeam echosounder data using machine learning techniques has been widely used in recent decades, such as Random Forest (RF), Artificial Neural Network (ANN), Support Vector Machine (SVM), and Nearest Neighbor (NN). This study combines the two most frequently used machine learning techniques to classify and map the seabed sediment types from multibeam echosounder data. The classification model developed in this study is a combination of two machine learning classification techniques, namely Support Vector Machine (SVM) and K-Nearest Neighbor (K-NN). This classification technique is called SV-KNN. Simply, SV-KNN adopts these two techniques to carry out the classification process. The SV-KNN technique begins with determining test data by specifying support vectors and hyperplanes, as was done on the SVM method, and executes the classification process using the K-NN. Clay, fine silt, medium silt, coarse silt, and fine sand are the five main classes produced by SVKNN. The SV-KNN method has an overall accuracy value of 87.38% and a Kappa coefficient of 0.3093.

Keywords: acoustic multibeam data, Jakarta Bay, multibeam echosounder, seabed classification, support vector machine, SV-KNN

1. Introduction

Underwater detection techniques are increasingly developing nowadays. Underwater detection utilizing acoustic technology (hydroacoustic) is increasingly being adopted for industrial activities and scientific purposes (Solikin et al., 2020). Various examples of underwater detection activities utilizing hydroacoustic technology encompass bathymetry (sea depth) measurements, location detection of fish assemblages (fishing), bottom substrate profiles, underwater communication, positioning, and tomographic networks (Lurton, 2002). One of the interesting subjects that can be examined through this hydroacoustic technology is the detection of the seabed surface.

Hydroacoustic technology, particularly the swath technique, is very convenient as the leading source of information on water bathymetry and seabed morphology (Tegowski et al., 2011). The hydroacoustic is one of the underwater remote sensing approaches that provides a foundation for identifying, classifying, and mapping resources on the seabed (Manik, 2012). Multibeam echosounder system (MBES) -as one of the hydroacoustic technology tools- is adopted for these multiple purposes because of the technological advances presented in this instrument, such as very broad area coverage, high resolution produced, and being able to reach deep and wide areas (Che Hasan et al., 2014).

Nonetheless, the evolution of well-established, consistent, quantitative, and objective methods for the analysis of swath acoustic data lags behind the remarkable strides made in the collection of high-quality acoustic multibeam data. Anderson et al. (2008) highlighted the pressing issue of the classification of acoustic seabed, pinpointing the glaring gap in statistical and objective methodologies. Currently, it remains a common practice to rely on the discerning 'eye' of experts for acoustic data interpretation. However, recent years have seen a growing interest in automated techniques, driven in part by the undeniable advantages of employing objective classification algorithms that mitigate subjectivity. A multitude of methods have been rigorously tested, ranging from random forests (Che Hasan et al., 2012a) and support vector machines (Che Hasan et al., 2012b; Lucieer et al., 2013) to artificial neural networks (Marsh & Brown, 2009; Huang et al., 2013), maximum likelihood classification (Brown et al., 2011), Bayesian decision rules (Simons & Snellen, 2009), and decision trees (Che Hasan et al., 2012). Remarkably, only a limited number of studies have undertaken the crucial task of comparing multiple automated seabed mapping techniques, with just five such studies published thus far (Che Hasan et al., 2012; Ierodiaconou et al., 2011; Diesing et al., 2014; Snellen et al., 2018; Cui el al., 2020). This indicates the ongoing need for further exploration and refinement of automated methodologies to advance our understanding of the acoustic seabed classification.

In this study, we exert the combination of two classification methods (SVM and NN) to predict seabed type in Jakarta Bay based on acoustic MBES data and seabed samples. This combination was named SV-KNN as a combination of the two recent classification methods. We try to compare two methods (SVM and this combination method) in terms of the thematic accuracy and spatial depiction of predicted seabed sediment types.

2. Methods

2.1 Study Area and Survey Information

The study area covers a small area in Jakarta Bay, specifically on the north side of the reclamation island named G-Island (Figure 1).

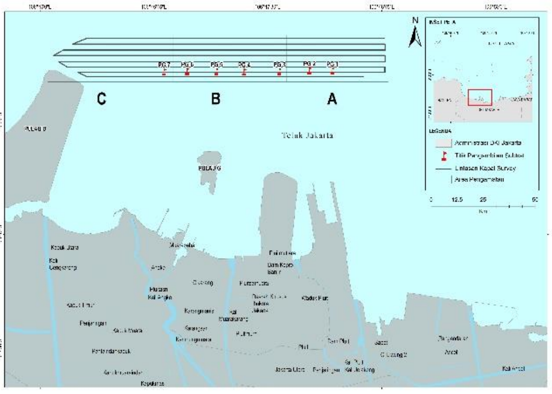

Figure 1. Survey area in The G-Island, Jakarta Bay. The data acquisition path using the MBES instrument is shown by the black line. The survey area is divided into three areas, namely areas A, B, and C

A marine survey was conducted in the area in 2016 (Solikin et al., 2020) to collect acoustic multibeam data and sediment samples. A fisherman vessel of 12 x 2.5 meters was used for the survey. There are 8 main survey lines and the space between lanes ranges from ±100 meters.

During the survey, the SIMRAD Kongsberg EM 2040 multibeam echo-sounder was used to collect both bathymetry and backscatter data in shallow waters, from 9 to 17 m depth with a vertical resolution of 1 cm. Teledyne TSS DMS-05 motion sensor was used to rectify the MBES data for vessel movement, bow aberration, and GPS delay correction. It has a slant and bobbing accuracy of 0.05°. Veripos DGPS was used to determine the position, and at a confidence level of 95%, it had a horizontal precision of 0.13 meters and a vertical accuracy of 0.32 meters. Automatic Data Logging (ADL) Hydro-Pro and Seafloor Information System (SIS) software were used as navigation systems and data collecting, respectively.

2.2 Sediment Sample Acquisition

In addition to the MBES data, sediment samples were also gathered to be used as validation data for the classification of sediments based on their geo-acoustic characteristics. Seven sediment samples were taken, and they were distributed at random around the study area. Sediment samples were taken using grab samplers. They are only taken from the water's bottom surface due to the grab sampler's restricted ability to reach deeper into the sediment layer. The procedure started with the collection of sediment samples, placing them in plastic samples, and a laboratory examination of the particle size.

Wet sieving was used to separate the sediment's grain size depending on the grain size fraction during the analysis of the sediment sample. Each fraction was divided into the following according to Shephard's triangle (Shepard, 1954) (Figure 2) where each fraction was divided into:

1. Gravel (gravel) fraction: a mixture of rock and gravel material
2. Sand fraction (sand): a mixture of fine sand to coarse sand material
3. Mud fraction: a mixture of clay and silt material

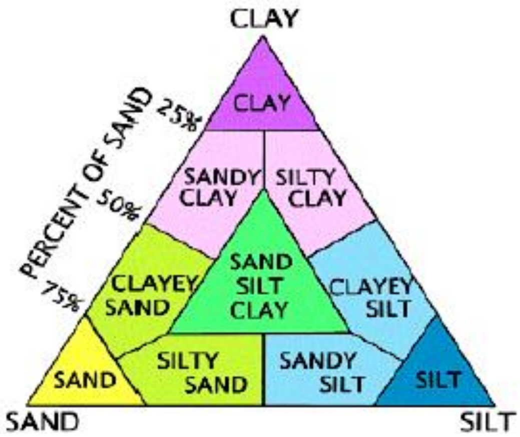

Figure 2. Shephard's triangle diagram
Source: Shepard (1954)

Secondary data was also collected from the Pusat Hidro-Oseanografi TNI AL (Pushidrosal) based on data from 2016 along with the sediment sample data to identify the types of sediment in Jakarta Bay waters that had been processed by Sediment Analysis Tools in CARIS 9.0 software. To add validation to the MBES data obtained, this secondary data was placed next to the original sediment sample data. According to data from the Indonesian Navy Hydrography and Oceanography Center from 2016, clay predominates in the waters of Jakarta Bay (Table 1).

Table 1. Number of points and sediment types in Jakarta Bay

Number	Sediment type	Number of points
1	Clay	3276
2	Clayey sand	14
3	Coarse silt	62
4	Fine silt	25
5	Medium silt	16
6	Sandy clay	15
7	Sandy mud	10
8	Sandy silt	89
9	Silty clay	13
10	Very fine sand	1

Source: Pushidrosal (2016)

2.3 Support Vector Machine K-Nearest Neighbor Classifications

Nearest Neighbor (NN) classification is one of the most popular classification methods to use (Wu & Kumar 2009). The NN algorithm performs classification based on the similarity of data with other data (Tan et al., 2006). SV-KNN classification is a combination of two methods, namely K-NN and Support Vector Machine (SVM). The advantage of this method is that it can reduce a large amount of data and improve classification results (Srisawat et al., 2006).

According to Prasetyo (2014), SV-KNN uses a Support Vector Machine (SVM) to obtain a support vector (SV) from the initial data to classify the data. Therefore, the SV obtained is a representation of the entire data, but the SV obtained from the SVM is a representation for each class in the high dimension (feature space) and some of the SV may not be suitable as input data for the K-NN classification.

To overcome this problem, it is necessary to re-examine the contribution of SV to the classification in the initial dimensions before using it to make classification predictions. The weight (w) is used to determine the contribution of each SV. SV with a higher weight becomes more reliable and will have a greater influence on K-NN predictions.

The weighting process begins by providing an algorithm to partition all SVs into pre-defined clusters (grouping data). After the data is grouped, each instance in the cluster is given a weight according to the proportion of the class labels of the instances in the cluster. The sample weight xi expressed by wi is given in Equation (1):

$$w_i = \frac{n(class(x_i))}{Total} \qquad (1)$$

i = 1, 2, …, m (m is the sum of all SVs), while n(class(xi)) is the number of samples in the cluster that have the same class label as sample xi. Total is the total number of samples in the cluster. The prediction process is carried out using the classic K-NN algorithm and only the weighted SV which is calculated in the previous process is used as K-NN input data. When data is entered, K-NN looks for the nearest K number of instances of the SV set. The classification framework using the SV-KNN model is shown in Figure 3.

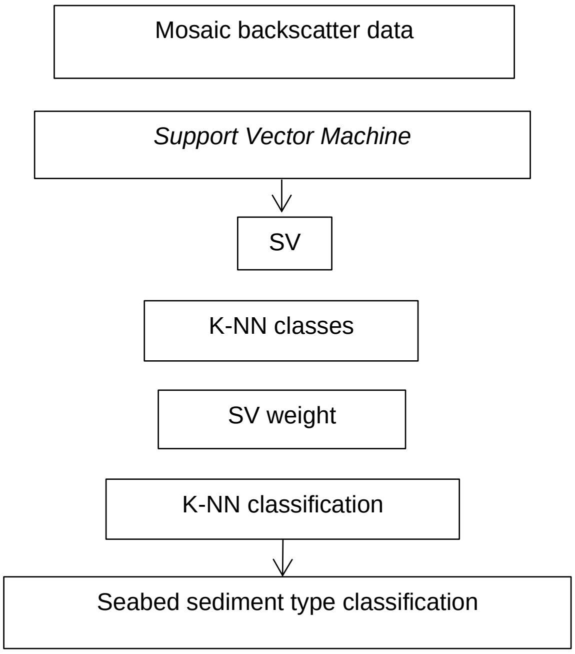

Figure 3. SV-KNN classification framework

2.4 Data and Model Correlation

The accuracy level of the classification models with ground truth data was compared using a statistical examination of the Kappa coefficient. The Kappa coefficient was invented by Cohen (1960). The Kappa coefficient, which was developed for use in remote sensing research in the early 1980s (Congalton & Mead, 1983; Congalton, 1991), has proven to be a valuable test for determining the accuracy of object classification results. The level of agreement between two evaluators when classifying items into categories, as well as the differences in agreement between novel techniques and old methods, can be analyzed using the Kappa coefficient. The equation to calculate the Kappa coefficient is Equation 2.

$$\kappa = \frac{N\sum_{i=1}^{r} X_{ii} - \sum_{i=1}^{r} X_{i+}X_{+i}}{N^2 - \sum_{i=1}^{r} X_{i+}X_{+i}} \qquad (2)$$

where

κ : Kappa coefficient;
N : number of observations;
Xii : observation in i-th row i-th column;
Xi+ : marginal total in i-th row;
X+i : marginal total in i-th column.

A simpler equation is shown in Equation (3):

$$\kappa = \frac{p_0 - p_e}{1 - p_e} \qquad (3)$$

where

p_0= accuracy of the agreement observed, $\frac{\sum X_{ii}}{N}$,

p_e= estimation of chance of agreement, $\frac{\sum X_{i+} X_{+i}}{N^2}$.

The Kappa coefficient value classification classifies class results into 5 classes: bad, fair, moderate, good, and very good (Altman, 1991).

Table 2 provides an interpretation of the types of kappa coefficient value distribution.

Table 2. Classification of closeness of agreement Kappa coefficient value

κ-value	***Strength of agreement***
< 0.20	*Poor*
0.21 – 0.40	*Fair*
0.41 – 0.60	*Moderate*
0.61 – 0.80	*Good*
0.81 – 1.00	*Very good*

Source: Altman (1991)

3. Results and Discussion

The classification model developed in this study is a combination of two machine learning classification techniques, namely Support Vector Machine (SVM) and K-Nearest Neighbor (K-NN). This classification technique is called SV-KNN. Simply put, SV-KNN adopts these two techniques to carry out the classification process. The SV-KNN technique begins with determining the test data by determining the support vector and hyperplane, as is done in the SVM method, and executing the classification process using the K-NN technique.

The initial stages are the same as the stages in SVM is reducing data, determining training data, and determining test data, until finally determining the class used in classification. Similar to the SVM technique, the SV-KNN technique also produces 5 main classes, namely clay, fine silt, medium silt, coarse silt, and fine sand.

After going through the initial stages, for the classification process, another classification technique is used, namely K-NN. The basic difference between K-NN and SVM in carrying out the classification process is the determination of the number of K-neighbors used. In contrast to SVM which does not use neighboring factors in the classification process, the K-NN technique makes neighboring factors a consideration in determining the classification of a point. In this study, after repeated iterations, the value of K as a neighbor is obtained, which is used as a consideration of 3 iterations.

Figure 4 displays labels for clay sediments in dark blue, fine silt in light blue, medium silt in green, coarse silt in orange, and fine sand in red. The SV-KNN classification technique was also examined for accuracy to evaluate how well the classification model matched the field data and how accurate it was. Similar to other categorization procedures, this was carried out. The accuracy test employed in this method is the Kappa coefficient test, which was also utilized in the previous classifications. The Kappa test-derived overall accuracy value for the SV-KNN algorithm is 0.87, with a 95% confidence level agreement of chance value of 0.82. The accuracy number for classification is quite high at 0.87, which indicates that this method was successful in correctly classifying the data for 87.38% of the data.

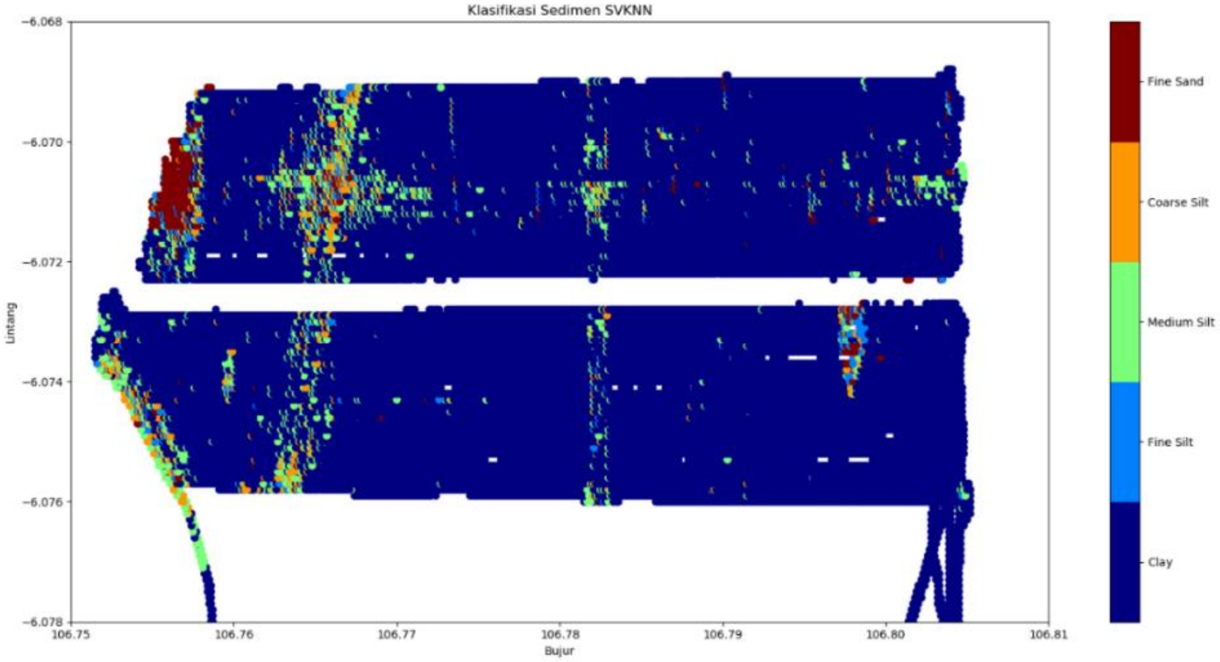

Figure 4. Classification of sediment types in the G-Island waters using the SV-KNN method

This study's Kappa coefficient value is 0.31. This value is classified into the fair class of the strength of agreement variable. The producers' accuracy and the user's accuracy numbers are also computed as part of this method's accuracy test. The producer's accuracy is classification accuracy as viewed from the producer's perspective. On the other side, the user's accuracy is the classification accuracy as perceived from the perspective of the map user, not the map maker.. It demonstrates how frequently the classification results are displayed accurately spatially or the likelihood that a benthic habitat in an area is categorized correctly. The user's accuracy demonstrates how frequently the situation depicted on the map corresponds to real conditions. Table 3 lists the classifications derived from the SVM classification results along with the producer's and user's accuracy values for each class.

The values of the producer's accuracy and the user's accuracy are not equal, as shown in Table 3. For instance, the producer's accuracy value for the clay class in this study's application of the SV-KNN method is 92.98%, whereas the user's accuracy value is 98.61%. As many as 98.61% of the points labeled "clay" actually fit the categorization for clays, indicating that 92.98% of the clay classes were correctly designated as "clay." The remaining four classes were incorrectly assigned to a total of 1.39% of the points. Naturally, this outcome is classified as having a rather high accuracy value. This is a consequence of the clay class predominates in the study region.

The producer's accuracy value for the coarse silt class is 37.10%, whereas the user's accuracy value is 16.08%. This shows that just 16.08% of areas classified as "coarse silt" actually matched into the category of coarse silt, whereas 37.10% of the coarse silt class were correctly identified as distinct. A total of 83.92% of the points (clay, fine silt, medium silt, and fine sand) were incorrectly assigned to other groups.

We compare this result with the previous research (Solikin et al., 2020), the SVM method produces high accuracy, although some classes cannot be logged. The level of accuracy achieved by this method is 80.25%. This high accuracy value is evidence of the success of SVM in classifying the most dominant class, namely clay. The Kappa coefficient value of this method is 0.2031. In the categorization of the closeness of the agreement, the value is classified into a fair or moderate class. SVM still necessitates to be tested in regions with a high degree of heterogeneity to prove the level of accuracy and closeness of agreement that has been reached so far. Meanwhile, the performance shown by the SV-KNN method is not much different from the SVM method, but there is progress in this method. This high accuracy value must be re-tested in different areas and with different instruments. It is intended to see the true potential of this method.

Table 3. Producer's accuracy and user's accuracy values from the class classification results from the SV-KNN method

Classes	Number of points classified correctly	Number of reference points	Number of classification points	*User's accuracy* (%)	*Producer's accuracy* (%)
Clay	3046	3276	3089	98.61	92.98
Fine Silt	2	25	45	4.44	8.00
Medium Silt	4	16	239	1.67	25.00
Coarse Silt	23	62	143	16.08	37.10
Fine Sand	1	1	5	20.00	100.00

4. Conclusions

The SVM method produces an accuracy rate of 80.25% with a Kappa coefficient of 0.2031. The SV-KNN method produces an accuracy rate of 87.38% with a Kappa coefficient value of 0.3093. From the accuracy level and Kappa coefficient of these two methods, it can be determined that the SV-KNN method produces the best classification process compared to the SVM. Based on the classification results and the accuracy value produced by the SV-KNN method, it can be concluded that this method is suitable for use as a machine learning-based classification method that can be applied in classifying seabed substrate types. The classification process is not too tough and the performance is quite good, making the SV-KNN method an alternative method in addition to the existing machine learning-based classification methods for classifying seabed sediment types.

Acknowledgments

The authors would like to thank Ministry of Research, Technology, and Higher Education of The Republic of Indonesia for funding this research through PMDSU Batch II scheme and we also would like to thank Pusat Hidro-Oseanografi TNI AL (Pushidrosal), for the collaboration in the survey activity and providing the secondary data in Jakarta Bay.

References

Altman DG. 1991. Practical statistics for medical research. London: Chapman and Hall.

Anderson JT, Van Holliday D, Kloser R, Reid DG, Simard Y. 2008. Acoustic seabed classification: current practice and future directions. ICES J. Mar. Sci. 65, 1004–1011.

Brown CJ, Todd BJ, Kostylev VE, Pickrill RA, 2011b. Image-based classification of multibeam sonar backscatter data for objective surficial sediment mapping of Georges Bank, Canada. Cont. Shelf Res. 31, S110–S119.

Che Hasan R, Ierodiaconou D, Laurenson L. 2012a. Combining angular response classification and backscatter imagery segmentation for benthic biological habitat mapping. Estuar Coast Shelf Sci, 97:1-9.

Che Hasan R, Ierodiaconou D, Monk J. 2012b. Evaluation of Four Supervised Learning Methods for Benthic Habitat Mapping Using

Backscatter from Multi-Beam Sonar. Remote Sens, 4: 3427–3443.

Che Hasan R, Ierodiaconou D, Laurenson L, Schimel A. 2014. Integrating multibeam backscatter angular response, mosaic and bathymetry data for benthic habitat mapping. PLoS ONE 9(5): e97339.

Cohen J. 1960. A coefficient of agreement for nominal scales. Educ Psychol Meas, 20(1): 37–46.

Congalton RG, Mead R. 1983. A quantitative method to test for consistency and correctness of photointerpretation. Photogramm Eng Remote Sensing, 49(1): 69-74.

Congalton RG. 1991. A review of assessing the accuracy of classifications of remotely sensed data. Remote Sens Environ, 37(1): 35-46.

Cui X, Liu H, Fan M, Ai B, Ma D, Yang F. 2020. Seafloor habitat mapping using multibeam bathymetric and backscatter intensity multi-features SVM classification framework. Appl. Acoust, 174: 107728.

Diesing M, Green SL, Stephens D, Lark RM, Stewart HA, Dove D. 2014. Mapping seabed sediments: Comparison of manual, geostatistical, object-based image analysis and machine learning approaches. Cont Shelf Res, 84: 107-119.

Huang Z, Siwabessy PJW, Nichol SL, Anderson T, Brooke BP. 2013. Predictive mapping of seabed cover types using angular response curves of multibeam backscatter data. Cont Shelf Res, 61-62: 12-22.

Ierodiaconou D, Monk J, Rattray A, Laurenson L, Versace VL. 2011. Comparison of automated classification techniques for predicting benthic biological communities using hydroacoustics and video observations. Cont. Shelf Res. 31, S28–S38.

Lucieer V, Hill NA, Barrett NS, Nichol S. 2013. Do marine substrates ‘look’ and ‘sound’ the same? Supervised classification of multibeam acoustic data using autonomous underwater vehicle images. Estuar Coast Shelf Sci, 117: 94–106.

Lurton X. 2002. An Introduction to Underwater Acoustics, Principles and Application. Chicester: Springer Praxis (UK).

Manik HM. 2012. Seabed identification and characterization using sonar. Adv Acoust Vib, vol. 2012, Article ID 532458, 5 pages.

Marsh I dan Brown C. 2009. Neural network classification of multibeam backscatter and bathymetry data from Stanton Bank (Area IV). Appl Acoust, 70: 1269-1276.

Prasetyo, Eko. 2014. DATA MINING, Mengolah Data Menjadi Informasi Menggunakan Matlab. Yogyakarta: Andi (INA).

Shepard FP. 1954. Nomenclature based on sand-silt-clay ratios. J. Sediment Petrol. 24: 151-158.

Simons DG dan Snellen M. 2009. A Bayesian approach to seafloor classification using multi-beam echo-sounder backscatter data. Appl Acoust, 70: 1258-1268.

Snellen M, Gaida TC, Koop L, Alevizos E, Simons DG. 2018. Performance of multibeam echosounder backscatter-based classification for monitoring sediment distributions using multitemporal large-scale ocean data sets. IEEE J. Oceanic Eng., 44(1): 142-155.

Solikin S, Manik HM, Pujiyati S, Susilohadi S. 2020. Support vector machine classification method for predicting Jakarta Bay bottom sediment type using multibeam echosounder data. Pertanika J. Sci. & Technol. 28(2): 477 – 491.

Srisawat A, Phienthrakul T, Kijsirikul B. 2006. “SV-KNNC: An Algorithm for Improving the Efficiency of K-Nearest Neighbor”. In: Qiang Yang, Geoffrey I. Webb. The 9th Pacific Rim International Conference on Artficial Intelligence (PRICAI-2006). Guilin, China. New York: Springer-Verlag.

Tan P, Steinbach M, dan Kumar V. 2006. Introduction to Data Mining. New York: Pearson Education (USA).

Tegowski J, Nowak J, Moskalik M, Szefler K. 2011. Seabed classification from multibeam echosounder backscatter data using wavelet transformation and neural network approach. Proceeding of 4th Intern. Conf. and Exhib. Greece, 20-24.

Wu XD, Kumar V. 2009. The top ten algorithm in data mining. Chapman & Hall/CRC, London.