World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:11, No:6, 2017

Summarizing Data Sets for Data Mining by Using Statistical Methods in Coastal Engineering

Yunus Doğan, Ahmet Durap

***Abstract*—**Coastal regions are the one of the most commonly used places by the natural balance and the growing population. In coastal engineering, the most valuable data is wave behaviors. The amount of this data becomes very big because of observations that take place for periods of hours, days and months. In this study, some statistical methods such as the wave spectrum analysis methods and the standard statistical methods have been used. The goal of this study is the discovery profiles of the different coast areas by using these statistical methods, and thus, obtaining an instance based data set from the big data to analysis by using data mining algorithms. In the experimental studies, the six sample data sets about the wave behaviors obtained by 20 minutes of observations from Mersin Bay in Turkey and converted to an instance based form, while different clustering techniques in data mining algorithms were used to discover similar coastal places. Moreover, this study discusses that this summarization approach can be used in other branches collecting big data such as medicine.

***Keywords*—**Clustering algorithms, coastal engineering, data mining, data summarization, statistical methods.

I. INTRODUCTION

COASTAL zones contain rich resources to produce goods and services for most commercial and industrial activities. For example, the long sea coasts in Europe, where almost half of the population live and supply the economic prosperity to the rest of the continent. Many activities such as fishing, shipping and tourism show that Europe's estimated competitiveness and labor are very extensive, and thus, a new workspace has emerged known as coastal engineering. Especially, after World War II, coastal engineering has emerged as a discipline that improves rapidly to resolve very complex hydrodynamics and morphological behavior under various environmental factors. While coast guards and analysis have been increased after 1950, it is assumed that the subject of coast guards and analysis prior to the 1950's due to incidents such as flooding and erosion. After the 1950's, new techniques and environmentally friendly policies were followed. Thus, the importance of need for protection of extensive coastal regions has increased. For example, it has been identified that significant destruction of coastal areas would be experienced due to rising sea levels, especially, the sea coasts of Venice in Italy, New Orleans in the US, Japan, the Netherlands and the Caspian. Measures have been taken to prevent possible destruction through the building of structures such as breakwaters.

Five direction-actions have been established as the coastal defense strategy, they are:

1) Management and planning of the appropriate desired objective;
2) Applying natural engineering ways for the existing structure and creating resistance blocks;
3) Protection of coastal areas through building walls;
4) Filling with the more favorable primary materials to protect of sea coasts;
5) A lot of vertical and high-rise buildings should be built-up along coastal areas in terms of minimizing so many events such as flooding.

In literature, there are studies containing data mining approaches about these five direction-actions.

It is becoming increasingly important to obtain meaningful information through data to predict reliably and make strategic decisions because of the rapid increase in the volume and number of databases. The tool fulfilling this function is data mining [1]. In other words, data mining is the process of uncovering the hidden rules, the patterns and models in large amounts of data stored [2]. Data mining performs this process by combining computer science, machine learning, database management, mathematical algorithms and statistics. Confidential information in the data stack can provide useful tips on the creation of strategic business plans as a discovered prediction for the future [3]. Data mining is a complementary approach for other statistical data analysis techniques such as Online Analytical Processing (OLAP) and basic data access [4], [5]. Data mining gives the opportunity to provide grouping on the basis of dimensions, the examination of the relationship between the dimensions, the data visualization and the presentation of results as graphs and reports. The obtained knowledge is a useful detection to help managers especially in strategic decision-making [6].

The process of predictive models and pattern recognition in data mining includes the approaches of classification, regression and time series. These models are differentiated on the basis of what is desired to predict. If it is prompted to predict the values of continuous output quality, regression analysis is preferred. If the distinctive features of the times are the subject, the approach of time series is preferred. If the predictions of the target feature for a new instance are required, a classification method must be preferred. Also, if a rule model for a specific data set with a target feature is required, a classification algorithm must be used [7], [8].

Clustering finds groups of data; association rule mining aims to obtain the behavior of the discovery association

Y. Dogan is with the Department of Computer Engineering, Dokuz Eylul University, Izmir, Turkey (e-mail: yunus@cs.deu.edu.tr).
A. Durap is with the Department of Civil Engineering, Gazi University, Ankara, Turkey (e-mail: ahmet.durap@gazi.edu.tr).

analysis and identification that reveals hidden sequences and rules. Clustering, of which the primary purpose is to group objects according to their specifications, is the technical of multivariate analysis. The grouping objects at clustering are performed according to a predetermined criterion. Clustering results should show a high degree of homogeneity in clusters and a high degree of heterogeneity between clusters [9].

In the literature, there are many clustering algorithms. Clustering algorithms are considered as two general types, including hierarchical and non-hierarchical. Some clustering algorithms belong to more than one type. Hierarchical algorithms have two sub-types; these are the agglomerative and the divisive approaches. The non-hierarchical types are partitioning, density-based, grid-based and other approaches. If used, the clustering algorithm creates clusters gradually and shows the data set as a tree formed of clusters, it is a hierarchical approach. The agglomerative approach is called the bottom-up approach. In this approach, instances are combined with each other and create clusters. The divisive approach has a process called top-to-bottom, conversely. BIRCH, CURE, Chameleon, ROCK algorithms belong to the hierarchical type. Hierarchical algorithms collect the most similar two instances between each other in a cluster. This process has a very high process cost because all instances are compared between each other before each collection. The divisive approach requires computing steps similar to the agglomerative approach. Also, Ward's minimum variance method can be given as an example of the hierarchical clustering method [10].

Clustering algorithms belong to the non-hierarchical type to find the clusters of each instance directly. Separation type of clustering is the most widely used in the non-hierarchical clustering approach. This type of algorithm generally changes the center of the clusters until the distances between all instances, and their related cluster centers, have minimum values. K-means is given as the most common example of the separation approach. Fuzzy C-means, K-modes and K-prototypes algorithms are kinds of k-means [11].

In a comparison of separation algorithms and hierarchical algorithms, separation approaches have less cost in terms of computation time, although these algorithms have an operation that calculates the distances between the centroid and all instances for each cluster until the values are the minimum sequentially. The distance criterion of the separation approaches is based on the formulas as the Euclidean distance. This usage limits the ability of these algorithms at finding spherical clusters. Furthermore, this type of algorithm has difficulty at finding random shaped clusters. The density-based clustering approach is another non-hierarchical clustering approach. Because the density-based clustering algorithm assumes that a density area with instances is a cluster, these algorithms can find non-linearly separable clusters. Typical examples of a density-based clustering algorithm are the DBSCAN, DENCLU and OPTICS algorithms. Grid-based clustering is an approach that accounts for cells besides instances, and because of this property, they are more efficient than other clustering algorithms computationally in general. STING, STING +, Waveclust, CLIQUE and GDILC are mentioned as examples of this approach [12].

The self-organizing maps (SOM) suggested by Kohonen, which is one of data mining techniques, has a wide range of applications [13]. SOM is the most common one of the artificial neural network (ANN) techniques, after standard feed forward networks. It can be described as a two-tier unsupervised neural network used for clustering and size reduction [14].

In this study, three types of clustering algorithms have been used for experimental studies. By usage of k-means, Single Linked Hierarchical Clustering Algorithm (SLHCA) and DBSCAN, an outlier has been detected and confirmed.

II. Related Works

In literature, it is seen that machine learning algorithms and statistical approaches are used in coastal engineering studies. Some of the prominent studies are described briefly in the following paragraphs.

The Bosphorus is a narrow, curved waterway connecting the Marmara Sea and the Black Sea. The two-layer complex flow structure in the Bosphorus is significantly affected by the changes of water level in these two seas. In the study of [15], the measurements of the water level at both entrances of Bosphorus have been evaluated separately and a tidal analysis has been done by using statistical methods such as Fourier analysis.

Van der Meer equations, which are widely used in the design of rock-fill breakwaters, are obtained from the results of hydraulic model experiments and are equations that can express the experimental data on average. For this reason, they affect the reliability of the building as a source of significant variability and indeterminacy. In the study of [16], for the preliminary design of the stone-filled coastal structures, an ANN including Van der Meer experiment data has been developed and applied for preliminary design of Mersin Yacht Port in Turkey. The obtained results show that the feed forward supervised neural networks can show a higher modeling ability than the Van der Meer equations.

In the study of [17], prediction potentials, uncertainties, and knowledge derivation of neural networks have been used for an algal metric to analyze with coastal big data and the computation of natural effects.

The paper of [18] addresses the existence of rare wave groups by examining time series data combined with data mining methods from a buoy. The Coastal Data Information Program (CDIP), University of California San Diego has operated the buoy.

The paper of [19] mentions a decision support system, PRISM-2 developed to evaluate salinity intrusion. The aim has been the examining of potential climate change along the South Carolina coast in southeastern United States. The decision support system has been disseminated using data mining methods such as Artificial Neural Network.

The paper of [20] describes a novel approach of integrating a shallow water semi-analytical model by using genetic algorithm.

The study of [21] is interested in applying machine-learning algorithms on water depth inversion from remote sensing images as a case study in Michigan lake area. The aim is evaluating the use of the public available LANDSAT (satellite) images on shallow water depth inversion.

In [22], the development of a Regional Neural Network for Water Level predictions is presented with an application to the coastal inlets along the South Shore of Long Island, New York. Neural network supplies an effective approach to correlate the non- linear input and output of water levels to recognize the historic patterns between them.

The study of [23] presents two different neural network strategies employed to forecast significant wave heights and zero-up-crossing wave periods 3, 6, 12 and 24 h in advance. Eight simple separate neural networks have been implemented to simulate every wave parameter over each prediction interval in the first approach. Two networks have provided simultaneous forecasts of these wave parameters for the four prediction intervals in the second approaches.

III. Methodology

There are some statistical methods, such as the maximum wave height (H_{max}), the maximum period of waves (T_{max}), the mean of the highest one-tenth of waves ($H_{1/10}$), the mean period of the highest one-tenth of waves ($T_{1/10}$), the mean of the highest one-third of waves ($H_{1/3}$), the mean period of the highest one-third of waves ($T_{1/3}$), the mean height of waves (H_{mean}) and the mean period of waves (T_{mean}) and the root-mean-square wave height (H_{rms}) in the wave spectrum to analysis of the big data in coastal engineering.

The wave height H is defined as the vertical distance between the highest and the lowest surface elevation in a wave. A wave has only one wave height and the mean wave height is $\overline{H}$ in a wave record with N waves defined as:

$$\overline{H} = \frac{1}{N}\ x\ \sum_{i=1}^{N} H_i \quad (1)$$

where i is the sequence number of the wave in the record.

If a quadratically weighted average value is used, it is called the root-mean-square wave height H_{rms} and it is defined as

$$H_{rms} = \sqrt{\frac{1}{N}\ x\ \sum_{i=1}^{N} H_i^2} \quad (2)$$

The wave energy is proportional to the square of the wave height; a measure of wave heights may be relevant for energy-related projects. These wave heights $\overline{H}$ and H_{rms} have been used in this study, but they are not enough to discover the characteristic of the wave data. Another wave height, called the significant wave height H_s has been used. The mean of the highest one-third of waves in the wave record is defined as:

$$H_{1/3} = \frac{1}{N/3}\ x \sum_{j=1}^{N/3} H_j \quad (3)$$

where j is not the sequence number in the record.

The rank number of the wave is based on the wave height. This is a different way to define a characteristic wave height; also the value of this wave height is close to the value of the visually estimated wave height.

Both the visually estimated characteristic wave height and this measured characteristic wave height are called the 'significant wave height'. The visually estimated significant wave height, to distinguish them from one another, is denoted as H_v, and the measured significant wave height as $H_{1/3}$. This significant wave height can be estimated from the wave spectrum denoted as H_{m0}. Also, the mean of the highest one-tenth of waves has been used to define $H_{1/10}$ as:

$$H_{1/10} = \frac{1}{N/10}\ x \sum_{j=1}^{N/10} H_j \quad (4)$$

There is not any obvious relation to the visually estimated significant wave height. If the waves are not too steep and not in very shallow water, it means that there is a constant ratio between the various characteristic wave heights [24].

$$H_{rms} = \frac{1}{2}\ x\ \sqrt{2}\ x\ H_{1/3} \quad (5)$$

The period T is defined as the time interval between the start and the end of the wave that is the interval between one zero-down crossing and the next. The mean of the zero-crossing wave period, T_0 is denoted as $\overline{To}$.

$$\overline{To} = \frac{1}{N}\ x\ \sum_{i=1}^{N} T_{0i} \quad (6)$$

where i is the sequence number of the wave in the time record.

The mean period of the highest one-third of waves in the wave record is defined as:

$$T_{1/3} = \frac{1}{N/3}\ x \sum_{j=1}^{N/3} T_{0j} \quad (7)$$

where j is not the sequence number in the time record [24].

The mean period of the highest one-tenth of waves has been used to define $T_{1/10}$ as:

$$T_{1/10} = \frac{1}{N/10}\ x \sum_{j=1}^{N/10} T_{0j} \quad (8)$$

In this study, base statistical methods exclusive of these wave measurements ($\overline{H}$, H_{rms}, $H_{1/3}$, $H_{1/10}$, $\overline{To}$, $T_{1/3}$ and $T_{1/10}$) have been used as the minimum wave height (min), the maximum wave height (max), the standard deviation (ð), the mode value (m), the median value (n), the range value (r), the variation value (v) and the coefficient of variation (c) to reveal the characteristic features of the selected coasts. There are 15 features, and all observation data have been reduced to these 15 features for a coast.

In the experimental studies, the tests and results have been described for a sample data from Mersin Bay in Turkey as a case study.

IV. Experimental Studies

In this study, clustering algorithms have been used and this data summarization approach was required for the data preprocessing phase. The clustering algorithms must take an instance-based data set as the input. The measurements of wave heights from six different areas of Mersin Bay in Turkey have been collected in order to determine the characteristic features of each area. This observation took 20 minutes, and during this time, 1536 wave heights have been recorded for each area separately. In Fig. 1, the samples of the wave heights of the six coasts are presented. These samples are for the first 100 wave heights.

A total of 9216 (1536 x 6) height values have been used to reveal the characteristic features of the coast line used in this study. Firstly, statistical values being mentioned in the previous section have been calculated and 9216 values have been reduced to 90 (15x6) values. The calculated values for the first five attributes as $\overline{H}$, H_{rms}, $H_{1/3}$, $H_{1/10}$ and $\overline{To}$ are presented in Table I. The calculated values for the second five attributes as $T_{1/3}$, $T_{1/10}$, ð, v and maximums are presented in Table II. Table III presents the calculated values for the third five attributes as minimums, r, m, n and c are presented.

Secondly, these values have been normalized by Z-score method to have equal weights between the 15 attributes while using clustering algorithms. The results are presented in Tables IV-VI.

$$\text{Z-score} = \frac{V-\mu}{\sigma} \tag{9}$$

where V is the original value, σ is the standard deviation value of the current attribute and μ is the average value of the current attribute.

TABLE I
The First Five Attributes and Values of the Prepared Data Set

	$\overline{H}$	H_{rms}	$H_{1/3}$	$H_{1/10}$	$\overline{To}$
A	-0.09	132.24	144.66	229.96	5.56
B	-0.01	92.45	100.81	161.73	4.94
C	-0.10	93.59	102.10	161.87	5.09
D	-0.01	90.65	98.67	98.67	5.56
E	0.06	78.08	85.93	137.75	4.94
F	-0.04	144.36	160.02	249.47	5.97

After the data preprocessing phase, three different types of clustering algorithms have been used to conduct an analysis. Firstly, k-means has been used for 2, 3 and 4 for the value of the k parameter; the sum of the square error obtained 3.5, 2.3 and 1.7, separately. Moreover, when the value of k has been 2, coasts B-C-D-E and F have been located in a cluster and only coast A has been located in the other cluster. It shows that the characteristic features of coast A to be more different than the others in the same bay. To confirm this result, SLHCA and DBSCAN have been used separately. The aim of using these algorithms has been that they do not need any parameter for the number of clusters.

After the analysis, SLHCA has been presented as a hierarchical structure where the coasts B-C-D-E-F have been located in a common dendrogram and coast A has been linked to them as the last piece. DBSCAN has discovered only one cluster the coast A has been named as a noise instance. As a result, it has been proved that the characteristic features of coast A are more different than coasts B-C-D-E-F in the same bay.

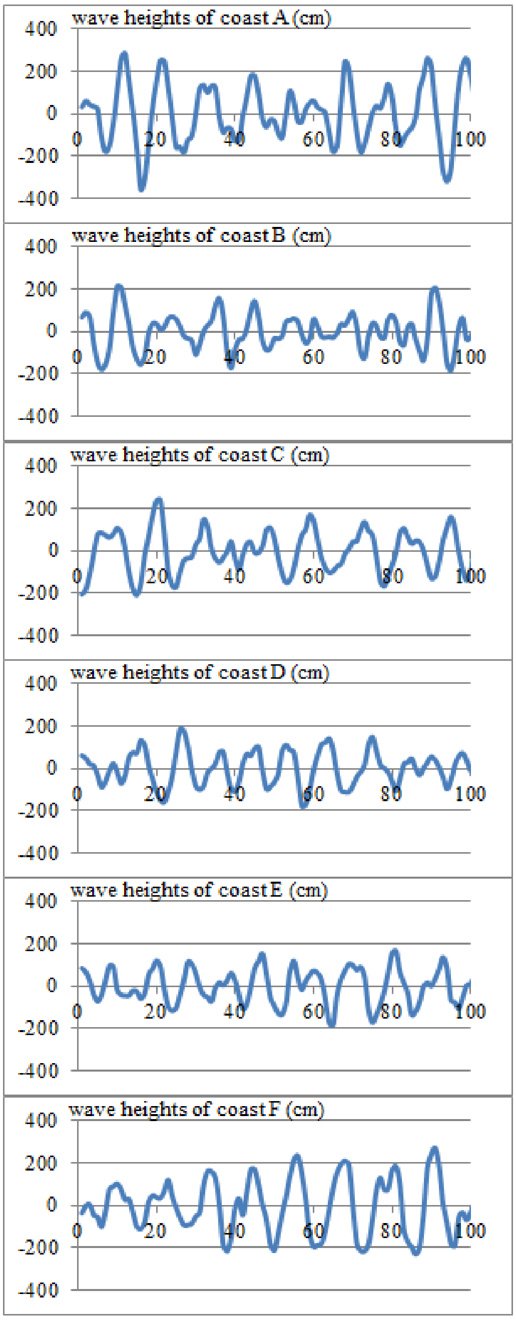

Fig. 1 Samples of wave heights for six different coast locations (named as A, B, C, D, E and F) in Mersin Bay in Turkey

TABLE II
THE SECOND FIVE ATTRIBUTES AND VALUES OF THE PREPARED DATA SET

	$T_{1/3}$	$T_{1/10}$	ð	v	max
A	7.98	8.84	132.24	17487.68	425.00
B	7.37	8.46	92.45	8547.13	301.00
C	7.77	9.57	93.59	8758.50	315.00
D	8.32	9.69	90.65	8217.50	272.00
E	7.26	8.59	78.08	6096.75	274.00
F	8.72	9.88	160.02	249.47	5.97

TABLE III
THE THIRD FIVE ATTRIBUTES AND VALUES OF THE PREPARED DATA SET

	min	r	m	n	c
A	-462.00	887.00	-37.00	-1.00	-149354.55
B	-308.00	609.00	61.00	0.00	-887526.42
C	-303.00	618.00	0.00	0.00	-97788.73
D	-278.00	550.00	24.00	4.00	-1160325.56
E	-262.00	536.00	1.00	-0.50	137854.57
F	-468.00	866.00	-27.00	0.00	-335974.29

TABLE IV
THE FIRST FIVE ATTRIBUTES AND VALUES OF THE NORMALIZED DATA SET

	$\bar{H}$	H_{rms}	$H_{1/3}$	$H_{1/10}$	$\overline{To}$
A	-1.10	1.12	1.08	1.09	0.56
B	-2.48	-0.38	-0.38	-0.39	-0.12
C	0.23	-0.37	-0.37	-0.39	-0.09
D	-2.57	-0.40	-0.40	-0.76	0.00
E	-4.61	-0.51	-0.51	-0.53	-0.11
F	-1.45	0.12	0.13	0.11	0.08

TABLE V
THE SECOND FIVE ATTRIBUTES AND VALUES OF THE NORMALIZED DATA SET

	$T_{1/3}$	$T_{1/10}$	ð	v	max
A	0.14	-0.59	1.12	1.07	1.58
B	-0.08	-0.04	-0.38	-0.77	-0.37
C	-0.03	0.08	-0.37	-0.75	-0.33
D	0.04	0.09	-0.40	-0.80	-0.46
E	-0.09	-0.03	-0.51	-0.98	-0.46
F	0.09	0.11	0.12	0.29	-0.08

TABLE VI
THE THIRD FIVE ATTRIBUTES AND VALUES OF THE NORMALIZED DATA SET

	min	r	m	n	c
A	-1.36	1.46	-1.25	-0.86	0.58
B	-0.44	-0.41	26.73	2.40	1.78
C	-0.46	-0.40	10.09	2.40	-0.12
D	-0.53	-0.50	16.64	12	2.43
E	-0.58	-0.52	10.36	1.20	-0.69
F	0.02	-0.03	2.73	2.40	0.45

V. CONCLUSION

Data mining algorithms need an instance based data set to conduct analysis. Wave heights in coastal engineering are the most important data to detect the specialties of a coast. After the analysis, coastal engineers decide the structure of the building that must be built at the coast. In the current study, only one coast is examined using wave spectrum methods. This study shows that comparisons between coasts can be done by converting sequential data sets to instance base data sets and then, by using data mining algorithms. Thus, the profiles of coasts can be revealed, and also, big data obtained through observations for minutes, hours or days is reduced to a summarized form.

This approach can be used for other sciences such as medicine. For example, in intensive care units, patients are observed with a lot of parameters such as heartbeat, diastolic-systolic blood pleasures and body temperature. The characteristic specialties for each parameter of patients can be detected by this summarization approach and the historic data lists transformed at an instance in the data set. Finally, these patient statistics can be compared and analyzed using data mining algorithms.

REFERENCES

[1] C. P. Wei, Y. H. Lee and C. M. Hsu, "Empirical comparison of fast partitioning-based clustering algorithms for large data sets", *Expert Systems with Applications*, vol. 24, no. 4, pp. 351-361, 2003.
[2] M. J. Shaw, C. Subramaniam, G. W. Tan and M. E. Welge, "Knowledge management and data mining for marketing", *Decision Support Systems*, vol. 1, no. 31, pp. 128-138, 2001.
[3] S. H. Liao, "Knowledge management technologies and applications - literature review from 1995 to 2002", *Expert Systems with Applications*, vol. 25, no. 2, pp. 157-167, 2003.
[4] E. W. M. Ma and T. W. S. Chow, "A new shifting grid clustering algorithm", *Pattern Recognition*, vol. 37, no. 3, pp. 503-513, 2004.
[5] J. A. McCarty and M. Hastak, "Segmentation approaches in data mining: A comparison of RFM, CHAID and Logistic Regression", *Journal of Business Research*, vol. 60, no. 6, pp. 656-666, 2009.
[6] G. G. Emel, C. Taskin and S. Kilicarslan, "An analysis at the process of steel production by using artificial neural network", *Journal of Dokuz Eylül University*, vol. 5, no. 1, pp. 206-207, 2004.
[7] C. Rygielski, J. C. Wang and D. C. Yen, "Data mining techniques for customer relationship management", *Technology in Society*, vol. 24, no. 4, pp. 488-498, 2002.
[8] H. M. Moshkovich, A. I. Mechitov and D. L. Olson, "Rule induction in data mining: Effect of ordinal scales", *Expert Systems with Applications*, vol. 22, no. 4, pp. 303-304, 2002.
[9] C. Budayan, I. Dikmen ve M. T. Birgonul, "Comparing the performance of traditional cluster analysis, self-organizing maps and fuzzy c-means method for strategic grouping", *Expert Systems with Applications*, vol. 36, no. 9, pp. 117-127, 2009.
[10] R. J. Kuo, L. M. Ho and C. M. Hu, "Cluster analysis in industrial market segmentation through artificial neural network", *Computers & Industrial Engineering*, vol. 42, no. 4, pp. 393-403, 2002.
[11] A. Likas, N. Vlassis and J. J. Verbeek, "The global k-means clustering algorithm", *Pattern Recognition*, vol. 36, no. 2, pp. 451-461, 2003.
[12] C. H. Hsu, "Data mining to improve industrial standards and enhance production and marketing: An empirical study in apparel industry", *Expert Systems with Applications*, vol. 36, no. 3, pp. 504-514, 2009.
[13] B. Hammer, A. Micheli, A. Sperduti and M. Strickert, "Recursive self-organizing network models", *Neural Networks*, vol. 17, no. 10, pp. 1061-1071, 2004.
[14] D. G. Roussinov and H. Chen, "Document clustering for electronic meetings: An experimental comparison of two techniques", *Decision Support Systems*, vol. 27, no. 2, pp. 70-80, 1999.
[15] B. Aydogan, B. Ayat, M. N. Ozturk, Y. Yuksel and E. O. Cevik, "Modeling of water level changes in the Bosphorus", *The 6th National Coastal Engineering Symposium*, Izmir, Turkey, 2007, pp. 271-278.
[16] M. L. Koc, C. E. Balas and A. Arsla, "Preliminary design of artificial neural networks of stone filled breakwaters", *IMO Technical Journal*, vol. 225, no. 11, pp. 3351-3375, 2004.
[17] D. F. Milliea, G. R. Weckmanc, W. A. Y. IId, J. E. Iveye, D. P. Friesf, E. Ardjmandc and G. L. Fahnenstielb, "Coastal 'big data' and nature-inspired computation: Prediction potentials, uncertainties, and knowledge derivation of neural networks for an algal metric", *Coastal and Shelf Science*, vol. 125, pp. 57–67, 2013.
[18] H. C. Seyffert and A. W. Troesch, "Data mining Pt. Reyes Buoy for rare wave groups", *Journal of Offshore Mechanics and Arctic Engineering*, vol. 138, no. 1, pp. 1-8, 2015.

[19] P. A. Conrads and E. A. Roehl, "The use of data-mining techniques for developing effective decision support systems: a case study of simulating the effects of climate change on coastal salinity intrusion", *The Geological Society of London*, vol. 408, 2015.

[20] C. H. Chang, C. C. Liu, H. W. Chung, L. J. Lee and W. C. Yang, "Development and evaluation of a genetic algorithm-based ocean color inversion model for simultaneously retrieving optical properties and bottom types in coral reef regions", *Applied Optics*, vol. 53, no. 4, pp. 605-617, 2014.

[21] S. Gao, "Shallow water depth inversion based on data mining models", *B.S., China University of Petroleum (East China)*, 2013.

[22] W. Huang, C. Murray, N. Kraus and J. Rosati, "Development of a regional neural network for coastal water level predictions", *Ocean Engineering*, vol. 30, pp. 2275–2295, 2003.

[23] O. Makarynskyya, A. A. Pires-Silvab, D. Makarynskaa and C. Ventura-Soaresc, "Artificial neural networks in wave predictions at the west coast of Portugal", Computers & Geosciences, vol. 31, pp. 415–424, 2005.

[24] L. H. Holthuijsen, "Waves in Oceanic and Coastal Waters", *Cambridge University Press*, pp. 27-28.