Journal of Applied Data Sciences
Vol. 6, No. 1, January 2025, pp. 247-258

Using Machine Learning Approach to Cluster Marine Environmental Features of Lesser Sunda Island

Evellin Dewi Lusiana[1,*], Suci Astutik[2], Nurjannah[3], Abu Bakar Sambah[4]

[1,2,3]Department of Statistics, Faculty of Mathematics and Natural Science, Universitas Brawijaya, Veteran Street, Malang, 65145, Indonesia

[4]Faculty of Fisheries and Marine Science, Universitas Brawijaya, Veteran Street, Malang, 65145, Indonesia

(Received: September 11, 2024; Revised: October 31, 2024; Accepted: November 8, 2024; Available online: December 28, 2024)

Abstract

Mapping marine ecosystems is acknowledged as a vital tool for implementing ecosystem services in practical situations. It provides a framework for effective marine spatial planning, enabling the designation of marine protected areas (MPAs) that consider ecological connectivity and habitat requirements. It also helps pinpoint areas of high biodiversity or ecological significance, allowing conservationists to prioritize these regions for protection and management. Numerous studies over decades have produced a vast amount of data that illustrates the features of the marine ecosystem. Therefore, the unsupervised learning is a promising technique to map marine ecosystem based on its environmental features. This study aims to compare unsupervised learning techniques to analyze marine environmental features in order to map marine ecosystem in Lesser Sunda waters. Eleven global environmental variables were accessed from global databases. The Lesser Sunda waters were delineated into groups according to their environmental characteristics using four unsupervised learning techniques: k-mean, fuzzy c-mean, self-organizing map (SOM), and density-based spatial clustering of applications with noise (DBSCAN). According to the findings, the Lesser Sunda waters can be divided into five to nine clusters, each with distinct environmental features. Moreover, the fuzzy c-mean method's clustering result outperformed the others based on the highest Silhouette (0.2204478) and Calinski-Harabasz (1741.099) Index. As an unsupervised learning technique, fuzzy c-mean clustering offered good performance in delineating Lesser Sunda Island marine waters with five clusters. The clustering results mostly consistent with existing conservation programs, even though there are several areas which needed international and multinational organization collaboration to effectively accomplish marine conservation objectives.

Keywords: Cluster Analysis, Conservation, Machine Learning, Marine Ecosystem, Protected Areas

1. Introduction

The marine ecosystem is one of the world's largest and most complex, supporting a diverse range of species and providing essential ecological services. The human need for services provided these ecosystems has grown in recent years due to the fast economic expansion and urbanization of coastal areas [1]. The marine environment is being destroyed at an increasing rate due to human demand for marine ecosystem services, and certain places are seeing a reduction in ecosystem services. With the depletion of marine resources worldwide and the degradation of the marine environment in general, the preservation and oversight of marine ecosystems has become an international concern and responsibility [2].

Understanding the marine ecosystem completely involves continuous monitoring due to its vastness and complex environment [3]. While terrestrial ecosystems are conveniently monitored since biotic communities serve as the major basis to classify them, marine systems lack set boundaries and are extremely dynamic in a tri-dimensional environment [4]. Within this scheme, mapping marine ecosystems is acknowledged as a vital tool for implementing ecosystem services in practical situations. They can be used to determine and design new networks of marine protected areas and restoration areas, to evaluate the effects of human pressures on ocean resources and ecosystem services, and to provide information for maritime spatial planning [5]. Advances in modern science led to an increase in both local and international research on the marine environment.

***Corresponding author: Evellin Dewi Lusiana (evellinlusiana@ub.ac.id)**
DOI: https://doi.org/10.47738/jads.v6i1.478

Numerous studies over decades have produced a vast amount of data that illustrates these features of the ecosystem [6]. Utilizing traditional numerical methods to evaluate large data indeed becomes more and more challenging. By fully utilizing large ocean data, humans can advance their understanding of how to respond to climate change, safeguard the ecological system, and avert natural calamities [7]. Classical statistical analysis, ocean model simulation, manual clustering and recognition, and other techniques are used in the processing and analysis of ocean data [8], [9], [10]. These techniques cannot fully capture the underlying information in the data since they are frequently influenced by subjective factors [11].

Marine environmental parameters such as sea surface temperature, chlorophyll, salinity, and so on are all significant ocean elements [12]. To avert disasters and safeguard the environment, it is crucial to analyze these factors. A multitude of machine learning algorithms may offer precise and effective techniques for examining oceanic properties [13]. Machine learning enables computers to aid humans in evaluating large and complex data sets. This field of study focuses on building models, analyzing data, classifying it, and making predictions using that information. Machine learning techniques are commonly employed to handle a variety of large-data problems, such as picture recognition and classification/clustering, as well as extreme events in complex systems [14].

Machine learning is separated into two categories based on whether the input data contains labels: supervised learning, which uses labeled sample data, and unsupervised learning, which uses unlabelled sample data [15]. The techniques could seek to group observations together according to a measure of similarities [16] to create more simple representations for the data whilst keeping essential properties, also referred to as reduction in dimensionality, or to create a model for the distribution of the data [17]. Therefore, unsupervised learning is a promising technique to map marine ecosystems based on their environmental features. However, such researches is still limited in practice.

This study aims to implement and compare unsupervised learning techniques to analyze marine environmental features to map marine ecosystems in Lesser Sunda waters. The region is a crucial marine area in Indonesia as it is located in the "intersection" of key global climate processes, resulting in the most vulnerable tropical environment in the Indonesian ocean region [18]. The Lesser Sunda Islands are becoming increasingly vulnerable to various human activities such as seismic oil exploration and production, tourism development, and domestic and industrial waste, all of which have resulted in environmental damage [19]. The study is important for monitoring ocean environments, which will support the growth and sustainability of the marine sector.

2. Research Methodology

This research process involved several key steps as shown in figure 1. The process begins with data collection from global databases focusing on marine environmental characteristics of Lesser Sunda Island marine waters. Then, an exploratory data analysis (EDA) was conducted to elucidate main characteristics of the study area. After that, data pre-processing stage was initiated to improve the data quality and analysis efficiency. Further, several ML algorithms were performed, such as K-means, Fuzzy C-means (FCM), Density-based spatial clustering of applications with noise (DBSCAN), and Self-organizing map (SOM), to classify or map the study area based on environmental variables. Subsequently, the results were compared in terms of cluster internal validation indices.

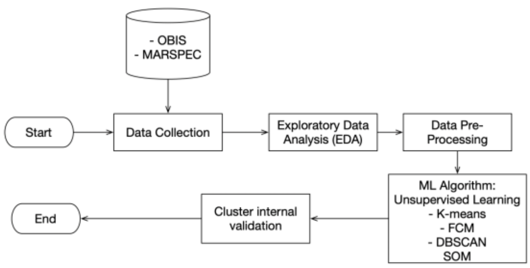

Figure 1. Research procedure workflow

2.1. Study Area and Data Collection

The study area took place in Lesser Sunda Island (figure 2), which is primarily in Indonesia and is separated into four provinces: West Nusa Tenggara, East Nusa Tenggara, Bali, and Maluku. It also includes areas inside Timor Leste. This area included in Coral triangle (CT), which is a triangular-shaped tropical marine area around Indonesia, the Philippines, Malaysia, Timor Leste, Papua New Guinea and the Solomon Islands. Due to its high species diversity and high endemicity, this place is considered the center of world marine biodiversity [20].

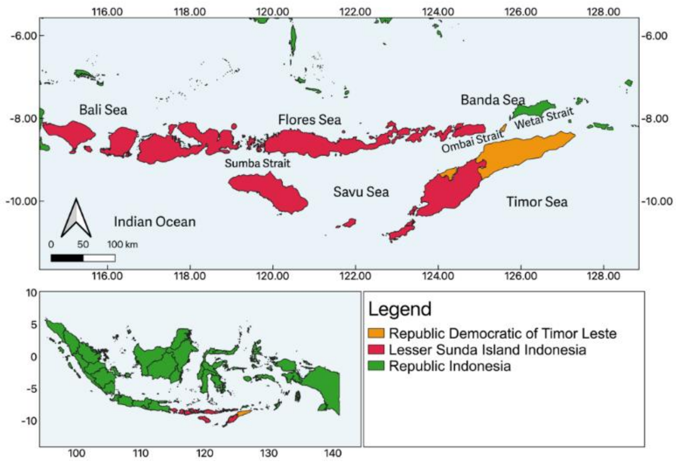

Figure 2. Geographical location of study area in Lesser Sunda island

Various marine environmental factors were examined as variables for describing the marine environment of the Lesser Sunda waters, and geographic information sources for such variables were obtained. The variables are including sea surface chlorophyll (SSC, mg/m3), sea surface temperature (SST, 0C), nitrate (μmol/m3), salinity (PSS), pH, depth (m), slope of the seabed (degree), distance to shore (km), current velocity (m/s), the eastness (aspect E-W) and northness (aspect N-S) of the slope (radians). SSC and nitrate represent the nutrient content in marine waters [21], [22], while aquatic organisms responded to SST shift due to global warming through changes in the physiology and phenology of organisms, as well as their populations and distribution [23]. Salinity and pH can reduce the concentration of calcium carbonate, which has a negative effect on calcareous species [24]. Factors such as distance to shore and current velocity are related to anthropogenic pressure on the coastal region [25]. On the other hand, the rest of the variables denote the local-scale habitat characteristics that affect the marine biodiversity [26]. These environmental data are sourced from the Marine Spatial Ecology (MARSPEC) and Bio-Oracle databases (http://www.bio-oracle.org). This global database offers up-to-date, satellite-based data with a spatial scale of 30 arc seconds (~ 1 km2) on the ocean's surface and seafloor [27]. While remote sensing databases provide valuable data, they can miss critical local variations or species-specific responses to environmental factors. Furthermore, The spatial resolution of Bio-Oracle data may not be fine enough for certain local studies, potentially obscuring important ecological patterns or gradients [28].

2.2. Exploratory Spatial Data Analysis

Exploratory spatial data analysis (ESDA) is a subset of exploratory data analysis that emphasizes the unique aspects of geographic information. It is a growing technique in geographic information science (GIS) that enables those who use it to explain and perceive spatial spreads, recognize peculiar areas or geographical anomalies, identify variations in spatial association, clusters, or hot spots, and propose spatial regimes or other types of spatial heterogeneity [29]. In this study, a chloropleth map was used to describe the spatial distribution [30] of the environmental features of the Lesser Sunda island.

2.3. Data Pre-Processing

Prior to the implementation of the machine learning algorithm, the data was pre-processed, including removal of missing values and data transformation. The listwise deletion technique was used to remove any rows that contain

ISSN 2723-6471

missing values. Meanwhile, the data were transformed using Z-score normalization so that they have a mean of 0 and a standard deviation of 1.

2.4. Machine Learning Approach: Unsupervised Learning Clustering Algorithm

Unsupervised learning is a subfield of machine learning for detecting patterns in datasets with unlabelled or unstructured data points. It infers underlying concealed patterns from the historical data. In this approach, a machine learning model searches for similar characteristics, differences, patterns, and structure in data on its own, with no need for human intervention. One of the most common algorithms in unsupervised learning is the clustering algorithm. The process of clustering splits a given set of unlabeled data into multiple clusters based on similarity criterion. There are several clustering algorithms within unsupervised learning, as shown in table 1.

Table 1. Clustering Algorithm of Unsupervised Learning

No	Algorithm	Description
1	K-means	An iterative technique that divides data into a set number of groups, or clusters, and minimizes the variation within each cluster [31]
2	Fuzzy c-means (FCM)	A data set is divided into numerous clusters, and every data point in the dataset belongs to each cluster for particular degree [32]
4	Density-based spatial clustering of applications with noise (DBSCAN)	Makes use a particular distance to distinguish high density clusters from low density noise [33]
5	Self-Organizing Map (SOM)	An artificial neural network wherein nodes inside a grid are designed to align with clusters of related data points [34]

2.5. Internal Validation Indices

There are several commonly used internal validation indices in clustering analysis, such as the Silhouette index [35], the Davies-Bouldin index, and the Calinski-Harabasz index [36]. The silhouette index ranges between −1 and 1, where a higher silhouette coefficient refers to a model with more coherent clusters. In other words, silhouette coefficients close to +1 mean the sample is far away from the neighboring clusters. A value of 0 means that the sample is on or very close to the decision boundary between two neighboring clusters. Finally, negative values indicate that the samples could have potentially been assigned to the wrong cluster. The Silhouette index can be calculated as follows [37].

$$\mathrm{SI_i} = \frac{\mathrm{b_i - a_i}}{\max\,(\mathrm{a_i, b_i})} \tag{1}$$

where a_i represents the average distance between sample i and other samples in its cluster and b_i represents the minimum average distance between sample i and samples in other clusters.

Meanwhile, The Davies–Bouldin index is a measure of uniqueness of the clusters and takes into consideration both cohesiveness of the cluster (distance between the data points and center of the cluster) and separation between the clusters. It is the function of the ratio of within-cluster separation to the separation between the clusters. The lower the value of the Davies–Bouldin index, the better the clustering. The index is defined in the following way [38]

$$\mathrm{DB_k} = \frac{1}{\mathrm{k}} \sum_{\mathrm{i=1}}^{\mathrm{k}} \max_{\mathrm{i \neq j}} \left\{ \frac{\mathrm{S_i + S_j}}{\mathrm{d(\bar{x}_i, \bar{x}_j)}} \right\} \tag{2}$$

Where k is the number of clusters, S_i is the average distance between the cluster's center and all of its elements, and $d(\bar{x}_i, \bar{x}_j)$ is distance of the cluster's i-th and j-th centers.

Calinski-Harabasz index is a classical cluster validity index as the ratio of the between-cluster to the within-cluster variance. hich is defined as the logarithmic of the ratio of the sum of the between-cluster squared distances (BSS) to the sum of the squared within-cluster distances (WSS). The ratio of BSS to WSS, which ranges from 0 (i.e., no difference among groups) to 1 (i.e., maximum difference among groups). This index measures the extent to which clusters are different from each other. The index is defined as follows [39]

$$CH = \frac{BSS/_{k-1}}{WSS/_{n-k}} \quad (3)$$

Where k is number of clusters and n is number of data samples.

3. Results and Discussion

3.1. Marine Environmental Features of Lesser Sunda Waters

The following figure 3 illustrates the characteristics of environmental variables in Lesser Sunda waters. It can be seen that pH, SST, and salinity values in the area were around 8, 29 0C, and 34 PSS, respectively. These numbers are considered in normal range for marine ecosystem [40]. On the other hand, the distribution of SSC was relatively uniform (<1 mg/m3) throughout the Lesser Sunda waters area. Higher SSC concentration was found in the coastal area. It results from a trophic effect, particularly from more carnivorous grazing in shallow waters [41]. Moreover, nitrate concentration in eastern part of Lesser Sunda waters was notably greater than its western part. However, these concentrations still in acceptable limit for marine aquatic life [42].

Current velocity and sea depth in research area was ranging from <0.2 – 0.8 m/s and <1000 – 5000 m, respectively. Currents are vital in marine ecosystems because they recirculate heat, water, nutrients, and oxygen throughout the ocean. Living things are invariably swept away by currents at the same moment [43]. In addition, the thickness of the water column determines the rate of cooling because the deeper the area, the more heat it stores and the slower the surface cools [44]. Because the depth of the sea reduces atmospheric pressure, marine life that lacks gas-filled cavities like swim bladders and lungs is not suited for the ocean depths. As a result, these species are exposed to the extreme pressures found in the deep ocean, which have the potential to destroy them and cause bodily harm [45].

Slope, eastness (aspect EW), and northness (aspect NS) are terrain variables of marine ecosystem. They represent the topographic of the ocean. The slope of seabed in Lesser Sunda waters was less than 1 degree. Therefore, it is considered as flat ocean [46]. Meanwhile, the eastness and northness of Lesser Sunda sea were vary from -0.5 to 0.5 radians. Topographic complexity is commonly thought to be closely related to habitat complexity and niche variety; nevertheless, complex topography does not guarantee habitat compatibility. Thus, topography may play a significant role in regulating both biotic and abiotic forces [47].

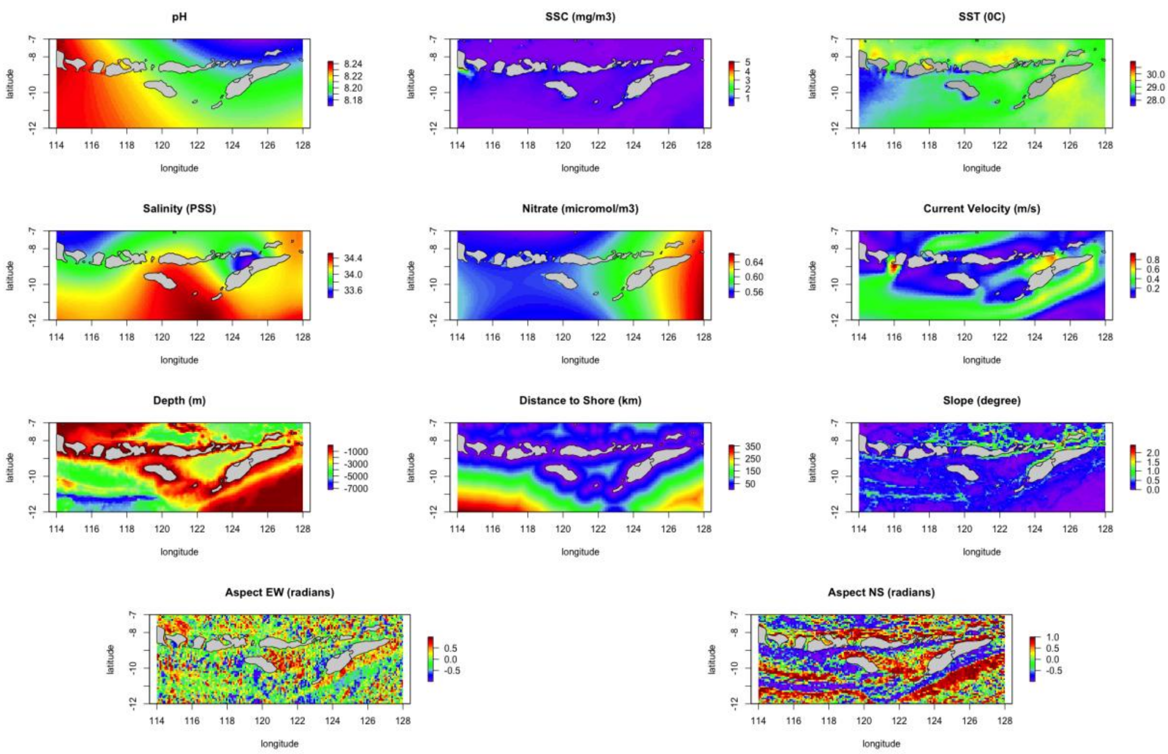

Figure 3. Environmental features of Lesser Sunda waters

3.2. Unsupervised Clustering Analysis Results

Clustering analysis result for K-Means algorithm suggested that number of clusters $k = 8$ as the optimum cluster in term of Silhouette, Davies-Bouldin, and Calinski-Harabasz index. On the other hand, the Fuzzy C-Means (FCM) technique showed that number of clusters $k = 5$ and fuzzy coefficient $m = 1.5$ gave the best result. Meanwhile, DBSCAN algorithm with $\varepsilon = 1.75$ and $k = 7$ produced preferable cluster. Finally, SOM's resulting in 9 cluster with mean distance to the closest unit in the map was equal to 3.549. These optimal number of clusters were determined in accordance to the internal validity indices. Then, comparison of the optimal clusters obtained from each algorithm were performed to decide which method that produced best clustering result, as shown in table 2.

Table 2. Comparison of unsupervised learning cluster evaluation indices

No	Algorithm	Optimum Clusters	Silhouette Index	Davies-Bouldin Index	Calinski-Harabasz Index
1	K-means	8	-0.1333740	1.473767	1701.544
2	Fuzzy c-means (FCM)	5	0.2204478	1.740454	1741.099
3	DBSCAN	7	-0.1940905	0.799283	35.77277
4	SOM	9	0.1561151	1.503326	277.3825

Table 2 shows the comparison of evaluation indices from each unsupervised clustering methods. It is visible that FCM method has the highest Silhouette and Calinski-Harabasz index, while DBSCAN method produced lowest Davies-Bouldin index. Hence, it is suggested that clusters produced by FCM method performs better than other techniques. This finding illustrates that having an extensive cluster does not equate to improved performance. It is because optimal cluster resulted from FCM method was the smallest. Variations in the cluster centroid value will have an impact on the threshold value that is determined, which will impact the classification outcomes [48]. Apart from the centroid cluster value, the threshold value determination method also has an impact on changes in threshold values [49].

The comparison result in this study is in line with those of [50], [51], [52], [53]. These previous studies related with ecology and environmental research suggested that FCM produced good performance and classification result in this field compared to traditional algorithm or hard clustering algorithm. Traditional unsupervised classification algorithms, such as K-means, known as hard clustering algorithms with a single classification basis. Hence the resulting categories are so difficult to control that the algorithms can easily fall into local optima and have classification uncertainties. Meanwhile, FCM algorithm considers the fuzzy characteristics between samples and classes in the membership degree, and completes the automatic classification by optimizing the objective function to obtain the membership degree. Fuzzy logic acknowledges that there are objects or area that cannot be clearly defined as belonging to one category or the other, but that have a degree of membership to a category/cluster. An area, with certain characteristics, cannot completely delineate into a separate category or cluster. It is because of the dynamics nature of marine ecosystem which involve the interaction between biological, chemical, geological, and physical factors of the ecosystem [54]. Furthermore, fuzzy logic approach takes into account the inherent uncertainty of environmental variables and enable the incorporation of ecological aspects such as the ecological gradient theory [55].

3.3. Marine Mapping of Lesser Sunda Waters

FCM clustering for marine mapping in the Lesser Sunda waters resulted in five clusters (figure 4). Cluster 1 and Cluster 2 represent Timor Sea and Bali Sea, respectively. Banda Sea and Flores Sea are covered by Cluster 3. Meanwhile, Cluster 4 denotes the Indian Ocean. Lastly, Cluster 5 primarily covers the Savu Sea.

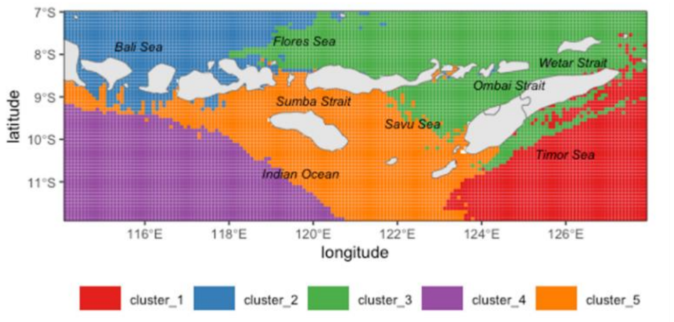

Figure 4. Marine mapping of Lesser Sunda Island waters based on FCM result

The research findings allow for the creation of geographical groups that describe the wide range of environmental conditions in the Lesser Sunda waters region. The zones can serve as a supplement to existing global maritime management schemes, such as Marine Ecoregions of the World (MEOW) systems [56]. Our findings revealed different environmental characteristics throughout the study area (Table 3). In comparison to the other groups, Cluster 4, representing the Indian Ocean, had the highest pH, current velocity, depth, and distance to shore. On the other hand, the Savu Sea (Cluster 5) is characterised by high levels of SSC and salinity; while the highest concentration of nitrate is found at Cluster 1 (Timor Sea). This demonstrates that a single marine protected zone in the Lesser Sunda Seas region is unable to sufficiently reflect their chemical and physical conditions, which determine aquatic organism distribution and richness.

Table 3. Environmental characteristics of Lesser Sunda waters' clusters

Variables	Statistics	Cluster 1 Timor Sea	Cluster 2 Bali Sea	Cluster 3 Banda and Flores Sea	Cluster 4 Indian Ocean	Cluster 5 Savu Sea
pH	Mean	8.2	8.22	8.19	8.23	8.21
	SD	0.010	0.010	0.010	0.000	0.010
	Range	8.18-8.22	8.18-8.24	8.17-8.21	8.22-8.24	8.19-8.24
SSC (mg/m3)	Mean	0.23	0.32	0.23	0.23	0.37
	SD	0.080	0.280	0.090	0.100	0.340
	Range	0.15-1.47	0.17-3.58	0.14-0.93	0.14-0.9	0.15-5.07
SST (oC)	Mean	29.4	29.2	29.3	28.5	28.7
	SD	0.140	0.410	0.240	0.220	0.300
	Range	28.64-30.96	27.63-30.91	28.3-30.74	28.03-29.09	27.57-29.67
Salinity (PSS)	Mean	34.2	33.7	34	34.2	34.3
	SD	0.080	0.130	0.190	0.120	0.180
	Range	33.79-34.47	33.49-34.15	33.41-34.29	33.88-34.45	33.65-34.52
Nitrate (micromol/m3)	Mean	0.63	0.56	0.59	0.56	0.57
	SD	0.020	0.010	0.030	0.000	0.010
	Range	0.56-0.66	0.54-0.6	0.55-0.66	0.56-0.57	0.56-0.61
Depth (m)	Mean	652	782	2634	4489	1574
	SD	763	809	1395	1025	1045
	Range	1 - 3236	2 - 3896	5 - 5415	1320 - 7203	3 - 4213
Distance to shore (km)	Mean	143	23.9	33.5	192	43.5
	SD	73.200	18.900	25.100	74.000	31.700
	Range	1-279	1-88	1-113	52-358	1-136
Slope (degree)	Mean	0.11	0.22	0.5	0.26	0.25

Variables	Statistics	Cluster 1 Timor Sea	Cluster 2 Bali Sea	Cluster 3 Banda and Flores Sea	Cluster 4 Indian Ocean	Cluster 5 Savu Sea
	SD	0.110	0.200	0.400	0.190	0.150
	Range	0-0.66	0-1.15	0.01-2.4	0.01-1.24	0.01-1.26
Current velocity (m/s)	Mean	0.23	0.16	0.23	0.29	0.18
	SD	0.190	0.150	0.150	0.090	0.130
	Range	0-0.6	0-0.93	0-0.74	0.01-0.5	0-0.77
Aspect NS (radians)	Mean	0.27	-0.21	0.05	0	-0.2
	SD	0.590	0.580	0.660	0.670	0.630
	Range	-0.99 - 0.99	-0.99 - 1	-0.99 - 1	-0.99 - 0.99	-0.99 - 1
Aspect EW (radians)	Mean	-0.09	0.13	0.04	-0.02	-0.09
	SD	0.430	0.410	0.440	0.450	0.480
	Range	-0.98 - 0.98	-0.95 - 0.97	-0.97 - 0.97	-0.98 - 0.99	-0.99 - 0.99

It is clear that the Indian Ocean Dipole (IOD) along with El Nino-Southern Oscillation (ENSO) are the two primary natural processes that impact the physical and chemical environment in the Indian Ocean. In addition, the Indian Ocean suffers considerable interannual shifts as an outcome of these phenomena' frequent simultaneous presence. The Indonesian Throughflow (ITF) transports low-salinity tropical waters from the Pacific to the Indian Ocean via Indonesian waterways. The ITF connects the ocean basins as well as is solely tropical oceanic conduit, making it critical to sea distribution and the globe's climate. Its annual variation is mostly caused by ENSO-related air current forcing across the Pacific waveguide. Yet, the IOD may alter the aftermath of the Pacific ENSO through wind fluctuations in the Indian Ocean and also the Indian Ocean waveguide [57]. Because the Indian Ocean is surrounded by multiple nations, its different protected areas are governed separately. According to findings, partnership among these nations is vital for the successful conservation of biological systems and living beings across the region's island and mainland countries [58]. The active role of multinational organizations such as the Wildlife Conservation Society (WCS), The Nature Conservancy (TNC), and the Coral Triangle Initiative could improve the effectiveness of conservation efforts in the study area.

Additionally, the ITF that passes through the Ombai Strait and enters the Savu Sea helps to maintain a physical ecosystem that provides resources for the local population as well as for marine life. Because of its distinct oceanography attributes, such as deep trenches, marine currents, and upwelling zones, the Savu Sea is a key sanctuary and migration channel for cetacean species and sea turtles, a crucial spawning ground, and a perfect habitat for reef systems [59]. Because of this, it has a remarkable diversity of species and many marine animals, especially mammals. It has extremely diversified coral reefs that serve as an important ecosystem for a wide range of marine species. Furthermore, it is a fishing area, which serves as a key source of income and promotes the economic success of the coastal communities [60].

4. Conclusion

Marine environmental mapping is critical to optimizing ecosystem services. Mapping can guide the prioritization of conservation efforts by highlighting areas that provide significant ecosystem services and are also at risk of degradation, as well as identifying the representativeness of the current protected area. Supported by numerous studies on the marine environment that created massive amounts of data, the use of unsupervised learning methods may provide precise and effective strategies for assessing oceanic characteristics. Our study found that the Lesser Sunda waters can be divided into five to nine groups based on significant environmental characteristics utilizing unsupervised clustering methods such as K-means, FCM, DBSCAN, and SOM. According to the internal validity indices, the FCM approach produced the best clustering results for mapping Lesser Sunda waters. The mapping proposed in this study demonstrated a certain extent of consistency with existing conservation efforts, such as MEOW. Furthermore, this result can become a basis for future research to create detailed maps of critical marine habitats, including coral reefs, seagrass beds, and mangroves, using high-resolution satellite imagery and GIS tools to classify habitats and monitor changes over time.

 ISSN 2723-6471

5. Declarations

5.1. Author Contributions

Conceptualization: E.D.L., S.A., N., and A.B.S.; Methodology: S.A.; Software: E.D.L.; Validation: S.A., and A.B.S.; Formal Analysis: E.D.L., S.A., and A.B.S.; Investigation: E.D.L.; Resources: S.A. and A.B.S; Data Curation: S.A. and H.; Writing – Original Draft Preparation: E.D.L., S.A., N., and A.B.S.; Writing – Review and Editing: S.A., E.D.L., and A.B.S.; Visualization: E.D.L.; All authors have read and agreed to the published version of the manuscript.

5.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

5.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

5.4. Institutional Review Board Statement

Not applicable.

5.5. Informed Consent Statement

Not applicable.

5.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] A. Lakshmi, "Coastal ecosystem services & human wellbeing," *Indian Journal of Medical Research*, vol. 153, no. 3, pp. 382–387, 2021.

[2] C. S. Campagne, L.-A. Roy, J. Langridge, J. Claudet, R. Mongruel, D. Beillouin, and É. Thiébaut, "Existing evidence on the impact of changes in marine ecosystem structure and functioning on ecosystem service delivery: a systematic map," *Environ Evid*, vol. 12, no. 1, pp. 1–29, 2023, doi: 10.1186/s13750-023-00306-1.

[3] K. P. Smit, A. T. F. Bernard, A. T. Lombard, and K. J. Sink, "Assessing marine ecosystem condition: A review to support indicator choice and framework development," *Ecol Indic*, vol. 121, no. 2, pp. 1–18, 2021

[4] M. Bozhilova, V. Georgiev, A. Georgieva, and R. Kostadinova, "Planning and Execution of Monitoring of The Marine And Terrestrial Environment for A Large Scale Infrastructure Project in Bulgaria," *15th International Multidisciplinary Scientific GeoConference SGEM*, vol. 1, no. 1, pp. 639–646, Jun. 2015, doi: 10.5593/SGEM.

[5] F. Douvere, "The importance of marine spatial planning in advancing ecosystem-based sea use management," *Mar Policy*, vol. 32, no. 5, pp. 762–771, 2008.

[6] T. Song, C. Pang, B. Hou, G. Xu, J. Xue, H. Sun, and F. Meng, "A review of artificial intelligence in marine science," *Front Earth Sci (Lausanne)*, vol. 11, no. 2, pp. 1–25, 2023.

[7] C. Qian, B. Huang, X. Yang, and G. Chen, "Data science for oceanography: from small data to big data," *Big Earth Data*, vol. 6, no. 2, pp. 236–250, Apr. 2022, doi: 10.1080/20964471.2021.1902080.

[8] E. A. Laws and J. W. Archie, "Appropriate use of regression analysis in marine biology," *Mar Biol*, vol. 65, no. 1, pp. 13–16, 1981, doi: 10.1007/BF00397062.

[9] P. Malanotte-Rizzoli and R. E. Young, "How useful are localized clusters of traditional oceanographic measurements for data assimilation?," *Dynamics of Atmospheres and Oceans*, vol. 17, no. 1, pp. 23–61, 1992.

[10] I. Boldina and P. G. Beninger, "Strengthening statistical usage in marine ecology: Linear regression," *J Exp Mar Biol Ecol*, vol. 474, no. 1, pp. 81–91, Jan. 2016, doi: 10.1016/j.jembe.2015.09.010.

[11] E. Vanem, T. Zhu, and A. Babanin, "Statistical modelling of the ocean environment – A review of recent developments in theory and applications," *Marine Structures*, vol. 86, no. 11, pp. 1–15, 2022.

[12] F. Rolle, F. R. Pennecchi, F. Durbiano, S. Pavarelli, C. Musacchio, G. Coppa, A. Merlone, and M. Sega, "Essential Ocean Variables for Marine Environment Monitoring: Metrological Case Studies," *J Mar Sci Eng*, vol. 11, no. 8, pp. 1–16, 2023, doi: 10.3390/jmse11081605.

[13] B. Sadaiappan, P. Balakrishnan, V. C.R., N. T. Vijayan, M. Subramanian, and M. U. Gauns, "Applications of Machine Learning in Chemical and Biological Oceanography," *ACS Omega*, vol. 8, no. 18, pp. 15831–15853, May 2023, doi: 10.1021/acsomega.2c06441.

[14] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," *SN Comput Sci*, vol. 2, no. 3, pp. 1–21, 2021, doi: 10.1007/s42979-021-00592-x.

[15] J. Wang and F. Biljecki, "Unsupervised machine learning in urban studies: A systematic review of applications," *Cities*, vol. 129, no. 10, pp. 1–20, 2022.

[16] M. Sonnewald, S. Dutkiewicz, C. Hill, and G. Forget, "Elucidating ecological complexity: Unsupervised learning determines global marine eco-provinces," *Sci Adv*, vol. 6, no. 22, pp. 1–11, Sep. 2024, doi: 10.1126/sciadv.aay4740.

[17] D. A. Kroodsma, J. Mayorga, T. Hochberg, N. A. Miller, K. Boerder, F. Ferretti, A. Wilson, B. Bergman, T. D. White, B. A. Block, P. Woods, B. Sullivan, C. Costello, and B. Worm, "Tracking the global footprint of fisheries," *Science (1979)*, vol. 359, no. 6378, pp. 904–908, Feb. 2018, doi: 10.1126/science.aao5646.

[18] Mahrup, M. Ma'shum, and Fahrudin, "Climate change in the Lesser Sunda Islands: The harsh region in the maritim continent of Indonesia," *IOP Conf Ser Earth Environ Sci*, vol. 824, no. 1, pp. 1–15, Jul. 2021, doi: 10.1088/1755-1315/824/1/012115.

[19] G. Perdanahardja and H. Lionata, "Nine years in Lesser Sunda," 2017, Accessed: Jul. 25, 2022. [Online]. Available: https://marineplanning.org/wp-content/uploads/2018/02/Nine_years_Lesser_Sunda.pdf. Link: https://marineplanning.org/wp-content/uploads/2018/02/Nine_years_Lesser_Sunda.pdf

[20] P. P. Linggi and A. Burhanuddin, "The role of coral triangle initiative on coral reefs, fisheries, and food securities in Indonesia's environmental conservation," *IOP Conf Ser Earth Environ Sci*, vol. 343, no. 1, pp. 1–7, Nov. 2019, doi: 10.1088/1755-1315/343/1/012092.

[21] M. Mahmudi, E. D. Lusiana, E. Y. Herawati, and L. G. Serihollo, "Environmental factors and seasonal effect on the potential harmful algae presence at Ambon Bay, Indonesia," *Biodiversitas*, vol. 21, no. 7, pp. 3101–3107, 2020, doi: 10.13057/biodiv/d210730.

[22] E. D. Lusiana, N. R. Buwono, M. Mahmudi, and P. P. Noviasari, "Nutrient limit estimation for eutrophication modelling at Sengguruh Reservoir, Malang, Indonesia," *Nature Environment and Pollution Technology*, vol. 19, no. 1, pp. 361–365, 2020.

[23] E. S. Poloczanska, M. T. Burrows, C. J. Brown, J. García Molinos, B. S. Halpern, O. Hoegh-Guldberg, C. V Kappel, P. J. Moore, A. J. Richardson, D. S. Schoeman, and W. J. Sydeman, "Responses of Marine Organisms to Climate Change across Oceans," *Front Mar Sci*, vol. 3, no. 5, pp. 1–21, 2016.

[24] S. C. Doney, M. Ruckelshaus, J. Emmett Duffy, J. P. Barry, F. Chan, C. A. English, H. M. Galindo, J. M. Grebmeier, A. B. Hollowed, N. Knowlton, J. Polovina, N. N. Rabalais, W. J. Sydeman, and L. D. Talley, "Climate Change Impacts on Marine Ecosystems," *Ann Rev Mar Sci*, vol. 4, no. 1, pp. 11–37, Dec. 2011, doi: 10.1146/annurev-marine-041911-111611.

[25] E. C. Heery *et al.*, "Urban coral reefs: Degradation and resilience of hard coral assemblages in coastal cities of East and Southeast Asia," *Mar Pollut Bull*, vol. 135, no. 10, pp. 654–681, 2018.

[26] F. B. R. Lasram, T. Hattab, G. Halouani, M. S. Romdhane, and F. Le Loc'h, "Modeling of beta diversity in Tunisian waters: Predictions using generalized dissimilarity modeling and bioregionalisation using fuzzy clustering," *PLoS One*, vol. 10, no. 7, pp. 1–16, 2015, doi: 10.1371/journal.pone.0131728.

[27] E. J. Sbrocco and P. H. Barber, "MARSPEC: ocean climate layers for marine spatial ecology," *Ecology*, vol. 94, no. 4, pp. 979–979, Apr. 2013.

[28] J. Morales-Barbero and J. Vega-Álvarez, "Input matters matter: Bioclimatic consistency to map more reliable species distribution models," *Methods Ecol Evol*, vol. 10, no. 2, pp. 212–224, Feb. 2019.

[29] S. Dall'erba, "Exploratory Spatial Data Analysis," *International Encyclopedia of Human Geography*, vol. 1, no. 1, pp. 683–690, 2009.

[30] L. Anselin, "Local Indicators of Spatial Association—LISA," *Geogr Anal*, vol. 27, no. 2, pp. 93–115, Apr. 1995.

[31] Y. Li and H. Wu, "A Clustering Method Based on K-Means Algorithm," *Phys Procedia*, vol. 25, no. 4, pp. 1104–1109, 2012.

[32] S. E. Hashemi, F. Gholian-Jouybari, and M. Hajiaghaei-Keshteli, "A fuzzy C-means algorithm for optimizing data clustering," *Expert Syst Appl*, vol. 227, no. 10, pp. 1–14, 2023.

[33] D. Deng, "DBSCAN Clustering Algorithm Based on Density," *2020 7th International Forum on Electrical Engineering and Automation (IFEEA)*, vol. 1, no. 1, pp. 949–953, 2020, doi: 10.1109/IFEEA51475.2020.00199.

[34] S. Licen, A. Astel, and S. Tsakovski, "Self-organizing map algorithm for assessing spatial and temporal patterns of pollutants in environmental compartments: A review," *Science of The Total Environment*, vol. 878, no. 6, pp. 1–17, 2023.

[35] K. R. Shahapure and C. Nicholas, "Cluster Quality Analysis Using Silhouette Score," *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, vol. 1, no. 1, pp. 747–748, 2020, doi: 10.1109/DSAA49011.2020.00096.

[36] T. Caliński and J. Harabasz, "A dendrite method for cluster analysis," *Communications in Statistics*, vol. 3, no. 1, pp. 1–27, Jan. 1974, doi: 10.1080/03610927408827101.

[37] H. Belyadi and A. Haghighat, "Chapter 4 - Unsupervised machine learning: clustering algorithms," in *Machine Learning Guide for Oil and Gas Using Python*, H. Belyadi and A. Haghighat, Eds., Gulf Professional Publishing, 2021, pp. 125–168.

[38] F. Ros, R. Riad, and S. Guillaume, "PDBI: A partitioning Davies-Bouldin index for clustering evaluation," *Neurocomputing*, vol. 528, no. 4, pp. 178–199, 2023.

[39] R. Todeschini, D. Ballabio, V. Termopoli, and V. Consonni, "Extended multivariate comparison of 68 cluster validity indices. A review," *Chemometrics and Intelligent Laboratory Systems*, vol. 251, no. 8, pp. 1–20, 2024.

[40] I. M. Chohan, A. Ahmad, N. Sallih, N. Bheel, W. M. Salilew, and A. H. Almaliki, "Effect of seawater salinity, pH, and temperature on external corrosion behavior and microhardness of offshore oil and gas pipeline: RSM modelling and optimization," *Sci Rep*, vol. 14, no. 1, pp. 1–18, 2024, doi: 10.1038/s41598-024-67463-2.

[41] K. Seenipandi, K. K. Ramachandran, P. Ghadei, and S. Shekhar, "17 - Ocean remote sensing of suspended sediment variability in Southern Indian coastal water region using Landsat 8 OLI images," in *Remote Sensing of Ocean and Coastal Environments*, M. Rani, K. Seenipandi, S. Rehman, P. Kumar, and H. Sajjad, Eds., Elsevier, 2021, pp. 297–313.

[42] J. A. Camargo, A. Alonso, and A. Salamanca, "Nitrate toxicity to aquatic animals: a review with new data for freshwater invertebrates," *Chemosphere*, vol. 58, no. 9, pp. 1255–1267, 2005.

[43] M. C. Curran, A. L. Ramsey, and A. S. Bower, "Learning about ocean currents one track at a time," *Sci Act*, vol. 58, no. 1, pp. 13–22, Mar. 2021, doi: 10.1080/00368121.2021.1885333.

[44] W. Schneider, D. Donoso, J. Garcés-Vargas, and R. Escribano, "Water-column cooling and sea surface salinity increase in the upwelling region off central-south Chile driven by a poleward displacement of the South Pacific High," *Prog Oceanogr*, vol. 151, no. 2, pp. 38–48, 2017.

[45] J. Rummer and W. Bennett, "Physiological Effects of Swim Bladder Overexpansion and Catastrophic Decompression on Red Snapper," *Trans Am Fish Soc*, vol. 134, no. 1, pp. 1457–1470, Nov. 2005, doi: 10.1577/T04-235.1.

[46] K. Muthukkumaran and K. S. Arun, "Effect of seabed slope on pile behaviour of fixed offshore platform under lateral forces," *J Ocean Eng Mar Energy*, vol. 1, no. 3, pp. 207–221, 2015, doi: 10.1007/s40722-015-0013-7.

[47] F. Bauer, A. M. Knights, M. E. Hanley, J. N. Griffin, A. Foggo, A. Brown, and L. B. Firth, "Topography-based modulation of environmental factors as a mechanism for intertidal microhabitat formation: A basis for marine ecological design," *Mar Pollut Bull*, vol. 202, no. 5, pp. 1–14, 2024.

[48] A. A. Abdulnassar and L. R. Nair, "Performance analysis of Kmeans with modified initial centroid selection algorithms and developed Kmeans9+ model," *Measurement: Sensors*, vol. 25, no. 2, pp. 1–7, 2023.

[49] H. Yan, L. Wang, and Y. Lu, "Identifying cluster centroids from decision graph automatically using a statistical outlier detection method," *Neurocomputing*, vol. 329, no. 2, pp. 348–358, 2019.

[50] S. A. Mingoti and J. O. Lima, "Comparing SOM neural network with Fuzzy c-means, K-means and traditional hierarchical clustering algorithms," *Eur J Oper Res*, vol. 174, no. 3, pp. 1742–1759, Nov. 2006, doi: 10.1016/j.ejor.2005.03.039.

[51] J. Heil, V. Häring, B. Marschner, and B. Stumpe, "Advantages of fuzzy k-means over k-means clustering in the classification of diffuse reflectance soil spectra: A case study with West African soils," *Geoderma*, vol. 337, no. 3, pp. 11–21, 2019.

[52] V. Lucieer and A. Lucieer, "Fuzzy clustering for seafloor classification," *Mar Geol*, vol. 264, no. 3, pp. 230–241, 2009.

[53] W. Zhao, J. Ma, Q. Liu, J. Song, M. Tysklind, C. Liu, D. Wang, Y. Qu, Y. Wu, and F. Wu, "Comparison and application of SOFM, fuzzy c-means and k-means clustering algorithms for natural soil environment regionalization in China," *Environ Res*, vol. 216, no. 1, pp. 1–10, 2023.

[54] M. De Cáceres, X. Font, and F. Oliva, "The management of vegetation classifications with fuzzy clustering," *Journal of Vegetation Science*, vol. 21, no. 6, pp. 1138–1151, 2010.

[55] M. A. E. Forio, A. Mouton, K. Lock, P. Boets, T. H. T. Nguyen, M. N. Damanik Ambarita, P. L. S. Musonge, L. Dominguez-Granda, and P. L. M. Goethals, "Fuzzy modelling to identify key drivers of ecological water quality to support decision and policy making," *Environ Sci Policy*, vol. 68, no. 2, pp. 58–68, 2017.

[56] M. D. Spalding, H. E. Fox, G. R. Allen, N. Davidson, Z. A. Ferdaña, M. Finlayson, B. S. Halpern, M. A. Jorge, A. Lombana, S. A. Lourie, K. D. Martin, E. McManus, J. Molnar, C. A. Recchia, and J. Robertson, "Marine ecoregions of the world: A bioregionalization of coastal and shelf areas," *Bioscience*, vol. 57, no. 7, pp. 573–583, Jul. 2007, doi: 10.1641/B570707.

[57] J. Sprintall and A. Révelard, "The Indonesian Throughflow response to Indo-Pacific climate variability," *J Geophys Res Oceans*, vol. 119, no. 2, pp. 1161–1175, Feb. 2014.

[58] N. Levin, M. Beger, J. Maina, T. McClanahan, and S. Kark, "Evaluating the potential for transboundary management of marine biodiversity in the Western Indian Ocean," *Australasian Journal of Environmental Management*, vol. 25, no. 1, pp. 62–85, Jan. 2018, doi: 10.1080/14486563.2017.1417167.

[59] M. Mujiyanto, A. S. Nastiti, and R. Riswanto, "EFFECTIVENESS OF SUB ZONE CETACEAN PROTECTION IN MARINE PROTECTED AREAS SAVU SEA NATIONAL MARINE PARK, EAST NUSA TENGGARA," *Coastal and Ocean Journal (COJ)*, vol. 1, no. 2, pp. 1–12, Nov. 2017, doi: 10.29244/coj.1.2.1-12.

[60] M. Maestro, M. L. Pérez-Cayeiro, J. A. Chica-Ruiz, and H. Reyes, "Marine protected areas in the 21st century: Current situation and trends," *Ocean Coast Manag*, vol. 171, no. 4, pp. 28–36, 2019.