A Model-based Analysis of Changes of the Coastal Urban Areas at Semarang City, Indonesia: A Comparison of Machine Learning Classifiers on Optical Satellite Imageries Data

R D L R Manik [1], **A W Wijayanto** [1,2,*]

[1] Department of Statistical Computing, Politeknik Statistika STIS, Jakarta, Indonesia
[2] BPS-Statistics Indonesia, Jakarta, Indonesia

*Corresponding author's e-mail: ariewahyu@stis.ac.id

Abstract. A coastal area is defined as the boundary between land and sea. Coastal urban areas are susceptible to various hazards that are becoming more severe, such as flooding, erosion, and subsidence due to a mix of man-made and natural factors, including urbanization and climate change. Regardless of the high importance of coastal area monitoring, conducting field surveys is expensive, time-consuming, and geographically limited to non-remote regions. Semarang City is one of the cities in Indonesia that is at risk of changes in its coastline and causes various natural problems. This research aims to estimate changes in the coastal land area in Semarang City. In observing the phenomenon of changes in area in coastal areas in Semarang City, remote sensing technology with Sentinel-2 satellite imagery was used. This research implements and compares the Random Forest (RF) and Support Vector Machine (SVM) machine learning methods in building classification models. From the results of land area in 2019, 2021, and 2023 with the best classification model, namely SVM, information was obtained on an increase in coastal area of 387.94 ha in 2021, then a change in area decrease of 417.32 ha in 2023.

1. Introduction

A coastal area is an area with the characteristics of a dry part and a wet part, which comes from loose sand and river deposits [1]. The coast is between the lowest point of seawater and the practical limit of the waves reaching the land. [1]. Coastal areas have much natural potential and are one of the areas with the most utilization [1]. In general, coastal areas can be used as settlements, tourism, ports, industry, and others to support community activities [2]. Coastal areas always experience line and morphological changes due to abrasion and accretion.

Coastal areas generally depend on the balance of rocks, waves, topography, and wind. Therefore, sooner or later, the coastal area will always experience changes. The coastline is the boundary between sea and land waters whose nature changes based on waves, currents, and tides [1]. The changing nature of the coastline impacts the coastal area, one of which is a change in the beach area. According to [3], coastal slope profile, inland water contours, and coastline morphology can influence the magnitude of coastline changes.

Line changes and coastal morphology occur due to abrasion and accretion events [4]. According to [5], abrasion is the loss of land in coastal areas due to changes in sediment transport balance. Accretion is the emergence of new land in coastal areas generally caused by nature and humans [5]. The phenomena of abrasion and accretion not only have an impact on regional conditions but also have an impact on social and economic problems in the community. Abrasion can result in community land loss, resulting in infrastructure damage. Likewise, accretion, which at first glance seems to benefit the community, but in reality, it causes conflicts over competition for land use and control within the community.

Semarang City is Central Java Province's capital, located north of Java Island, at the coast of the Java Sea. Semarang City is one of the cities vulnerable to the threat of changes in the coastline [6]. In [7], According to reports, the coastline of Semarang City has experienced significant changes. Specifically, the abrasion phenomenon has affected an area of 46.77 hectares, causing 45.72 kilometers of coastline to shift. Additionally, the accretion phenomenon has impacted an area of 165.95 hectares. Since 1972, the city of Semarang has been predicted to experience a shift in coastline and will continue to increase until now [6]. In addition, in [6], it was noted that 49.54 m of the coastline in Semarang City had shifted in 12 years, and there had been a change in the area of 241.9 Ha due to beach reclamation [8]. This case is a severe threat regarding natural conditions in Semarang. Tidal flooding is one of the significant phenomena and problems in the city of Semarang due to a shift in the coast [6].

The recent advances in remote sensing satellite imagery are unique and objective in accurately and efficiently identifying physical and geographical elements from a wide range of scales [10,11]. It provides abundant valuable resources [12,13]. Lower prices, periodic updates, and granularity of area-covered representation are advantages of employing these satellite pictures [14,15]. The ability to gather and preprocess remote sensing satellite imagery, as well as to speed up classification and clustering activities [15,16], has made it possible to use a multitude of cloud-based platforms for analyzing and storing Earth observation data in recent times [11,13].

In this study, we map changes in the coastal area of Semarang City in 2023 using remote sensing based on the described conditions. Remote sensing technology can be applied to study area changes in coastal areas. To the best of our knowledge, this is the first study that models coastal area change analysis of coastal urban areas at Semarang City, Indonesia by comparing machine learning classifiers on optical satellite imagery data. We compared the beach area in 2019 as a reference, 2021 and 2023. This mapping used Sentinel-2 satellite imagery and combined multispectral bands with different composite indices, including Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), and Normalized Difference Built-Up Index (NDBI). Remote sensing using Sentinel-2 satellite imagery is also more suitable than Landsat imagery with a smaller resolution [17]. In [17], satellite imagery with a higher resolution can reveal smaller landforms to provide more accurate results. In addition to using satellite imagery, this research implements a machine learning algorithm to find the best model for classification between land and water areas.

2. Study Area

This study is focused on Semarang City, Central Java Province, Indonesia, on northern Java Island bordered by the Java Sea. Semarang City was selected as it is one of the cities in Indonesia that is vulnerable to the threat of changes in the area of the coast and causes various natural problems [6,7]. The city spans 373.70 km2. The city of Semarang consists of 16 districts and 177 urban villages with astronomical boundaries located at 6° 50′ – 7° 10′ south latitude and 109° 35′ – 110° 50′ east longitude [9]. The areas of Semarang City directly adjacent to the Java Sea are Tugu sub-district, West Semarang sub-district, North Semarang sub-district, and Genuk sub-district. Apart from the Java Sea, Semarang City is directly adjacent to Demak Regency to the east, Kendal Regency to the west, and Semarang Regency to the south [9].

Figure 1. Focus Study Area of Semarang City, Indonesia

3. Research Methods

3.1. Research Workflows

In this study, we use remote sensing methods with satellite imagery to map changes in coastal areas in Semarang. A logical flow is established to conduct thorough research, starting with the issue of annual changes in coastal areas. We propose solutions by applying remote sensing and machine learning, with several research objectives, to use accuracy, precision, recall, and F1-Score as evaluation indicators. The research framework is depicted in Figure 2, referencing M. Berndtsson et al. [18].

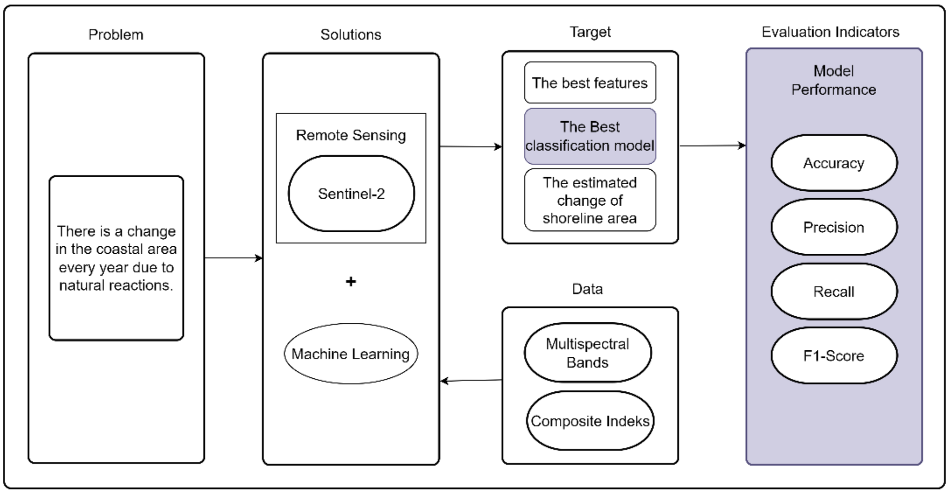


Figure 2. Research Framework

3.2. Data Collection

In this study, the data used is Sentinel-2 Multispectral Level-2A imagery data recorded in July 2019, June 2021, and June 2023, where the timing was made considering less cloud cover. The image data is obtained using the Google Earth Engine.

The image data obtained is followed by cloud masking to reduce cloud cover and shadows on the satellite image data and facilitate the process of object labeling. Furthermore, the image data is clipped to filter the object geometry that intersects with the given object geometry. The image data results clipped are Tugu sub-district, West Semarang sub-district, North Semarang sub-district, and Genuk sub-district and parts of the Java Sea area.

3.3. Feature Identification

Feature identification is performed by calculating the mean value of each multispectral band and composite index for the land and ocean classes. The Natural Break algorithm will group the results of the mean into three categories, namely "Low," "Medium," and "High." The results of categorizing the multispectral band mean and the composite index is compared and identified regarding which features can distinguish between land and ocean classes.

Table 1. Multispectral Bands and Composite Index of Sentinel-2 Satellite Imagery

Description	Multispectral Band and Composite Index
Blue	B2
Green	B3
Red	B4
Vegetation Red Edge 1	B5
Vegetation Red Edge 2	B6
Vegetation Red Edge 3	B7
Vegetation Red Edge 4	B8A
Near Infrared (NIR)	B8
Water Vapour	B9
Shortwave Infrared (SWIR) 1	B11
SWIR2	B12
Normalized Differenced Vegetation Index (NDVI)	$\frac{B8-B4}{B8+B4}$
Normalized Differenced Water Index (NDWI)	$\frac{B3-B8}{B3+B8}$
Normalized Differenced Built-Up Index (NDBI)	$\frac{B11-B8}{B11+B8}$

3.4. Machine Learning Classifiers

Machine Learning is one of the algorithms and methods applied to design and implement systems by learning from existing data to conclude the results obtained from the data obtained [19]. Based on data availability and expected output results, machine learning algorithms are divided into supervised, unsupervised, semi-supervised, and reinforcement learning [19]. The use of machine learning can assist in the process of collecting data using remote sensing techniques. It can determine the relationship between components such as reflectance and features of interest [24].

The machine learning algorithms used in this research are Support Vector Machine (SVM) and Random Forest. Random Forest is one of the most widely applied algorithms in classification research with various types of data, where creating more than one decision tree can improve the classification accuracy of the algorithm [20]. In machine learning, SVM designs a hyperplane as a decision boundary to maximize the distance between samples and divide classes [21]. One of the advantages of using the SVM algorithm is that it can achieve high accuracy values with minimal training data sets, which in turn can also help reduce the burden of collecting testing data [21].

4. Result

In this study, 2 categories were classified, namely the classification of land and sea areas. The land category is the environment that is above the water surface, consisting of land, buildings, roads, plantations, fields (whether covered with plants or not), and other areas that do not interact directly with the ocean. In the ocean category is a vast expanse of sea to the north of Java Island, namely the Java Sea.

The mean value of the multispectral band and composite index for each class was calculated to identify the best features in recognizing land and sea areas. The mean results are calculated using the Natural Break algorithm and grouped into 2 categories: "Low" and "High." The features that distinguish the land class from the sea well can be identified from these results.

Table 2. Results of Categorization of Mean Feature Values in Sentinel-2 Satellite Imagery

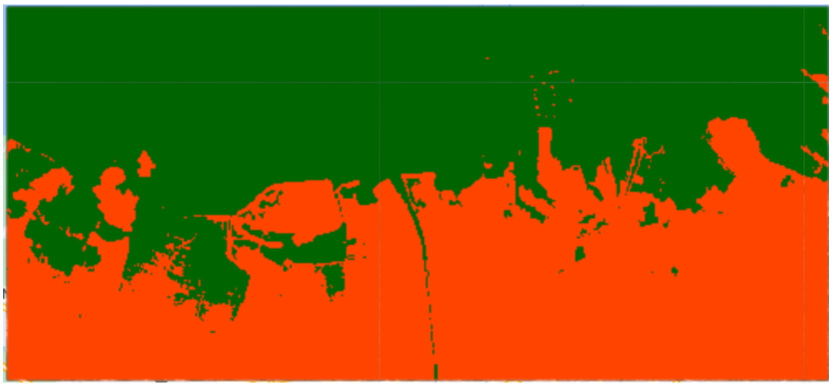
Features	Land Areas	Seas
Blue (B2)	High	Low
Green (B3)	High	Low
Red (B4)	High	Low
Vegetation Red Edge 1 (B5)	High	Low
Vegetation Red Edge 2 (B6)	High	Low
Vegetation Red Edge 3 (B7)	High	Low
NIR (B8)	High	Low
Vegetation Red Edge 4 (B8A)	High	Low
Water Vapour (B9)	High	Low
SWIR 1 (B11)	High	Low
SWIR 2 (B12)	High	Low
NDVI	High	Low
NDWI	Low	High
NDBI	Low	High

Based on the categorization results in Table 2, information is obtained that the features in the Sentinel-2 image that can differentiate images of land areas from ocean areas are Blue Ribbon (B2), Green (B3), Red (B4), Vegetation Red Edge 1 (B5), Vegetation Red Edge 2 (B6), Vegetation Red Edge 3 (B7), NIR (B8), Vegetation Red Edge 4 (B8A), Water Vapor (B9), SWIR 1 (B11), SWIR 2 (B12) and Composite Index NDVI.

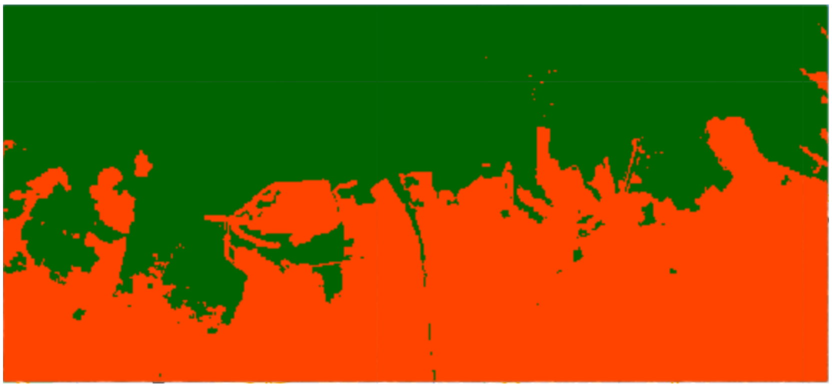
After analyzing the feature identification, the supervised classification uses a machine learning algorithm. The algorithms used are Random Forest and SVM. Then, validation was carried out using cross-validation to overcome the overfitting of the training data. The classification results of the two algorithms can be seen in Figure 2, where the green color shows the ocean area and the red color shows the land area in Semarang City in 2019.

Table 3. Comparison of Accuracy and F1-Score

Algorithm	Overall Accuracy	*F1-Score*
Random Forest (RF)	98.67 %	98.67 %
Support Vector Machine (SVM)	99.33 %	99.33 %



(a)

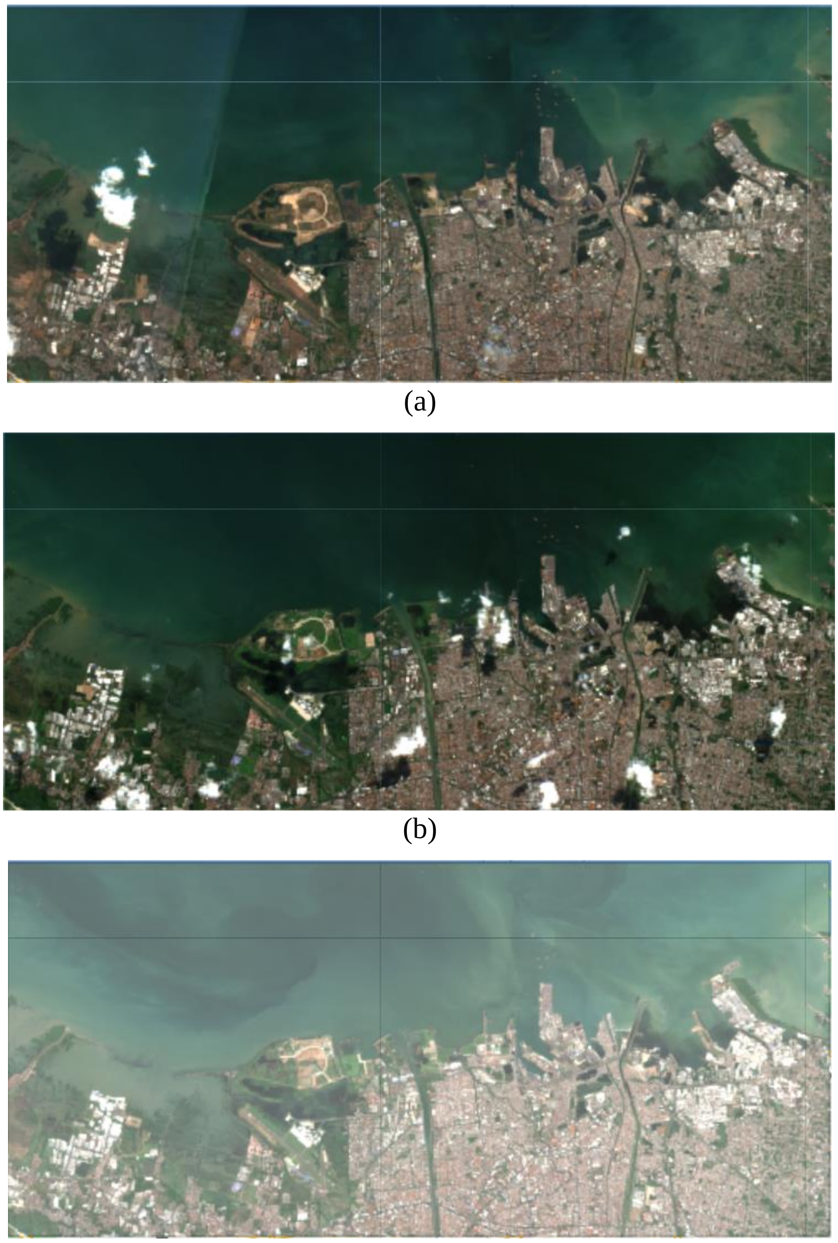


(b)

Figure 3. Land and sea classification results using (a) Random Forest (RF) and (b) Support Vector Machine (SVM) methods

Based on the results in Table 3, the Support Vector Machine (SVM) classification method is the method with the highest accuracy in classifying land class, shown by the high Accuracy value of 99.33% and the F1-Score, which is also 99.03%. The Random Forest (RF) classification method has lower accuracy in classifying land classes. However, this method still shows high accuracy and f1-score results in terms of an Accuracy value of 98.67% and an F1-Score of 98.67%. The classification model formed will then become the basis for estimating changes in coastal areas.

After obtaining the best model, the number of pixels and land area in Semarang City will be calculated in 2019, 2021, and 2023. The number of pixels and land area obtained will then be compared to see changes in the area in 2019, 2021, and 2023.



(a)

(b)

(c)

Figure 4. View of the land and sea areas of Semarang City in (a) 2019, (b) 2021, and (c) 2023

From the land area results in 2019, 2021, and 2023, information was obtained that there was a change in coastal area, increasing by 387.94 ha in 2021 from 2019. Then, in 2023, the coastal area experienced a change in area, decreasing by 417.32 ha. Changes in increasing and decreasing the line and area of the coast can be caused by abrasion and accretion in coastal areas, especially reclamation or addition of land in the city of Semarang.

5. Conclusion

Based on the written results and discussion, the best identification features were obtained for recognizing land and sea images. Band features Blue (B2), Green (B3), Red (B4), Vegetation Red Edge 1 (B5), Vegetation Red Edge 2 (B6), Vegetation Red Edge 3 (B7), NIR (B8), Vegetation Red Edge 4 (B8A), Water Vapor (B9), SWIR 1 (B11), SWIR 2 (B12) and the NDVI Composite Index can recognize land images better. In contrast, the NDWI and NDBI Composite Index features can better recognize ocean area images than others. The best land and sea area classification model is obtained by identifying the best features, namely the Support Vector Machine (SVM) algorithm. This algorithm provides accuracy and f1-score of 0.993. This model calculates the land area by counting the number of pixels and then converting it to hectares. In 2021, the land area will increase from 2019 to 387.94 hectares. However, in 2023, the land area will also decrease from 2021, amounting to 417.32 hectares.

References

[1] Opa E T 2011 Perubahan Garis Pantai Desa Bentenan Kecamatan Pusomaen, Minahasa Tenggara *Jurnal Perikanan dan Kelautan Tropis* **7** 109-114

[2] Isdianto A, Asyari I M, Haykal M F, Adibah F, Irsyad M J and Supriyadi S 2020 Analisis Perubahan Garis Pantai Dalam Mendukung Ketahanan Ekosistem Pesisir *Jukung (Jurnal Teknik Lingkungan)* **6**

[3] Islam H S, Suryoputro A A D and Handoyo G 2023 Studi Perubahan Garis Pantai 2017–2021 di Pesisir Kabupaten Batang, Jawa Tengah *Indonesian Journal of Oceanography* **4** 19-33

[4] Singa J C B G, Nuarsa I W and Putra I G N 2023 Deteksi Perubahan Garis Pantai Menggunakan Citra Satelit Sentinel-2 di Kabupaten Klungkung, Bali *Journal of Marine and Aquatic Sciences* **9** 70-81

[5] Munandar M and Kusumawati I 2017 Studi analisis faktor penyebab dan penanganan abrasi pantai di wilayah Pesisir Aceh Barat *Jurnal Perikanan Tropis* **4** 47-56

[6] Amalia F, Zairion Z and Atmadipoera A S 2023 Perubahan Garis Pantai Selama 20 Tahun (2001-2021) dan Prediksi dan Adaptasi Masyarakat Pesisir Tahun 2041 *Jurnal Sains dan Teknologi* **12** 102-110

[7] Sardiyatmo S, Supriharyono S and Hartoko A 2013 Dampak Dinamika Garis Pantai Menggunakan Citra Satelit Multi Temporal Pantai Semarang Provinsi Jawa Tengah (Study of the Dynamics of Image Using Satellite Beach Line Multi-Temporal Beach Semarang Central Java Province) *Saintek Perikanan: Indonesian Journal of Fisheries Science and Technology* **8** 33-37

[8] Marques J N and Khakhim N 2016 Kajian perubahan garis pantai menggunakan citra landsat multitemporal di Kota Semarang *Jurnal Bumi Indonesia* **5**

[9] BPS Kota Semarang. (2022). Kota Semarang Dalam Angka 2022. Semarang: Badan Pusat Statistik Kota Semarang

[10] Saadi T D T and Wijayanto A W 2021 Machine learning applied to Sentinel-2 and Landsat-8 multispectral and medium-resolution satellite imagery for the detection of rice production areas in Nganjuk, East Java, Indonesia *International Journal of Remote Sensing and Earth Sciences* **18** 19-32

[11] Wijayanto AW, Triscowati DW, Marsuhandi AH 2020 Maize Field Area Detection in East Java, Indonesia: An Integrated Multispectral Remote Sensing and Machine Learning Approach. 2020 12th International Conference on Information Technology and Electrical Engineering (ICITEE).

[12] Nurmasari Y, Wijayanto AW 2021 Oil Palm Plantation Detection in Indonesia using Sentinel-2 and Landsat-8 Optical Satellite Imagery (Case Study: Rokan Hulu Regency, Riau Province), International Journal of Remote Sensing and Earth Sciences (IJReSES), 18, 1, 1-18, LAPAN

[13] Putri SR, Wijayanto AW 2022 Learning Bayesian Network for Rainfall Prediction Modeling in Urban Area using Remote Sensing Satellite Data (Case Study: Jakarta, Indonesia), Proceedings of The International Conference on Data Science and Official Statistics, 2021, 1, 77-90

[14] Afira N, Wijayanto AW 2022 Mono-temporal and multi-temporal approaches for burnt area detection using Sentinel-2 satellite imagery (a case study of Rokan Hilir Regency, Indonesia), Ecological Informatics, 69, 101677, Elsevier

[15] Wijayanto AW, Afira N, Nurkarim W 2022 Machine Learning Approaches using Satellite Data for Oil Palm Area Detection in Pekanbaru City, Riau, Proceedings of the 2022 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom).

[16] Damayanti AR, Wijayanto AW 2021 Comparison of Hierarchical and Non-Hierarchical Methods in Clustering Cities in Java Island using the Human Development Index Indicators year 2018, Eigen Mathematics Journal, 4, 1, Universitas Mataram

[17] Zaidan R R, Suryono C A, Pratikto I and Taufiq-Spj N 2022 Penggunaan Citra Satelit Sentinel 2A untuk Mengevaluasi Perubahan Garis Pantai Semarang Jawa Tengah *Journal of Marine Research* **11** 105-113

[18] Berndtsson, Hansson J, Olsson B and Lundell B 2008 *Thesis Projects— A Guide for Students in Computer Science and Information System 2nd Edition* Springer London

[19] Tsiakos C A D and Chalkias C 2023 Use of Machine Learning and Remote Sensing Techniques for Shoreline Monitoring: A Review of Recent Literature *Applied Sciences* **13** 3268

[20] Junaid M, Sun J, Iqbal A, Sohail M, Zafar S and Khan A 2023 Mapping LULC Dynamics and Its Potential Implication on Forest Cover in Malam Jabba Region with Landsat Time Series Imagery and Random Forest Classification *Sustainability* **15** 1858

[21] Tariq A, Jiango Y, Li Q, Gao J, Lu L, Soufan W, Khalid F. Almutairi and Habib-ur-Rahman M 2023 Modelling, mapping and monitoring of forest cover changes, using support vector machine, kernel logistic regression and Naive Bayes tree models with optical remote sensing data *Heliyon* **9**

[22] Sharifani K and Amini M 2023 Machine Learning and Deep Learning: A Review of Methods and Applications *World Information Technology and Engineering Journal* **10** 3897-3904

[23] Shirmard H, Farahbakhsh E, Muller R B and Chandra R 2022 A Review of Machine Learning in Processing Remote Sensing Data for Mineral Exploration *Remote Sensing of Environment* **268**