This version of the article has been accepted for publication, after peer review (when applicable) and is subject to Springer Nature's AM terms of use, but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections.

The Version of Record is available online at: https://doi.org/10.1007/s40808-024-01957-w

Advancing Coastal Habitat Mapping in Bahrain: A Comparative Study of Remote Sensing Classifiers

Manaf Alkhuzaei[1*] and Dr Matthew Brolly[2]

[1*]Centre for Earth Observation Science, University of Brighton, Lewes Rd, Brighton, BN2 4GJ, East Sussex, UK.

[2]Centre for Earth Observation Science, University of Brighton, Lewes Rd, Brighton, BN2 4GJ, East Sussex, UK.

*Corresponding author(s). E-mail(s): mkhuzaei94@gmail.com;
Contact: +44(0)7502372323;
Contributing authors: m.brolly@brighton.ac.uk;

Abstract

This study explores the veracity of remote sensing-based classification of Bahrain's coastal water habitats, referencing data from three prior studies conducted in Bahrain waters. The objective is to illuminate the limitations of remote sensing for habitat mapping and evaluate the proficiency of Maximum Likelihood, Support Vector Machines, and SoftMax Regression classifiers using expanded pre-processing methodologies. Our reanalysis yielded Maximum Likelihood accuracies ranging from 45-61%, significantly influenced by the choice of data source and ground truth selection. In contrast, the Support Vector Machines classifier consistently demonstrated superior accuracy, achieving up to 80% by utilizing diverse band combinations. Meanwhile, SoftMax Regression classifiers reached peak accuracies of 58-69%. These findings emphasize the critical importance of selecting robust data sources, comprehending classifier limitations, and ensuring accurate ground truth verification to enhance the reliability of habitat classification maps. Such maps are instrumental for effective coastal management, conservation efforts, and marine biodiversity assessments, facilitating sustainable resource use and conservation strategies. Given the rapid advancement in remote sensing technologies, optimizing the precision of habitat classification maps is paramount for sustainable coastal ecosystem management and biodiversity preservation, highlighting the relevance and significance of this study in the broader context of environmental research and management.

Keywords: Habitat Classification, Support Vector Machines, SoftMax Regression

1 Introduction

Marine ecosystems, possessing a vast biodiversity network and ecological significance, have been at the forefront of scientific research in recent decades. Paramount to this research is understanding benthic habitats - including seagrasses, coral reefs, and macroalgae - which are indispensable for effective ecosystem management (Traganos and Reinartz 2018; Vahtmäe et al. 2020b). As vital components of the Earth's biosphere, marine benthic habitats harbour diverse life and serve as crucial indicators of broader ecological health. Understanding these habitats' spatial distribution and characteristics is paramount for conservation and resource management, particularly in regions experiencing rapid environmental change.

In the context of the Persian/Arabian Gulf, the mapping of benthic habitats acquires additional significance due to the unique ecological characteristics of this region. Recent studies have underscored the importance of such efforts. For instance, Purkis and Riegl (2005) quantified the spatial and temporal dynamics of Arabian Gulf coral assemblages using remote-sensing and in situ monitoring data, providing valuable insights into the ecological health of these vital ecosystems. Similarly, Kabiri et al. (2018) demonstrated the potential of high-resolution satellite imagery for mapping coral habitats around Hendorabi Island in the Persian Gulf, coupling Worldview-2 standard imagery with field observations to achieve detailed mapping. Furthermore, Grizzle et al. (2016) contributed to our understanding of the current status of coral reefs in the United Arab Emirates, including their distribution, extent, and community structure. These studies highlight the crucial role of remote sensing in advancing our knowledge of marine habitats in the Persian/Arabian Gulf and the implications for management and conservation efforts. The significance of such mappings is particularly pronounced in Bahrain, where the marine ecosystem plays a pivotal role in the country’s economy and ecology.

Benthic habitat mapping has recently emerged as a pivotal tool for understanding and managing the marine ecosystems along the Arabian Peninsula. Using innovative technologies, researchers uncover new insights into the spatial distribution, morphology, and composition of coral reefs, seagrass, macroalgal habitats, and associated benthic assemblages. In Saudi Arabia's Red Sea, Watts (2022) employed high-resolution acoustic technology alongside remotely operated vehicles to map and classify the reefs in Thuwal's coastal area. This groundbreaking study filled a critical information gap, revealing 23 different benthic habitat types and uncovering previously understudied reef morphologies in the Red Sea, thereby improving our comprehension of the spatial distribution of benthic communities and contributing essential data for conservation and future research efforts in the region.

Further west, Gholoum et al. (2019) harnessed the capabilities of high spatial resolution satellite images (WorldView-2 (WV-2) and Pleiades-1B) to explore coral density mapping in the State of Kuwait. By integrating the Multiple Linear Regression (MLR)

model with multispectral satellite sensors, they established a new image classification approach demonstrating varying degrees of reliability. Their work emphasizes the potential to detail coral density more accurately, though noting the challenges posed by the resolution capabilities of remote sensors and the intrinsic complexity of coral reef ecosystems. Their findings also suggest the possible enhancement of accuracy by including factors such as water turbidity, temperature, and salinity, highlighting a path for future development in the mapping process.

Furthermore, utilising high-resolution remote sensing, Butler et al. (2020) accomplished a comprehensive benthic habitat mapping of the coastal zone and Halul Island in Qatar. By contrasting pixel- and object-based classifiers, they successfully developed a country-wide map covering 4500 km^2, unveiling the complex interplay of various marine habitats. Their research stands as a benchmark in the region, illustrating the efficacy of object-based methods over pixel-based classifiers and delivering a valuable resource for ecosystem-based management decision-making.

These studies illustrate the rapid evolution and growing significance of benthic habitat mapping in the Arabian region. From the detailed acoustic examination of the Red Sea's reefs to advanced satellite imaging across Kuwait and Qatar's coastal zones, the innovative methodologies and findings delineate a vibrant picture of the area's marine habitats. Integrating new technologies and analytical approaches enhances understanding of these complex ecosystems and offers essential conservation, management, and ongoing exploration tools. This unfolding knowledge underscores the importance of continued research and technological advancement in benthic habitat mapping, which is crucial for the sustainable stewardship of the region's marine environment.

Comprehensive studies of its benthic ecosystems have been relatively scarce in the Kingdom of Bahrain, a nation whose economy and ecology are intimately tied to its surrounding marine environment. This lack of focused research underscores the importance of the pioneering surveys conducted by Vousden (1995), Geomatec (2006), and (Aljenaid et al. 2017), each offering unique insights into Bahrain's marine habitats through innovative methodologies. Vousden's groundbreaking 1995 thesis provided an in-depth study, employing satellite remote sensing, aerial photography, and in-field 'ground-truthing.' Geomatec's 2006 report aimed to create a Marine Environmental Geographic Information System (MARGIS II), merging remote sensing and GIS techniques. Aljenaid et al. (2017) further advanced the field, employing Landsat-8 imagery and achieving an overall accuracy of 84.1% in mapping diverse habitats.

Benthic habitat mapping provides comprehensive insights into these habitats' spatial distribution and dynamic changes. Remote sensing, with its benefits of cost-effectiveness, broad coverage, and real-time data provision, has emerged as a valuable tool for studying shallow marine and freshwater benthic ecosystems (Hedley et al. 2016; Ha et al. 2020). Despite these benefits, questions about the accuracy and reliability of these techniques remain (Park et al. 2019), particularly in Bahrain's coastal waters,

characterized by high suspended sediment and turbidity levels due to natural processes and intensified anthropogenic activities (Vousden 1995; Ali 2022).

Previous studies in Bahrain have made strides in this area, utilizing unique methodologies. However, each has limitations, from overlooking potential sensor and atmospheric disturbances to missing complex habitat variations. Despite these enhancements, Aljenaid et al.'s (2017) usage of k-means unsupervised classification did not always yield the best results due to its susceptibility to spectral overlap among different habitat types. This study builds on these previous works by introducing a methodology that enhances the pre-processing methods and employs multiple classification approaches. Through the evaluation of Maximum Likelihood (ML), Support Vector Machines (SVM), and SoftMax Regression (SMR) classifiers in Bahrain's dynamic coastal waters, this research aims to revisit and expand upon previously conducted benthic habitat surveys in Bahrain's waters, leveraging ground truth data and enhancing pre-processing methods to recreate and refine the existing maps. By scrutinizing the data collection and application of known pre-processing techniques, this study seeks to investigate whether the accuracy of these habitat maps can be improved. This endeavour contributes significantly to the existing body of knowledge and offers practical applications for developing more precise and reliable marine resource management strategies, conservation initiatives, and coastal development planning. Recognizing that well-documented data coupled with advanced pre-processing can enhance classification accuracy, this study ultimately seeks to pave the way for the sustainable management and preservation of marine ecosystems in Bahrain and potentially other complex coastal environments.

Collectively, these seminal studies have laid critical groundwork for understanding the marine environment in Bahrain, recognizing that well-documented data coupled with advanced pre-processing can enhance classification accuracy. The insights and methodologies provided represent pioneering efforts in this direction and warrant further exploration, expansion, and application within Bahrain's unique marine environment, paving the way for the sustainable management and preservation of marine ecosystems in Bahrain and potentially other complex coastal environments.

2 Study Area

The Kingdom of Bahrain is an archipelago with approximately 40 natural and human-made islands of varying sizes, as shown in Figure 1. The country's total area is about 782 km^2, with a total coastal length exceeding 537 km. The marine area encompasses more than 9,200 km^2 of the Arabian Gulf. Bahrain is situated in the southern part of the Arabian Gulf, approximately between longitudes 50°16' and 51°00' east and latitudes 25°33' and 27°12' north(Information and eGovernment Authority 2020). This partially enclosed sea is often called the Gulf of Bahrain, with its southern extension being the Gulf of Salwah (Dawhat Salwah) (Vousden 1995). The winds in Bahrain are predominantly northerly and north-westerly throughout the year, influencing the

marine environments through the wave action generated. These winds have a substantial impact due to friction and tidal velocity reduction.

Additionally, small waves generated in shallow water depths cause net motion in the direction of the wind, modifying the flow regime in speed and direction (Alkuzai et al. 2009). The Arabian Gulf is a shallow body of water with an average depth of around 35m (Alkuzai et al. 2009). The temperature ranges from 23 to 26°C in winter and 30 to 32°C in summer, with a salinity range of 36.5 to 37.2 PSU (Stéphane and Pascal 2015). These conditions and the limited exchange with the Indian Ocean have resulted in low species diversity. Despite these harsh conditions, the Gulf is known to be relatively rich in habitats and biota. Although Bahrain has a small land area, its territorial waters support numerous valuable ecosystem services related to seagrass beds, coral reefs, mangrove swamps, and mudflats (Naser 2016). These contribute significantly to producing marine resources in the Arabian Gulf (Vousden 1995; Khan et al. 2002; Naser 2011). Most of Bahrain's productive benthic habitats are located in Bahrain's north, northeast, and east

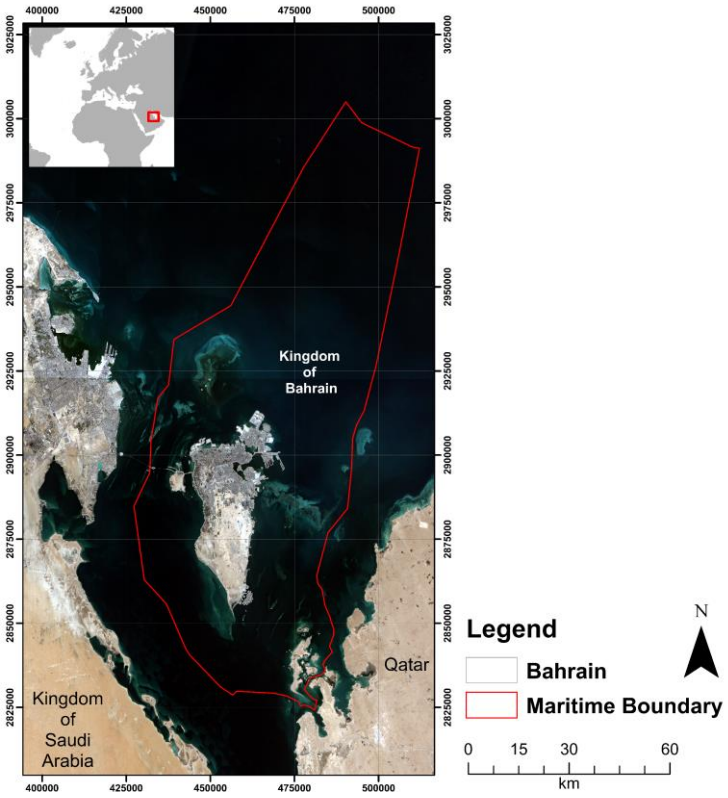

Fig. 1 Geographical overview of Bahrain's maritime borders and proximity to neighbouring countries globally.

coasts(Vousden 1995). These locations correspond with high concentrations of human impacts in the region, as most industrial and human activities are along these coastlines. The east coast is home to many industrial and sewage outfalls(Naser 2011), while the north and east coasts experience numerous reclamation activities and coastal development projects, such as Durrat Al Bahrain in the south and Amwaj Island in the northeast. The east coast is also associated with various sand dredging and oil exploration activities in offshore areas (Zainal et al. 2012). These activities serve as significant drivers of ecological impact and degradation.

In addition to the above considerations, the clarity and turbidity of the waters surrounding Bahrain are crucial factors in this study. Recent satellite-based research, such as Al Kaabi et al. (2013) work, has developed methodologies for water turbidity mapping in the Arabian Gulf, emphasizing the regional variability in water clarity. Furthermore, the use of MODIS data for mapping Secchi disk depth, as demonstrated by Al Kaabi et al. (2016), provides insights into the clarity of water in the shallow parts of the Arabian Gulf. More recent advancements in this field are illustrated by Kabiri (2023), who utilized Sentinel-3/OLCI satellite data to retrieve and validate Secchi disk depth values in the Persian Gulf and the Gulf of Oman. This research underscores the importance of considering water turbidity in Bahrain's marine and coastal environments, which are influenced by various human activities and natural processes.

3 Methodology

3.1 Data Acquisition

This investigation employs Landsat TM, Landsat 7, and Landsat 8 satellite imagery to characterize marine habitats, meticulously targeting spectral bands of blue (0.45-0.52 μm), green (0.52-0.60 μm), red (0.63-0.69 μm), near-infrared (NIR) (0.77-0.90 μm), and the coastal aerosol band (0.43-0.45 μm) from Landsat 8. The consistent 30m spatial resolution of the Landsat series offers an integrated view across four decades.

Data selection was guided by synchronization with previous landmark surveys conducted in February 1986, August 2005, and June 2015 (Vousden 1995; Geomatec 2006; Aljenaid et al. 2017). These surveys provided invaluable ground truth points and site descriptions, anchoring this work in a rich historical context.

Utilizing satellite imagery from USGS Earth Explorer, acquisition dates were aligned to mirror the survey timelines, creating a seamless temporal tapestry spanning significant ecological milestones. This harmonized approach facilitated the generation of updated classification maps of marine habitats that resonate with established chronicles.

Data was processed with precision using ENVI 5.6.1/IDL 8.8.1 and ArcGIS. Analytical rigour was ensured, matching the finesse of the data selection. The methodology, concise yet robust, is graphically elucidated in Figure 2.

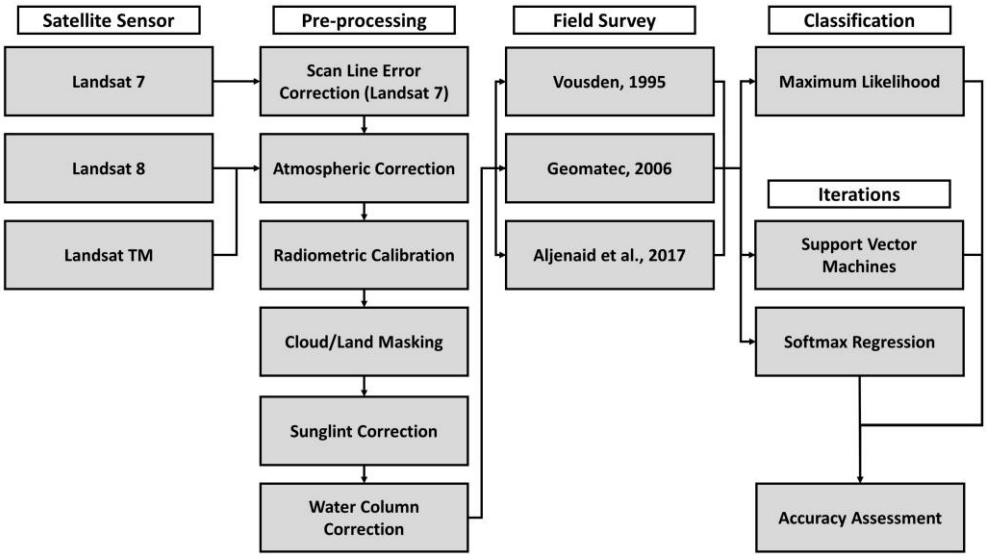

Fig. 2 Schematic representation of the methodological steps followed in this study.

3.2 Pre-processing Multispectral Imagery

3.2.1 Radiometric and Atmospheric Correction

Radiometric correction is essential for minimizing noise-induced inaccuracies in multispectral satellite data, enhancing the subsequent processing and quantitative analysis of digital images (Hadjimitsis et al. 2010b). Similarly, atmospheric correction is pivotal in accurately retrieving biophysical parameters from aquatic environments (Chander et al. 2009). This study used and corrected Level 1 Landsat products to yield Level 2 data, characterized by full calibration and atmospheric correction.

For the calculation of Top of Atmosphere (TOA), spectral radiance in Landsat data, band-specific multiplicative (ML), and additive rescaling factors (AL) were used in conjunction with pixel values from a calibrated standard product (Qcal). The formula is given by:

$$\boldsymbol{L_\lambda = ML \times Q_{cal} + AL} \qquad \textbf{(1)}$$

Here, (L_λ) represents the spectral radiance at the sensor's aperture. ML is the slope of the conversion from radiance to Digital Number (DN), and AL corresponds to ($LMIN_\lambda$), the minimum spectral radiance scaled to ($Qcal_{\mathrm{Min}}$).

Various atmospheric correction methodologies have been employed in the literature, ranging from Dark Object Subtraction (DOS) (Wu et al. 2005) to more sophisticated models like Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) and ATmospheric CORrection (ATCOR) (Pu et al. 2014). Comparative studies have indicated that model-based techniques generally surpass image-based methods in

performance (Adler-Golden et al. 2002; Malthus et al. 2003; Hadjimitsis et al. 2010a; Ilori et al. 2019a; Zhang et al. 2019; Schroeder et al. 2022).

In light of the increasing availability of Level-2 data from sources like the US Geological Survey (USGS) and the European Space Agency's Sen2Cor processor (Vermote et al. 2016, 2018; Main-Knorn et al. 2017; Wulder et al. 2019), this study evaluated both Dark Pixel Subtraction and FLAASH for their effectiveness in atmospheric correction. Dark Pixel Subtraction depends on the presence of a pixel with zero reflectance, such as deep clear water, and its effectiveness varies based on the availability and quality of such targets (Wicaksono and Hafizt 2018). FLAASH, on the other hand, models the transfer of solar radiation in the atmosphere to correct for molecular and particulate scattering and absorption. Notably, recent studies have highlighted the precision of ACOLITE, especially for atmospheric correction of Landsat imagery in coastal and near-shore areas. While this study did not utilize ACOLITE, it is recognized as a valuable alternative method for atmospheric correction, offering specific advantages in processing imagery of aquatic environments (Maciel and Pedocchi 2022).

3.2.2 Neighbourhood Similar Pixel Interpolation (NSPI) method

The Scan Line Corrector (SLC), which offsets the Landsat 7 satellite's forward motion, malfunctioned on May 31, 2003, with the failure being irreversible. The sensor's line of sight follows the satellite ground track in a zigzag pattern without an operational SLC. The outcome is a replica of the captured region with increasing width at the scene's edge. The duplicated portions are eliminated while processing the Level-1 data, creating data gaps. Despite having 78 per cent of their original pixels once the duplicated sections are removed, these pictures nonetheless include some of the most precise geometric and radiometric data available from any civilian satellite in the world (USGS 2019).

This study incorporates Landsat 7 data post-scan line corrector failure, so a gap-filling method is employed. For estimating the values of un-scanned pixels in the scan-line corrector (SLC-off) ETM+ image (Landsat 7), the NSPI method is simple but effective. It assumes that the same-class neighbouring pixels surrounding the un-scanned pixel have comparable spectral characteristics and that these neighbouring pixels and the un-scanned pixels have similar temporal change patterns. The NSPI approach can restore the value of un-scanned pixels and perform well in heterogeneous areas. Another benefit of the NSPI approach is that it can work well even if the time interval between the fill and target images is relatively long or features noticeable spectral differences (Chen et al. 2011). The NSPI method includes several additional sub-procedures, such as selecting neighbouring similar pixels, calculating weights for similar pixels, and calculating the target pixel value given temporal differences (Chen et al. 2012). Figure 3 illustrates the scan line errors in the Landsat 7 and NSPI-corrected images. The NSPI method can improve the quality of remote sensing images and enhance the classification process by filling in missing data. However, potential biases and limitations may arise due to its reliance on neighbouring pixels with similar spectral characteristics, temporal differences, spectral similarity assumptions, and limitations in complex landscapes. To

mitigate biases, it is crucial to validate classification results with ground truth data or independent sources, ensuring that the NSPI method does not compromise the accuracy of the outcomes (Chen et al. 2012).

The two initial predictions for the spectro-spatial and spectro-temporal information required for the gap filling were calculated using equations 2 and 3, respectively:

$$\boldsymbol{L_1(x, y, t_2, b) = \sum_{j=1}^{N} W_i \times L(x_i, y_i, t_2, b)} \qquad \textbf{(2)}$$

$$\boldsymbol{L_2(x, y, t_2, b) = L(x, y, t_1, b) + \sum_{j=1}^{N} W_i \times (L(x_i, y_i, t_2, b) - L(x_i, y_i, t_1, b)} \qquad \textbf{(3)}$$

Where $L(x_i, y_i, t_1, b)$ represents the value of the (i^{th}) similar pixel in the band (b) at the date (t_1), and $L(x_j, y_j, t_2, b)$ is the same but for the date (t_2); $L_1(x, y, t_2, b)$ is the prediction of the target pixel based on the spectro-spatial information, and $L_2(x_i, y_i, t_1, b)$ is the prediction pixel based on the spectro-temporal information. The (N) is the number of similar pixels, and (W_i) is the weight of the (i^{th}) similar pixel. NSPI uses a threshold and an adaptive moving window to identify identical pixels. The weights (W_i) are determined by multiplying a spatial distance (D_i) by a spectral distance ($RMSD_i$), where (D_i) is the Euclidean distance between the (i^{th}) comparable pixel and the target pixel, and $RMSD_i$ can be calculated as follows:

$$\boldsymbol{RMSD_i = \sqrt{\frac{\sum_{b=1}^{n} (L(x_i, y_i, t_1, b) - L(x, y, t_1, b))^2}{n}}} \qquad \textbf{(4)}$$

Where (n) is the band count. The final forecast is then calculated using a weighted combination of the two initial guesses. The degree of landscape homogeneity and spectral change between the input and target images within the moving window determine the weights (T_1 and T_2) (Chen et al. 2011). Consequently, the following conclusion can be drawn:

$$\boldsymbol{L(x, y, t_2, b) = T_1 \times L_1(x, y, t_2, b) + T_2 \times L_2(x, y, t_2, b)} \qquad \textbf{(5)}$$

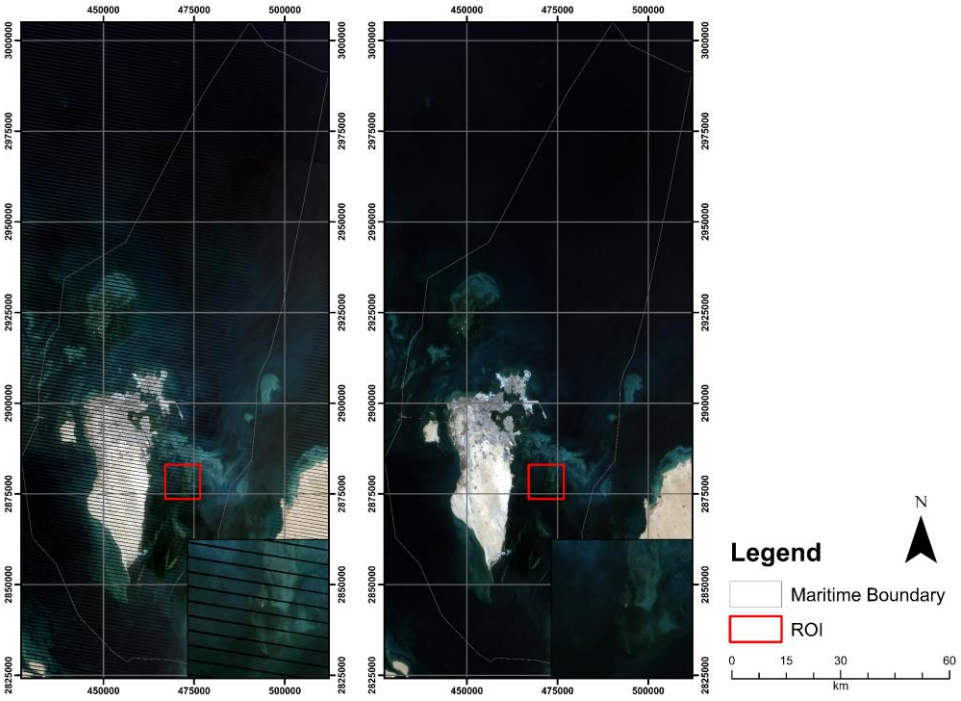

Fig. 3 Comparison of 2005 Landsat 7 images before (left) and after (right) applying the Neighbourhood Similar Pixel Interpolation (NSPI), focusing on regions of interest with scan lines. The images display bands 3, 2, and 1 for red, green, and blue, respectively.

3.2.3 Glint removal

Sunglint, characterized as sunlight reflected off the water surface towards the sensor, commonly affects satellite imagery of marine environments, particularly in shallow waters. The phenomenon introduces a radiation component that often saturates the pixel values and masks the water-leaving signal (Hochberg et al. 2003; Kay et al. 2009; Deidda and Sanna 2012; Streher et al. 2013). The presence of sunglint can compromise the accuracy of benthic habitat classification, necessitating correction methods tailored for marine substrate data (Aljenaid et al. 2017). The current study created a mask using near-infrared bands to isolate marine areas for more effective sunglint correction.

Various methodologies exist for sunglint correction, each aiming to estimate and remove the glint component from the sensor-received radiance. One approach, developed by Hochberg et al. (2003), operates under two assumptions. First, the near-infrared (NIR) brightness is comprised solely of sunglint and a spatially constant 'ambient' NIR component without varying benthic contributions. Second, the amount of sunglint in the visible bands is linearly related to the NIR brightness. This method eliminates a

consistent 'ambient' NIR brightness level from all pixels but is sensitive to outlier pixels, requiring careful land and cloud masking.

Hedley et al. (2005) offer an advancement to the Hochberg method, which focuses on individual pixel correction. This method assumes that the glint-free NIR radiance for a pixel is the same as the minimum value in the sample regions. A least-squares regression then adjusts the visible band based on this minimum NIR value. The equation used for this correction is:

$$L_i' = L_i - r(L_{NIR} - NIR_{Min}) \qquad \textbf{(6)}$$

Here, (L_i') represents the deglinted radiance, (L_i) is the original radiance, (r) is the regression slope, (L_{NIR}) is the NIR radiance, and ($\mathrm{NIR}_{\mathrm{Min}}$) is the minimum NIR radiance from the region of interest (Hedley et al. 2005). The effectiveness of this method is contingent upon the appropriate selection of a sample region, ideally situated in deep, dark locations exhibiting a noticeable glint effect (Green et al. 2000).

3.2.4 Water column correction

The penetration of light into aquatic environments is governed by an exponential attenuation with increasing depth, primarily due to scattering and absorption processes (Mumby et al. 2004). The principal absorbers in the marine environment—algae, suspended particles, dissolved organic compounds, and water itself—significantly absorb longer wavelengths while having minimal impact on blue light. Additionally, more turbid waters intensify light scattering due to a higher concentration of suspended particles.

Water column correction aims to derive a Depth Invariant Index (DII) from spectral band combinations rather than estimating seabed reflectance directly. This approach is most effective in clear waters and significantly enhances image interpretability when water characteristics are consistent across the scene. Variability of in-depth and Optically Active Constituents can blur distinctions between bottom types, necessitating water column correction to improve remote sensing reflectance (Zoffoli et al. 2014).

The adopted methodology for this study was the DII method developed by Lyzenga (1978), which eliminates the need for in situ water column data, a significant advantage, especially in under-studied reef areas (Mederos-Barrera et al. 2022). This algorithm is one of the earliest and remains widely used for this purpose. Alternatives such as those proposed by Tassan (1996) and Sagawa et al. (2010) cater to clear-water situations but require comprehensive bathymetric data.

The fundamental premise of the DII model is that light attenuation decreases exponentially with depth. A pair of spectral bands undergoes processing to produce a single depth-invariant band. Initially, the relationship between radiance and exponential attenuation is linearized, followed by calculating the coefficient attenuation ratio using a biplot of converted radiance in two bands (Seribu et al. 2021).

The DII equation is formulated as follows:

$$\text{Depth Invariant Index} = \ln(b_i) - \left[\left(\frac{k_i}{k_j}\right)\ln(b_j)\right] \quad \textbf{(7)}$$

Here, (b_i) *and* (b_j) are the radiances of bands (i) and (j), and (k_i/k_j) is the ratio of their attenuation coefficients. The attenuation reflectance of the water surface is calculated using the perpendicular offset method (Sordillo et al. 2014). The equations for determining the attenuation coefficients involve covariance and variance calculations, enhancing the method's precision.

3.3 Classification

3.3.1 Maximum likelihood classification

The maximum likelihood classifier features the same class labels as in previous studies to perform image segmentation in the focus years of 1985, 2005, and 2015. This supervised classification approach is trained using 80% of the ground-truth data from *in-situ* observations collected as part of the original Vousden (1995) study in February 1985 that surveyed 250 points. The Geomatec (2006) study, in August 2005, surveyed 735 sampling sites (294 subtidal and 441 intertidal. Lastly, the Aljenaid *et al.* (2017) study in June 2015 surveyed 176 field points. The training data is used to create signatures representative of the chosen classes while retaining the other 20% of data to validate the resulting classified maps. A stratified sampling strategy was used to guarantee that all observed ground-truth classes were proportionally represented in the training and validation. This was achieved by randomly sampling 20% of segments from each ground-truth code within each station without replacement. The population (the complete classification image or all ROIs) is divided into homogenous subgroups (the individual classes or ROIs) using this method, also known as proportionate or quota random sampling, before a simple random sample is taken in each subgroup. There are two variations of stratified random sampling: proportional and disproportionate. Disproportionate sampling enables you to specify each sample size in detail. Proportionate sampling was used where the sample sizes produced are closely correlated with the size of the classes (that is, the larger the class, the more samples will be drawn from it) (Cochran et al. 1954; Lynn 2016). The maximum likelihood classification was set up to classify each pixel, with none being rejected due to low confidence and all classes having an equal probability of being assigned. For the accuracy assessment of the maximum likelihood classification maps, contingency tables were produced using 20% of the point data withheld from the training set. The proposed outcomes are deemed to offer improved classification accuracy upon previously developed products by Vousden (1995), Geomatec (2006), and Aljenaid *et al.* (2017) by incorporating the previously stated methodologies.

3.3.2 Support Vector Machines Classification

Support Vector Machines, or SVM for short, is a classification method that is founded on the principle of statistical learning. It has been successfully employed in a variety of domains, including text categorisation, handwritten character recognition, image classification, bio sequence analysis, and others (Ma and Guo 2014). A decision surface that is optimised to maximise the margin between the classes is used by the SVM to differentiate between the classes. The hyperplane that lies closest to the surface is referred to as the support vectors, while the surface itself is sometimes referred to as the ideal hyperplane. The training set is missing some essential components without the support vectors. By utilising nonlinear kernels, the Support Vector Machine (SVM) can be modified to function as a nonlinear classifier (Cervantes et al. 2020). It is possible to use SVM as a multiclass classifier by combining many binary SVM classifiers, which will result in the creation of a binary classifier for each conceivable pair of classes. The pairwise classification approach is frequently used when multiclass classification has to be performed. The decision values of each pixel for each class are the result of the SVM classification method. These decision values are used for the estimation of probabilities. The probability values indicate "real" probability in the sense that each probability lies within the range of 0 to 1, and the sum of these values for each pixel equals 1. Additionally, each probability falls within the range of 0 to 1. After that, classification is accomplished by selecting the possibility with the highest likelihood. SVM has a penalty parameter that allows for a certain degree of misclassification. This is especially crucial for training sets that cannot be separated into their individual components. The penalty parameter determines the balance that should be struck between forbidding trainee mistakes and enforcing strict margins. It produces a fuzzy boundary that allows for some misclassifications, such as permitting some training points to be located on the incorrect side of the hyperplane. Increasing the value of the penalty parameter raises the cost of misclassifying points and compels the development of a more precise model, which may not be generalised as well as its less stringent counterpart. A linear kernel function, a polynomial kernel function, a radial basis function (RBF) kernel function, and a sigmoid kernel function are some of the various types of SVM classifier kernel functions (Hsu et al. 2003).

In addition to doing a maximum likelihood classification, a classification based on an SVM was carried out using Interactive Data Language (IDL) on a loop consisting of 100,000 iterations (due to machine capabilities and time constraints, while the ideal iterations would be much higher and different for each survey due to the different number of sampled points), where a different combination of 80:20 training/testing split of the ground truth data is implemented. The IDL code was used to normalise and extract samples from the DII raster using the ground truth points. After that, the instances that were retrieved are mixed up and divided into training and testing points in an 80:20 ratio. The classifier is then trained using the training points, and the DII raster is classified using the knowledge gained from the training. At last, the testing points will be utilised in the evaluation of the categorisation. This operation is carried out a hundred times by the code, and at the end of each cycle, a confusion matrix is

generated. This code was utilised to evaluate each of the DII rasters that were produced in 1985, 2005, and 2015. The DII band combinations that were selected were based on the correlation between the bands, where four band combinations were selected for all the years 1985, 2005 and 2015. The DII bands for 1985 Landsat 5TM, 2005 and Landsat 7 were all the same BG/BR (Blue-Green/Blue-Red), BG/GR (Blue-Green/Green-Red), BR/GR (Blue-Red/Green-Red) and BG/BR/GR (Blue-Green/Blue-Red/Green-Red), while the DII bands for 2015 Landsat 8 were CB/BG (Coastal-Blue/Blue-Green), CB/GR (Coastal-Blue/Green-Red), BG/GR(Blue-Green/Green-Red) and CB/BG/GR (Coastal-Blue/Blue-Green/Green-Red).

3.3.3 Softmax Regression Classification

Softmax Regression Classifier (SMR) is a form of multinomial logistic regression. It is capable of predicting the probability of the classes based on the input characteristics after weighting those features according to the relative importance they hold in the prediction (Wolfe *et al.,* 2017). Regression using Softmax is most useful when applied to classifications with several categories, particularly when the categories cannot be combined and there are more than two distinct outcomes that could occur (Greene 2012). An illustration of this would be separating the elements of an image into various categories like cloud, water, asphalt, and vegetation. Pal and Foody (2012) found that the classification results of multinomial logistic regression were just as accurate as those of SVM but required fewer data to train on.

Multiple classifications were performed on a loop for 100,000 iterations using IDL. The code written in IDL was used to normalise and extract examples using the ground truth points from the DII raster. Examples extracted are shuffled and split into 80:20 training and testing points. Training points are then used to train the classifier and perform the classification on the DII raster. Lastly, the testing points are used to evaluate the classification. The code repeats this process a hundred times and produces a confusion matrix for each cycle. This code was used to evaluate all the DII images created in 1985, 2005 and 2015. Similar to the SVM examination, only four band combinations were selected based on the highest correlation demonstrated between the bands, and they are the same for both classifiers.

4 Results

4.1 Overview

The fundamental objective of this study was to critically assess the outputs from previous classifications by comparing them with existing maps improved with current technological and software advancements. These enhanced maps retain the original ground truth data but are interpreted with modern tools to allow for a direct and meaningful comparison with the initial classification efforts.

The classification and subsequent mapping of the benthic habitat were executed using ML-supervised classification across four sets of input data spanning the critical periods

under investigation. Notably, the original habitat classification schemes varied across different surveys and have been harmonised through this ML approach. In updating the original classifications, the survey data from initial studies were employed as ground truth, thereby ensuring the consistent production of class numbers.

A comprehensive analysis of the derived classification maps of the benthic habitats around Bahrain revealed a somewhat disappointing performance. The overall accuracies were determined at 44.87% for the reclassified survey by Vousden (1995), 48.28% for areas deeper than 3m, and 41.18% for areas shallower than 3m for the survey conducted by Geomatec (2006). Furthermore, the survey by Aljenaid et al. (2017) yielded an accuracy of 61.05%.

For contextual understanding, the prior survey by Vousden (1995) reported an accuracy of 86.80% using ML classification. This result was achieved without any discernible pre-processing of the satellite imagery utilised in the study. Conversely, the Geomatec product underwent noise correction, resulting in accuracies of 81.00% for shallow areas (0-3m) and 75.00% for areas deeper than 3m after atmospheric correction and the implementation of the Lyzenga (1978) method for water column correction (restricted to the >3m part).

Interestingly, neither of these studies included sunglint correction, and the Geomatec study encountered additional complexities, with the scene bifurcated into two depth-dependent parts. The classification approach also varied, employing ML for shallow water and minimum distance for deeper sections. Aljenaid et al. (2017) achieved an accuracy of 84.10% by incorporating atmospheric and sunglint corrections and applying an unsupervised k-means classification, albeit without water column correction.

The meticulous examination of these previous surveys infers their reproducibility, contingent on the accuracy and precision of the reported data and methodologies. Such a comparative analysis sets the foundation for future research, affording opportunities to standardise practises and further comprehend inaccuracies. It opens new avenues for building upon and augmenting existing knowledge, thereby fostering an enriched understanding of the underlying phenomena and enhancing the potential for more accurate benthic habitat classifications in subsequent studies.

4.2 Vousden (1995) Classification and Comparison

4.2.1 Limitations and Biases

The classification of benthic habitats in Bahrain, as presented by Vousden (1995), utilised an image captured on February 16, 1985. Taken during the winter, this image was processed alongside a summer field study, necessitating an additional field investigation in the spring to adjust for seasonal variations.

In creating the classification, a false-colour image of the region was produced for habitat characterisation, using Landsat 5 TM bands 1, 2, and 3 and applying maximum likelihood classification. Any background noise that could be attributed to factors such as wave

action, atmospheric disturbances, or minor target shifts was eliminated through a specialised filtering technique. The methodology also considered the maximum effective penetration of the TM bands, ranging from 10 to 25 meters in clear water, and limitations imposed by high sediment levels. As a result, a deepwater class was introduced to include all water depths greater than 12 meters, with potential variations depending on localised sediment values.

A total of 235 sites were analysed by Vousden (1995), with 56 subsequently rejected, resulting in 179 usable sites and an accuracy of 87%. Of the discarded sites, 60% were excluded due to proximity to land, leading to interference, and the remaining 40% were disqualified for unreliable positioning.

However, several limitations and biases in Vousden's methodology warrant a critical analysis. Firstly, the selection of field sites was based on biological significance rather than an equitable distribution across the study area (Mentges et al. 2021; Zhang et al. 2021). This decision led to underrepresentation in areas of lesser biological significance, and consequently, the biased selection could impact the overall assessment of habitat distribution. It might also hinder accurately depicting the entire study area since regions of lesser significance received less attention during accuracy testing. A more balanced set of sampling points, encompassing areas of varying biological significance, would be required to comprehensively understand the classification's accuracy.

Secondly, the discrepancy between the seasons of image capture and ground study could introduce biases affecting habitat interpretation (de Souza et al. 2021; Rhif et al. 2022). Thirdly, the technological constraints, such as the specific use of certain bands and the absence of corrections for high sediment levels, might have contributed to an oversimplification or misclassification of habitats (Vahtmäe et al. 2020a; Mederos-Barrera et al. 2022).

Vousden's classification, while a significant step in understanding the benthic habitats surrounding Bahrain, bears marked limitations that must be recognised and carefully considered. Future studies might benefit from striving for balanced sampling and employing more refined techniques to arrive at nuanced and representative classifications. Such improvements could lead to more precise and comprehensive habitat maps, which would significantly enhance our understanding of the marine ecosystem in the region.

4.2.2 Classifications and Methodological Differences

The juxtaposition between the newly generated classification map and the original version by Vousden (1995) reveals insights that speak to continuity and divergence in the represented benthic habitat categories. The comparison, illustrated in Figure 4, highlights key classes such as Coarse sands, Coral-dominated, Deepwater gravels, Deepwater muds, Rocks with soft veneer, and Rocks/Sand with Coral/Algae/and Seagrass. These classes demonstrate notable congruence in spatial location and area coverage, with corresponding details in Table 1.

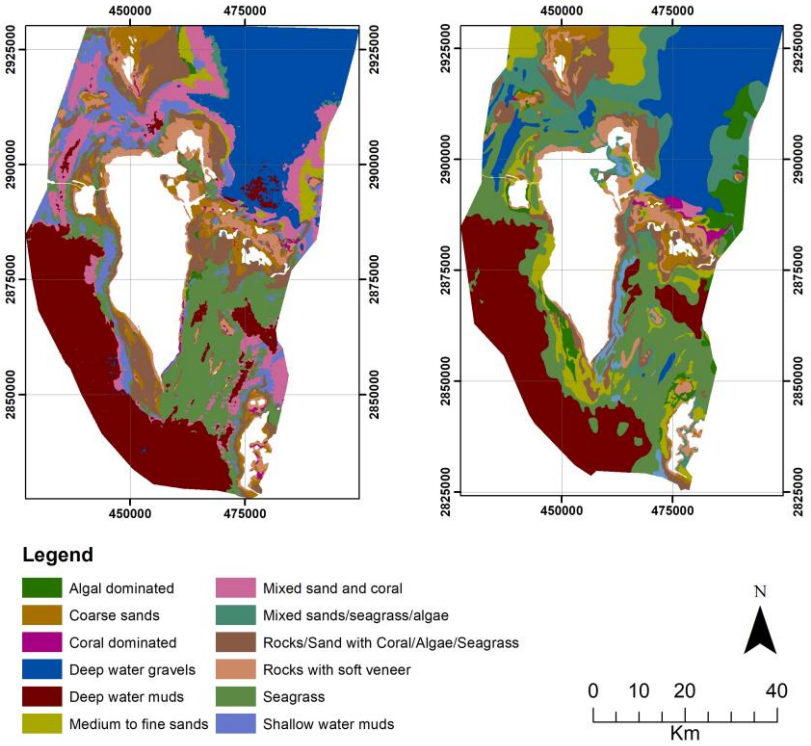

Fig. 4 Comparison of Vousden (1995) Maximum Likelihood classification and the newly generated classification using the proposed methodology, both in UTM Zone 39 coordinate system. The original classification by Vousden (1995) is displayed on the right with an accuracy of either 86.80% or 93.50%, while the new methodology yields a classification accuracy of 44.87%, presented on the left.

Table 1 Comparative Analysis of Class Areas in Vousden (1995) Classification Map.

Class	Area km²		Change	
	New	Old	km²	%
Algal Dominated	15	219	-203	-93
Coarse Sands	246	131	115	88
Coral Dominated	24	25	-1	-4
Deep Water Gravels	843	794	49	6
Deep Water Muds	973	940	33	4
Medium to Fine Sands	102	393	-290	-74
Mixed Sand and Coral	564	13	550	4231
Mixed Sands Seagrass Algae	23	507	-484	-95
Rocks with Soft Veneer	267	264	3	1
Rocks/Sand with Coral/Algae/Seagrass	563	412	152	37
Seagrass	722	956	-235	-25
Shallow Water Muds	368	77	291	378

However, the classification results underscore pronounced differences beneath these similarities, especially concerning the methodological approaches employed. Utilizing the Landsat 5TM imagery and applying a maximum likelihood classifier with a BG/GR band combination yielded an overall accuracy of 45%, a marked decline from the original 87% reported by Vousden (1995). This discrepancy between the two classifications raises questions about the underlying methodology.

One potential explanation for this divergence may lie in Vousden's (1995) omission of key pre-processing steps, such as atmospheric, sunglint, or column corrections. Such omissions can introduce significant biases and may contribute to the observed differences in classification accuracy (Jensen 2015).

Additionally, the original study's lack of transparency regarding selecting data points for accuracy calculation raises concerns. The report does not delineate which of the 250 points were employed to calculate the accuracy, thereby potentially indicating that the classification was trained and tested with the same data points. This approach, if confirmed, could undermine the validity and repeatability of the classification process, as the risk of overfitting is well-established in classification literature (Hastie et al. 2009).

The comparison between the two classification efforts highlights alignment and divergence, with the latter stemming from potential methodological inconsistencies and omissions. The observed discrepancies underscore the importance of transparent, rigorous, and repeatable methods, ensuring that advancements in classification techniques are grounded in scientifically robust practices. The findings of this comparison offer a valuable foundation for further refining benthic habitat mapping in the region, with the lessons learned potentially informing future methodological enhancements.

4.2.3 Classifier Performance and Implications

The in situ points were tested using SVM and SMR for 100,000 iterations, cycling through the points in an 80:20 split for training and testing. The outcome of the process for the 1985 classification is illustrated in Figure 5 for SVM and SMR. The results demonstrate a fairly normal distribution for all the band combinations and highlight the absence of high-accuracy classification outcomes based on the collected ground data. The SVM classifier yielded the most accurate results using the BR/GR (Blue-Red/Green-Red) depth invariant index (DII) band combination. This combination outperformed other combinations, achieving a maximum accuracy of 77.4%. Although all combinations exhibited relatively low mean accuracies and confidence levels (CL95%), the BR/GR combination showed the highest mean accuracy of 36.8% with a CL95% of 36.8 ± 8.

The SMR classifier reported lower values, with the highest mean accuracy of 22.5% for the BG/GR (Blue-Green/Green-Red) band combination and a CL95% of 22.5 ± 7. The

BG/GR combination attained the highest possible accuracy of 58%. The superior performance of the BR/GR combination can be attributed to its ability to account for varying water depths and water column attenuation while distinguishing between different benthic features. The combination leverages the properties of blue light (better water penetration) and red light (sensitivity to vegetation), allowing for improved discrimination of submerged habitats and substrates. However, it is essential to consider the relatively low mean accuracies, indicating that further enhancements in data pre-processing, feature selection, or classifier tuning may be necessary to achieve higher overall accuracy. The results from both classifiers indicate that the classification maps' accuracy relies on selecting the training and testing points. The data presented demonstrates that accuracy is inconsistent and that some points hold a higher weight in affecting the accuracy of the classification. This could mean that the initial classification map may have overestimated the accuracy, particularly if the data was not split into training and testing, which results in better accuracy.

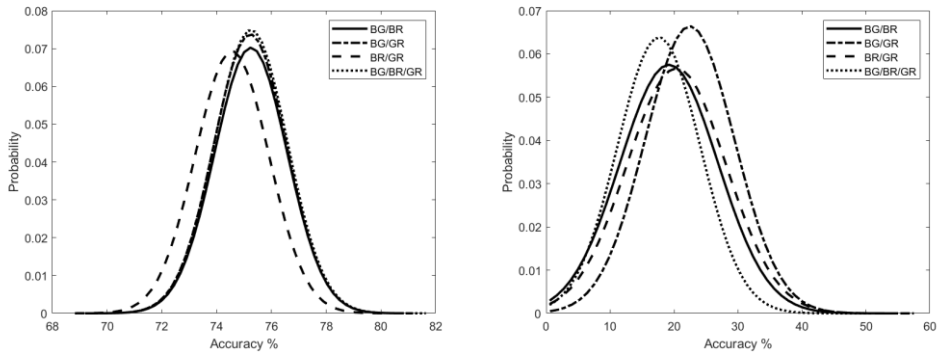

Fig. 5 Comparative analysis of accuracy probabilities for Vousden (1995) using 1985 survey points, employing SVM and SMR classifiers. Results are based on Landsat 5 TM data across various DII band combinations. SVM results are displayed on the left, and SMR results are on the right.

4.3 Geomatec (2006) Classification and Reanalysis

4.3.1 ASTER Limitations

Utilizing Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) images for mapping the coastal waters of Bahrain presents several challenges. The inconsistency in the acquisition times for each granule captured by the ASTER sensor may create misalignment between the survey and satellite image, leading to potential inaccuracies (Foody 2002). ASTER's 16-day revisit time further compounds this issue in dynamic coastal environments, limiting the detection of transient phenomena like algal blooms or sediment plumes (Abrams and Hook 2002).

The sensor's limited spectral bands constrain its ability to differentiate between various water constituents or submerged aquatic vegetation types, affecting the detailed

classification of complex ecosystems (Wang and Shi 2005). Additionally, ASTER's suboptimal signal-to-noise ratio (SNR) in optically complex waters can obstruct the detection of subtle water quality changes or specific features such as submerged reefs (Kutser et al. 2006).

These collective factors, including temporal alignment, revisit frequency, spectral capability, and SNR, can significantly impact the accuracy and reliability of the resulting maps. Future studies should consider integrating other data sources or methodological adjustments to enhance the robustness of coastal water monitoring in Bahrain (Abrams & Hook 2002).

4.3.2 Reanalysis of Geomatec (2006) Classification with Landsat 7

The reanalysis of the Geomatec (2006) classification, as detailed in Figure 6, posed several challenges, including methodological differences, data quality, and sensor selection. The original study employed ASTER data, requiring the mosaicking of multiple granules to cover the entire survey area in Bahrain's waters. This method introduced variances in data quality, condition, and temporal alignment, most notably leading to issues in image visibility due to the use of granules from different acquisition dates (Kutser et al. 2006; Ilori et al. 2019). In the reanalysis, Landsat 7 data, corrected for scan line errors, were used to achieve comprehensive coverage. However, the results were unsatisfactory, with overall accuracy rates dropping to 48.28% for areas deeper than 3m and 41.18% for shallower areas less than 3m (Wang and Shi 2005; Zhang et al. 2020). Table 2 compares class areas, showing significant discrepancies between the original and reanalyzed classifications. These differences suggest that the original Geomatec (2006) classification might have been based on additional, undisclosed data, complicating efforts for reproducibility and reliability (Wulder et al. 2008, 2019). Overall, this reanalysis underscores the challenges associated with using different sensors and methodologies, emphasizing the need for transparent reporting to improve the reliability and reproducibility of future mapping efforts.

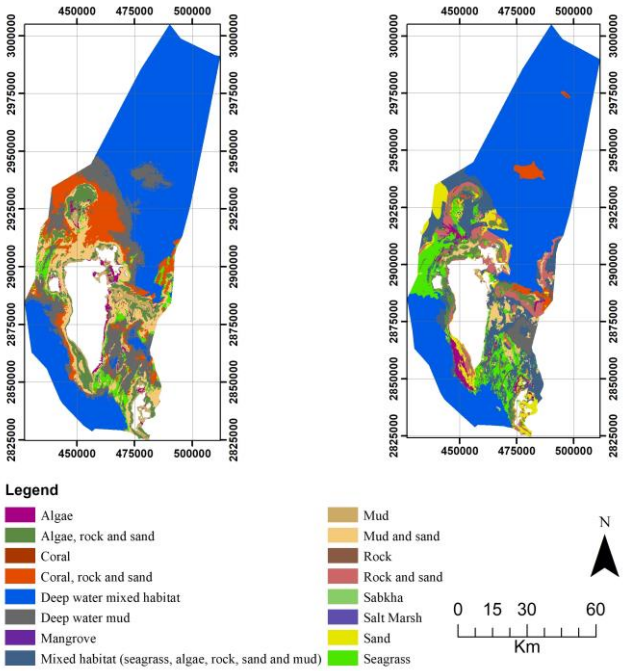

Fig. 6 Side-by-side comparison of Geomatec (2006) Maximum Likelihood classification with the newly generated classification based on the proposed methodology, both in UTM Zone 39 coordinate system. The original Geomatec classification is shown on the right, with an accuracy of 81.00% for shallow areas (0-3m) and 75.00% for areas deeper than 3m. The new classification, depicted on the left, yields an accuracy of 25.26%.

Table 2 Comparative analysis of class areas in the Geomatec (2006) classification map.

Class	Area km²		Change	
	New	Old	km²	%
Algae	164	133	-31	-23
Algae; Rock and sand	425	322	-103	-32
Coral	252	5	-247	-4931
Coral; rock and sand	590	134	-456	-340
Deepwater mixed habitat	3643	4332	689	16
Deepwater muds	1010	110	-900	-818
Mixed habitat (seagrass algae, rock sand and mud)	132	953	821	86
Mud	212	13	-199	-1529
Mud and sand	538	193	-345	-179
Rock and sand	110	466	356	76
Sand	86	327	241	74
Seagrass	321	526	205	39
Mangrove		0		
Rock		2		
Sabkha		6		
Salt Marsh		1		

4.3.3 SVM and SMR Classifier Performance

The SVM classifier performed on the 2005 Geomate (2006) data using Landsat 7 is presented in Figure 7 for deep and shallow areas. The results demonstrate a clear difference between the deep and shallow areas. The SVM conducted on the deep areas generated the highest accuracy of 80% using the BG/GR band combination. However, the mean accuracy is similarly low as the other band combinations. The highest mean accuracy using the BG/GR band combination was 25.9%, and it is close to the value of the other band combinations, while the CL95% was 25.9 ± 7.4.

On the other hand, the SVM classifier performed on the shallow regions did not have as high of accuracy as the deep regions, but the mean values are within the same range. The highest accuracy for the shallow regions was acquired using the same band combination BG/GR, and an equal value was generated using the BR/GR combination, where the accuracy value was 57.1%. However, the highest mean accuracy was generated using BG/GR, which resulted in a mean accuracy of 23.4%, and the CL95% was 23.4 ± 7.2. The deep regions SVM classifier might have resulted in higher accuracy, but the mean accuracy between the two regions is relatively close, with a 2.5% difference.

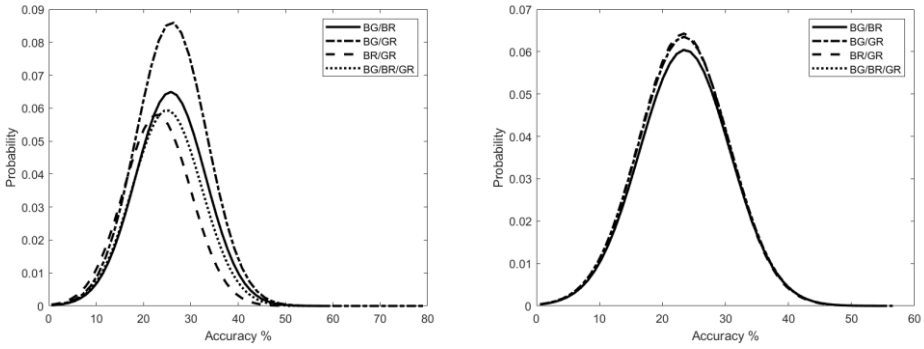

Fig. 7 Probability of accuracy for Geomatec (2006) using 2005 survey points and SVM classifier. Results are based on Landsat 7 data across various DII band combinations. Deep regions are shown on the left, and shallow regions are displayed on the right.

The SMR classifier was also performed and is presented in Figure 8 for the deep and shallow regions. The SMR classifier resulted in lower accuracy values than the SVM. For the deep regions, the highest accuracy was achieved using the BG/GR and the BG/BR/GR band combination, where both combinations resulted in a max accuracy of 60%. However, the highest mean was generated using the BG/BR/GR combination, resulting in a 20.9% accuracy with a CL95% of 20.9 ± 8.3. The SMR for the shallow regions had comparable values with a max accuracy of 60% using BG/BR and BG/GR, while the best mean accuracy was generated using BG/GR combination, resulting in a value of 20.6% and CL95% of 20.6 ± 8.3.

The best results for benthic mapping are obtained with the BG/GR band combination because it effectively reduces water column attenuation and normalises the impact of water depth. This combination makes distinguishing between various benthic habitats and features in deep and shallow areas possible. Blue and green wavelengths more effectively penetrate water, and the ratio of these bands lessens the influence of variations in depth on the spectral data. Compared to other band combinations, the BG/GR band combination produces better benthic classification accuracies, though further advancements in data pre-processing, feature selection, or classifier tuning may still be required for increased accuracy.

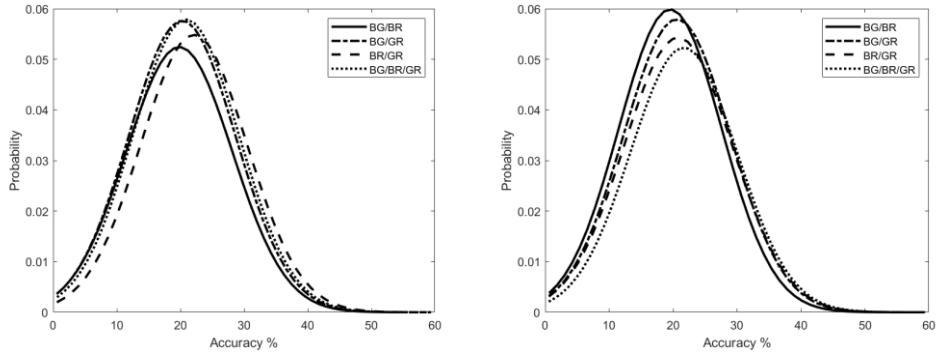

Fig. 8 Probability of accuracy for Geomatec (2006) using 2005 survey points and SMR classifier. Analyses are performed with Landsat 7 data across various DII band combinations. Deep regions are presented on the left, and shallow regions on the right.

4.4 Aljenaid et al. (2017) Classification

Aljenaid et al. (2017) conducted the third mapping case study, which gathered 176 field points to produce a map of Bahrain's benthic ecosystems. The original habitat classification identified 20 independent habitat classes using k-means classification, later consolidated into seven main classes. Despite the limitations of Landsat 8's spatial and spectral resolutions, the overall accuracy reached 84% in the study.

The reproduction of Aljenaid et al. (2017) classification using updated pre-processing methods, particularly with maximum likelihood classification (using CB/BG band combination), yielded promising results. It achieved an accuracy of 61.05%, showing overall similarities between the original survey and the recreation. The recreated classes exhibited comparable results, highlighting the best outcome among the three earlier studies.

4.4.1 Limitations and Reproduction of Aljenaid et al. (2017) Classification

Despite these successes, specific limitations were observed in reproducing the classification of Aljenaid et al. (2017). The recreated classes showed discrepancies concerning Deepwater, Deepwater sand, and sand classifications. Figure 9 and Table 3

illustrate these differences, notably in the northern parts of the region where the new classification failed to recognize rock and Deepwater sand. This failure aligns with similar issues observed in the original classification.

Unsupervised classification, such as the k-means method employed by Aljenaid et al. (2017), inherently presents challenges. Unsupervised classification algorithms do not utilize prior information or labelled examples to guide the classification process, leading to a lack of control in class determination (Umbaugh 2022). This can result in overly generalized or specific classes and inconsistencies in class assignments across different runs of the algorithm (Xu and Wunsch 2005; Xu and Tian 2015). In complex environments like coastal ecosystems, where features can be subtle and highly variable, the unsupervised approach may struggle to identify meaningful classes corresponding to real-world phenomena(Ahmad and Quegan 2013; Ahmad et al. 2018).

The noted limitations in spatial and spectral resolutions and discrepancies in some recreated classes underline the need for careful consideration when adapting this approach to different environments or implementing updated methodologies. While the Aljenaid et al. (2017) unsupervised classification demonstrated significant success in consolidating 20 classes into 7, these specific challenges, as backed by literature, may inform further refinements or necessitate supplementary data to enhance result reliability in future applications.

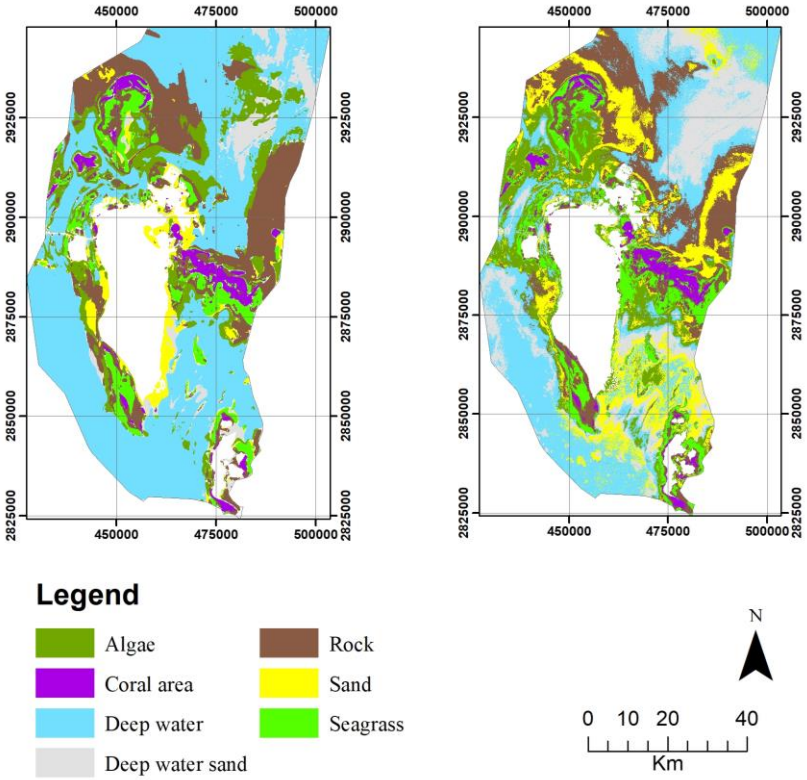

Fig. 9 Comparison between Aljenaid et al. (2017) Maximum Likelihood classification (right) and the new classification using the proposed methodology (left), both in UTM Zone 39 coordinate system. The original Aljenaid et al. classification demonstrates an accuracy of 84.10%, while the new classification shows an accuracy of 61.05%.

Table 3 Comparative analysis of class areas in the Aljenaid et al. (2017) classification map.

Class	Area km²		Change	
	New	Old	km²	%
Algae	758	661	97	15
Coral area	150	161	-11	-7
Deepwater	2608	1535	1073	70
Deepwater sand	147	856	-709	-83
Rock	1243	1071	172	16
Sand	294	859	-565	-66
Seagrass	380	449	-69	-15

4.4.2 SVM and SMR Classifier Results

The SVM and SMR classifier results are presented in Figure 10. The SVM classifier resulted in high accuracies for all the band combinations, but the highest accuracies were acquired using the CB/BG and CB/BG/GR band combinations with a value of 80%.

Regardless of the high accuracies generated, the mean accuracies for all the combinations are far lower, with the best mean accuracy being 47.6 and CL95% 47.6 ± 8.2 using the CB/BG/GR band combination. Unlike the SVM classifier, the SMR classifier resulted in low accuracy values. The best results were generated using the CB/BG/GR band combination, where the maximum accuracy was 68.6% and a mean accuracy of 30%, while the CL95% was 30 ± 7.4. The combinations exploit blue light's water penetration and red light's vegetation sensitivity, along with green light, to improve submerged habitat discrimination. However, relatively low mean accuracies suggest a need for further refinements in data pre-processing, feature selection, or classifier tuning to enhance overall benthic mapping accuracy.

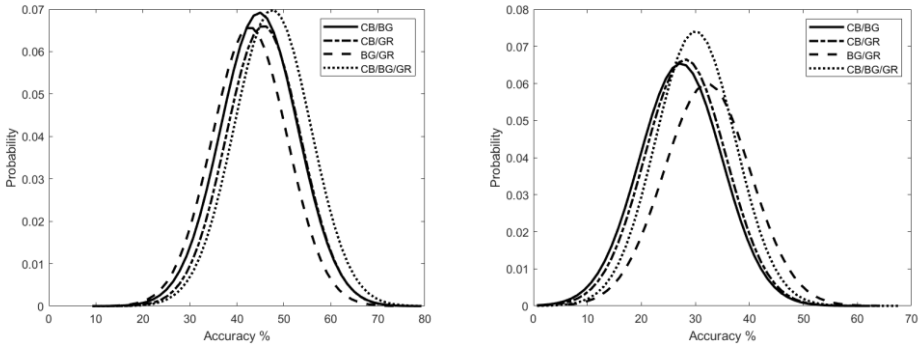

Fig. 10 Probability of accuracy for Aljenaid et al. (2017) based on 2015 survey points using SVM and SMR classifiers. The analysis utilizes Landsat 8 data across different DII band combinations. Results from the SVM classifier are shown on the left, and those from the SMR classifier are displayed on the right.

5 Discussion and Summary

The objective of this study was to scrutinize the accuracy and reliability of various remote sensing classification techniques employed for mapping coastal marine habitats in Bahrain. The focus was laid on three seminal works, namely Vousden (1995), Geomatec (2006), and Aljenaid et al. (2017), with an effort to reanalyse their results through the lens of contemporary pre-processing methods and alternative classifiers.

In the original studies by Vousden (1995) and Geomatec (2006), Landsat sensors were used, and classification accuracy was 86.80% in 1985, 81% for shallow regions, and 75% for deep regions in 2005. Despite not applying pre-processing methods to the satellite imagery, these studies achieved higher accuracy compared to cases where enhanced pre-processing was employed in this current study. This unexpected outcome could be attributed to the absence of training and testing data splits in the original classifications, which may have led to overfitting and overestimation of the classification accuracies, as can be seen in the SVM and SMR figures, where considerable variability in the accuracies can be seen when selecting the ground truth points. Additionally, enhanced pre-

processing might not have effectively addressed inherent data quality issues, or other factors like feature selection and classifier tuning could have influenced the lower accuracies.

On the other hand, Aljenaid et al. (2017) utilised the Landsat 8 sensor and applied pre-processing methods such as atmospheric correction, Lyzenga's sun glint correction, and column correction, resulting in an accuracy of 84.1%. The expansion of pre-processing methods, in addition to the use of K-means classification, contributed to the closer reproduction in this study. However, similar to the previous surveys, the SVM and SMR figures illustrate considerable variability in the accuracy when selecting the ground truth points.

The composite analysis of these three studies elucidates the challenges and limitations of using remote sensing data for coastal water monitoring and mapping. The recurring themes across the studies underscore the need for meticulous selection of appropriate remote sensing data sources and classifiers, along with robust and accurate ground-truth data. Moreover, it opens up vistas for future research to address existing limitations through technological advancements, such as deploying new sensors with higher spatial and spectral resolutions.

Accurate habitat mapping holds immense value for various stakeholders in coastal zone management, conservation, biodiversity assessment, and climate change research. These maps are instrumental in shaping sustainable practices and conservation strategies and advancing our understanding of complex marine ecosystem dynamics.

While the study has its merits, it is not without limitations. A noticeable constraint was the limited availability of literature focusing on the marine ecosystems of Bahrain. This paucity of resources inevitably impacts the scope and depth of the comparative analysis, making it less comprehensive than desired. Another concern is the methodological questions surrounding the replicability of the methods used. Tailored to local conditions, the broader application of these methodologies remains an open question, requiring further scrutiny. Moreover, the study's historical classifications primarily rely on singular or restricted data sources, which introduces an element of limitation that should be considered when assessing the robustness of our findings. Given these limitations, pursuing more comprehensive, robust, and generalizable outcomes would benefit from an expanded literature base, additional scrutiny on the adaptability of methodologies, and diversification of data sources. Incorporating these elements can significantly improve the quality and scope of future work in this domain, thereby making a meaningful contribution to the conservation and management of marine habitats in Bahrain and beyond.

In sum, as remote sensing technologies continue to develop, the scientific community must remain committed to refining methodologies to achieve greater accuracy in habitat classification maps. Such endeavours will enhance our comprehension of marine habitats and contribute significantly to the sustainable stewardship of coastal

ecosystems and the broader preservation of biodiversity. With its critical examination and synthesis of past works, this study contributes a vital step in that direction.

Declarations

Funding and/or Conflicts of Interests/Competing Interests

The authors declare that they have no financial or personal relationships with individuals or organizations that could inappropriately influence this study (bias). There were no funding sources for this research, and the authors have no conflicts of interest to disclose.

References

Abrams M, Hook S (2002) ASTER User Handbook Version 2. Jet Propulsion 2003:

Ahmad A, Hashim UKM, Mohd O, et al (2018) Comparative analysis of support vector machine, maximum likelihood and neural network classification on multispectral remote sensing data. International Journal of Advanced Computer Science and Applications 9:. https://doi.org/10.14569/ijacsa.2018.090966

Ahmad A, Quegan S (2013) Comparative analysis of supervised and unsupervised classification on multispectral data. Applied Mathematical Sciences 7:. https://doi.org/10.12988/ams.2013.34214

Al Kaabi MR, Zhao J, Charron C, et al (2013) Developing satellite-based tool for water turbidity mapping in the Arabian Gulf: Abu Dhabi case study. In: OCEANS 2013 MTS/IEEE - San Diego: An Ocean in Common

Al Kaabi MR, Zhao J, Ghedira H (2016) MODIS-based mapping of Secchi disk depth using a qualitative algorithm in the shallow Arabian Gulf. Remote Sens (Basel) 8:. https://doi.org/10.3390/rs8050423

Ali TS (2022) Long-term assessment of water quality in Askar coast, east of Bahrain. Scientific J Research Rev 3:1–13

Aljenaid S, Ghoneim E, Abido M, et al (2017) Integrating remote sensing and field survey to map shallow water benthic habitat for the Kingdom of Bahrain. J Environ Sci Eng 6:176–200. https://doi.org/10.17265/2162-5263/2017.04.002

AlKuzai J, Sheppard CRR, Abdulqader EAA, AlKuzai SA, Loughland RA (2009) Subtidal habitats. In: Loughland RA, Zainal AJM (eds) Marine Atlas of Bahrain. Geomatic Bahrain Centre for Studies and Research, Bahrain, pp 113–69

Cervantes J, Garcia-Lamont F, Rodríguez-Mazahua L, Lopez A (2020) A comprehensive survey on support vector machine classification: Applications, challenges and trends. Neurocomputing 408:189–215. https://doi.org/10.1016/J.NEUCOM.2019.10.118

Chander G, Markham BL, Helder DL (2009) Summary of current radiometric calibration coefcients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. Remote Sens Environ 113:893–903. https:// doi.org/10.1016/j.rse.2009.01.007

Chen F, Zhao X, Ye H, Karakehayov Z (2012) Making use of the Landsat 7 SLC-off ETM+ image through different recovering approaches. Data Acquisition Applications

Chen J, Zhu X, Vogelmann JE, et al (2011) A simple and effective method for filling gaps in Landsat ETM+ SLC-off images. Remote Sens Environ 115:1053–1064. https://doi.org/10.1016/j.rse.2010.12.010

Cochran WG, Mosteller F, Tukey JW (1954) Principles of Sampling. J Am Stat Assoc 49:13–35. https://doi.org/10.2307/2281032

de Souza R, Buchhart C, Heil K, et al (2021) Effect of time of day and sky conditions on different vegetation indices calculated from active and passive sensors and images taken from uav. Remote Sens (Basel) 13:. https://doi.org/10.3390/rs13091691

Deidda M, Sanna G (2012) Pre-processing of high resolution satellite images for sea bottom classification. Eur J Remote Sens. https://doi.org/10.5721/ItJRS20124417

Foody GM (2002) Status of land cover classification accuracy assessment. Remote Sens Environ 80:185–201

Geomatec (2006) Marine environmental Geographic information system (MARGISII), fnal report. Manama, Bahrain, Geomatec, Bahrain Center for studies and research, Kingdom of Bahrain, p 72

Gholoum M, Bruce D, Alhazeem S (2019) A new image classification approach for mapping coral density in State of Kuwait using high spatial resolution satellite images. Int J Remote Sens 40:. https://doi.org/10.1080/01431161.2019.1574991

Green EP, Mumby PJ, Edwards AJ, Clark CD (2000). Remote sensing handbook for tropical coastal management. UNESCO, Paris, pp 36–89. https://lib.riskreductionafrica.org/bitstream/handle/12345 6789/258/3558.Remote%20Sensing.%20Handbook%20for%20Tro pical%20Coastal%20Management.pdf?sequence

Greene WH (2012) Econometric analysis, 7th Edition. Boston et al: Pearson International Edition

Grizzle RE, Ward KM, AlShihi RMS, Burt JA (2016) Current status of coral reefs in the United Arab Emirates: Distribution, extent, and community structure with implications for management. Mar Pollut Bull 105:. https://doi.org/10.1016/j.marpolbul.2015.10.005

Ha NT, Manley-Harris M, Pham TD, Hawes I (2020) A comparative assessment of ensemble-based machine learning and maximum likelihood methods for mapping seagrass using sentinel-2 imagery in Tauranga Harbor, New Zealand. Remote Sens (Basel) 12:355

Hastie T, Tibshirani R, Friedman J (2009) Springer series in statistics: The elements of statistical learning: Data mining, inference and prediction

Hedley JD, Harborne AR, Mumby PJ (2005) Technical note: Simple and robust removal of sun glint for mapping shallow-water benthos. Int J Remote Sens 26:2107–2112. https://doi.org/10.1080/ 01431160500034086

Hedley JD, Roelfsema CM, Chollett I, et al (2016) Remote sensing of coral reefs for monitoring and management: a review. Remote Sens (Basel) 8:118

Hochberg EJ, Andréfouët S, Tyler MR (2003) Sea surface correction of high spatial resolution Ikonos images to improve bottom mapping in near-shore environments. IEEE transactions on geoscience and remote sensing 41:1724–1729. https://doi.org/10.1109/TGRS.2003.815408

Hsu C-W, Chang C-C, Lin C-J (2003) A practical guide to support vector classifcation Technical Report. Department of Computer Science and Information Engineering, University of National Taiwan, Taipei, pp 1–12

Ilori CO, Pahlevan N, Knudby A (2019) Analyzing Performances of Different Atmospheric Correction Techniques for Landsat 8: Application for Coastal Remote Sensing. Remote Sensing 2019, Vol 11, Page 469 11:469. https://doi.org/10.3390/RS11040469

Information and eGovernment Authority (2020) Bahrain open data portal. In: Central Organization Statistics. https://www.iga.gov.bh/en/. Accessed 6 Mar 2023

Jensen JR (University of SC (2015) Introductory digital image processing : a remote sensing perspective (4th ed.)

Kabiri K (2023) Retrieval and validation of the Secchi disk depth values (Zsd) from the Sentinel-3/OLCI satellite data in the Persian Gulf and the Gulf of Oman. Environmental Science and Pollution Research 30:. https://doi.org/10.1007/s11356-023-27625-7

Kabiri K, Rezai H, Moradi M (2018) Mapping of the corals around Hendorabi Island (Persian Gulf), using WorldView-2 standard imagery coupled with field observations. Mar Pollut Bull 129:. https://doi.org/10.1016/j.marpolbul.2018.02.045

Kay S, Hedley JD, Lavender S (2009) Sun glint correction of high and low spatial resolution images of aquatic scenes: a review of methods for visible and near-infrared wavelengths. Remote Sens (Basel) 1:697–730. https://doi.org/10.3390/rs1040697

Khan NY, Munawar M, Price ARG (2002) The gulf ecosystem: Health and sustainability. Backhuys Leiden, The Netherlands

Kutser T, Metsamaa L, Strömbeck N, Vahtmäe E (2006) Monitoring cyanobacterial blooms by satellite remote sensing. Estuar Coast Shelf Sci 67:. https://doi.org/10.1016/j.ecss.2005.11.024

Lynn P (2016) Principles of Sampling. Research Methods for Postgraduates: Third Edition 244–254. https://doi.org/10.1002/9781118763025.CH24

Lyzenga DR (1978) Passive remote sensing techniques for mapping water depth and bottom features. Appl Opt 17:379. https://doi. org/10.1364/ao.17.000379

Ma Y, Guo G (2014) Support vector machines applications. Springer

Maciel FP, Pedocchi F (2022) Evaluation of ACOLITE atmospheric correction methods for Landsat-8 and Sentinel-2 in the Río de la Plata turbid coastal waters. Int J Remote Sens 43:. https://doi.org/10.1080/01431161.2021.2009149

Main-Knorn M, Pflug B, Louis J, et al (2017) Sen2Cor for Sentinel-2. PROCEEDINGS OF SPIE SPIEDigitalLibrary.org/conference-proceedings-of-spie. https://doi.org/10.1117/12.2278218

Mederos-Barrera A, Marcello J, Eugenio F, Hernández E (2022) Seagrass mapping using high resolution multispectral satellite imagery: A comparison of water column correction models. International Journal of Applied Earth Observation and Geoinformation 113:102990. https://doi.org/10.1016/J.JAG.2022.102990

Mentges A, Blowes SA, Hodapp D, et al (2021) Effects of site-selection bias on estimates of biodiversity change. Conservation Biology 35:. https://doi.org/10.1111/cobi.13610

Mumby PJ, Skirving W, Strong AE, et al (2004) Remote sensing of coral reefs and their physicalenvironment. Mar Pollut Bull 48:219–228. https://doi.org/10.1016/j.marpolbul.2003.10.031

Naser H (2011) Human impacts on marine biodiversity: macrobenthos in Bahrain, Arabian Gulf. INTECH Open Access Publisher

Naser HA (2016) Management of Marine Protected Zones–Case Study of Bahrain, Arabian Gulf. Applied studies of coastal and marine environments 323

Pal M, Foody GM (2012) Evaluation of SVM, RVM and SMLR for accurate image classification with limited ground data. IEEE J Sel Top Appl Earth Obs Remote Sens 5:1344–1355

Park SJ, Achmad AR, Syifa M, Lee C-W (2019) Machine learning application for coastal area change detection in gangwon province, South Korea using high-resolution satellite imagery. J Coast Res 90:228–235

Pu R, Landry S, Zhang J (2014) Evaluation of atmospheric correction methods in identifying urban tree species with WorldView-2 imagery. IEEE J Sel Top Appl Earth Obs Remote Sens 8:1886–1897

Purkis SJ, Riegl B (2005) Spatial and temporal dynamics of Arabian Gulf coral assemblages quantified from remote-sensing and in situ monitoring data. Mar Ecol Prog Ser 287

Rhif M, Abbes A Ben, Martinez B, et al (2022) Detection of trend and seasonal changes in non-stationary remote sensing data: Case study of Tunisia vegetation dynamics. Ecol Inform 69

Sagawa T, Boisnier E, Komatsu T, et al (2010) Using bottom surface refectance to map coastal marine areas: a new application method for Lyzenga's model. Int J Remote Sens 31:3051–3064

Seribu K, Dki JM, Hamidah RA, et al (2021) Accuracy assessment of relative and absolute water column correction methods for benthic habitat mapping in Parang Island. IOP Conf Ser Earth Environ Sci 686:012034. https://doi.org/10.1088/1755-1315/686/1/012034

Sordillo LA, Pu Y, Pratavieira S, et al (2014) Deep optical imaging of tissue using the second and third near-infrared spectral windows. J Biomed Opt 19:056004. https://doi.org/10.1117/1.jbo.19.5.056004

Stéphane P, Pascal L (2015) A model of the general circulation in the Persian Gulf and in the Strait of Hormuz: Intraseasonal to interannual variability. 94:55–70. https://doi.org/10.1016/j.csr.2014.12.008

Streher AS, Goodman JA, Galvão LS, et al (2013) Sunglint removal in high spatial resolution hyperspectral images under different viewing geometries. Anais XVI Simpósio Brasileiro de Sensoramiento Remoto-SBSR 7958–7965

Tassan S (1996) Modifed Lyzenga's method for macroalgae detection in water with nonuniform composition. Int J Remote Sens 17:1601–1607

Traganos D, Reinartz P (2018) Interannual change detection of mediterranean seagrasses using RapidEye image time series. Front Plant Sci 9:1–15. https://doi.org/10.3389/fpls.2018.00096

Umbaugh SE (2022) Digital image processing and analysis: digital image enhancement, restoration and compression, 4th edn. CRC Press. https://doi.org/10.1201/9781003221142

USGS (2019) Landsat 7 (L7) Data Users Handbook. https://www.usgs.gov/media/fles/landsat-7data-usershandbook

Vahtmäe E, Kutser T, Paavel B (2020a) Performance and applicability of water column correction models in optically complex coastal waters. Remote Sens (Basel) 12:. https://doi.org/10.3390/rs12111861

Vahtmäe E, Paavel B, Kutser T (2020b) How much benthic information can be retrieved with hyperspectral sensor from the optically complex coastal waters? J Appl Remote Sens 14:16504

Vermote E, Justice C, Claverie M, Franch B (2016) Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. Remote Sens Environ 185:46–56. https://doi.org/10.1016/J.RSE.2016.04.008

Vermote E, Roger JC, Franch B, Skakun S (2018) LASRC (Land Surface Reflectance Code): Overview, application and validation using MODIS, VIIRS, LANDSAT and Sentinel 2 data's. International Geoscience and Remote Sensing Symposium (IGARSS) 2018-July:8173–8176. https://doi.org/10.1109/IGARSS.2018.8517622

Vousden HP (1995) Bahrain Marine Habitats and Some Environmental Efects on Seagrass Beds: A Study of the Marine Habitats of Bahrain With Particular Reference to the Efects of Water Temperature, Depth and Salinity on Seagrass Biomass and Distribution. Ph.D. thesis, University of Wales (Bangor), Bangor, UK, 261. https://e.bangor.ac.uk/4257/. Accessed 20 Feb 2016

Wang M, Shi W (2005) Estimation of ocean contribution at the MODIS near-infrared wavelengths along the east coast of the U.S.: Two case studies. Geophys Res Lett 32:. https://doi.org/10.1029/2005GL022917

Watts MAE (2022) Benthic Habitat Mapping of Thuwal's Reefs Using High-Resolution Acoustic Technologies and Imaging Data [KAUST Research Repository]. https://doi.org/10.25781/ KAUST-61418

Wicaksono P, Hafizt M (2018) Dark target effectiveness for dark-object subtraction atmospheric correction method on mangrove above-ground carbon stock mapping. IET Image Process 12:582–587. https://doi.org/10.1049/iet-ipr.2017.0295

Wolfe J, Jin X, Bahr T, Holzer N (2017) Application of softmax regression and its validation for spectral-based land cover mapping. The international archives of photogrammetry. Remote Sens Spat Inf Sci 42:455

Wu J, Wang D, Bauer ME (2005) Image-based atmospheric correction of QuickBird imagery of Minnesota cropland. Remote Sens Environ 99:315–325

Wulder MA, Loveland TR, Roy DP, et al (2019) Current status of Landsat program, science, and applications. Remote Sens Environ 225:127–147. https://doi.org/10.1016/J.RSE.2019.02.015

Wulder MA, White JC, Goward SN, et al (2008) Landsat continuity: Issues and opportunities for land cover monitoring. Remote Sens Environ 112:955–969

Xu D, Tian Y (2015) A Comprehensive Survey of Clustering Algorithms. Annals of Data Science 2:. https://doi.org/10.1007/s40745-015-0040-1

Xu R, Wunsch D (2005) Survey of clustering algorithms. IEEE Trans Neural Netw 16

Zainal K, Al-Madany I, Al-Sayed H, et al (2012) The cumulative impacts of reclamation and dredging on the marine ecology and land-use in the Kingdom of Bahrain. Mar Pollut Bull 64:1452–1458. https://doi.org/10.1016/j.marpolbul.2012.04.004

Zhang H, Kang J, Xu X, Zhang L (2020) Accessing the temporal and spectral features in crop type mapping using multi-temporal Sentinel-2 imagery: A case study of Yi'an County, Heilongjiang province, China. Comput Electron Agric 176:. https://doi.org/10.1016/j.compag.2020.105618

Zhang W, Sheldon BC, Grenyer R, Gaston KJ (2021) Habitat change and biased sampling influence estimation of diversity trends. Current Biology 31:. https://doi.org/10.1016/j.cub.2021.05.066

Zofoli ML, Frouin R, Kampel M (2014) Water column correction for coral reef studies by remote sensing. Sensors 14(9):16881–16931