Comparison of Conventional Machine Learning and Convolutional Deep Learning models for Seagrass Mapping using Satellite Imagery

A. Mederos-Barrera, L. Albors, F. Marques, *Senior Member, IEEE,* J. Marcello, *Senior Member, IEEE,* G. Martinez, and F. Eugenio, *Senior Member, IEEE*

***Abstract*— Benthic communities, such as seagrass meadows, play a crucial environmental role in marine ecosystems and provide socio-economic benefits. Satellite remote sensing is currently used for their monitoring, and Deep Learning (DL) techniques offer improvements in mapping quality compared to traditional Machine Learning (ML). This study compares conventional ML and convolutional DL models for mapping *Cymodocea nodosa* meadows in El Río, Canary Islands, using WorldView-2 satellite imagery. An in-situ measurement campaign was conducted to generate an open dataset for segmentation. Evaluated models include Decision Trees, Gaussian Naïve Bayes, Support Vector Machines, K-Nearest Neighbors, Subspace KNN, Feedforward Neural Networks, U-Net, Attention U-Net, and Pix2Pix models. Results show that DL models significantly outperform conventional ML models in detecting *Cymodocea nodosa*. The best model (U-net) achieved an Intersection over Union (IoU) of 83% overall and 74% for *Cymodocea nodosa*, while the best ML model (FNN) only reached 62% and 23%, respectively. IoU was highlighted for its sensitivity to minor mapping changes. Additionally, a temporal analysis revealed a dramatic 96% reduction in *Cymodocea nodosa* coverage over 21 years, from 245.32 ha in 2001 to 9.31 ha in 2022. This study not only compares conventional ML and convolutional DL techniques for benthic habitat mapping but also provides a valuable methodology and dataset for future marine ecosystem monitoring research.**

***Index Terms*—Cymodocea nodosa, machine learning, deep learning, seagrass mapping.**

I. Introduction

THE ocean encompasses a complex network of physical, chemical, biological, and geological processes. Specifically, coastal marine zones are highly dynamic and intricate, influenced by these processes and significantly impacted by anthropogenic factors [1]. It is estimated that over 60% of the global population resides within 100 kilometers of the coast, making these areas not only ecologically but also economically and socially vital [2]. Coastal zones host a diverse array of habitats, including seagrass meadows, coral reefs, mangroves, and estuaries, which are home to numerous plant and animal species [3]. The sustainable management of these coastal areas is crucial to preserving their biodiversity and ensuring they continue to provide essential ecological services [4].

Benthic communities, including seagrass meadows, corals, and algae, play a crucial role in aquatic ecosystems. These habitats provide significant ecological and socio-economic benefits, such as coastal protection, carbon fixation, oxygen production, and serve as nurseries for commercially important species [5], [6]. However, these ecosystems are under threat from anthropogenic activities and the impact of climate change, making their sustainable management and monitoring essential [7], [8].

Given the alarming decline of seagrass meadows and their ecological significance, it is necessary to implement monitoring strategies that are not only accurate but also scalable and cost-effective. Traditional in-situ surveys, while highly precise, are limited by their high operational costs and logistical complexity, which makes them unsuitable for large-scale or frequent assessments. This limitation has driven the adoption of remote sensing technologies as a fundamental tool for environmental monitoring. Remote sensing enables the acquisition of consistent, repeatable, and spatially extensive data [9], which is critical for detecting temporal changes in benthic habitats. Furthermore, when combined with advanced computational techniques, remote sensing can provide detailed and timely information to support conservation and management decisions.

Specifically, remote sensing has proven to be useful for monitoring shoreline changes, suspended matter concentrations, chlorophyll levels, pollutant discharges, the state of coastal and marine habitats and changes in bathymetry [10], [11], [12]. Additionally, remote sensing is used to measure sea surface temperature, salinity, and surface roughness. Despite the availability of numerous approaches for measuring these variables in oceanic environments (Case I waters), coastal waters (Case II) present unique challenges due to their complexity [13]. Consequently, developing reliable methods to characterize these environments remains a significant research area.

This work was supported by the European Project MAC-CLIMA, MAC2/3.5b/254 (Corresponding author: A. Mederos-Barrera)

A. Mederos-Barrera, J. Marcello, and F. Eugenio are with the Institute of Oceanography and Global Change, University of Las Palmas de Gran Canaria (ULPGC), 35001, Las Palmas de Gran Canaria, Spain (e-mail: antonio.mederos@ulpgc.es, javier.marcello@ulpgc.es, francisco.eugenio@ulpgc.es)

L. Albors, and F. Marques are with the Department of Signal Theory and Communications, Technical University of Catalonia (UPC), 08034, Barcelona, Spain (e-mail: laia.albors@upc.edu, ferran.marques@upc.edu)

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Technological advancements in the early 2000s, such as the launch of sensors like Moderate Resolution Imaging Spectroradiometer (MODIS) and Medium Resolution Imaging Spectrometer (MERIS), enabled the study of coastal environments [14]. However, their spatial resolution was insufficient for monitoring areas close to the coastline. The launch of the WorldView-2 multispectral satellite in 2009 marked a significant milestone, offering less than 2-meter resolution in its eight multispectral bands [15]. This satellite introduced an unusually high number of channels, including a high-penetration blue band, enhancing the monitoring capabilities of coastal waters up to depths of 20-30 meters [16].

In this context, the classification of benthic habitats is highlighted, allowing the detection of changes in the populations of seagrass meadows and sediments of the seabed [10]. The benthic habitats mapping using remote sensing data involves several challenges, such as the presence of turbulent water, the maximum detectable depth, or the accuracy of model estimates, where the choice of classification algorithms is critical [17]. Regarding estimation methods, conventional Machine Learning (ML) techniques have been traditionally used for seabed classification, where traditional models such as Decision Tree (DT) [18] and Gaussian Naïve Bayes (GNB) [19], common models such as Support Vector Machine (SVM) [20] and K-Nearest Neighbors (KNN) [21], or more advanced models such as the use of Ensemble Learning [22] or the application of Feedforward Neural Network (FNN) [21] can be highlighted. Among the prominent models that have historically been used, the SVM stands out, since it obtains satisfactory results even if the quantity and quality of the training data is not high [23]. However, with the advancement of techniques, FNN may also be a promising candidate. Nevertheless, these algorithms often classify at the pixel level without exploiting the surrounding spatial information [24].

On the other hand, Deep Learning (DL) techniques have recently attracted attention for their ability to exploit the spectral and spatial components of remotely sensed imagery [25]. The applications of DL methods have grown significantly in recent times, with diverse applications ranging from agriculture [26], [27], food quality [28], or medical imaging [29], [30], highlighting the versatility of deep learning for segmentation tasks. It is worth mentioning the encoder-decoder architectures where, from an image, a semantic segmentation is obtained, such as U-Net model [31]. In addition to the classic U-Net, variants have been proposed to improve performance in complex environments. For example, the use of attention mechanisms [32], like in the Attention U-Net model [33] that introduces Attention Gates (AGs) in the skip connections, allowing the network to focus on the most relevant regions of the image while suppressing irrelevant or noisy features. This mechanism is particularly useful in heterogeneous benthic habitats, where class boundaries can be subtle.

Another relevant architecture is the conditional Generative Adversarial Network (cGAN) [34]. For example, the Pix2Pix model [35] is based on a U-Net architecture for segmentation, and the discriminator, known as PatchGAN, evaluates the realism of local patches rather than the entire image. This adversarial training strategy encourages the generation of more accurate and spatially consistent segmentation maps.

TABLE I
SURVEY ON SEAGRASS CLASSIFICATION AND SEGMENTATION USING REMOTE SENSING IMAGERY.

Authors	Platform	Models	Metrics	Maximum depth
Li and Xiao, 2011 [37]	Landsat	DT, NB, SVM	Accuracy (86.6, 86.8, 85.9%)	Shallow, average depth of 1.35 m
Pu et al., 2014 [38]	Landsat	MLC	Accuracy (79.99%), Kappa (0.745)	4 m
Topouzelis et al., 2016 [39]	Sentinel-2	MLC	Accuracy (92.3%)	10 m
Marcello et al., 2018 [40]	WorldView-2	MLC, SVM	Accuracy (86.70%, 91.66%)	20 m
Traganos and Reinartz, 2018 [41]	Sentinel-2	SVM	Accuracy (99.5%)	16.5 m
Bayyana et al., 2020 [42]	Sentinel-2	RF, SVM, KNN	Accuracy (99.0, 96.0, 96.0%)	20 m
Coffer et al., 2020 [43]	WorldView-2 and RapidEye	DCNN	Accuracy (97%)	12 m
Mederos-Barrera et al., 2022 [44]	WorldView-2 and 3	NB, SVM, KNN, S-KNN	Accuracy (89.8, 95.2, 95.4, 88.6%)	25 m
Scarpetta et al., 2022 [45]	WorldView-2 and 3	U-Net	IoU (0.87)	Shallow
Meister and Qu, 2024 [46]	Sentinel-2	NB, CART, SVM, RF	Accuracy (65.3, 82.5, 86.8, 87.4%)	Shallow, seagrass in depth less than 2 m
Wang et al., 2024 [47]	Landsat	SegNet, U-Net	Accuracy (95.1, 94.6%), Precision (83.8, 77.7%), F_1 Score (0.74, 0.69)	Shallow, average depth less than 1.5 m

* Decision Tree (DT), Naïve Bayes (NB), Support Vector Machine (SVM), Maximum Likelihood Classifier (MLC), Random Forest (RF), K-Nearest Neighbors (KNN), Subspace KNN (S-KNN), Classification and Regression Tree (CART), Deep Convolutional Neural Network (DCNN).

However, despite their potential, DL models also present challenges, e.g. they require elaborate datasets for training, the fitting of hyperparameters has a higher complexity, or that there is a high interdependency of the results to the architecture, the data and the hyperparameters used [36]. Therefore, a comprehensive DL analysis would serve a useful purpose.

In Table I, a review of the literature on seagrass mapping using satellite images is shown. The study reveals a trend towards using commercial platforms like WorldView-2/3 and open data from Sentinel-2 or Landsat satellites. Most studies focus on clear and shallow waters, employing classifiers such as DT, Naïve Bayes (NB), SVM, and KNN. In addition, the Maximum Likelihood Classifier (MLC) model is also employed, which, like GNB, is a probabilistic model based on Bayes' theorem. The difference between the two is that GNB assumes class independence, as well as a Gaussian probability

density function. Furthermore, Random Forest (RF) is shown [48], which is an application of Ensemble Learning techniques to DTs. On the other hand, recent works have begun exploring DL techniques, especially Convolutional Neuronal Networks (CNN) [49], where the presence of the U-Net model is highlighted. In addition, the widespread use of the Accuracy metric is noticeable, as is the recent use of the Intersection over Union (IoU). In this case, the need to compare different evaluation metrics is also apparent.

Despite the growing use of conventional ML and DL in remote sensing, several limitations have been identified in the literature, particularly in marine environments. Traditional ML models, such as SVM or KNN, often rely solely on spectral information and may struggle to capture spatial patterns critical for benthic habitat mapping. On the other hand, DL models like U-Net, while powerful, require large, annotated datasets, are computationally intensive, and their performance is highly sensitive to hyperparameter tuning and data quality [50]. Moreover, DL models are often criticized for their lack of interpretability, which can hinder ecological understanding and decision-making [51]. These challenges are further exacerbated in optically complex coastal waters (Case II), where water column effects and variable turbidity can distort spectral signatures. As highlighted by Maxwell et al. [52], the lack of standardized evaluation practices and the risk of overfitting in DL models call for more systematic and reproducible methodologies. Therefore, a critical and systematic evaluation of these techniques is essential to guide their effective application in marine ecosystem monitoring.

In light of these challenges, this work presents a comprehensive methodological study specifically adapted to the complexities of marine environments. A detailed comparison is conducted between multiples conventional Machine Learning models for pixel classification—such as GNB, DT, KNN, S-KNN, SVM and FNN—and Deep Learning models—such as U-Net, Attention U-Net, and Pix2Pix —, which are specifically designed for semantic segmentation. An incremental training strategy is proposed for the U-Net model, evaluating multiple loss functions, learning rates, scheduling techniques, and regularization methods. Additionally, a comparative assessment of evaluation metrics is performed to identify the most sensitive indicators for benthic habitat mapping. The main contributions of this work can be summarized as follows:

- An in-situ measurement campaign was carried out which, to the authors' knowledge, generated the first database for the segmentation of *Cymodocea nodosa* meadows with WorldView-2 images. (Information available at https://github.com/MederosBarrera-Antonio/CymoHub).
- A comprehensive comparative study of conventional Machine Learning models and Deep Learning model is conducted, including an incremental analysis to identify techniques that enhance benthic habitat mapping. To the best of our knowledge, no previous studies have applied Attention U-Net and Pix2Pix architectures for seagrass mapping using multispectral satellite imagery.
- Different evaluation metrics are analyzed to determine the most appropriate for benthic habitat mapping.

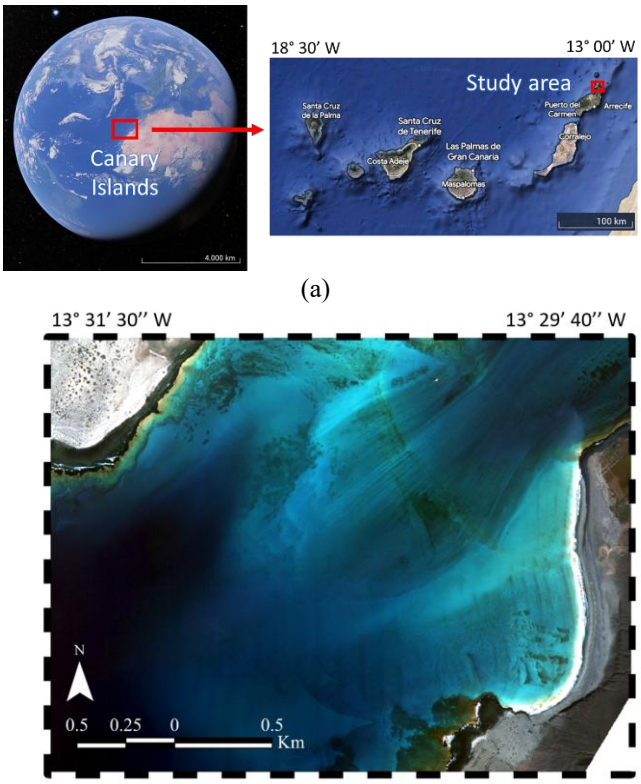


Fig. 1. Study area: (a) El Río strait between the islands of La Graciosa in the north and Lanzarote in the south, Google Earth©, (b) WorldView-2 image on 22 January 2022.

This work addresses the lack of systematic comparisons between conventional ML and DL models for seagrass mapping in optically complex waters, introducing a dataset and evaluating multiple performance metrics to guide future research.

II. Area of Study and Data

A. Area of Study

In particular, the specific study area is in the strait called El Río, which separates the islands of La Graciosa and Lanzarote, in the Canary Islands, shown in Fig. 1(b). In this region, the morphology of the site and the global marine currents generate a strong current in the strait, generally in a north-south direction. Fig. 1(b) shows lines of varying intensity created by suspended matter due to these currents, assuming non-clear Case II waters. It should be noted that the image presented is the result of extensive preprocessing, which is described in the methodology. In addition, another fundamental aspect is the depth range, which is challenging up to 15 m depth. In this case, the high complexity of the study area is observed.

The study area benefits from a high level of environmental protection, which has contributed to the long-term stability of its benthic habitats. Specifically, it is part of the Chinijo Archipelago Marine Reserve, the largest marine reserve in the European Union, with a surface of 700 km² approximately. The region is also designated under several conservation frameworks: UNESCO Biosphere Reserve (since 1993), Site of Community Importance (SCI) and Special Area of

Conservation (SAC) under the Natura 2000 network (since 2001), Natural Park of the Chinijo Archipelago (since 1986), and Special Protection Area for Birds (SPA) (since 1994). These overlapping protections have significantly limited direct anthropogenic pressures.

Biological objectives in this region include the monitoring of species and populations associated with the seabed. In particular, the historic underwater meadows of *Cymodocea nodosa* [53], [54], known locally in the Canary Islands as "sebadales", which are of vital ecological importance and productivity in the area. In particular, the declaration of the SCI and SAC protection of the area is due to the presence of this species. However, the meadows are in decline. Previous studies in the area concluded that there was an 81% reduction in populations in 10 years, between 2001 and 2010 [54]. Therefore, monitoring the *Cymodocea nodosa* population is of great importance. On the other hand, there are also other species of importance in the area, such as *Caulerpa prolifera* and filamentous algae [54]. Fig. 2 shows examples of the most representative species in the study area. The images show that the population density of the marine flora is low in the study area, which shows the complexity of its detection using remote sensing.

B. Satellite Data

In this study, a very-high-resolution multispectral image from the WorldView-2 (WV-2) satellite of 22 January 2022 has been used. Fig. 1(b) shows the natural color composite of the chosen image. WV-2 images are composed of 8 channels ranging from blue to infrared, from about 400 to 1040 nm. In this case, the spatial resolution is 2.0 m, the Off Nadir angle is 26.1°, and the radiometric resolution is 11-bits.

On this image, preprocessing was initially applied to obtain sea surface reflectance before using it for dataset generation, training and map estimation. The preprocessing techniques will be explained in more detail in the next section.

C. In-situ measurement campaign

In this work, supervised models are used, so that reference data for training and validation of the models are necessary to obtain seabed maps. However, nowadays there is no openly available datasets that can be used for the characterization of *Cymodocea nodosa* using satellite images. For this reason, a campaign of in-situ measurements was initially carried out to characterize the seabed in the study area. Subsequently, the database was generated by manually segmenting the image in the areas where measurements exist, thus, providing, as far as our knowledge, the first public database with these characteristics.

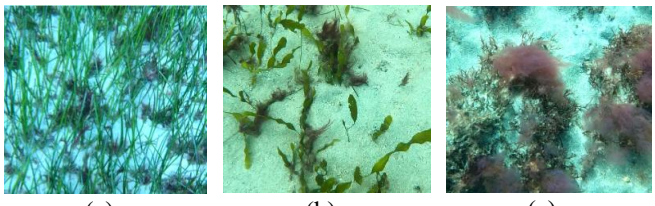
Regarding the in-situ campaign, it was carried out on 29 October 2023 and consisted of filming 79 videos in different locations. For this purpose, an inflatable boat, a Gopro Hero 9 camera, and a Garmin eTrex 10 Global Positioning System (GPS) were used. To capture the videos, the camera was submerged up to the seabed at each location. It should be noted that the GPS has the function of georeferencing the measurement points, as well as generating a track of the boat,

Fig. 3. Distribution of database images for seabed characterization.

which is useful for the analysis of the videos. Examples of seabed types recorded are in Fig. 2.

The videos captured were then analyzed, classifying the type of substrate and the flora community on the seabed. The analysis highlighted the joint presence of *Cymodocea nodosa* and *Caulerpa prolifera* on sandy seabeds, as well as red filamentous algae on all types of seabeds, both sandy and rocky. In addition, the presence of sandy and rocky bottoms without vegetation was also noted. On the other hand, the presence of *Dictyotales sp.*, *maërl* and *Lobophora sp.* was detected but to a minor extent than the previous seabed types.

Finally, it should be noted that the satellite image used in this study was acquired on January 22, 2022, and the in-situ measurement campaign was carried out on October 29, 2023. In this case, it is considered that the temporal difference does not significantly affect the validity of the training and validation of the models. In this case, the waters around the Canary Islands experience relatively mild temperature variations, due to the influence of the cold Canary Current [55]. This relative thermal stability may contribute to less pronounced seasonal changes in underwater vegetation compared to regions with more extreme temperature shifts. In addition, the study area has multiple figures of environmental protection, as mentioned above, which has contributed to a remarkable stability of the main benthic habitats over time. Although changes in the distribution of some species, such as *Cymodocea nodosa*, have been detected, these appear to be related to long-term environmental factors, such as climate change, rather than direct anthropogenic impacts. In



(a) (b) (c)

Fig. 2. Examples of the flora present in the study area: (a) *Cymodocea nodosa*, (b) *Caulerpa prolifera*, and (c) filamentous red algae.

addition, it should be noted that the in-situ data were not used directly as training labels, but as a reference to perform a manual segmentation of the satellite image, thus ensuring spatial consistency with the patterns observed in the image.

D. Database Generation

Using the information from the campaign, a manual image segmentation was performed in the proximity of the filming location. Afterwards, the manual segmentation was divided into 51 images to generate the database, as presented in Fig. 3. The size of each image is 70x70 pixels, or 140x140 m given the spatial resolution of the image (2 m). The choice of size was based on maximizing the surface area considering the spatial limitations of the camera recordings, as the camera was dropped to the seabed.

On the other hand, the final classes chosen for the study were: *Cymodocea nodosa*, which is mixed to a minor extent with *Caulerpa prolifera*; filamentous red algae; sand without vegetation; and rock without vegetation. It is worth mentioning that the land class has been added, and that the rest of the classes are not sufficiently extensive in population to be appreciated, as they consist of isolated individuals.

In addition, it can be seen in Fig. 3 that the images have been grouped into different zones, which correspond to the presence of different classes or seabeds: rock without vegetation, land and sand in zone 1; filamentous red algae and sand in zones 2, 3, 6, and 9; sandy seabed in zone 4; *Cymodocea nodosa* in zones 5 and 8; and rocky and sandy seabed in zone 7.

E. Training and validation distribution

A further fundamental aspect is the partitioning of the database for the generation of the training and validation datasets. In this case, a 70/30% distribution has been considered, respectively, to increase the amount of data in validation due to the limited number of images.

The prior partitioning of the database into different zones is useful for the division, since each zone corresponds mainly to a predominant class, allowing to avoid overfitting by considering a higher class representability between the training and validation datasets. For this purpose, in each zone the images were split approximately with 70/30% proportion. In addition, the bathymetric gradient was also considered for the splitting. This aspect is also fundamental as depth affects the observed radiance at different wavelengths. This behavior can be modelled by the diffuse attenuation coefficient in the radiative equations.

The images chosen for validation are 6, 10, 9, 7, 13, 17, 19, 25, 29, 34, 37, 42, 46, and 50, using the nomenclature in Fig. 3. On the other hand, the remaining images correspond to the training dataset. Therefore, finally there is an approximate distribution of 72.5/27.5% that considers the distribution of classes and depth.

F. Data augmentation

Finally, it is important to note that rotations have been applied as data augmentation. The rotations chosen were 90º, 180º and 270º. In addition, they have been applied to each training and test partition separately to avoid information leakage between sets.

III. METHODOLOGY

To ensure accuracy and reliability in mapping benthic habitats using passive remote sensing imagery, techniques must initially be applied to reduce noise on useful information. That is, phenomena that interfere with the observation of the very small amount of radiation reflected by the seafloor must be removed. Therefore, pre-processing techniques are applied to eliminate distortions caused by the satellite sensor, the atmosphere and the water surface [56].

Subsequently, seabed type maps can be obtained. In this case, two groups of models are studied. Firstly, Machine Learning models, which are based on pixel classification, where only the spectral information of each pixel is analyzed individually, without considering neighboring pixels. In this group, the GNB, DT, SVM, KNN and FNN models are compared. On the other hand, the Deep Learning U-Net, Attention U-Net, and Pix2Pix models are studied, based on image segmentation, in which, in addition to the information of each pixel, the neighboring spatial information is considered. In this case, three common experiments were carried out, consisting of the analysis of different loss functions and learning rates (LR), the study of the use of the Exponential scheduler for LR, and the application of L2 regularization [57]. In addition, an exclusive experiment was carried out for the Pix2Pix model, where the impact of a specific hyperparameter of its loss function was observed.

Finally, the maps obtained are analyzed. In this case, the study of five evaluation metrics should be highlighted. Fig. 4 presents a diagram of the aforementioned methodology.

A. Preprocessing

To ensure the accuracy and reliability in benthic habitat mapping using multispectral passive remote sensing images, different pre-processing steps are undertaken. As mentioned, the main objective is to reduce the noise generated by the optical medium in both downward and upward directions. For this purpose, atmospheric correction and sunglint correction is used. However, other pre-processing steps are also used, such as georeferencing, so that the geographic information is correct; masking of the study area and water, so that the presentation is more appropriate; and radiometric correction, to deal with effects inherent in the sensor capture.

Initially, georeferencing correction is performed to address spatial deviations in the original WV-2 image [58]. This involves using reference ground points and applying a first-order polynomial transformation. This pre-processing is crucial to be able to compare different maps.

Next, the study area and water masking are addressed, with the objective of only retaining water pixels in the study area [56]. This is useful for the presentation of the results. For water masking, the Normalized Difference Water Index (NDWI) was used [59].

Radiometric calibration follows, converting digital numbers from the satellite sensor into radiance values at the top of the atmosphere (ToA) using the linear model of analog-to-digital converters [60]. The values of the linear model are based on metadata information provided by the IMD metadata file. This step ensures that the images reflect physical values accurately

and avoid other banding phenomena due to different sensor properties.

An atmospheric correction is then applied to correct atmospheric effects such as absorption or scattering. This process converts ToA radiance values to reflectance values at the bottom of the atmosphere. The Second Simulation of a Satellite Signal in the Solar Spectrum (6S) model [61], [62] is used for this correction, employing a maritime aerosol profile, a mid-latitude summer atmospheric profile, a clear water reflectance profile, a surface altitude of 0 m, as this is a marine application, and an Atmospheric Optimal Thickness (AOT) value of 0.11 retrieved from the MODIS Aqua satellite through NASA's Giovanni service [63].

Subsequently, sunglint correction is performed to mitigate the effect of sunlight reflecting off waves, which causes brightness in the images. The Hedley et al. algorithm [64], which uses a linear regression between visible and near-infrared (NIR) bands, is applied. The two independent multispectral sensors (MS1 and MS2) of WV-2 satellite are utilized, with NIR1 correcting the blue, green, and red channels, and NIR2 correcting the coastal blue, yellow, and red edge channels. This avoids displacement artefacts in the image.

Finally, it is worth mentioning that correction of the water column has not been applied since, as previously studied [44], it does not improve seabed mapping results in turbulent waters and notable depths, as in this case study.

B. Mapping: conventional Machine Learning models

Conventional ML models, which are based on the classification and estimation of the spectral information of each pixel, are the first category to be studied. In this case, ML models contemplated can be grouped into four approaches. Initially, there are the (i) traditional models, which are the GNB, a probabilistic model based on Bayes' rule, and the DT, a model based on the construction of a tree-like graph. Next, the (ii) conventional models, which are based on the division of the feature space, such as the KNN and the SVM. Subsequently, the model used by (iii) Ensemble Learning, Subspace KNN, is used. Finally, the (iv) FNN model is considered, which is positioned as the initial model for Deep Learning models. This last model is based on obtaining an artificial neural network. Next, the different algorithms are briefly explained.

· *Gaussian Naïve Bayes*

Gaussian Naïve Bayes (GNB) [65] is a probabilistic classification model based on Bayes' theorem, which assumes conditional independence between the features. In this model, the feature vector $X = (x_1, \dots x_n)$ represents the spectral channels, and the class variable Y represents different seabed types. The conditional probability $P(y|X)$ is proportional to the product of the conditional probabilities of each feature given the class. In GNB, the probability density function of each feature is assumed to be Gaussian. Maximum A Posteriori (MAP) estimation [66] is used to estimate the prior probability $P(Y)$ and the likelihood $P(x_i|Y)$. The Gaussian assumption simplifies the computation of the likelihood, making GNB a computationally efficient model suitable for high-dimensional data.

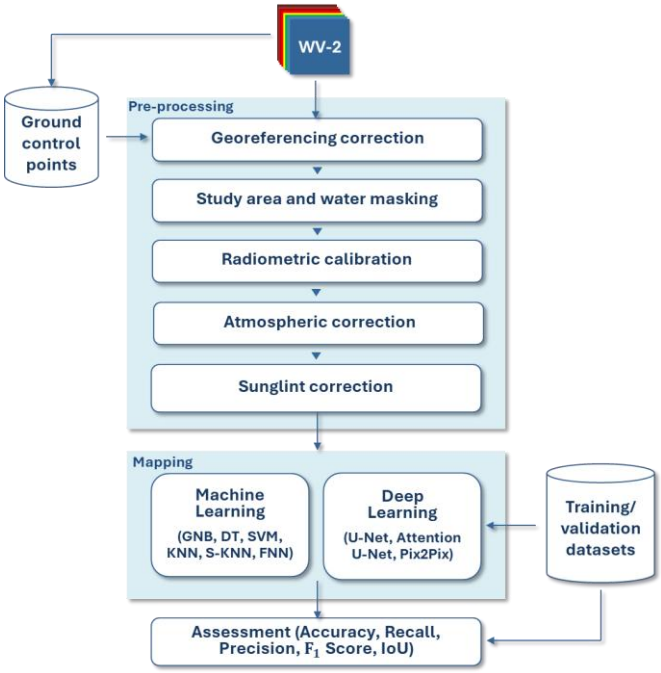


Fig. 4. Methodology for the generation of seabed type maps.

· *Decision Tree*

Decision Trees (DT) [67] are multivariate classification models that use conditional statements to categorize inputs into a finite set of outputs. The structure of a DT consists of an initial node (root), intermediate nodes (decision points), and final nodes (leaves) that represent the classification outcomes. Each intermediate node splits the data into subsets based on specific criteria, refining the classification at each level. The training process involves creating bifurcations to generate more homogeneous subsets, with the goal of increasing the homogeneity of each subset compared to the original dataset. The Gini Index (GI) [68] is commonly used to quantify the homogeneity of the splits. In this case, the maximum number of splits is set to 100.

· *Support Vector Machine*

Support Vector Machines (SVM) [69] are supervised learning models used for classification tasks. They aim to find the optimal hyperplane that separates different classes in the feature space. The SVM algorithm constructs a hyperplane or set of hyperplanes in a high-dimensional space, which can be used for classification, regression, or other tasks. For non-linear classification, SVMs use kernel functions to transform the feature space into a higher dimension where a linear separation is possible. In this case, the Gaussian Radial Basis Function (RBF) kernel is used [70]. This kernel function allows the SVM to create a non-linear decision boundary in the original feature space.

· *K-Nearest Neighbors*

K-Nearest Neighbors (KNN) [71] is a supervised ML classification method that assigns a point in the feature space to a group based on distance. In this method, new data is classified according to the group with the highest presence among the k nearest neighbors, where k is a hyperparameter. In this case, 10

neighbors are chosen. Another hyperparameter is the distance metric used to determine the k neighbors. In this work, the Euclidean distance is used. For a new data point, the Euclidean distance to all other points is calculated, and only the k nearest neighbors is considered. The new data point is then classified based on the majority class among these neighbors, a process known as majority voting.

· Subspace *KNN*

Ensemble Learning [72] techniques are used in Machine Learning to reduce data variance and improve system performance by combining multiple models. For a new input, each trained model provides a result, and the final output is determined by voting for classification among the models' outputs. Key hyperparameters include how the dataset is sampled and divided, and the number of subsets and models to train. One method for generating subsets is Random Subspace [73], which involves uniform sampling without replacement, ensuring that the intersections of generated subsets are empty, meaning no data points are repeated. In the context of K-Nearest Neighbors (KNN), Random Subspace can be applied to create diverse training subsets, enhancing the robustness of the KNN model. The resulting model is Subspace KNN [74]. By using different subsets of features, the KNN algorithm can better handle variations in the data and improve classification accuracy. This approach helps in managing the computational complexity and prevents overfitting by ensuring that each model in the ensemble learns from a unique subset of the data. In this case, after a preliminary assessment, 30 models are used for Ensemble Learning, as well as a Subspace dimension of 5.

· *Feedforward Neuronal Network*

Feedforward Neural Networks (FNNs) [75] are a type of Artificial Neural Network (ANN) where information flows in one direction, from input to output, without feedback loops. An FNN consists of an input layer, hidden layers, and an output layer, with each neuron in a layer connected to all neurons in the next layer. The input layer receives data, hidden layers perform computations using activation functions like Sigmoid, hyperbolic tangent (Tanh), or Rectified Linear Unit (ReLU), and the output layer produces the final classification. Training involves adjusting weights to minimize error using backpropagation. Key hyperparameters include the number of hidden layers, neurons per layer, and activation functions. Various configurations were analyzed, showing that these choices significantly impact accuracy and generalization. FNNs are effective for tasks like image classification but require careful tuning and computational resources. In this case, 78 different combinations of network sizes by modifying the number of hidden layers (1 to 3), and the number of neurons in each layer were analyzed, where the model with 3 hidden layers and 100, 50 and 50 neurons obtained the best results. In addition, the Sigmoid, TanH and ReLU activations functions were analyzed, and TanH obtained better results.

C. Mapping: Deep Learning models

For applications with large data dimensions, such as computer vision, the number of parameters in FNN models increases significantly. Thus, Convolutional Neural Networks (CNNs) [49] address this inconvenience by using convolution and pooling functions to reduce the number of parameters. Initially, CNNs used a FNN at the last stage to classify the entire image, which was ineffective for segmentation. Therefore, CNNs with encoder-decoder architecture were developed [76]. The encoder uses convolution to extract features, and the decoder uses transposed convolution to restore spatial dimensions, reducing the number of image channels and enabling effective image segmentation without an FNN. Additionally, other techniques were applied in DL models, such as attention mechanisms [32] that allow models to focus on more relevant parts of images. On the other hand, the use of generative adversarial neural networks has increased, especially conditional (cGAN) [34], in which a generator, such as a segmentation encoder-decoder network, faces a discriminator, where they mutually try to deceive each other by improving their performance in a symbiotic approach. In this case, the following models are implemented: U-Net [31], an encoder-decoder CNN; Attention U-Net [33], which applies Attention Gates (AGs) in the skip connections of U-Net; and Pix2Pix [35], where the generator is a basic U-Net and the discriminator is a PatchGAN.

· *U-Net model*

A CNNs with encoder-decoder architecture widely used for semantic segmentation is the U-Net. In addition to the above aspects, U-Net also has skip connections where the image generated in the convolution section is copied to the corresponding level of the same channel depth in the transposed convolution section. Skip connections allow the network to access features at different scales. The U-Net used in this work is shown in Fig. 5(a), where each block refers to an image, the vertical size to its dimension, and the number at the top corresponds to the number of channels in the image. In this case, Attention Gates are not used. Originally, the input image is based on an 8-channel image, corresponding to the WV-2 sensor. On the other hand, the output image contains 5 channels, corresponding to the 4 types of seabed types mentioned above, plus an extra one corresponding to the land. It should be noted that, in the segmentation, the image must be completely annotated. Therefore, the land pixels must be annotated in coastal areas, as is the case of Zone 1 (Fig. 3), even if they are zero due to the water mask preprocessing.

There are many hyperparameters inherent to the network and its training, such as the number of convolution and pooling stages, the number of convolution filters, the activation functions, the type of pooling, the loss functions, the optimizer, the number of training epochs, the initialization of the weights, or the learning rate values [57]. Additionally, there are techniques that can improve the training or estimation, such as the application of early stopping [77], the use of scheduler for the learning rate [78], or the application of regularization techniques [79]. The high complexity of choosing optimal hyperparameters, compared to conventional ML models, is highlighted. In this case, an incremental methodology is applied, based on three experiments where, at each step, new techniques are applied to the best models of the previous experiment.

This article has been accepted for publication in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. This is the author's version which has not been fully edited content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2025.3642923

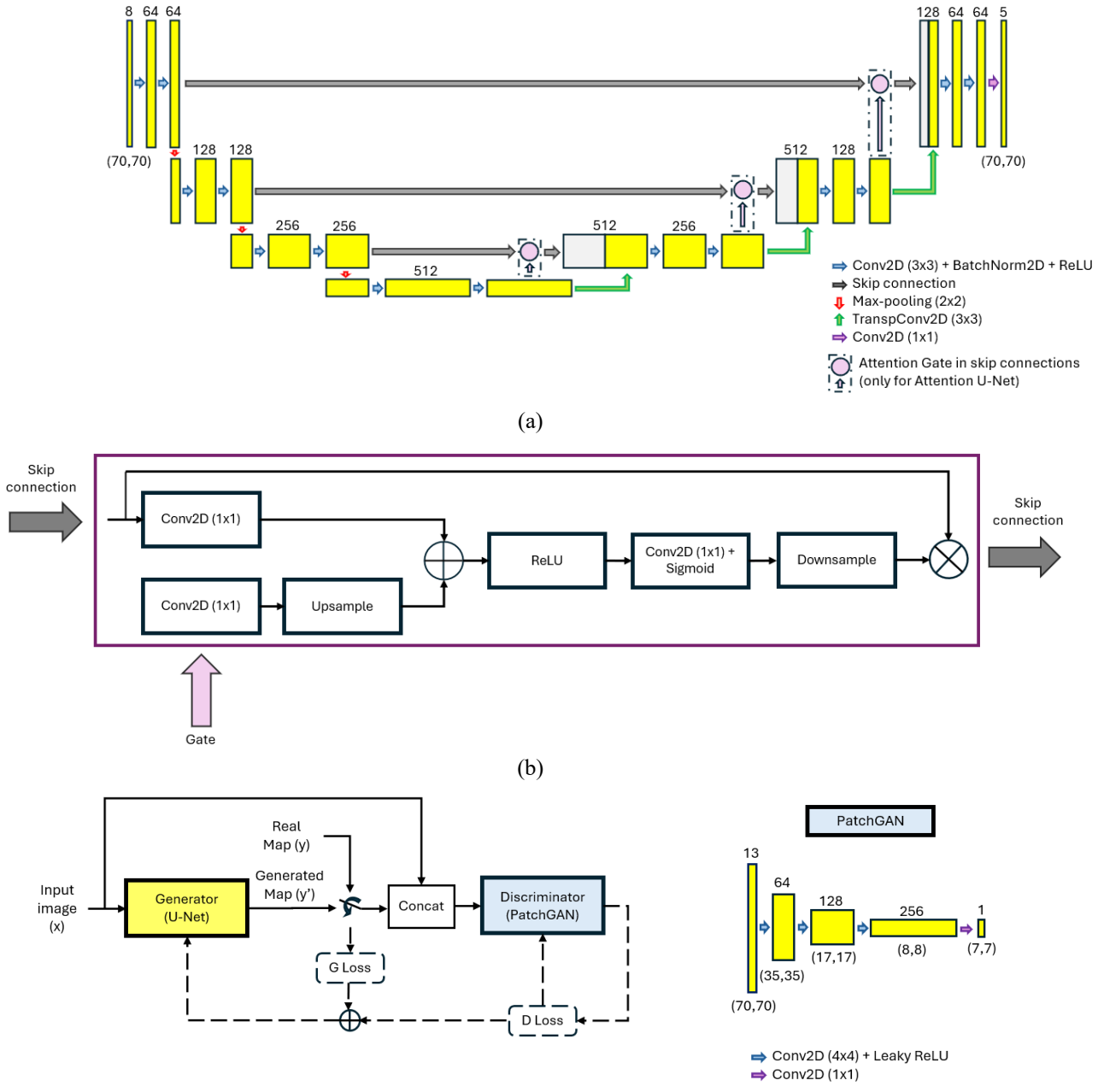


Fig. 5. Deep Learning architectures used for seagrass segmentation: (a) U-Net architecture with optional Attentions Gates (only used for Attention U-Net). (b) Attention Gate architecture. (c) Pix2Pix and PatchGAN discriminator architectures.

It is worth noting that there are common hyperparameters. Initially, the number of convolution and pooling stages has been reduced to 3 with respect to the original U-Net, due to the limitations of the spatial dimensions of the image, as well as to reduce the number of parameters to facilitate training. Remember that a limited database is available. Moreover, the number of convolution filters are presented in Fig. 5. In addition, the early stopping has been applied to avoid temporal divergence in training. In this case, the value of the loss function in the evaluation has been chosen as the metric, as well as 140 epochs of patience to stop the training. The hyperparameters chosen in general for all the experiments, as well as the specific hyperparameters in each one, are presented in Table II.

· *Attention U-Net model*

The Attention U-Net model is an extension of the classic U-Net that incorporates attention mechanisms in the skip connections between the encoder and decoder. Specifically, Attention Gates (AGs) are used, which allow the model to focus on relevant regions of the information transmitted in the skip connection, suppressing irrelevant or noisy activations. Instead of directly transferring the features from the encoder to the decoder, AGs filter important information based on the context of the regions, which improves segmentation in complex environments such as benthic habitats. This architecture is especially useful in images with high spatial variability. Fig. 5(b) shows the structure of the AGs used in this model. It should

be noted that AGs introduce parameters to the model, increasing its size.

· Pix2Pix model

Pix2Pix is a deep learning model based on conditional Generative Adversarial Networks (cGAN), designed for image translation tasks such as semantic segmentation. In this architecture, the generator consists of a U-Net model that produces segmentation maps from input images, while the discriminator, known as PatchGAN in this case, evaluates whether the received segmentations are real or generated by the generator. The relationship between generator and discriminator is shown in Fig. 5(c), as well as the architecture of the PatchGAN used. Unlike traditional discriminators that evaluate the entire image, PatchGAN evaluates regions or patches of the image, allowing it to discover small details that can be produced by the generator. Training is performed adversarially, i.e., the generator tries to trick the discriminator by producing realistic segmentations, while the discriminator improves its ability to distinguish between real and fake maps. In this case, the number of network parameters is also increased, as an additional CNN is introduced to U-Net. Furthermore, it is worth mentioning that, in this case, the network hyperparameters are also increased, not only by PatchGAN, but also by the adversarial loss function.

· Experiment 1: Loss functions and learning rates

The first step in the incremental experiments is the choice of the loss function with the best results. The loss functions studied for U-Net and Attention U-Net are Binary Cross-Entropy (BCE) [80], Cross-Entropy (CE) [81], Weighted Cross-Entropy (WCE) [81], Dice Loss (DL) [82], and Generalized Dice Loss (GDL) [82]. In the case of the Pix2Pix model, adversarial loss is based on BCE, and L1 reconstruction loss functions is used [35]. In addition, for each loss function, different learning rates are used during training, as shown in Table II. To ensure a comprehensive exploration of the learning rate hyperparameter, a wide range of values was tested, including both coarse (e.g., 0.1) and fine (e.g., 0.00001) adjustments. This strategy aimed to identify optimal configurations under different training dynamics and model sensitivities.

Initially, for the BCE case, two different aspects should be noted. First, it should be highlighted that the approach is one versus all, i.e., BCE is applied to each class, considered as 1, and the rest of the classes are considered as 0. Finally, the mean over the classes is obtained. On the other hand, in this case, the sigmoid function is applied on the logit outputs of the network. This allows several classes to have high probabilities. The expression used is given as:

$$BCE = -\frac{1}{C}\sum_{i=1}^{C}\sum_{j=1}^{N} y_{i,j}\log\left(\hat{y}_{i,j}\right) + \left(1-y_{i,j}\right)\log\left(1-\hat{y}_{i,j}\right), \quad (1)$$

where y refers to the reference, $\hat{y}$ to the estimation, C to the number of classes and N to the number of examples.

Subsequently, CE, which is widely used, is also applied, as well as WCE where the results of each class are weighted. This avoids overfitting due to class imbalance. In this case, the weights are used as the normalized inverse of the frequency of each class. In addition, the Softmax function is applied to the logits output of the network, so the sum of probabilities of all classes is 1. The WCE expression is given below. It should be noted that, to obtain CE, we just arrange all the weights of each class w as 1.

$$WCE = -\frac{1}{C}\sum_{i=1}^{C} w_i \sum_{j=1}^{N} y_{i,j}\log\left(\hat{y}_{i,j}\right). \quad (2)$$

On the other hand, DL and GDL loss functions are also studied, which, like CE and WCE, GDL is a generalization of DL considering class weights. In this case it is of special mention that the use of DL is not as widespread in the remote sensing literature as the rest of the loss functions. Moreover, it is based on the use of the Dice-Sørensen coefficient as a loss function. In this case, the Sofmax function is also applied on the logits outputs of the network. The GDL expression contemplates DL when the weights are all 1.

$$GDL = 1 - 2\frac{\sum_{i=1}^{C} w_i \sum_{j=1}^{N} y_{i,j}\,\hat{y}_{i,j}}{\sum_{i=1}^{C} w_i \sum_{j=1}^{N} y_{i,j} + \hat{y}_{i,j}}. \quad (3)$$

Finally, it should be noted that different loss functions are used for the generator G (L_G) and discriminator D (L_D) in the Pix2Pix model. In the case of the discriminator (PatchGAN), the mean binary cross-entropy (BCE) loss is used to distinguish between real and fake image pairs. On the other hand, the generator (U-Net) is trained using the sum of the adversarial BCE loss (L_{GAN}) and the L1 loss (L_{L1}). The L1 loss is used as a reconstruction loss, encouraging the generated image to resemble the target image. The L1 loss is weighted by a hyperparameter λ to control its impact during training.

$$L_D = \frac{1}{N}\sum_{i=1}^{N}\log\left(D(x_i,y_i)\right) + \log(1-D(x_i,G(x_i))), \quad (4)$$

$$L_{GAN} = -\frac{1}{N}\sum_{i=1}^{N}\log\left(D\left(x_i,G(x_i)\right)\right),$$
$$L_{L1} = \sum_{i=1}^{N}\sum_{j=1}^{H}\sum_{k=1}^{W}\sum_{l=1}^{C}\left|y_{i,j,k,l} - G\left(x_{i,j,k,l}\right)\right|, \quad (5)$$
$$L_G = L_{GAN} + \lambda\, L_{L1}$$

where H, W, and C describe the height, width, and channels of the input image.

· Experiment 2: Contribution of λ in generator loss

As mentioned above, the hyperparameter λ allows modeling the contribution of L1 loss in the Pix2Pix model. In this case, several values have been tested to observe their impact on the generation of benthic habitat maps. The values of λ are shown in Table II.

· *Experiment 3: Exponential scheduler of learning rate*

With the best previous models, the application of a learning rate scheduler is studied to improve the result and to make a fine adjustment of the studied values. In this case, the Exponential scheduler is used where, starting from the initial learning rate value, its value is decreased with a negative real exponential. The rate of decrease is controlled by the γ factor, where its values are shown in Table II.

· *Experiment 4: L2 regularization*

Finally, with the best models with the Exponential scheduler, the L2 regularization [57] is applied to prevent high parameter values, which can improve the portability of the model in areas other than the training and validation zones. In this case, a term is added to the loss function consisting of the quadratic sum of the parameters. This term is multiplied by hyperparameter weight decay which controls the regularization. The chosen weight decay values are presented in Table II.

D. Assessment of benthic mapping

Regarding the assessment of the models, the Accuracy, Recall, Precision, F_1 Score and Intersection over Union (IoU) metrics over the validation dataset are used. In this case, all metrics have been calculated as the average of each class, to observe their global behavior and avoid overfitting due to class imbalance. Additionally, the metrics for the *Cymodocea nodosa* class are also presented independently due to the importance of this seagrass species in the area.

Regarding the metrics, as indicated in Table I, Accuracy has been widely used in literature. In addition, Recall, Precision, and F_1 Score are also presented. On the other hand, IoU is included, which is a metric that has been recently incorporated extensively, especially in the specific field of artificial intelligence.

The formulation of the validation metrics discussed are detailed below:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \tag{6}$$

$$Precision = \frac{TP}{TP + FP}, \tag{7}$$

$$Recall = \frac{TP}{TP + FN}, \tag{8}$$

$$F_1\ Score = 2\frac{Precision \cdot Recall}{Precision + Recall}, \tag{9}$$

$$IoU = \frac{TP}{TP + FP + FN}, \tag{10}$$

where TP indicates the number of true positive pixels, TN the true negative pixels, FP the false positive pixels, and FN the false negative pixels.

In addition to using the metrics to compare the quality of the models, they are compared with each other to see which metric is more sensitive to changes. For this purpose, the standard deviation, normalized to the mean value, is used for the results of all models for each metric. In this case, the standard deviation is desired to be as large as possible, since it indicates that there is greater variability in the results, indicating better sensitivity to changes in the results.

On the other hand, other parameters have also been considered to standardize the comparison between models. These are the execution time, consisting of the combined training/validation and estimation time; the number of trainable parameters in gradient descent (GD); and the number of hyperparameters in the models.

Finally, in addition to obtaining the metrics for quantitative analysis, the qualitative analysis of the benthic maps is done to analyze the portability of the model outside the training and validation dataset.

TABLE II
DEEP LEARNING MODELS HYPERPARAMETERS.

Hyperparameter	Values
General	
Convolution kernel	3x3 pixels
Convolution padding	1 pixel
Pooling	Max-pooling
Activation function	ReLU
Normalization	Batch Normalization
Optimizer	Adam
Early stopping patience	140 epochs
Experiment 1: Loss functions and learning rates	
Loss functions for U-Net and Attention-U-Net	Binary Cross-Entropy (BCE), Cross-Entropy (CE), Weighted Cross-Entropy (WCE), Dice Loss (DL), and Generalized Dice Loss (GDL)
Loss function for Pix2Pix	BCE for adversarial loss, and L1 reconstruction loss
Learning rates	0.1, 0.01, 0.001, 0.0001, and 0.00001
Experiment 2 (only for Pix2Pix): Contribution of λ in generator loss	
λ	50, 100, 150, 200, 250
Experiment 3: Exponential scheduler of learning rate	
Scheduler of learning rate	Exponential
Exponential decay (γ) parameter of Exponential scheduler	0.9, 0.95, 0.99, 0.995, 0.999, 0.9995, and 0.9999
Experiment 4: L2 regularization	
Regularization	L2
Weight decays of L2 regularization	0.0001, 0.00025, 0.0005, 0.00075, 0.001, 0.0025, 0.005, 0.0075, 0.01, 0.025, 0.05, 0.075, and 0.1

IV. RESULTS

In this section the main results are presented where, initially, the conventional ML model results are shown, as well as an analysis of the FNNs study. Next, the results of the four experiments will be presented. Finally, the results regarding the comparison of the evaluation metrics used are included.

TABLE III
EVALUATION METRICS OF MACHINE AND DEEP LEARNING MODELS.

		Cymodocea nodosa					*Overall (mean of classes)*				
	Model	***Accuracy*↑**	***Precision*↑**	***Recall*↑**	***F_1 Score*↑**	***IoU*↑**	***Accuracy*↑**	***Precision*↑**	***Recall*↑**	***F_1 Score*↑**	***IoU*↑**
ML	*GNB*	0.7728	0.1245	0.8006	0.2155	0.1208	0.8232	0.6329	0.7069	0.5901	0.4582
	KNN	0.9372	0.2806	0.3907	0.3266	0.1952	0.9272	0.7482	0.6771	0.7006	0.5793
	DT	0.9333	0.2703	0.4191	0.3286	0.1966	0.9247	0.7471	0.7018	0.7151	0.5992
	S-KNN	0.9382	0.2826	0.3803	0.3243	0.1935	0.9296	0.7560	0.7022	0.7224	0.6106
	SVM	0.9541	0.3963	0.3376	0.3646	0.2230	0.9386	0.7842	0.6896	0.7296	0.6144
	FNN	0.9495	0.3595	0.3784	0.3687	0.2260	0.9379	0.7723	0.7112	0.7370	0.6242
DL	*U-Net with BCE*	**0.9896**	**0.8994**	0.8050	**0.8496**	**0.7386**	0.9710	0.9210	0.8784	0.8983	0.8179
	U-Net with GDL	0.9866	0.8516	0.7671	0.8071	0.6767	**0.9726**	**0.9255**	**0.8862**	**0.9047**	**0.8317**
	Attention U-Net	0.9840	0.7675	0.8057	0.7861	0.6476	0.9611	0.8807	0.8581	0.8677	0.7710
	Pix2Pix	0.9886	0.8549	**0.8283**	0.8414	0.7262	0.9661	0.9094	0.8640	0.8837	0.7944
Std ↑ (% over mean)		0.0681	0.5859	0.3768	0.5024	**0.6677**	0.0466	0.1201	0.1178	0.1374	**0.1865**

TABLE IV
OTHER METRICS OF MACHINE AND DEEP LEARNING MODELS.

	Model	***Execution time (min)***	***Number of trainable parameters in GD***	***Number of hyperparameters***
ML	*GNB*	0.03	No	2
	KNN	0.82	No	3
	DT	0.61	No	4
	S-KNN	3.64	No	4
	SVM	22.23	No	3
	FNN	10.19	$8.7 \cdot 10^3$	8
DL	*U-Net*	39.08	$31.0 \cdot 10^6$	11
	Attention U-Net	77.14	$31.6 \cdot 10^6$	12
	Pix2Pix	94.94	$31.7 \cdot 10^6$	16

The quantitative analysis, presented in Table III, examines the different metrics in general, with the mean of all the classes, and the *Cymodocea nodosa* class in particular, since it is the class of greatest interest in the study area. Additionally, Table IV provides supplementary parameters that facilitate a more comprehensive comparison among the models. As can be seen, the number of parameters between models differs. Therefore, all comparisons are contextualized with execution time, number of trainable parameters, and hyperparameters. In addition, all U-Nets are restricted to three polling steps given the limited size of the input images and the scarcity of training and validation datasets. To obtain the results, a NVIDIA GeForce RTX 3050 Ti GPU was used. Moreover, the MATLAB's Classification Learning Toolbox was used to implement conventional ML models, and the PyTorch library in Python with CUDA acceleration was used for the DL model.

On the other hand, Fig. 6 illustrates the qualitative comparison of seabed mapping results obtained with the different models analyzed in this study. Fig. 6 (a) shows a RGB representation of the WorldView-2 satellite image, as well as the isobaths at 1 m resolution in depth. It should be noted that changes in the seabed are detected up to 15 m, which is the maximum detectable depth in this image. Figs. 6 (b) to (g) corresponds to conventional ML models, while Figs. 6 (h) to (k) shows DL-based approaches. The reference bathymetry and the original WorldView-2 image are included for context. Color codes indicate the predicted classes: *Cymodocea nodosa*, red algae, sand, rock, and land. This visual comparison highlights the superior spatial consistency of DL models, particularly U-Net and Pix2Pix, in detecting seagrass patches and reducing noise in rocky areas. In addition, as a guide to the results, the presence of *Cymodocea nodosa* is mainly present in the upper central zone, as well as in the lower zone. The discussions include historical maps showing Cymodocea nodosa seagrass beds in these regions and the decrease in their extent in recent years. A visual inspection of Fig. 6 reveals that misclassification errors are more frequent among the vegetation classes (*Cymodocea nodosa* and red filamentous algae) with respect to the sand class, as it misclassifies the dark sand pixels. This is due to the low reflectance of the classes in near deep water and dark areas due to marine currents, as shown in the central right part of the maps. These errors are more prominent in the conventional ML models where only spectral information is used. In contrast, the U-Net model, which has the advantage of spatially analyzing the image, allows them to correctly separate all classes considering the spatial distribution of marine species, with a distinguished horizontal growth. In addition, vegetation classes show high separability in all maps, both in the conventional ML models and the DL model.

A. Machine Learning

Regarding the conventional ML models, initially, using the Accuracy metric for the overall classes, the best performance value is given by the FNN model with 93.79%. In this case, the other models, except for GNB, show similar values, with an absolute error of less than 1.32%. On the other hand, analyzing

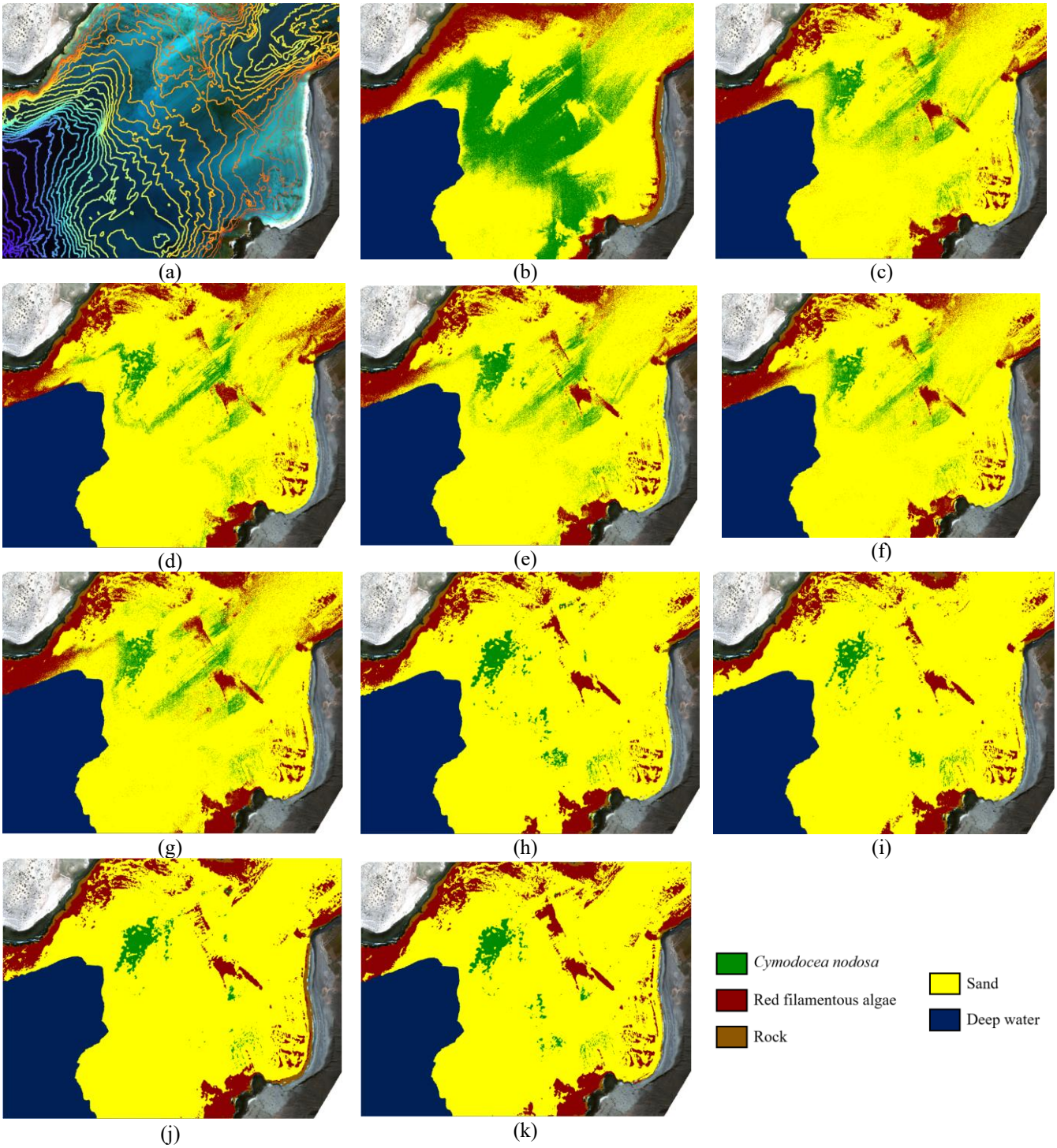


Fig. 6. Qualitative comparison of the Machine and Deep Learning models: (a) WV-2 image of January 22 with the reference isobaths up to 30 m, (b) GNB, (c) KNN, (d) DT, (e) S-KNN, (f) SVM, (g) FNN, (h) U-Net (GDL), (i) U-Net (BCE), (j) Attention U-Net (Attention Gates in skip layers), and (k) Pix2Pix (U-Net and PatchGAN).

the *Cymodocea nodosa* class, it can be highlighted that SVM presents the best Accuracy with a value of 95.41%. As in the previous case, the rest of the models, regardless of GNB, have similar values, with an absolute error of less than 2.08%. This analysis can be corroborated in Fig. 6 (b) to (g), where all the results are similar, except for GNB where a notable overestimation is observed for the *Cymodocea nodosa* class. In addition, it should be noted that the values obtained are quite high, so their use for analysis is complex to distinguish models.

Secondly, analyzing the results of the Precision metric, the best model is SVM, for the *Cymodocea nodosa* and the mean of classes. In this case, the overall values are similar between models, highlighting the SVM with a value of 78.42%, and an

absolute error with respect to the rest of the models of less than 3.71%, except for GNB with worse values. However, using this metric for the *Cymodocea nodosa* class, three groups of models can be seen: FNN and SVM; S-KNN, DT, and KNN; and GNB. In general, the best model is SVM with a metric value of 39.63%, with a difference of 3.68% with respect to FNN. In this case, it is observed that the quality of the different models can be grouped using this metric, at least in the *Cymodocea nodosa* class, as well as a greater absolute difference with respect to Accuracy, and metric values somewhat far from the maximum value. Therefore, the use of Precision may be adequate to distinguish results in *Cymodocea nodosa*.

Thirdly, the results using the Recall metric can be analyzed. The best overall is achieved by the FNN model with a value of 71.12%. In this case, all models, including GNB, offer similar results, with an absolute value of less than 3.41%. Analyzing the metrics in the *Cymodocea nodosa* class, GNB model has a much higher value than the rest of the metrics, with a value of 80.06%, and a very high absolute difference with respect to the second best model, DT, of 38.15%. This value is due to the high overestimation of *Cymodocea nodosa*, which can be observed in Fig. 6 (b). Note the definition of Recall in Equation 6, where False Negative is used. Therefore, when all the classes are considered, similar Recall values are obtained among all models. For this reason, the use of Recall is not advisable in the presence of overestimation.

Finally, the IoU metric has the same structure as the Precision and F_1 Score metrics, that is, in general, all the models have similar values, except for GNB, while in the *Cymodocea nodosa* class there are three groups. In this case, the best model is FNN, with values of 62.42% in overall results, and 22.60% in the *Cymodocea nodosa* class. In addition, the absolute error among the overall, except for GNB, is 4.49%. In this case, there is a greater difference between the models, compared to the Precision and F_1 Score, so their use is advised above these two metrics. Another fundamental aspect is that the values of the metrics in this case are further away from the maximum of 100%, so there may be a greater linear margin of improvement of the metric if they are analyzed with other models.

In general, the use of IoU is recommended, followed by F_1 Score and Precision metrics. Additionally, the use of Accuracy is discouraged, as well as Recall. Furthermore, with respect to the results, it is appreciated that the FNN and SVM models are the best Machine Learning models and offer similar results, followed by the S-KNN, DT and KNN. Moreover, the use of GNB is discouraged in this case.

Regarding the metrics on the computational and temporal impact of the models, the information in Table IV can be used. In this case, a direct relationship between execution time and improvement in metrics can be observed, except for SVM, which is a costly model. In this sense, the best relationship between execution time and performance is found in the FNN model. On the other hand, it can be observed that in all ML models, except for the FNN, gradient descent has not been used, but rather analytical and heuristic methods. It should be highlighted that Sequential Minimal Optimization (SMO) was used for the SVM model. Finally, a relationship between the number of hyperparameters and performance can also be seen, with the FNN being the model with the highest number of hyperparameters.

B. Deep Learning

Regarding the DL models, the second part of Table III shows the best U-Net, Attention U-Net and Pix2Pix models. In general, DL models significantly improve all the metrics presented with respect to ML models. For the overall IoU, an improvement over the best conventional ML model of more than 20% is observed, with a high metric value of 0.83 approximately, indicating an excellent result. In the case of the IoU for the *Cymodocea nodosa* class, the improvement is even more remarkable, of more than 51%, with an IoU value of ~0.74. On the other hand, for the second group of metrics, Precision and F_1 Score, an overall improvement of more than 14% and 16% respectively is observed, as well as an improvement in the *Cymodocea nodosa* class of more than 50% and 48% respectively. Finally, with respect to the Accuracy and Recall metrics, improvements of more than 3% and 17% respectively were observed in general, as well as improvements of more than 4% and 42% for the *Cymodocea nodosa* class. In the last case, it is observed that Accuracy shows less sensitivity in comparison of the models. It is worth mentioning that the improvement is notable in the *Cymodocea nodosa* class, the species of greatest interest in the area. Furthermore, in this case it is observed that the IoU has the highest sensitivity, since the absolute error is greater than the rest of the metrics. In general, it is observed that DL models improve considerably accuracy in the studied seagrass meadows compared to the conventional ML models.

Furthermore, with respect to the maps obtained, in Fig. 6, it can be seen that the estimation of the classes has higher accuracy. It is observed that *Cymodocea nodosa* has no noise and is well defined. In addition, the decrease in noise is also observed in the rocky bottoms, especially to the north, west and south. Therefore, in this study case, the use of the DL model is recommended in comparison with the analyzed conventional ML models.

In particular, the basic U-Net obtains the best results, which are very similar to the results obtained with the Pix2Pix model, with a difference of approximately 0.0445 for IoU. Regarding Attention U-Net, it can be observed that it is better than the ML models, but it has the worst performance of the DL models. Qualitatively, the same can be observed in Fig. 6, where the U-Net model presents the highest quality map. Regarding the metrics presented in Table IV, it can be observed that Deep Learning models have a significantly higher number of trainable parameters, with several orders of magnitude greater than in conventional Machine Learning models. This suggests more complex architectures with a greater capacity to capture the non-linearities of the dataset. A similar trend can be seen in execution time and the number of hyperparameters, where Pix2Pix has the highest computational cost and complexity, while the classic U-Net architecture is the most efficient. Given the limitations imposed by the small size of the dataset for benthic habitat classification, the standard U-Net appears to be the most suitable option. An optimal relationship between hyperparameters and efficiency is observed. However, for larger datasets and/or images, models with more hyperparameters should be reanalyzed. The two U-Net models, one presenting the best overall results, and the other for the *Cymodocea nodosa* class, were the results of the three

experiments explained above. This dual presentation allows for a more nuanced evaluation: while the overall model ensures general robustness across all seabed types, the *Cymodocea nodosa*-optimized model highlights the potential for targeted improvements in ecologically critical classes.

Finally, in the next experiments, the IoU is used as an analysis metric as it has a higher sensitivity to map changes. The results of each experiment are presented below.

· *Experiment 1: Loss functions and learning rates*

Regarding the first experiment, different loss functions and learning rates were analyzed, as shown in Table II. For U-Net models, the results showed that the best models in each loss function were BCE with LR of 0.01 obtained a general IoU of 0.85 and IoU in the *Cymodocea nodosa* class of 0.76; CE with LR 0.01 obtained 0.72 and 0.58 respectively; WCE with LR 0.001 of 0.78 and 0.66; DL with LR of 0.001 obtained 0.78 and 0.67; and GDL with LR of 0.001 obtained 0.82 and 0.72 respectively. Fig. 7 shows the distribution of values to assist in comparison.

In this case, it is observed that the best models are obtained with GDL and BCE loss functions with differences of 3% and 4% in overall IoUs and *Cymodocea nodosa* class. Moreover, as in previous studies, the overall metric is higher than for the seagrass class. Finally, it is highlighted that the use of class weights in the loss functions helps to contemplate class imbalance, improving the results. This occurs both for CE, with WCE, and DL, with GDL, so their use is recommended.

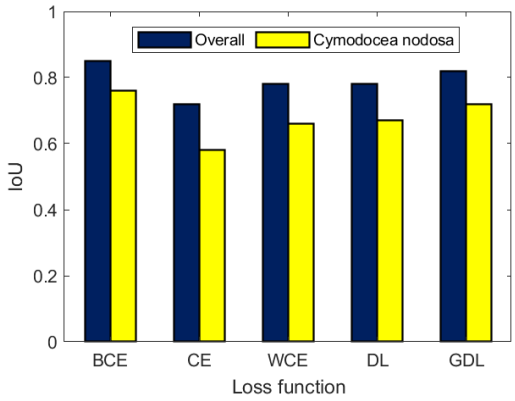


Fig. 7. Distributions of IoU values by loss function for the U-Net model in Experiment 1.

The same behavior occurs in the Attention U-Net model, where the BCE and GDL loss functions obtain the best results. In this case, for both loss functions, a learning rate of 0.0001 has been found to be optimal.

Finally, in the case of Pix2Pix, only the loss function mentioned above has been implemented. In this case, the best LR obtained was 0.001. This model has a hyperparameter exclusive to its loss function, which will be analyzed in the following experiment.

· *Experiment 2: Contribution of λ in generator loss for the Pix2Pix model*

Secondly, the λ hyperparameter exclusive to the loss function of the Pix2Pix model has been studied. This parameter controls the contribution of the L1 reconstruction loss in the generator's loss function. From the values studied, shown in Table II, it has been found that the lambda value 100 is optimal, providing the best results.

· *Experiment 3: Exponential scheduler of learning rate*

Thirdly, the Exponential scheduler of learning rate is applied to the best models of experiment 1. In this case, the previous LRs are used as starting LRs and different exponential decays (γ) values are applied, as shown in Table II. Regarding the results, the best models were in U-Net for BCE with γ of 0.9995 obtaining an overall IoU of 0.84 and IoU in the *Cymodocea nodosa* class of 0.77, and for GDL with γ of 0.999 obtaining 0.82 and 0.75, respectively. In this case, the scheduler applied does not seem to improve the BCE results, getting 1% worse, which is not statistically significant. However, in the case of GDL, improvements are observed in general, so its use is recommended in U-Net model. On the other hand, for both the Attention U-Net model, with BCE and GDL, and for Pix2Pix, the use of a learning rate scheduler worsens the results, so its use is not recommended in these cases.

· *Experiment 4: L2 regularization*

Finally, L2 regularization is applied on the best previous models. For the L2 regularization, the weight decays presented in Table II were used. For U-Net models, the results show that, for BCE without scheduler and with a weight decay of 0.0005, an overall IoU of 0.82 and an IoU in the *Cymodocea nodosa class* of 0.74 were obtained. On the other hand, in the case of GDL with scheduler and weight decay of 0.01, IoUs of 0.84 and 0.68 were obtained, respectively. These two models have been the models finally chosen for U-Net model, where the rest of the metrics are presented in Table III. It should be noted that both models are included since GDL is better overall, while BCE obtains the best results in *Cymodocea nodosa* class, that is of greatest interest in the study.

Regarding the Attention U-Net model, it should be noted that L2 regularization was not available in the AGs, as it worsened the results. This may be because it can decrease the attention capacity of the AGs, reducing the mask values. In this case, the optimal weight decay value was 0.01 using GDL and 0.001 using BCE. For Attention U-Net, the best final result was obtained with the GDL loss function.

Finally, regarding the Pix2Pix model, L2 regularization with weight decay of 0.0001 was proven to achieve the best results. Therefore, it was observed that, in general for all DL models, the learning rate scheduler is a technique that does not always succeed and must be studied in each particular case, while the use of L2 regularization is recommended.

C. Evaluation metrics

As mentioned above, the IoU metric has a higher sensitivity, so its use is recommended. In addition, the Precision and F_1 Score metrics can also be recommended, although their sensitivity is lower. However, Recall and, especially, Accuracy metrics are not recommended in this case. To formalize this analysis, the standard deviation values, in percentage with respect to the mean, of the metrics are highlighted in the last row of Table III. The values were obtained for each metric in the different models. Regarding the results, the most sensitive

metric, that is, the one with the highest standard deviation, is the IoU both for the *Cymodocea nodosa* class and in general, with standard deviation of 66.77% and 18.65% respectively. Next comes the group of Precision, Recall and F_1 Score metrics, where Precision has the highest sensitivity in the *Cymodocea nodosa* class and F_1 Score in general. Finally, the extensively used Accuracy metric is the worst by far with respect to the other metrics and its use is discouraged in this work. As discussed, there is a notable improvement with the use of IoU, and its use is encouraged.

V. Discussions

In this section, two analyses are presented. Initially, the work is discussed in relation to other works. Subsequently, the U-Net results are used to study the temporal evolution of the *Cymodocea nodosa* meadows in the area with respect to previous studies.

A. Discussions on models and metrics

Regarding the results obtained in this work, it can be highlighted that the use of Deep Learning, with encoder-decoder architecture, significantly improves the results with respect to the use of conventional and novel Machine Learning models. It should be noted that, in this case, conventional ML models only analyze the spectral information of each pixel without including the spatial and spectral information of its neighborhood. In contrast, the DL models used consider spectral information and the morphology of the spatial object, which may be a cause of the model improvement. Additionally, another aspect of the enhanced results may be that DL models have higher nonlinearity compared to the conventional ML models [83]. In addition, with respect to ML models it is of special interest to note that FNNs can be a competitor to the extensive SVM model.

In addition, the performance differences observed among conventional ML models can be attributed to their inherent structural characteristics. For instance, GNB assumes independence and Gaussian distributions, leading to poor generalization. In contrast, SVM and FNN can model non-linear boundaries more effectively, but still lack spatial context. The superior performance of U-Net is due to its encoder-decoder architecture with skip connections, which allows it to capture both local textures and global spatial patterns. Moreover, the use of convolutional layers enables the model to learn hierarchical features, which is particularly advantageous in heterogeneous benthic environments.

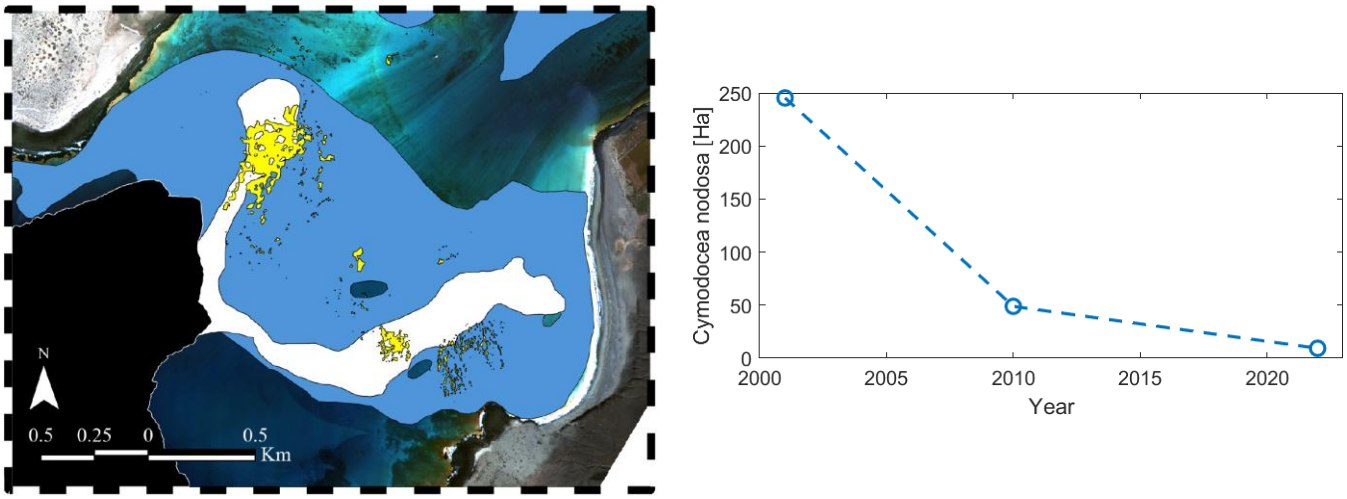
On the other hand, with respect to DL models, it is important to note that in this study, the dataset is highly constrained, consisting of only 51 original patches (204 after data augmentation) extracted from a single WorldView-2 image, with small dimensions of 70×70 pixels and 8 spectral channels. This limited size and low spatial diversity significantly increase the risk of overfitting, particularly for models with many parameters. Under these conditions, the classic U-Net architecture offers a more favorable trade-off between complexity and generalization. Attention U-Net introduces additional parameters through Attention Gates, which can enhance feature selection in large and diverse datasets but may degrade performance when data is scarce and lacks variability. Similarly, Pix2Pix, based on a conditional GAN framework, requires training both a generator and a discriminator, further increasing model complexity and sensitivity to data limitations. This adversarial setup is powerful for generating realistic benthic habitats maps but is highly dependent on abundant and diverse training samples to stabilize training. In contrast, the standard U-Net, with fewer parameters and a simpler optimization process, is better suited for small, homogeneous datasets, reducing the risk of overfitting while still leveraging spatial and spectral information effectively. In addition, the Attention U-Net and Pix2Pix models are more complex in terms of optimal hyperparameter selection. However, it is important to highlight that the results are much better than those of ML models, where Pix2Pix offers very similar results to classic U-Net. In any case, for studies with limited data, the classic U-Net is recommended. However, if the training dataset is increased, it should be re-evaluated, as Pix2Pix and Attention U-Net can better leverage their potential. Finally, it would be necessary to comment on the limitations of using DL compared to traditional ML, where there is a larger number of hyperparameters, as well as a greater dependence on the training data [87]. In complex cases, where spectral separability is compromised, DL models are of great use. However, in simpler cases, the use of conventional ML models can provide great results.

Regarding the assessment metrics, it was concluded that the use of IoU, as well as F_1 Score and Precision, are of special interest due to their sensitivity to map changes.

To the authors' knowledge, this is the first analysis of the best metrics for benthic habitat mapping using passive multispectral satellite remote sensing. Of particular interest, however, is the work of Maxwell et al. [52], where an analysis of the impact of the use of different assessment metrics on land classification is performed. As in this work, the use of the IoU and F_1 Score was highlighted as they provide a way to summarize the performance of the models.

B. Temporal evolution of Cymodocea nodosa meadows

To study the temporal evolution of *Cymodocea nodosa* meadows in El Río, La Graciosa, the result of the U-Net model with the BCE loss function is used, since it has the best quality in the studied class. Based on the result, it is concluded that, in the study area and up to 15 meters depth, corresponding to the maximum detectable depth in this case, the extent of *Cymodocea nodosa* seagrass meadows is approximately 9.31 Ha for 2022. Other studies with in-situ data have previously been carried out in the study area, where the 2001 and 2010 studies stand out [54]. Fig. 8(a) shows the historical maps compared to the new distribution, where a remarkable deterioration of the seagrass meadows can be observed. Specifically, in 2001 there was a population of about 245.32 Ha (blue), reduced to 48.84 Ha in 2010 (white). In this case, by 2022 only 9.31 Ha are detected (yellow), which represents a reduction on the population of 96.2% with respect to 2001 approximately. It should be highlighted that the decline has occurred in a period of only 21 years. On the other hand, Fig. 8(b) shows the temporal evolution of the number of hectares of *Cymodocea nodosa*. In this case, if the trend continues, in the coming years it is expected that *Cymodocea nodosa* will disappear completely in the study area.



(a) (b)

Fig. 8. Evolution of the *Cymodocea nodosa* population in El Rio: (a) Comparison of the historical maps of 2001 (blue) and 2010 (white), and the new distribution of 2022 (yellow). Deep water up to 15 m is shown (black mask). (b) Temporal evolution of the population in the study area.

Additionally, it is worth noting that, for the comparison to be correct, the deep water pixels have been masked, as shown in Fig. 8(a). The depth limit is a limitation in all water passive remote sensing studies. However, it should be noted that the *Cymocodea nodosa* meadows in this case are abundant in shallow water. On the other hand, the high accuracy in the maps generated stands out, due to the high spatial resolution of the satellite image used.

Regarding the factors that can affect seagrass meadows deterioration, natural effects, as well as direct anthropogenic effects are highlighted. Natural impacts include intense marine currents, which cause massive sediment movement and root upwelling, and rainfall runoff, which increases water turbidity and reduces light penetration, affecting plant photosynthesis [88], [89].

On the other hand, anthropogenic impacts are more frequent and severe, especially the construction of coastal infrastructures such as ports and artificial beaches, wastewater and brine discharges from desalination plants or potable water that change salinity, and trawling [90]. These factors increase turbidity, plant burial and pollution, leading to loss of habitat and biodiversity [90]. In these aspects, the study area is supported primarily by tourism and fishing, as well as there have been breaks in the pipeline that transports potable water from Lanzarote to La Graciosa [91], seriously endangering the health of marine flora and fauna. In addition, there are plants that can flourish with the change of water characteristics, so the populations of *Cymodocea nodosa* also must deal with competition from endemic and invasive species [92]. In this case, filamentous red algae, present at all measurement points in 2022, was previously not detected in the study area (2010). Unfortunately, all the commented aspects lead to the disappearance of historical seagrass meadows, endangering biodiversity and ecosystem stability.

C. Future directions

Based on the results obtained, several directions for model improvement are suggested. Initially, when the dataset is limited, as in this case, the Self-Supervised Learning (SSL) technique [93] is particularly noteworthy. This approach is based on generating robust representations by pretextual tasks, such as predicting temporal sequences or reconstructing images, which drastically reduces dependence on manual annotations and improves generalization in environments with high spatial and temporal variability. For example, [93] demonstrates that SSL is particularly effective in scenarios with high temporal and spatial variability, such as environmental monitoring and multitemporal analysis. In addition, the use of Vision Transformer (ViT) architectures [94], adaptations of Transformers to images, may also be interesting, where images are treated as sequences of patches. However, the impact of reduced datasets on their performance must be carefully studied, as these architectures often require large volumes of data to avoid overfitting. Therefore, the use of pre-trained models based on ViT architecture is particularly promising, notably the Segment Anything Model (SAM) [95] and Prithvi-EO-2.0 [96].The last one is especially interesting because it's a trained multitemporal ViT model that also includes self-supervised strategies on large volumes of satellite data. So, it can be adjusted with transfer learning using minimal amounts of labeled data, achieving near state-of-the-art performance in tasks like segmentation and classification in complex geospatial environments. Consequently, the integration of these techniques and models in future studies could mitigate current limitations arising from data limitations and improve accuracy in benthic mapping. Additionally, more advanced DL encoder-decoder architectures can also be implemented, such as SegNet or DeepLabv3+, which may better capture spatial hierarchies. Incorporating auxiliary data sources (e.g., bathymetry or hyperspectral imagery) could enhance class separability. Additionally, transfer learning and attention mechanisms may improve generalization in data-scarce environments. These

strategies could further reduce misclassification in complex benthic habitats.

VI. Conclusions

This study presents a comprehensive comparison between conventional Machine Learning and convolutional Deep Learning models for mapping seagrass meadows using high-resolution WorldView-2 satellite imagery. For this study, the species *Cymodocea nodosa* in El Río, Canary Islands, were analyzed, and, up to the authors' knowledge, the first dataset for the segmentation of these meadows was generated using WorldView-2 satellite imagery is presented.

The results obtained demonstrate that DL models, U-Net, Attention U-Net and Pix2Pix, particularly the classic U-Net architecture, significantly outperform conventional ML models of Gaussian Naïve Bayes (GNB), Decision Tree (DT), K-Nearest Neighbors (KNN), Subspace KNN, Support Vector Machine (SVM), and Feedforward Neural Network (FNN). For example, the U-Net model with the Binary Cross-Entropy (BCE) loss function achieved an overall Intersection over Union (IoU) of 82% and a value of 74% specifically for *Cymodocea nodosa*, improving by more than 50% the results obtained with the best conventional ML models. This was especially evident in the detection and segmentation of *Cymodocea nodosa*, where the DL models showed a greater ability to handle the complexity and variability of the data.

In addition, an incremental study has been carried out for the training of the DL model, modifying the hyperparameters that control the network, as well as different loss functions and additional techniques, such as exponential scheduler of learning rate and L2 regularization. In addition, further study was conducted for the Pix2Pix model, where the hyperparameter λ in the generator's loss function was studied. Regarding the models, the U-Net model is recommended among the models studied for datasets with limited sizes, as in this study. However, the Attention U-Net and Pix2Pix models should be studied with more extensive datasets. On the other hand, a detailed analysis of different evaluation metrics has been conducted, with the metrics of Accuracy, Precision, Recall, F_1 Score, and Intersection over Union (IoU). The IoU metric stands out for its high sensitivity to map changes.

The study also reveals an alarming reduction in the extent of *Cymodocea nodosa* meadows in the area, representing a decline of 96.2% over a 21-year period. This drastic reduction underscores the urgent need to implement conservation and sustainable management measures.

The findings of this study highlight the importance of using advanced DL techniques to improve the accuracy and efficiency of seagrass mapping, thus contributing to the conservation and management of these vital ecosystems. Also, the usefulness of satellite remote sensing in monitoring coastal seagrass meadows can be observed.

References

[1] C. D. Woodroffe, N. Evelpidou, I. Delgado-Fernandez, D. R. Green, A. Karkani, and P. Ciavola, "Coastal Systems: The Dynamic Interface Between Land and Sea," in Research Directions, Challenges and Achievements of Modern Geography, Springer, 2023, pp. 207–229. DOI: 10.1007/978-981-99-6604-2_11.

[2] L. Jiang, T. Yang, X. Wang, J. Yu, J. Liu, and K. Zhang, "Research on integrated coastal zone management from past to the future: a bibliometric analysis," Front. Mar. Sci., vol. 10, 2023. DOI: 10.3389/fmars.2023.1201811.

[3] B. P. Hayden, G. C., Ray, and R. Dolan, "Classification of Coastal and Marine Environments," Environmental Conservation, Cambridge, vol. 11, no. 3, pp. 199-207, 1984. DOI: 10.1017/S0376892900011449.

[4] S. Refulio-Coronado, K. Lacasse, T. Dalton, A. Humphries, S. Basu, H. Uchida, and E. Uchida, "Coastal and Marine Socio-Ecological Systems: A Systematic Review," Frontiers in Marine Science, vol. 8, 2021. DOI: 10.3389/fmars.2021.1234567.

[5] K. S. Tew, J. Kuo, J.-O. Cheng, F.-C. Ko, P.-J. Meng, and A. B. Mayfield, "Impacts of Seagrass on Benthic Microalgae and Phytoplankton Communities in an Experimentally Warmed Coral Reef Mesocosm," Front. Mar. Sci., vol. 8, 2021. DOI: 10.3389/fmars.2021.679683.

[6] M. do A. C. Lima, T. F. Bergamo, R. D. Ward, and C. B. Joyce, "A review of seagrass ecosystem services: providing nature-based solutions for a changing world," Hydrobiologia, vol. 850, pp. 2655–2670, 2023. DOI: 10.1007/s10750-023-05244-0.

[7] M. Unsworth, L. Cullen-Unsworth, J. Pretty, and R. Smith, "Seagrass meadows provide coastal protection and support higher biodiversity," Nature Communications, vol. 10, 2019. DOI: 10.1038/s41467-019-12345-6.

[8] J. M. Pandolfi, S. R. Connolly, D. J. Marshall, and A. L. Cohen, "Projecting coral reef futures under global warming and ocean acidification," Science, vol. 333, no. 6041, pp. 418–422, 2011. DOI: 10.1126/science.1204794.

[9] S. Vitousek, D. Buscombe, K. Vos, P. L. Barnard, A. C. Ritchie, and J. A. Warrick, "The future of coastal monitoring through satellite remote sensing," Cambridge Prisms: Coastal Futures, vol. 1, 2023. DOI: 10.1017/cft.2022.4.

[10] M. J. McCarthy, K. E. Colna, M. M. El-Mezayen, A. E. Laureano-Rosario, P. Méndez-Lázaro, D. B. Otis, G. Toro-Farmer, M. Vega-Rodriguez, and F. E. Muller-Karger, "Satellite Remote Sensing for Coastal Management: A Review of Successful Applications," Environmental Management, vol. 60, pp. 323–339, 2017. DOI: 10.1007/s00267-017-0880-x.

[11] F. Eugenio, J. Marcello, A. Mederos-Barrera, and F. Marqués, "High-resolution satellite bathymetry mapping: Regression and machine learning-based approaches," IEEE Transactions on Geoscience and Remote Sensing, 60, 1-14, 2021. DOI: 10.1109/TGRS.2021.3135462.

[12] F. Eugenio, J. Marcello, and J. Martin, "High-resolution maps of bathymetry and benthic habitats in shallow-water environments using multispectral remote sensing imagery," IEEE Transactions on Geoscience and Remote Sensing, 53(7), 3539-3549, 2015. DOI: 10.1109/TGRS.2014.2377300.

[13] S. Niculescu, J. Xia, and D. Roberts, "Remote Sensing for Coastal and Aquatic Ecosystems' Monitoring and Biodiversity Management," Remote Sens., vol. 16, no. 1, 2024. DOI: 10.3390/rs16010179.

[14] H. J. van der Woerd and M. R. Wernand, "True Colour Classification of Natural Waters with Medium-Spectral Resolution Satellites: SeaWiFS, MODIS, MERIS and OLCI," Sensors, vol. 15, no. 10, pp. 25663-25680, 2015. DOI: 10.3390/s151025663.

[15] M. K. Heenkenda, K. E. Joyce, S. W. Maier, and R. Bartolo, "Mangrove Species Identification: Comparing WorldView-2 with Aerial Photographs," Remote Sens., vol. 6, no. 7, pp. 6064-6088, 2014. DOI: 10.3390/rs6076064.

[16] F. Eugenio, J. Marcello, J. Martin, and D. Rodríguez-Esparragón, "Benthic Habitat Mapping Using Multispectral High-Resolution Imagery: Evaluation of Shallow Water Atmospheric Correction Techniques," Sensors, vol. 17, no. 11, 2017. DOI: 10.3390/s17112639.

[17] H. Mohamed, K. Nadaoka, and T. Nakamura, "Towards Benthic Habitat 3D Mapping Using Machine Learning Algorithms and Structures from Motion Photogrammetry," Remote Sens., vol. 12, no. 1, 2020. DOI: 10.3390/rs12010127

[18] C. Frederick, S. Villar, and Z.-H. Michalopoulou, "Seabed classification using physics-based modeling and machine learning," J. Acoust. Soc. Am., vol. 148, no. 2, pp. 859-872, 2020. DOI: 10.1121/10.0001728.

[19] X. Yan, T. Zhang, W. Du, Q. Meng, X. Xu, and X. Zhao, "A Comprehensive Review of Machine Learning for Water Quality Prediction over the Past Five Years," J. Mar. Sci. Eng., vol. 12, no. 1, 2024. DOI: 10.3390/jmse12010159.

[20] G. Cerulli, "Discriminant Analysis, Nearest Neighbor, and Support Vector Machine," in Machine Learning Techniques, Springer, 2023. DOI: 10.1007/978-3-031-41337-7_4.

[21] G. Chaubey, P. R. Gavhane, D. Bisem, and S. K. Arjaria, "Customer purchasing behavior prediction using machine learning techniques," J. Ambient Intell. Humaniz. Comput., vol. 13, pp. 1-15, 2022. DOI: 10.1007/s12652-022-03837-6.

[22] R. Polikar, "Ensemble Learning," in Ensemble Machine Learning: Methods and Applications, Springer, 2012, pp. 1-34. DOI: 10.1007/978-1-4419-9326-7_1.

[23] J. Marcello, F. Eugenio, J. Martín, and F. Marqués, "Seabed mapping in coastal shallow waters using high resolution multispectral and hyperspectral imagery," Remote Sensing, 10(8), 1208, 2018. DOI: 10.3390/rs10081208.

[24] M. Li, S. Zang, B. Zhang, S. Li, and C. Wu, "A Review of Remote Sensing Image Classification Techniques: the Role of Spatio-contextual Information," European Journal of Remote Sensing, vol. 47, pp. 389-411, 2014. DOI: 10.5721/EuJRS20144723.

[25] B. Jiang, X. An, S. Xu, and Z. Chen, "Intelligent Image Semantic Segmentation: A Review Through Deep Learning Techniques for Remote Sensing Image Analysis," Journal of the Indian Society of Remote Sensing, vol. 51, pp. 1865–1878, 2023. DOI: 10.1007/s12524-022-01496-w.

[26] Eliwa, E. H. I., & Abd El-Hafeez, T. (2025). Advancing crop health with E. Hassan, S. A. Ghazalah, N. El-Rashidy, T. A. El-Hafeez, and M. Y. Shams, "Sustainable deep vision systems for date fruit quality assessment using attention-enhanced deep learning models," Frontiers in Plant Science, vol. 16, Art. no. 1521508, 2025. doi: 10.3389/fpls.2025.1521508.

[27] E. Hassan, S. A. Ghazalah, N. El-Rashidy, T. A. El-Hafeez, and M. Y. Shams, "Sustainable deep vision systems for date fruit quality assessment using attention-enhanced deep learning models," Frontiers in Plant Science, vol. 16, Art. no. 1521508, 2025. doi: 10.3389/fpls.2025.1521508.

[28] M. Y. Shams, W. M. Elmessery, A. A. T. Oraiath, A. Elbeltagi, A. Salem, P. Kumar, T. M. El-Messery, T. Abd El-Hafeez, M. F. Abdelshafie, G. G. Abd El-Wahhab, I. S. El-Soaly, and A. E. Elwakeel, "Automated on-site broiler live weight estimation through YOLO-based segmentation," Smart Agricultural Technology, vol. 10, Art. no. 100828, 2025. doi: 10.1016/j.atech.2025.100828.

[29] W. M. Elmessery, D. V. Maklakov, T. M. El-Messery, D. A. Baranenko, J. Gutiérrez, M. Y. Shams, T. Abd El-Hafeez, S. Elsayed, S. K. Alhag, F. S. Moghanm, M. A. Mulyukin, Y. Y. Petrova, and A. E. Elwakeel, "Semantic segmentation of microbial alterations based on SegFormer," Frontiers in Plant Science, vol. 15, Art. no. 1352935, 2024. doi: 10.3389/fpls.2024.1352935.

[30] G. Mostafa, H. Mahmoud, T. Abd El-Hafeez, and M. E. ElAraby, "Feature reduction for hepatocellular carcinoma prediction using machine learning algorithms," Journal of Big Data, vol. 11, no. 1, Art. no. 88, 2024. doi: 10.1186/s40537-024-00944-3.O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Springer, 2015, pp. 234–241. DOI: 10.1007/978-3-319-24574-4_28.

[31] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Springer, 2015, pp. 234–241. DOI: 10.1007/978-3-319-24574-4_28.

[32] M.-H. Guo, T.-X. Xu, J.-J. Liu, Z.-N. Liu, P.-T. Jiang, T.-J. Mu, S.-H. Zhang, R. R. Martin, M.-M. Cheng, and S.-M. Hu, "Attention mechanisms in computer vision: A survey," Computational Visual Media, vol. 8, no. 3, pp. 331–368, 2022. doi: 10.1007/s41095-022-0271-y.

[33] O. Oktay, J. Schlemper, L. Le Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: Learning where to look for the pancreas," arXiv preprint, 2018. doi: 10.48550/arXiv.1804.03999.

[34] A. Jabbar, X. Li, and B. Omar, "A survey on generative adversarial networks: Variants, applications, and training," ACM Computing Surveys, vol. 54, no. 8, pp. 1–49, 2021. doi: 10.1145/3463475.

[35] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2017, pp. 1125–1134. doi: 10.48550/arXiv.1611.07004.

[36] J. Richards, "Remote Sensing Digital Image Analysis," Springer, 2013. DOI: 10.1007/978-3-642-30062-2.

[37] F. Li and B. Xiao, "Aquatic vegetation mapping based on remote sensing imagery: An application to Honghu Lake," *2011 International Conference on Remote Sensing, Environment and Transportation Engineering*, Nanjing, 2011, pp. 4832-4836, DOI: 10.1109/RSETE.2011.5965394.P

[38] R. Pu, S. Bell, and C. Meyer, "Mapping and assessing seagrass bed changes in Central Florida's west coast using multitemporal Landsat TM imagery," Estuarine, Coastal and Shelf Science, vol. 149, pp. 68-79, 2014. DOI: 10.1016/j.ecss.2014.07.014.

[39] K. Topouzelis, S. C. Spondylidis, A. Papakonstantinou, and N. Soulakellis, "The use of Sentinel-2 imagery for seagrass mapping: Kalloni Gulf (Lesvos Island, Greece) case study," in Fourth International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2016), vol. 9688, pp. 460-466, Aug. 2016. DOI: 10.1117/12.2242887.

[40] J. Marcello, F. Eugenio, J. Martín, and F. Marqués, "Seabed mapping in coastal shallow waters using high resolution multispectral and hyperspectral imagery," Remote Sensing, vol. 10, no. 8, p. 1208, 2018. DOI: 10.3390/rs10081208.

[41] D. Traganos and P. Reinartz, "Machine learning-based retrieval of benthic reflectance and Posidonia oceanica seagrass extent using a semi-analytical inversion of Sentinel-2 satellite data," International Journal of Remote Sensing, vol. 39, no. 24, pp. 9428-9452, 2018. DOI: 10.1080/01431161.2018.1519289.

[42] S. Bayyana, S. Pawar, S. Gole, S. Dudhat, A. Pande, D. Mitra, K. Sivakumar, "Detection and mapping of seagrass meadows at Ritchie's archipelago using Sentinel 2A satellite imagery," Current Science, vol. 118, no. 8, pp. 1275-1282, 2020. DOI: 10.18520/cs/v118/i8/1275-1282.

[43] M. M. Coffer, B. A. Schaeffer, R. C. Zimmerman, V. Hill, J. Li, K. A. Islam, and P. J. Whitman, "Performance across WorldView-2 and RapidEye for reproducible seagrass mapping," Remote Sensing of Environment, vol. 250, p. 112036, 2020. DOI: 10.1016/j.rse.2020.112036.

[44] A. Mederos-Barrera, J. Marcello, F. Eugenio, and E. Hernández, "Seagrass mapping using high resolution multispectral satellite imagery: A comparison of water column correction models," International Journal of Applied Earth Observation and Geoinformation, vol. 113, p. 102990, 2022. DOI: 10.1016/j.jag.2022.102990.

[45] M. Scarpetta, P. Affuso, M. De Virgilio, M. Spadavecchia, G. Andria, and N. Giaquinto, "Monitoring of Seagrass Meadows Using Satellite Images and U-Net Convolutional Neural Network," in 2022 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), Ottawa, ON, Canada, 2022, pp. 1-6. DOI: 10.1109/I2MTC48687.2022.9806535.

[46] M. Meister and J. J. Qu, "Quantifying Seagrass Density Using Sentinel-2 Data and Machine Learning," Remote Sensing, vol. 16, no. 7, p. 1165, 2024. DOI: 10.3390/rs16071165.

[47] L. Wang, H. Liang, S. Wang, D. Sun, J. Li, H. Zhang, and Y. Yuan, "Estimating four-decadal variations of seagrass distribution using satellite data and deep learning methods in a marine lagoon," Science of The Total Environment, vol. 919, p. 170936, 2024. DOI: 10.1016/j.scitotenv.2024.170936.

[48] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5-32, 2001. DOI: 10.1023/A:1010933404324.

[49] K. O'Shea and R. Nash, "An Introduction to Convolutional Neural Networks," arXiv preprint arXiv:1511.08458, 2015. DOI: 10.48550/arXiv.1511.08458.

[50] S. E. Whang and J.-G. Lee, "Data Collection and Quality Challenges for Deep Learning," Proc. VLDB Endowment, vol. 13, no. 12, pp. 3429–3432, 2020. DOI: 10.14778/3415478.3415562.

[51] A. Somani, A. Horsch, and D. K. Prasad, Interpretability in Deep Learning, Springer, 2023. DOI: 10.1007/978-3-031-20639-9.

[52] A. E. Maxwell, T. A. Warner, y L. A. Guillén, "Accuracy assessment in convolutional neural network-based deep learning remote sensing studies—Part 1: Literature review," Remote Sensing, vol. 13, no. 13, p. 2450, 2021. DOI: 10.3390/rs13132450.

[53] F. Espino, F. Tuya, I. Blanch, and R. Haroun, "Los sebadales de Canarias. Oasis de vida en los fondos arenosos," BIOGES, Universidad de Las Palmas de Gran Canaria, 2008. [Online]. Available: https://accedacris.ulpgc.es/handle/10553/107045. [Accessed: August 19, 2025].

[54] L. Martín-García, I. Brito-Izquierdo, and A. Brito-Hernández, "Bionomía bentónica de las Reservas Marinas de Canarias (España). Comunidades y hábitats bentónicos del infralitoral," Vieraea, vol. 44, pp. 181-200, 2016. [Online]. Available: https://www.mapa.gob.es/dam/mapa/contenido/pesca/publicaciones/libro_bionomiareservasmarinas_27_dic_prot.pdf. [Accessed: August 19, 2025].

[55] E. Mason, F. Colas, J. Molemaker, A. F. Shchepetkin, C. Troupin, J. C. McWilliams, and P. Sangrà, "Seasonal variability of the Canary Current: A numerical study," J. Geophys. Res. Oceans, vol. 116, no. C6, Jun. 2011, Art. no. C06001, doi: 10.1029/2010JC006665.

[56] J. Smith, "Preprocessing Techniques for WorldView-2 Imagery," Remote Sensing, vol. 12, no. 4, pp. 567-580, Apr. 2020. DOI: 10.3390/rs12040567.

[57] S. Ruder, An Overview of Gradient Descent Optimization Algorithms, 1st ed. Springer, 2020. DOI: 10.1007/978-3-030-50732-9.

[58] A. Hackeloeer, K. Klasing, J. M. Krisp, and L. Meng, "Georeferencing: a review of methods and applications," Annals of GIS, vol. 20, no. 1, pp. 61-69, Mar. 2014. DOI: 10.1080/19475683.2013.868826.

[59] Y. Du et al., "Water Bodies' Mapping from Sentinel-2 Imagery with Modified Normalized Difference Water Index at 10-m Spatial Resolution Produced by Sharpening the SWIR Band," Remote Sensing, vol. 8, no. 4, Apr. 2016. DOI: 10.3390/rs8040354.

[60] A. Jenerowicz, D. Wierzbicki, and M. Kedzierski, "Radiometric Correction with Topography Influence of Multispectral Imagery Obtained from Unmanned Aerial Vehicles," Remote Sensing, vol. 15, no. 8, Apr. 2023. DOI: 10.3390/rs15082059.

[61] E. Vermote et al., "Second Simulation of a Satellite Signal in the Solar Spectrum (6S)," Applied Optics, vol. 35, no. 15, pp. 2758-2776, May 1997. DOI: 10.1364/AO.35.002758.

[62] J. Marcello, F. Eugenio, U. Perdomo, and A. Medina, "Assessment of atmospheric algorithms to retrieve vegetation in natural protected areas using multispectral high resolution imagery," Sensors, 16(10), 1624, 2016. DOI: 10.3390/s16101624.

[63] NASA, "Giovanni: A System for Rapid Access, Visualization and Analysis of Earth Science Data," NASA Technical Reports Server, Jan. 2005. Available: https://giovanni.gsfc.nasa.gov/giovanni/.

[64] J. D. Hedley, A. R. Harbone, and P. J. Mumby, "Simple and robust removal of sun glint for mapping shallow-water benthos," International Journal of Remote Sensing, vol. 26, no. 10, pp. 2107-2112, May 2005. DOI: 10.1080/01431160500034086."

[65] E. M. K. Reddy, A. Gurrala, V. B. Hasitha, and K. V. R. Kumar, "Introduction to Naive Bayes and a review on its subtypes with applications," in Bayesian Reasoning and Gaussian Processes for Machine Learning Applications, 1st ed., New York, NY, USA: Chapman and Hall/CRC, 2022, pp. 1-14. DOI: 10.1201/9781003164265-1.

[66] J. Berger, "Statistical Decision Theory and Bayesian Analysis," 2nd ed., Springer, 1985, pp. 98-105. DOI: 10.1007/978-1-4757-4286-2.

[67] L. Breiman, J. Friedman, R. Olshen, and C. Stone, Classification and Regression Trees, 1st ed. CRC Press, 1984. DOI: 10.1201/9781315139470.

[68] C. Gini, "Measurement of Inequality of Incomes," The Economic Journal, vol. 31, no. 121, pp. 124-126, Mar. 1921. DOI: 10.2307/2223319.

[69] V. Vapnik, The Nature of Statistical Learning Theory, 2nd ed. Springer, 1999. DOI: 10.1007/978-1-4757-3264-1.

[70] B. Schölkopf and A. Smola, Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond, MIT Press, 2002. DOI: 10.7551/mitpress/4175.001.0001.

[71] T. Cover and P. Hart, "Nearest Neighbor Pattern Classification," IEEE Transactions on Information Theory, vol. 13, no. 1, pp. 21-27, Jan. 1967. DOI: 10.1109/TIT.1967.1053964.

[72] T. Dietterich, "Ensemble Methods in Machine Learning," Multiple Classifier Systems, vol. 1857, pp. 1-15, 2000. DOI: 10.1007/3-540-45014-9_1.

[73] T. K. Ho, "The Random Subspace Method for Constructing Decision Forests," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 8, pp. 832-844, Aug. 1998. DOI: 10.1109/34.709601.

[74] T. K. Ho, "Nearest Neighbors in Random Subspaces," Advances in Pattern Recognition, Lecture Notes in Computer Science, vol. 1451, pp. 640-648, 1998. DOI: 10.1007/BFb0033288.

[75] S. Haykin, Neural Networks and Learning Machines, 3rd ed. Pearson, 2008. DOI: 10.5555/1516246.

[76] J. C. Ye and W. K. Sung, "Understanding Geometry of Encoder-Decoder CNNs," arXiv preprint arXiv:1901.07647, 2019. DOI: 10.48550/arXiv.1901.07647.

[77] L. Prechelt, "Early Stopping — But When?," in Neural Networks: Tricks of the Trade, 2nd ed., Springer, 2012, pp. 53-67. DOI: 10.1007/978-3-642-35289-8_5.

[78] O. V. Johnson et al., "Learning Rate Schedules and Optimizers, A Game Changer for Deep Neural Networks," in Advances in Intelligent Computing Techniques and Applications, Springer, 2024, pp. 327-340. DOI: 10.1007/978-3-031-59711-4_28.

[79] I. Nusrat and S.-B. Jang, "A Comparison of Regularization Techniques in Deep Neural Networks," Symmetry, vol. 10, no. 11, 2018. DOI: 10.3390/sym10110648.

[80] P. Hurtik, S. Tomasiello, J. Hula, and D. Hynar, "Binary cross-entropy with dynamical clipping," Neural Computing and Applications, vol. 34, pp. 12029–12041, Mar. 2022. DOI: 10.1007/s00521-022-07091-x.

[81] A. Mao, M. Mohri, and Y. Zhong, "Cross-Entropy Loss Functions: Theoretical Analysis and Applications," arXiv preprint arXiv:2304.07288, 2023. DOI: 10.48550/arXiv.2304.07288.

[82] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso, "Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations," in Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Springer, 2017, pp. 240–248. DOI: 10.1007/978-3-319-67558-9_28.

[83] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436-444, 2015. DOI: 10.1038/nature14539.

[84] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, y J. Liang, "UNet++: A Nested U-Net Architecture for Medical Image Segmentation," en 4th Deep Learning in Medical Image Analysis (DLMIA) Workshop, 2018. DOI: 10.48550/arXiv.1807.10165.

[85] V. Badrinarayanan, A. Kendall, y R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 12, pp. 2481-2495, 2017. DOI: 10.48550/arXiv.1511.00561.

[86] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, y H. Adam, "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation," en ECCV 2018, 2018. DOI: 10.48550/arXiv.1802.02611.

[87] I. Goodfellow, Y. Bengio, y A. Courville, "Deep Learning," MIT Press, 2016. DOI: 10.5555/3086952.

[88] B. Daru and B. M. Rock, "Climate Change Impacts on Seagrass Meadows," Nature Plants, vol. 6, pp. 123-134, Jun. 2023. DOI: 10.1038/s41477-023-01034-5.

[89] C.-F. Boudouresque, A. Blanfuné, G. Pergent, and T. Thibaut, "Restoration of Seagrass Meadows in the Mediterranean Sea: A Critical Review of Effectiveness and Ethical Issues," Water, vol. 13, no. 8, 2021. DOI: 10.3390/w13081034.

[90] M. Román, E. Fernández, J. Zamborain-Mason, and G. Méndez, "Anthropogenic Impact on Zostera noltei Seagrass Meadows (NW Iberian Peninsula) Assessed by Carbon and Nitrogen Stable Isotopic Signatures," Estuaries and Coasts, vol. 42, pp. 987-1000, Mar. 2019. DOI: 10.1007/s12237-019-00549-7.

[91] I. A. Hernández, "La Graciosa, sin agua: seis roturas de una tubería submarina en ocho meses," El Diario, Aug. 2022. [Online]. Available: https://www.eldiario.es/canariasahora/sociedad/graciosa-agua-seis-roturas-tuberia-submarina-ocho-meses_1_9259941.html [Accessed: August 19, 2025].

[92] M. Najdek et al., "Effects of the Invasion of Caulerpa cylindracea in a Cymodocea nodosa Meadow in the Northern Adriatic Sea," Frontiers in Marine Science, vol. 7, Dec. 2020. DOI: 10.3389/fmars.2020.602055.

[93] I. Dumeur, S. Valero, and J. Inglada, "Self-supervised spatio-temporal representation learning of satellite image time series," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 17, pp. 4350–4367, 2024. doi: 10.1109/JSTARS.2024.3358066.

[94] Y. Liu, Y. Zhang, Y. Wang, F. Hou, J. Yuan, J. Tian, Y. Zhang, Z. Shi, J. Fan, Z. He, "A survey of visual transformers," IEEE Transactions on Neural Networks and Learning Systems, vol. 35, no. 6, pp. 7478–7498, 2023. doi: 10.48550/arXiv.2111.06091.

[95] A. Kirillo, E. Mintun, N. Ravi, H. Mao, C. Rolland, and L. Gustagson, "Segment anything," in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), 2023, pp. 4015–4026. doi: 10.1109/ICCV51070.2023.00371.

[96] D. Szwarcman et al., "Prithvi-EO-2.0: A Versatile Multi-Temporal Foundation Model for Earth Observation Applications," arXiv preprint, 2024. doi: 10.48550/arXiv.2412.02732.