Ecological Informatics 86 (2025) 103066

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

Cetacean feeding modelling using machine learning: A case study of the Central-Eastern Mediterranean Sea

Carla Cherubini [a,b,c], Giulia Cipriano [d], Leonardo Saccotelli [c], Giovanni Dimauro [e], Giovanni Coppini [c], Roberto Carlucci [d], Carmelo Fanizza [f], Rosalia Maglietta [a,c,*]

[a] Institute of Intelligent Industrial Technologies and Systems for Advanced Manufacturing, CNR-STIIMA, Bari, Italy
[b] Department of Electrical and Information Engineering, Polytechnic of Bari, Bari, Italy
[c] CMCC Foundation - Euro-Mediterranean Centre on Climate Change, Lecce, Italy
[d] Department of Biosciences, Biotechnology and Environment, University of Bari Aldo Moro, Bari, Italy
[e] Department of Computer Science, University of Bari, Bari, Italy
[f] Jonian Dolphin Conservation, Taranto, Italy

ARTICLE INFO

Keywords:
Machine learning
Random Forest
Species distribution models
Cetacean conservation
Feeding behaviour
Behavioural science

ABSTRACT

Investigating environmental drivers of cetacean feeding behaviour is essential for effective marine resource management, especially in the Mediterranean Sea, a biodiversity hotspot heavily impacted by human activities and climate change. This study realized a pioneer assessment of feeding activity related to the marine environment for three cetacean species - striped dolphin, common bottlenose dolphin, and Risso's dolphin - in the Gulf of Taranto (Northern Ionian Sea, Central-eastern Mediterranean) using an innovative Machine Learning (ML) approach. Behavioural data from April 2016 to October 2023, coupled with 20 environmental variables from Copernicus Marine Service and EMODnet-bathymetry datasets, were used to build Cetacean Feeding Models (CFMs) for the target species using Random Forest and RUSBoost algorithms. Multiple subsets of environmental predictors—physiographic, physical, inorganic, and bio-chemical—were employed to develop and evaluate ML models tailored to feeding prediction. Risso's dolphin resulted to be the best modelled species, with the bio-chemical model based on the RUSBoost algorithm achieving a Balanced Classification Rate (BCR) of 94 %, primarily influenced by 3D chlorophyll-a concentrations, a close proxy for prey availability. The second-best model was the physical one for the common bottlenose dolphin with a BCR of 72 %, influenced by salinity, currents speed, and temperature. These differences in predictive performance might reflect the distinct trophic niches of the studied odontocetes. Finally, simulated predictive maps of Risso's dolphin feeding habitats for summer months were realized in the Gulf of Taranto, providing actionable insights for conservation and sustainable management. The developed CFMs enhance understanding of cetacean feeding preferences and offer a versatile framework for integrating behavioural processes into species distribution models to inform area-based conservation measures, with significant potential for application across other Mediterranean areas.

1. Introduction

A central theme in cetacean conservation is the need to understand how they utilize their environment and how they are affected by its changes over time (Nowacek et al., 2016). For years this has been the crucial point in the challenge of managing and conserving these species. The difficulty of studying the behaviour of these elusive and wide-ranging species in marine ecosystems is due to their significant temporal and spatial flexibility in the habitat use, migratory patterns for foraging and reproduction purposes, and to their complex social interactions (King and Jensen, 2023; Schorr et al., 2022). Therefore, this issue has been approached in different ways. One of this is aimed to understand how to discriminate and characterize the different behaviours and, specifically, feeding, which directly influence cetacean daily survival, reproduction and fitness. Cetaceans have high metabolic demands and presumably move by following the movements of prey, which are known to influence their habitat selection, alongside environmental conditions in patchy environments (Riekkola et al., 2019). Additionally, the feeding is the activity most likely to conflict with anthropogenic activities, such as fishing and aquaculture, leading to

---

* Corresponding author at: Institute of Intelligent Industrial Technologies and Systems for Advanced Manufacturing, CNR-STIIMA, Bari, Italy.
E-mail address: rosalia.maglietta@cnr.it (R. Maglietta).

https://doi.org/10.1016/j.ecoinf.2025.103066
Received 10 July 2024; Received in revised form 5 February 2025; Accepted 5 February 2025

potential negative impacts on cetacean populations and direct or indirect competition with fisheries (Bonizzoni et al., 2022; Methion and Díaz López, 2019). Hence, gaining more comprehension into the environmental drivers that shape the feeding behaviour of these top predators is highly valuable for two main reasons. The first is related to the pivotal ecological role they play, delivering many ecosystem services and contributing to the stability and resiliency of marine ecosystem structure (Kiszka et al., 2022; Roman et al., 2014); the second one is linked to aiding in the prioritization of specific areas for conservation in marine regions heavily exploited by intensive human activities (Carlucci et al., 2021a, 2022; Ricci et al., 2021, 2023). To date, the study of feeding behaviour in cetaceans has been approached through a multifaceted lens, reflecting the complexity of their ecological interactions and the challenges posed by obtaining direct observations of their movement and behaviour. The ethological method, supported by descriptive statistics, is a cornerstone of behavioural ecology, offering critical erspective into species' habitat use and ecological needs through direct observations of free-ranging animals (Altmann, 1974; Mann, 1999). This methodology has been instrumental in advancing our understanding of the foraging ecology of cetaceans, both in the Mediterranean Sea and in other regions. In the Mediterranean basin, it has made it possible to study the spatial and temporal foraging patterns of coastal species like common bottlenose dolphins (i.e., Blasi et al., 2015; Pace and Pedrazzi, 2024) and the environmental requirements of fin whale feeding grounds (i.e., Arcangeli et al., 2014); but also to quantify the diving and feeding behaviours of deep-divers as sperm whales (i.e., Drouot et al., 2004). Beyond the Mediterranean region, it has been used, for example, to investigate the feeding habits of odontocetes like killer whales in the northern Pacific (i.e., Van Cise et al., 2024). In recent years, advances in bioacoustics and biologging technologies have transformed the study of cetacean feeding behaviour by enabling the collection of fine-scale, high-resolution data previously unattainable. Bioacoustics has been widely used to explore cetaceans prey detection and foraging strategies (i.e. Carlucci et al., 2024; Cipriano et al., 2022a; Sol et al., 2024), while biologging devices equipped with different sensors (i.e. accelerometers, depth, bioacoustics ones) have allowed continuous tracking of movements and behaviours, providing detailed information into foraging events, prey capture attempt and energy expenditure (i.e. Clayton et al., 2023; Pérez-Jorge et al., 2023). However, despite their effectiveness, these technologies come with limitations, as they are often expensive, and in the case of biologging, the deployment on cetaceans is invasive and raise ethical concerns. On the other hand, Species Distribution Models (SDMs) serve as powerful tools in behavioural ecology for indirectly providing insights into cetacean habitat suitability, including areas likely to support feeding activities (Pasanisi et al., 2024). This methodology, indeed, relies primarily on species sightings and it offers a non-invasive approach to studying the species distribution and their environmental requirements. By linking field observations to explanatory environmental descriptors through statistically derived response functions, SDMs have proven effective in capturing patterns of habitat suitability and ecological dynamics (Claro et al., 2020; Guisan et al., 2013; Menegotto and Rangel, 2018; von Schuckmann et al., 2018). However, challenges related to data resolution and uncertainty remain, as direct behavioural observations can offer critical context for a more comprehensive understanding of habitat use. In this context, Machine learning (ML) has emerged as a cutting-edge tool in ecological research, offering the power to analyse complex and large-scale datasets, collected using potentially all the above mentioned methodologies. In behavioural ecology, ML models have been increasingly employed to classify animal behaviours based on movement data from accelerometers and satellite telemetry (Chambault et al., 2021; Koger et al., 2023; Leoni et al., 2020; Maglietta et al., 2024; Ngô et al., 2021; Wijeyakulasuriya et al., 2020). Similarly, in SDMs, ML-based approaches have been integrated to enhance the predictive accuracy of habitat suitability, leveraging complex environmental datasets to identify cetacean environmental requirements (Carlucci et al., 2018b; Maglietta et al., 2023b; Arcangeli et al., 2024; for an update review: Pasanisi et al., 2024). Despite the plethora of advanced methodological approaches aimed at studying cetacean feeding from various perspectives, many questions regarding the feeding ecology of odontocetes remain unresolved, especially in the Mediterranean basin. In fact, the application of ML to classify behaviour employing environmental predictors remains relatively underrepresented in the literature (Bergen et al., 2023; Carrasco et al., 2021). In particular, to the best of current knowledge, there is a gap in the utilization of ML models applied to the ecological study of cetacean feeding behaviour, which indirectly (Cazau et al., 2021) or directly (Cherubini et al., 2023; Stredulinsky et al., 2023) predict the suitable feeding habitat of the target species.

To address these gaps, the current study proposes a complementary strategy, integrating the spatial approach of SDMs, the computational and predictive capabilities of ML, and the quantitative methods of behavioural science to dissect the environmental drivers of cetacean feeding activity (see Fig. 1). Therefore, this research aimed to design, develop and validate a pioneer framework of Cetacean Feeding Models (CFMs), namely ML-based models tailored to classify cetacean feeding activity. Specifically, Random Forest and RUSBoost algorithms were built to investigate the role of environmental variables as predictors of the feeding behaviour in three odontocete species sighted in the Gulf of Taranto (Northern Ionian Sea, Central-eastern Mediterranean Sea). This region is a key study area due to intense anthropogenic activities, including fisheries, shipping, and military operations, which overlap with critical cetacean habitats that still lack adequate management or conservation measures (Carlucci et al., 2021b). The target species—striped dolphin (*Stenella coeruleoalba*), common bottlenose dolphin (*Tursiops truncatus*), and Risso's dolphin (*Grampus griseus*)—are highly exposed to these pressures, with their feeding activities potentially competing with fisheries (Carlucci et al., 2021a, 2022; Ricci et al., 2021, 2023). This approach ultimately aims to address the gap in the literature regarding the feeding habitat suitability of the target odontocetes in a study area where, despite numerous studies on the presence of cetacean species of conservation interest, there remains a notable absence of effective management strategies for these flag species.

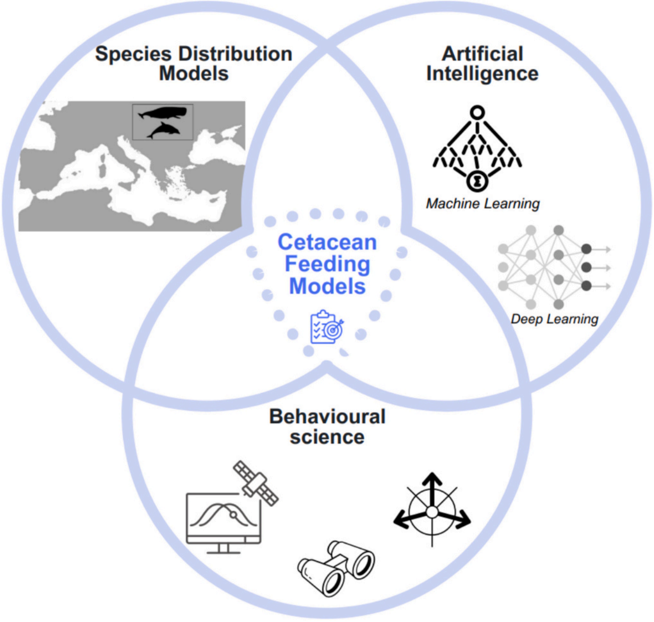

Fig. 1. Diagram elucidating the interdisciplinary nature of the study: the proposed Cetacean Feeding Models framework is derived from the convergence of SDMs, traditional behavioural studies, and the innovative application of Artificial Intelligence.

2. Materials and methods

2.1. Study area

The Gulf of Taranto is the northernmost part of the Ionian Sea in the Central-eastern Mediterranean basin. It spans over 14,000 km^2 from Santa Maria di Leuca to Punta Alice with depths reaching 2000 m (Fig. 2). Its seabed morphology is complex, with an expansive continental shelf that gradually descends toward the continental slope in the eastern sector, and a steep shelf that drops abruptly into deeper waters in the western one. Moreover, an underwater canyon called “Taranto Valley” occurs in the central part of the basin. This articulated morphology system includes several priority habitats in shallow such as seagrass meadows and coralligenous outcrops, and deep-sea areas such as submarine canyons, shoals and cold-water coral banks (Capezzuto et al., 2010; Carlucci et al., 2018a) These features together with the occurrence of upwelling currents ensure favourable conditions for the support of a high biological diversity and provide diverse ecological services (Carlucci et al., 2021a; Carlucci et al., 2022). Specifically, the Gulf of Taranto has been identified as a crucial habitat for three species of odontocetes, the common bottlenose, the striped and Risso’s dolphin (Carlucci et al., 2016, 2018c, 2020; Cipriano et al., 2022b; Maglietta et al., 2018, 2020, 2022, 2023a, 2023b), which are pivotal in controlling the food web dynamics and facilitating trophic cascades (Ricci et al., 2020). However, the Gulf of Taranto faces several anthropogenic pressures such as fishery, marine traffic, habitat degradation and industrialization. The striped dolphin (*S. coeruleoalba*) is the most abundant species in the study area, predominantly pelagic and recently classified by IUCN Red List as Least Concern. It consumes small pelagic and mesopelagic fishes, but also cephalopods (ACCOBAMS, 2021) and although there is no demonstrated interaction, a higher overlap between its trophic niches and trawling has already been estimated. The common bottlenose dolphin (*T. truncatus*) is typically neritic with a high phenotypic plasticity due to the existence of both inshore and offshore ecotypes in the western (Cañadas and Hammond, 2008) and central Mediterranean Sea (Cipriano et al., 2022b). It has been down-listed by the IUCN Red from Vulnerable to Least Concern List (ACCOBAMS, 2021). Also, this odontocete is considered a generalist and feeds on a variety of prey, including fishes, cephalopods and occasionally crustaceans, sharing these preys with trawl and passive nets. Finally, the Risso’s dolphin (*G. griseus*) is a pelagic species preferentially occurring on the continental slope and outer shelf, especially in areas with submarine canyons. Its IUCN conservation status was recently updated from Data Deficient to Endangered (Lanfredi et al., 2021) and it is a teutophagous species with a diet restricted on meso- and benthopelagic cephalopods (Luna et al., 2022; Peda et al., 2015).

2.2. Data description

2.2.1. Behavioural data collection

The collection of behavioural data occurred from April 2016 to October 2023 during standardized vessel-based surveys carried out on a study area of about 960 km^2 in the Gulf of Taranto. The Conventional Line Transect Distance Sampling was applied according to Thomas et al. (2010) and Carlucci et al. (2016). The predominant behavioural activity state was collected using the focal group method with instantaneous scan sampling (Neumann, 2001). Surface behavioural states observed during the surveys were categorized into feeding, resting, socializing and travelling (Carlucci et al., 2018c; Shane, 1990).

The number of observations for each behavioural class and, respectively, for the three studied species, is shown in Table 1. Among the 1410 sightings, 163 were Risso’s dolphin records, 260 of common bottlenose dolphin and 987 of striped dolphin. Since the focus of this study is on feeding activity state, the remaining class were grouped into a macro-category renamed as *Other*. Indeed, the ML models were specifically built to accomplish the task of binary classification between the *Feeding* class and the *Other* class. Details on the monthly distribution of the behavioural sightings data for each species are shown in Table S1.

2.2.2. Environmental predictors

It is now well-established knowledge that there is a strong relationship between cetacean distribution and several environmental descriptors that are hypothesized drivers of their habitat selection. Behavioural observations were matched to dynamic and static environmental variables, listed in Table 2.

The temporal resolution of the variables is daily, with extraction of the environmental predictors corresponding to the day and coordinates of the sightings. The environmental variables used for modelling the cetacean feeding behaviour were selected and processed following the experimental procedure reported by Maglietta et al. (2023b). The predictors classes used in this study were:

- Physiographic variables: these included the distance from the sighting location to the coastline (*Distance_From_Coast*, km) and high-resolution bathymetry data (*Depth*, m). The latter was obtained from the EMODnet dataset (https://emodnet.ec.europa.eu/en/bathymetry#bathymetry-products), while the sighting-to-coastline distance was calculated using geographic coordinates.

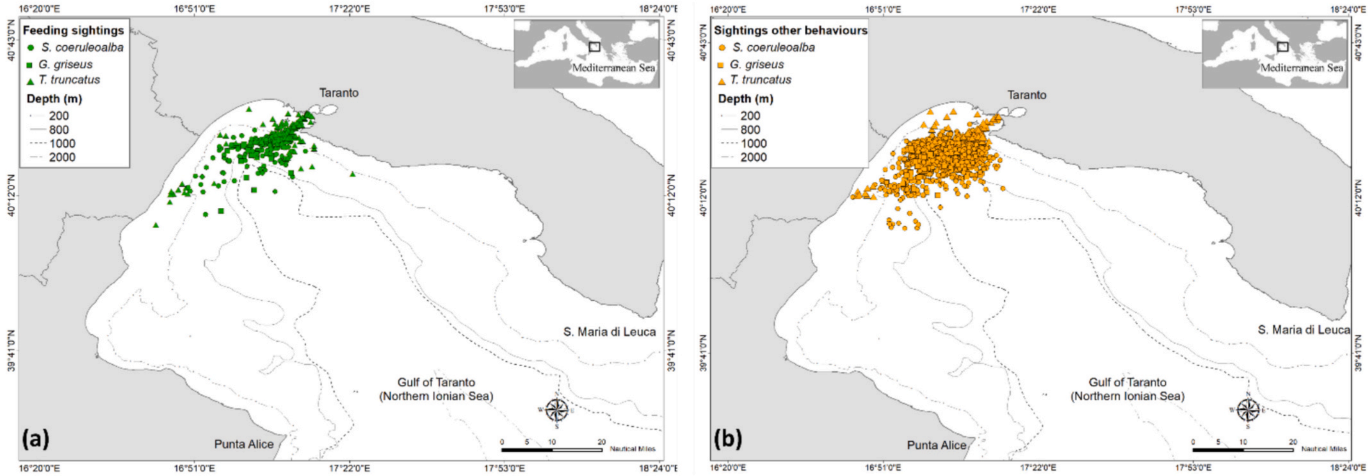

Fig. 2. Maps of the study area with indication of sightings of the three target species during which feeding (a) and other (b) behavioural observations have been recorded from April 2016 to October 2023.

Table 1
The number of sightings for each species recorded during the study period is reported, grouped by activity states in the columns. The last column shows the number of records aggregated under the category *Other*.

	N. of sightings					
Species	Feeding	Resting	Socializing	Travelling	Other (R + S + T)	Total
G. *griseus*	20	44	17	82	143	163
T. *truncatus*	142	10	22	86	118	260
S. *coeruleoalba*	137	125	207	518	850	987
Total	299	179	246	686		1410

Table 2
List and main characteristics (description, source, resolution, variable name) of the predictor variables selected in this study and distinguished into 4 classes. The exact number of the features used to model the feeding behaviour of the three target cetacean species in the Gulf of Taranto is enumerated in column N, for a total number of twenty features. The * symbol refers to layered variables.

Class of variables	Environmental descriptor	Variable name	N	Source	Resolution
Physiographic variables	Distance sighting-coastline (Km)	Distance_From_Coast	1	Computed	–
	High-resolution bathymetry (m)	Depth	1	EMODnet-bathymetry dataset	ca 4 km
Physical variables	Ocean Temperature (°C)	Temperature*	2	Mediterranean Sea Physics reanalysis provided by CMS	ca 4–5 km
	Salinity (psu)	Salinity*	2		
	Currents speed (m/s)	Currents_intensity*	2		
	Mixed layer depth (m)	Mixed_Layer_Depth	1		
	Depth of the max Squared Brunt–Väisälä frequency (m)	DepthofMaxN2	1	Mediterranean Sea Physics reanalysis provided by CMS - Computed	
Inorganic nutrients	Concentration of nitrate ($mmol/m^3$)	Nitrate*	2	Mediterranean Sea biogeochemical reanalysis delivered by CMS	ca 4–5 km
	Concentration of phosphate ($mmol/m^3$)	Phosphate*	2		
Bio-chemical variables	Primary production ($mg/m^3/day$)	Primary_production*	2	Mediterranean Sea biogeochemical reanalysis delivered by CMS	ca 4–5 km
	Phytoplankton carbon biomass ($mmol/m^3$)	PHYC*	2		
	3D chlorophyll-a (mg/m^3)	CHL3D*	2		

- Physical variables: this class consisted of key oceanographic parameters such as ocean temperature (*Temperature*, °C), salinity (*Salinity*, psu), currents speed (*Currents_intensity*, m/s), mixed layer depth (*Mixed_Layer_Depth*, m) and the depth at which the Brunt–Väisälä frequency squared reaches its maximum (*DepthofMaxN2*, m). These features were derived from the Mediterranean Sea Physics reanalysis produced by CMCC (Clementi et al., 2019a; Clementi et al., 2019b) and delivered by Copernicus Marine Service (CMS, https://marine.copernicus.eu/it), with data resolution at approximately 1/24° (4–5 km).
- Biogeochemical variables: metrics for primary production (*Primary_production*, mg/m^3/day), phytoplankton carbon biomass (*PHYC*, $mmol/m^3$), and 3D chlorophyll-a (*CHL3D*, mg/m^3) were utilized. This macro category included inorganic nutrients like nitrate (*Nitrate*) and phosphate (*Phosphate*) concentrations ($mmol/m^3$). These features were sourced from the Mediterranean Sea biogeochemical reanalysis by OGS (IT) and delivered by CMS (Cossarini et al., 2021). The product at 1/24° horizontal resolution (ca. 4–5 km) is produced using the MedBFM3 model system. To build the ML models, inorganic nutrients were treated as a distinct class.

This comprehensive suite of 20 environmental features underpins the ML models, offering a robust framework for understanding cetacean habitat preferences for feeding behaviour. Moreover, for temperature, salinity, currents speed, concentration of nitrate and phosphate, primary production, phytoplankton carbon biomass and 3D chlorophyll-a (referred to as ‘layered variables’ in Table 2) data were extracted at multiple depths up to 200 m (i.e. at the surface and at 10 m, 20 m, 30 m, 40 m, 50 m, 100 m, and 200 m) to capture vertical environmental gradients. The data were then averaged over specified depth intervals to create two representative features: one for the upper 40 m (top layer) and another for the 50 to 200 m depth range (bottom layer). The variables have been limited up to 200 m of depth because of the stability and low variability of the water column below. This dual-tiered approach provided a comprehensive view of the marine environment, essential for modelling cetacean behaviour near the surface and at depth. Then, a sea-overland extrapolation procedure (De Dominicis et al., 2014) was used to prevent the presence of missing values interpolating the oceanic fields over each cetacean sightings record.

2.3. Machine learning models

The ML models were implemented in MATLAB (Inc., T. M, 2023a, 2023b, 2023c) using the Statistics and Machine Learning ToolBox (Inc., T. M, 2023a, 2023b, 2023c). Random Undersampling boosting (RUSBoost) (Seiffert et al., 2009) is a variant of the AdaBoost algorithm (Wang, 2012). It is effective for dealing with imbalanced datasets where one class significantly outnumbers the other. In such datasets, one class (e.g., *Feeding*) is significantly smaller than the other (e.g., *Other* behaviours), which is common in ecological studies. To address this imbalance, RUSBoost reduces the larger class by randomly undersampling it to match the size of the smaller class. Specifically, the RUS technique takes the smallest class size within the training dataset, denoted by N, as the foundational benchmark for sampling procedures. Classes exceeding this size are subjected to a reduction process whereby only N instances from each class are selected for analysis. Consequently, for every constituent weak learner that forms part of the ensemble, the RUSBoost algorithm systematically extracts a stratified sample encompassing N instances from each of the classes. Then, the AdaBoost algorithm builds an ensemble of weak learners, where each learner focuses on the mistakes made by the previous ones, thereby improving the overall accuracy of the model. RUSBoost models were used to absolve a binary classification task using environmental predictors, namely to predict the *Feeding* behaviour of a target cetacean species against the other type of behaviours, pooled together in the class *Other*. The models have been built using the *fitcensemble* function (Inc., T. M, 2023a), which allows automatic optimization of hyperparameters to be carried out. The optimizable hyperparameters for RUSBoost are the maximum number of splits (MaxNumSplits), the number of ensemble learning cycles (NumLearningCycles) and the learning rate (LearnRate).

Random Forest (RF) is a powerful tool for ecological studies because it can handle complex environmental data with many variables, for example useful to model habitat suitability based on various factors (Kaveh et al., 2023; Maglietta et al., 2023b; Pasanisi et al., 2024). By combining many decision trees, it provides robust predictions, even when data is noisy, making it particularly useful for analyzing ecological patterns. In particular, RF is an ensemble learning technique that combines multiple decorrelated decision trees to make predictions for categorical variables, when it is used for classification tasks (Breiman, 2001; Cutler et al., 2012). Each decision tree is grown independently from a distinct subset of data and features, obtained by randomly sampling the dataset with replacement (i.e. feature bagging). Each tree in RF operates independently, making a decision based on its subset, which yields the ensemble prediction determined by the majority voting process. As well as for the RUSBoost approach, RF models were built using the *fitcensemble* function and the automatic hyperparameter optimization. The optimizable hyperparameters included MaxNumSplits, NumLearningCycles, the minimum leaf size (MinLeafSize) and number of predictors to sample (NumVariablesToSample).

2.4. Experimental framework for cetacean feeding models

The Machine Learning experimental process designed and developed in this study is illustrated in Fig. 3. It is structured into three main phases, each one focusing respectively on data processing, the construction of ML models and model evaluation, with the selection of the optimal model followed by an emphasis on its ecological interpretability. This experimental process has been applied to the three cetacean species under investigation.

2.4.1. Datasets generation via data splitting

In this initial phase, the original dataset, named *D*, consists of 20 predictor features, listed in Table 2. It has been manipulated to obtain four subsets of data, each containing only one class of environmental variables, tailored for building the ML models. The datasets generated are: the *D_phg* dataset made up of the 2 physiographic features (distance sighting-coastline and high-resolution bathymetry); the *D_phy* made up of the 8 physical features (temperature, salinity and currents speed, each with both top and bottom layers, mixed layer depth and the depth at which the Brunt–Väisälä frequency squared reaches its maximum); the *D_ino* consisting of the 4 inorganic features (nitrate and phosphate each with both top and bottom layers) and the *D_bio* dataset, with the 6 biochemical features (primary production, phytoplankton carbon biomass,3D chlorophyll-a, each with both top and bottom layers). Five ML models were built on each of these datasets and on the complete dataset, respectively, using the methodology described in the following section.

2.4.2. Building and evaluation of the CFMs

Both RUSBoost and Random Forest algorithms employed automated hyperparameter optimization and cross-validation techniques, according to the experimental methodology illustrated in Fig. 4. Automated hyperparameter optimization helps the model automatically find the best settings for its parameters, which improves its accuracy without the need for manual tuning. The input dataset was split into training and test set and hyperparameter tuning was automatically carried out on the training set using the Bayesian Optimization Method (BOM) in parallel (Eggensperger et al., 2013; Snoek et al., 2012). The BOM determines the next hyperparameter value for evaluating the actual objective function based on the previous results of tested hyperparameter values, which avoids numerous unnecessary evaluations. Moreover, the cross-validation techniques were applied to ensure robust model evaluation, minimizing the risk of overfitting and providing more reliable generalization to unseen data. In particular, the stratified k-fold cross-validation method was employed (Refaeilzadeh et al., 2016). Through this process, we identified the hyperparameter configurations that produced the highest level of model accuracy. Next, we applied the model, tuned with these selected hyperparameters, on the independent test data set to evaluate the classifier predictive performances.

More specifically, the performance of the model was evaluated using a set of quantitative classification metrics (Vujović, 2021). The first metric is the accuracy, and it quantifies the overall proportion of correct predictions out of all predictions made:

$$accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}$$

Then, the sensitivity measures the proportion of actual positives correctly identified by the model to the total number of positive samples:

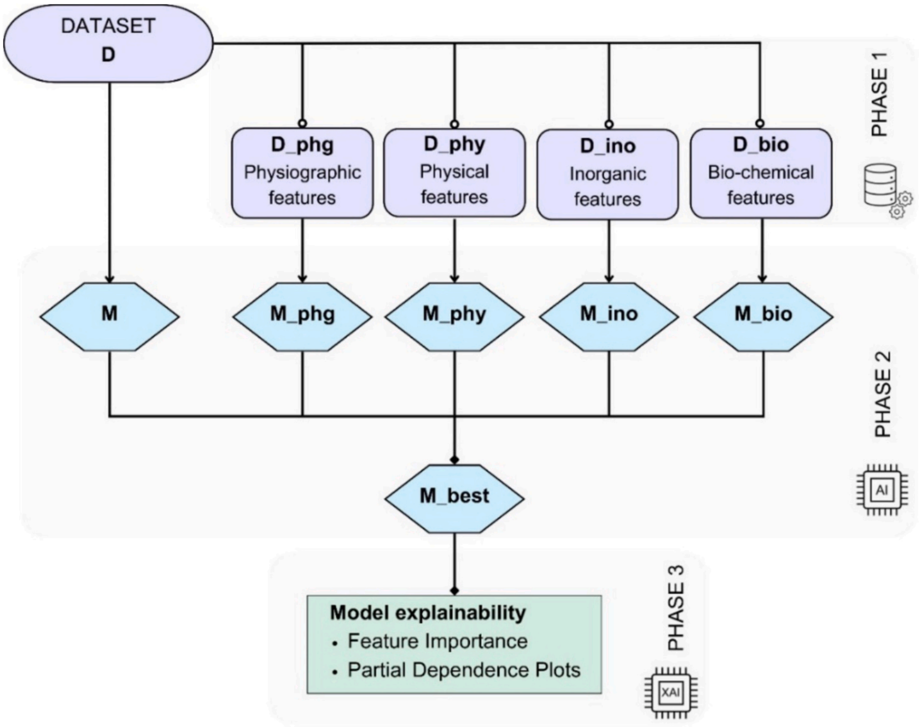

Fig. 3. The flowchart outlines the three-phase experimental process used in the study.

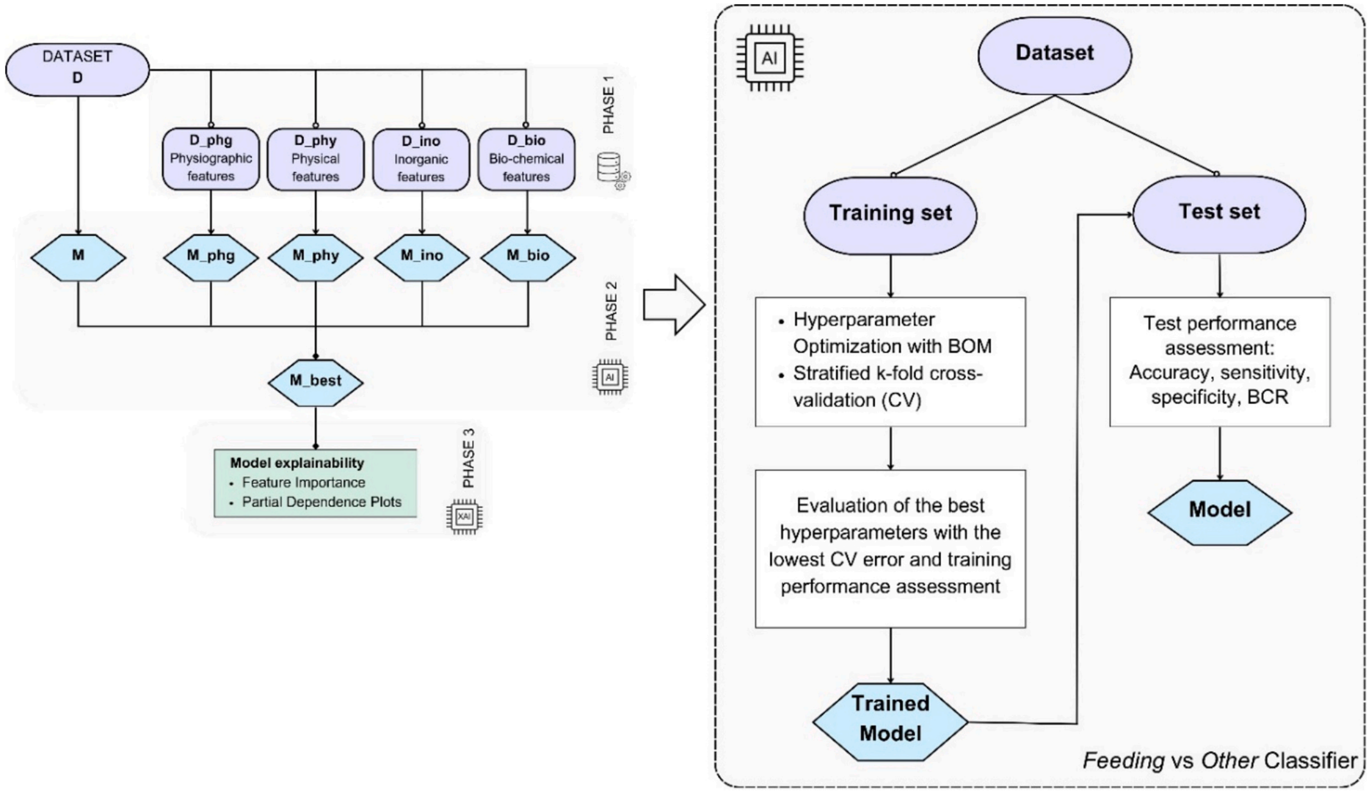

Fig. 4. Flowchart of the experimental process used for training and testing the ML models to predict the *Feeding* behaviour for each of the three cetacean species studied. It delves into the details of the Phase 2 shown in Fig. 2.

$$sensitivity = \frac{TP}{TP + FN}$$

On the other hand, specificity assesses the proportion of actual negatives that are correctly identified by the model to the total number of negative samples:

$$specificity = \frac{TN}{TN + FP}$$

However, when working with unbalanced datasets it is more appropriate to consider more suitable metrics such as Balanced Classification Rate (BCR). BCR is calculated as the average of sensitivity and specificity:

$$\mathrm{BCR} = \frac{\mathrm{sensitivity} + \mathrm{specificity}}{2}$$

By balancing the contribution of sensitivity and specificity, BCR provides a more comprehensive assessment of a model's effectiveness across all classes.

2.4.3. Models interpretability

One of the most debated aspects of Machine Learning models applied in the ecological field concerns their interpretability. Domain experts are no longer just responsible for developing algorithms to optimize model performance. Rather, their research is increasingly moving toward understanding the mechanisms by which an algorithm generates a specific output (Lucas, 2020). This type of information constitutes the essence of ML-based ecological modelling, as it enables comprehensible ecological understanding to be obtained. Hence, this knowledge becomes crucial in enhancing management and conservation measures for vulnerable species. In particular, in the context of species distribution models, explainable artificial intelligence (XAI) facilitates improved ecological interpretability of otherwise 'black-box' models (Castelvecchi, 2016).

The Feature Importance was evaluated using the best-performing models. This technique computes a score for each predictor variable based on its contribution to the model learning process (Saarela and Jauhiainen, 2021). The variables are ordered according to the Feature Importance scores and the median of these scores was calculated and used as a threshold (Carrasco et al., 2021). Features surpassing this threshold were selected for Partial Dependence Plots (PDPs) (Goldstein et al., 2015), a valuable tool for enhancing the model's interpretability by examining the relationships between predictor variables and the response variable. PDPs offer insight into whether these relationships are linear or more complex, thereby providing an understanding of how the evaluated feature affects the probability of predicting feeding behaviour. In brief, the PDPs provide a visual representation of a variable's effect on the model's average prediction, while also considering the distribution of all other variables included in the model (i.e., marginalizing with respect to the influence of the other variables).

3. Results

3.1. Environmental characterization of the feeding behaviour

A two-sided two-sample *t*-test (Drummond and Tom, 2011) was utilized to investigate the presence of statistically significant differences between the two classes of target behaviours, *Feeding* and *Other*, in relation to the values of each of the 20 environmental features. This test evaluated the null hypothesis that the population means are equal against the alternative hypothesis of any difference in the means. The significance level was initially set to 0.05. Given the large number of tests that were performed for each cetacean species, a Bonferroni correction was applied to adjust the significance threshold accordingly. The significant outcomes of this analysis are shown in the Table 3. In addition, the type of mean comparison (*Feeding* $>$ *Other* or *Feeding* $<$ *Other*) was assessed after confirming that the difference was statistically

Table 3
Results of the comparative analysis between *Feeding* and other observed behaviours (*Other*) for each of the 20 predictive environmental variables are reported across the three target species. Variables of the same class are highlighted with the same colour. The test used is the two-sided two-sample *t*-test, with a Bonferroni correction. Only comparisons with a *p*-value less than the adjusted significance level are reported in the table.

Species	Mean comparison	Class of variable	Features	*p*-values
G. griseus	*Feeding* < *Other*	Physical	*Temperature_Top*	≤0.001
	Feeding > *Other*	Physical	*Currents_Intensity_Top*	≤0.001
		Physical	*Currents_Intensity_Bottom*	≤0.001
		Bio-chemical	*Primary_Production_Top*	≤0.001
		Bio-chemical	*Phyc_Top*	≤0.001
		Bio-chemical	*Phyc_Bottom*	≤0.001
		Bio-chemical	*CHL3D_Top*	≤0.001
		Bio-chemical	*CHL3D_Bottom*	≤0.001
T. truncatus	*Feeding* < *Other*	Physical	*Currents_Intensity_Top*	≤0.001
		Physical	*Currents_Intensity_Bottom*	≤0.001
		Inorganic	*Nitrate_Top*	≤0.001
		Inorganic	*Nitrate_Bottom*	≤0.001
	Feeding > *Other*	Physical	*Salinity_Top*	≤0.001
		Physical	*Salinity_Bottom*	≤0.001
S. coeruleoalba	*Feeding* < *Other*	Physiographic	*Distance_From_Coast*	≤0.001
	Feeding > *Other*	Inorganic	*Nitrate_Top*	≤0.001

significant and reported in the table.

The feeding of the Risso's dolphin was significantly associated with a total of five bio-chemical and three physical features. In particular, among these statistically significant variables, the feeding occurred at lower values of the temperature at top layer, compared to other behaviours. In addition, primary production at the top and, respectively, the currents speed, phytoplankton carbon biomass and 3D chlorophyll-a at both top and bottom layers exhibited higher mean values for feeding than for other behaviours. The feeding behaviour of the common bottlenose dolphin in the study area appeared to be significantly associated with a total of four physical and two inorganic features. In particular, feeding behaviour occurred at lower mean values for currents speed and nitrate concentration at both top and bottom layers. Instead, feeding occurred at higher mean values of salinity at top and bottom layers. Finally, the feeding behaviour of the striped dolphin was associated with a lower mean value of the variable distance from the coast, while feeding was associated with higher mean values of the nitrate concentration at the top. In total, one physiographic and one inorganic feature were found to be statistically significant based on two-sided two-sample t-test.

3.2. Exploring feeding activity with CFMs

For each odontocetes species, the procedure described in section 2.5.2 was employed to construct five different CFMs. Among the investigated species, only *G. griseus* and *S. coeruleoalba* had unbalanced behavioural classes distributions. Therefore, only the RUSBoost algorithm was used for these two species because of its well-known ability to handle class imbalance. Conversely, for *T. truncatus*, which did not present this issue, both RUSBoost and Random Forest algorithms were applied and the one achieving the best performance was chosen. In each experiment, the input dataset was split into a training set, collecting 80 % of the data samples and a test set that included the other 20 %. Moreover, to reduce the model's overfitting, k-fold cross-validation with $k = 5$ was integrated during the hyperparameter optimization process. The test performances for all the models are listed in Table 4, while the optimal hyperparameter values, used for model building, along with the training performances of all models are presented in Table S2 and S3.

The datasets used to build ML models for the *G. griseus* presented a total of 163 observations. Among the five built models, the one trained with bio-chemical input variables, *M_bio*, outperformed the others for all classification metrics. Specifically, *M_bio* predicted the feeding behaviour with an accuracy of 88 %, a sensitivity of 87 %, a specificity of 100 % and a BCR of 94 %. However, the model trained on all input features, *M*, also demonstrated high performance, emerging as the second-best model (accuracy of 88 %, a sensitivity of 78 %, a specificity of 100 % and a BCR of 89 %). Conversely, the other three models proved less capable of carrying out the classification task, with lower values of precision metrics. To assess which variables mostly influenced the performance of the RUSBoost classifier in predicting Risso's dolphin feeding behaviour, the Feature Importance analysis on the *M_bio* model was

Table 4
Performance metrics such as accuracy, sensitivity, specificity and BCR of ML models for the *Feeding* classification, trained on five different datasets for each species. For the common bottlenose dolphin only the best model for each dataset is reported in the table, after the comparison between RUSBoost and Random Forest performances. Bold characters are employed to highlight the best model performance for each species.

		Test performance				
Species	Algorithm	Type of Model	Accuracy	Sensitivity	Specificity	BCR
G. griseus	RUSBoost	*M*	0.88	0.78	1	0.89
		M_phg	0.67	0.50	0.70	0.6
		M_phy	0.72	0.75	0.6	0.68
		M_ino	0.79	0.89	0.33	0.61
		M_bio	**0.88**	**0.87**	**1**	**0.94**
T. truncatus	RUSBoost	*M*	0.69	0.71	0.67	0.69
	RF	*M_phg*	0.70	0.70	0.71	0.70
	RUSBoost	***M_phy***	**0.72**	**0.76**	**0.68**	**0.72**
	RF	*M_ino*	0.70	0.76	0.63	0.69
	RUSBoost	*M_bio*	0.58	0.46	0.69	0.58
S. coeruleoalba	RUSBoost	***M***	**0.67**	**0.70**	**0.52**	**0.61**
		M_phg	0.54	0.54	0.52	0.53
		M_phy	0.67	0.71	0.44	0.58
		M_ino	0.66	0.68	0.50	0.59
		M_bio	0.58	0.60	0.48	0.54

conducted. The results of this analysis are depicted in Fig. 5. The main important features were 3D chlorophyll-a at the top, phytoplankton carbon biomass both at the top and bottom layers, and the PDPs for these variables are shown in Fig. 6. In particular, the PDP in Fig. 6a visualizes the relationship between 3D chlorophyll-a at the top and the predicted behaviours of the species, showing a sharp drop of the predicted probability of feeding for 3D chlorophyll-a values less than 0.13 mg/m^3.

For the *T. truncatus*, the best-performing model was the RUSBoost trained with only the input features from the physical dataset, the *M_phy*. It predicted *Feeding* with an accuracy of 72 %, a sensitivity of 76 % and a specificity of 68 %, achieving a BCR of 72 %. On the other hand, the other two models that exhibited slightly lower performance were the RF models trained on physiographic (*M_phg*) and inorganic nutrients (*M_ino*) and the RUSBoost trained on the full dataset (*M*) with BCR of 70 % the first and 69 % both the others. Lastly, the model built only on biological variables (*M_bio)* did not achieve satisfactory performance, as their BCR metrics range between 51 % and 58 %. Results of the Feature Importance are reported in Fig. 7. The features that exceeded the median threshold were salinity both at the top and bottom layers, currents speed at the top, temperature at the bottom and the PDPs for these variables are shown in Fig. 8. Finally, the PDP realized for salinity at the bottom (see Fig. 8a) showed stable feeding probabilities across a range of values greater than 38.87 psu and then dropping sharply, suggesting a threshold effect for this value. The PDP for salinity at the top (see Fig. 8b) displayed more variability in the feeding prediction probabilities, but the same drop around 38.7 psu.

Regarding *S. coeruleoalba*, the model that performed best for this species is the *M*, with an accuracy of 67 %, a sensitivity of 70 %, a specificity of 52 %, and a BCR of 61 %. As for the other four developed models, the overall test performances were found to be significantly lower. The accuracy of these models ranged from a minimum of 54 % for the model trained only on physiographic variables (*M_phg*) to a maximum of 67 % for the model built only on physical variables (*M_phy*). However, the BCR varied within a range from 53 % for the *M_phg* model to 59 % for the model trained with only inorganic variables (*M_ino*). Lastly, the best model Feature Importance was computed and showed in Fig. 9. The median value of the predictor importance score was computed for the top 10 features with respect to the total 20 features used to build the model and it was set as a threshold. The features that achieved a score higher than this threshold, were distance from the coast, depth, currents speed at the top, salinity at the top, temperature at the top, and the PDPs for these variables are shown in Fig. 10. Focusing on the PDP realized for the distance from the coast, the feeding behaviour probability fluctuates significantly with closer proximity to the coast, showing a notable decrease in predicted feeding behaviour as distance increases more than 10 km (see Fig. 10a).

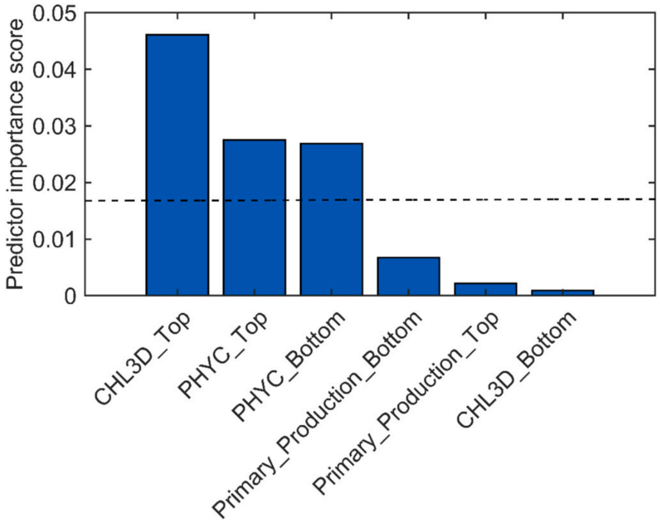

Fig. 5. Feature Importance computed for the RUSBoost of the Risso's dolphin, trained only with bio-chemical features (*M_bio*). The dotted line indicates the threshold of the predictor importance score.

As a final analysis, the best-performing model comparing all species resulted to be the *M_bio* of the *G. griseus*. This model was selected to extend the feeding suitability prediction across the entire Gulf of Taranto (see Fig. 11), using data from both outside and the study area. The simulation was conducted for June, July, and August of 2024. These months were chosen because they correspond to the months in which most of the sighting data of Risso's dolphin were collected (Table S1–1). Daily predictions for each location cell of the area with a resolution of circa 4 km were made and subsequently counting the number of days predicted as feeding days for a given month at each location. A location was designated as a feeding cell only if the count of feeding days surpassed the count of days predicted for other behaviours. The model in June suggested that almost the entire area could be considered a potentially suitable feeding habitat. However, in July, the model predicted a gradual decrease in the extent of suitable feeding habitats, although these areas remained predominant compared to those suitable for other behaviours. For the month of August, the model predicted that almost the entire Gulf of Taranto would be a non-feeding area. These results should be interpreted with caution, as the model was trained on data from a specific area, and its application to regions outside of this area involves extrapolation. While it is reasonable to assume that there may be a relationship between the study area and the surrounding zones, the reliability of the predictions in these extrapolated regions remains uncertain. Future data, ideally collected from an expanded study area and in the 2024 year, will be crucial for further validating the model and assessing its robustness in broader contexts and in forecasting application.

4. Discussion and conclusions

This study develops and validates a novel ML-based modelling framework, the Cetacean Feeding Models (CFM), designed to investigate feeding activity and predict suitable habitats for small odontocetes in the Northern Ionian Sea (Central-eastern Mediterranean Sea). This represents, for the best of current knowledge, the first approach which integrate environmentally mediated behavioural processes into a spatial modelling strategy, enhanced by the computational power of Artificial Intelligence (see Fig. 1). Unlike traditional SDMs, which typically relate species occurrences to environmental descriptors, CFMs directly incorporate behavioural observations as the model response variable, providing a more behaviour-specific and ecologically meaningful perspective. This method aims to fill the gap of previous studies on SDMs (e.g., Melo-Merino et al., 2020; Pasanisi et al., 2024) by explicitly linking environmental predictors to cetacean behavioural states, specifically the feeding one. Cherubini et al. (2023) reported a preliminary approach to this issue, using both environmental variables and group size to model the behaviours of 3 odontocetes species, pulled together.

4.1. Ecological insights on feeding behaviour from the statistical analysis and CFMs

The Risso's dolphin resulted to be the best modelled species (see Table 4**)**. Its feeding behaviour was significantly associated with three physical and five bio-chemical variables, as resulted from the two-sided two-sample *t*-test (see Table 3). Among these statistically significant variables, the feeding occurred at higher concentrations of 3D chlorophyll-a, phytoplankton carbon biomass and primary productivity. Therefore, the bio-chemical model (*M_bio*) proved to be the best-performing ML-model for this species, with 3D chlorophyll-a as the most influential predictor, followed by phytoplankton carbon biomass, as shown in Fig. 5. These findings are consistent with previous research in the Pacific region, where chlorophyll-a was identified as a key driver

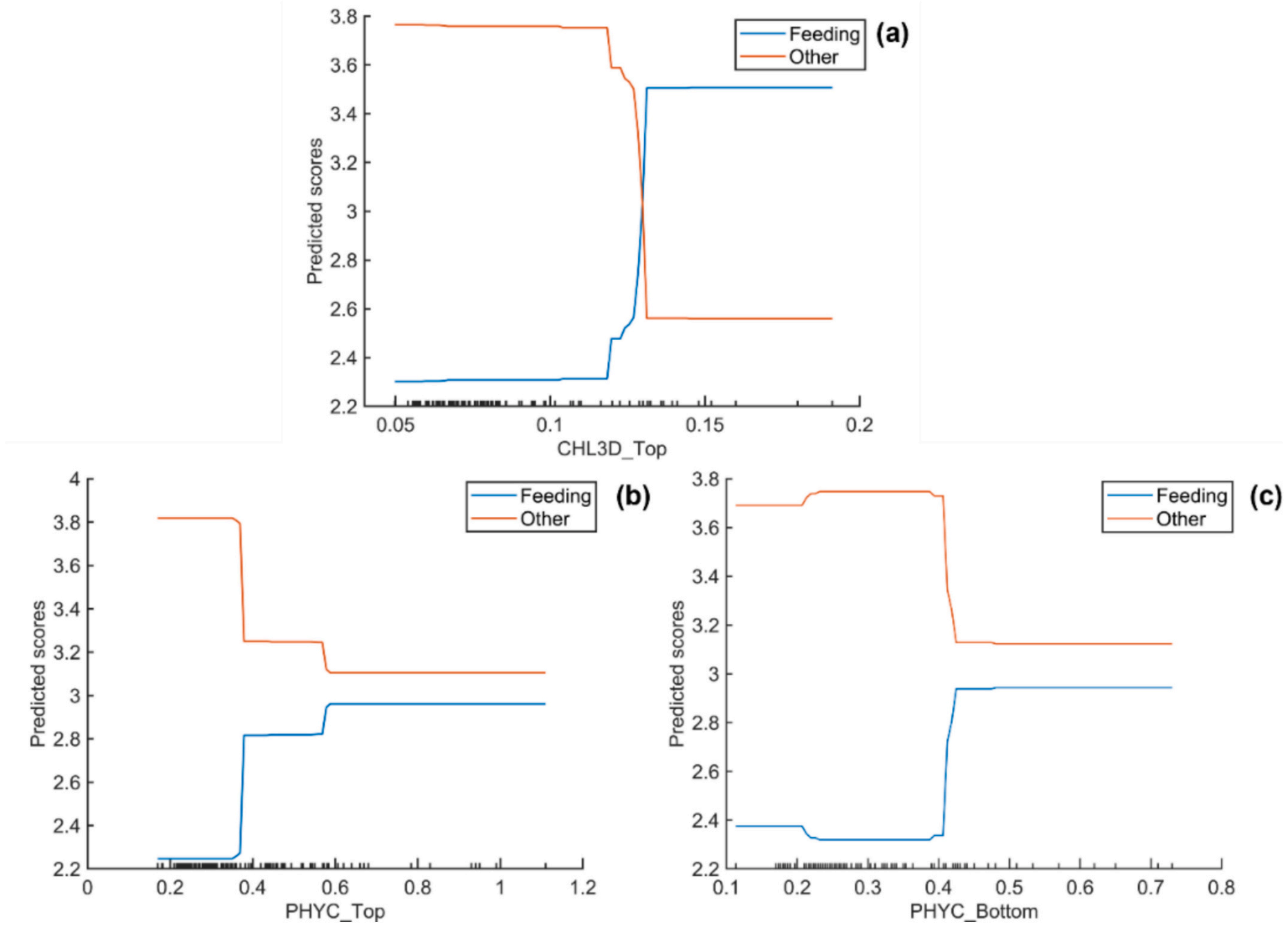

Fig. 6. Partial dependence plot from the RUSBoost *M_bio* showing the influence of the 3D chlorophyll-a at the top (a), the phytoplankton carbon biomass both at the top (b) and bottom (c) layers on the predicted probability of *Feeding* versus *Other* behaviours in Risso's dolphin.

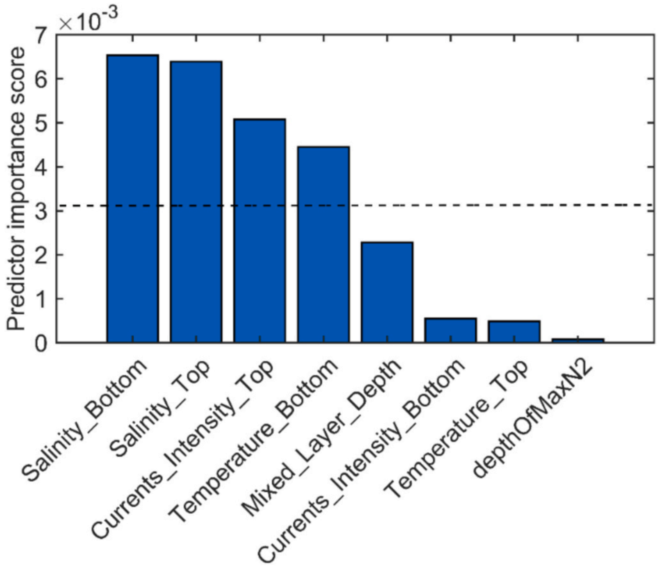

Fig. 7. Feature Importance computed for the RUSBoost of the common bottlenose dolphin, trained only with physical features (*M_phy*). The dotted line indicates the threshold of the predictor importance score.

of feeding habitat for this species (Soldevilla et al., 2011) and this variable has proven to be a significant predictor for improving the performance of SDMs in predicting the distribution of the species (Fiedler et al., 2023). In the Mediterranean basin, a major knowledge gap remains on the feeding ecology of Risso's dolphin, however some studies have demonstrated correlations between its encounter rates and peak chlorophyll-a concentrations (i.e. Azzellino et al., 2017) and others proved this variable to be one of the main factors influencing the species habitat selection (i.e. Arcangeli et al., 2024). This reliance on biochemical variables makes these predictors highly effective for modelling feeding occurrences of deep-diving predators that predominantly feeds on cephalopods and squids, linking its feeding behaviour to lower trophic-level productivity (Luna et al., 2022; Ricci et al., 2021). Specifically, the partial dependence plot for the *M_bio* model demonstrated a sharp increase in the probability of feeding at 3D chlorophyll-a concentrations above a specific threshold (see Fig. 6). This favourable trophic condition verified in the deep waters in the study area where the anticyclonic gyre occurs (Pinardi et al., 2016) and the presence of the submarine canyon enhances nutrient upwelling and primary productivity, supporting aggregation of prey and top predators (Fernandez-Arcaya et al., 2017). Evidence of such productivity-related patterns and dynamics might be found in the predictive maps represented in Fig. 11, where feeding suitability was simulated for the entire Gulf of Taranto for the summer months of 2024. The maps reveal that the canyon system stands out as the only area consistently predicted as suitable for feeding also for the month of July. This result supports the hypothesis that the canyon plays a crucial role in sustaining prey availability for Risso's

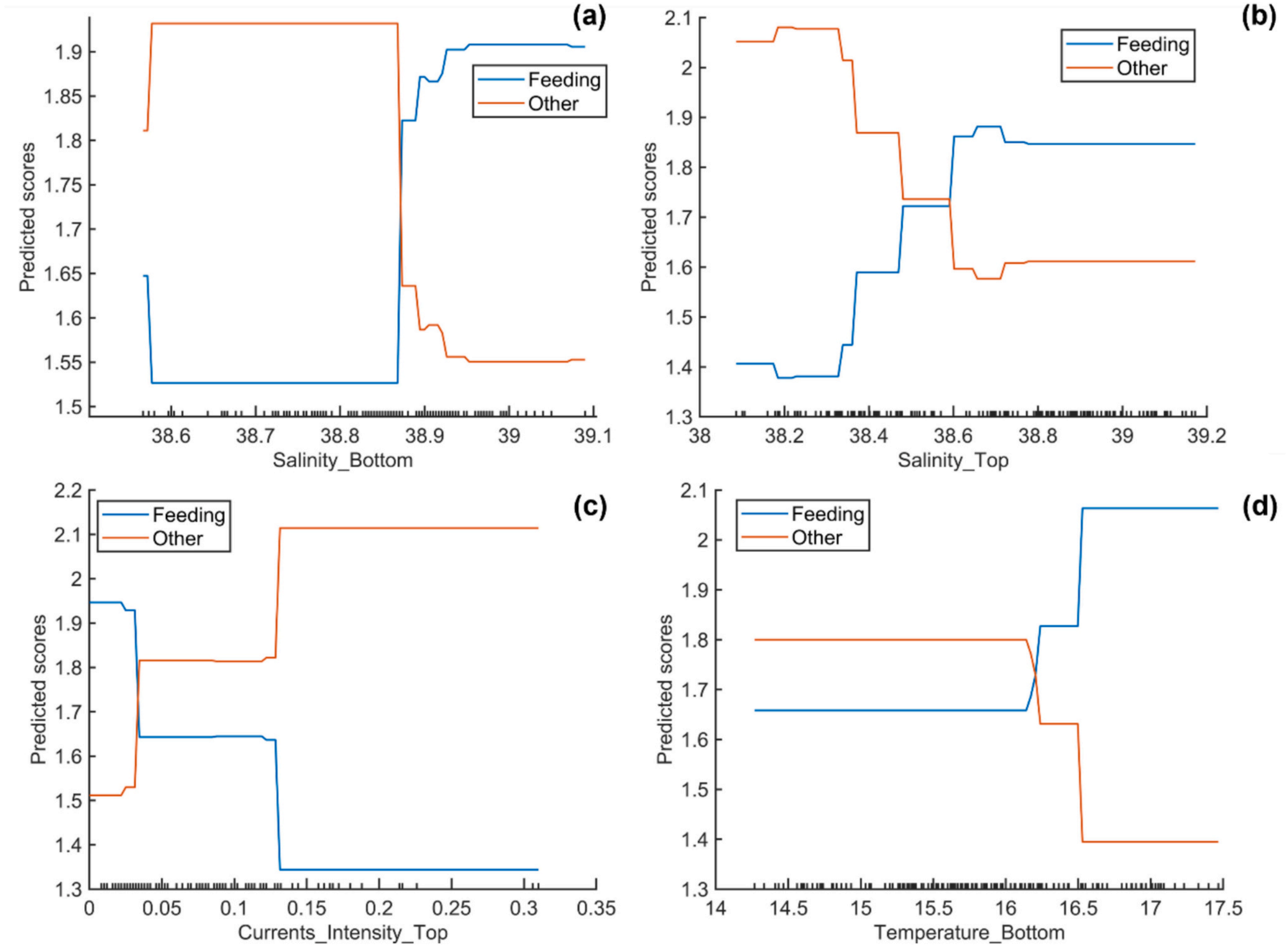

Fig. 8. Partial dependence plots from the RUSBoost *M_phy* showing the influence of salinity at the bottom (a) and top (b) layers, currents speed at the top (c) and temperature at the bottom (d) on the predicted probability of *Feeding* versus *Other* behaviours in common bottlenose dolphin.

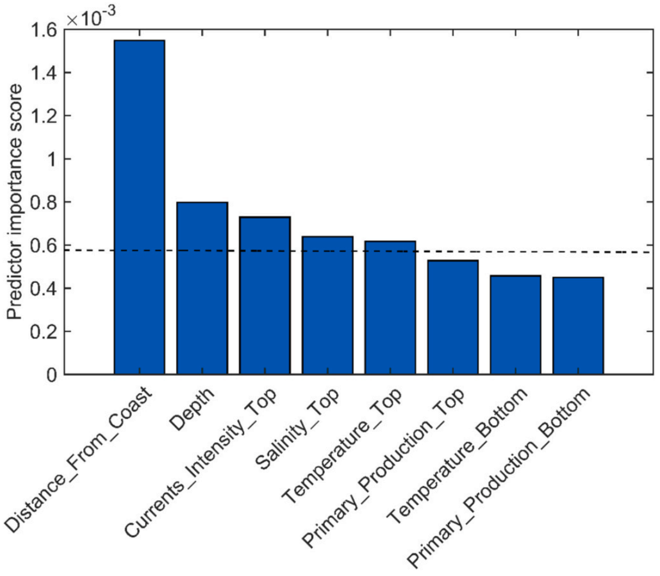

Fig. 9. Feature Importance computed for the RUSBoost of the striped dolphin, trained with the complete dataset (*M*). The dotted line indicates the threshold of the predictor importance score.

dolphins and other deep-diving predators, aligning with findings by Ricci et al. (2019), which emphasize the role of shelf break zones for this area marine food web. However, caution is warranted in interpreting these results, as the general extrapolation to regions beyond the study area requires further validation with additional data, so further studies are needed to validate these results.

For common bottlenose dolphins, this study highlights that the feeding behaviour was significantly linked to higher salinity values and lower currents speed and nitrate values then other behaviours (see Table 3). In accordance with the t-test results, the physical model (M_phy) showed the highest performance for this species, with salinity, currents speed at the top and temperature at the bottom identified as the most influential predictors (see Table 4 and Figs. 7,8). These findings are somehow consistent with previous studies in the Gulf of Mexico, where salinity was positively correlated with the encounter rates of *T. truncatus* in coastal bays (Hornsby et al., 2017; Mintzer and Fazioli, 2021) and where it was one of the most important predictors of the species habitat suitability (Pitchford et al., 2016). However, in the Mediterranean basin, this variable does not appear to be an efficient predictor in habitat suitability studies employing SDMs (i.e. Pace et al., 2022). Literature suggests that the distribution of this coastal species and its group size are more often predicted by factors such as chlorophyll-a, sea surface temperature and physiographic variables (Carlucci et al., 2018b; La Manna et al., 2016; La Manna et al., 2023; Maglietta et al., 2023b; Pace et al., 2022). These outcomes may be attributed to two main factors. First, most SDMs studies tend not to include a wide range of variables (i.e. currents speed, mixed layer depth, salinity), as done in this research,

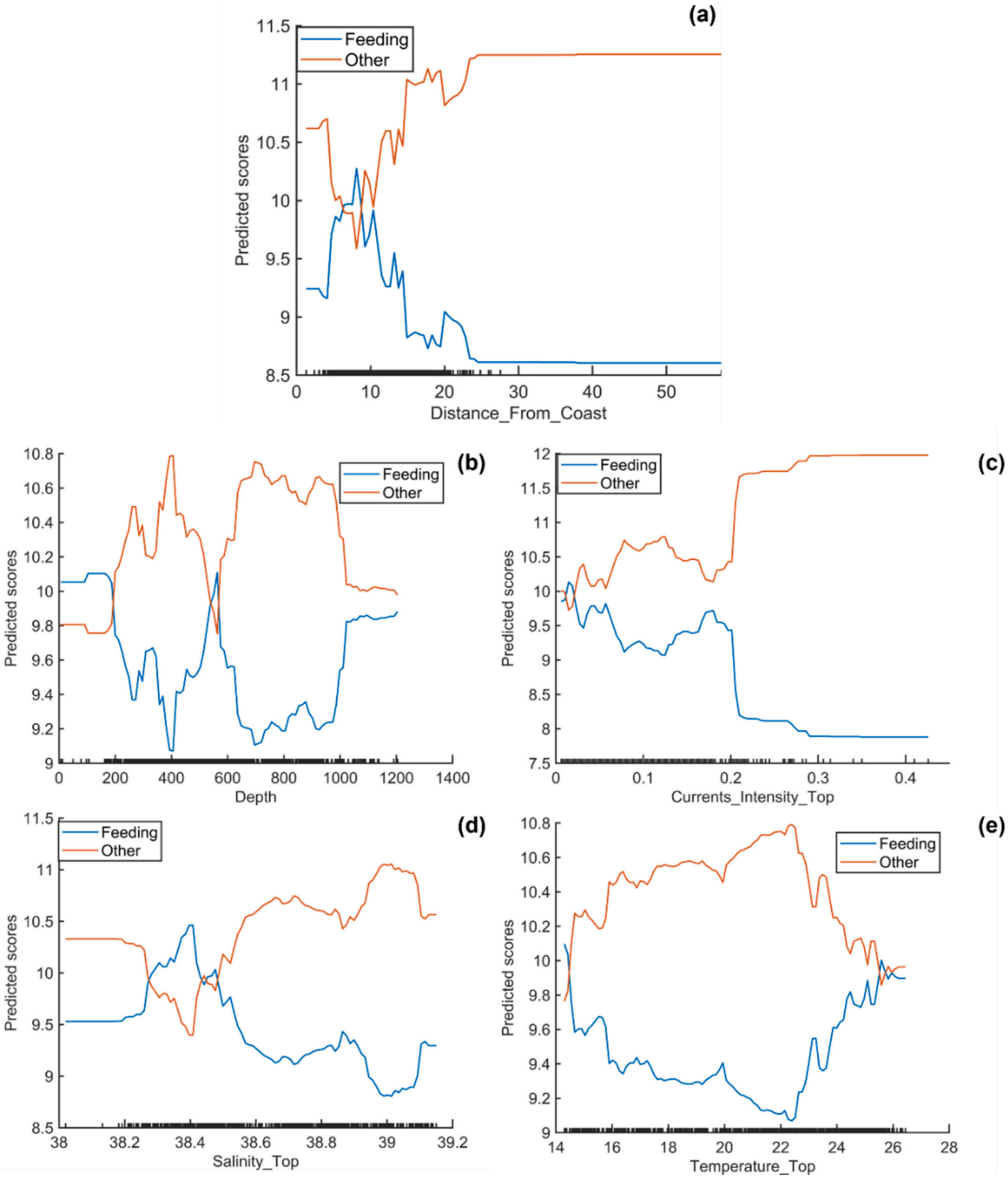

Fig. 10. Partial dependence plot from the RUSBoost *M* showing the influence of distance from the coast (a), depth (b), currents speed at the top (c), salinity at the top (d), temperature at the top (e) on the predicted probability of *Feeding* versus *Other* behaviours in striped dolphin.

resulting in limited research on the effects of these specific environmental predictors. Second, there is a predominant reliance on surface-layer data, overlooking the three-dimensional complexity of the seascape, including depth-specific conditions like temperature and salinity, which are essential for a thorough understanding of habitat use and species behaviour, as underlined by the recent review of Pasanisi et al. (2024). Moreover, compared to the models developed for Risso's dolphin, the CFMs for the common bottlenose dolphin showed moderate

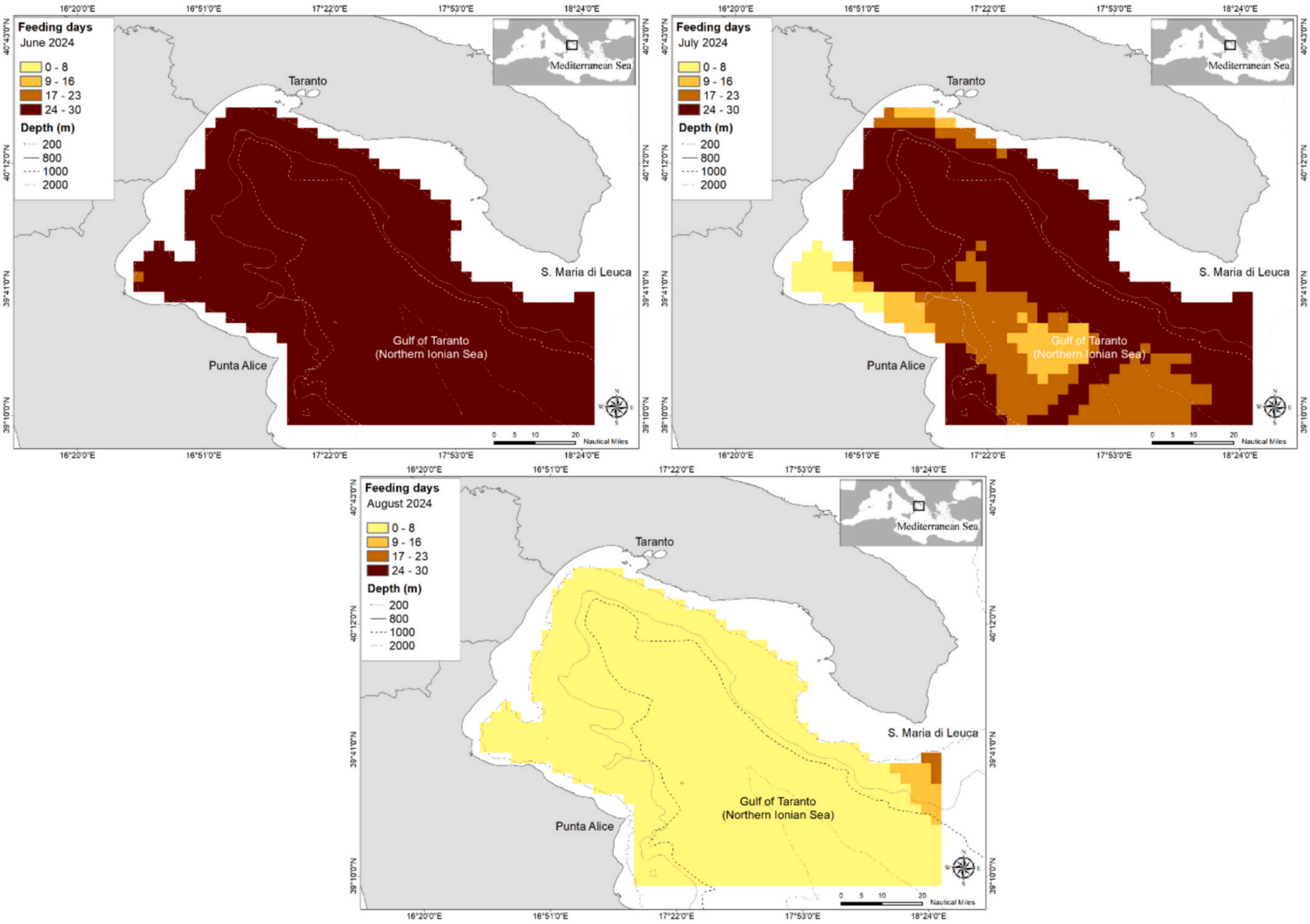

Fig. 11. Predictive maps of suitable feeding habitats for Risso's dolphins in the Gulf of Taranto for June (a), July (b) and August 2024 (c), with different colours indicating the number of days predicted as suitable for feeding, as detailed in the legend.

predictive accuracy. This suggests that incorporating additional variables, such as other biotic descriptors (e.g. prey abundance) or proxies for anthropogenic activities (e.g., distance from fisheries as in Carlucci et al. (2018a, 2018b, 2018c), aquaculture, or noise levels as in La Manna et al. (2023)), could enhance the model's ability to capture the complexity of *T. truncatus* feeding ecology. Evidence from previous studies supports this approach, as human activities, particularly fisheries, have been shown to significantly influence the foraging strategies of common bottlenose dolphins (Methion and Díaz López, 2019; Pace and Pedrazzi, 2024; Ricci et al., 2021, 2023).

Finally, the feeding behaviour of striped dolphin was less strongly associated with environmental variables compared to the other species. Only nitrate concentration at the top and distance from the coast emerged as statistically significant predictor of feeding occurrences (see Table 3). Consistently, the full model (M) was the best-performing CFMs configuration for striped dolphins, with distance from the coast and depth as the first two most important predictors (see Table 4 and Figs. 9,10). These findings show accordance with prior research showing that nitrate levels positively influence the group size of *S. coeruleoalba* in the eastern tropical Pacific Ocean (Redfern et al., 2008) and in the Mediterranean study area (Maglietta et al., 2023b). However, the model's lower accuracy compared to that for Risso's dolphin highlights the species' opportunistic feeding strategy, which likely involves a wider range of prey types and less reliance on specific environmental conditions. In fact, more generalized and sometimes opportunistic species like the common bottlenose and striped dolphins have varied diets, increasing the lag with lower trophic levels (Borrell et al., 2021). For these species, and specifically for the striped dolphin, strong associations have been demonstrated, in SDMs studies, between their presence and more physiographic and physical variables (i.e. depth, distance from the coast, and slope, Chavez-Rosales et al., 2019; Azzolin et al., 2020; Canales-Cáceres et al., 2023; Martino et al., 2021; Maglietta et al., 2023b), as confirmed also by this study's CFMs. However, the limited effectiveness in predicting feeding behaviours only from the environmental variables unveiled the need, as seen for the common bottlenose dolphin, to include in future research other type of variables (e.g. prey abundance, noise levels), as recommended also by Pasanisi et al. (2024).

4.2. Conservation and management implications

The Mediterranean Sea is a biodiversity hotspot where cetaceans face a unique combination of threats including habitat fragmentation, ship collisions, underwater noise, chemical pollution, and climate change (Carlucci et al., 2021b; Lauriano et al., 2014; Micheli et al., 2013). Such a complex environment requires conservation strategies that are both robust and precisely tailored to the basin's ecological and socio-economic context. EU policies, including the Habitat Directive (EEC, 1992), the Marine Strategy Framework Directive (EC, 2008), and the Marine Spatial Planning Directive (EC, 2014), emphasize the importance of targeted actions to protect marine biodiversity. A pillar of these efforts is the identification and protection of Cetacean Critical Habitat (CCH), defined as "those parts of a cetacean's range that are essential for day-to-day well-being and survival, as well as for maintaining a healthy population growth rate" (ACCOBAMS-MOP8/2022/Doc30, 2025). In this context, the development and application of cutting-edge technologies, such as the CFMs are strategical. By providing detailed maps of the environmental requirements of feeding habitats, the CFMs offer critical tools for supporting the identification of CCH and informing conservation strategies. Addressing literature gap on feeding habitat suitability, this study contributes to overcome the absence of effective conservation measures for odontocetes in the Northern Ionian Sea, despite extensive efforts given in recent years to investigate their distribution and ecological roles in this area (Carlucci et al., 2018b, 2021a, 2021b, 2022; Maglietta et al., 2023b).

4.3. Limitations and future directions

The CFMs approach aimed to advance beyond the conventional ecological models and methods of behavioural ecology, suggesting that certain behavioural activities can be effectively and directly modelled using predictors such as the environmental variables, thus providing an exhaustive comprehension of cetacean habitat use. Compared to even the most advanced studies in biologging and bioacoustics, this approach appear to be complementary yet equally effective, as it overcomes the costs, ethical concerns, and logistical limitations associated with technologies like satellite tags by relying on direct observation (Stredulinsky et al., 2023).

To summarize the outcome of the research, it becomes clear that the experimental framework developed is robust, as it integrates both a process of multiple models construction on different datasets to reduce feature noise, and the adopted a consolidated ML strategy (Eggensperger et al., 2013; Maglietta et al., 2023b; Maglietta et al., 2024; Snoek et al., 2012). However, there are some limitation to be discussed. It emerged a variability in model performance across different species, with lower accuracy for common bottlenose and striped dolphins and this could be linked to the model assumption that all individuals within a species will respond similarly to environmental factors (Davies et al., 2023), which may not be accurate. This uncertainty poses a challenge for behavioural-based studies like the present one and underscores the importance of collecting more behavioural data to improve ML models' generalizability. However, the amount of data alone may not fully account for the differences in performance, as the model employs a robust training strategy with hyperparameter optimization and cross-validation, effectively mitigating overfitting. Recent studies (Pasanisi et al., 2024; Virgili et al., 2021) suggest that, to accurately model the habitat use of top predators as cetaceans, it would be more appropriate to incorporate abiotic variables directly linked to their prey dynamics (i.e. prey abundance) rather than relying on indirect proxies such as primary productivity. Furthermore, integrating variables that reflect potential conflicts or disturbances from increasing anthropogenic activities could enhance the model's ability to represent habitat use comprehensively. For this reason, future efforts will be directed toward the implementation of this approach.

To conclude, this study contributes to the growing field of ecological informatics by demonstrating the potential of ML-based models to predict cetacean feeding behaviour and habitat suitability. The CFMs provide a robust framework applicable to other regions and species, offering a tool to support conservation efforts in marine ecosystems under anthropogenic pressure. In the Mediterranean Sea, where endangered cetaceans face unique threats, such approaches are critical for achieving tailored, effective conservation strategies.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT 3.5 to improve language and readability. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

CRediT authorship contribution statement

Carla Cherubini: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology. **Giulia Cipriano:** Writing – review & editing, Writing – original draft. **Leonardo Saccotelli:** Writing – review & editing, Software. **Giovanni Dimauro:** Writing – review & editing. **Giovanni Coppini:** Data curation. **Roberto Carlucci:** Writing – review & editing. **Carmelo Fanizza:** Data curation. **Rosalia Maglietta:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Conceptualization.

Declaration of competing interest

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

Acknowledgements

The authors would like to thank Michele Attolico for his technical assistance.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecoinf.2025.103066.

Data availability

Data and software are freely available here: https://github.com/che7carla/Cetacean-feeding-modelling.git

References

ACCOBAMS, Notarbartolo di Sciara, G., Tonay, A.M. (Eds.), 2021. Conserving Whales, Dolphins and Porpoises in the Mediterranean Sea, Black Sea and Adjacent Areas: An ACCOBAMS Status Report, (2021). ACCOBAMS, Monaco. By. 160 p.

Cetacean critical habitat identification process, 2025. ACCOBAMS-MOP8/2022/Doc30.

Altmann, J., 1974. Observational study of behaviour: sampling methods. Behaviour 49 (3–4), 227–266.

Arcangeli, A., Orasi, A., Carcassi, S.P., Crosti, R., 2014. Exploring thermal and trophic preference of *Balaenoptera physalus* in the Central Tyrrhenian Sea: a new summer feeding ground? Mar. Biol. 161, 427–436.

Arcangeli, A., Azzolin, M., Babey, L., David, L., Garcia-Garin, O., Moulins, A., Rosso, M., Scuderi, A., Tepsich, P., Vighi, M., 2024. Looking for reliable species distribution models for low-density cetacean species: compared effectiveness of SDMs for G. *Griseus*, *G. Melas*, *Z. Cavirostris* in the Mediterranean Sea based on long-term fixed-transect data. Aquatic conservation: marine and freshwater ecosystems 34 (3), e4115.

Azzellino, A., Airoldi, S., Lanfredi, C., Podestà, M., Zanardelli, M., 2017. Cetacean response to environmental and anthropogenic drivers of change: results of a 25-year distribution study in the northwestern Mediterranean Sea. Deep-Sea Res. II Top. Stud. Oceanogr. 146, 104–117.

Azzolin, M., Arcangeli, A., Cipriano, G., Crosti, R., Maglietta, R., Pietroluongo, G., Saintingan, S., Zampollo, A., Fanizza, C., Carlucci, R., 2020. Spatial distribution modelling of striped dolphin (*Stenella coeruleoalba*) at different geographical scales within the EU Adriatic and Ionian Sea region, Central-Eastern Mediterranean Sea. Aquatic Conservation: Marine and Freshwater Ecosystems 30 (6), 1194–1207.

Bergen, S., Huso, M.M., Duerr, A.E., Braham, M.A., Schmuecker, S., Miller, T.A., Katzner, T.E., 2023. A review of supervised learning methods for classifying animal behavioural states from environmental features. Methods Ecol. Evol. 14 (1), 189–202.

Blasi, M.F., Giuliani, A., Boitani, L., 2015. Influence of trammel nets on the behaviour and spatial distribution of bottlenose dolphins (*Tursiops truncatus*) in the Aeolian archipelago, Southern Italy. Aquatic Mammals 41 (3), 295.

Bonizzoni, S., Hamilton, S., Reeves, R.R., Genov, T., Bearzi, G., 2022. Odontocete cetaceans foraging behind trawlers, worldwide. Rev. Fish Biol. Fish. 32 (3), 827–877.

Borrell, A., Gazo, M., Aguilar, A., Raga, J.A., Degollada, E., Gozalbes, P., García-Vernet, R., 2021. Niche partitioning amongst northwestern Mediterranean cetaceans using stable isotopes. Prog. Oceanogr. 193, 102559.

Breiman, L., 2001. Random forests. Machine Learning 45, 5–32.

Cañadas, A., Hammond, P.S., 2008. Abundance and habitat preferences of the short-beaked common dolphin *Delphinus delphis* in the southwestern Mediterranean: implications for conservation. Endanger. Species Res. 4 (3), 309–331.

Canales-Cáceres, R., Gomariz-Castillo, F., Alonso-Sarría, F., Abel, I., Giménez-Casalduero, F., 2023. Distribution and behaviour of striped dolphins in the southwestern Mediterranean Sea based on whale-watching data. Reg. Stud. Mar. Sci. 68, 103256.

Capezzuto, F., Carlucci, R., Maiorano, P., Sion, L., Battista, D., Giove, A., Indennidate, A., Tursi, A., D'Onghia, G., 2010. The bathyal benthopelagic fauna in the north-western Ionian Sea: structure, patterns and interactions. Chem. Ecol. 26 (S1), 199–217.

Carlucci, R., Fanizza, C., Cipriano, G., Paoli, C., Russo, T., Vassallo, P., 2016. Modelling the spatial distribution of the striped dolphin (*Stenella coeruleoalba*) and common bottlenose dolphin (*Tursiops truncatus*) in the Gulf of Taranto (northern Ionian Sea, Central-Eastern Mediterranean Sea). Ecol. Indic. 69, 707–721.

Carlucci, R., Bandelj, V., Ricci, P., Capezzuto, F., Sion, L., Maiorano, P., Tursi, A., Solidoro, C., Libralato, S., 2018a. Exploring spatio-temporal changes in the demersal and benthopelagic assemblages of the north-western Ionian Sea (Central Mediterranean Sea). Mar. Ecol. Prog. Ser. 598, 1–19.

Carlucci, R., Cipriano, G., Paoli, C., Ricci, P., Fanizza, C., Capezzuto, F., Vassallo, P., 2018b. Random Forest population modelling of striped and common-bottlenose dolphins in the Gulf of Taranto (northern Ionian Sea, Central-Eastern Mediterranean Sea). Estuar. Coast. Shelf Sci. 204, 177–192.

Carlucci, R., Bas, A.A., Maglietta, R., Renò, V., Fanizza, C., Rizzo, A., Crugliano, R., Cipriano, G., 2018c. Site fidelity, residency and habitat use of the Risso's dolphin *Grampus griseus* in the Gulf of Taranto (northern Ionian Sea, Central-Eastern Mediterranean Sea) by photo-identification. IEEE International Workshop on Metrology for the Sea; Learning to Measure Sea Health Parameters (MetroSea) 2018, 173–177.

Carlucci, R., Baş, A.A., Liebig, P., Renò, V., Santacesaria, F.C., Bellomo, S., Fanizza, C., Maglietta, R., Cipriano, G., 2020. Residency patterns and site fidelity of *Grampus griseus* (Cuvier, 1812) in the Gulf of Taranto (northern Ionian Sea, Central-Eastern Mediterranean Sea). Mammal Research 65, 445–455.

Carlucci, R., Capezzuto, F., Cipriano, G., D'Onghia, G., Fanizza, C., Libralato, S., Maglietta, R., Maiorano, P., Sion, L., Tursi, A., 2021a. Assessment of cetacean–fishery interactions in the marine food web of the Gulf of Taranto (northern Ionian Sea, Central Mediterranean Sea). Reviews in Fish Biology and Fisheries 31, 135–156.

Carlucci, R., Manea, E., Ricci, P., Cipriano, G., Fanizza, C., Maglietta, R., Gissi, E., 2021b. Managing multiple pressures for cetaceans' conservation with an ecosystem-based marine spatial planning approach. J. Environ. Manage. 287, 112240.

Carlucci, R., Cipriano, G., Cascione, D., Ingrosso, M., Russo, T., Sbrana, A., Fanizza, C., Ricci, P., 2022. Application of a multi-species bio-economic modelling approach to explore fishing traits within eligible cetacean conservation areas in the northern Ionian Sea (Central Mediterranean Sea). Front. Mar. Sci. 9, 1005649.

Carlucci, R., Cipriano, G., Bonato, M., Buscaino, G., Crugliano, R., Fanizza, C., Gatto, S., Maglietta, R., Papetti, C., Pelagatti, M., Ricci, P., 2024. Anthropogenic noise effects on Risso's dolphin vocalizations in the Gulf of Taranto (northern Ionian Sea, Central Mediterranean Sea). Ocean & Coastal Management 254, 107177.

Carrasco, J., Lisón, F., Weintraub, A., 2021. RUSBoost: A suitable species distribution method for imbalanced records of presence and absence. A case study of twenty-five species of Iberian bats. BioRxiv, 2010–2021.

Castelvecchi, D., 2016. Can we open the black box of AI? Nature News 538 (7623), 20.

Cazau, D., Nguyen Hong Duc, P., Druon, J.-N., Matwins, S., Fablet, R., 2021. Multimodal deep learning for cetacean distribution modelling of fin whales (*Balaenoptera physalus*) in the western Mediterranean Sea. Mach. Learn. 1–22.

Chambault, P., Fossette, S., Heide-Jørgensen, M.P., Jouannet, D., Vély, M., 2021. Predicting seasonal movements and distribution of the sperm whale using machine learning algorithms. Ecol. Evol. 11 (3), 1432–1445.

Chavez-Rosales, S., Palka, D.L., Garrison, L.P., Josephson, E.A., 2019. Environmental predictors of habitat suitability and occurrence of cetaceans in the western North Atlantic Ocean. Sci. Rep. 9, 5833.

Cherubini, C., Saccotelli, L., Caccioppoli, R., Fanizza, C., Santacesaria, F.C., Lecci, R., Causio, S., Federico, I., Cipriano, G., Dimauro, G., 2023. Machine learning to predict cetacean behaviour using social and environmental features. In: 2023 IEEE International Workshop on Metrology for the Sea; Learning to Measure Sea Health Parameters (MetroSea), pp. 289–293.

Cipriano, G., Gatto, S., Cherubini, C., Crugliano, R., Santacesaria, F.C., Bellomo, S., Fanizza, C., Papale, E., Buscaino, G., Maglietta, R., Ricci, P., 2022a. Variation in the emission rates of striped dolphin's vocalizations recorded in the Gulf of Taranto (northern Ionian Sea, Central Mediterranean Sea). In: 2022 IEEE International Workshop on Metrology for the Sea; Learning to Measure Sea Health Parameters (MetroSea). IEEE, pp. 236–241.

Cipriano, G., Santacesaria, F.C., Fanizza, C., Cherubini, C., Crugliano, R., Maglietta, R., Ricci, P., Carlucci, R., 2022b. Social structure and temporal distribution of *Tursiops truncatus* in the Gulf of Taranto (Central Mediterranean Sea). Journal of Marine Science and Engineering 10 (12), 1942.

Claro, B., Pérez-Jorge, S., Frey, S., 2020. Seafloor geomorphic features as an alternative approach into modelling the distribution of cetaceans. Eco. Inform. 58, 101092.

Clayton, H., Cade, D.E., Burnham, R., Calambokidis, J., Goldbogen, J., 2023. Acoustic behaviour of gray whales tagged with biologging devices on foraging grounds. Front. Mar. Sci. 10, 1111666.

Clementi, E., Grandi, A., Escudier, R., Lyubartsev, V., 2019a. Mediterranean Sea Production Centre MEDSEA_ANALYSIS_FORECAST_PHYS_006_013.

Clementi, E., Pistoia, J., Escudier, R., Delrosso, D., Drudi, M., Grandi, A., Lecci, R., Cretí, S., Ciliberti, S., Coppini, G., 2019b. Mediterranean Sea Analysis and Forecast (CMEMS MED-Currents, EAS5 System).

Cossarini, G., Feudale, L., Teruzzi, A., Bolzon, G., Coidessa, G., Solidoro, C., Di Biagio, V., Amadio, C., Lazzari, P., Brosich, A., 2021. High-resolution reanalysis of the Mediterranean Sea biogeochemistry (1999–2019). Front. Mar. Sci. 8, 1537.

Cutler, A., Cutler, D.R., Stevens, J.R., 2012. Random forests. Methods and Applications, Ensemble Machine Learning, pp. 157–175.

Davies, S.C., Thompson, P.L., Gomez, C., Nephin, J., Knudby, A., Park, A.E., Friesen, S.K., Pollock, L.J., Rubidge, E.M., Anderson, S.C., 2023. Addressing uncertainty when projecting marine species' distributions under climate change. Ecography 2023 (11), e06731.

De Dominicis, M., Falchetti, S., Trotta, F., Pinardi, N., Giacomelli, L., Napolitano, E., Fazioli, L., Sorgente, R., Haley Jr., P.J., Lermusiaux, J., 2014. A relocatable ocean model in support of environmental emergencies: the Costa Concordia emergency case. Ocean Dyn. 64, 667–688.

Drouot, V., Gannier, A., Goold, J.C., 2004. Diving and feeding behaviour of sperm whales (*Physeter macrocephalus*) in the northwestern Mediterranean Sea. Aquat. Mamm. 30 (3), 419–426.

Drummond, G.B., Tom, B.D.M., 2011. Statistics, probability, significance, likelihood: words mean what we define them to mean. Adv. Physiol. Educ. 35 (4), 361–364.

EC, 2008. European Commission. Directive 2008/56/EC of the European Parliament of the council of 17 June 2008 establishing a framework for community action in the field of marine environmental policy (marine strategy framework directive). Off. J. Eur. Union L164, 19±40.

EC, 2014. European Commission. Directive 2014/89/EC of the European Parliament of the council of 23 July 2014 establishing a framework for maritime spatial planning. Off. J. Eur. Union L257/135.

EEC, 1992. European Commission. EU Council Directive 92/43/EEC on the conservation of natural habitats and of wild fauna and flora. In: Orkesterjournalen L 7–50, 206, 22.7.1992.

Eggensperger, K., Feurer, M., Hutter, F., Bergstra, J., Snoek, J., Hoos, H., Leyton-Brown, K., 2013. Towards an empirical foundation for assessing bayesian optimization of hyperparameters. NIPS Workshop on Bayesian Optimization in Theory and Practice 10 (3).

Fernandez-Arcaya, U., Ramirez-Llodra, E., Aguzzi, J., Allcock, A.L., Davies, J.S., Dissanayake, A., Harris, P., Howell, K., Huvenne, V.A., Macmillan-Lawler, M., Martín, J., 2017. Ecological role of submarine canyons and need for canyon conservation: a review. Front. Mar. Sci. 4, 5.

Fiedler, P.C., Becker, E.A., Forney, K.A., Barlow, J., Moore, J.E., 2023. Species distribution modeling of deep-diving cetaceans. Mar. Mamm. Sci. 39 (4), 1178–1203.

Goldstein, A., Kapelner, A., Bleich, J., Pitkin, E., 2015. Peeking inside the black box: visualizing statistical learning with plots of individual conditional expectation. J. Comput. Graph. Stat. 24 (1), 44–65.

Guisan, A., Tingley, R., Baumgartner, J.B., Naujokaitis-Lewis, I., Sutcliffe, P.R., Tulloch, A.I.T., Regan, T.J., Brotons, L., McDonald-Madden, E., Mantyka-Pringle, C., 2013. Predicting species distributions for conservation decisions. Ecol. Lett. 16 (12), 1424–1435.

Hornsby, F.E., McDonald, T.L., Balmer, B.C., Speakman, T.R., Mullin, K.D., Rosel, P.E., Wells, R.S., Telander, A.C., Marcy, P.W., Klaphake, K.C., Schwacke, L.H., 2017. Using salinity to identify common bottlenose dolphin habitat in Barataria Bay, Louisiana, USA. Endanger. Species Res. 33, 181–192.

Inc., T. M, 2023a. Fitcensemble.

Inc., T. M, 2023b. MATLAB version: 9.14.0 (R2023a).

Inc., T. M, 2023c. Statistics and Machine Learning Toolbox: 12.5 (R2023a).

Kaveh, N., Ebrahimi, A., Asadi, E., 2023. Comparative analysis of random forest, exploratory regression, and structural equation modelling for screening key environmental variables in evaluating rangeland above-ground biomass. Eco. Inform. 77, 102251.

King, S.L., Jensen, F.H., 2023. Rise of the machines: integrating technology with playback experiments to study cetacean social cognition in the wild. Methods in Ecology and Evolution 14 (8), 1873–1886.

Kiszka, J.J., Woodstock, M.S., Heithaus, M.R., 2022. Functional roles and ecological importance of small cetaceans in aquatic ecosystems. Front. Mar. Sci. 9, 803173.

Koger, B., Deshpande, A., Kerby, J.T., Graving, J.M., Costelloe, B.R., Couzin, I.D., 2023. Quantifying the movement, behaviour and environmental context of group-living animals using drones and computer vision. J. Anim. Ecol. 92 (7), 1357–1371.

La Manna, G., Ronchetti, F., Sarà, G., 2016. Predicting common bottlenose dolphin habitat preference to dynamically adapt management measures from a marine spatial planning perspective. Ocean & coastal management 130, 317–327.

La Manna, G., Ronchetti, F., Perretti, F., Ceccherelli, G., 2023. Determinants of group size in the common bottlenose dolphins: the role of water temperature and noise. Behav. Ecol. Sociobiol. 77 (4), 45.

Lanfredi, C., Arcangeli, A., David, L., Holcer, D., Rosso, M., Natoli, A., 2021. *Grampus griseus* (Mediterranean subpopulation). The IUCN red list of threatened species 2021: e.T16378423A190737150. https://doi.org/10.2305/IUCN.UK.2021-3.RLTS.T16378423A190737150.en.

Lauriano, G., Di Guardo, G., Marsili, L., Maltese, S., Fossi, M.C., 2014. Biological threats and environmental pollutants, a lethal mixture for mediterranean cetaceans? J. Mar. Biol. Assoc. U. K. 94 (6), 1221–1225.

Leoni, J., Tanelli, M., Strada, S.C., Berger-Wolf, T., 2020. Ethogram-based automatic wild animal monitoring through inertial sensors and GPS data. Eco. Inform. 59, 101112.

Lucas, T.C.D., 2020. A translucent box: interpretable machine learning in ecology. Ecological monographs 90 (4), e01422.

Luna, A., Sánchez, P., Chicote, C., Gazo, M., 2022. Cephalopods in the diet of Risso's dolphin (*Grampus griseus*) from the Mediterranean Sea: a review. Mar. Mamm. Sci. 38 (2), 725–741.

Maglietta, R., Renò, V., Cipriano, G., Fanizza, C., Milella, A., Stella, E., Carlucci, R., 2018. DolFin: an innovative digital platform for studying Risso's dolphins in the northern Ionian Sea (north-eastern Central Mediterranean). Sci. Rep. 8 (1), 17185.

Maglietta, R., Caccioppoli, R., Seller, E., Bellomo, S., Santacesaria, F.C., Colella, R., Cipriano, G., Stella, E., Hartman, K., Fanizza, C., Dimauro, G., 2020. Convolutional neural networks for Risso's dolphins identification. IEEE Access 8, 80195–80206.

Maglietta, R., Carlucci, R., Fanizza, C., Dimauro, G., 2022. Machine learning and image processing methods for cetacean photo identification: a systematic review. IEEE Access 10, 80195–80207.

Maglietta, R., Bussola, A., Carlucci, R., Fanizza, C., Dimauro, G., 2023a. ARIANNA: a novel deep learning-based system for fin contours analysis in individual recognition of dolphins. Intelligent Systems with Applications 18, 200–207.

Maglietta, R., Saccotelli, L., Fanizza, C., Telesca, V., Dimauro, G., Causio, S., Lecci, R., Federico, I., Coppini, G., Cipriano, G., 2023b. Environmental variables and machine learning models to predict cetacean abundance in the Central-Eastern Mediterranean Sea. Sci. Rep. 13 (1), 2600.

Maglietta, R., Caccioppoli, R., Piazzolla, D., Saccotelli, L., Cherubini, C., Scagnoli, E., Piermattei, V., Marcelli, M., De Lucia, G.A., Lecci, R., Causio, S., Dimauro, G., De Franco, F., Scuro, M., Coppini, G., 2024. Habitat suitability modeling of loggerhead sea turtles in the central-eastern Mediterranean Sea: a machine learning approach using satellite tracking data. Front. Mar. Sci. 11, 1493598.

Mann, J., 1999. Behavioral sampling methods for cetaceans: a review and critique. Mar. Mamm. Sci. 15 (1), 102–122.

Martino, S., Pace, D.S., Moro, S., Casoli, E., Ventura, D., Frachea, A., Silvestri, M., Arcangeli, A., Giacomini, G., Ardizzone, G., Jona Lasinio, G., 2021. Integration of presence-only data from several sources: a case study on dolphins' spatial distribution. Ecography 44 (10), 1533–1543.

Melo-Merino, S.M., Reyes-Bonilla, H., Lira-Noriega, A., 2020. Ecological niche models and species distribution models in marine environments: a literature review and spatial analysis of evidence. Ecol. Model. 415, 108837.

Menegotto, A., Rangel, T.F., 2018. Mapping knowledge gaps in marine diversity reveals a latitudinal gradient of missing species richness. Nat. Commun. 9 (1), 4713.

Methion, S., Díaz López, B., 2019. Natural and anthropogenic drivers of foraging behaviour in bottlenose dolphins: influence of shellfish aquaculture. Aquatic Conservation: Marine and Freshwater Ecosystems 29 (6), 927–937.

Micheli, F., Halpern, B.S., Walbridge, S., Ciriaco, S., Ferretti, F., Fraschetti, S., Lewison, R., Nykjaer, L., Rosenberg, A.A., 2013. Cumulative human impacts on Mediterranean and Black Sea marine ecosystems: assessing current pressures and opportunities. PloS One 8 (12), e79889.

Mintzer, V.J., Fazioli, K.L., 2021. Salinity and water temperature as predictors of bottlenose dolphin (*Tursiops truncatus*) encounter rates in upper Galveston Bay, Texas. Frontiers in Marine Science 8, 754686.

Neumann, D.R., 2001. Activity budget of free-ranging common dolphins (*Delphinus delphis*) in the northwestern Bay of Plenty. New Zealand. Aquatic Mammals 27 (2), 121–136.

Ngô, M.C., Selvan, R., Tervo, O., Heide-Jørgensen, M.P., Ditlevsen, S., 2021. Detection of foraging behavior from accelerometer data using U-net type convolutional networks. Eco. Inform. 62, 101275.

Nowacek, D.P., Christiansen, F., Bejder, L., Goldbogen, J.A., Friedlaender, A.S., 2016. Studying cetacean behaviour: new technological approaches and conservation applications. Anim. Behav. 120, 235–244.

Pace, D.S., Pedrazzi, G., 2024. Bottlenose dolphin predation on flathead grey mullets using barrier feeding techniques: a visual and acoustic case study in the Mediterranean Sea. Ethol. Ecol. Evol. 1–9.

Pace, D.S., Panunzi, G., Arcangeli, A., Moro, S., Jona-Lasinio, G., Martino, S., 2022. Seasonal distribution of an opportunistic apex predator (*Tursiops truncatus*) in marine coastal habitats of the Western Mediterranean Sea. Front. Mar. Sci. 9, 939692.

Pasanisi, E., Pace, D.S., Orasi, A., Vitale, M., Arcangeli, A., 2024. A global systematic review of species distribution modelling approaches for cetaceans and sea turtles. Eco. Inform. 82, 102700.

Peda, C., Battaglia, P., Scuderi, A., Voliani, A., Mancusi, C., Andaloro, F., Romeo, T., 2015. Cephalopod prey in the stomach contents of odontocete cetaceans stranded in the western Mediterranean Sea. Mar. Biol. Res. 11 (6), 593–602.

Pérez-Jorge, S., Oliveira, C., Rivas, E.I., Prieto, R., Cascão, I., Wensveen, P.J., Miller, P.J., Silva, M.A., 2023. Predictive model of sperm whale prey capture attempts from time-depth data. Movement Ecology 11 (1), 33.

Pinardi, N., Lyubartsev, V., Cardellicchio, N., Caporale, C., Ciliberti, S., Coppini, G., De Pascalis, F., Dialti, L., Federico, I., Filippone, M., Grandi, A., 2016. Marine rapid environmental assessment in the Gulf of Taranto: a multiscale approach. Nat. Hazards Earth Syst. Sci. 16 (12), 2623–2639.

Pitchford, J.L., Howard, V.A., Shelley, J.K., Serafin, B.J., Coleman, A.T., Solangi, M., 2016. Predictive spatial modelling of seasonal bottlenose dolphin (*Tursiops truncatus*) distributions in the Mississippi sound. Aquat. Conserv. Mar. Freshwat. Ecosyst. 26 (2), 289–306.

Redfern, J.V., Barlow, J., Ballance, L.T., Gerrodette, T., Becker, E.A., 2008. Absence of scale dependence in dolphin–habitat models for the eastern tropical Pacific Ocean. Mar. Ecol. Prog. Ser. 363, 1–14.

Refaeilzadeh, P., Tang, L., Liu, H., 2016. Cross-Validation, 1–7, 10. Springer New York, New York, NY, pp. 971–978.

Ricci, P., Libralato, S., Capezzuto, F., D'Onghia, G., Maiorano, P., Sion, L., Tursi, A., Solidoro, C., Carlucci, R., 2019. Ecosystem functioning of two marine food webs in the North-Western Ionian Sea (Central Mediterranean Sea). Ecol. Evol. 9 (18), 10198–10212.

Ricci, P., Ingrosso, M., Carlucci, R., Fanizza, C., Maglietta, R., Santacesaria, F.C., Tursi, A., Cipriano, G., 2020. Quantifying the dolphins-fishery competition in the gulf of Taranto (northern Ionian Sea, Central Mediterranean Sea). Proceedings IMEKO Metrology for the Sea 135–140.

Ricci, P., Manea, E., Cipriano, G., Cascione, D., D'Onghia, G., Ingrosso, M., Fanizza, C., Maiorano, P., Tursi, A., Carlucci, R., 2021. Addressing cetacean–fishery interactions to inform a deep-sea ecosystem-based management in the Gulf of Taranto (northern Ionian Sea, Central Mediterranean Sea). Journal of Marine Science and Engineering 9 (8), 872.

Ricci, P., Serpetti, N., Cascione, D., Cipriano, G., D'Onghia, G., De Padova, D., Fanizza, C., Ingrosso, M., Carlucci, R., 2023. Investigating fishery and climate change effects on the conservation status of odontocetes in the northern Ionian Sea (Central Mediterranean Sea). Ecol. Model. 485, 110500.

Riekkola, L., Andrews-Goff, V., Friedlaender, A., Constantine, R., Zerbini, A.N., 2019. Environmental drivers of humpback whale foraging behaviour in the remote Southern Ocean. J. Exp. Mar. Biol. Ecol. 517, 1–12.

Roman, J., Estes, J.A., Morissette, L., Smith, C., Costa, D., McCarthy, J., Nation, J.B., Nicol, S., Pershing, A., Smetacek, V., 2014. Whales as marine ecosystem engineers. Front. Ecol. Environ. 12 (7), 377–385.

Saarela, M., Jauhiainen, S., 2021. Comparison of feature importance measures as explanations for classification models. SN Applied Sciences 3, 1–12.

Schorr, G.S., Hanson, M.B., Falcone, E.A., Emmons, C.K., Jarvis, S.M., Andrews, R.D., Keen, E.M., 2022. Movements and diving behavior of the eastern North Pacific offshore killer whale (*Orcinus orca*). Front. Mar. Sci. 9, 854893.

Seiffert, C., Khoshgoftaar, T.M., Van Hulse, J., Napolitano, A., 2009. RUSBoost: a hybrid approach to alleviating class imbalance. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans 40 (1), 185–197.

Shane, S.H., 1990. Behavior and ecology of the bottlenose. The Bottlenose Dolphin 245.

Snoek, J., Larochelle, H., Adams, R.P., 2012. Practical bayesian optimization of machine learning algorithms. Adv. Neural Inf. Proces. Syst. 25.

Sol, M., Ollier, C., Boisseau, O., Ridoux, V., Virgili, A., 2024. Temporal patterns in dolphin foraging activity in the Mediterranean Sea: insights from vocalisations recorded during the ACCOBAMS survey initiative. Front. Mar. Sci. 11, 1378524.

Soldevilla, M.S., Wiggins, S.M., Hildebrand, J.A., Oleson, E.M., Ferguson, M.C., 2011. Risso's and Pacific white-sided dolphin habitat modelling from passive acoustic monitoring. Mar. Ecol. Prog. Ser. 423, 247–260.

Stredulinsky, E.H., Toews, S., Watson, J., Noren, D.P., Holt, M.M., Thornton, S.J., 2023. Delineating important killer whale foraging areas using a spatiotemporal logistic model. Global Ecology and Conservation 48, e02726.

Thomas, L., Buckland, S.T., Rexstad, E.A., Laake, J.L., Strindberg, S., Hedley, S.L., Bishop, J.R., Marques, T.A., Burnham, K.P., 2010. Distance software: design and analysis of distance sampling surveys for estimating population size. J. Appl. Ecol. 47 (1), 5–14.

Van Cise, A.M., Hanson, M.B., Emmons, C., Olsen, D., Matkin, C.O., Wells, A.H., Parsons, K.M., 2024. Spatial and seasonal foraging patterns drive diet differences among North Pacific resident killer whale populations. Royal Society open Science 11 (9), rsos240445.

Virgili, A., Hedon, L., Authier, M., Calmettes, B., Claridge, D., Cole, T., Corkeron, P., Dorémus, G., Dunn, C., Dunn, T.E., Laran, S., 2021. Towards a better characterisation of deep-diving whales' distributions by using prey distribution model outputs? PloS One 16 (8), e0255667.

von Schuckmann, K., Le Traon, P.-Y., Smith, N., Pascual, A., Brasseur, P., Fennel, K., Djavidnia, S., Aaboe, S., Fanjul, E.A., Autret, E., 2018. Copernicus marine service ocean state report. J. Oper. Oceanogr. 11 (sup1), S1–S142.

Vujović, Ž., 2021. Classification model evaluation metrics. Int. J. Adv. Comput. Sci. Appl. 12 (6), 599–606.

Wang, R., 2012. AdaBoost for feature selection, classification and its relation with SVM, a review. Phys. Procedia 25, 800–807.

Wijeyakulasuriya, D.A., Eisenhauer, E.W., Shaby, B.A., Hanks, E.M., 2020. Machine learning for modeling animal movement. PloS One 15 (7), e0235750.