Fully convolutional neural networks applied to large-scale marine morphology mapping

Riccardo Arosio[1,2*†], Brandon Hobley[3†], Andrew J. Wheeler[1,2,4], Fabio Sacchetti[5], Luis A. Conti[6], Thomas Furey[5] and Aaron Lim[1,2,7]

[1]Marine Geoscience Group, School of Biological, Earth and Environmental Sciences, University College Cork, Cork, Ireland, [2]Environmental Research Institute, University College Cork, Cork, Ireland, [3]School of Computing Sciences, University of East Anglia, Norwich, United Kingdom, [4]SFI Centre for Research In Applied Geosciences, University College Cork, Cork, Ireland, [5]Marine Institute, Galway, Ireland, [6]Escola de Artes Ciências e Humanidades, Universidade de São Paulo, São Paulo, Brazil, [7]Department of Geography, University College Cork, Cork, Ireland

In this study we applied for the first time Fully Convolutional Neural Networks (FCNNs) to a marine bathymetric dataset to derive morphological classes over the entire Irish continental shelf. FCNNs are a set of algorithms within Deep Learning that produce pixel-wise classifications in order to create semantically segmented maps. While they have been extensively utilised on imagery for ecological mapping, their application on elevation data is still limited, especially in the marine geomorphology realm. We employed a high-resolution bathymetric dataset to create a set of normalised derivatives commonly utilised in seabed morphology and habitat mapping that include three bathymetric position indexes (BPIs), the vector ruggedness measurement (VRM), the aspect functions and three types of hillshades. The class domains cover ten or twelve semantically distinct surface textures and submarine landforms present on the shelf, with our definitions aiming for simplicity, prevalence and distinctiveness. Sets of 50 or 100 labelled samples for each class were used to train several U-Net architectures with ResNet-50 and VGG-13 encoders. Our results show a maximum model precision of 0.84 and recall of 0.85, with some classes reaching as high as 0.99 in both. A simple majority (modal) voting combining the ten best models produced an excellent map with overall F1 score of 0.96 and class precisions and recalls superior to 0.87. For target classes exhibiting high recall (proportion of positives identified), models also show high precision (proportion of correct identifications) in predictions which confirms that the underlying class boundary has been learnt. Derivative choice plays an important part in the performance of the networks, with hillshades combined with bathymetry providing the best results and aspect functions and VRM leading to an overall deterioration of prediction accuracies. The results show that FCNNs can be successfully applied to the seabed for a morphological exploration of the dataset and as a baseline for more in-depth habitat mapping studies. For example, prediction of semantically distinct classes

as "submarine dune" and "bedrock outcrop" can be precise and reliable. Nonetheless, at present state FCNNs are not suitable for tasks that require more refined geomorphological classifications, as for the recognition of detailed morphogenetic processes.

1 Introduction

In the fast-expanding field of marine habitat and geomorphological mapping, with an increasing influx of data at high spatial resolution being gathered by geophysical and remote sensing surveys ("Big Data"), rapid, machine-based and cost-effective methods that capture the nuances of the highly varying seabed environments have become essential. Thus, computer-based supervised and unsupervised mapping methods have become progressively more popular, demonstrating equivalence or superiority to traditional manual mapping (Micallef et al., 2012; Diesing et al., 2014; Ismail et al., 2015). Presently, the leading supervised mapping approach is a combination of object-based image analysis (OBIA) (Blaschke et al., 2014) and conventional machine learning models (e.g. Decision Trees, Support Vector Machines, Random Forests etc.). In this approach, OBIA first segments raw data, for example imagery, into a suitable internal representation of descriptive objects, then a machine learning sub-system detects statistical patterns in extracted descriptive features in order to distinguish different class domains. When the raw data are digital surface models (e.g. from multibeam echosounders or Lidar data), as it happens for most marine-based habitat mapping studies, the segmentation and statistics are largely based on morphological and substrate attributes (e.g. relative depth, roughness, backscatter etc.), and habitat prediction is strictly linked to morphology. Recent applications range between identification and analysis of coral mounds (Diesing and Thorsnes, 2018; Conti et al., 2019; de Oliveira et al., 2021), sediment wave characterisation (Summers et al., 2021) to general marine mapping (Ierodiaconou et al., 2018; Linklater et al., 2019). However, the OBIA method still requires careful engineering and considerable domain expertise and manual intervention, which increases processing time and effectiveness.

In the last decade, Deep Learning (DL), and in particular Convolutional Neural Networks (CNNs) have supported more traditional approaches, and have shown state of the art results on a wide range of imaging problems (Long et al., 2014; He et al., 2017; Krizhevsky et al., 2017). Fully Convolutional Neural Networks (FCNNs) are a variant of CNNs that can perform per-pixel classification. Contrarily to traditional machine learning, FCNNs allow for hierarchical feature learning, which in effect combines learning features and training a classifier in one optimisation (LeCun et al., 2015). Furthermore, FCNNs can leverage semi-supervised strategies whereby subsets of labelled data are used for optimisation; this approach can be beneficial for practical applications of FCNNs for marine geomorphology and ecology mapping where the quantity and distribution of labelled data may be limited due to associated costs of *in situ* surveying (Leitão et al., 2018; Hobley et al., 2021). While interest in DL has been shown early on by the marine community for ecological and habitat mapping (Gazis et al., 2018; Yasir et al., 2021), only a few studies have been focused on automated identification with DL of seabed geomorphological features or textures (McClinton et al., 2012; Valentine et al., 2013; Juliani, 2019; Keohane and White, 2022; Lundine et al., 2023), even though the significance of geomorphology for habitat distribution is widely acknowledged (Brown et al., 2011; Lecours et al., 2016; Harris and Baker, 2020). Deep Learning in geomorphology has found instead a more fertile ground in coastal and geohazard studies (Ma and Mei, 2021; Buscombe et al., 2023), and in outer space, in particular for Martian or Lunar geology, where several studies have taken advantage of the high resolution optical imagery available and attempted to separate specific landforms from a background (Foroutan and Zimbelman, 2017; Palafox et al., 2017; Wang et al., 2017; Rubanenko et al., 2021), or more generally characterise the ground surface to identify optimal landing spots or assess rover traversability (Wilhelm et al., 2020; Barrett et al., 2022). Barrett et al. (2022) in particular have demonstrated the potential of large-scale exploratory morphological mapping, where machine learning assists the geomorphologist to isolate sections of interest in the dataset, sifting through an enormous dataset.

Following on this latter example, and transposing it to the marine realm, in this study we explore the potential of FCNNs to map distinctive morphologies on the seabed, generate the prospective to create an automated, streamlined method to greatly increase the efficiency of many seabed mapping workflows including data exploration of the main morphological signatures, preliminary domain segmentation for ground-truthing campaigns and the identification of areas of interest or generalised habitat predictions. We align the exercise to recurrent situations and practices in seabed mapping, and we test the capability of the FCNNs to their limits, feeding the bare minimum usually available to researchers:

1) we furnish only bathymetry and bathymetry-derived surface functions as input layers (contrarily to the optical imagery used for the planetary studies) and do not include multibeam backscatter data as it is sometimes unreliable.

Elevation as main input in itself poses challenges as DL methods are designed for optical imagery.

2) we provide only a very limited amount of labelled data, as the creation of large amount of labels would defeat the purpose of automation and time saving. This constitutes a second challenge, as in spite of CNNs success, these models perform best with very large, labelled training datasets (Tarvainen and Valpola, 2017). Labels are a pivotal concern in many real-world scenarios as CNNs are optimised based on an objective error metric between model outcomes and known outcomes.

In the next sections of the paper, firstly we describe the dataset utilised and the classification systems adopted, which include two different classifications and label sets. Secondly, we concatenate various combinations of bathymetry and derivative layer inputs to create pseudo-images and assess the value of the different derivatives in the predictions. In parallel, we trial two FCNN encoders and semi-supervision techniques to gauge their effectiveness with the non-standard input data (i.e. bathymetry). Finally, we discuss the results from the point of view of applicability to marine seabed or habitat mapping studies, including challenges behind finding the optimal set of semantic morphological classes, the impact of mapping landforms with diverging dimensions and the importance of selecting appropriate derivatives for modelling neural networks on bathymetry-derived data.

2 Materials and methods

2.1 Input layers: bathymetry and derivatives

The multibeam echosounder (MBES) bathymetry utilised in this study was obtained from the INFOMAR hydrographic dataset, which is freely accessible on the INFOMAR website (https://www.infomar.ie) (Figure 1). Bathymetric data at 10 m resolution were downloaded and processed using ESRI ArcMap v 10.6. Firstly, fine holes in the dataset were filled with the mean of the surrounding 5x5 pixel neighbours. A general median filter (5x5 rectangle) was applied to remove 'salt-and-pepper' imperfections and fine artefacts before re-gridding using a nearest neighbour algorithm. For the purpose of this large-scale mapping, a resolution of 25 m/pixel was deemed a good compromise between morphological detail, partial suppression of acquisition artefacts in the INFOMAR dataset (especially at the outer beam) and computing speed. Bathymetry derivatives were calculated using ArcMap built-in algorithms or with the help of the Benthic Terrain Modeller (BTM) toolbox version 3.0 (Walbridge et al., 2018). The derivatives created include three bathymetric position indexes, a vector ruggedness measurement, two aspect functions (eastness and northness) and three types of hillshades (Table 1). The aspect functions rasters were smoothed using a Gaussian filter (5x5 rectangle) to simplify the signal and reduce salt-and-pepper effects.

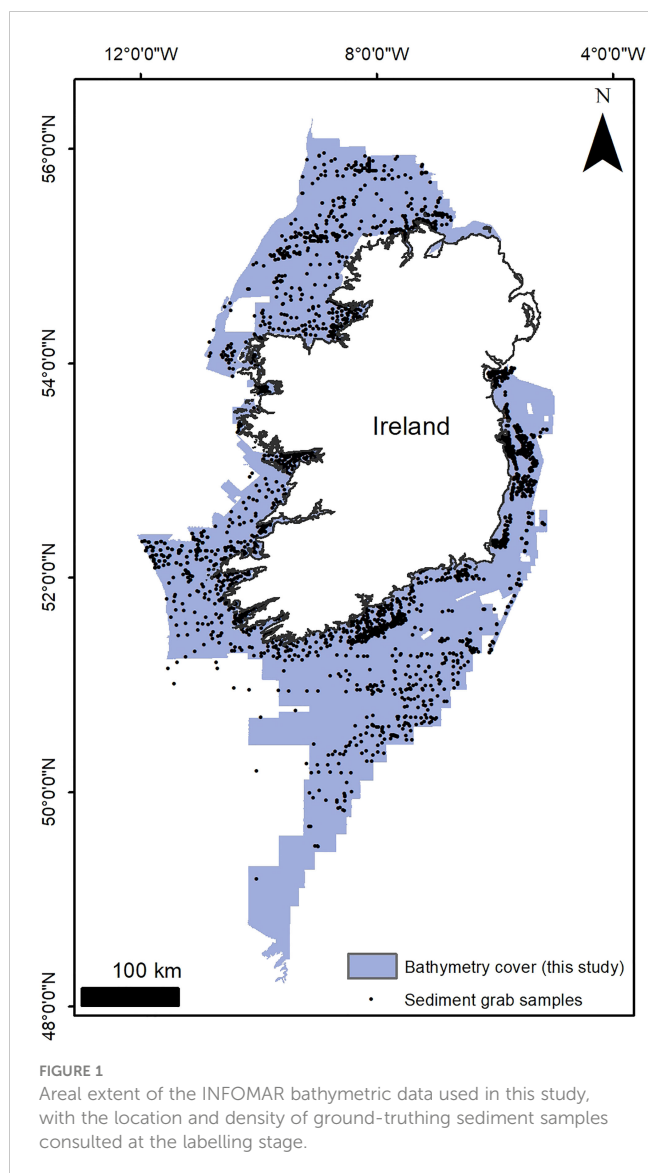For the purposes of FCNN model training, the bathymetry and derivatives were normalised to a double precision value between 0 and 1 based on the minimum and maximum value recorded or calculated in the case for derivative layers.

2.2 Classification system and labelling

To train the weakly supervised convolutional neural network, we had to define a dataset from which the model could learn the relationship between bathymetry and derivative data and the landforms present. So firstly, a suitable classification system had to be chosen. The classification system adopted is derived from the Mareano-INFOMAR-Maremap-Geoscience Australia (MIM-GA) two-part marine geomorphology scheme, a standardised seabed mapping glossary aimed to enable more consistent seabed classifications (Dove et al., 2016; Dove et al., 2020). This framework independently describes seabed features according to their observed physical structure (Morphology), and the more subjective interpretation of their origin and evolution (Geomorphology). The separation between physical structure and genesis aligned well with the scope of the machine learning-based mapping of this study, where classes were defined based upon the textural characteristics of the surface rather than apparent geological nature or proper geomorphological definitions. This mapping approach was chosen both because of the general lack of geological ground-truthing for novel marine datasets, and for the exploratory nature of the exercise. In general, defined classes describe archetypal seabed textures which, in various combinations, form seabed landforms. The basic distinction between sediment and rock landforms was nonetheless retained (see Table 2), and the INFOMAR sediment grab dataset (https://www.infomar.ie/maps/interactive-maps/seabed-and-sediment) was consulted at the labelling stage (Figure 1). Three principles for classification were adopted (following the advice in Barrett et al., 2022): (a) classes had to be representative of the diversity of seabed morphologies encountered on the Irish continental shelf. This is a "completeness" rule; FCNNs classify pixels in maximum-likelihood fashion, therefore it is essential to fully capture the problem domain as the FCNNs cannot create a new class, or leave a space blank, if an unknown type of seabed is encountered. (b) The classification sets were kept simple and short, as a comprehensive, lengthy list of classes would potentially create difficulties in the training process and especially create more subjective inconsistencies during the labelling work carried out by the expert, and (c) last, but most importantly, classes had to be distinct so that their differences could be confidently isolated visually by the human mapper. This step is critical as the labelling stage may introduce subjectivity and inconsistencies in class delineation that can affect the capabilities of the networks. Therefore, care was taken to semantically define each class, making sure that delineation could be performed with a high level of confidence notwithstanding the limited geological knowledge.

2.2.1 Terrain and landform classes

A list of 10 classes (Table 2, Figure 2) was considered sufficient to capture the morphological domain of the study area. These classes include three types (or textures) of hard substrate – *Fissured,*

FIGURE 1
Areal extent of the INFOMAR bathymetric data used in this study,
with the location and density of ground-truthing sediment samples
consulted at the labelling stage.

Hummocky and *Layered (rock)*, which comprise bedrock outcrops of meta-/sedimentary and igneous nature but can also include rough or rubbly glacial surfaces which are often hardly distinguishable from bedrock. *Corrugated* and *Ridge (sediment)* capture the extensive current-induced bedform fields, respectively the short wavelength megaripples, sediment ribbons or dunes and the larger dunes of different type (transverse, linear, trochoidal etc.) which occur especially in the Irish and Celtic seas (Van Landeghem et al., 2009; Creane et al., 2022). The Irish shelf glacial vestiges, which include prevalently moraines (Ó Cofaigh et al., 2012) and drumlin fields (Benetti et al., 2010) are captured in the *Large Ridge* and *Hummocky (sediment)* classes respectively, although we included the Celtic "megaridges" (Lockhart et al., 2018) and sediment banks in the *Large Ridge* class. The *Depression* class includes the bathymetric lows on the shelf, which are prevalently channel-like features including scouring, palaeofluvial channels/tunnel valleys (Giglio et al., 2022) and some isolated cases of large pockmarks. Finally, finer scale, elongated depressions or incisions as iceberg ploughmarks and furrowing are represented by the *Grooved*

(sediment) class. The *Plane* class act as filler for the areas of smooth and featureless terrain. A second list of 12 classes (Table 2, Figure 2) was created to test the performance of the FCNNs with a slightly more complex problem. The second classification set was established increasing the detail for *Large Ridge* and *Depression*, splitting them respectively into *Bank (sediment)* and *Relict Ridge* (thus dividing the sediment banks from the glacial ridges), and *Depression (enclosed)* and *Depression (elongated)* (thus separating circular or quasi circular scouring and pockmarks from channels and elongated scours).

2.2.2 Labelling procedures

The labelling of seabed classes was carried out by a single human annotator (the first author) utilising expert judgement with the support of published studies and sediment grain size data for ground-truthing (GT samples in Figure 1) available from the INFOMAR website. Classes were labelled by manually digitising polygons on ArcGIS 10.6 and making sure they contained only the landforms or terrain textures of interest, partially or completely, regardless of their dimensions. Therefore, naturally larger landforms (e.g. the class *Large Ridge*) are defined by larger labels. Two sets of labels were created, one containing 50 labels per class, and a second with 100 labels per class. The labelled areas constitute only a very small proportion of the total study area (97,526 km^2), with the 100-label set covering only 3.2% of the total, the 50 label (12 classes) 2.8% and the 50 labels (10 classes) 2.57%.

Each digitised polygon contains a unique semantic value associated to the landform or terrain texture class. FCNNs were trained with rasterised labels that contain one-to-one mappings of pixels from input layers (Long et al., 2014). The rasterised labels employed to train FCNNs were created using the geographic coordinates stored in each digitised polygon label and converting real-world coordinates for each vertex to image-coordinates. Training pseudo-images were created by centre-cropping 256 x 256 pixel (Figure 3) blocks containing multi-layered raster (bathymetry and derivatives) data. By centre-cropping digitised label polygons, the edges of each pseudo-image may possess a number of unlabelled pixels, which in turn allows for semi-supervised approaches to be leveraged (see Section 2.3). Two factors are behind the specific dimension of the pseudo-images. Firstly, the power of two (256 = 2^8) grants numerical ease in image resizing (i.e., the blocks are divided/multiplied by 2) with sequential pooling operations and up-sample. Secondly, the image size is appropriate for GPU memory constraints and mini-batch optimisation. For instance, 256x256 may allow 8 images per batch which was found to be optimal for neural network optimisation, whereas an image of 512x512 allows only 1 to 2 images per batch and converges the neural network incorrectly. In the case digitised polygons covered a region that extended past the 256 x 256 area, centre-crops were split into several 256 x 256 blocks. This process generated 553 training images for the 50 label 12-class, 543 training images for the 50 label 10-class and 1134 training images for the 100 label 10-class. For each dataset, the training imagery was randomly subdivided into mutually exclusive training (90%) and validation (10%) sets.

TABLE 1 List of derivatives and production parameters utilised in this study.

Derivative	Name	Focal statistics parameters			Hillshade parameters			Description
		type	dimension 1	dimension 2	Azimuth	Altitude	Exaggeration	
Bathymetric position index	BPI 1	annulus	2	10	na	na	na	BPI represents the positional difference of a pixel compared to the mean of its neighbours (defined by an annulus in this case).
	BPI 2	annulus	5	50	na	na	na	
	BPI 3	annulus	25	250	na	na	na	
Vector ruggedness measurement	VRM	circle	3	na	na	na	na	VRM is a measure of surface roughness as the variation in three-dimensional orientation of grid cells within a neighbourhood. Vector analysis is used to calculate the dispersion of vectors (orthogonal) to grid cells within the specified neighbourhood.
Aspect Northness [cos(aspect)]	Northness	na	na	na	na	na	na	Aspect identifies the downslope direction of the maximum rate of change in value from each cell to its neighbours. Eastness and northness are the sine and cosine functions of aspect
Aspect Eastness [sin(aspect)]	Eastness							
Hillshade	HS 1	na	na	na	multidirectional		10	A hillshade function produces a grayscale pseudo-3D image of the bathymetry, with the sun's relative position taken into account for shading the image. Different sun positions give different shading effects.
	HS 2	na	na	na	315	45	10	
	HS 3	na	na	na	45	45	10	

2.3 Fully Convolutional Neural Networks

Fully Convolutional Neural Networks (Long et al., 2014) are an extension of traditional CNN architectures (Krizhevsky et al., 2017) adapted for semantic segmentation. CNNs comprise a series of layers that process lower layer inputs through repeating convolution and pooling operations followed by a final classification layer. Each convolution and pooling layer transform the input image, or in this case bathymetric data, into higher level abstracted representations. FCNNs can be broken down into two networks: an encoder and a decoder network. The encoder network is identical to a CNN, except the final classification layer is removed. The decoder network applies transposed convolutions in order to up-sample feature maps back to the original input size, and each decoding stage combines corresponding feature maps created by the encoder network. The final classification layer utilizes 1 by 1 convolution kernels (Lin et al., 2014) to transform the original bathymetric data and derivatives source into a set of dense probabilities using a softmax transfer function. Network weights and biases are adjusted through gradient descent by minimizing the loss function between network outputs and the ground truth pixel labels.

The overall architecture of the FCNN used in this study (Figure 3) is based on a U-Net (Ronneberger et al., 2015) and the encoder networks are VGG-13 and ResNet50 (Simonyan and Zisserman, 2015; He et al., 2017). Residual learning using ResNet encoders has proven to surpass very deep neural networks such as VGG, but for completeness in results we experimented with both encoder networks. The decoder network applies a transposed 2 by 2 convolution for a learnt up-sample track and concatenates feature maps from the encoding network at appropriate resolutions followed by a final 3 by 3 convolution. The final 1 by 1 convolution condenses feature maps to have the same number of channels as the total number of classes in the dataset (Figure 3).

Semi-supervision is the process of incorporating unlabelled image samples for the optimisation of deep neural networks. This branch of deep learning methods is more applicable when unlabelled data are readily available, while labelled instances are often hard, expensive, and time-consuming to collect. Semi-supervised methods can be capable of building better classifiers that compensate for the lack of labelled training data and therefore present a cost-effective solution to label acquisition. In this study, where semantic segmentation was achieved with a pixel classifier, the masks that were used to label pixels did not cover entire 256x256 pseudo-images and therefore every pseudo-image had pixels that were left unlabelled. This condition allowed for unsupervised loss terms to be added into the optimisation process and thus for semi-supervision to be implemented. The supervised loss term is calculated by processing a mini batch of images $X \in \mathbb{R}^{B \times C \times H \times W}$ and corresponding segmentation maps $Y \in \mathbb{R}^{B \times C \times H \times W}$, where B, C, H and W are batch size, number of input channels, height and width. The network produces per-pixel logits $\bar{Y} \in \mathbb{R}^{B \times K \times H \times W}$ where K is the number of target classes. The softmax transfer function (1) converts network scores into probabilities by normalizing all K scores for each pixel to sum to one:

$$P_k(x) = \frac{e^{\overline{Y_k}(x)}}{\sum_{k'=1}^{K} e^{\overline{Y_{k'}}(x)}} \quad (1)$$

TABLE 2   Classifications adopted in this study with either 10 or 12 classes and their correspondence to the MIM-GA classification system (Dove et al., 2020).

MIM-GA class	Class (10)	Class (12)	Morphological description	Geomorphology
Plane	**Plane**	**Plane**	Flat and smooth surface with either absent or imperceptible relief.	Mainly represents areas of the seabed that are covered with modern Holocene soft, potentially mobile sediments (sands and muds)
Ridge	**Ridge (sediment)**	**Ridge (sediment)**	Surfaces characterised by a pattern of generally regular corrugation or ondulation, with pronounced linear or curvilinear crests. These are usually medium to high frequency and medium-high relief.	Areas of developed and active or pristine dune fields
Hummocks/ Hill	**Hummocky (sediment)**	**Hummocky (sediment)**	Distinct elongated ridges or hills with relatively smooth surface, often occurring in swarms	High relief drumlin hills or other relict hilly terrain
Ridge	**Corrugated (sediment)**	**Corrugated (sediment)**	Surfaces characterised by regular or irregular corrugation or ondulation. These are generally high frequency and small-scale.	Areas of subdued dune fields or megaripples, scouring or sorted bedforms
Lineation	**Layered (rock)**	**Layered (rock)**	Surfaces characterised by quasi regular or irregular patterns of repeated crests. These are generally high frequency and low relief.	Bedrock exposure where the bedding or tectonic fabric is well-developed and regular.
Groove	**Grooved (sediment)**	**Grooved (sediment)**	Planar or quasi planar surfaces traversed by linear, shallow and narrow incisions or grooves, that can be regular and sub-parallel or irregular and criss-crossing.	Iceberg ploughmarks, elongated furrows or trawling areas
Blocks	**Fissured (rock)**	**Fissured (rock)**	Polygonal and quasi-regular fractured seabed platforms, showing high relief compared to surrounding seabed, often breaking up into fragmented polygonal blocks, usually flat-topped.	Jointed and fragmented bedrock outcrops and platforms
Hummocks	**Hummocky (rock)**	**Hummocky (rock)**	Rough, low-relief hummocky surfaces, often adjacent to *Fissured (rock)* that have been interpreted as exposed or only partially covered bedrock.	Bedrock exposure where the bedding or tectonic fabric is cryptic or absent. Areas of consolidate or partially consolidated irregular glacial sediment (e.g. till).
Bank/Ridge	**Large ridge**	**Bank (sediment)**	Bank or large ridge-like feature with a generally smooth surface.	Sediment bank or mounded sediment accumulation
		Relict ridge (sediment)	large, low and broad ridge or smaller irregular arcuate ridge-like feature with a generally rough surface	Glacial moraines or grounding zone wedges
Hole	**Depression**	**Depression (enclosed)**	Closed-contour bathymetric depression below the surrounding seabed	Various negative relief including scours and pockmarks
Channel		**Depression (elongated)**	Elongated bathymetric depression below the surrounding seabed	Various negative relief including scours and palaeo-channels

Where, $x \in \Omega; \Omega \subseteq \mathbb{Z}^2$ is a pixel location and $P_k(x)$ is the probability for the kth channel at pixel location x. The negative log-likelihood loss is calculated between segmentation maps and network probabilities:

$$L = -\sum_{k=1}^{K} \mathbf{Y_k(x)} \, \log(\mathbf{P_k(x)}) \quad (2)$$

For each image, the supervised loss is the sum of all losses for each pixel using Equation (2) and averaged according to the number of labelled pixels in Y. Full details on the use of semi-supervision can be found in Hobley et al. (2021). The training parameters and convergence of FCNNs was analysed by testing multiple settings for learning rate and batch size, and assessing computed confusion matrices over several consecutive runs of the algorithm. This ensured that a fair range of different convergence approaches was evaluated. Furthermore, for the semi-supervised approach, several different loss weights were experimented to tune for the unsupervised loss term. The best performing networks were trained for 300 epochs with a batch-size of 12 using AdamW optimiser with a learning rate set to 0.001. With regards to the semi-supervised approach, the unsupervised loss was

scaled down by a factor of 10 and the confidence threshold for teacher prediction was set to 0.97. All FCNNs were implemented and trained using Pytorch version 10.2, the code is freely available on our GitHub repository: https://github.com/BrandonHobley/geomorph_deep.

2.4 Quality assessment

The quantitative metrics of interest to evaluate a classification algorithm are precision and recall (Equations (3) and (4)). These metrics are adequate to test classification algorithms over different datasets as well as their capability to detect false positives and false negatives. Precision and recall are metrics that can show how a classifier performs for each specific class, where precision measures the ability of the model to identify only the relevant instances, while recall measures the ability to detect correctly the occurrence of a class of interest. For instance, in a dataset with 17 confirmed landforms and 121 false landforms, an algorithm that detects every case as false would have an accuracy of 87%, but at the same time it would have an extremely poor recall of 13%. The F1-score (Equation (5)) is the
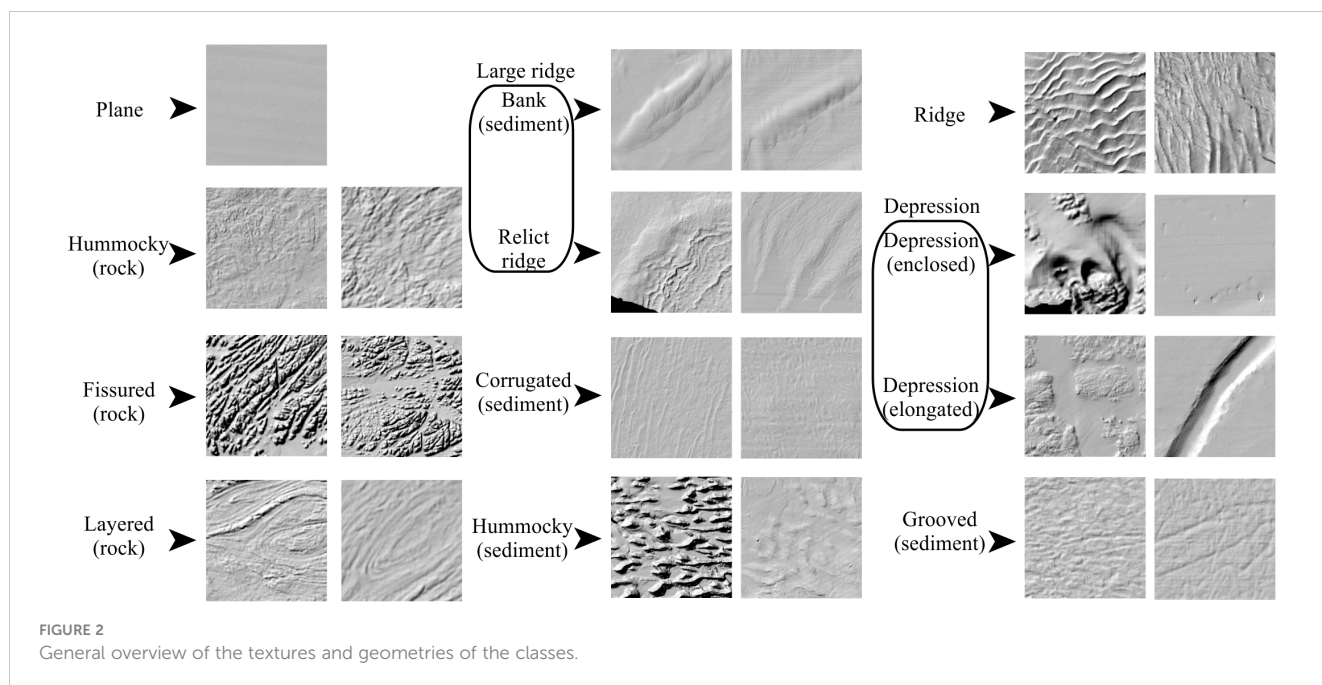
FIGURE 2
General overview of the textures and geometries of the classes.

harmonic mean of recall and precision, giving a suitable generalised single figure of merit to convey the performance of a classifier.

$$precision = \frac{true\_positive}{true\_positive+false\_positive} \quad (3)$$

$$recall = \frac{true\_positive}{true\_positive+false\_negative} \quad (4)$$

$$F1\_score = 2 \times \frac{recall \times precision}{recall+precision} \quad (5)$$

The quantitative evaluation metrics listed above are valid if the dataset is labelled, which in our study covers a small subset of the total surface mapped. Therefore, these results can give an indication of a particular classifier performance, but visual inspection is still required to fully grasp the capabilities of FCNNs for bathymetric data.

3 Results

3.1 Model performance

For this study, 40 different FCNN model runs were carried out, and total mean of their performances are presented in Table 3.

FIGURE 3
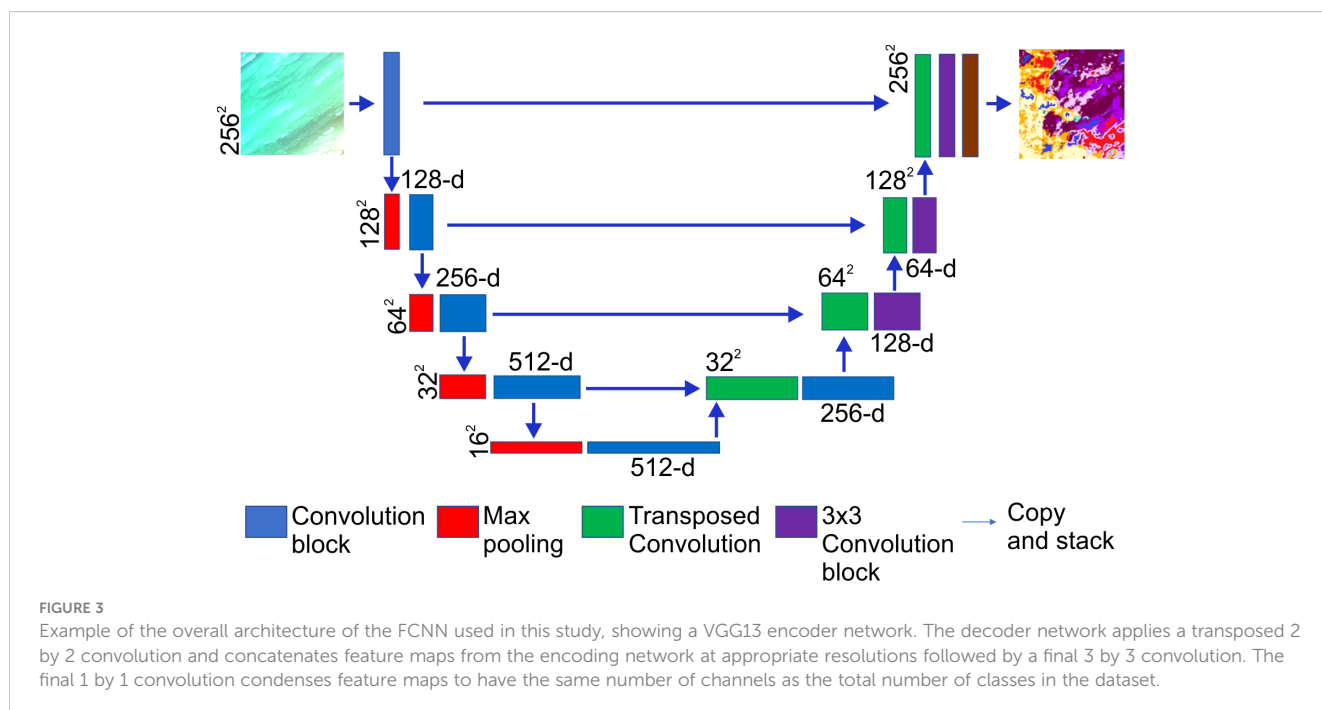Example of the overall architecture of the FCNN used in this study, showing a VGG13 encoder network. The decoder network applies a transposed 2 by 2 convolution and concatenates feature maps from the encoding network at appropriate resolutions followed by a final 3 by 3 convolution. The final 1 by 1 convolution condenses feature maps to have the same number of channels as the total number of classes in the dataset.

Precision, recall and F1 scores for each class and model are instead given in the Supplementary Material.

The first set of results (models #1 to #12) shows the initial tests carried out on the two different encoders (VGG13 and ResNet50) comparing their efficacy and assessing the utility of semi-supervision and some preliminary combinations of input layers. Overall, the scores show that neither VGG13 nor ResNet50 outperforms the other, although ResNet50 produces slightly better scores at the second decimal point, with increases between 0.01 and 0.05 (e.g. compare models #6 and #7 or #10 and #11).

The use of semi supervision does not improve significantly nor consistently the results, contributing to positive or negative fluctuations. For example, ResNet50 model #11 acquires 0.053 points in Precision compared to non-supervised #12 (Figure 4), with no change in Recall. ResNet50 model #8 gains only 0.004 points in Precision and loses 0.004 points in Recall compared to unsupervised model #7. VGG13 models seem instead to suffer more the application of semi-supervision, leading to higher loss in scores (e.g. compare model outputs #5 and #6).

The best model results were achieved using the complete set of input layers, with ResNet50 models #11 and #12 giving F1 scores of 0.724 and 0.688 respectively. Evaluation metrics are supported by the visual assessment of the resulting thematic maps, where models #11 and #12 show the most visually pleasing results (Figure 4). Nonetheless, high scores were also obtained limiting the input to a combination of bathymetry, BPIs, VRM and Aspect functions (models #3 and #8, with F1 scores of 0.677 and 0.660) or bathymetry and hillshades (model #4, F1 score of 0.641). In order to test the contribution of the input layers to the model predictions, a series of additional model runs were carried out using both encoders but without implementing the semi-supervision -which the previous results reveal to be relatively erratic, and increasing the number of labels to 100 per class, to gauge the effect of boosting label number to model performance.

The results of this series of tests are presented in Table 3, model numbers #19 to #40. As expected, the scores show an overall improvement caused by the increase in the number of labels from 50 to 100. While the time effort required to create labels for the classes is doubled (from 500 to 1000 labels in total), the improvements are significant, up to ~0.19 points, i.e., from 0.641 to 0.833 when comparing the F1 scores of the best VGG13 bathymetry and hillshade results (#4 vs #29).

Once again ResNet50 runs are slightly more successful in evaluation metrics compared to VGG13, with ResNet50 scoring higher in F1 8 times out of 11 model runs.

The exploration of the usefulness of each input layer in model performance provides strong indications that the hillshades are the most valuable set of layers for a correct prediction of morphological classes. Models that utilise hillshades have consistently higher scores than those that do not (cf. for example models #21 with #31 or #33 with #37, Table 3). The use of hillshades alone provides very good results (model #26 F1 score: 0.774), although the combination of layers with different azimuths is essential and a single hillshade is insufficient to produce an accurate map. Overall, the combination of the other derivatives alone or with bathymetry leads to substantially

inferior predictions, with Precision scores consistently under 0.59 and Recall scores only slightly better. Aspect functions and VRM do not seem to provide useful information to the models, on the contrary their addition is detrimental to their performance. For example, bathymetry as input layer alone (#19 and #20) contributes to a better score than bathymetry combined with aspect functions and VRM (#33 and #34). While the bathymetric position indexes improve the predictions of the bathymetry baseline, they do not seem to enhance significantly the performance of the hillshade layers, with oscillating results when comparing the "HS full" baselines (models #25 and 26) and the "BPI + all HS" (models #31 and 32). The only layer that does improve the predictions of the hillshades alone is the bathymetry, with model runs #29 and 30 presenting the highest scores obtained in this study (VGG13, Precision 0.845, Recall 0.850). Visually comparing the map outputs of hillshades alone against bathymetry-supported hillshades shows improvement in score metrics obtained by the latter as reflected in the outlook of the map (Figure 5), although the crispness of the boundaries is somewhat diminished, creating more "padded" class interfaces and generalisations.

The effect of the combination of all the input layers is given in runs #39 and 40, where the scores are only slightly superior to the model runs of the hillshades, but inferior to the bathymetry and hillshades runs. Overall, *Large Ridge, Plane* and *Fissured (rock)* are the three most successfully identified classes by all 10-class models, with an average F1 score of 0.884, 0.772 and 0.716 respectively. *Corrugated (sediment), Hummocky (sediment)* and *Layered (rock)* score instead the lowest across all models, with average F1 scores of 0.471, 0.466 and 0.423. The confusion matrices (see the Supplementary Material) show that the prevalent misinterpretation is related to Type II errors (false negatives) where *Layered (rock)* is classified as *Plane, Hummocky (sediment)* as *Depression* and *Corrugated (sediment)* as *Large Ridge*. Probable causes for these misinterpretations are treated in the discussion.

Finally, models #13 to #18 (Table 3) show the results of separate tests carried out to investigate the performance of the FCNNs with an increased number of labels. All the results show a substantial decrease in all the scores when moving from the 10 class to the 12 class problem, with Precision and Recall ranging between 0.442-0.567 and 0.451-0.584. Confusion matrices show a decline in accuracy in all classes, and not only those that were split. *Depression (enclosed)* and *Bank (sediment)* scored the lowest amongst the classes, showing that the separation from the original and more general *Depression* and *Large Ridge* classes (10 class division) weakens the training.

3.2 Modal voting and combined map

The use of several permutations and combinations of different input layers allows for an ensemble learning scenario to be leveraged. We have tested this hypothesis with a simple modal voting of FCNN pixel classifications for the 10 best performing models (both in terms of scores and visual quality), which produced an excellent map with an overall F1 score of 0.96 and class precisions and recalls superior to
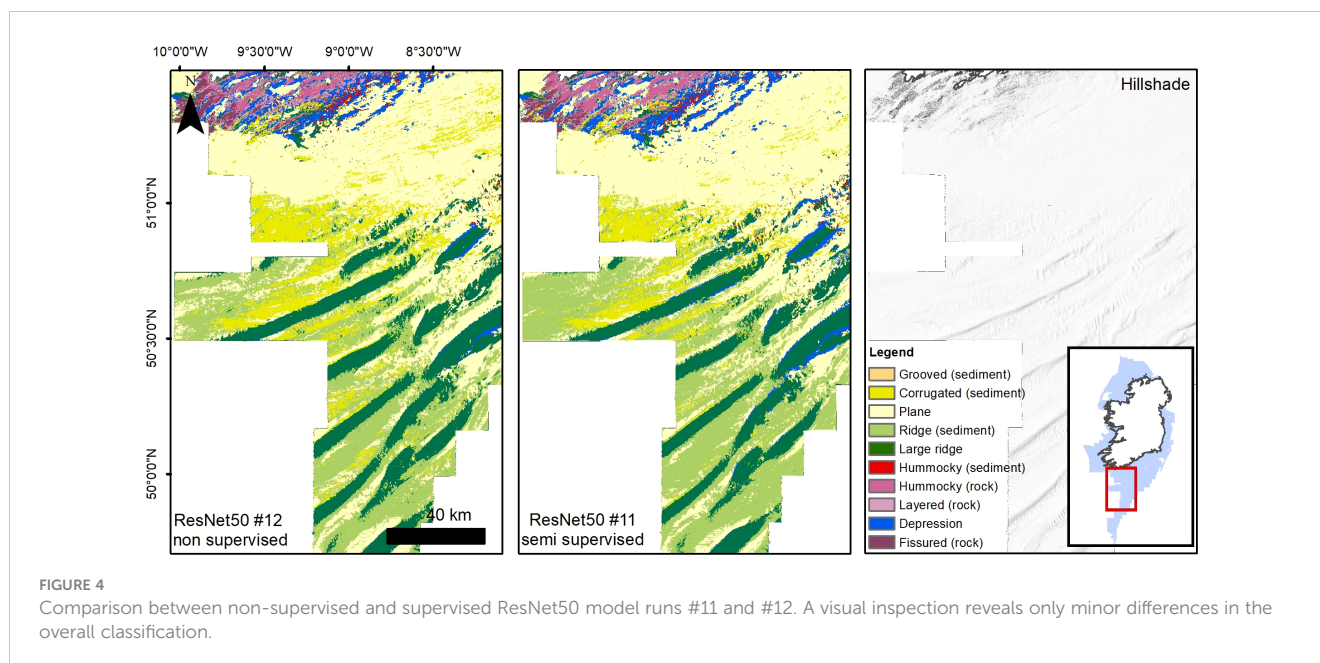
TABLE 3 Complete list of model results from this study.

#	Name	Encoder	Epoch	Semi supervision	Class number	Labels per class	Bathy	BPI1	BPI2	BPI3	VRM	Northn	Eastn	HS1	HS2	HS3	Precision	Recall	F1 score
											Input layers								
1	bbd1	VGG13	128	n	10	50	y	y	y	y	y	y	y	n	n	n	0.675	0.629	0.644
2	bbd2	VGG13	140	n	10	50	y	y	y	y	y	y	y	n	n	n	0.694	0.674	0.680
3	bbd3	ResNet50	84	n	10	50	y	y	y	y	y	y	y	n	n	n	0.717	0.660	0.677
4	bhsc1	VGG13	200	n	10	50	y	n	n	n	n	n	n	y	y	y	0.694	0.649	0.641
5	bbd4	VGG13	227	y	10	50	y	y	y	y	y	y	y	n	n	n	0.646	0.638	0.622
6	bbd4	VGG13	246	n	10	50	y	y	y	y	y	y	y	n	n	n	0.647	0.682	0.638
7	bbd5	ResNet50	240	n	10	50	y	y	y	y	y	y	y	n	n	n	0.697	0.688	0.668
8	bbd6	ResNet50	282	y	10	50	y	y	y	y	y	y	y	n	n	n	0.701	0.684	0.660
9	cs1	VGG13	199	n	10	50	y	y	y	y	y	y	y	y	y	y	0.707	0.766	0.656
10	cs2	VGG13	263	y	10	50	y	y	y	y	y	y	y	y	y	y	0.722	0.763	0.676
11	cs3	ResNet50	228	y	10	50	y	y	y	y	y	y	y	y	y	y	0.771	0.775	0.724
12	cs4	ResNet50	253	n	10	50	y	y	y	y	y	y	y	y	y	y	0.718	0.775	0.688
13	bbd7	VGG13	279	y	12	50	y	y	y	y	y	y	y	n	n	n	0.442	0.507	0.430
14	bbd8	VGG13	287	n	12	50	y	y	y	y	y	y	y	n	n	n	0.505	0.584	0.521
15	bbd9	ResNet50	275	n	12	50	y	y	y	y	y	y	y	n	n	n	0.520	0.549	0.513
16	bbd10	ResNet50	291	y	12	50	y	y	y	y	y	y	y	n	n	n	0.491	0.548	0.482
17	cs5	VGG13	111	n	12	50	y	y	y	y	y	y	y	y	y	y	0.567	0.491	0.460
18	cs6	VGG13	320	n	12	50	y	y	y	y	y	y	y	y	y	y	0.549	0.451	0.451
19	b1	VGG13	228	n	10	100	y	n	n	n	n	n	n	n	n	n	0.343	0.385	0.274
20	b2	ResNet50	298	n	10	100	y	n	n	n	n	n	n	n	n	n	0.299	0.455	0.302
21	bp1	VGG13	161	n	10	100	n	y	y	y	n	n	n	n	n	n	0.457	0.584	0.460
22	bp2	ResNet50	149	n	10	100	n	y	y	y	n	n	n	n	n	n	0.405	0.595	0.434
23	hs1	VGG13	235	n	10	100	n	n	n	n	n	n	n	y	n	n	0.489	0.661	0.508
24	hs2	ResNet50	243	n	10	100	n	n	n	n	n	n	n	y	n	n	0.612	0.672	0.619
25	hsc1	VGG13	186	n	10	100	n	n	n	n	n	n	n	y	y	y	0.665	0.757	0.683

(Continued)

TABLE 3 Continued

#	Name	Encoder	Epoch	Semi supervision	Class number	Labels per class	Bathy	BPI1	BPI2	BPI3	VRM	Northn	Eastn	HS1	HS2	HS3	Precision	Recall	F1 score
26	hsc2	ResNet50	103	n	10	100	n	n	n	n	n	n	n	y	y	y	0.786	0.788	0.774
27	bb1	VGG13	243	n	10	100	y	y	y	y	n	n	n	n	n	n	0.521	0.656	0.544
28	bb2	ResNet50	286	n	10	100	y	y	y	y	n	n	n	n	n	n	0.530	0.655	0.555
29	bhsc2	VGG13	240	n	10	100	y	n	n	n	n	n	n	y	y	y	0.845	0.850	0.833
30	bhsc3	ResNet50	252	n	10	100	y	n	n	n	n	n	n	y	y	y	0.825	0.848	0.830
31	bphs1	VGG13	280	n	10	100	n	y	y	y	n	n	n	y	y	y	0.731	0.779	0.733
32	bphs2	ResNet50	165	n	10	100	n	y	y	y	n	n	n	y	y	y	0.772	0.801	0.770
33	bd1	VGG13	38	n	10	100	y	n	n	n	y	y	y	n	n	n	0.287	0.291	0.291
34	bd2	ResNet50	45	n	10	100	y	n	n	n	y	y	y	n	n	n	0.212	0.266	0.262
35	bpd1	VGG13	268	n	10	100	n	y	y	y	y	y	y	n	n	n	0.365	0.530	0.393
36	bpd2	ResNet50	240	n	10	100	n	y	y	y	y	y	y	n	n	n	0.545	0.629	0.565
37	hscd1	VGG13	150	n	10	100	n	n	n	n	y	y	y	y	y	y	0.598	0.753	0.614
38	hscd2	ResNet50	296	n	10	100	n	n	n	n	y	y	y	y	y	y	0.771	0.776	0.745
39	cs6	VGG13	176	n	10	100	y	y	y	y	y	y	y	y	y	y	0.720	0.823	0.735
40	cs7	ResNet50	280	n	10	100	y	y	y	y	y	y	y	y	y	y	0.749	0.818	0.765

FIGURE 4
Comparison between non-supervised and supervised ResNet50 model runs #11 and #12. A visual inspection reveals only minor differences in the overall classification.

0.87. The results and full map are presented in Table S1 and Figure S1 in the Supplementary Material.

4 Discussion

Scores and qualitative assessment of the results have shown that both ResNet50 and VGG13 encoders can achieve good accuracy, with performances driven mostly by the nature of the input layers and the quantity and precision of the labelling. The unsuccessful attempt with 12 classes is most likely caused by a fallacy in the semantic definition of these classes more than weakness of the networks, and it shows that FCNNs can be very susceptible to deceptive labelling. In the first set of tests the best score result was given by ResNet50 model #11, that included all input layers; however, our subsequent analysis of layer contribution shows that the best results are achieved with hillshades and bathymetry only. It must be said that this discrepancy relies on the comparison with a single observation in the first set (i.e. model #4 vs models #9 to 12), and if we take the worst performing model with all input layers (model #9), its scores are not too different from those of hillshade-based model #4 (only Recall being significantly higher in #9). The limitation in sample comparison coupled with the consistent observation that non-hillshade derivatives do not enhance the performance even in the best of cases, support the conclusion that either model #4 is an underperforming outlier or that the doubling of labels has substantially improved the prediction performance based on hillshades. The evaluation metrics improvement generated by the addition of the bathymetry layer to the hillshades input is possibly partly due to the nature of the offshore physiography, where some classes are preferentially found at specific bathymetric ranges. For example, bedrock outcrops are focussed close to the coastline, and unusually high F1 scores for
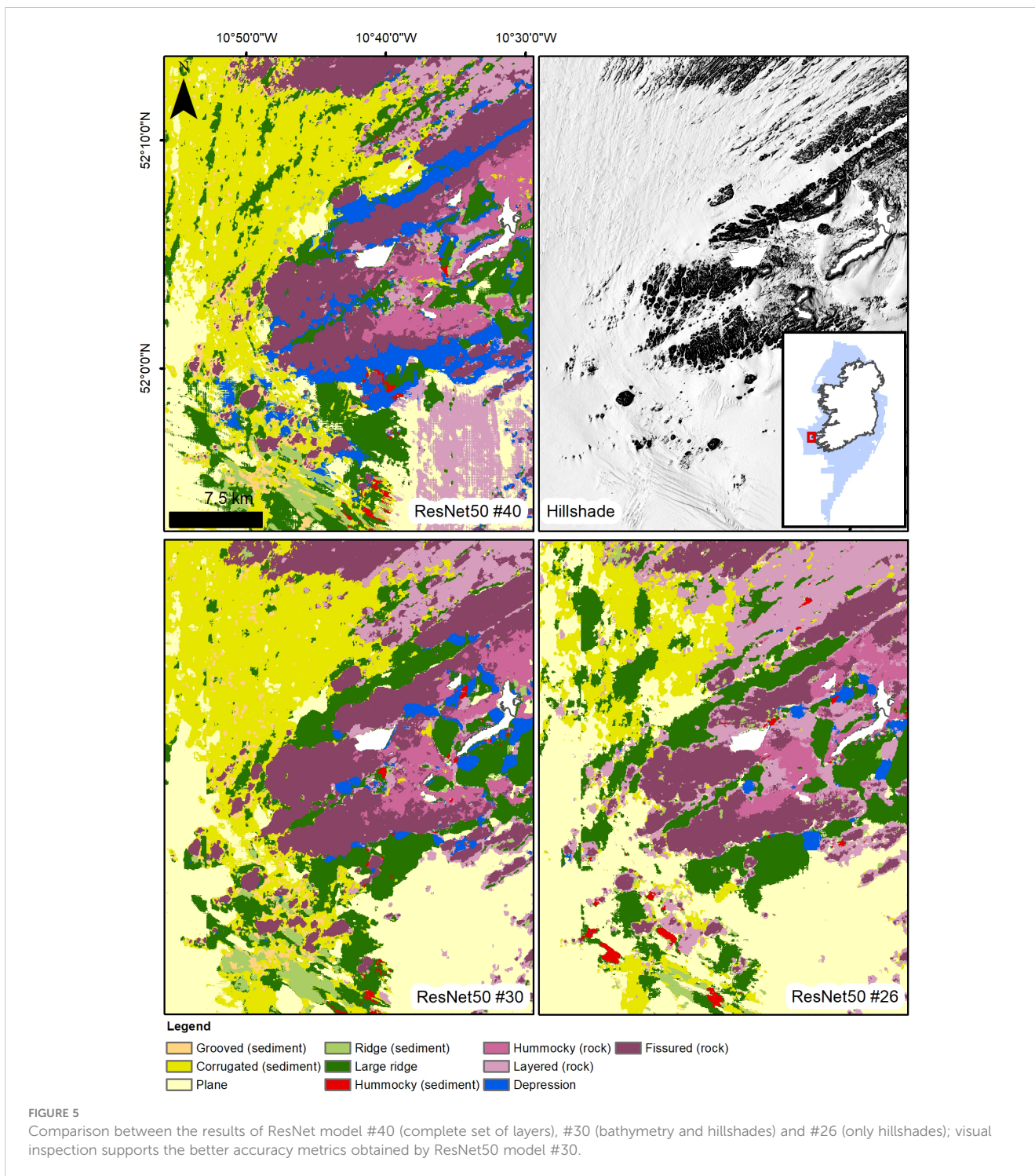
Fissured (rock) and *Hummocky (rock)* in the bathymetry-based models (#19, #20, see Table S1 in Supplementary Material) strengthen the suspicion of a regional bias. Therefore, the utility of the bathymetry input is potentially lower in different datasets.

Figures 6–8 give an overview of the results provided by the best performing model (#30) and the combined modal vote map. A qualitative assessment of the maps shows that slightly better performances are sometimes achieved to the detriment, in places, of boundary crispness and detail. The evaluation metrics, calculated on a pixel basis, give a good approximation of the effectiveness of a model, however in order to fully assess the models' performance and potential for seabed mapping studies, we need to consider the results in term of boundary position, nature of misclassifications, type of class misclassified and general distribution of errors.

4.1 Sources of error and uncertainty

In the breakdown of evaluation metrics for each class (Table S1 in Supplementary Material) the three most recurring weakest predictions are linked to the classes *Corrugated (sediment)*, *Hummocky (sediment)* and *Layered (rock)*. Coupling the observations of class type misinterpretation (see Results) and the qualitative assessment of the map outputs has led to the identification of three main types of errors or uncertainty.

Misclassifications linked to liminal spaces between classes is the first type of ambiguities we discuss (Figures 9A, B). This misclassification is reflected in the significant confusion between *Layered (rock)* and *Plane* or *Hummocky (sediment)* and *Depression*. Stratified, gently dipping bedrock possesses significant extents of planar features within them (bedding planes), that transition into fine elongated and often isolated ridges. This texture is sometimes misidentified as *Plane*, but in unlabelled data can also be observed as

FIGURE 5

Comparison between the results of ResNet model #40 (complete set of layers), #30 (bathymetry and hillshades) and #26 (only hillshades); visual inspection supports the better accuracy metrics obtained by ResNet50 model #30.

Layered (rock) in areas of sorted bedforms, that possess similar geometry. A similar case is provided by the *Hummocky (sediment)* class, which includes the occurrences of drumlins (oval shaped, glacial-flow aligned, moraine hills formed beneath fast-moving ablating ice flows). The drumlins are surrounded by depressed areas, the "connecting surface" between the high relief landforms. Models tended to confuse the proximal interconnecting surface as *Depression* instead of "drumlin", leading to the lower score. In

defence of the networks, it is often very difficult even for a geomorphologist to find the "correct" place to draw a boundary to define a landform (Smith and Mark, 2003). One major reason that labelling was carried out by a single expert, was to try to achieve maximum consistency in delineation, as another geomorphologist might introduce subjective bias and training conflicts for the networks. Moreover, complex terrains or where class assignment felt ambiguous were deliberately not labelled, leaving effectively the
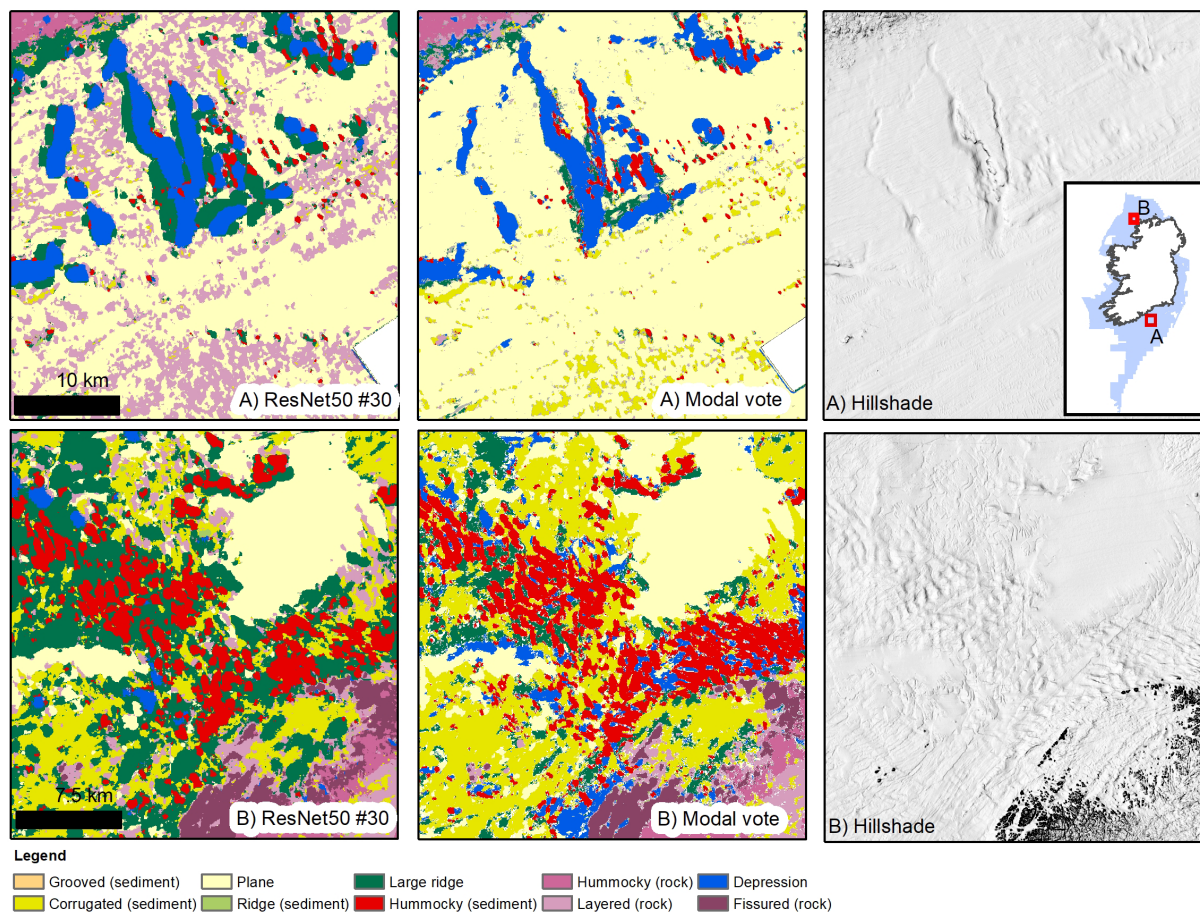
FIGURE 6
Results from best performing ResNet50 model (#30 – bathymetry and hillshades only) and the modal vote map. While producing overall the best precision and recall scores amongst the model runs, model #30 has underperformed in the detection of the *Layered (rock)* class (F1 score 0.584), completely misinterpreting the sorted bedforms in the Celtic Sea as rock **(A)**. The modal vote map is instead effective in recognising the bedforms, having better efficacy in identifying *Layered (rock)* (F1 score 0.90). The glacial streamlined terrain in **(B)** is well captured by model #30, with only minor mixing between *Large Ridge* and *Hummocky (sediment)* where Rogen moraines become larger and are intertwined with larger underlying morainic ridges. While the modal vote map gives also a fair depiction of the area, it overestimates the presence of *Depressions*, probably due to the interference of the BPI layers.

model to decide. We have stressed in the Methods section that good care was taken in the definition of distinctive semantic classes, however these errors indicate that morphological textures form part of a spectrum that is fundamentally difficult to compartmentalise (e.g., at what scale and configuration does a corrugation become a hummock or vice-versa)?, and the shortcomings of the FCNNs are at least partly by-products of natural variability and the inability of a set of classes to fully capture it. Without using a more complex set of classes and fuzzy classifiers it is not possible to treat any existing terrain variation.

The second type of ambiguity is related to scale (Figures 7, 8, 10). Our classifications included the class *Large Ridge* (or *Bank (sediment)* and *Relict ridge*), which can be significantly bigger than other terrain or landform classes. This factor of scale ambiguity was introduced wittingly into the models, as we wanted to explore the "style" and ability of the networks to disentangle the problem of multiscale classification, which is very common in geomorphology and habitat mapping. If restricted to create a single map layer with a

small number of classes, the human mind would prioritise the assignation of a class depending on what they think is the most important attribute to classify. So, for example, a large moraine which is covered by a boulder field might be preferentially mapped as "moraine", even though both classes identify a correct characteristic of the ground. The hierarchical nature of BTM (Walbridge et al., 2018; Goes et al., 2019; de Oliveira et al., 2020) perpetuates this problem. In our results, this multiscale ambiguity is well reflected in the misclassification of *Corrugated (sediment)* as *Large Ridge*; corrugated surfaces such as smaller dunes or sorted bedforms occur extensively on the shelf and can overprint larger features, such as sediment banks or large moraines. The networks preferentially choosing the classification as *Large Ridge* might reduce the scores in the evaluation metrics but do not technically produce a wrong interpretation, rather a partial one. In some instances (e.g. the on shelf edge, see Figure 10), model predictions have dissected longer wavelength dunes (i.e. large underlying landforms) interpreting them partially as *Large Ridges* and
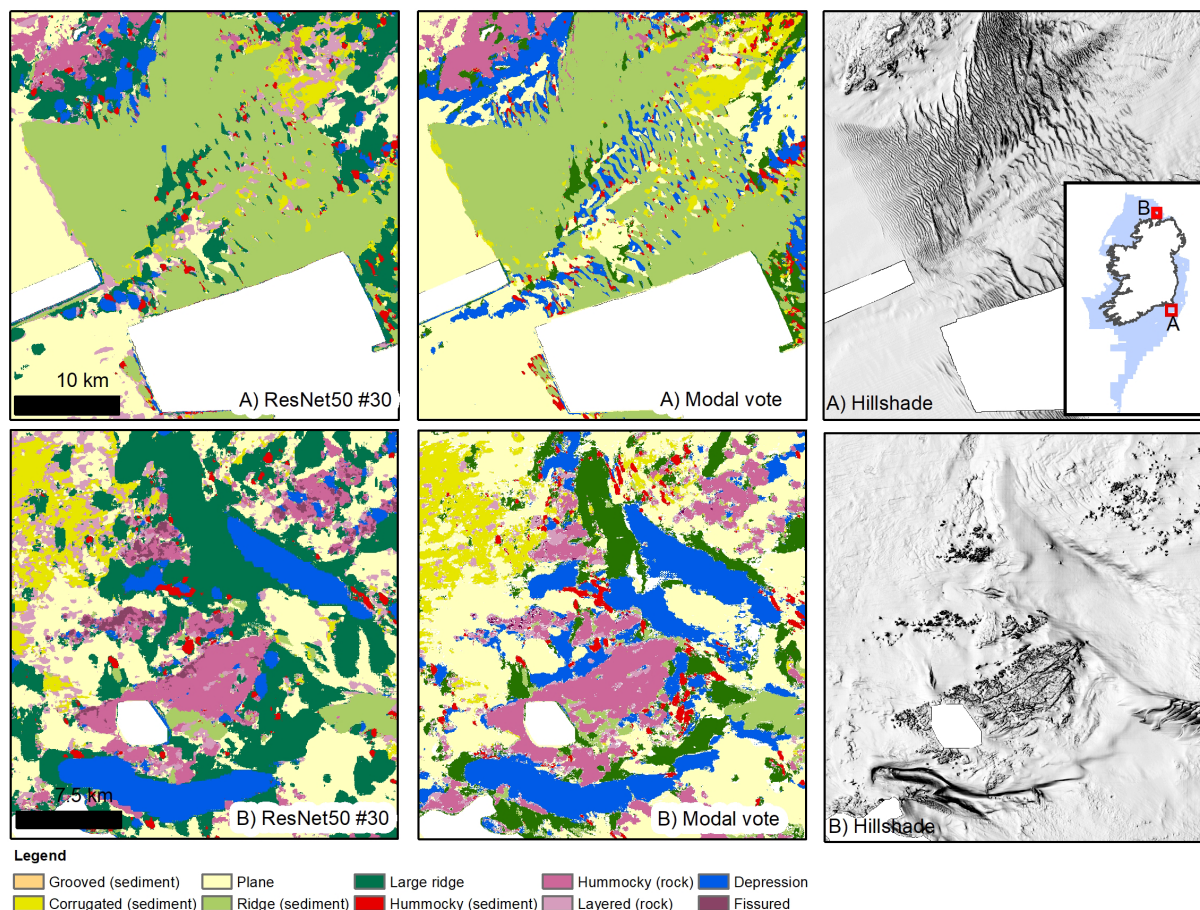
FIGURE 7
Results from best performing ResNet50 model (#30 – bathymetry and hillshades only) and the modal vote map. **(A)** Model #30 classifies correctly the extent of the large dune field, although once again the *Layered (rock)* class is erroneously predicted in liminal places. Both **(A)** and **(B)** show well the higher detail provided by the modal vote map, for example in **(A)** singular dune ridges are mapped correctly at the centre of the field, while for model #30 they are generalised with the surrounding flat or depressed terrain.

partially as *Corrugated (sediment)*, where the superficial sorted bedforms are more pronounced. Class prioritisation seems to be dependent on the way the model has learned the classes and boundaries, which in turn depends on adjusted weights and biases the model has learned during model training. However, understanding the individual activations and the internal workings of the neural network would require a study of class activation maps or the visualisation of deconvolutional layers (Noh et al., 2015).

Finally, a third type of recurring errors is connected to an inherent problem of the input layers: namely artefacts. MBES data can present many type of artefacts mostly caused by the limits of the instrument, the motions of the survey vessel (dynamic systematic errors), poor tidal or water sound velocity control causing vertical shifts and sound refraction. These artefacts are difficult to eliminate completely and a common obstacle in automated marine mapping (Lecours et al., 2017). Artefacts are recurrent in the extensive INFOMAR MBES bathymetry dataset, which is a combination of data from hundreds of different surveys with an array of vessels and survey operators, acquired with

different (improving) instrumentation, in the space of about 25 years. The topographic variability introduced may consist in pixelation (salt-and-pepper effects), undulation along the swath, striping effects and cliff-like edges, and the vertical difference is often comparable with real features at seabed (e.g. megaripples or furrowing) (Figures 9C, D). Additionally, our hillshades are particularly susceptible to this kind of "topographic noise", as they are vertically exaggerated to enhance the visibility of faint terrain patterns, which diminishes considerably their effectiveness. While a study of the effect of artefacts was outside the scope of this paper, it is reasonable to affirm that much stronger predictions can be achieved with a "cleaner" dataset.

4.2 Habitat mapping applications

Morphological maps provide the backbone for seabed habitat mapping studies, with classifications commonly obtained using semi-automated techniques as OBIA, BTM or other GIS tools (Harris et al., 2014; Goes et al., 2019; Linklater et al., 2019; Arosio
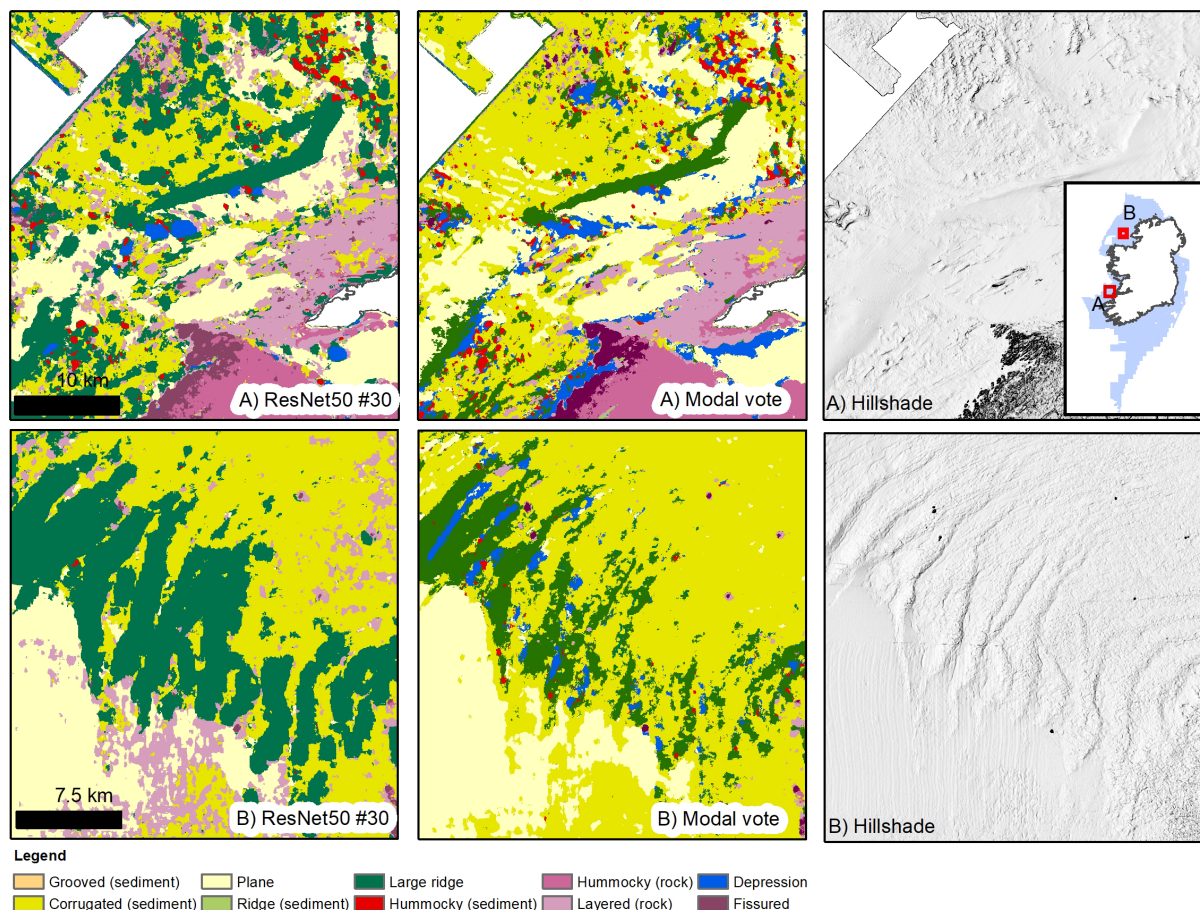
FIGURE 8
Results from best performing ResNet50 model (#30 – bathymetry and hillshades only) and the modal vote map. **(A)** this inset shows the overinterpreted *Depressions* for the modal vote map adjacent to the rocky outcrops, possibly caused by the BPI layers and totally absent for model #30. **(B)** model #30 correctly identifies the series of moraines in Donegal Bay, while the modal vote map produces a result which is a mixture of textural interpretation (corrugated seabed over the moraines) and larger features interpretation. The bathymetry artefacts that cover the otherwise featureless seabed in the southern portion of the inset have caused misinterpretations in both models; in particular model #30 shows again the confusion in predicting the location of *Layered (rock)*, assigning the artefacts that value.

et al., 2021) that segment the seabed in discrete parcels subsequently classified on the basis of pixel group statistics or geometrical characteristics. While grounded on mathematical rules and granting replicability, these techniques lack flexibility (e.g. how to treat morphological exceptions or near-isomorphisms) and require a good measure of engineering. Moreover, rules applied in one seabed region do not necessarily work elsewhere, so each dataset might need to be treated differently. On the contrary, FCNNs can provide the flexibility needed to capture any instance of discrete landforms or terrain textures without requiring *ad hoc* segmentation protocols (OBIA) or formulation of classification rulesets (BTM).

A semi-quantitative assessment of the effectiveness of the FCNN predictions for habitat mapping can be made comparing bedrock or sediment texture substrates to existing maps. We compared the predicted FCNN "bedrock" classes (*Fissured, Layered* and *Hummocky (rock)*), with the bedrock substrate layer produced by INFOMAR and available on the INFOMAR portal

(INFOMAR, 2022). In Figure 11 we take the models with best scores in "bedrock" prediction (model #29 and the modal vote map) and overlap the INFOMAR layer. We limit the comparison area to a subsection of the entire dataset (indicated in Figure 11C), as parts of the INFOMAR layer are mapped at very low resolution (e.g. the areas in Figure 11C pointed by the red arrows), introducing further deviations, and in other zones the *Hummocky (rock)* class includes also rough glacial till substrate. The best comparison is provided by the modal vote map, with a total bedrock area of 2721 km^2 (INFOMAR = 2336 km^2) and an overlap of 77%. Model #29 has a slightly better overlap (~78%), but has also a larger area mapped as rock (3276 km^2). Most of this excess bedrock is caused by misinterpretation of *Layered (rock)* (Figure 11D), which is over-represented in the model (F1 score 0.58). These numbers have to be taken with a pinch of salt, the mapping approaches are different (e.g. in the INFOMAR dataset the fissures in the bedrock outcrops are given another class), at a slightly different resolution and using different input layers (the INFOMAR map relies abundantly on
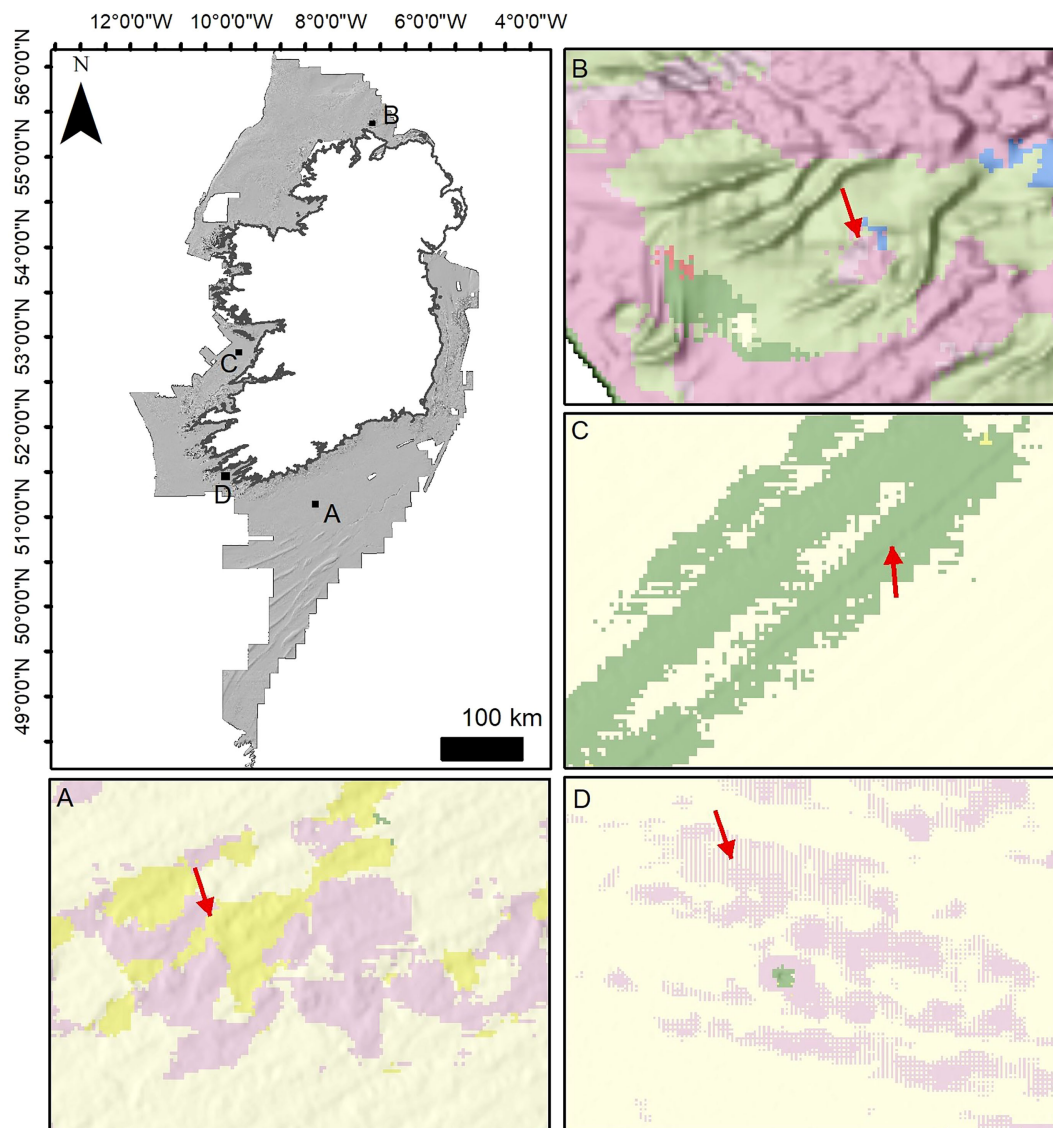
FIGURE 9
Types of error and ambiguities encountered in the maps. **(A, B)** sharp class transitions/interfaces and misclassification due to the ambiguous nature of the terrain. This is especially evident in **(B)**, where the dunes cross a rugged bedrock terrain with a similar signature. **(C)** bathymetry artefacts caused by MBES swath merging and correction that leads to a striping effect (misclassified as *Large Ridge*). **(D)** bathymetry artefacts and pixelation produced by low quality older MBES data.

backscatter data). Nonetheless, there is a broad agreement between the two, and the FCNNs consistently predict bedrock where it has been effectively mapped (see Figure 11). Moreover they give further information on the texture of the bedrock, which can be useful for habitat predictions (Novaczek et al., 2017). A similar comparison can be made with submarine dune fields. In Figure 12 we compare the general location of submarine dune ridges extracted using semi-automated techniques and checked manually (Arosio et al., 2023) with the class *Ridge (sediment)* in the best performing models (#12 and the modal vote map). Once more the results show an overall agreement, with *Ridge (sediment)* predictions corresponding with dune field areas (Figures 12A, B). In some places the FCNN is more efficient in identifying subtler ridges (e.g. Figure 12F), however in

parts the related classes *Large Ridge* and *Corrugation (sediment)* were preferentially selected (e.g. Figures 12E, F). The models show higher levels of confusion in the presence of trochoidal dunes (Figure 12D) that are often misclassified as *Fissured (rock)* indicating that the labelling is not effective enough to train for this particular morphological distinction.

4.3 Final considerations

This exploratory study has shown that FCNNs have considerable potential for the creation of large scale seabed landforms and terrain textures map, and that even with relatively
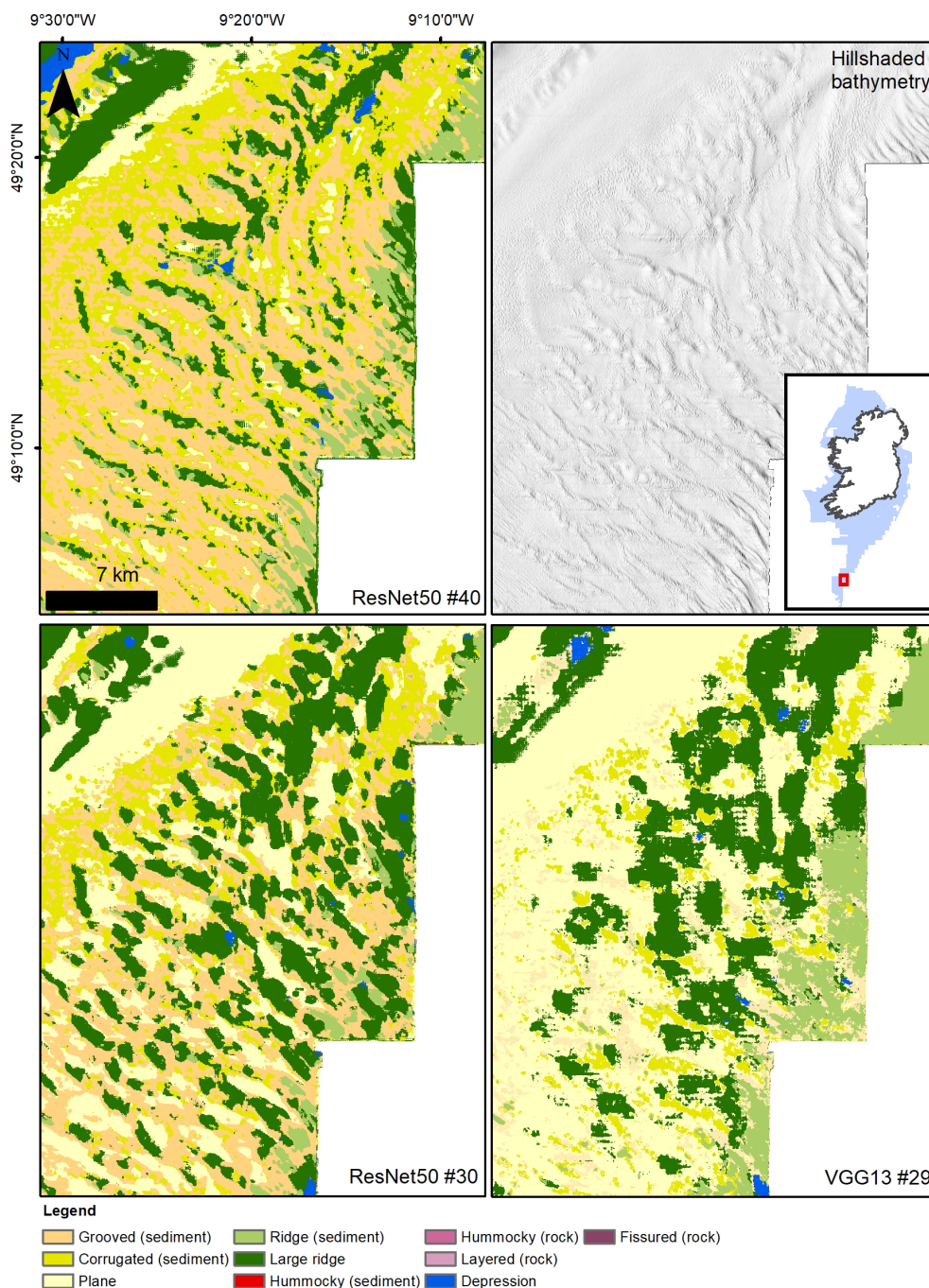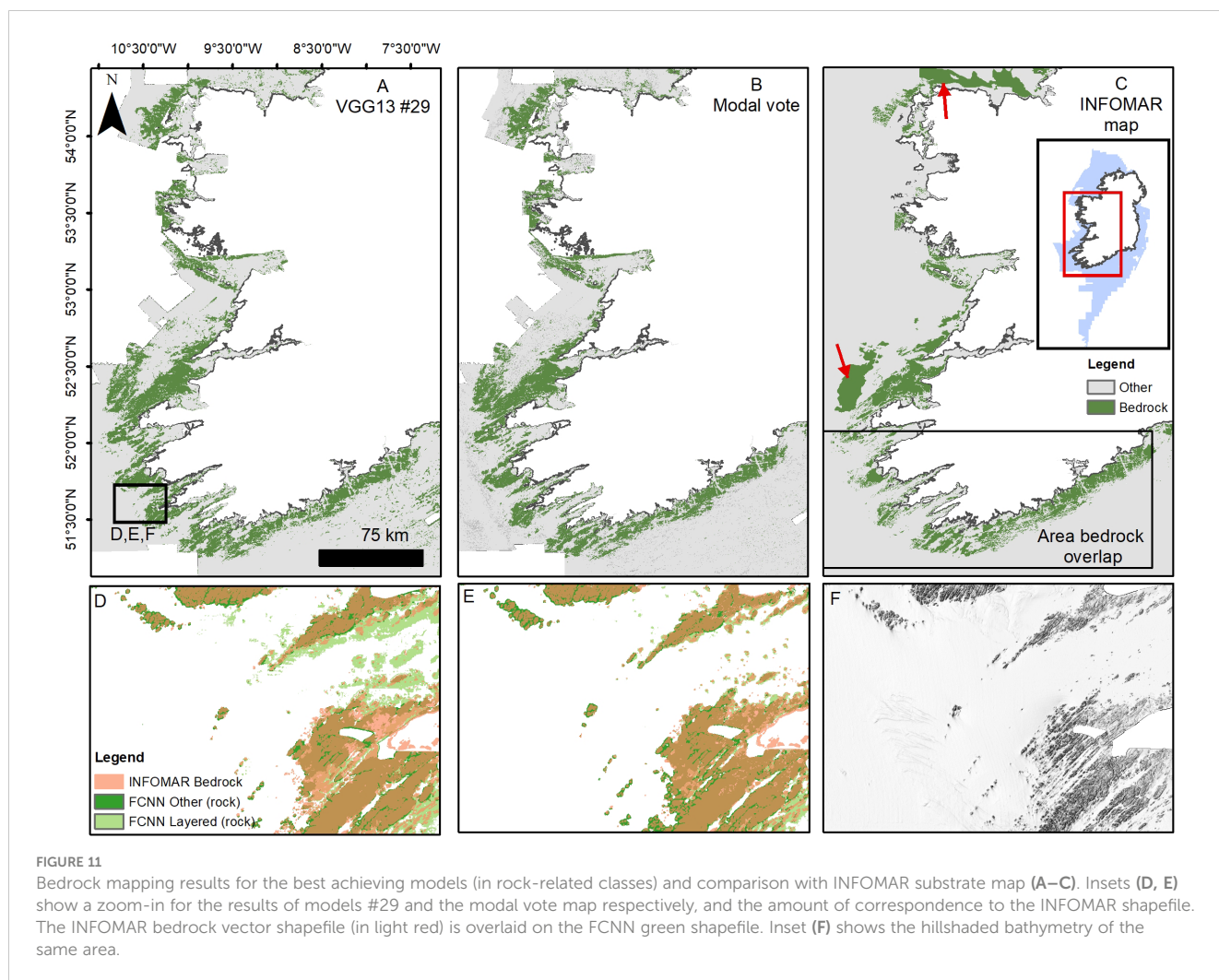
FIGURE 10
Representation of the different classification styles adopted by the networks when dealing with "nested" bedforms with different dimensions (large dunes, finer megaripples and sorted bedforms) using discrete and non-overlapping classes. All models map the most visible class in an area, reaching different competing results. Models #30 and #40 produce good alternative representations, while model #29 fails to reach a proper depiction of the area.

modest human input the results can be satisfactory. A clear semantic class definition and label delineation (including numerous boundary cases) will improve the accuracy of the classification, while a more rigorous consistency in mapping scale will most likely reduce ambiguity. Our results show that the

optimisation of derivative selection helps the model outputs, and a combination of hillshaded layers contribute substantially to prediction improvements. Further insights on the contribution of each layer could be obtained using techniques based on feature importance, as saliency maps (e.g. Simonyan et al., 2014). The

FIGURE 11
Bedrock mapping results for the best achieving models (in rock-related classes) and comparison with INFOMAR substrate map **(A–C)**. Insets **(D, E)** show a zoom-in for the results of models #29 and the modal vote map respectively, and the amount of correspondence to the INFOMAR shapefile. The INFOMAR bedrock vector shapefile (in light red) is overlaid on the FCNN green shapefile. Inset **(F)** shows the hillshaded bathymetry of the same area.

ensemble voting map, which constituted the best outcome of these experiments, clearly shows the utility of using learnt biases on different subsets of input data, and that assembling predictions from several 'weak' learners outperforms a single 'expert' network, which is the premise of ensemble learning (Ganaie et al., 2022). For further work, several FCNNs could be trained concurrently on different subsets of input data, and a loss could be calculated based on the confidence of individual networks (Goyal et al., 2020; Zhou et al., 2021). The latter is akin to several Decision trees in a Random Forest in classical machine learning (Cutler et al., 2012).

From a habitat mapper's perspective, the use of FCNNs can be successfully applied to seabed maps for morphological characterisation, and very good results and flexibility can be achieved provided the model is well trained and furnished with clean data. Very large scale mapping endeavours, as that presented in Harris et al. (2014), could be easily replicated and improved upon using FCNNs. Moreover previously trained models could be applied on the new datasets that are being collected and gathered for Seabed 2030. If a sufficient volume of labelled classes is cooperatively assembled in a "dictionary" and made publicly available, it could be used by the community to predict morphological classes across

different datasets, improving upon map objectivity and inter-comparison. The time invested in creating such a dictionary would be considerable but worthwhile, as the FCNN method will be eventually better, quicker and easily repeatable compared to semi-automated or manual digitisations. We shared our labelled dataset on GitHub (https://github.com/BrandonHobley/geomorph_deep) as a starting point. While discrete computer power it is necessary, the code is open source and requires a relatively basic level of coding expertise to be run, allowing for a widespread adoption.

FCNNs have also their significant drawbacks. Firstly, they are essentially a blackbox whose internal workings are not fully understood. Secondly, labelling and training at one determinate pixel resolution is most likely not transferable to a different one. So having mapped at 25m/pixel our dataset is probably ineffective to map at 2m/pixel, and more *ad hoc* labelling will be required. Finally, at this stage of sophistication, FCNNs fail to recognize complex geomorphological processes, especially in cases of isomorphism, so human intervention is still required. This limitation is also caused by the input types themselves, as bathymetry-derived raster data alone are often insufficient (for human geomorphologists too!) to
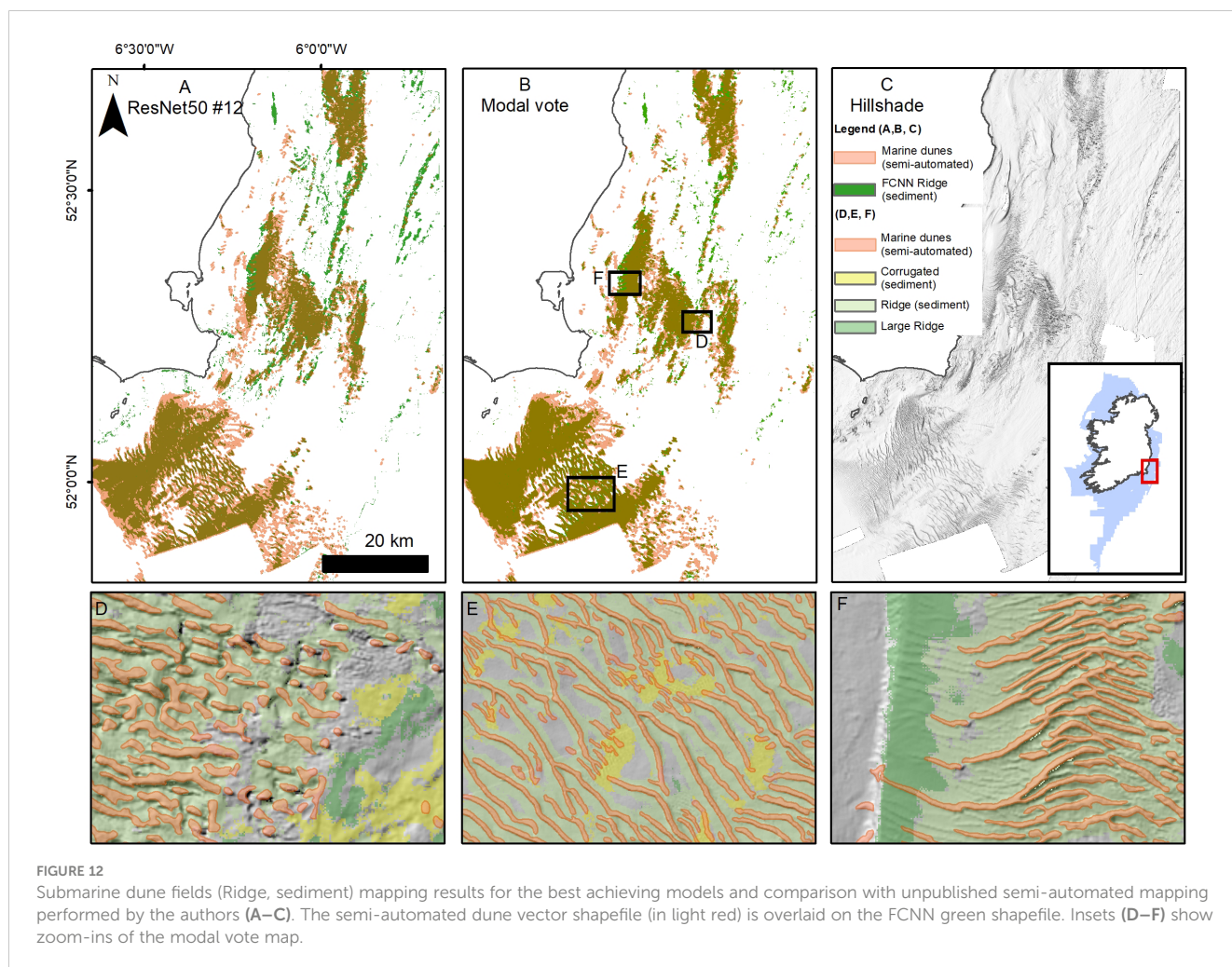
FIGURE 12
Submarine dune fields (Ridge, sediment) mapping results for the best achieving models and comparison with unpublished semi-automated mapping performed by the authors **(A–C)**. The semi-automated dune vector shapefile (in light red) is overlaid on the FCNN green shapefile. Insets **(D–F)** show zoom-ins of the modal vote map.

unequivocally identify seabed landforms. Only when different types of datasets (seismic lines, ground-truthing etc.) can be included in the predictions, will machine learning be useful for more complex seabed geological interpretations.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

RA conceived the original idea, performed the labelling and analysed the results. BH wrote the neural networks and run the models. RA and BH developed the approach and wrote the manuscript. AL and AW obtained the funding and reviewed the manuscript. FS, LC, and TF reviewed the manuscript. All authors provided useful feedback and helped shape the research.

All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2023.1228867/full#supplementary-material

References

Arosio, R., Mitchell, P., Hawes, J., Bolam, S., Benson, L., and Sperry, J. (2021). *Small island developing states (SIDS) and the sea: creating high resolution habitat maps to support effective marine management in St. Lucia* (Vienna: Presented at the EGU). doi: 10.5194/egusphere-egu21-102

Arosio, R., Wheeler, A. J., Sacchetti, F., Guinan, J., Conti, L. A., Furey, T., et al. (2023). *The NOMANS_TIF map: ireland's first complete shallow seabed geomorphology map* (Saint-Gilles-Les-Bains: Presented at the GeoHab 2023, La Réunion). doi: 10.5281/zenodo.7890332

Barrett, A. M., Balme, M. R., Woods, M., Karachalios, S., Petrocelli, D., Joudrier, L., et al. (2022). NOAH-h, a deep-learning, terrain classification system for Mars: results for the ExoMars rover candidate landing sites. *Icarus* 371, 114701. doi: 10.1016/j.icarus.2021.114701

Benetti, S., Dunlop, P., and Cofaigh, C.Ó. (2010). Glacial and glacially-related features on the continental margin of northwest Ireland mapped from marine geophysical data. *J. Maps* 6, 14–29. doi: 10.4113/jom.2010.1092

Blaschke, T., Hay, G. J., Kelly, M., Lang, S., Hofmann, P., Addink, E., et al. (2014). Geographic object-based image analysis – towards a new paradigm. *ISPRS J. Photogrammetry Remote Sens.* 87, 180–191. doi: 10.1016/j.isprsjprs.2013.09.014

Brown, C. J., Smith, S. J., Lawton, P., and Anderson, J. T. (2011). Benthic habitat mapping: a review of progress towards improved understanding of the spatial ecology of the seafloor using acoustic techniques. *Estuarine Coast. Shelf Sci.* 92, 502–520. doi: 10.1016/j.ecss.2011.02.007

Buscombe, D., Wernette, P., Fitzpatrick, S., Favela, J., Goldstein, E. B., and Enwright, N. M. (2023). A 1.2 billion pixel human-labeled dataset for data-driven classification of coastal environments. *Sci. Data* 10, 46. doi: 10.1038/s41597-023-01929-2

Conti, L. A., Lim, A., and Wheeler, A. J. (2019). High resolution mapping of a cold water coral mound. *Sci. Rep.* 9, 1016. doi: 10.1038/s41598-018-37725-x

Creane, S., Coughlan, M., O'Shea, M., and Murphy, J. (2022). Development and dynamics of sediment waves in a complex morphological and tidal dominant system: southern Irish Sea. *Geosciences* 12, 431. doi: 10.3390/geosciences12120431

Cutler, A., Cutler, D. R., and Stevens, J. R. (2012). "Random forests," in *Ensemble machine learning.* Eds. C. Zhang and Y. Ma (New York, NY: Springer New York), 157–175. doi: 10.1007/978-1-4419-9326-7_5

de Oliveira, N., Bastos, A. C., da Silva Quaresma, V., and Vieira, F. V. (2020). The use of benthic terrain modeler (BTM) in the characterization of continental shelf habitats. *Geo-Mar Lett.* 40, 1087–1097. doi: 10.1007/s00367-020-00642-y

de Oliveira, L. M. C., Lim, A., Conti, L. A., and Wheeler, A. J. (2021). 3D classification of cold-water coral reefs: a comparison of classification techniques for 3D reconstructions of cold-water coral reefs and seabed. *Front. Mar. Sci.* 8. doi: 10.3389/fmars.2021.640713

Diesing, M., Green, S. L., Stephens, D., Lark, R. M., Stewart, H. A., and Dove, D. (2014). Mapping seabed sediments: comparison of manual, geostatistical, object-based image analysis and machine learning approaches. *Continental Shelf Res.* 84, 107–119. doi: 10.1016/j.csr.2014.05.004

Diesing, M., and Thorsnes, T. (2018). Mapping of cold-water coral carbonate mounds based on geomorphometric features: an object-based approach. *Geosciences* 8, 34. doi: 10.3390/geosciences8020034

Dove, D., Bradwell, T., Carter, G., Cotteril, C., Gafeira, J., Green, S., et al. (2016). *Seabed geomorphology: a two-part classification system (Marine geosciences programme open report no. OR/16/001)* (Edinburgh: British Geological Survey).

Dove, D., Nanson, R., Bjarnadóttir, L. R., Guinan, J., Gafeira, J., Post, A., et al. (2020). *A two-part seabed geomorphology classification scheme (v.2); part 1: morphology features glossary* (Zenodo). doi: 10.5281/ZENODO.4075248

Foroutan, M., and Zimbelman, J. R. (2017). Semi-automatic mapping of linear-trending bedforms using 'Self-organizing maps' algorithm. *Geomorphology* 293, 156–166. doi: 10.1016/j.geomorph.2017.05.016

Ganaie, M. A., Hu, M., Malik, A. K., Tanveer, M., and Suganthan, P. N. (2022). Ensemble deep learning: a review. *Eng. Appl. Artif. Intell.* 115, 105151. doi: 10.1016/j.engappai.2022.105151

Gazis, I. Z., Schoening, T., Alevizos, E., and Greinert, J. (2018). Quantitative mapping and predictive modeling of Mn nodules' distribution from hydroacoustic and optical AUV data linked by random forests machine learning. *Biogeosciences* 15, 7347–7377. doi: 10.5194/bg-15-7347-2018

Giglio, C., Benetti, S., Sacchetti, F., Lockhart, E., Hughes Clarke, J., Plets, R., et al. (2022). A Late Pleistocene channelized subglacial meltwater system on the Atlantic continental shelf south of Ireland. *Boreas*, 51, 118–135. doi: 10.1111/bor.12536

Goes, E. R., Brown, C. J., and Araújo, T. C. (2019). Geomorphological classification of the benthic structures on a tropical continental shelf. *Front. Mar. Sci.* 6. doi: 10.3389/fmars.2019.00047

Goyal, M., Oakley, A., Bansal, P., Dancey, D., and Yap, M. H. (2020). Skin lesion segmentation in dermoscopic images with ensemble deep learning methods. *IEEE Access* 8, 4171–4181. doi: 10.1109/ACCESS.2019.2960504

Harris, P. T., and Baker, E. (2020). *Seafloor geomorphology as benthic habitat* (Elsevier). doi: 10.1016/C2017-0-02139-0

Harris, P. T., Macmillan-Lawler, M., Rupp, J., and Baker, E. K. (2014). Geomorphology of the oceans. *Mar. Geology* 352, 4–24. doi: 10.1016/j.margeo.2014.01.011

He, K., Gkioxari, G., Dollar, P., and Girshick, R. (2017). "Mask R-CNN," 2017 IEEE International Conference on Computer Vision (ICCV). (Venice, Italy) 2980–2988. doi: 10.1109/ICCV.2017.322

Hobley, B., Arosio, R., French, G., Bremner, J., Dolphin, T., and Mackiewicz, M. (2021). Semi-supervised segmentation for coastal monitoring seagrass using RPA imagery. *Remote Sens.* 13, 1741. doi: 10.3390/rs13091741

Ierodiaconou, D., Schimel, A. C. G., Kennedy, D., Monk, J., Gaylard, G., Young, M., et al. (2018). Combining pixel and object based image analysis of ultra-high resolution multibeam bathymetry and backscatter for habitat mapping in shallow marine waters. *Mar. Geophys Res.* 39, 271–288. doi: 10.1007/s11001-017-9338-z

INFOMAR (2022). Seabed sediment classification Irish waters WGS84 shapefile.

Ismail, K., Huvenne, V. A. I., and Masson, D. G. (2015). Objective automated classification technique for marine landscape mapping in submarine canyons. *Mar. Geology* 362, 17–32. doi: 10.1016/j.margeo.2015.01.006

Juliani, C. (2019). Automated discrimination of fault scarps along an Arctic mid-ocean ridge using neural networks. *Comput. Geosciences* 124, 27–36. doi: 10.1016/j.cageo.2018.12.010

Keohane, I., and White, S. (2022). Chimney identification tool for automated detection of hydrothermal chimneys from high-resolution bathymetry using machine learning. *Geosciences* 12, 176. doi: 10.3390/geosciences12040176

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Commun. ACM* 60, 84–90. doi: 10.1145/3065386

Lecours, V., Devillers, R., Edinger, E. N., Brown, C. J., and Lucieer, V. L. (2017). Influence of artefacts in marine digital terrain models on habitat maps and species distribution models: a multiscale assessment. *Remote Sens. Ecol. Conserv.* 3, 232–246. doi: 10.1002/rse2.49

Lecours, V., Dolan, M. F. J., Micallef, A., and Lucieer, V. L. (2016). A review of marine geomorphometry, the quantitative study of the seafloor. *Hydrol. Earth Syst. Sci.* 20, 3207–3244. doi: 10.5194/hess-20-3207-2016

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539

Leitão, P. J., Schwieder, M., Pötzschner, F., Pinto, J. R. R., Teixeira, A. M. C., Pedroni, F., et al. (2018). From sample to pixel: multi-scale remote sensing data for upscaling

aboveground carbon data in heterogeneous landscapes. *Ecosphere* 9, e02298. doi: 10.1002/ecs2.2298

Lin, M., Chen, Q., and Yan, S. (2014). Network in network. *International Conference on Learning Representations (ICLR) (Banff)*. Available at: http://arxiv.org/abs/1312.4400.

Linklater, M., Ingleton, T. C., Kinsela, M. A., Morris, B. D., Allen, K. M., Sutherland, M. D., et al. (2019). Techniques for classifying seabed morphology and composition on a subtropical-temperate continental shelf. *Geosciences* 9, 141. doi: 10.3390/geosciences9030141

Lockhart, E. A., Scourse, J. D., Praeg, D., Van Landeghem, K. J. J., Mellett, C., Saher, M., et al. (2018). A stratigraphic investigation of the celtic Sea megaridges based on seismic and core data from the Irish-UK sectors. *Quaternary Sci. Rev.* 198, 156–170. doi: 10.1016/j.quascirev.2018.08.029

Long, J., Shelhamer, E., and Darrell, T. (2014). *Fully convolutional networks for semantic segmentation* (Boston, MA, USA: IEEE Conference on Computer Vision and Pattern Recognition (CVPR)), 3431-3440. doi: 10.1109/CVPR.2015.7298965

Lundine, M. A., Brothers, L. L., and Trembanis, A. C. (2023). Deep learning for pockmark detection: implications for quantitative seafloor characterization. *Geomorphology* 421, 108524. doi: 10.1016/j.geomorph.2022.108524

Ma, Z., and Mei, G. (2021). Deep learning for geological hazards analysis: data, models, applications, and opportunities. *Earth-Science Rev.* 223, 103858. doi: 10.1016/j.earscirev.2021.103858

McClinton, T. J., White, S. M., and Sinton, J. M. (2012). Neuro-fuzzy classification of submarine lava flow morphology. *Photogrammetric Eng. Remote Sens.* 78, 605–616. doi: 10.14358/PERS.78.6.605

Micallef, A., Le Bas, T. P., Huvenne, V. A. I., Blondel, P., Hühnerbach, V., and Deidun, A. (2012). A multi-method approach for benthic habitat mapping of shallow coastal areas with high-resolution multibeam data. *Continental Shelf Res.* 39–40, 14–26. doi: 10.1016/j.csr.2012.03.008

Noh, H., Hong, S., and Han, B. (2015). *Learning deconvolution network for semantic segmentation* (Santiago, Chile: IEEE International Conference on Computer Vision (ICCV)), 1520-1528. doi: 10.1109/ICCV.2015.178

Novaczek, E., Devillers, R., Edinger, E., and Mello, L. (2017). High-resolution seafloor mapping to describe coastal denning habitat of a Canadian species at risk: Atlantic wolffish (Anarhichas lupus). *Can. J. Fish. Aquat. Sci.* 74, 2073–2084. doi: 10.1139/cjfas-2016-0414

Ó Cofaigh, C., Dunlop, P., and Benetti, S. (2012). Marine geophysical evidence for late pleistocene ice sheet extent and recession off northwest Ireland. *Quaternary Sci. Rev.* 44, 147–159. doi: 10.1016/j.quascirev.2010.02.005

Palafox, L. F., Hamilton, C. W., Scheidt, S. P., and Alvarez, A. M. (2017). Automated detection of geological landforms on Mars using convolutional neural networks. *Comput. Geosciences* 101, 48–56. doi: 10.1016/j.cageo.2016.12.015

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-Net: convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention – MICCAI 2015, lecture notes in computer science*. Eds. N.

Navab, J. Hornegger, W. M. Wells and A. F. Frangi (Springer International Publishing, Cham), 234–241. doi: 10.1007/978-3-319-24574-4_28

Rubanenko, L., Perez-Lopez, S., Schull, J., and Lapotre, M. G. A. (2021). Automatic detection and Segmentation of barchan dunes on Mars and earth using a convolutional neural network. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* 14, 9364–9371. doi: 10.1109/JSTARS.2021.3109900

Simonyan, K., Vedaldi, A., and Zisserman, A. (2014). Deep inside convolutional networks: visualising image classification models and saliency maps. *Workshop at International Conference on Learning Representations*. doi: 10.48550/arXiv.1312.6034

Simonyan, K., and Zisserman, A. (2015). *Very deep convolutional networks for Large-scale image recognition* (San Diego: International Conference on Learning Representations (ICLR)). Available at: https://arxiv.org/abs/1409.1556.

Smith, B., and Mark, D. M. (2003). Do mountains exist? towards an ontology of landforms. *Environ. Plann B Plann Des.* 30, 411–427. doi: 10.1068/b12821

Summers, G., Lim, A., and Wheeler, A. J. (2021). A scalable, supervised classification of seabed sediment waves using an object-based image analysis approach. *Remote Sens.* 13, 2317. doi: 10.3390/rs13122317

Tarvainen, A., and Valpola, H. (2017). "Mean teachers are better role models: weight-averaged consistency targets improve semi-supervised deep learning results," in *31st Conference on Neural Information Processing System. Presented at the NIPS*, Long Beach, USA.

Valentine, A. P., Kalnins, L. M., and Trampert, J. (2013). Discovery and analysis of topographic features using learning algorithms: a seamount case study. *Geophysical Res. Lett.* 40, 3048–3054. doi: 10.1002/grl.50615

Van Landeghem, K. J. J., Uehara, K., Wheeler, A. J., Mitchell, N. C., and Scourse, J. D. (2009). Post-glacial sediment dynamics in the Irish Sea and sediment wave morphology: data–model comparisons. *Continental Shelf Res.* 29, 1723–1736. doi: 10.1016/j.csr.2009.05.014

Walbridge, S., Slocum, N., Pobuda, M., and Wright, D. J. (2018). Unified geomorphological analysis workflows with benthic terrain modeler. *Geosciences* 8, 94. doi: 10.3390/geosciences8030094

Wang, Y., Di, K., Xin, X., and Wan, W. (2017). Automatic detection of Martian dark slope streaks by machine learning using HiRISE images. *ISPRS J. Photogrammetry Remote Sens.* 129, 12–20. doi: 10.1016/j.isprsjprs.2017.04.014

Wilhelm, T., Geis, M., Püttschneider, J., Sievernich, T., Weber, T., Wohlfarth, K., et al. (2020). DoMars16k: a diverse dataset for weakly supervised geomorphologic analysis on Mars. *Remote Sens.* 12, 3981. doi: 10.3390/rs12233981

Yasir, M., Rahman, A. U., and Gohar, M. (2021). Habitat mapping using deep neural networks. *Multimedia Syst.* 27, 679–690. doi: 10.1007/s00530-020-00695-0

Zhou, T., Lu, H., Yang, Z., Qiu, S., Huo, B., and Dong, Y. (2021). The ensemble deep learning model for novel COVID-19 on CT images. *Appl. Soft Computing* 98, 106885. doi: 10.1016/j.asoc.2020.106885