Habitat classification using convolutional neural networks and multitemporal multispectral aerial imagery

Sara Pérez-Carabaza[a,*] **Oisín Boydell**[a] **and Jerome O'Connell**[b]
[a]University College Dublin, CeADAR, Ireland's Centre for Applied AI, Dublin, Ireland
[b]ProvEye, Dublin, Ireland

Abstract. The monitoring of threatened habitats is a key objective of European environmental policies. Due to the high cost of current field-based habitat mapping techniques, there is keen interest in proposing solutions that can reduce cost through increased levels of automation. Our study aims to propose a habitat mapping solution that benefits both from the merits of convolutional neural networks (CNNs) for image classification tasks, as well as from the high spatial, spectral, and multitemporal unmanned aerial vehicle image data, which shows great potential for accurate vegetation classification. The proposed CNN-based method uses multitemporal multispectral aerial imagery for the classification of threatened coastal habitats in the Maharees (Ireland) and shows a high level of classification accuracy. © *2021 Society of Photo-Optical Instrumentation Engineers (SPIE)* [DOI: 10.1117/1.JRS.15.042406]

Keywords: habitat mapping; unmanned aerial vehicle imagery; multitemporal imagery; convolutional neural networks.

Paper 210199SS received Mar. 31, 2021; accepted for publication Jun. 18, 2021; published online Jul. 2, 2021.

1 Introduction

The European Union Habitat Directive,[1] which addresses the conservation of natural habitats and wild fauna and flora, directs EU member states to take measures in order to maintain the favorable conservation status of threatened habitats. Namely, the EU requires member states to periodically produce maps for change detection and conservation status assessments of priority habitat types specified in the EU Habitats Directive.[1] The Habitats Directive ensures the conservation of a wide range of rare, threatened, or endemic animal and plant species listed in the directive's annexes (Annex I covers habitats; Annexes II to V, species). The coastal habitat under study is one of the threatened and protected habitats listed in Annex I of the Habitats Directive, which are often simply referred as Annex I habitats. At present, Ireland reports the conservation status of its threatened habitats based on ecological field data. The current field-based mapping and assessment methodology requires significant time and financial resources. The development of automatic habitat mapping tools would allow one to considerably reduce the financial and man-power cost, offering a great benefit that has motivated this work. This work looks to leverage the potential offered through the rapid advances in unmanned aerial vehicle (UAV) platforms along with deep learning techniques to address the classification of threatened coastal habitats in Ireland.

The outstanding results that convolutional neural networks (CNNs) have achieved in other domains have motivated the remote sensing community to apply them in image classification problems dealing with satellite or aerial imagery. Despite the challenges involved in the application of CNNs to remote sensing, in recent years deep learning approaches have achieved significant breakthroughs in remotely sensed image classification, offering exciting opportunities for research.[2] For instance, CNN-based solutions have been successfully applied for the detection of plant species,[3] crop classification,[4] or vegetation mapping,[5] showing in many instances better performance than object-based image analysis (OBIA) methods[3] or traditionally machine

*Address all correspondence to Sara Perez-Carabaza, sara.perezcarabaza@ucd.ie

learning methods.[4,5] Contrary to traditional machine learning models or OBIA methods that are dependent on engineered features, CNNs have the ability to automatically extract the features. This end-to-end learning approach empowers CNNs ability to automatically learn contextual features from raw input images, making them highly effective for large-scale image recognition and semantic segmentation tasks.[2] Remote sensing images are georeferenced and may have high spectral dimensionality [multispectral (MS) or hyperspectral images] or multitemporal information (which is especially relevant in remote sensing tasks dealing with crops or vegetation). Therefore, as most of the deep learning models for image data related tasks are designed to work with multimedia [red–green–blue (RGB)] images commonly encountered in the computer vision literature, there is a need to adapt deep learning-based models to utilize MS and multitemporal information. Another common challenge in the application of deep learning to land mapping is the high cost and expert knowledge frequently required for obtaining labeled training data. Furthermore, while urban environments tend to have well-defined scene context, in several remote sensing problems including the coastal dune habitat of study in this work, the images present a potentially unlimited continuous space, where boundaries between classes are transitional, increasing the complexity of the image classification problems. For instance, the spatial borders of different coral reef classes are hard to separate due to their tendency to appear in groups, which makes its classification a challenging task that requires aid from marine biologists.[6]

Currently, there is a huge need for precise land cover information in various fields such as crop monitoring or conservation. Images from optical sensors mounted on UAVs can provide very detailed information in comparison to satellite acquired imagery thanks to their higher spatial resolution. In fact, UAVs are increasingly used as an efficient tool for rapid monitoring of land resources.[7] Multitemporal UAV imagery has the advantage of providing high spatial resolution as well as phenological information that has greater potential for accurate vegetation classification problems, such as crop classification.[5] The benefits that multitemporal information has shown in other remote sensing problems that involve vegetation such as crop classification,[8] in addition to the coastal dunes characteristics, which are a very dynamic environment changed constantly as a result of waves and wind, has motivated this work to propose a CNN-based model for habitat mapping that incorporates temporal information.

In summary, this work proposes a multitemporal MS CNN-based model that classifies with high accuracies several priority coastal habitats located in the Maharees, Ireland. The experimental analysis over the study site shows the benefits of including multitemporal aerial imagery for the habitat classification problem.

The remainder of this paper proceeds as follows. Section 2 provides a review of machine-learning solutions for related remote sensing problems. Section 3 begins with a description of the study site, the data collection process and finishes with an overview of CNNs. Section 4 presents the proposed deep learning-based model to classify four habitat types present in the Maharees coastal area. Section 5 evaluates the performance of the proposed multitemporal MS approach and compares it to an MS CNN-based model. Section 6 discusses about the results obtained by the proposed approach, its benefits, and possible improvements. Finally, Sec. 7 summarizes the main conclusions of this work and describes possible future research directions.

2 Related Work

In the habitat mapping state-of-the-art, we can find several recently solutions that propose the use of machine learning to deal with the need for cost-effective habitat mapping tools. These related works are listed in Table 1 along with their main characteristics: type of habitat mapping problem, type of remotely sensed imagery, type of data, the machine learning technique used, and whether the works exploit MS and temporal information.

Habitat mapping problems can be divided into the works that classify terrestrial[3,7,9–11] or marine habitats.[6,12–14] Among the first group, Kobler et al.[9] dealt with a forest habitat mapping problem where ten different habitats included in the Habitat Directive[1] are considered. Rezaee et al.[11] tackled a wetland classification problem where eight different wetland classes specified by the Canadian wetland classification system are mapped. Timm and McGarigal[10] addressed the

Table 1 Machine-learning solutions for habitat mapping.

Work	Problem	Imagery	Data type	Model	Multi-spectral	Temporal info.
9	Habitat classification	Satellite (IKONOS 2)	MS	DT	✓	
10	Habitat classification	Satellite (QuickBird)	MS, DEM	RF	✓	
11	Habitat classification	Satellite (RadpidEye)	MS	2D-CNN (Alexnet)		
3	Habitat classification	Satellite (Worldview-2)	RGB	2D-CNN (ResNet)		
7	Habitat classification	Aerial	RGB	FCN (UNet)		
12	Marine habitat classification	Underwater	Sonar	RF		
13	Marine habitat classification	Underwater	Sonar	2D-CNN		
14	Marine habitat classification	Satellite (WorldView-2)	MS	FCN	✓	
6	Marine habitat classification	Underwater	RGB	2D-CNN (DenseNet)		
This work	Habitat mapping	Aerial	MS	2D-CNN	✓	✓

classification of coastal dune and salt marsh ecosystems. Guirado et al.[3] proposed a deep learning-based solution for detection of *Ziziphus lotus* shrubs, a priority habitat under the Habitat Directive.[1] And Kattenborn et al.[7] addressed a binary canopy classification problem where herbaceous vegetation communities are mapped from aerial imagery. Among the marine habitat mapping problems, both Berthold et al.[13] and Diesing et al.[12] mapped seabed sediment habitats, considering four different sediment types. Finally, Yasir et al.[6] dealt with coral reef marine habitats, distinguishing among four coral types and four non-coral habitat types.

These habitat mapping models make use of a variety of imagery data types captured from satellites,[3,10,11,14] aerial vehicles,[7] or underwater sensors[6,7,13] by means of different sensor technologies such as RGB cameras,[3,6,7] high-resolution MS sensors,[10,11,14] or sound navigation and ranging (sonar).[13] For instance, Kattenborn et al.[7] and this work employed UAV aerial imagery, taking advantage of the big development that UAVs have experienced during the last decade, whose increased level of autonomy and decreased cost have facilitated their use in a wide range of applications. In addition, the models proposed by Refs. 10, 11, and 14 as well as this work employ MS imagery, which involves the acquisition of visible, near-infrared, and short-wave infrared images in several broad wavelength bands. The consideration of near-infrared (NIR) spectral bands in combination with the visible bands allows to increase the interclass variability among different types of vegetation types or crops. Therefore, MS imagery is used by many of the reviewed habitat mapping works[9–11,14] as well as in related crop classification problems.[4,8,15,16] In the case of the habitat mapping model proposed by Timm and McGarigal[10] for salt marsh and coastal dunes habitats, it combines the MS imagery with the information provided by a digital elevation model (DEM) to deal with the complex classification of this type of habitats.

The machine learning-based techniques employed for habitat mapping can be divided among those that use traditional machine-learning-based algorithms[9,10,12] and the ones based on deep learning models.[3,6,7,11,13,14] Among the first group, Diesing et al.[12] proposed the use of the random forest (RF) algorithm for marine habitat mapping using features extracted from sonar images. Timm and McGarigal[10] made use of an RF model to classify the features extracted from MS satellite imagery in combination with DEM data. Kobler et al.[9] proposed a decision tree (DT)-based method that assigns a label to each pixel of a satellite image considering the spectral values of a kernel (window of adjacent pixels) and its adjacency-event matrix. The above-mentioned studies are mainly based on low-level, manually designed features (e.g., textures and

roughness) as input to machine-learning classifiers, which require significant domain expertise and can be prone to demonstrate poor performance in obtaining high-level representative features.[5] On the contrary, deep learning provides end-to-end learning and has shown on multiple occasions better performance than traditional-machine learning methods (e.g., RF) on related problems such as crop monitoring.[14,17] Among the deep learning-based solutions for habitat mapping, we can distinguish two approaches: the use of CNNs architectures for image classification or the use of fully connected networks (FCN) for pixel-wise segmentation. Both approaches typically use patch-based methods where the patches extracted from the georeferenced mosaics are used for training the deep learning models. In CNN-based models for remote sensing, each patch (typically of smaller size) has assigned a unique label, and the deep learning models are trained to predict a class, or a probability distribution over all classes, for each input image patch. However, in FCN-based approaches, each pixel in the patch can have a different label and thus the FCN-based models return the predicted labels for all pixels in the patch. FCN-based approaches are proposed by Kattenborn et al.[7] for terrestrial habitat mapping and by Li et al.[14] for marine habitat mapping. CNNs-based approaches are proposed in Refs. 3, 6, 11, and 13 for terrestrial or marine habitat mapping, where Refs. 3, 6, and 11 employ transfer learning through the use of a pretrained network indicated between brackets in the model column of Table 1. This work proposes a 2D-CNN-based model that classifies multitemporal MS aerial imagery. To the authors' knowledge, this work presents the first deep learning solution for habitat mapping that leverages the spatial, MS, and temporal information of aerial imagery. However, in the state-of-the-art of related vegetation classification problems such as crop classification, we can find several recent works that exploit the temporal information of remotely sensed imagery.[4,5,8,15,17] When compared with monotemporal classification, the incorporation of temporal information may boost the model accuracy, enhancing the interclass variability thanks to the characteristic growth patterns of the target vegetation.[5,17] This has motivated us to propose a CNN-based model for habitat mapping that incorporates temporal information. For instance, among the deep learning solutions for crop classification that take into account temporal information, Rußwurm and Korner[15] and Ndikumana et al.[17] made use of recurrent neural networks for learning the temporal and spectral features of pixel-wise time series, achieving high classification accuracies without exploiting the spatial information of the imagery. Another example is the 3D-CNN-based model proposed by Ji et al.[8] to extract the spatiotemporal features from multitemporal MS satellite images. Finally, both Benedetti et al.[4] and Feng et al.[5] proposed hybrid approaches that combine RNN architectures for learning temporal features with CNN architectures for learning spatial features. This work proposes a 2D-CNN architecture that makes use of grouped convolutions with the intention of extracting spatial features by independent temporal groups, reducing the computational complexity of the network, and making the network less likely to overfit with limited training data.

3 Materials and Methods

This section starts by the introduction of the study area and habitats and follows with a description of the habitat data collection and characteristics and ends with an introduction to CNNs.

3.1 *Study Area*

The study area is located in the Maharees, a 5-km long tombolo located in the Dingle Peninsula in south-west of Ireland. Figure 1 represents the study area delimited with a red polygon over a map at different scales. The study area consists of a sand dune ecosystem that contains several annexed habitats of the EU Habitats Directive,[1] which lists the habitats that are considered to be of most nature conservation importance at a European level. The four habitats considered for the classification and their respective Habitat Directive[1] Annex I habitat codes are: white dunes (2120), gray dunes (2130), embryonic shifting dunes (2170), and dune slacks (2190). Sample images of the four habitat types are displayed in Fig. 2. The white dunes (also known as marran dunes) are partially stabilized hills or ridges of sand that occur along the seaward edge of the main sand dune system. The gray dunes are a type of fixed dunes, stabilized ridges, or hills of

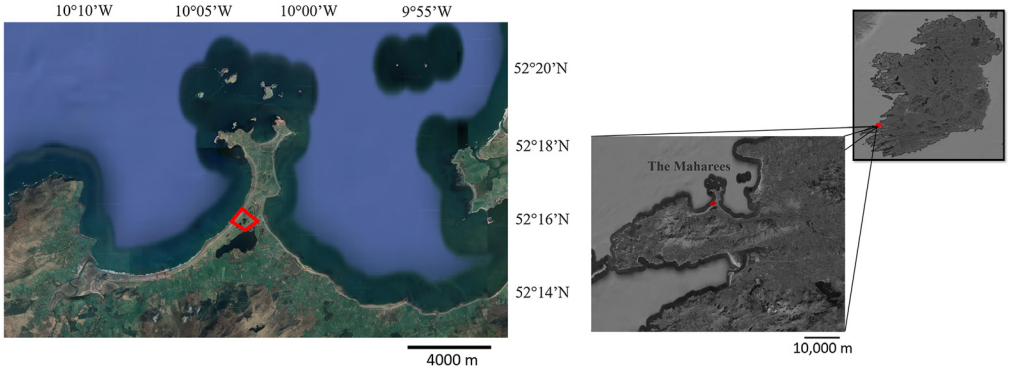

Fig. 1 Study area located in the Maharees, Co. Kerry Ireland. Images obtained thanks to Google Earth Engine.[18]

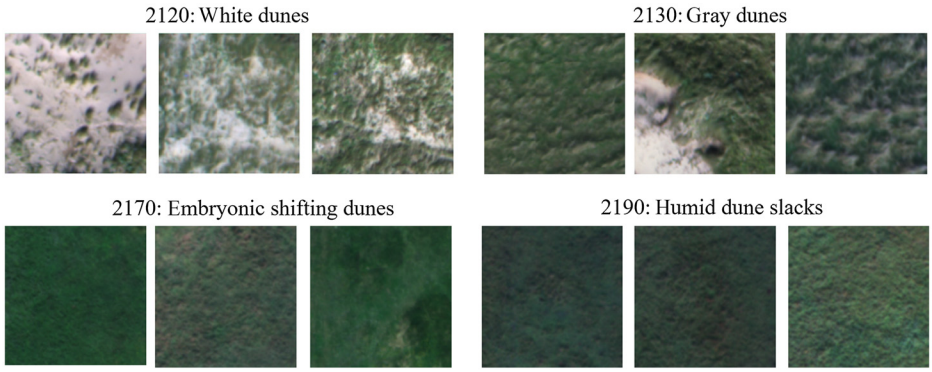

Fig. 2 Sample images of the four Annex I habitats present in the Maharees.

sand with a more or less complete cover of vegetation and a variable species composition usually characterized by grassland or heath communities. The embryonic or shifting dunes are unstable low hills or mounds of sand that occur on the upper extreme of the littoral zone or seashore. And finally, dune slacks are nutrient-enriched wet areas that occur in hollows or depressions either behind or between dune ridges or in blowouts in the sand dunes.[19]

3.2 *Habitat Imagery and Reference Data*

This study considers three MS mosaics of the Maharees study site built from the UAV aerial imagery collected by an MS sensor on the May 26, July 28, and October 8, 2020. The MS images in UTM/WGS84 projection cover a 1×0.7 km^2 area with a spatial resolution of 5 cm and contain the reflectance values of five spectral bands: blue, green, red, red-edge, and NIR. The MS mosaics were obtained by the combination of the image tiles collected by the MS sensor[20] using ProvEye proprietary software.[21] In summary, this software uses back and forward projection in combination with automated algorithms for feature detection of control points that are employed for a correct mosaicking (fusion) process. The software uses upward facing irradiance values to normalize for inflight illumination variation by creating a correction coefficient. Surface reflectance is derived using calibration panels of known albedo taken before and after each flight. This ensures that radiometric variation between mosaics is minimized while still maintaining the spectral integrity of the data. As reference data, this work considers a shapefile with the information about the types of habitats provided by a team of ecologists that labeled multiple sample points across the study site with *in situ* measurements.[22,23] Each on-the-spot habitat label has associated a square polygon that delimits the extent of the habitat label provided by the expert. The reason why the labeling process was done on the field is the difficulty of the classification of the habitats from the remotely sensed images (such as the ones shown in Fig. 2)

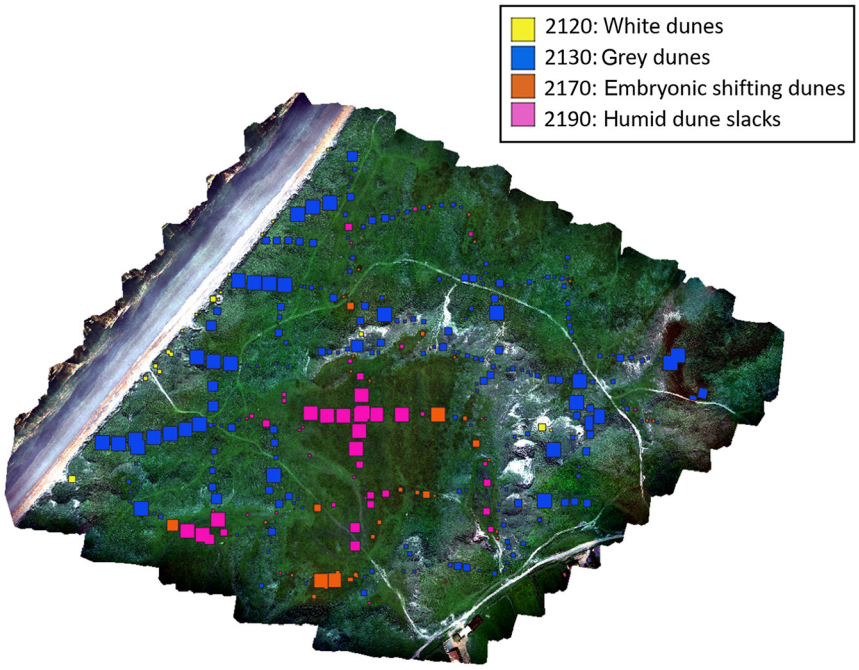

Fig. 3 A true color composite of the Maharees mosaic collected in July 2020 and habitat labels based on field data. These data were kindly provided by the EPA-funded iHabiMap project.[22,23]

caused by the similarity of the habitats characteristics. Figure 3 represents a true color representation of the MS mosaic in July 2020 with the on-the-spot habitat labels represented in different colors according to the four habitat types. The heterogeneous distribution of the habitat types can be seen in Fig. 3, where habitats 2130 and 2190 cover 67% and 23% of the total study area, whereas the minority habitats of 2120 and 2170 only correspond to 2% and 8%, respectively.

In short, the Maharees dataset consists of three MS mosaics corresponding to the same area in the Maharees (Ireland) obtained at three different dates during May, July, and October of 2020. A shapefile with training information on point-based habitat labels was also available information. Both the Maharees MS mosaics and habitat information was provided through the iHabiMap project.[22,23]

3.3 *Convolutional Neural Networks*

CNNs are variants of artificial neural networks that exploit the structural features of the data (e.g., spatial, temporal, or spectral) through the use of convolutional layers.[24] CNNs have shown outstanding performance across a wide range of image classification problems, both at an image level (image classification problems) and at a pixel level (semantic segmentation problems). This is done by successively convolving 1D, 2D, or 3D filters, respectively, along one, two, or three dimensions of the input information. As most of the image classification applications in computer vision deal with monotemporal multimedia images, the majority of the research employs 2D-CNNs.[24] In this case, 2D convolutions are applied along the two spatial dimensions of the images (without considering any type of temporal information) with the objective of learning the spatial patterns that characterize the different classes.

CNN architectures for image classification problems are typically formed by several convolutional layers that learn the patterns (working as a feature extraction function) and followed by one or several fully connected (FC) layers that learn to discriminate the different classes of the problem.[24] CNN architectures can include other types of layers with the intention of improving the performance and convergence of the network, such as max-pooling layers, which downsample the input representation.[24] For a comprehensive description of CNNs and the different types of layers, the reader is referred to Ref. 24.

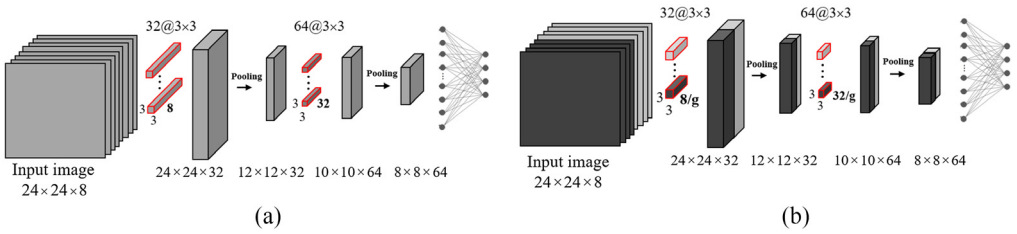

Fig. 4 (a) CNN architecture and (b) CNN architecture with grouped convolutions ($g = 2$).

In this work, we make use of grouped convolutions that refer to convolutional layers whose convolutional filters are divided into g groups and each of them are applied to a portion of the layer input. The concept of grouped convolution was first introduced in AlexNet[25] for distributing the model over two graphics processing units. Namely, the input feature map is divided into g groups and n/g filters convolve separately the input groups (being n the number of filters of the convolutional layer). In this way, the numbers of filters n of the grouped convolutional layer stays the same, but the depth of the filters and the number of parameters of the convolutional layer is reduced by g in comparison with the convolutional layers that do not consider grouped convolutions (which is equivalent to consider a unique group $g = 1$).

For a better illustration of typical 2D-CNN architectures for image classification, Fig. 4 shows an example of two analogous CNN architectures (a) without grouped convolutions ($g = 1$) and (b) with grouped convolutions ($g = 2$), for the classification of eight-channel input image in four different classes. Both CNNs consist of two convolutional layers (whose filters are represented by gray cuboids with red borders), two pooling layers and two FC layers. In the example architecture shown in Fig. 4(b), the 32 filters of the first convolutional layer are split into two 4-depth filter groups, where each group convolves four of the eight channels of the input image. Analogously, the 64 filters of the second convolutional layer are split into two 16-depth filter groups that are applied to each of the two groups of the input feature maps.

4 Multitemporal Multispectral Deep Learning-Based Model for Habitat Classification

This section describes the proposed habitat mapping methodology to classify multitemporal MS images by means of a novel CNN architecture.

4.1 *Processing of the Remotely Sensed Imagery*

In order to extract the multitemporal MS images to train the deep learning model from the Maharees mosaics, the patch-based extraction method described below was followed with the intention of generating more samples for training the model and of mitigating the imbalance of the training dataset. From each polygon p, n_p patches of dimensions $w \times h$ are extracted by randomly sampling the same area from the three MS mosaics of the Maharees corresponding to May, July, and October 2020. The number of patches extracted from each polygon p is equal to $c_j \cdot \text{area}_p/(w \cdot h)$, where area_p is the area of polygon p, and c_j a constant value associated to habitat j. Namely, the patch width w and height h are both set to 25. And with the purpose of mitigating the imbalance of the training dataset, the constant c_j that controls the number of images extracted from a polygon p labeled as class j is set to $c_{1:4} = \{6,1,4,2\}$, extracting higher number of images from the least frequent habitats (white dunes and embryonic shifting dunes). No oversampling/undersampling is applied to the test dataset, where $c_j = 2$ for the four classes. Next, the resulting five-band sampled patches from the three MS mosaics are concatenated resulting in image of size $w \times h \times 15$ (with the fifteen-depth dimension corresponding to the five spectral bands at three different dates) that can be fed into the proposed 2D-CNN model described in Sec. 4.2.

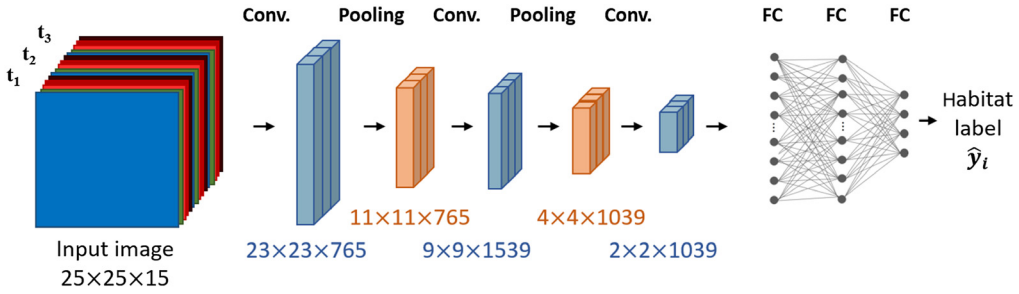

Fig. 5 The CNN architecture proposed for habitat classification from multitemporal MS images.

4.2 *Multitemporal Multispectral CNN-Based Architecture*

The proposed 2D-CNN classifier takes as input a multitemporal MS image and predicts the most probable habitat label $\hat{y}_i$ among the four considered habitat types. The proposed CNN architecture, which is sketched in Fig. 5, is composed of three convolutional layers (with 765 and 1539 filters), two pooling layers and three FC layers (with 1023, 513, and 4 units). The grouped convolutional layers apply three groups ($g = 3$) of 3×3 kernels with no padding and a stride of 1. Note that the grouped convolutions are applied considering three temporally independent groups, where each group of filters is applied to feature maps that belong to the same time (either May, July or October). The two pooling layers consider 2×2 average pooling strategy that reduces by half the spatial dimensions of input feature maps.[24] In addition, all the layers have associated with a rectified linear unit with the exception of the last FC layer, which considers the softmax activation function that normalizes the output of the network to a probability distribution over the four predicted output classes.[24]

The categorical cross entropy loss function is optimized using the Adagrad optimizer and considering a weighting of the loss function based on the number of samples of each class,[26] which is suitable for unbalanced datasets. Moreover, during the training process, 10% of the training data is reserved for validation, and the model is trained following a validation-based early stopping strategy during at most 700 epochs considering a batch size of 32. The model was implemented by utilizing Python and Tensorflow Keras libraries.[27]

5 Results

This section analyzes the performance of the proposed approach for habitat classification with multitemporal MS imagery. Moreover, the proposed approach is compared with an MS CNN-based model in order to analyze the effect of including temporal information on the classification performance.

5.1 *Evaluation Strategy*

To reduce variability and ensure consistent evaluation of the model performance, we apply k-fold cross validation with k equal to two. In this way, the labeled polygons (described in Sec. 4.1) are divided in a randomized stratified fashion into two complementary subsets and during each cross validation round a deep learning model is trained on one subset leaving out the other subset for testing the model performance. This approach ensures there is no overlap between the images used for training and testing the deep learning models, testing in this way the performance of the models over unseen images extracted from labeled polygons unused for training the models.

Furthermore, in order to prove the statistical improvement in the performance of the proposed model with state-of-the-art techniques, a 5×2 cross validation paired t test proposed by Dietterich et al.[28] is considered. In the 5×2 cross validation paired t hypothesis test, five replications of twofold cross validation are carried out and their classification errors are used to calculate the $\tilde{t}$ estimate, which under the null hypothesis of no statistical difference between the models follows a t distribution with 5 degrees of freedom. The results of this analysis can be found in Sec. 5.3.

Table 2 Average confusion matrix obtained through twofold cross validation by the proposed model over the Maharees test images.

		Predicted habitats			
Reference habitats	Habitat	White dunes	Gray dunes	Embryonic shifting dunes	Humid dune slacks
	White dunes	**0.96**	0.04	0.00	0.00
	Gray dunes	0.02	**0.92**	0.03	0.03
	Embryonic shifting dunes	0.00	0.14	**0.69**	0.18
	Humid dune slacks	0.00	0.04	0.16	**0.79**

The parametrization and training options described in Sec. 4.2 are considered for all the experiments presented in this work.

5.2 *Analysis of the Model Performance*

The multitemporal MS images that compose the training and test sets are extracted from the labeled polygons following the batch extraction strategy defined in Sec. 4.1.

The following metrics, which are obtained by averaging the individual metrics obtained over the five replications over the two-fold cross validation test sets, summarize the performance of the proposed CNN-based classifier; overall accuracy: 0.88 and macroaveraged $F1$-score: 0.78. Hence, on average, 88% of the images from the test datasets were correctly identified by the proposed model. These metrics show a high-prediction performance for a complex habitat classification problem. Table 2 shows the average confusion matrix obtained over the test sets through five replications of twofolds cross validations by the comparison of the reference ground truth labels and predicted labels. The four types of dunes are classified with high accuracies, including the least frequent habitat white dunes (corresponding to Annex I code 2120), which obtained 95% of the image patches correctly classified. The class embryonic shifting dunes is the one with lower accuracies due to its similarity to humid dune slacks.

5.3 *Analysis of the Influence of Temporal Information*

This section analyzes the benefits of including temporal information into the habitat classification problem. To this end, the performance of the proposed multitemporal MS 2D-CNN-based model is compared with an MS 2D-CNN-based model represented in Fig. 6. 2D-CNN MS-based models have been previously used by several habitat mapping state-of-the-art works.[3,6,11] In addition, a 2D-CNN monospectral (grayscale) model was employed by Berthold et al.[13] for habitat classification from sonar imagery. The MS model used for the comparison classifies MS image patches that contain the B, G, R, red-edge, and NIR spectral values corresponding to the same area and time. Hence, the depth of the input image patch is reduced by a third in

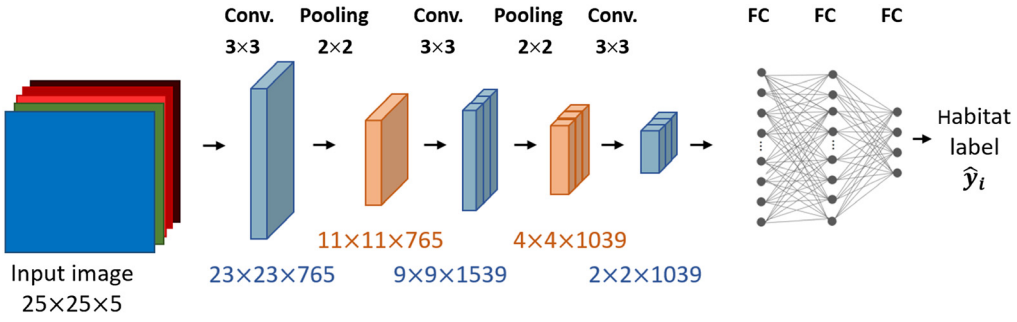

Fig. 6 CNN architecture for habitat classification from MS images.

Table 3 Comparison of the 2D-CNN multitemporal model with 2D-CNN monotemporal models.

Model	Samples	Overall accuracy		p value
		1	2	
Multitemporal MS model (Fig. 5)	May–July–October	0.88 ± 0.02	0.88 ± 0.01	—
MS model (Fig. 6)	May	0.82 ± 0.03	0.83 ± 0.01	0.03
	July	0.82 ± 0.02	0.82 ± 0.01	0.03
	October	0.77 ± 0.02	0.79 ± 0.03	0.04

comparison with the multitemporal MS image patch used by the proposed model (represented in Fig. 5). In order to keep the number of parameters of the MS model equal to the number of parameters of the proposed model (17,501,815 parameters), the number of groups of filters g of the first convolutional layer is set to one. The second and third convolutional layers consider $g = 3$, although contrary to the proposed architecture the groups do not attend to a temporal dimension but instead all the feature maps correspond to the same time (either May, July, or October). Keeping the same number of parameters in the proposed and comparative models allows us to analyze what is the advantage of including temporal information for the habitat mapping classification problem avoiding undesired effects in the performance of the model caused by a change of the number of parameters of the model.

The results of the comparison are shown in Table 3, whose columns inform about: the model being used, the temporal information of the image patches used, the average overall accuracy obtained over the five replications of the twofold cross validations, and the p value of 5×2 the hypothesis test[28] that compares each model with the proposed multitemporal MS model. As shown by the p values in the last column of Table 3, the hypothesis tests allow rejection with a significance level of 0.05 of the null hypotheses of equal performance of the proposed multitemporal model and the monotemporal models that employ the spectral information obtained in May (p value $= 0.03$), in July (p value $= 0.03$), and in October (p value = 0.04). The better results obtained by the proposed multitemporal MS model in comparison with the ones obtained by the MS model show the benefits of incorporating temporal information through the proposed CNN-based architecture.

6 Discussion

The mapping of coastal dune habitats is necessary for monitoring ongoing changes on coastal dunes habitats in order to develop/update the best management practices to mitigate possible climatic or anthropogenic negative impacts on this high-environmental value habitats. Machine learning-based methods for habitat mapping like the one proposed in this work provide a great advantage over the current field-based mapping methodology by the reduction of the financial and man-power cost.

The results of this analysis show the power of deep learning and UAV imagery to classify the Annex I coastal dune habitats in the Maharees, Ireland. The proposed multitemporal MS CNN-based model achieves an overall accuracy of 88% over the Maharees test dataset. This is a high classification accuracy for a complex classification problem, whose low interclass variability requires on the field classification made by environmental experts, and a good result in comparison with the accuracies achieved by other habitat mapping models,[11,13] including the model proposed by Timm and McGarigal[10] for the mapping of a coastal dune habitat in Massachusetts, United States. Furthermore, the experimental analysis indicates the advantage of including temporal information, allowing to improve around 5% the overall classification accuracy over MS 2D-CNN-based models, a technique used by several state-of-the-art habitat mapping works.[3,6,11]

In order to improve the classification performance of the proposed method, we consider two possible directions. On the one hand, the gathering of a higher number of sample points from the embryonic shifting dunes habitat could help to reduce the number of misclassifications with

the similar habitat humid dune slacks. On the other hand, the inclusion of topographical information by the deep learning model could help to improve the results, as was the case in the RF model proposed by Timm and McGarigal[10] for a coastal dune habitat in Massachusetts, United States.

7 Conclusions and Future Research Lines

This work proposes a deep learning approach using CNNs to classify four coastal habitat types from multitemporal MS aerial imagery of the Maharees (Ireland) corresponding to three different dates in 2020. The proposed approach is tested over unseen multitemporal MS aerial imagery following a cross-validation approach, obtaining an overall classification accuracy of 88%. Moreover, the experimental results show that the proposed approach benefits the inclusion of multitemporal imagery.

As future work, we intend to make use of the new aerial imagery from Maharees and different Irish habitats data that is planned to be gathered by the EPA-funded iHabiMap project.[23] This new imagery will allow us to assess the model's performance not only over unseen areas but also with data corresponding to a different year than the one used for training. In addition, the new data from other habitats would allow us to test similar deep learning models on additional types of Annex I habitats located in different regions of Ireland.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 Research and Innovation program under the Marie Skłodowska-Curie Grant Agreement No. 847402. The authors would like to thank the EPA-funded iHabiMap project[22,23] for providing the data used in this work. We thank the anonymous reviewers whose comments and suggestions helped improve and clarify this manuscript. The authors declare no conflicts of interest.

References

1. H. Directive, "Council directive 92/43/eec of 21 may 1992 on the conservation of natural habitats and of wild fauna and flora," *Off. J. Eur. Union* **206**, 7–50 (1992).
2. J. E. Ball, D. T. Anderson, and C. S. Chan, "Comprehensive survey of deep learning in remote sensing: theories, tools, and challenges for the community," *J. Appl. Remote Sens.* **11**(4), 042609 (2017).
3. E. Guirado et al., "Deep-learning versus OBIA for scattered shrub detection with Google Earth imagery: ziziphus lotus as case study," *Remote Sens.* **9**(12), 1220 (2017).
4. P. Benedetti et al., "M^3Fusion: a deep learning architecture for multiscale multimodal multitemporal satellite data fusion," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **11**(12), 4939–4949 (2018).
5. Q. Feng et al., "Multi-temporal unmanned aerial vehicle remote sensing for vegetable mapping using an attention-based recurrent convolutional neural network," *Remote Sens.* **12**(10), 1668 (2020).
6. M. Yasir, A. U. Rahman, and M. Gohar, "Habitat mapping using deep neural networks," *Multimedia Syst.* (2020).
7. T. Kattenborn, J. Eichel, and F. E. Fassnacht, "Convolutional neural networks enable efficient, accurate and fine-grained segmentation of plant species and communities from high-resolution UAV imagery," *Sci. Rep.* **10**(1), 17656 (2019).
8. S. Ji et al., "3D convolutional neural networks for crop classification with multi-temporal remote sensing images," *Remote Sens.* **10**(1), 75 (2018).
9. A. Kobler, S. Džeroski, and I. Keramitsoglou, "Habitat mapping using machine learning-extended kernel-based reclassification of an Ikonos satellite image," *Ecol. Modell.* **191**(1), 83–95 (2006).
10. B. C. Timm and K. McGarigal, "Fine-scale remotely-sensed cover mapping of coastal dune and salt marsh ecosystems at Cape Cod National Seashore using random forests," *Remote Sens. Environ.* **127**, 106–117 (2012).

11. M. Rezaee et al., "Deep convolutional neural network for complex wetland classification using optical remote sensing imagery," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **11**(9), 3030–3039 (2018).
12. M. Diesing et al., "Mapping seabed sediments: comparison of manual, geostatistical, object-based image analysis and machine learning approaches," *Continent. Shelf Res.* **84**, 107–119 (2014).
13. T. Berthold et al., "Seabed sediment classification of side-scan sonar data using convolutional neural networks," in *IEEE Symp. Ser. Comput. Intell. (SSCI)*, IEEE, pp. 1–8 (2017).
14. A. S. Li et al., "NASA NeMO-Net's convolutional neural network: mapping marine habitats with spectrally heterogeneous remote sensing imagery," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **13**, 5115–5133 (2020).
15. M. Rußwurm and M. Korner, "Temporal vegetation modelling using long short-term memory networks for crop identification from medium-resolution multi-spectral satellite images," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit. Workshops*, pp. 11–19 (2017).
16. C. Pelletier, G. I. Webb, and F. Petitjean, "Temporal convolutional neural network for the classification of satellite image time series," *Remote Sens.* **11**(5), 523 (2019).
17. E. Ndikumana et al., "Deep recurrent neural network for agricultural classification using multitemporal SAR Sentinel-1 for Camargue, France," *Remote Sens.* **10**(8), 1217 (2018).
18. N. Gorelick et al., "Google Earth engine: planetary-scale geospatial analysis for everyone," *Remote Sens. Environ.* **202**, 18–27 (2017).
19. J. A. Fossitt, *A Guide to Habitats in Ireland*, Heritage (2000).
20. P4 multispectral, https://www.dji.com/ie/p4-multispectral.
21. P. 2021, "ProvUAV v1.3, ProvEye Limited, Ireland," http://proveye.ie/ (2021).
22. EPA Research Programme 2014–2020, "iHabiMap: habitat mapping, monitoring and assessment using high-resolution imagery. Science Foundation Ireland (SFI) Grant No. 12/RC/2289-P2 and 16/SP/3804."
23. C. Cruz et al., "iHabiMap: habitat mapping, monitoring and assessment using high-resolution imagery," in *13th Irish Earth Obs. Symp.* (2019).
24. L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: a technical tutorial on the state of the art," *IEEE Geosci. Remote Sens. Mag.* **4**(2), 22–40 (2016).
25. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Adv. Neural Inf. Process. Syst.* **25**, 1097–1105 (2012).
26. Y. Cui et al., "Class-balanced loss based on effective number of samples," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 9268–9277 (2019).
27. M. Abadi et al., "Tensorflow: a system for large-scale machine learning," in *12th USENIX Symp. Oper. Syst. Design and Implementation*, pp. 265–283 (2016).
28. T. G. Dietterich, "Approximate statistical tests for comparing supervised classification learning algorithms," *Neural Comput.* **10**(7), 1895–1923 (1998).

Sara Pérez-Carabaza received her PhD in physics from the University Complutense of Madrid financed by Airbus Group in 2019. She is a Marie Skłodowska-Curie Career-FIT fellow at CeADAR at the University College of Dublin. She is currently interested in the application of deep learning to remote sensing classification problems, and she has collaborated at the Joint Research Centre for the classification of agricultural parcels from Sentinel-2 satellite imagery.

Oisín Boydell received his PhD in computer science from the University College Dublin. He is a principal data scientist and a head of the Applied Research Group at CeADAR, where among other activities, he leads research projects in conjunction with industry partners in exploring innovative applications of AI and machine learning applied to Earth observation data.

Jerome O'Connell received his PhD from the School of Biosystems Engineering at the University College Dublin in applied remote sensing. He is the founder and the managing director of ProvEye, a company that specializes in image processing from UAVs and satellites in agriculture and natural resources. He has more than 15 years of experience developing image processing methods for remote sensing, working with a host of commercial institutes and organizations throughout the world.