Automated detection of large brown macroalgae using machine learning algorithms-a case study from Island Bay, Wellington New Zealand

Technical Report · June 2021

4 authors, including:

Roberta D'Archino
National Institute of Water and Atmospheric Research
71 PUBLICATIONS   **775** CITATIONS

SEE PROFILE

Alexandre Carmelo Gregory Schimel
37 PUBLICATIONS   **1,074** CITATIONS

SEE PROFILE

Tara Anderson
Independent Researcher
32 PUBLICATIONS   **776** CITATIONS

SEE PROFILE

Fisheries New Zealand
Tini a Tangaroa

Automated detection of large brown macroalgae using machine learning algorithms—a case study from Island Bay, Wellington

R. D'Archino,
A.C.G. Schimel,
C. Peat,
T. Anderson

New Zealand Government

TABLE OF CONTENTS

EXECUTIVE SUMMARY

D'Archino, R.; Schimel, A.C.G.; Peat, C.; Anderson, T. (2021). Automated detection of large brown macroalgae using machine learning algorithms—a case study from Island Bay, Wellington.

New Zealand Aquatic Environment and Biodiversity Report No. 263. **36 p.**

This study was initiated to assess and improve previously developed machine learning (ML) algorithm for automated detection and identification of three habitat-forming brown macroalgae, using underwater towed video imagery, within the Island Bay to Houghton Bay study region (Taputeranga Marine Reserve) in Wellington.

The target taxa were selected to represent key habitat forming taxa across a range of taxonomic levels, including: two species (*Ecklonia radiata* and *Lessonia variegata*), one genus (*Carpophyllum* spp.), and a grouping of all macroalgae present.

A total of 32 towed video transects were surveyed. Twelve were acquired in January 2019, as part of a ML pilot/feasibility study funded by NIWA's Strategic Science Investment Fund and the additional 20 transects were surveyed in August 2020 as part of the current study. The surveyed area was designed to include a broad variety of reef sites with variable macroalgal cover and species composition, rather than overlapping for temporal comparisons.

The acquisition of additional video footage provided the data necessary for additional training (15%) and algorithm validation (85%). The performance of the models was assessed by evaluating them on their respective training dataset and validation dataset, compiling the predicted results against actual categories as confusion matrices, and calculating several common metrics used as accuracy, precision, recall, F1-score, and Matthew's correlation coefficient.

NIWA's ML models were successful at identifying and distinguishing all four macroalgal groups across the entire survey, with overall models attaining high performance accuracies of between 80.6 to 87.1% for individual species and genera, and 97.2% for the 'all-macroalgae' group.

Two distribution maps (presence/absence) for each of the four target taxon levels were produced: an overall prediction map that indicated the highest certainty for determining the distributions; and an overall uncertainty map to examine where rare occurrences of each taxon occurred.

This study showed that ML can provide a significantly faster and cost-effective approach to post-processing video imagery which will improve as more data are collected. ML models performed well at detecting and distinguishing between closely related species (*Lessonia variegata* and *Ecklonia radiata*), highlighting the accurate fine-tuned performance of these models. The next step in ML coding would be to extend the ML algorithm to estimate the percentage cover for the four macroalgal groups and to detect other key macroalgal taxa.

Kelp forests are susceptible to changing environmental conditions, including changes in water temperatures, increasing turbidity, sediment deposition, and ocean acidification. Monitoring data are critical to documenting changes in our coastal communities and for management. Decline of kelp beds is likely to affect subcanopy species (including pāua and rock lobster) that will lose the three-dimensional habitat and be affected by changed light and water motion regimes.

1.   INTRODUCTION

1.1   Background

Large, canopy-forming brown macroalgae (Laminariales and Fucales, commonly known as kelp) are globally recognised as critical components of coastal ecosystems (e.g., Krumhansl et al. 2016, Coleman & Wernberg 2017). They provide three-dimensional structures that create habitat, shelter, and nursery grounds for a diverse range of organisms, including commercially important species of fish and shellfish as well as diverse and productive macroalgal assemblages (Mann 1973, Dayton 1985, Graham et al. 2007). Kelp forests are highly productive and are considered to be one of the most productive systems on earth (Mann 1973, Graham et al. 2007). As primary producers, macroalgae use solar energy to convert inorganic material to organic matter through photosynthesis by using CO_2. Kelps modify their surrounding environment by altering the water flow and the light and sedimentation in the subcanopy, and they buffer wave energy, reducing coastal erosion (Layton et al. 2019, Murie & Bourdeau 2020). Finally, kelp forests have important cultural and socio-economic value and support tourism and recreational activities (Bennett et al. 2016).

In New Zealand, rocky reefs are dominated by diverse algal communities which form large beds with species composition differing with geographic distribution. The Laminariales includes three native genera (*Ecklonia*, *Lessonia*, and *Macrocystis*) and one introduced genus (*Undaria*), and the Fucales includes 10 genera (*Carpophyllum, Cystophora, Durvillaea, Hormosira, Landsburgia, Marginariella, Notheia* (an obligate epiphyte, primarily found on *Hormosira*), *Phyllotricha, Sargassum, Xiphophora*) and 29 species.

International studies have reported on the loss of subtidal macroalgal forests in temperate and subtropical marine ecosystems, particularly on urbanised coasts. Large brown macroalgae are now globally considered key indicators of ecosystem change on coasts where they are found (D'Archino et al. 2019). In New Zealand, the lack of baseline data makes it difficult to assess if large brown algae are in decline as in other regions in temperate waters (e.g., Australia, Europe, Japan). Recently the loss of *Durvillaea* spp. was documented in the South Island as a result of the heatwave in 2017–2018 (Thomsen et al. 2019) and there is increasing concern about the decline of *Macrocystis pyrifera.* The conservation status of *Durvillaea poha, Durvillaea antarctica*, and *Macrocystis pyrifera* has been assessed as 'at risk – declining' (Nelson et al. 2019). Because the loss of canopy-forming algae is likely to be associated with a significant loss of associated species and ecological functions (Teagle et al. 2017), it has become clear that monitoring canopy-forming algal beds is important in order to assess changes due to climate change and anthropogenic stressors (e.g., pollution, sedimentation, eutrophication. The importance of acquiring baseline data on habitat-forming species was widely discussed by (D'Archino et al. (2019).

This lack of baseline data is partly due to the significant time and cost associated with coastal benthic surveys, which rely heavily on expert SCUBA-divers (D'Archino et al. 2019). A solution may be found in the automation of underwater video footage analysis, which could constitute a step-change in terms of cost and time effectiveness of coastal benthic surveys. Not only would this allow the acquisition of the necessary baseline data, but also this would increase the coverage and frequency of monitoring and change-detection efforts. Bewley et al. (2012) and Seiler et al. (2012) demonstrated how, given a sufficiently large training dataset of expert-labelled frames of underwater video footage, the application of standard machine learning classification algorithms (respectively, Support Vector Machine and Random Forest) to local image features could accurately predict the presence or absence of *Ecklonia radiata* and map the distribution of temperate rocky reef habitats. The method was developed further into the CoralNET software (Beijbom et al. 2015), which was successfully used, for example, to estimate habitat-forming algae in Australia (Griffin et al. 2017).

There have been significant recent advances in Artificial Intelligence (AI): machines can now use knowledge from image and video recognition and classification to detect the world in a similar manner to that of humans. Advancements in computer vision with deep learning have resulted in Convolutional Neural Networks (CNNs), pattern recognition algorithms that can assess an image, detect

aspects/objects in the image, assign importance (trained to include weights and biases), and differentiate one object/aspect from another (Rawat & Wang 2017). As such, CNN is fast becoming the method of choice for visual tasks such as image detection and classification (Rawat & Wang 2017, Gu et al. 2018), with direct and significant applications for marine underwater video surveys (Moniruzzaman et al. 2017, Malde et al. 2020). CNNs can achieve a high level of performance in image analysis, but their main drawback is that a large amount of training data (in the order of 10^4+ images) is generally required for these models to learn the spatial features (Yamashita et al. 2018). Advances in transfer learning approaches have allowed the use of CNNs without the required training dataset size.

In a pilot machine learning study (D'Archino et al. 2019, Appendix 8), the authors investigated the potential of the 'feature extraction' transfer learning method, which consists of passing the video frames to a pre-trained, deep CNN to produce 'deep features' that can be then used in input to a traditional classifier. In this pilot study, Inception-ResNet-v2 was used as the pre-trained neural network, and a linear Support Vector Machine (SVM) classifier. Mahmood et al. (2020) demonstrated that this methodology (using the deep residual network ResNet-50 instead) allows accurate identification of *Ecklonia radiata* in images collected by an autonomous underwater vehicle in Australia. Following the success of this pilot study, a project was funded by NIWA's Strategic Science Investment Fund to (1) acquire additional videos and establish best camera settings, (2) test different approaches for preparing the training dataset, (3) develop alternative deep-learning classification approaches, and (4) have an expert identifying macroalgae in order to develop the training dataset.

1.2   Project focus

The current project aims to assess and improve the previously developed machine learning (ML) algorithm for automated-detection and identification of three habitat-forming brown macroalgae, using underwater towed video imagery, within the Island Bay to Houghton Bay study region in Wellington. Previously collected imagery used to train the model is combined with additional new imagery (representing new sites and time of sampling) to train (15%) or evaluate (85%) the ML models.

The acquisition of additional video footage on the south coast of Wellington provided data necessary for additional training, algorithm validation, and the completion of maps of macroalgal distribution in the area between Island Bay and Houghton Bay (Taputeranga Marine Reserve).

1.3   Research objectives

Objective 1: Turn the machine learning method into a practical system that applies readily-trained neural networks to new videos and provides results in both .csv files and maps.

Objective 2: Acquire baseline data of large brown algal distribution along transects in Island Bay and Houghton Bay (Taputeranga Marine Reserve) for monitoring. Undertake one day of field survey to acquire video transects; apply the ML algorithm to these video data to produce a distribution map of large brown algae.

2.   Methods

2.1   Study species and site

The target taxa were selected to represent key habitat forming taxa across a range of taxonomic levels, including: two species (*Ecklonia radiata* and *Lessonia variegata*), one genus (*Carpophyllum* spp.), and a grouping of all macroalgae present (Figure 1). These species/genera occur around much of New Zealand's coast (D'Archino et al. 2019), provide important 3-dimensional biogenic habitat for a plethora of other benthic and demersal species (Anderson et al. 2019), and, due to their physiological sensitivities to changing environmental conditions, are excellent indicator species for monitoring environmental and ecosystem health (D'Archino et al. 2019, Anderson et al. 2019).

Lessonia variegata (here after referred to as *Lessonia*) is found in the upper subtidal zone through to deep water on rock on exposed coasts, and forms extensive beds in some places. The thalli are large, with long, parallel-sided, strap-like blades growing from perennial, solid, dichotomously branched axes, up to 1 m or more in height. There are seven species of *Lessonia* currently recognised in New Zealand, three of which have restricted geographic distributions (e.g., Chatham Islands, Snares Islands, and subantarctic islands). *Lessonia* was reported to grow around the North, South, and Stewart islands, but recent research has revealed that there is greater genetic diversity than previously recognised. *Lessonia* is confirmed to occur in the Cook Strait region, with an additional three species found around mainland New Zealand yet to be described (Zuccarello & Martin 2016).

Ecklonia radiata (here after referred to as *Ecklonia*) is found in the low intertidal to subtidal zones, on rocky reefs, occasionally on cobbles and shells, on moderately sheltered shores, and on exposed coasts. The thalli are large, up to 1 m or more high, with a smooth, cylindrical, unbranched stipe and a flattened blade with lateral lobes. In New Zealand there is only one species in this genus, but the morphology can vary considerably.

Carpophyllum is an endemic genus commonly found fringing the low intertidal margins of rocky reefs. There are four species in New Zealand and two occur in Wellington: *Carpophyllum flexuosum*, which has the widest geographic range and extends into deeper waters (to depths of up to 20 m), and *Carpophyllum maschalocarpum,* which is very common in the low intertidal and upper subtidal zones, in many areas forming a horizontal band at low water, and occurs on open coasts as well as in sheltered harbours, lagoons, and tide pools. Because the two species share a similar morphology (e.g., thalli have flattened main axes bearing leaves of various shapes and sizes), they have been treated as *Carpophyllum* spp. in this study.

Figure 1:   The four categories of large brown macroalgae examined in this study. These included two species: a) *Ecklonia radiata*, b) *Lessonia variegata*; one genus: c) *Carpophyllum* spp.; and d) a category of all macroalgae present.

This study was conducted along an approximately 1.7 km section of exposed coastline encompassing Island Bay and Houghton Bay, on the Wellington south coast (Figure 2a). This section of coast lies inside the Taputeranga Marine Reserve, a marine protected area that extends 3.8 km along shore and 2.3 km offshore.

2.2 Survey design

In this study, a total of 32 towed video transects were surveyed within the Island Bay to Houghton Bay survey area (Figure 2b). Twelve transects were surveyed on 31st January 2019, as part of a ML pilot/feasibility study, with seven transects in Houghton Bay and five in Island Bay (black lines in Figure 2b) (Table 1). To provide more comprehensive spatial coverage within the study area, an additional 20 transects were surveyed on 3rd August 2020. Because this survey aimed to assess the capability of AI to automatically detect kelp taxa using the newly-developed ML algorithms, spatial extent was optimised to include a broad variety of reef sites with variable macroalgal cover and species composition, rather than any overlap for temporal comparisons.
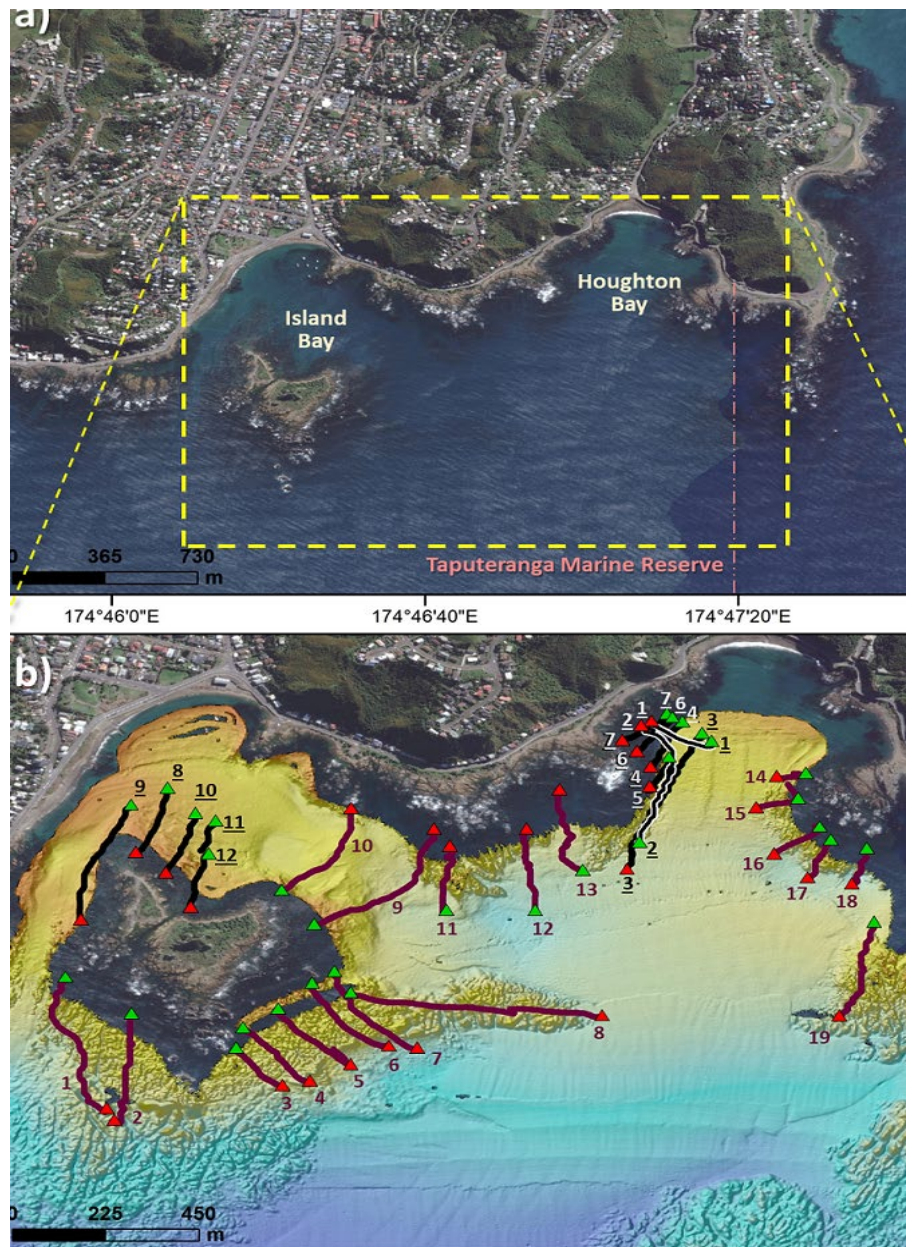
Figure 2: **Survey area and design of towed video survey of benthic macroalgae within Island Bay to Houghton Bay on Wellington's south coast, inside the Taputeranga Marine Reserve. a) Yellow dotted rectangle = extent of survey area; rose-coloured dashed line = the eastern boundary of the Taputeranga Marine Reserve. b) Study area, with locations of towed video transects surveyed in 2019 (black track lines) and 2020 (burgundy track lines); green and red triangles = start and end of towed video transects, respectively; transect numbers = 2019 (underlined black or grey) and 2020 (burgundy).**

Table 1:   Use of transects in this study.

| Survey year	Transect numbers in survey (total)	Used for model training (T), model evaluation (E), or not labelled (NL)	Used for model predictions (map)
2019	1–12 (12)	T	Yes
	13–20 (8)	T	No (Breaker Bay)
2020	11, 18 (2)	T	Yes
	2, 8, 10 (3)	E	Yes
	1, 3–7, 9, 12–17, 19, 20 (15)	NL	Yes

NIWA's Seaweed-Cam small towed video camera system (Figure 3) and methodology were used for both the 2019 and 2020 surveys. The tow frame was fitted with a high-definition forward-facing (45° angle) video camera (Splashcam Deep-Blue HD-1080p video camera - Figure 3a-b) to see both the seabed and oncoming objects, with real-time video feed via a coaxial cable to the surface vessel[1]. The high-resolution (1080p) video footage was viewed and recorded on a topside video monitor/recorder (HDMI Atomos Ninja Blade 5" fitted with a 240GB Solid State Disk Drive). Two LED 1500 lumen dive lights provided some close-to-the-seabed illumination, and paired lasers, spaced 10 cm apart, provided size reference for objects and organisms seen on the seabed (Figure 3b). The system received a GPS fix from a small GPS antenna affixed to the vessel, collecting a low-accuracy satellite-referenced position every 1–3 seconds. Each video frame was stamped with the GPS position (latitude and longitude in decimal degrees), site or transect number (manually entered), and local date and time (e.g., Figure 4, Figure 5), which ensured that the video imagery and the corresponding metadata (GPS position and time) were permanently synchronized.

The Seaweed-Cam was deployed by hand off the side of NIWA's RV *Rukawai* and was towed approximately 1 m above the seabed at a speed of approximately 0.5 to 1 knot. Video transects at each site were surveyed perpendicular to the coast and island shores, extending wherever possible from the immediate subtidal (about 2 m) across the width of the reef and out beyond the reef edge. Underwater videos were acquired at depths between 2 m and 16 m, and the transect length varied from about 76 m to 482 m, indicating how far each reef extended offshore. All towed video transects were surveyed when underwater visibility was over 3 m to ensure video imagery was of adequate quality to distinguish macroalgae. Transects generally started in deep water with the boat proceeding backwards to the shore, which allowed a better view of upcoming emerging or shallower rocks and minimised the drift length of the cable. During each transect, the real-time video footage displayed on the Atomos Ninja Blade monitor allowed the cable operator to see on-coming obstacles and adjust the cable length accordingly to avoid collision. All video footage was backed up to NIWA's archive drive upon completion of the day's field survey.

---

[1] The tow-frame also has an attached underwater housing for a GoPro Hero4 video camera (stand-alone camera used as a backup system), but not used in this survey.
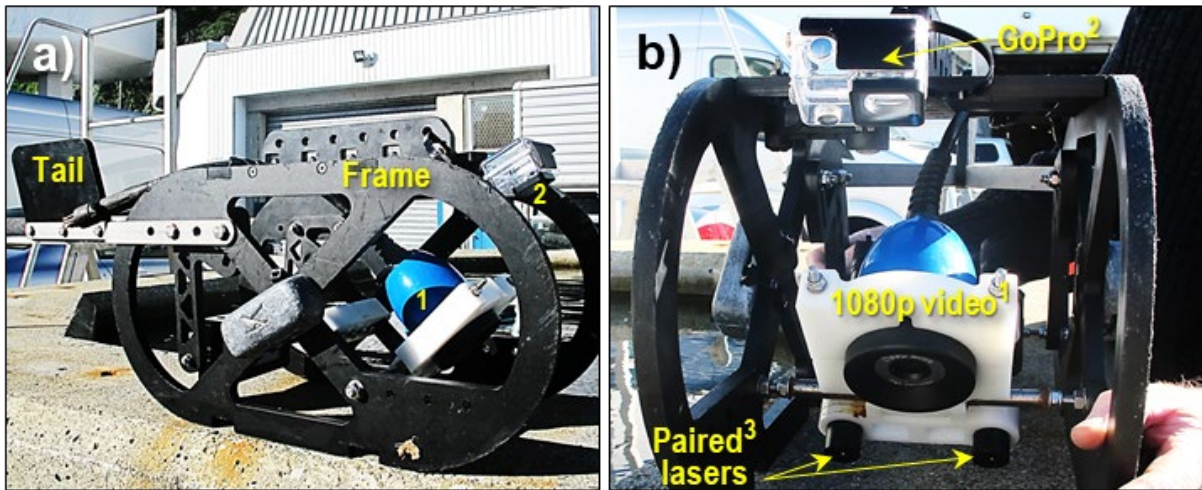
Figure 3: The Seaweed-Cam—NIWA's small towed video camera system. a) Wide angle side view of the towed video frame (85L x 24W x 24H cm) and configuration; b) close-up front view of the towed video setup. Superscript numbers denote: 1 = High-definition video camera (Splashcam's Deep-Blue HD-1080p video camera) with real-time video feed via a coaxial cable to the surface vessel; 2 = Underwater housing fitted with a GoPro Hero4 video camera (stand-alone camera used as a backup system); 3 = Paired lasers are spaced 10 cm apart to provide size reference for objects and organisms seen on the seabed.
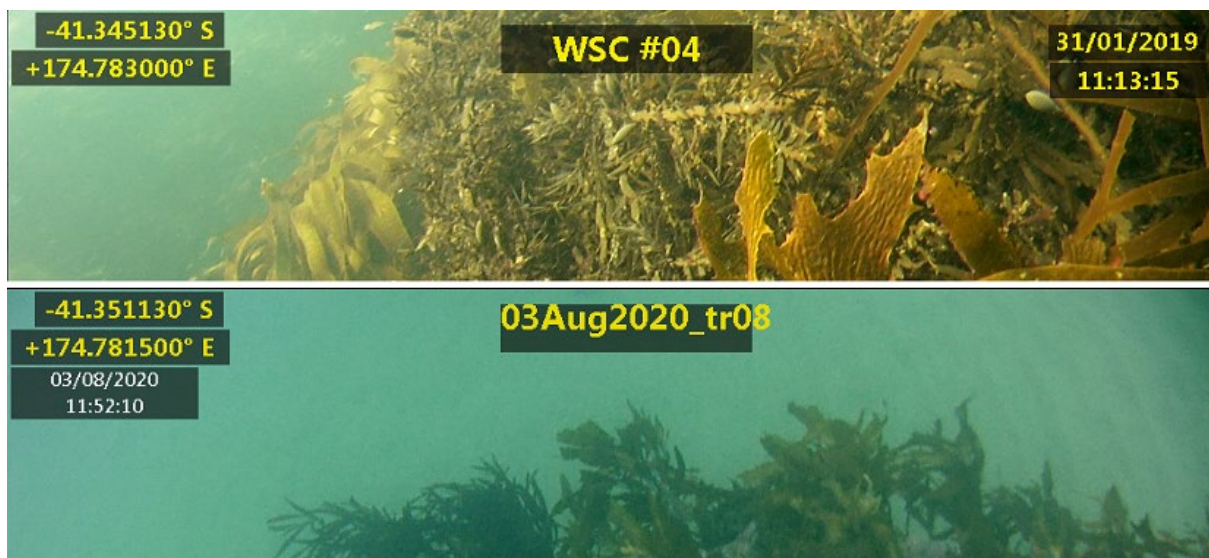
Figure 4: Examples of metadata overlaid at the top of the video footage. Examples from the 2019 (top) and 2020 (bottom) surveys.

Figure 5: **Examples of seafloor imagery collected within the study area (Island Bay to Houghton Bay).**

2.3 Pre-model processing

2.3.1 Metadata and image extraction

A Python algorithm developed during the pilot study, to automatically read the metadata overlaid on the videos, was reused for this project (see detailed description in appendix 8 of D'Archino et al. 2019). It was modified to account for the difference in the location and style of the date and time overlay in more recent footage.

These extracted data were then exported in comma-delimited files, along with relevant information, including the source video filename, corresponding frame number in the video, and time since the beginning of the video (e.g., Table 2; also see appendix 8 of D'Archino et al. 2019 for a detailed description of this method). The metadata algorithm was modified slightly in this study to account for the change in screen position and text style of the 2020 date and time overlay.

Table 2: **Example of metadata extracted from video overlay. Showing the results for the first few seconds of the video corresponding to transect 4 of the August 2020 survey**.

Data lines	Date	Time	Video time	Latitude	Longitude	Prefix	Frame
0	3/08/2020	11:08:51	0:00:00	-41.3527	174.7755	SWC2001_004	0
1	3/08/2020	11:08:52	0:00:01	-41.3527	174.7755	SWC2001_004	30
2	3/08/2020	11:08:53	0:00:02	-41.3527	174.7755	SWC2001_004	60
3	3/08/2020	11:08:54	0:00:03	-41.3527	174.7755	SWC2001_004	90
4	3/08/2020	11:08:55	0:00:04	-41.3528	174.7755	SWC2001_004	120
5	3/08/2020	11:08:56	0:00:05	-41.3528	174.7755	SWC2001_004	150
6	3/08/2020	11:08:57	0:00:06	-41.3528	174.7755	SWC2001_004	180

2.3.2 Expert data labelling

Twenty-five video transects were selected to provide adequate video imagery of the four macroalgal taxa/groups to train and to evaluate the models' ability to accurately detect each taxon/group. This included all 12 of the 2019 survey, 8 additional transects from the 2019 survey outside the study area (specifically, Breaker Bay, south coast Wellington), and five of the 20 transects surveyed in 2020 (Table 1). Individual images for the training library were extracted from each of the video transects at the rate of 1 per second (1 image every 30 frames; equivalent to the metadata extraction) and saved as 'png' files.

A macroalgal specialist (R. D'Archino) then visually examined each image and recorded her expert assessment in a set of columns of the corresponding comma-delimited file. The suitability of the image for expert analysis was recorded in the 'good-frame' column, using '0' for unsuitable frames (due to the camera being still on the boat, or too far from the bottom, in too turbid waters, obstructed by debris, too close to the bottom, stuck on the reef, too much motion blur, etc.) and '1' for otherwise suitable frames (where species presence/absence can be confidently assessed). In other columns corresponding to a given macroalgal type (such as '*Ecklonia*', '*Lessonia*', and '*Carpophyllum*') or category of interest (e.g., 'macroalgae' or 'sand'), the possible labels were '0' for absent, '1' for present (irrespective of density), or '2' for uncertain.

2.4 Machine learning models

Five individual ML models were generated using the training images with their associated expert identifications (i.e., labelled data library): (1) frame suitability, (2) *Ecklonia*, (3) *Lessonia*, (4) *Carpophyllum*, and (5) macroalgae.

The frame suitability model was trained and validated on all images using the 'good frame' label as the target variable. The purpose was to generate a robust indicator of the quality and suitability of video footage for analysis, which would provide an additional degree of confidence on the predictions from the other specific models.

The four taxa models were trained and validated on suitable images (i.e., good frame = 1) using the corresponding identification label (respectively '*Ecklonia*', '*Lessonia*', '*Carpophyllum*', and 'macroalgae') as the target variable, ignoring the images where the target was uncertain (i.e., expert label = 2). The rationale was that the exclusion of 'unsuitable' images (i.e., where the seabed was not clearly visible) or images where the macroalgal expert could not clearly see or identify taxa from images

(i.e., assessment was uncertain), would allow the model training to be focused on the specific features that differentiate the presence or absence of a species/taxon/category.

2.4.1 Training and evaluation dataset

To accurately assess a model's capability for generalisation on future datasets, the evaluation must be performed on a dataset that is independent of the dataset that was used for its training. Since there is a high risk of correlation between the image contents of consecutive frames in any given transect, several specific transects were selected to form the evaluation dataset rather than a random subset of frames. After verifying that they contained instances of all classes for all models, the labelled data of transects 2, 8, and 10 of the 2020 survey were reserved for evaluation (n=3). The training dataset thus consisted of the labelled data for all 20 transects of the 2019 survey (including the 8 transects outside the study area) and 2 transects from the 2020 survey (transects 11 and 18, see Table 3).

Based on this split, the training dataset for the frame suitability model totalled 7839 frames, and its evaluation dataset totalled 1790 frames (Table 4). For the other models, the removal of unsuitable frames and frames with uncertain assessment (label = 2) led to slightly smaller datasets, with training datasets in the 5863–6137 range and evaluation datasets in the 1366–1517 range (Table 4).

Table 3: **Allocation of towed video transects for ML models.**

Survey year	Transect numbers in survey (n)	Use in model	Used for model predictions (map)
2019	1–12 (12)	Training set	model predictions
	13–20 (8)	Training set	Not used (Breaker Bay)
2020	11, 18 (2)	Training set	model predictions
	2, 8, 10 (3)	Evaluation set	model predictions
	1, 3–7, 9, 12–17, 19, 20 (15)	not labelled	model predictions

Table 4: **Size and class counts of the training and validation datasets for all five models. Each model consists of a binary classification with 'Negative (0)' and 'Positive (1)' referring respectively to the 'unsuitable' and 'suitable' classes for the frame suitability model, and to the 'absent' and 'present' classes for all other models.**

Model	Training			Validation		
	Negative (0)	Positive (1)	Total	Negative (0)	Positive (1)	Total
Frame suitability	1 694	6 145	7 839	273	1 517	1 790
Ecklonia	4 235	1 628	5 863	815	551	1 366
Lessonia	4 198	1 699	5 897	888	559	1 447
Carpophyllum	4 311	1 596	5 907	1 128	357	1 485
Macroalgae	976	5 161	6 137	73	1 444	1 517

2.4.2 Model architecture and training parameters

Given the larger training dataset compared with that in the pilot study (D'Archino et al. 2019), the original 'features extraction' transfer learning approach (transfer learning = a machine learning method where a model developed for a task is reused as the starting point for a model on a second task) was upgraded to the 'fine tuning' method, in which the pre-trained deep neural network was partially retrained (fine-tuned) on new data, rather than simply used to extract image features. The InceptionV3 CNN pre-trained on the ImageNet dataset was used, as available in the Tensorflow-Keras (v1.12) Python library, as a base network.

Each of the five models used the same configuration. The last 1000 classes classification layer of the InceptionV3 base network was removed and the following were added; a fully-connected (dense) layer of 128 units, a dropout layer with a dropout rate of 0.2, and a fully-connected layer of 2 units, suitable as a final classification layer for the binary classification tasks.

Model training used a batch size of 32 and class weights equal to the inverse of the class frequencies to compensate for the class frequency imbalance. Images were transformed using augmentation with parameters randomly updating (see next section) and resized to 299 x 299 pixels to match the input size of InceptionV3. Accuracy was selected as the evaluation metric and categorical cross-entropy as the evaluation (loss) function. Optimisation followed a Stochastic Gradient Descent with an initial learning rate of 0.001, a momentum of 0.5, and a decay rate of 0.0001.

A maximum number of training epochs of 150 was set but included an early-stopping function called at the end of each epoch to interrupt the training if accuracy over the validation dataset had not increased after 20 consecutive epochs. The model weights at the epoch with maximum accuracy were retained as the final weights.
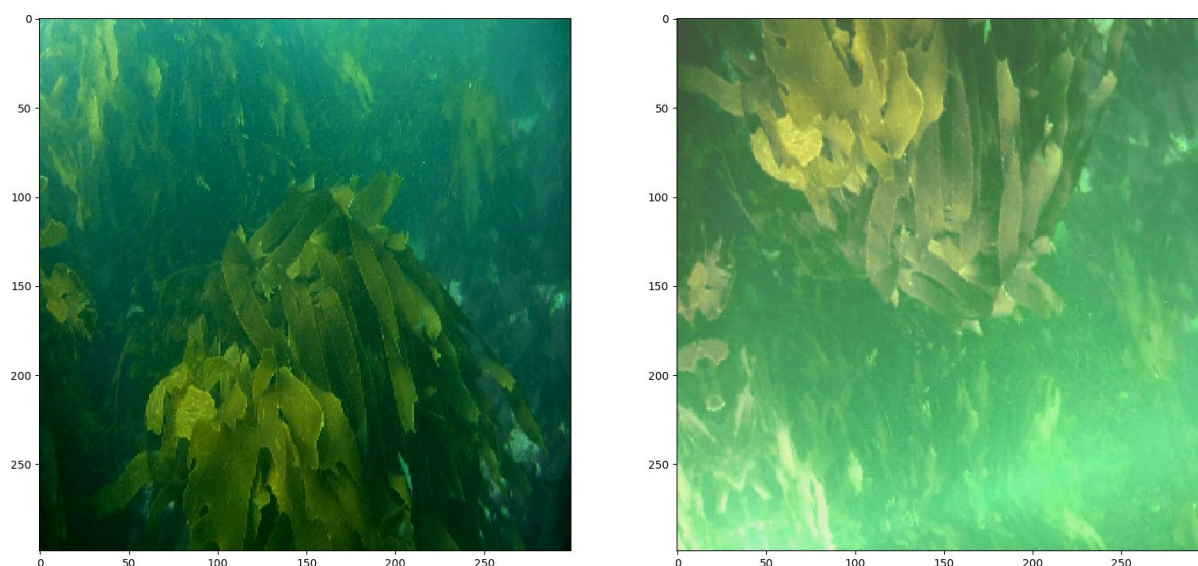
2.4.3   Image augmentation

A common approach to address the issues associated with limited data in training CNNs (e.g., overfitting) is to use 'augmented duplicate images' (Lei et al. 2019, Shorten & Khoshgoftaar 2019). Instead of using the raw images, the models were trained on transformed images (e.g., rotated, stretched, flipped, translated, etc.) with the transformations controlled by parameters that update randomly within a desired range at each epoch (Table 5, Figure 6). Although the resulting images were still highly correlated with the originals, this approach ensured that the neural network was never exposed to the exact same image twice during training, which then allowed the network to generalise to these transformations. Table 5 lists the transformations and the range of the corresponding parameters.

Underwater photography is characterised by variations in colour and brightness with sun azimuth, cloud cover, depth, distance from the seafloor, and water turbidity (Warrant & Locket 2004). This non-random variability in the resulting images can introduce bias in CNN image recognition. For example, since species of macroalgae tend to be distributed by depth, the model could end up learning to recognise the presence of a species not on instances of the algae themselves but on the level of brightness or blue light that characterise the depth range at which it occurs. To compensate for this risk, many of these lighting variations can be simulated by scaling the hue, saturation, and/or lightness of the image. As part of the image augmentation process, a function that transformed the RGB channels to an HSV representation and introduced random scaling was implemented, thus allowing the CNN to generalise to variations in a wide range of lighting conditions.

Table 5:   Transformation types and corresponding parameter ranges used for image augmentation.

Augmentation type	Parameter range
Rotation by a random angle	Between -30 and 30 degrees
Height stretch by a random factor	Between -10% and 10%
Width stretch by a random factor	Between -10% and 10%
Randomly horizontal flipping	N/A
Randomly vertical flipping	N/A
Randomly scaling each channel (RGB)	Up to 30% (each channel has its own scaling factor)
Randomly scaling the Hue, Saturation, and Value	Up to 30%

Figure 6: **Example of a seafloor image augmented for analysis**. **The unaltered image (left-image) has been augmented by enhancing the brightness, flipping its vertical orientation, and adding rotation (right-image).**

2.4.4   Model evaluation

The performance of the models was assessed by evaluating them on their respective training dataset and validation dataset, compiling the predicted results against actual categories as confusion matrices (Table 6), and calculating several common metrics used on such binary classification problems: accuracy, precision, recall, F1-score, and Matthew's correlation coefficient (Chicco & Jurman 2020).

Table 6: **The confusion matrix, listing the total number of samples classified by their true and predicted categories. The total number of correct positive predictions is termed True Positive (TP), total number of correct negative predictions is True Negative (TN), total number of negative incorrectly classified as positive is the False Positive (FP), and total number of positive incorrectly classified as negative is the False Negative (FN).**

		Actual samples:	
		Actual positives (1):	Actual negatives (0):
Predicted samples:	Predicted positive (1):	True Positive (TP)	False Positive (FP)
	Predicted negative (0):	False Negative (FN)	True Negative (TN)

Accuracy, the most widely used classifier performance metric, is calculated by dividing the number of correct predictions (true positive and true negative) by the total number of predictions:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

The main default of this metric is that it is not suitable if data are strongly unbalanced. As a result, it has become common to calculate the alternative metrics Precision, Recall, and F1-score. Precision is the proportion of positive predictions that were correct, and Recall is the proportion of actual positives that were correctly predicted as such. The F1-score is the harmonic mean of those two terms. They are calculated as follows:

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

$$Recall = \frac{\text{TP}}{\text{TP+FN}} \tag{3}$$

$$F1\ score = 2 \cdot \frac{\text{precision·recall}}{\text{precision+recall}} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \tag{4}$$

Despite its popularity as an alternative to accuracy, the F1-score has recently been criticised for providing misleading information where a prediction displays many true positives but few true negatives, or many true negatives but few true positives (Chicco & Jurman 2020). Accordingly, the unbiased, normalised Matthew's correlation coefficient (nMCC) was also calculated, following Chicco & Jurman (2020):

$$MCC = \frac{\text{TP·TN−FP·FN}}{\sqrt{(TP+FP) \cdot (TP+FN) \cdot (TN+FP) \cdot (TN+FN)}} \tag{5}$$

$$nMCC = \frac{\text{MCC+1}}{2} \tag{6}$$

All five trained models were applied to all frames of all videos, producing prediction values between 0 and 1 for each pair of classes, summing to 1 (Table 7). Whichever class has the highest value (>0.5) is usually taken as the predicted class.

Table 7: **Example of predictions from models applied to video. The predictions from the macroalgae model applied to the first five frames of the first transect of the 2019 survey. For each class ('present' and 'absent'), the model outputs a number between 0 and 1, with the two numbers summing to 1. Whichever class is superior to 0.5 is usually taken as the class predicted.**

Frame	Prefix	0_absent	1_present
0	SWC1901_001	0.162671	0.837329
1	SWC1901_001	0.086542	0.913458
2	SWC1901_001	0.066806	0.933194
3	SWC1901_001	0.074461	0.925539
4	SWC1901_001	0.052704	0.947296

2.4.5 GPS interpolation

The precision of the GPS recorded was unbalanced, i.e., though at the latitude of Wellington, the latitudes recorded had a precision of 5 decimal places (approximately 1.1 m), the longitude measurements had a precision of 4 decimal places (approximately 8.4 m). Before being used for post-processing, both latitude and longitude outputs (from the metadata extraction process) were filtered (forward and backward) using a low-pass Butterworth filter of the 4th order with a critical frequency of 0.05 half-cycles per sample.

2.4.6 Averaging predictions and uncertainty

Because trained models are applied to all frames in a video, any given model produces 30 predictions for each second of footage. Moreover, these predictions are likely to be highly correlated because the tow camera and its field of view may not move significantly between each frame. Therefore, consecutive model predictions can be averaged to produce an output that is both more robust than the raw predictions per frame, and more suitable for analysis or display in a map.

The simplest averaging approach would be to retain whichever class was predicted the most frequently (i.e., the mode) over all frames included in the averaging period. However, the fact that model

predictions are graduated in their certainty (e.g., a frame for which macroalgae was predicted 'present' with a value of 0.99 is more certain to show macroalgae than a frame for which the value was 0.51), as well as the availability of frame suitability predictions (e.g., higher trust may be put in the result of any presence/absence model over a frame that was predicted to be 'suitable' with a value of 0.99 than a frame with a suitability value of 0.51), create an opportunity for improved forms of averaging.

Considering a set F of 120 consecutive frames of video footage (i.e., 4 seconds), the corresponding set of predicted 'presence' values p=$(p_1, p_2, \ldots, p_{120})$ (taking possible values between 0 and 1) produced by a given model (e.g., macroalgae, *Ecklonia, Lessonia,* or *Carpophyllum*) and the corresponding set of predicted 'frame suitability' s=$(s_1, s_2, \ldots, s_{120})$ (taking possible values between 0 and 1) produced by the frame suitability model, an overall presence value P_F and an overall frame suitability value S_F were calculated as:

$$P_F = \frac{\sum s_i(2p_i - 1)}{\sum s_i} \tag{7}$$

$$S_F = \frac{\sum s_i}{120} \tag{8}$$

The sign of P_F informs the overall class prediction for this set of frames:

$$C_F = \begin{cases} present, & P_F \geq 0 \\ absent, & P_F < 0 \end{cases} \tag{9}$$

$$U_F = \min(|P_F|, S_F) \tag{10}$$

U_F takes values between 0, indicating 'uncertain' and 1, indicating 'certain'. U_F is only close to 1 if it fits all three conditions: 1) the individual predicted presence values p_i for the set of frames tend to be certain (i.e., close to 0 or 1, not close to 0.5); 2) the individual predicted presence values p_i for the set of frames tend to agree with each other (i.e., a clear majority of predictions indicate either 0 or 1, not an even mix), and; 3) the individual predicted frame suitability values s_i for the set of frames tend to be high (i.e., S_F tends towards 1). Any unfulfilled condition will result in pulling the value of U_F towards 0.

2.4.7 Post processing and mapping

The metadata files previously exported (every 30 frames, i.e., every second) were reused. The navigation (latitude, longitude) were read, interpolated (as described previously), and averaged for each group of 120 consecutive frames, then joined with

1. the overall predictions C_F wherever $U_F > 0.5$, which indicated the highest certainty for determining the distributions of each species/group; and,

2. the overall uncertainty U_F was plotted, colour coded by three levels of certainty to examine where rare occurrences (within the 4-sec range) of each taxon, occurred.

The model outputs for both the 'overall predicted' and the 'overall uncertainty values' for each of the four species/groups were then exported as coma-delimited (csv) files. Shapefiles and distribution maps were then created in ESRI ArcGIS version 6.6, for both the 'overall predicted' and the 'overall uncertainty values' for all four taxon categories.

3. RESULTS

3.1 Model performance

The ML models performed well, with high detection accuracy and high levels of generalisation performance (Table 8 and Table 9, respectively), along with an overall accuracy of 97.2% achieved for the identification of macroalgae (F1-score: 0.985, nMCC: 0.876), and between 80.6 and 87.1% for individual species/groups (F1-score range: 0.738–0.815, nMCC range: 0.796–0.865, see Table 8). For all models, false negative (FN) predictions (i.e., predicting it wasn't there when it was) were more common than false positive (FP) predictions (i.e., predicting it was there when it wasn't), which translated in a Precision score higher than the Recall score.

Table 8: Confusion matrix and performance metrics from the evaluation of each model on the respective training dataset. TP: True Positive. TN: True Negative. FP: False Positive. FN: False Negative. Acc: Accuracy. Prec.: Precision. F1: F1-score. MCC: Matthew's correlation coefficient. nMCC: normalised Matthew's correlation coefficient. All performance metrics except MCC range between 0 for complete misclassification to 1 for perfect classification. The corresponding range for MCC is [-1:1].

Model	TP	TN	FP	FN	Acc.	Prec.	Recall	F1	MCC	nMCC
Frame suitability	6028	1532	162	117	**0.964**	0.974	0.981	0.977	0.894	0.947
Ecklonia	1556	4126	109	72	**0.969**	0.935	0.956	0.945	0.924	0.962
Lessonia	1629	4012	186	70	**0.957**	0.898	0.959	0.927	0.897	0.949
Carpophyllum	1485	4156	155	111	**0.955**	0.905	0.930	0.918	0.887	0.943
Macroalgae	5128	951	25	33	**0.991**	0.995	0.994	0.994	0.965	0.982

Table 9: Confusion matrix and performance metrics from the evaluation of each model on the respective validation dataset. TP: True Positive. TN: True Negative. FP: False Positive. FN: False Negative. Acc: Accuracy. Prec.: Precision. F1: F1-score. MCC: Matthew's correlation coefficient. nMCC: normalized Matthew's correlation coefficient. All performance metrics except MCC range between 0 for complete misclassification to 1 for perfect classification. The corresponding range for MCC is [-1:1].

Model	TP	TN	FP	FN	Acc.	Prec.	Recall	F1	MCC	nMCC
Frame suitability	1348	232	41	169	**0.883**	0.970	0.889	0.928	0.637	0.818
Ecklonia	373	728	87	178	**0.806**	0.811	0.677	0.738	0.592	0.796
Lessonia	410	851	37	149	**0.871**	0.917	0.733	0.815	0.729	0.865
Carpophyllum	245	1034	94	112	**0.861**	0.723	0.686	0.704	0.614	0.807
Macroalgae	1410	65	8	34	**0.972**	0.994	0.976	0.985	0.751	0.876

3.2 Time comparison

The main purpose for the development of ML algorithms for determining macroalgal presence/absence is the potential cost/time saving compared to manual post-processing. In this project, the authors estimated that the determination of macroalgal presence/absence from video imagery (i.e., in one image

per second of footage) took a specialist observer between 1 and 3 hours per transect, depending on transect length, the number of algae species to identify, and the amount of algae present.

This would result in any new survey of 40 video transects (i.e., approximately two days of survey for one site) requiring approximately 68 hours of staff time to process manually.

With the ML approach, efforts are split between active staff time (coding, expert labelling of training/testing dataset) and automated computing runtime (metadata extraction, training, inferring, post-processing), with the bulk of staff efforts (coding, and initial training dataset) having now been done and not needing to be repeated. For a new survey, up to 5 video transects may need expert labelling for testing and possibly retraining (see discussion in section 4), which would take up to 10 hours of staff time. All other efforts (inferring using trained models, testing, re-training) are automated and thus only necessitate computing runtime, with negligible staff involvement (i.e., copying videos to a suitable place, selecting the code parameters, and starting the algorithms).

Aside from the number and length of transects, the computing time for the post-processing of a new survey (i.e., metadata extraction, inference, results combination and interpolation, assuming no testing and re-training) is dependent on the number of models to be used and the desired prediction rate (i.e., all or a fraction of the frames in the videos), as well as the computing performance of the machine used. For this project, seeking predictions from all 5 models for all frames of 40 video transects, and using a single Nvidia Tesla P100 GPU from Māui AncillaCFry Nodes, the post-processing took approximately 10 hours of computing runtime and could thus be run overnight.

3.3 Distribution map

3.3.1 *Lessonia variegata*

Lessonia was prevalent on rocky reefs, in water depths of 3–10 m, throughout most of the study area (Figure 7a-b), and was the dominant species in most transects, with the exception of those on the northern, more sheltered side of Taputeranga Island, where the transects were instead dominated by *Carpophyllum*. *Lessonia* was not present on the deeper sections of the reefs, which were instead characterised by *Ecklonia* beds (Figure 8a). *Lessonia* was abundant in the exposed Houghton Bay and dominant in the western-most transect on the south-west side of Taputeranga Island. Artificial intelligence misclassification might be expected between *Lessonia* and *Marginariella* spp. due to similar blade morphologies, both having long strap like blades. However, although both *Lessonia* and *Marginariella* spp. were present in the survey area (*Marginariella* albeit not overly common), no misclassification occurred—indicating that the AI identification was very robust in distinguishing these two species. The blades of *Marginariella* have a distinct serrated margin and are a darker brown colour, compared to the smooth non-serrated yellower blades of *Lessonia*, which the AI classification accurately identified 100% as not *Lessonia*.

3.3.2 *Ecklonia radiata*

Ecklonia was common on exposed rocky reefs in water depths of 8–15 m (Figure 8a-b), growing either in monospecific beds or with other large brown seaweeds. *Ecklonia* was common at sites along the south-east coast out from Houghton Bay and in two transects off nearby Princess Bay. In contrast, *Ecklonia* was rarely found in the sheltered areas on the northern side of Taputeranga Island (the large island in the middle of Island Bay), or along the western-most transect on the south-west side of Taputeranga Island (Figure 8), which was instead dominated by *Lessonia* (Figure 7a-b). Clear shifts from dense *Lessonia* beds (shallower than10 m) into beds dominated by *Ecklonia* (deeper than10 m) were common at numerous sites but were most pronounced along the south-eastern side of Taputeranga Island (Figure 8 vs. Figure 7).

Model performance (based on detection certainty over 0.5) for *Ecklonia* was notably lower (80.6%) than for *Lessonia* or *Carpophyllum* (87.1 and 86.1%, respectively). Targeted examination of video footage, examining uncertainty/misclassifications of *Ecklonia* in the final predicted plots, identified that

the majority of these misclassifications reflected individual or small clusters of *Ecklonia* plants growing partially obscured within the *Lessonia* or mixed macroalgal beds. However, examination of the uncertainty plots for *Ecklonia* found that the ML model had correctly detected these cryptic plants as *Ecklonia* (and not *Lessonia*), but had given these identifications a low certainty value of 0.3–0.49, and even rarer occurrences, also correctly detected, were given a certainty value of under 0.3.

3.3.3 *Carpophyllum* spp.

Less common than *Ecklonia* and *Lessonia*, *Carpophyllum* was detected mostly on transects along the northern and north-west side of Taputeranga Island and in Houghton Bay (Figure 9a-b). *Carpophyllum* could have been confused with other branched brown macroalgal species, such as *Landsburgia* and/or *Sargassum*. These species were also present within the study area, including within transects where *Carpophyllum* occurred. Based on visual observations of the video footage at several sites, *Landsburgia* in particular looked quite similar to *Carpophyllum* plants (to an untrained eye), particularly where present within mixed macroalgal beds. The predicted distributions of *Carpophyllum* were therefore carefully examined to ensure that the ML models had not incorrectly included any *Landsburgia*. The final predicted distributions of *Carpophyllum* spp. not only accurately detected this genus, but also accurately distinguished *Carpophyllum* from other species including *Landsburgia*, with no *Landsburgia* included in the predicted *Carpophyllum* plots.

3.3.4 All macroalgae

Macroalgae were prevalent across the entire survey area occurring in over 90% of all towed video images (Figure 10a-b). This reflected the dominance of both monospecific and mixed macroalgal beds over the extensive rocky reefs in this region (Figure 10a). A diverse range of species was observed, with reef areas dominated by *Lessonia, Ecklonia*, and *Carpophyllum* (as described above) and several species of green macroalgae (specifically: *Ulva*, *Caulerpa flexilis*, and *C. brownii*), along with localised or rare occurrences of other large brown seaweed (e.g., *Cystophora*, *Sargassum*, *Landsburgia*, *Marginariella*, *Macrocystis*, and *Undaria*). In contrast to the rocky reefs, areas devoid of macroalgae were characterised by mobile sands and gravels that extended out beyond and between these rocky reefs. However, even in these mobile sediment habitats, the model accurately detected macroalgae, albeit with low certainty (under 0.3), based on the rare occurrence of drift algae. Locally dense patches of *Marginariella* that were recorded in several 2020 sites were examined to determine if these plants were misclassified in the species/taxon specific ML models; however, no misclassification due to *Marginariella* was found, instead it was correctly and consistently classified within 'all-macroalgae'.
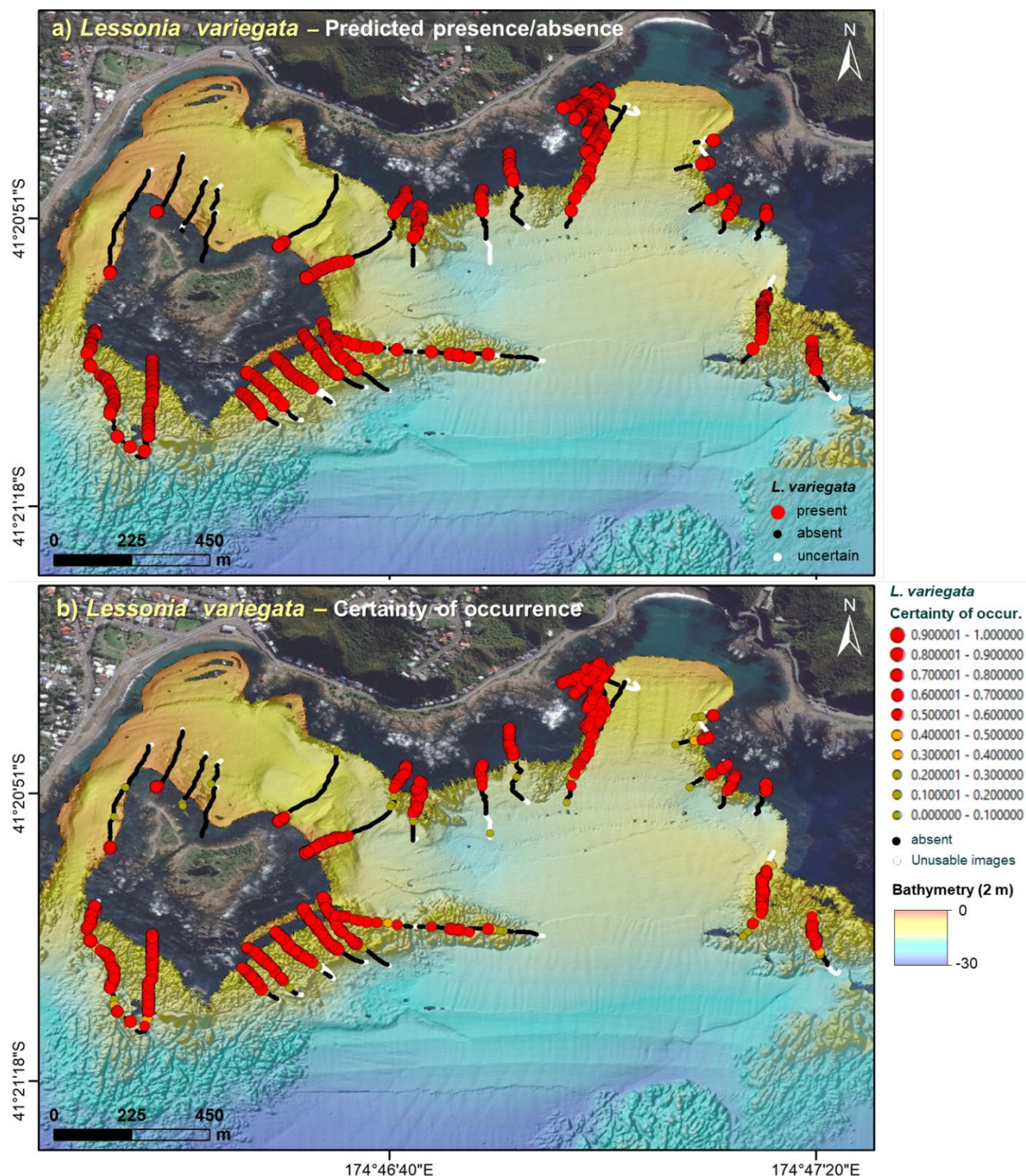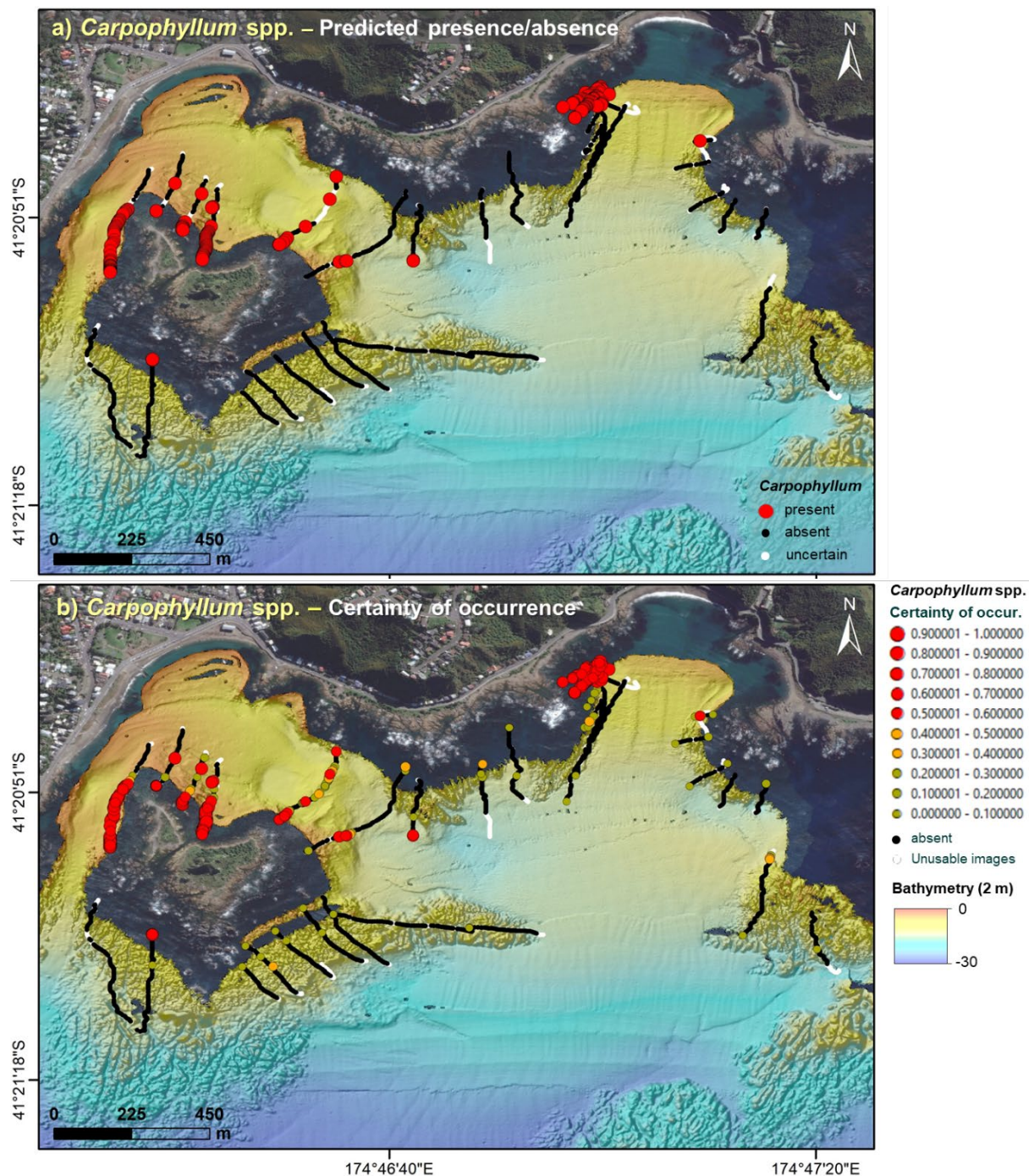
Figure 7: Baseline spatial distribution of *Lessonia variegata* between Houghton Bay and Island Bay, based on automated detection of presence/absence in video footage using the deep CNN developed by NIWA (model generalisation performance: 87.1% overall accuracy). a) Distribution of occurrence with certainty threshold at 0.5. Red circles = present with certainty > 0.5; black dots/lines = absent with certainty> 0.5; white lines = low certainty (< 0.5) due to poor quality imagery (e.g., topside, water column, too high or too close to the substratum, or poor water clarity), poor predictions, or mixed results within 4-sec. windows. b) Certainty of occurrence. Bubble size and colour depict the certainty of the model in detecting occurrence, where red = good certainty (> 0.5); orange = low-moderate certainty (0.3–0.5); muddy-green = very low certainty (> 0 and < 0.3).
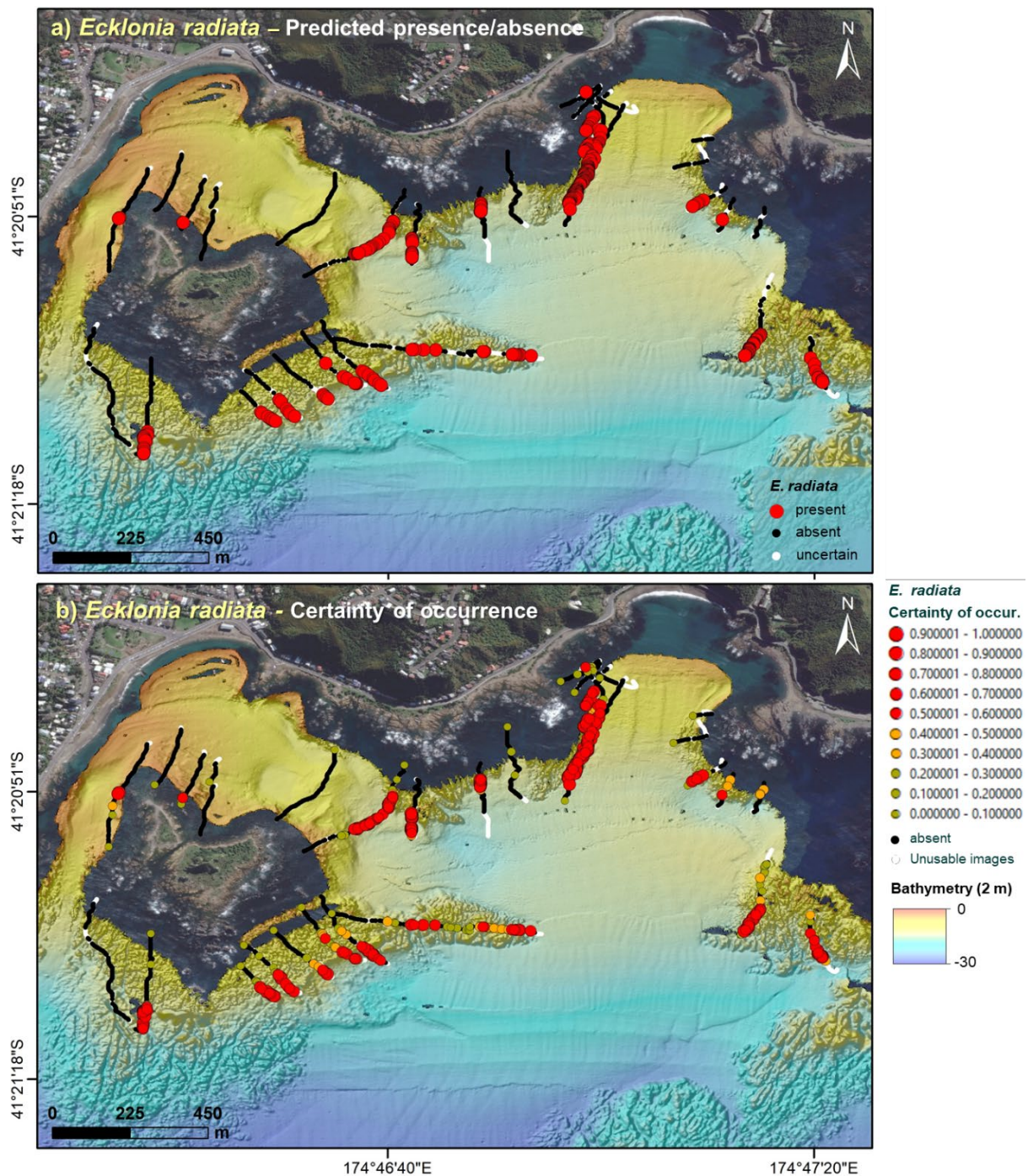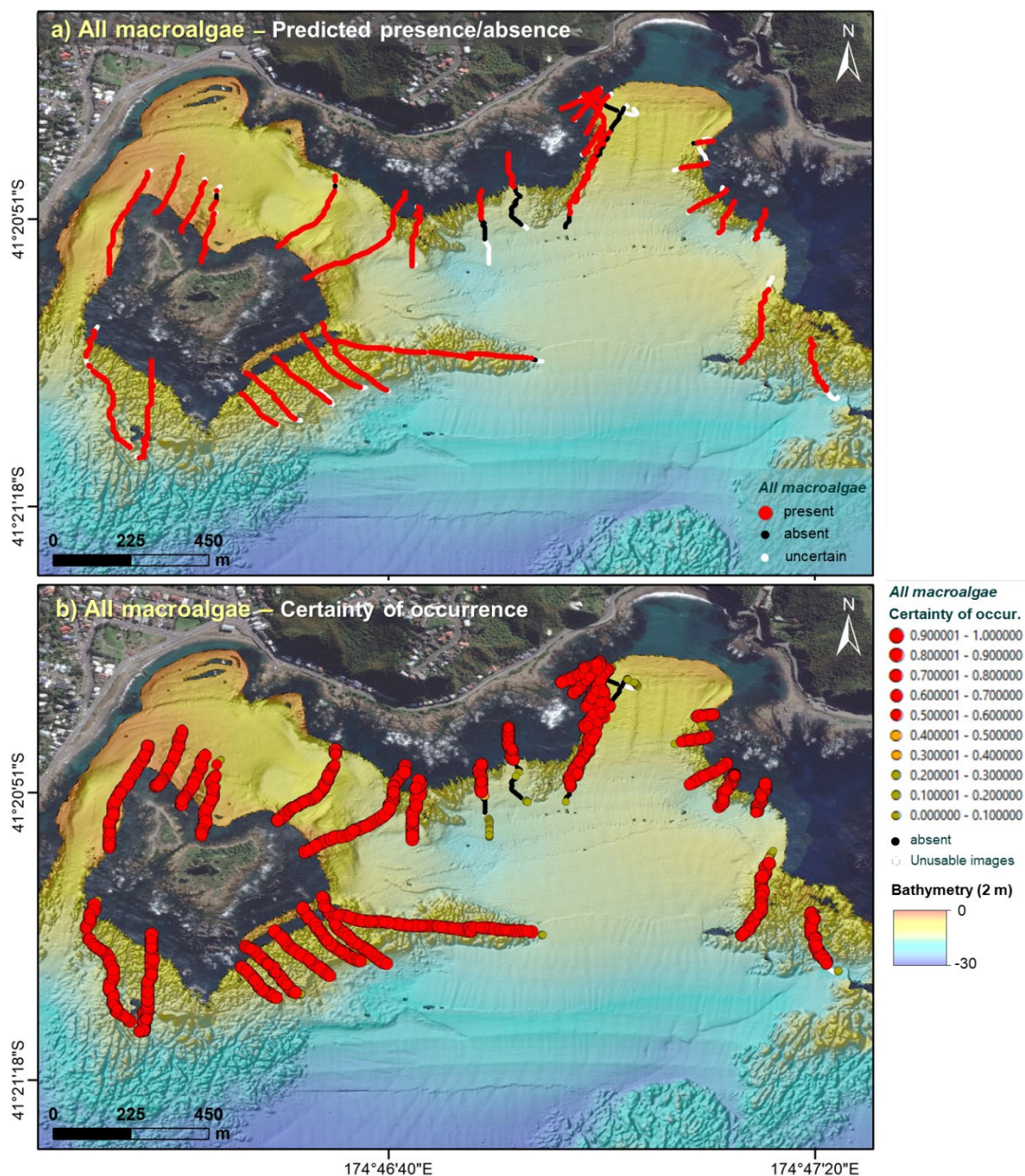
Figure 8: Baseline spatial distribution of *Ecklonia radiata* between Houghton Bay and Island Bay, based on automated detection of presence/absence in video footage using the deep CNN developed by NIWA (model generalisation performance: 80.6% overall accuracy). a) Distribution of occurrence with certainty threshold at 0.5. Red circles = present with certainty > 0.5; black dots/lines = absent with certainty > 0.5; white lines = low certainty (< 0.5) due to poor quality imagery (e.g. topside, water column, too high or too close to the substratum, or poor water clarity), poor predictions, or mixed results within 4-sec. windows. b) Certainty of occurrence. Bubble size and colour depict the certainty of the model in detecting occurrence, where red = good certainty (> 0.5); orange = low-moderate certainty (0.3–0.5); muddy-green = very low certainty (> 0 and < 0.3).

Figure 9: Baseline spatial distribution of *Carpophyllum* spp. between Houghton Bay and Island Bay, based on automated detection of presence/absence in video footage using the deep CNN developed by NIWA (model generalisation performance: 86.1% overall accuracy). a) Distribution of occurrence with certainty threshold at 0.5. Red circles = present with certainty > 0.5; black dots/lines = absent with certainty > 0.5; white lines = low certainty (< 0.5) due to poor quality imagery (e.g., topside, water column, too high or too close to the substratum, or poor water clarity), poor predictions, or mixed results within 4-sec. windows. b) Certainty of occurrence. Bubble size and colour depict the certainty of the model in detecting occurrence, where red = good certainty (> 0.5); orange = low-moderate certainty (0.3–0.5); muddy-green = very low certainty (> 0 and < 0.3).

Figure 10: **Baseline spatial distribution of all macroalgae between Houghton Bay and Island Bay, based on automated detection of presence/absence in video footage using the deep CNN developed by NIWA (model generalisation performance: 97.2% overall accuracy). a) Distribution of occurrence with certainty threshold at 0.5. Red circles = present with certainty > 0.5; black dots/lines = absent with certainty > 0.5; white lines = low certainty (< 0.5) due to poor quality imagery (e.g., topside, water column, too high or too close to the substratum, or poor water clarity), poor predictions, or mixed results within 4-sec. windows. b) Certainty of occurrence. Bubble size and colour depict the certainty of the model in detecting occurrence, where red = good certainty (> 0.5); orange = low-moderate certainty (0.3–0.5); muddy-green = very low certainty (> 0 and < 0.3).**

4. DISCUSSION

4.1 Machine learning model performance

In this study, NIWA's newly developed ML algorithms for detecting macroalgae (developed during the 'proof of concept' pilot study) were applied to determine whether AI could be used to successfully identify and distinguish different species, genera, and taxonomic groups, and whether this AI approach could be used to rapidly detect and map the distributions (presence/absence) of these macroalgal groups. NIWA's ML models were successful at identifying and distinguishing all four macroalgal groups across the entire survey, with overall models attaining high performance accuracies of between 80.6 to 87.1% for individual species and genera, and 97.2% for the 'all-macroalgae' group. Post-processing of macroalgal presence/absence was time consuming, and, in comparison, ML automated detection allowed a much greater volume of imagery to be processed in a much shorter timeframe. Although there was considerable time involved in the initial development of the ML algorithms (see D'Archino et al. 2019), and additional time required to set up the ML models for new datasets, once the setup was completed the ML models provided a rapid and accurate method to detect and map the distribution (presence/absence) of macroalgae at all four taxonomic levels.

Model performance varied among the four taxa/groups. Although model performance for individual species categories was high (80.6% to 87.1%), the lumped 'all-macroalgae' model produced the highest overall accuracy (97.2%). This is also true of the F1-score (0.985 for macroalgae, compared with 0.704–0.815 for individual species/taxon models) (see Chicco & Jurman (2020) for comparison of F1 and Matthew's coefficient). Although the immediate interpretation of this result is that the task of identifying the presence of macroalgae compared to its absence (e.g., bare sand) may be more trivial than identifying individual species, it is also very likely that these high scores reflect the class frequency imbalance in the validation dataset (1444 instances of 'present' samples for only 73 instances of 'absent', see Table 4). Using the unbiased normalised Matthew's correlation coefficient, the macroalgae model achieves a score (nMCC=0.876) that is more in line with the individual species/taxon models (nMCC ranging from 0.796 to 0.865).

Detection success was consistently higher for *Lessonia* (accuracy 87.1%, F1-score 0.815, nMCC 0.865) and *Carpophyllum* (accuracy 86.1%, F1-score 0.704, nMCC 0.807) than *Ecklonia* (accuracy 80.6%, F1-score 0.738, nMCC 0.796). This difference in model performance among taxa could be due to several factors. First, detection success might reflect differences in image quality with water depth, with improved species detection in shallower water with better natural light illumination. Both *Lessonia* and *Carpophyllum* spp. were recorded on shallow reefs, whereas *Ecklonia* (lower performing model) was more dominant on deeper reefs. It is unlikely, however, that depth related image quality is a key factor in the final models, because the process of image augmentation (duplicating and altering the colour and illumination of images in the training library) acts to negate any general biases of image quality due to factors such as depth, day, or seasonal water clarity/quality differences. In addition, misclassifications of *Ecklonia* occurred in both shallow and deeper depths. Second, within the survey area, *Lessonia* commonly occurred in relatively dense and extensive beds, whereas *Ecklonia* occurred in patchy beds or isolated clusters of plants in amongst other kelps. Numerous plants per image, along with numerous plants in subsequent images allows a particular model to gain higher certainty of identification (such as in the *Lessonia* model), whereas the identification of an isolated plant would register a low certainty in the model (e.g., *Ecklonia*). Targeted examination of video footage, examining misclassifications of *Ecklonia*, identified that most of these reflected individual, or small clusters of *Ecklonia* within and often beneath larger beds of *Lessonia*. It is unsurprising then that these isolated plants might be missed. However, examination of the uncertainty plots for *Ecklonia* found that the ML model had correctly detected these cryptic plants as *Ecklonia* (and not *Lessonia*), but these identifications had registered a low certainty rating of 0.3–0.49 (isolated small clusters of plants), or 0.1–0.29 (isolated individual plants). This indicates that the ML model accurately detected *Ecklonia*, and the uncertainty plots (based on the averaging of positive identifications for the '120-consecutive frames') provided additional and

very valuable information on the relative abundance of these beds (over 0.5) compared with the occurrence of isolated or infrequent subcanopy plants (under 0.5).

Lessonia and *Ecklonia* are closely related laminarian kelp species and were included in the current study to determine if the ML algorithms could accurately distinguish between kelp species. Inspection of both datasets found models did not misclassify *Lessonia* as *Ecklonia* or vice versa. Instead both training libraries were very efficient at distinguishing the species-specific features of these two species, highlighting the accurate fine-tuned performance of these models. Although misclassification with other macroalgal species was also possible, the findings highlight that these ML models were clearly able to distinguish between macroalgal species that have similar growth forms and colours, indicating that AI capabilities are rapidly approaching those of the human eye.

Both AI and ML models are being applied to the detection of marine species. Evaluation of these tools for detecting benthic marine species include studies on cold water corals (Purser et al. 2009), oysters (Ridge et al. 2019), and seagrass fishes (Ditria et al. 2020). In western Australia, one study has examined the use of AI in detecting large canopy and sub-canopy kelps (Mahmood et al. 2020). Mahmood et al. (2020) used a feature extraction approach to transfer learning to identify large mixed kelps, including *Ecklonia*, with an accuracy of 75.2–90.0% (ignoring the worst-performing models) and corresponding F1-score of 0.75–0.80. Mahmood et al. (2020) identified the presence of large canopy and sub-canopy kelp (both mixed genera) from 50 random points per image using the CATAMI scheme, while this study identified finer taxonomic (i.e., species, genera) and coarser taxonomic (all-macroalgae) categories from the entire surface of each image. Although comparison between these studies needs to be made cautiously due to methodological and modelling differences, the validation scores in Mahmood et al. (2020) for their lumped "large-canopy forming macroalgae" were in a similar range to the findings presented here, albeit slightly lower than the lumped 'all-macroalgae' category, both of which were higher than the scores for individual species/genera. Although the study presented here has been the only one published to date to use ML methods to detect and distinguish individual kelp species, both studies highlight the value of a ML approach to mapping ecologically relevant community indices.

Identifying the presence of macroalgae is not a trivial task, because species may occur at any density from a single plant to a large patch, and any level of community assemblage from monospecific to mixed algal habitats. Assessing absence also needs to be accurately identified. The post-processing methodology used in this study, which included averaging 120 consecutive predictions accounting for their respective uncertainty, proved to be a robust and simple approach to remove uncertain records. For example, rare and uncertain presence records would be dropped where no other presence records were recorded within the adjacent 120 images (i.e., over 0.5 certainty). In contrast, repeated but uncertain presence records in the adjacent 120 images would increase the probability that the taxon was indeed present. An unforeseen benefit of this approach was that this ability to calculate the average over the four-second window (i.e., the average estimate of occurrence over the 120 frames/4-sec neighbouring frames) provided an excellent means to distinguish rare plants (under 0.5 certainty) from beds of kelp for each taxon—here *Ecklonia, Lessonia*, and *Carpophyllum* (over 0.5 certainty). This two-step approach provided more information than a simple presence/absence.

4.2   Including new datasets to the ML models

An important consideration for future ML modelling is how to ensure the performance of this system remains adequate as new survey imagery is included. Variations between surveys, due for example to changes in seasonal water conditions or change in substratum type and location, imply that there is no guarantee that a model trained on past surveys, each with a particular suite of conditions, will perform as well on new surveys conducted under different suites of conditions. Some additional training, using imagery from a new survey, will likely be required to 'prime' the model for these new conditions. In the initial ML pilot study, only 20 transects surveyed in 2019 were available. Fifteen of these transects were initially labelled (i.e., identified by an observer), of which 12 were used for training (and the remaining three used for validation). In the current study, five of the additional 20 transects surveyed in 2020 were labelled and used to prime the model (i.e., include examples of these new conditions).

Two transects were used for training, which boosted the training dataset to 22 transects (all 20 transects of the 2019 survey, plus 2 of the 2020 survey), and the remaining three transects were used for validation, leaving the 15 unlabelled transects for inference. This approach effectively increased the training library to best capture image variability between the two surveys, ensuring better colour and morphology detection for each macroalgal species/group.

Based on the findings of this study, the approach below is recommended as general guidance for processing any additional video imagery datasets.

1. Label 25% of the new transects (e.g., 5 out of 20).

2. Evaluate the existing model on this dataset and calculate performance metrics.

    A. If performance is adequate, apply model to the rest of the survey to create survey results.

    B. If performance is inadequate, add some of the transects (e.g., 1–3 out of 5) to the pool of training data and retrain the model to prime it for the new survey. Evaluate new model performance on the remaining labelled transects.

        i. If performance is adequate, apply model to the rest of the survey.

        ii. If performance is inadequate, label more transects and repeat the training/validation process.

This iterative but conditional approach would result in increasing amounts of labelled data in the training datasets (to better account for spatial and temporal image variability) thus improving overall model performance. At the same time, the 'initial evaluation of performance' ensures that retraining and revalidation only occurs when necessary. Surveys in areas similar to those previously surveyed (e.g., nearby sites on the Wellington south coast) may require fewer training data to be allocated from each new survey as the data library increases (e.g., about 10%), whereas new and/or more remote locations with different kelp morphologies (e.g., other geographic areas) may require more images to be allocated for training (e.g., about 30%). Importantly, however, new macroalgal video datasets would sequentially add to the number of training and validation images in the model, and thus slowly increase both the generality and predictive capability of these ML models.

The ML models programmed using the NIWA's 'seaweedai' Python package will require ongoing maintenance and updating. New versions of the training libraries may include more recent (and improved) base classification architectures, as well as new or improved training tools. However, the seaweedai code is also flexible to implement minute modifications to capture these and other changes as required.

4.3   Changes in model performance

In the initial pilot study (D'Archino et al. 2019), ML models were developed to identify presence/absence of *Lessonia*, *Ecklonia*, *Carpophyllum*, and all-macroalgae. Model performance for these taxon/groups in the pilot study were comparable with those achieved in the current study, with 85% for *Lessonia* (compared with 87.1% now); 86.4% for *Carpophyllum* (compared with 86.1% now); and 92.7% and 98.8% (two results were obtained on two different validation datasets in the pilot study) compared with 97.2% in the current study. Model performance for *Ecklonia*, however, was lower in the current study (80.6%) than the initial pilot study (89%). As discussed above, lower performance of the *Ecklonia* model appears to reflect the inclusion of the 2020 sites, where isolated or small clusters of *Ecklonia* were present within kelp beds dominated by *Lessonia* but were accurately detected and plotted in the uncertainty plots.

Other factors may also alter model performance. The previous pilot study used a training dataset of limited size for the task, with the four models having been trained on only 836 (*Ecklonia*), 797 (*Lessonia*), 807 (*Carpophyllum*), and 1400 (macroalgae) images, respectively (D'Archino et al. 2019).

The present study was carried out, in part, to obtain more training data to boost the models' performance. The authors suggest several reasons why the measured performance did not increase in the new study:

    1.   The training datasets were built differently.

In the pilot study, the training/validation datasets were built by a non-macroalgae expert (C. Peat) selecting out of a pool of images instances of 'present' and 'absent' classes. Therefore, the datasets were built with the most evident examples of both classes. Although this approach is sound for a pilot study meant to assess the discriminating capabilities of a classifier, it is less suitable for the development of a practical system for survey footage analysis, which would include a wider range of uncertain cases. In addition, this approach does not easily scale up with additional videos and surveys.

The building of the datasets in the current study was more adapted to the development of such a system. The datasets were built by having a macroalgae expert (R. D'Archino) labelling all images extracted at a rate of 1 frame per second, from selected survey transects. This approach resulted in more uncertain cases, with only the most uncertain being discarded from the labelling process (by use of the '2' or 'uncertain' label in the csv files). This variety of cases is necessary to train a robust model with the capacity to work with more uncertain imagery, but a loss of generalisation performance could be expected, compared with the pilot study and its more ideal training dataset.

    2.   New models were not evaluated on the same validation dataset.

For an unbiased comparison of the performance of two models, these would ideally be evaluated on the same validation dataset. This was not done here because a new validation dataset was developed from the expert labelling (Table 4). As demonstrated in the pilot study by the wide difference between the two accuracy scores (respectively 92.7% and 98.8%) obtained from the validation of the macroalgae model on two different validation datasets (respectively 138 images from transects at Moa Point, and 90 images from transects at Palmer Head), the evaluation of the performance of a model is highly dependent on each validation dataset. In addition, the sizes of the validation datasets in the pilot study were perhaps too small (respectively 100, 100, and 103 images from both Moa Point and Palmer Head for the *Ecklonia*, *Lessonia*, and *Carpophyllum*) for the performance metrics to be precise (compared with 1366, 1447, and 1485 respectively in the new validation datasets, see Table 4).

    3.   Differences in methodology.

Besides the difference in the build and size of the training/validation datasets, the new methodology is also different from that of the pilot study, with the use of a different CNN as a base model (InceptionV3 vs. Inception-Resnet-V2 in the pilot study) and the transfer learning approach (fine tuning vs. features extraction in the pilot study). In the future, different models and parameters (hyper-parameter tuning) will be tested and compared to elucidate whether in fact these differences have a significant impact on performance, compared with simply the dataset building method and size.

    4.   Modelling increased natural complexity.

Measured performance may decrease with increasing dataset size, because large data libraries will likely include a larger universe of data, and therefore by definition increased variation for each target taxon. In the current study, additional sites included new areas with numerous sites surveyed around Taputeranga Island, along with more sites across deeper offshore reefs. These new video sites included a broader range of plant morphologies, sizes, conditions, and bed densities, and more variations in image quality (e.g., colour, illumination, altitude above the seabed). Updated ML models need to learn to account for this increased variability. Larger image libraries are also likely to include more images of other co-occurring species of macroalgae (and their inherent variations), which need to be distinguished from the target groups. Consequently, although performance may drop as a reflection of this increased variation (i.e., as some of this variation is missed or misclassified), these updated models will become

more realistic representations of the natural ecosystems and will therefore become more robust at predicting new untrained sites (i.e., improved generality).

4.4   Future recommendations/next steps

This study selected four taxonomic categories to examine the potential of using AI to map macroalgal species distributions. Following the success of this approach, the authors highly recommend extending the ML algorithms to detect other key macroalgal taxa. This will allow the monitoring of changes in key habitat forming taxa given that increasing temperature and ocean acidification will impact macroalgal species differently (Hepburn et al. 2011, Cornwall 2012, 2017). It would be possible to include, for example, other dominant taxa that were present within the study area including green algae (e.g., *Caulerpa*, *Ulva*), other large brown algae (e.g., *Sargassum*, *Marginariella*, and *Landsburgia*), and other key indicator species (e.g., *Macrocystis* and *Durvillaea*). Although *Macrocystis* and *Durvillaea* were not particularly abundant within the current survey area, both species are common along the south coast of Wellington and play critical roles in the coastal marine ecosystems.

Although presence/absence data provide valuable information on the distribution of these macroalgae, quantifying and monitoring the abundance (or percentage cover) of key species would be more informative. Monitoring the abundance and the species composition of kelp forests would provide reliable baseline data to detect future modifications due to climate change or anthropogenic impacts. The underwater videos acquired nowadays could then be re-analysed with improved algorithms and larger image library. Consequently, the next logical step in ML coding would be to extend the ML algorithms to estimate percentage cover for the four macroalgal groups. This could be done as a relative estimate, i.e., the amount of each algal type within the image or standardised by the area of seafloor seen within the image. For the latter to be calculated, it would be necessary first for the AI code to detect the two, 20-cm spaced laser points within each image and then calculate the distance between the lasers, to then use this information to determine the altitude off the seafloor and the area of seafloor within the frame. Although ML code already exists to detect and measure laser points from video imagery (e.g., Schoening 2015), this or similar code would need to be applied and fine-tuned for the macroalgal-imagery and ML models.

The success of this ML approach means that this method can now be used to survey other locations and undertake repeated sampling at sites. Another recommendation is the inclusion of new areas where other key taxa are present, such as *Macrocystis* and *Durvillaea* which are more common outside of the Taputeranga Marine Reserve. These additional surveys could then be used with the current study to compare macroalgal occurrence and distributions between areas inside and outside the Taputeranga Marine Reserve (and other marine reserves). This approach would be extremely valuable to compare the ongoing effects of environmental change and different management and conservation strategies.

Improvements between exporting ML output and plotting resulting in GIS software could also be streamlined by integrating the output of the ML python code with R-coding so that macroalgal distribution plots, based on a designed template, are automatically produced rather than plotted individually in GIS software. This would further streamline map production and reduce laborious staff hours, especially where numerous taxa are being plotted.

Importantly, all video images from NIWA surveys (along with imagery collected by other institutes and agencies) are archived, and as more species are included in the ever-expanding training libraries, previously archived imagery can be easily and quickly analysed to detect these and other newly-detected taxa, with no need to spend months reprocessing these images by eye.

4.5   Kelp monitoring and management

Kelp forests are susceptible to changing environmental conditions, including changes in water temperatures, increasing turbidity, sediment deposition, and ocean acidification (D'Archino et al. 2019, Anderson et al. 2019, Reed et al. 2016). The ML auto-detection models developed in this current study

can now be readily applied to ongoing monitoring programmes to rapidly determine and map the distributions of key macroalgal indicator species along coastlines, particularly targeting areas where environmental change is likely or may already be occurring. Monitoring has traditionally been a costly staff-intensive undertaking that in most situations has prohibited its use. However, monitoring data are critical to documenting changes in our coastal communities, particularly the shifting baselines that many of our marine environments are facing. For example, the decline of kelp forests is a serious issue affecting many temperate rocky reefs globally (see D'Archino et al. 2019). Large scale monitoring is necessary for coastal resource management to make informed decisions. This study highlights that ML auto-detection of macroalgal occurrence is a powerful tool to determine spatial distributions for species, genera, and group level identifications with great accuracy, and thus it provides a feasible tool to monitor and map future contraction or expansion of macroalgal distributions over both local and national scales. The ML auto-detection approach is also rapid and cost-efficient making it readily applicable to mapping and monitoring macroalgal communities over broader spatial and temporal scales and, as such, is likely to play an important role in future management and conservation of marine ecosystems.

5.   CONCLUSIONS

- This study successfully shows that ML is significantly faster than traditional video analysis and is a cost-effective approach to post-process video imagery, and that will improve as more data are collected.
- The ML models performed well at detecting and distinguishing between closely related species (*Lessonia* and *Ecklonia*), highlighting the accurate fine-tuned performance of these models.

6.   ACKNOWLEDGMENTS

7.   REFERENCES

Anderson, T.J.; Morrison, M.; MacDiarmid, A.; D'Archino, R.; Nelson, W.; Tracey, D.; Clark, M.; Gordon, D.; Read, G.; Morrisey, D.; Kettles, H.; Wood, A.; Anderson, O.; Smith, A.M.; Page, M.; Paul-Burke, K.; Schnabel, K.; Wadhwa. S. (2019). Review of New Zealand's Key Biogenic Habitats. (NIWA Client Report 2018139WN, prepared for the Ministry for the Environment.) 184 p. https://environment.govt.nz/assets/Publications/Files/NZ-biogenic-habitat-review.pdf

Beijbom, O.; Edmunds, P.J.; Roelfsema, C.; Smith, J.; Kline, D.I.; Neal, B.P; Dunlap, M.J.; Moriarty, V.; Fan, T.-Y.; Tan, C.-J.; Chan, S.; Treibitz, T.; Gamst, A.; Mitchell, B.G.; Kriegman, D. (2015). Towards Automated Annotation of Benthic Survey Images: Variability of Human Experts and Operational Modes of Automation. *PLoS ONE 10(7)*: e0130312. https://doi.org/10.1371/journal.pone.0130312

Bennett, S.; Wernberg, T.; Connell, S.D.; Hobday, A.J.; Johnson, C.R.; Poloczanska, E.S. (2016). The 'Great Southern Reef': social, ecological and economic value of Australia's neglected kelp forests. *Marine and Freshwater Research 67(1)*: 47–56. https://doi.org/10.1071/MF15232

Bewley, M.S.; Douillard, B.; Nourani-Vatani, N.; Friedman, A.; Pizarro, O.; Williams, S.B. (2012). Automated species detection: An experimental approach to kelp detection from sea-floor AUV images. Proceedings of Australasian Conference on Robotics and Automation, 3–5 Dec 2012, Victoria University of Wellington, New Zealand. 10 p.

Chicco, D.; Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics* 2020 *21*: 6 https://doi.org/10.1186/s12864-019-6413-7

Coleman, M.A.; Wernberg, T. (2017). Forgotten underwater forests: The key role of fucoids on Australian temperate reefs. *Ecology and Evolution 7*: 8406–8418. https://doi.org/10.1002/ece3.3279

Cornwall, C.; Revill, A.; Hall-Spencer, J. . (2017). Inorganic carbon physiology underpins macroalgal responses to elevated CO_2. *Scientific Reports 7*: 46297. https://doi.org/10.1038/srep46297

Cornwall, C.E.; Hepburn, C.D.; Pritchard, D.; Currie, K.I.; McGraw, C.M.; Hunter, K.A. Hurd, C.L. (2012). Carbon-use strategies in macroalgae: differential responses to lowered pH and implications for ocean acidification. *Journal of Phycology 48*: 137-144. https://doi.org/10.1111/j.1529-8817.2011.01085.x

D'Archino, R.; Neill, K.F.; Nelson, W.A.; Fachon, E.; Peat, C. (2019). New Zealand macroalgae: distribution and potential as national scale ecological indicators. *New Zealand Aquatic Environment and Biodiversity Report No. 207*. 217 p.

Ditria, E.M.; Lopez-Marcano, S.; Sievers, M.; Jinks, E.L.; Brown, C.J.; Connolly, R.M. (2020). Automating the Analysis of Fish Abundance using Object Detection: Optimizing Animal Ecology with Deep Learning. *Frontiers in Marine Science 7(429)*: 1–9. 10.3389/fmars.2020.00429

Dayton, P.K. (1985). Ecology of kelp communities. *Annual Review of Ecology & Systematics 16*: 215–245. https://doi. org/10.1146/annurev.es.16.110185.001243

Graham, M.H.; Vásquez, J.A.; Buschmann, A.H. (2007). Global ecology of the giant kelp *Macrocystis*: from ecotypes to ecosystems. *Oceanography and Marine Biology: An Annual Review 45*: 39–88.

Griffin, K.J.; Hedge, L.H.; Gonzalez-Rivero, M.; Hoegh-Guldberg, O.I.; Johnston, E.L. (2017). An evaluation of semi-automated methods for collecting ecosystem-level data in temperate marine systems. *Ecology and Evolution 7(13)*: 4640–4650. http://dx.doi.org/10.1002/ece3.3041

Gu, J.; Wang, Z.; Kuen, J.; Ma, L.; Shahroudy, A.; Shuai, B.; Liu, T.; Wang, X.; Wang, G.; Cai, J.; Chen; T. (2018). Recent advances in convolutional neural networks. *Pattern Recognition 77*: 354–377.

Hepburn; C.D.; Pritchard, D.W.; Cornwall, C.E.; Mcleod, R.J.; Beardall, J.; Raven, J.A.; Hurd, C.L. (2011). Diversity of carbon use strategies in a kelp forest community: implications for a high CO_2 ocean. *Global Change Biology 17*: 2488–2497. https://doi.org/10.1111/j.1365-2486.2011.02411.x

Krumhansl, K.A.; Okamoto, D.K.; Rassweiler, A.; Novak, M.; Bolton, J.J.; Cavanaugh, K.C.; Connell, S.D.; Johnson, C.R.; Konar, B.; Ling, S.D.; Micheli, F.; Norderhaug, K.M.; Perez-Matus, A.; Sousa-Pintol, I.; Reed, D.C.; Salomon, A.K.; Shears, N.T.; Wernberg, T.; Anderson, R.J.; Barrett, N.S.; Buschmanns, A.H.; Carr, M.H.; Caselle, J.E.; Derrien-Courtel, S.; Edgar, G.J.; Edwards, M.; Estes, J.A.; Goodwin, C.; Kenner, M.C.; Kushner, D.J.; Moy, F.E.; Nunn, J.; Stenecka, R.S.; Vsquezb, J.; Watsonc, J.; Witmand, J.D.; Byrnese, J.E.K. (2016). Global patterns of kelp forest change over the past half-century. *Proceedings of the National Academy of Sciences of the United States of America 113(48)*: 13785–13790. http://dx.doi.org/10.1073/pnas.1606102113

Layton, C.; Shelamoff V.; Cameron, M.J.; Tatsumi, M.; Wright, J.T.; Johnson, C.R. (2019). Resilience and stability of kelp forests: The importance of patch dynamics and environment-engineer feedbacks. *PLoS ONE 14(1)*: e0210220. https://doi.org/10.1371/journal.pone.0210220

Lei, C.; Hu, B.; Wang, D.; Zhang, S.; Chen, Z. (2019). A Preliminary Study on Data Augmentation of Deep Learning for Image Classification. *Proceedings of the 11th Asia-Pacific Symposium on Internetware.* Fukuoka, Japan: Association for Computing Machinery.

Mahmood, A.; Ospina, A.G.; Bennamoun, M.; An, S.; Sohel, F.; Boussaid, F.; Hovey, R.; Fisher, R.B.; Kendrick, G.A. (2020). Automatic Hierarchical Classification of Kelps Using Deep Residual Features. *Sensors 20*: 447.

Malde, K.; Handegard, N.O.; Eikvil, L.; Salberg, A.B. (2020). Machine intelligence and the data-driven future of marine science. *ICES Journal of Marine Science 77*: 1274–1285.

Mann, K.H. (1973). Seaweeds: Their Productivity and Strategy for Growth. *Science 182(4116)*: 975–981. http://dx.doi.org/10.1126/science.182.4116.975

Moniruzzaman, M.; Islam, S.M.S.; Bennamoun, M.; Lavery, P. (2017). Deep Learning on Underwater Marine Object Detection: A Survey. *In*: Proceedings of the International Conference on Advanced Concepts for Intelligent Vision Systems. 150 p.

Murie, K.A.; Bourdeau, P.E. (2020). Fragmented kelp forest canopies retain their ability to alter local seawater chemistry. *Scientific Reports 10***: 11939.

Nelson, W.A.; Neill, K.; D'Archino, R.; Rolfe, J.R. (2019). Conservation status of New Zealand macroalgae, 2019. *New Zealand Threat Classification Series 30*. Department of Conservation, Wellington. 33 p.

Purser, A.; Bergmann, M.; Lundälv, T.; Nattkemper, T. (2009). Use of machine-learning algorithms for the automated detection of cold-water coral habitats: A pilot study. *Marine Ecology Progress Series* 397: 241–251.

Rawat; W.; Wang, Z. (2017). Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. *Neural Computation 29*: 2352–2449.

Reed, D.; Washburn, L.; Rassweiler, A.; Miller, R.; Bell, T.; Harrer, S. (2016). Extreme warming challenges sentinel status of kelp forests as indicators of climate change. *Nature Communications 7*. http://dx.doi.org/10.1038/ncomms13757

Ridge, J.T.; Gray, P.C.; Windle, A.E.; Johnston, D.W. (2019). Deep learning for coastal resource conservation: automating detection of shellfish reefs. *Remote Sensing in Ecology and Conservation 6(4)*: 431–440. doi.org/10.1002/rse2.134.

Schoening, T. (2015). Automated detection in benthic images for megafauna classification and marine resource exploration: supervised and unsupervised methods for classification and regression tasks in benthic images with efficient integration of expert knowledge. PhD Thesis. Bielefeld: Universitätsbibliothek Bielefeld: 185.

Seiler, J.; Friedman, A.; Steinberg, D.; Barrett, N.; Williams, A.; Holbrook, N.J. (2012). Image-based continental shelf habitat mapping using novel automated data extraction techniques. *Continental Shelf Research 45*: 87–97.

Shorten, C.; Khoshgoftaar, T.M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data 6*: 60.

Teagle, H.; Hawkins, S.J.; Moore, P.J.; Smale, D.A. (2017). The role of kelp species as biogenic habitat formers in coastal marine ecosystems. *Journal of Experimental Marine Biology and Ecology 492*: 81–98. http://dx.doi.org/https://doi.org/10.1016/j.jembe.2017.01.017

Thomsen, M.S.; Mondardini, L.; Alestra, T.; Gerrity, S.; Tait, L.; South, P.M.; Lilley, S.A.; Schiel, D.R. (2019). Local Extinction of Bull Kelp (*Durvillaea* spp.) Due to a Marine Heatwave. *Frontiers in Marine Science 6*. https://doi.org/10.3389/fmars.2019.00084

Warrant, E.J.; Locket, N.A. (2004). Vision in the deep sea. *Biological Reviews 79*: 671–712. doi:10.1017/S1464793103006420

Yamashita, R.; Nishio, M.; Do, R.K.G.; Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. *Insights Imaging 9*: 611–629. https://doi.org/10.1007/s13244-018-0639-9

Zuccarello, G.C.; Martin, P. (2016). Phylogeography of the *Lessonia variegata* species complex (Phaeophyceae, Laminariales) in New Zealand. *ALGAE 31(2)*: 91–103.

GLOSSARY

AEBR	Aquatic Environment & Biodiversity Report
CNN	Convolutional Neural Network
FPS	Frames Per Second
GPS	Global Positioning System
HDMI	High-Definition Multimedia Interface
HSV	Hue-Saturation-Value/Brightness
MCC	Matthew's correlation coefficient
ML	Machine Learning
MPI	Ministry for Primary Industries
nMCC	normalised Matthew's correlation coefficient
NIWA	National Institute of Water and Atmospheric Research
RGB	Red-Green-Blue
SVM	Support Vector Machine

APPENDIX

Seaweed AI - Identifying macroalgae in towed underwater video with Deep Learning

seaweedai is a Python package to identify macroalgal species in underwater video footage from NIWA's SeaweedCam, using deep-learning image classification. It implements a transfer learning technique using the Inception V3 deep neural network pretrained for computer vision, made available in the TensorFlow/Keras libraries.

1. Context

1.1 Raw data

Raw data are videos from the SeaweedCam shallow water drop-camera, are in the QuickTime File Format (.MOV). The main repository for the videos is R:\National\Datasets\Seaweed_Video_Data\raw_data.

Each survey with the SeaweedCam is coded as NIWA voyage codes using **SWC** as the "vessel" code, aka SWCXXYY with XX being the year and YY the survey number that year. The main repository for videos contains preliminary surveys from testing the system and two actual surveys: SWC1901 and SWC2001, both covering the Wellington south coast (Island Bay, Houghton Bay, Breaker Bay).

Each transect is recorded as a separate video. The name of the video file has the transect number as a suffix (e.g. SWC1901_001.MOV for transect #1). Occasionally, the system closes the video file and starts a new one DURING a transect. In these cases, the video files are appended with a letter (e.g. SWC1901_009.MOV and SWC1901_009B.MOV for transect #9).

1.2 Pre-processed data

The videos are pre-processed into two types of data:

- Individual frames extracted from the videos at regular intervals (1 every second), to use for labelling and training the models. They are recorded as png files named after the video and the frame number in the video (prefix_frame.png, e.g. SWC1901_001_270.png for frame #270 in video SWC1901_001.MOV).

- One csv file per video containing, for each extracted frame, the metadata hard-coded in the frame ('date', 'time', 'latitude', 'longitude'), the calculated time since start of video ('video_time'), or frame information ('prefix' and 'frame').

1.3 Labels

To inform on the content of the frames for training (labelling), we append additional columns to the csv files and complete each row.

Currently, the standard additional columns are: 'comment', 'good-frame', 'turfing_or_foliose_algae_on_sand', 'sand', 'algae', 'Lessonia', 'Carpophyllum', 'Ecklonia', 'Ulva', 'Cystophora', 'Macrocystis', 'Sargassum', 'Marginariella', 'Caulerpa_flexilis', 'Caulerpa_brownii', 'Undaria', 'Landsburgia', 'cobbles'.

Others can be added. Ideally, respect the nomenclature:
- No spaces. Use underscores between words (e.g., good-frame)

- No capitalisation for non-species (e.g., algae, cobbles)

- First word is capitalised for species (e.g., Lessonia, Caulerpa_brownii)

'comment' is a field for free use. All others only accept a limited number of entries: 0, 1, or 2 (or stay empty)

For 'good-frame':

- '0' (or empty) means 'this is a bad/unusable frame'. Typically applies to footage acquired while on the boat, too far from the bottom, stuck against the reef, view obstructed by algae, etc.
- '1' means 'this is a good/usable frame'.
- '2' means 'Unsure. Ignore this frame in training.'

For all others:

- '0' (or empty) means 'absent'.
- '1' means 'present'.
- '2' means 'Unsure. Ignore this frame in training.'

The main repository for the prepared and labelled dataset is R:\National\Datasets\Seaweed_Video_Data\ai\data.

2. Dependencies

This package uses dependencies that were up-to-date at the time of development but now outdated. Including:

- Python 3.6.8
- TensorFlow 1.12

It will eventually need upgrading, especially to TensorFLow 2.0.

All dependencies are listed in the ./conda_env/seaweed.yml file. Using conda, you can create the appropriate environment with:

conda env create --file ./conda_env/seaweed.yml

3. Pre-processing

Pre-processing is the task of extracting pre-processed data (.png frame files and .csv metadata files) from raw data (.MOV videos). These are performed with functions seaweedai.preprocess.preprocess_videos.preprocess_video (for one video) or seaweedai.preprocess.preprocess_videos.preprocess_videos (for multiple videos).

preprocess_video opens an input video and for every Nth frame, an can do either or both of (1) extracting the metadata that is hard-coded in the frame, reformatting it (lat, long, date, time), adding information, and saving it all in a single csv file; and (2) saving the frames as individual png files.

preprocess_videos runs preprocess_video over videos found in a root folder specified in input. Substrings can be specified to limit the process to desired videos.

Python script used for this project:

from seaweedai.preprocess.preprocess_videos import preprocess_videos

input_videos_root_dirpath = r'R:\National\Datasets\Seaweed_Video_Data\raw_data0'

input_videos_file_substrings = ['SWC1901', 'SWC2001']

every_nth_frame = 30

verbose = 1

output_dirpath = r'R:\National\Datasets\Seaweed_Video_Data\ai\data0'

preprocess_videos(input_videos_root_dirpath, input_videos_file_substrings, every_nth_frame, verbose, output_dirpath)

4. Labelling

Once frames and the corresponding csv file are created, the labels must be manually informed.

4.1 csv approach

The main approach is to edit the csv file produced. Add categories as new columns and for each row, complete the corresponding cell with "0" (or leave empty) for absent, "1" for present, or "2" for uncertain.

You can use seaweedai.label.add_categories_to_csv_files.add_categories_to_csv_file to automatically add columns in one csv file, or seaweedai.label.add_categories_to_csv_files.add_categories_to_csv_files for all csv files within a root folder (recursively). Both functions can optionally prefill the cells with a set value. Not specifying the categories in input will add the default list of categories ('comment', 'good-frame', 'turfing_or_foliose_algae_on_sand', 'sand', 'algae', 'Lessonia', 'Carpophyllum', 'Ecklonia', 'Ulva', 'Cystophora', 'Macrocystis', 'Sargassum', 'Marginariella', 'Caulerpa_flexilis', 'Caulerpa_brownii', 'Cyathea_brownii', 'Undaria', 'Landsburgia', 'cobbles').

Python script used for this project:

from seaweedai.label.add_categories_to_csv_files import add_categories_to_csv_files

input_root_dirpath = r'R:\National\Datasets\Seaweed_Video_Data\ai\data'

add_categories_to_csv_files(input_root_dirpath)

4.1 folder approach

The issue with the approach of labelling in a csv file is that mistakes are difficult to correct. It is not easy to spot a wrong number in an array of numbers. It would be much easier to spot mistakes if frames with the same label were grouped into individual, distinct folders.

This alternative approach of labelling is possible:

1. Copy the entire set of frames obtained from pre-processing into a new folder named after a category of interest.
2. In this folder, create subfolders "0_absent", "1_present", and "2_uncertain".
3. Inspect frames and move them into the appropriate folder.

The default of this alternative approach is that if you use several categories, you need to copy the whole set of frames as many times as there are categories. This undesirable duplication of data is the reason why training operates on the labelled csv files, and not on folders of duplicated frames. It's also much slower as you go over each frame as many times as there are categories

A solution is to use both approaches:

1. Add categories to the csv file and label this file.
2. Turn this labelled csv into classified folders (using seaweedai.label.csv_to_folder.csv_to_folder).
3. Examine the contents of each folder and correct it if necessary (by moving frames between the "0_absent", "1_present", and "2_uncertain" subfolders of a category).

4. Turn this edited folder labelling into a new csv file (using seaweedai.label.folder_to_csv.folder_to_csv).

seaweedai.label.csv_to_folder.csv_to_folder automatically copies (or moves) png frames from an input folder, into the subfolders of an output folder containing subfolders named as per the labelling info in a csv file.

seaweedai.label.folder_to_csv.folder_to_csv writes and fills in a new csv file as per organization of frames in the subfolders (labels) of a category folder of a root folder. The function needs to know the original csv file to copy the metadata.

5. Training

Once the dataset is prepared (folders of frames + labelled csv file), one can run the training function, specifying (among other parameters) the path to the dataset and the desired category to run training on.

seaweedai.train.trainer_categorical.trainer_categorical performs this task. It will find the csv files in the subfolders of the data root folder (data_root_dirpath) and read the data for the category of interest (category) and relevant labels (labels). Datasets for the training and testing phases can be specified (train_file_substrings and test_file_substrings). The datasets can be resampled in different ways including limiting the dataset size (max_datasize), shuffling the dataset (shuffle_dataset), or equalizing the frequency of labels through subsampling and oversampling (balance_dataset). The function can also take several parameters for the model (image_dim) or training hyperparameters (batch_size, num_epoch). Outputs, including terminal log, tensorboard logs, and weights of the best model are saved into a folder named after the category of interest, in a main output folder (experiment_root_dirpath).

Command used on HPC for this project (repeat for all categories: good-frame, algae, Lessonia, Ecklonia, Carpophyllum):

python trainer_categorical.py \

--data_root_dirpath /nesi/project/niwa02671/data \

--experiment_root_dirpath /nesi/project/niwa02671/models \

--category good-frame \

--labels 0 1 \

--train_file_substrings SWC1901 SWC2001_011 SWC2001_018 \

--test_file_substrings SWC2001_002 SWC2001_008 SWC2001_010 \

--shuffle_dataset True \

--balance_dataset False \

--image_dim 299 \

--batch_size 32 \

--num_epoch 150

trainer_categorical only works for one category. Ideally, modify the code to allow looping on several training. But you will need to instantiate several tensorflow graphs.

Each instance of training is given a unique code based on the date and time of the start of training: exp_YYYMMDD-HHMMSS. This code is showing on the screen when starting the training, and all outputs are saved in folders bearing this code.

The training creates three types of outputs:

- A terminal log
  (<experiment_root_dirpath>/<category>/terminal_logs/<training_code>.log e.g.
  R:\National\Datasets\Seaweed_Video_Data\ai\models\algae\terminal_logs\exp_202009
  07-123524.log) — a text file containing (some of) the screen output of the command.

- The best model weights
  (<experiment_root_dirpath>/<category>/model_weights/<training_code>/bestmodelwe
  ights.h5 e.g.
  R:\National\Datasets\Seaweed_Video_Data\ai\models\algae\model_weights\exp_20200
  907-123524\bestmodelweights.h5) — an HDF file containing the weights of the best
  model trained, to allow rebuilding the model for later inference.

- The tensorboard logs (in
  <experiment_root_dirpath>/<category>/tensorboard_logs/<training_code> e.g.
  R:\National\Datasets\Seaweed_Video_Data\ai\models\algae\tensorboard_logs\exp_202
  00907-123524) — one or several files for TensorBoard, a tool for providing the
  measurements and visualizations needed during the machine learning workflow.

Tensorboard is a web app, started from your machine. In the terminal, type in the command line:

tensorboard --logdir path/to/tensorboard_logs

Then click on the link provided. This will open TensorBoard in your browser. Note this does not work from the HPC so you will have to copy over the logs from the HPC to your local machine.

All training outputs are on the HPC on /nesi/project/niwa02671/models.

They were copied to R:\National\Datasets\Seaweed_Video_Data\ai\models.

6. Inferencing

A trained model can be used for inferencing. One only needs the model weights file (.h5) that was produced by the training.

The main function for this is seaweedai.inference.classify_videos.classify_videos, which takes a root directory of videos (videos_root_dirpath) and a list of substrings (videos_file_substrings, to choose the videos in the folder), and apply a model with specified weights (model_weights_filepath) to predict the class of every Nth frame (every_nth_frame). The result is a csv file for each video, bearing the same name as the video, all saved in a folder (output_csv_dirpath). The csv file contain columns prefix (video name), frame (frame number) and the probabilities for each class, between 0 and 1.

Like the training, this function only works for one category at a time. Ideally, modify the code to allow looping on several models but you will need to instantiate several tensorflow graphs to separate the models.

HPC command used for this project (repeated for each model, with appropriate path to model weights and output dirpath):

python classify_videos.py \

--videos_root_dirpath /nesi/nobackup/niwa02671/raw_data \

--model_weights_filepath   /nesi/project/niwa02671/models/good-
frame/model_weights/exp_20200907-112706/bestmodelweights.h5 \

--output_csv_dirpath /nesi/project/niwa02671/outputs/good-frame

The results were initially saved on the HPC at /nesi/project/niwa02671/outputs.

They were copied to R:\National\Datasets\Seaweed_Video_Data\ai\outputs.

6.1 Processing results

seaweedai.inference.process_results.process_results takes the results from the inference stage and metadata and process them into a single file suitable for import into a GIS.

Namely, the function:

- reads all prediction csv files for all categories requested;
- reads all metadata csv files provided and interpolate latitude and longitude for each file;
- merge predictions and (interpolated) metadata;
- for each video, operate a process combining a number of consecutive frames and taking into account uncertainty;

write all results into a single csv file.

The combination process consists of a weighted average where the model prediction value in the 0–1 range for a category (algae, or specie) is weighted by the usability of the frame (good-frame) in the 0–1 range. The weighted average is then turned into a hard category (aka "absent" if <0.5, or "present" if >0.5).

An estimate of uncertainty is also produced, using the minimum between 1) the absolute value of that weighted average, and 2) the average of the good-frame value. If that value is low (towards 0), it means the prediction for this group of frame is uncertain (either mostly unusable frames, or frames with mix of classes, or frames with low model confidence). If the value is high (towards 1), it means the prediction for this group of frame is rather certain (mostly usable frames AND consistent classification in the group AND high models confidence).

Python script used for this project:

```
from seaweedai.process.process_results import process_results

preds_root_dirpath = r'R:\National\Datasets\Seaweed_Video_Data\ai\outputs'

categories = ['algae', 'Ecklonia', 'Lessonia', 'Carpophyllum']

metadata_dirpath = r'R:\National\Datasets\Seaweed_Video_Data\ai\outputs\metadata'

output_csv_filepath = r'R:\National\Datasets\Seaweed_Video_Data\ai\outputs\out.csv'

interpolate_navigation = True

average_n_frames = 120

process_results(preds_root_dirpath, categories, metadata_dirpath, output_csv_filepath, interpolate_navigation, average_n_frames)
```