

Article

A Multi-Stage Approach Combining Very High-Resolution Satellite Image, GIS Database and Post-Classification Modification Rules for Habitat Mapping in Hong Kong

Ivan H. Y. Kwong ^{1,2,*}, Frankie K. K. Wong ¹, Tung Fung ^{1,2}, Eric K. Y. Liu ³, Roger H. Lee ³ and Terence P. T. Ng ³

¹ Department of Geography and Resource Management, The Chinese University of Hong Kong, Hong Kong; kkit@cuhk.edu.hk (F.K.K.W.); tungfung@cuhk.edu.hk (T.F.)

² Institute of Future Cities, The Chinese University of Hong Kong, Hong Kong

³ Agriculture, Fisheries and Conservation Department, 5/F, Cheung Sha Wan Government Offices, 303 Cheung Sha Wan Road, Kowloon, Hong Kong; eric_ky_liu@afcd.gov.hk (E.K.Y.L.); roger_h_lee@afcd.gov.hk (R.H.L.); pun_tung_ng@afcd.gov.hk (T.P.T.N.)

* Correspondence: khoyinivan@link.cuhk.edu.hk

Supplementary Materials

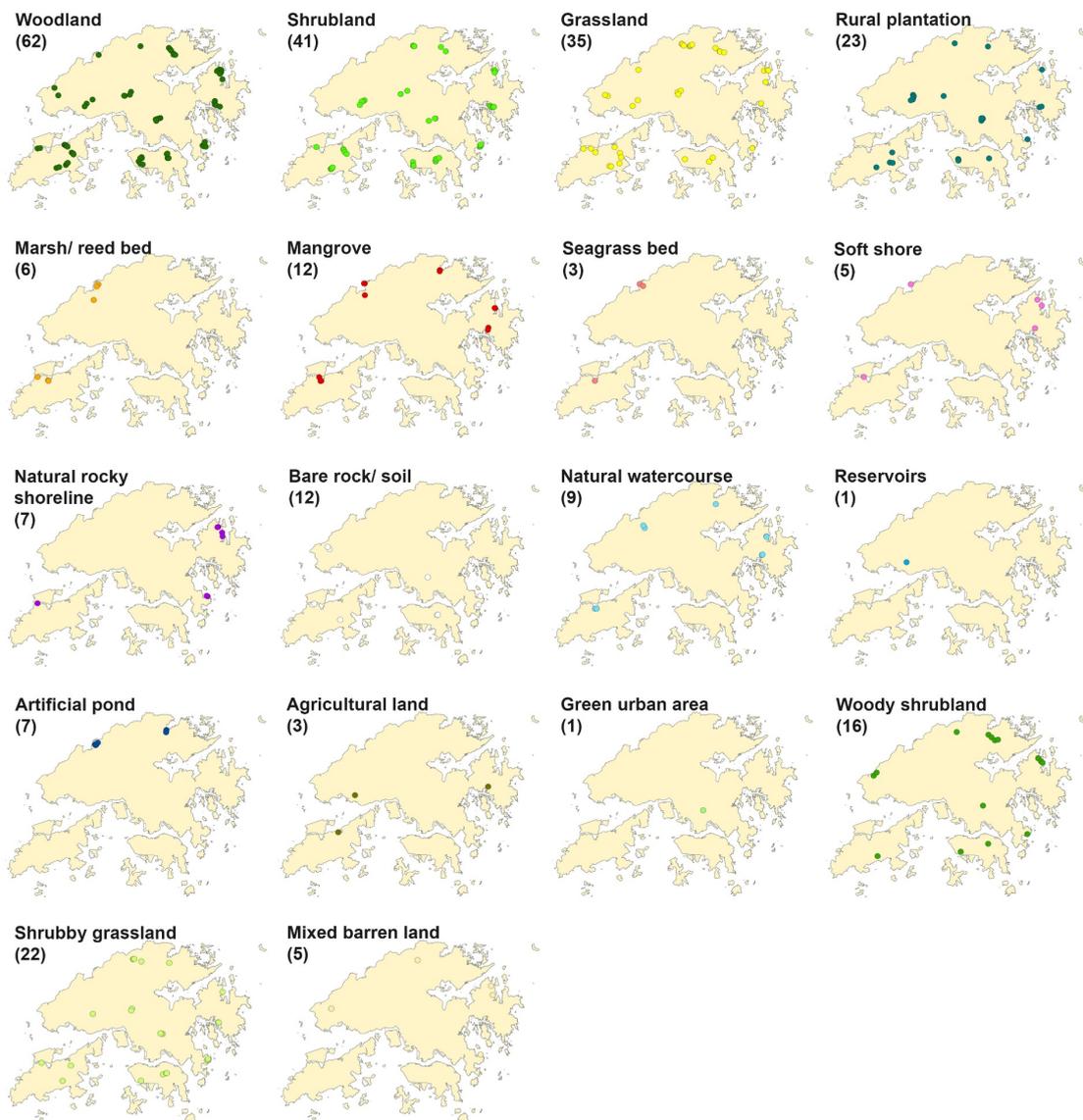


Figure S1. Distribution of survey points used for training data. The survey points were then transformed to part of the training pixels/objects as classification input (Figure S3). The number in parentheses refers to the number of survey points for the habitat. The difference in numbers of points among habitats was a consequence of the survey objective and their coverages over the entire territory, which affected the likelihood of identifying some habitats along the survey trails.

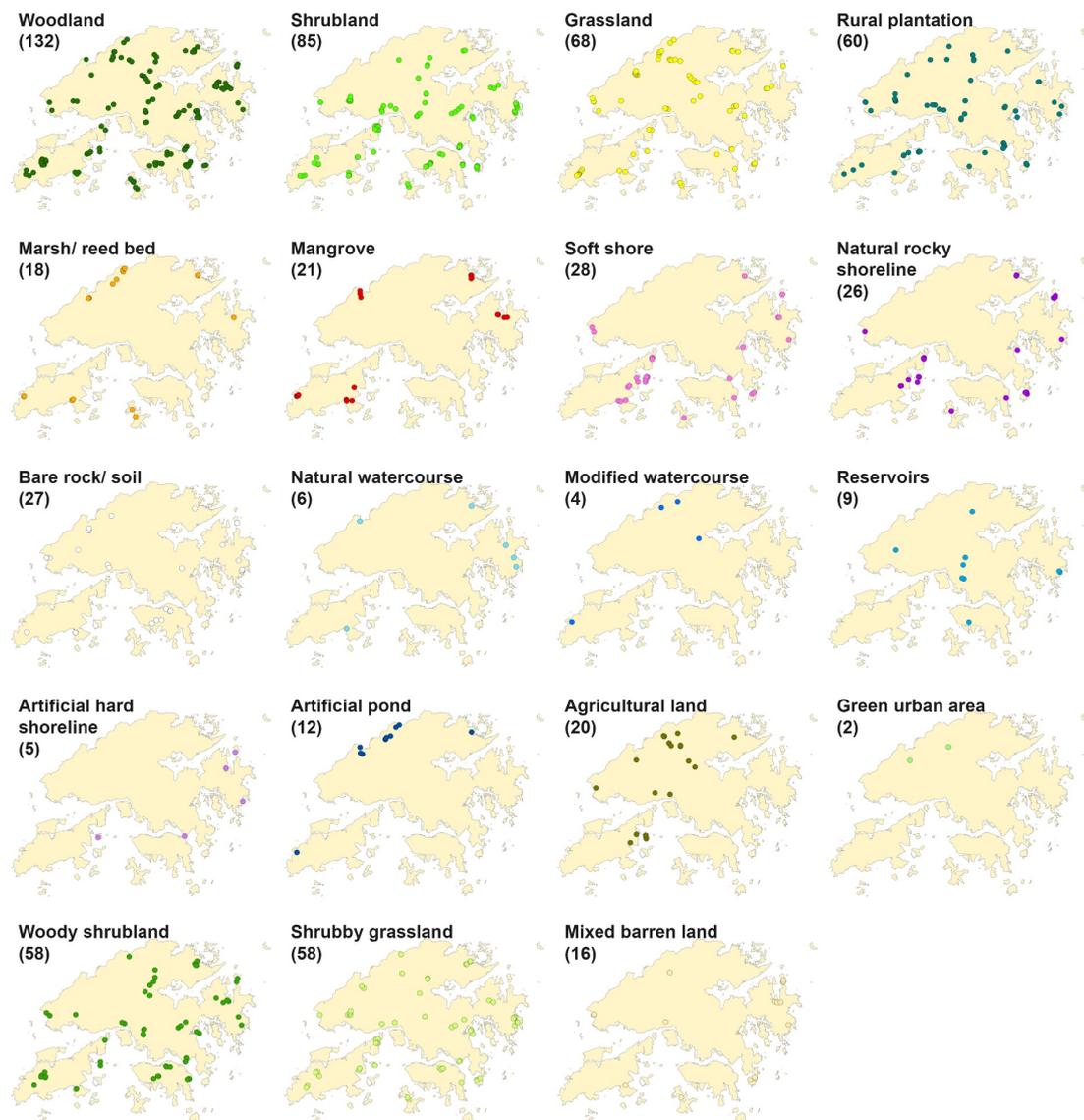


Figure S2. Distribution of survey points used for validation data. The number in parentheses refers to the number of survey points for the habitat. The difference in numbers of points among habitats was a consequence of the survey objective and their coverages over the entire territory, which affected the likelihood of identifying some habitats along the survey trails.

Table S1. Summary of terrain heights of each habitat according to field survey points.

Habitat	Number of points	Terrain height (m)		
		Minimum	Mean	Maximum
Woodland	198	2.84	159.63	854.78
Shrubland	128	15.78	222.47	725.54
Grassland	107	1.14	158.17	926.76
Rural plantation	84	3.22	166.99	475.53
Marsh/ reed bed	24	1.13	2.81	4.09
Mangrove	33	0.00	1.38	4.08
Seagrass bed	3	0.00	0.93	2.37
Soft shore	33	0.00	1.07	3.83
Natural rocky shoreline	33	0.00	1.96	4.80
Bare rock/ soil	39	6.10	166.26	481.16
Natural watercourse	15	0.00	8.41	124.19
Modified watercourse	4	0.00	4.75	17.73
Reservoirs	10	1.65	95.61	185.90
Artificial hard shoreline	5	0.00	1.79	4.24
Artificial pond	19	0.30	2.46	5.08
Agricultural land	23	2.76	39.46	243.20
Green urban area	3	5.97	45.12	93.75
Woody shrubland	74	7.93	184.96	501.95
Shrubby grassland	82	16.12	212.30	918.19
Mixed barren land	21	2.68	58.96	419.66

Table S2. Examples of species with the highest occurrences in vegetation-related habitats during field survey.

Habitat	Species	Number of survey points that the species were found
Woodland	<i>Schefflera heptaphylla</i>	97
	<i>Psychotria asiatica</i>	54
	<i>Sterculia lanceolata</i>	38
	<i>Mallotus paniculatus</i>	35
	<i>Machilus chekiangensis</i>	30
Shrubland	<i>Rhodomyrtus tomentosa</i>	83
	<i>Dicranopteris pedata</i>	56
	<i>Polyspora axillaris</i>	39
	<i>Melastoma sanguineum</i>	38
	<i>Litsea rotundifolia</i>	35
Grassland	<i>Dicranopteris pedata</i>	36
	<i>Rhodomyrtus tomentosa</i>	22
	<i>Miscanthus sinensis</i>	21
	<i>Bidens alba</i>	16
	<i>Baeckea frutescens</i>	13
Rural plantation	<i>Acacia confusa</i>	44
	<i>Lophostemon confertus</i>	24
	<i>Acacia auriculiformis</i>	14
Marsh/ reed bed	<i>Phragmites australis</i>	7
	<i>Acrostichum aureum</i>	6
	<i>Eichhornia crassipes</i>	4
Mangrove	<i>Kandelia obovata</i>	26
	<i>Aegiceras corniculatum</i>	12
	<i>Acanthus ilicifolius</i>	10
Agricultural land	<i>Musa acuminata</i>	10
	<i>Dimocarpus longan</i>	9
	<i>Carica papaya</i>	7
Green urban area	<i>Acacia confusa</i>	2
	<i>Bombax ceiba</i>	1
	<i>Aleurites moluccana</i>	1
Woody shrubland	<i>Rhodomyrtus tomentosa</i>	36
	<i>Dicranopteris pedata</i>	25
	<i>Schefflera heptaphylla</i>	21
Shrubby grassland	<i>Dicranopteris pedata</i>	65
	<i>Rhodomyrtus tomentosa</i>	50
	<i>Baeckea frutescens</i>	27
Mixed barren land	<i>Dicranopteris pedata</i>	3
	<i>Bidens alba</i>	2
	<i>Rhodomyrtus tomentosa</i>	2

Table S3. Numbers of training pixels/objects used in pixel-based and object-based classification. The training pixels and objects were generated from the same set of training sites. The difference in numbers was caused by the transformation of each site (usually 10×10 m) to dozens of pixels but only a few objects.

Habitat class	Pixels			Objects		
	Field survey	Visual selection	Total	Field survey	Visual selection	Total
Woodland	2,272	200	2,472	122	10	132
Shrubland	1,765	600	2,365	98	30	128
Grassland	2,126	300	2,426	106	20	126
Marsh/ reed bed	401	500	901	14	35	49
Mangrove	1,089	200	1,289	52	12	64
Soft shore	699	400	1,099	29	28	57
Natural rocky shoreline	292	800	1,092	16	40	56
Bare rock/ soil	473	600	1,073	24	32	56
Other urban area	28	1,000	1,028	2	55	57
Water	990	300	1,290	22	45	67
Shadow	0	1,000	1,000	0	44	44
<i>Total</i>	<i>10,135</i>	<i>5,900</i>	<i>16,035</i>	<i>485</i>	<i>351</i>	<i>836</i>

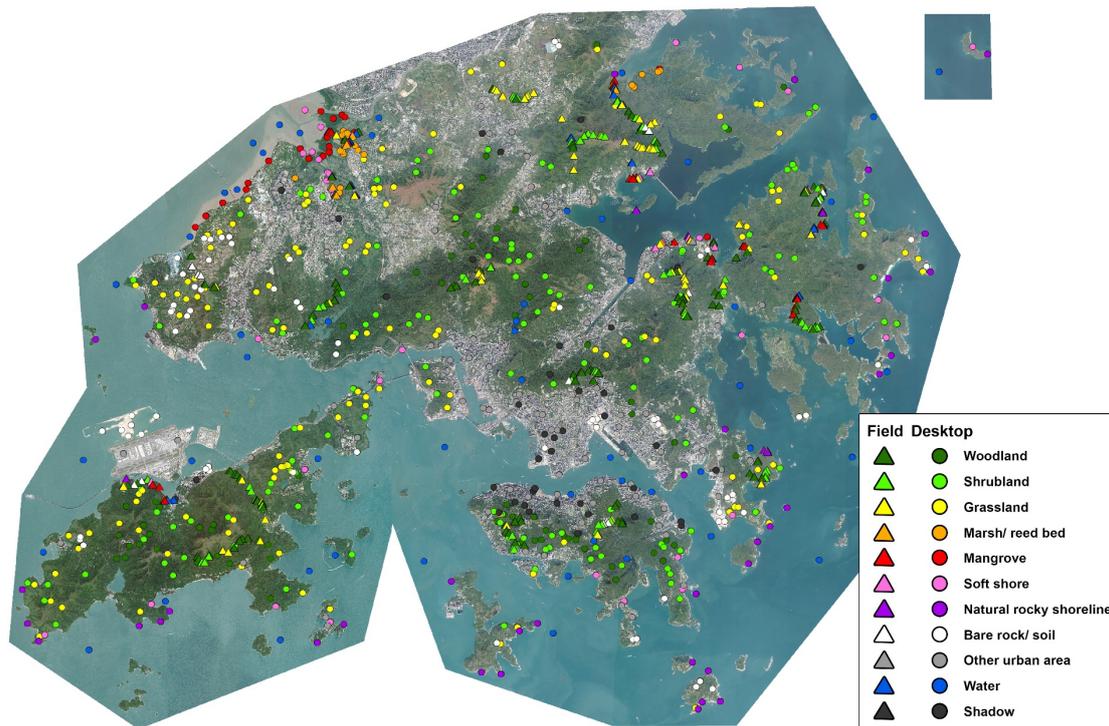


Figure S3. Distribution of training data over the study area.

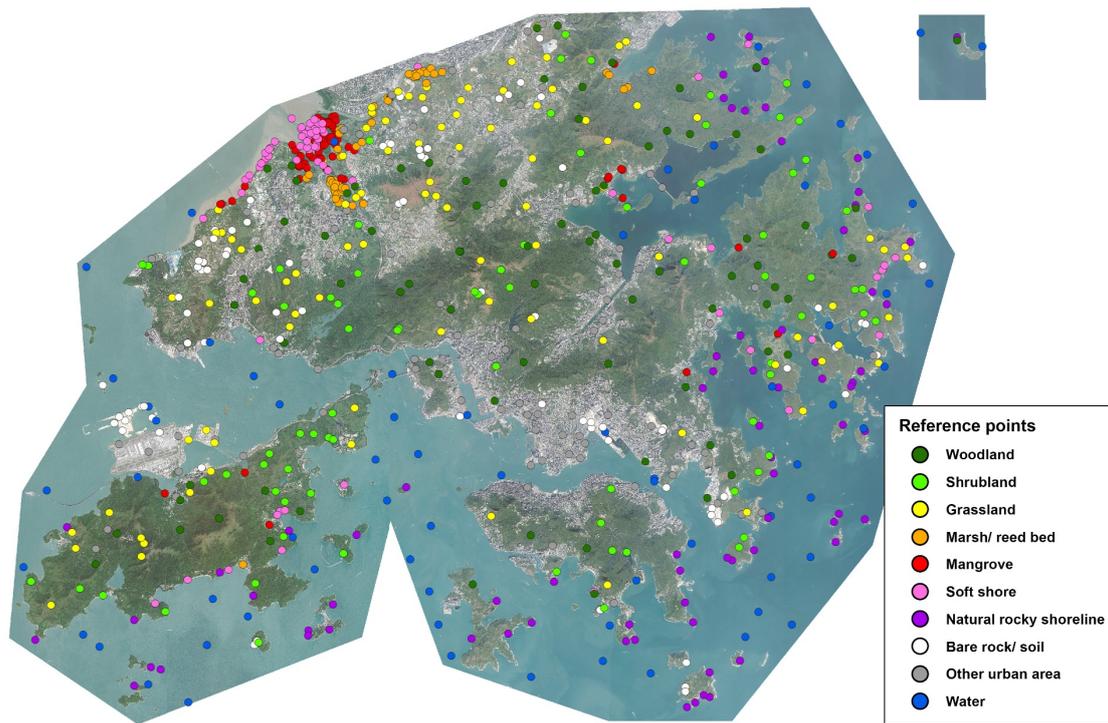


Figure S4. Distribution of randomly sampled points used for accuracy assessment in stage 2.

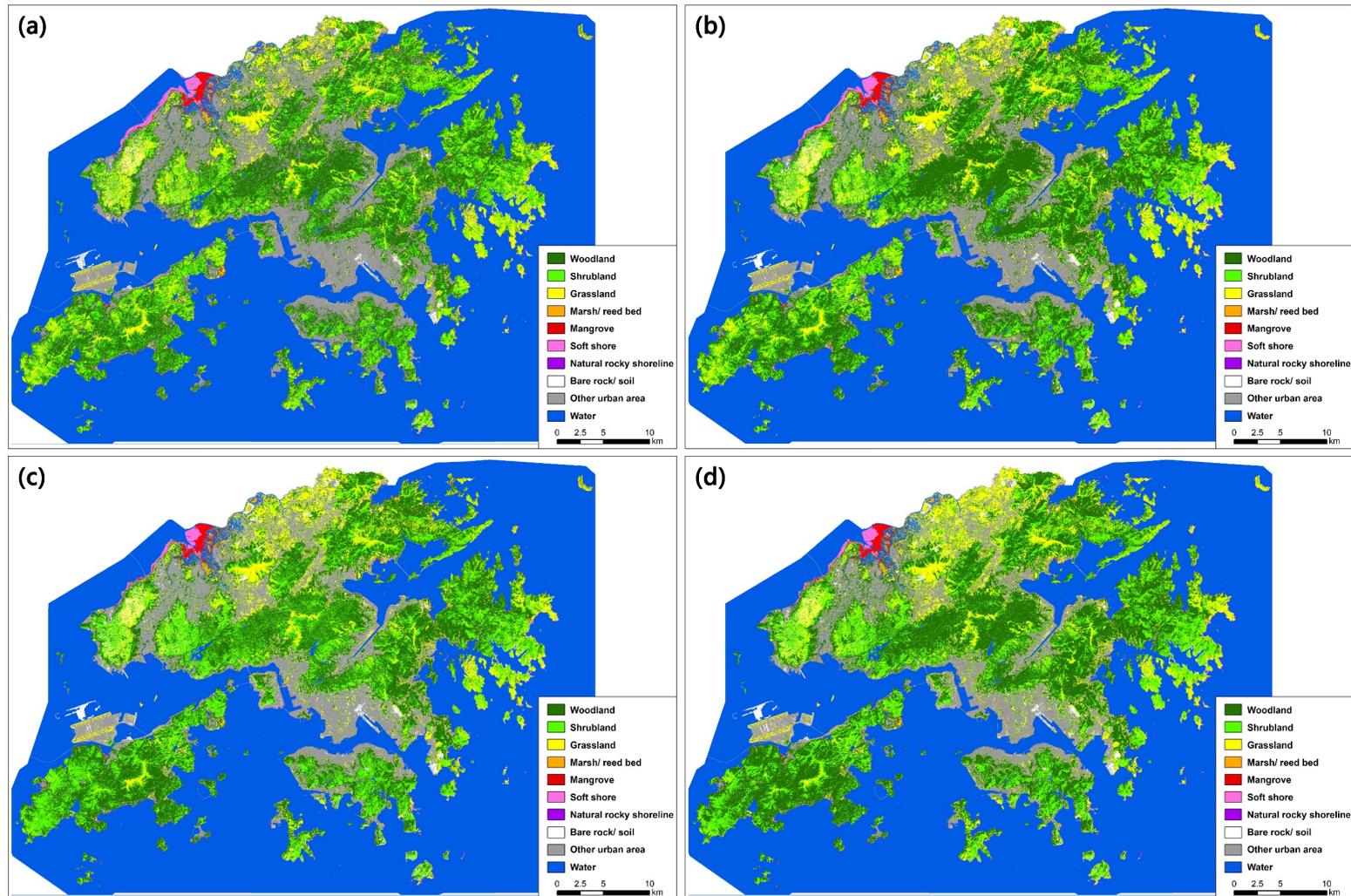


Figure S5. Refined classification maps with 10 classes in stage 2 obtained using the four classification methods, including (a) pixel-based support vector machine (SVM), (b) pixel-based random forest (RF), (c) object-based SVM, and (d) object-based RF.

Table S4. Confusion matrix of pixel-based SVM classification result (Overall accuracy = 76.0%; Kappa statistics = 0.73; 95% confidence interval [CI] = $\pm 3.9\%$).

Pixel-SVM		Reference data										Total	User's accuracy	95% CI
		Woodland	Shrubland	Grassland	Marsh/ reed bed	Mangrove	Soft shore	Natural rocky shoreline	Bare rock/ soil	Other urban area	Water			
Classified result	Woodland	65	18	3	0	1	0	0	0	1	0	88	73.9%	$\pm 9.2\%$
	Shrubland	9	47	14	0	1	0	0	0	0	0	71	66.2%	$\pm 11.1\%$
	Grassland	1	13	60	8	1	0	0	3	1	0	87	69.0%	$\pm 9.8\%$
	Marsh/ reed bed	0	0	6	47	3	0	0	0	0	0	56	83.9%	$\pm 9.7\%$
	Mangrove	0	0	2	1	60	0	0	0	0	0	63	95.2%	$\pm 5.3\%$
	Soft shore	0	0	0	0	0	64	1	0	0	0	65	98.5%	$\pm 3.0\%$
	Natural rocky shoreline	1	0	0	0	0	2	56	1	7	0	67	83.6%	$\pm 8.9\%$
	Bare rock/ soil	0	0	1	0	0	2	6	50	10	0	69	72.5%	$\pm 10.6\%$
	Other urban area	8	0	7	4	1	6	10	25	68	1	130	52.3%	$\pm 8.6\%$
	Water	0	0	0	1	0	2	3	0	0	68	74	91.9%	$\pm 6.3\%$
	Total	84	78	93	61	67	76	76	79	87	69	770		
Producer's accuracy	77.4%	60.3%	64.5%	77.0%	89.6%	84.2%	73.7%	63.3%	78.2%	98.6%				
95% CI	$\pm 5.7\%$	$\pm 9.1\%$	$\pm 10.0\%$	$\pm 3.6\%$	$\pm 23.6\%$	$\pm 13.4\%$	$\pm 6.7\%$	$\pm 6.9\%$	$\pm 6.0\%$	$\pm 0.4\%$				

Table S5. Confusion matrix of pixel-based RF classification result (Overall accuracy = 84.0%; Kappa statistics = 0.82; 95% confidence interval [CI] = $\pm 3.1\%$).

Pixel-RF	Reference data										Total	User's accuracy	95% CI
	Woodland	Shrubland	Grassland	Marsh/ reed bed	Mangrove	Soft shore	Natural rocky shoreline	Bare rock/ soil	Other urban area	Water			
Woodland	73	7	0	1	1	0	0	1	1	0	84	86.9%	$\pm 7.3\%$
Shrubland	9	65	11	2	2	0	0	1	1	0	91	71.7%	$\pm 9.3\%$
Grassland	2	6	72	5	2	0	1	1	3	0	92	78.3%	$\pm 8.5\%$
Marsh/ reed bed	0	0	2	49	4	1	0	0	0	1	57	86.0%	$\pm 9.1\%$
Mangrove	0	0	3	2	58	0	0	0	0	0	63	92.1%	$\pm 6.7\%$
Soft shore	0	0	0	0	0	67	1	0	0	0	68	98.5%	$\pm 2.9\%$
Natural rocky shoreline	0	0	0	2	0	3	65	0	7	0	77	84.4%	$\pm 8.2\%$
Bare rock/ soil	0	0	5	0	0	2	7	69	14	0	97	71.1%	$\pm 9.1\%$
Other urban area	0	0	0	0	0	2	0	7	61	0	70	87.1%	$\pm 7.9\%$
Water	0	0	0	0	0	1	2	0	0	68	71	95.8%	$\pm 4.7\%$
Total	84	78	93	61	67	76	76	79	87	69	770		
Producer's accuracy	86.9%	83.3%	77.4%	80.3%	86.6%	88.2%	85.5%	87.3%	70.1%	98.6%			
95% CI	$\pm 3.6\%$	$\pm 10.2\%$	$\pm 8.2\%$	$\pm 6.3\%$	$\pm 20.3\%$	$\pm 26.6\%$	$\pm 20.0\%$	$\pm 16.0\%$	$\pm 6.3\%$	$\pm 0.0\%$			

Table S6. Confusion matrix of object-based SVM classification result (Overall accuracy = 77.1%; Kappa statistics = 0.75; 95% confidence interval [CI] = $\pm 4.2\%$).

Object-SVM	Reference data										Total	User's accuracy	95% CI
	Woodland	Shrubland	Grassland	Marsh/ reed bed	Mangrove	Soft shore	Natural rocky shoreline	Bare rock/ soil	Other urban area	Water			
Woodland	59	20	2	0	3	0	0	0	0	0	84	70.2%	$\pm 9.8\%$
Shrubland	13	53	23	3	2	0	0	1	0	0	95	55.8%	$\pm 10.0\%$
Grassland	3	5	56	13	0	1	1	8	3	0	90	62.2%	$\pm 10.1\%$
Marsh/ reed bed	1	0	6	44	1	0	0	0	0	1	53	83.0%	$\pm 10.2\%$
Mangrove	0	0	3	1	61	0	0	0	0	0	65	93.8%	$\pm 5.9\%$
Soft shore	0	0	0	0	0	65	2	1	2	0	70	92.9%	$\pm 6.1\%$
Natural rocky shoreline	2	0	0	0	0	2	59	2	2	0	67	88.1%	$\pm 7.8\%$
Bare rock/ soil	1	0	1	0	0	2	7	61	12	0	84	72.6%	$\pm 9.6\%$
Other urban area	5	0	2	0	0	0	5	6	68	0	86	79.1%	$\pm 8.7\%$
Water	0	0	0	0	0	6	2	0	0	68	76	89.5%	$\pm 7.0\%$
Total	84	78	93	61	67	76	76	79	87	69	770		
Producer's accuracy	70.2%	67.9%	60.2%	72.1%	91.0%	85.5%	77.6%	77.2%	78.2%	98.6%			
95% CI	$\pm 5.9\%$	$\pm 9.7\%$	$\pm 9.4\%$	$\pm 3.2\%$	$\pm 16.9\%$	$\pm 5.1\%$	$\pm 10.8\%$	$\pm 12.9\%$	$\pm 3.4\%$	$\pm 0.0\%$			

Table S7. Confusion matrix of object-based RF classification result (Overall accuracy = 76.6%; Kappa statistics = 0.74; 95% confidence interval [CI] = $\pm 4.1\%$).

Object-RF	Reference data										Total	User's accuracy	95% CI
	Woodland	Shrubland	Grassland	Marsh/ reed bed	Mangrove	Soft shore	Natural rocky shoreline	Bare rock/ soil	Other urban area	Water			
Woodland	60	19	2	2	3	0	0	1	0	0	87	69.0%	$\pm 9.8\%$
Shrubland	13	53	27	2	3	0	0	1	3	0	102	52.0%	$\pm 9.7\%$
Grassland	9	6	59	9	1	3	2	9	2	0	100	59.0%	$\pm 9.7\%$
Marsh/ reed bed	0	0	1	44	2	0	0	0	0	0	47	93.6%	$\pm 7.1\%$
Mangrove	0	0	2	3	58	0	0	0	0	0	63	92.1%	$\pm 6.7\%$
Soft shore	0	0	0	0	0	64	2	1	1	0	68	94.1%	$\pm 5.6\%$
Natural rocky shoreline	1	0	0	1	0	3	65	2	3	1	76	85.5%	$\pm 8.0\%$
Bare rock/ soil	0	0	0	0	0	1	3	58	17	0	79	73.4%	$\pm 9.8\%$
Other urban area	1	0	2	0	0	2	0	7	61	0	73	83.6%	$\pm 8.6\%$
Water	0	0	0	0	0	3	4	0	0	68	75	90.7%	$\pm 6.6\%$
Total	84	78	93	61	67	76	76	79	87	69	770		
Producer's accuracy	71.4%	67.9%	63.4%	72.1%	86.6%	84.2%	85.5%	73.4%	70.1%	98.6%			
95% CI	$\pm 5.1\%$	$\pm 10.2\%$	$\pm 9.0\%$	$\pm 3.1\%$	$\pm 14.2\%$	$\pm 8.3\%$	$\pm 8.9\%$	$\pm 11.6\%$	$\pm 5.4\%$	$\pm 0.0\%$			

Table S8. Confusion matrix of habitat map against field survey points (WO: Woodland; SH: Shrubland; GR: Grassland; RP: Rural plantation; MR: Marsh/reed bed; MA: Mangrove; SS: Soft shore; RS: Natural rocky shoreline; BR: Bare rock/soil; NW: Natural watercourse; MW: Modified watercourse; RE: Reservoirs; AS: Artificial hard shoreline; AP: Artificial ponds; AL: Agricultural land; GU: Green urban area; OU: Other urban area; WS: Woody shrubland; SG: Shrubby grassland; MB: Mixed barren land; PA: Producer's accuracy; UA: User's accuracy; CI: 95% confidence interval of PA/UA).

	Reference data																				Total	UA (%)	CI (±%)
	WO	SH	GR	RP	MR	MA	SS	RS	BR	NW	MW	RE	AS	AP	AL	GU	OU	WS	SG	MB			
WO	111	1	1	10	0	2	0	0	0	0	0	0	0	0	1	1	0	1	0	0	128	86.7	5.8
SH	0	62	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	4	0	72	86.1	8.0
GR	0	3	58	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	4	1	71	81.7	9.1
RP	11	0	1	45	0	0	0	0	0	0	0	0	0	0	1	0	0	3	0	0	61	73.8	11.1
MR	0	0	0	0	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	100.0	0.0
MA	0	0	0	0	0	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	19	100.0	0.0
SS	0	0	0	0	0	0	24	1	0	1	0	0	0	0	0	0	0	0	0	0	26	92.3	10.5
RS	0	0	0	0	0	0	2	25	0	0	0	0	0	0	0	0	0	0	0	0	27	92.6	10.1
BR	0	0	0	0	0	0	2	0	26	0	0	0	0	0	0	0	0	0	1	1	30	86.7	12.4
NW	0	0	0	0	0	0	0	0	0	5	0	1	0	1	0	0	0	0	0	0	7	71.4	36.2
MW	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	3	100.0	0.0
RE	0	0	0	0	0	0	0	0	0	0	0	8	0	0	0	0	0	0	0	0	8	100.0	0.0
AS	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	5	100.0	0.0
AP	0	0	0	0	0	0	0	0	0	0	0	0	0	11	0	0	0	0	0	0	11	100.0	0.0
AL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	17	0	0	0	0	0	17	100.0	0.0
GU	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	100.0	NA
OU	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0.0	NA
WS	10	9	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	47	0	0	71	66.2	11.1
SG	0	10	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	49	0	67	73.1	10.7
MB	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	14	17	82.4	18.7
Total	132	85	68	60	18	21	28	26	27	6	4	9	5	12	20	2	0	58	58	16	655		
PA (%)	84.1	72.9	85.3	75.0	72.2	90.5	85.7	96.2	96.3	83.3	75.0	88.9	100.0	91.7	85.0	50.0	NA	81.0	84.5	87.5			
CI (±%)	4.4	8.7	8.5	12.5	19.6	32.7	23.4	5.6	3.6	25.9	0.0	2.8	0.0	6.4	9.8	NA	NA	7.0	8.3	24.4			

Table S9. Land coverage of each habitat type mapped in this study. The area is separated into above and below high water mark (defined as 2.3 m above the Hong Kong Principal Datum) according to the coastline layer. The occurrence of the habitats below the high water mark was affected by the satellite viewing condition and tidal level at the time of observation. 95% confidence intervals (CI) were computed to quantify the sampling variability of the estimated areas resulting from the field surveys. Note that the numbers presented here represent the mapping results from the proposed methodology and can be slightly different from the finalized numbers adopted by the Hong Kong government.

Habitat	Area (km ²)			95% CI	Percentage
	Above high water mark	Below high water mark	Total		
Woodland	270.13	0.29	270.42	±20.24	24.02%
Shrubland	96.15	0.03	96.18	±16.57	8.54%
Grassland	82.03	0.28	82.31	±10.97	7.31%
Rural plantation	57.02	0.04	57.06	±16.56	5.07%
Marsh/ reed bed	3.19	0.24	3.43	±4.96	0.30%
Mangrove	1.67	4.54	6.21	±5.83	0.55%
Seagrass bed	0.26	0.12	0.38	NA	0.03%
Soft shore	1.59	6.43	8.02	±3.48	0.71%
Natural rocky shoreline	6.64	3.07	9.71	±1.17	0.86%
Bare rock/ soil	35.12	0.45	35.57	±4.55	3.16%
Natural watercourse	1.03	0.63	1.66	±0.99	0.15%
Modified watercourse	1.89	0.04	1.93	±0	0.17%
Reservoirs	22.37	0.00	22.37	±0.63	1.99%
Artificial hard shoreline	1.35	0.29	1.64	±0	0.15%
Artificial pond	9.41	0.00	9.41	±0.63	0.84%
Agricultural land	40.22	0.00	40.23	±4.68	3.57%
Green urban area	41.26	0.01	41.27	NA	3.67%
Other urban area	195.96	1.03	197.00	NA	17.50%
Woody shrubland	143.36	0.06	143.41	±18.02	12.74%
Shrubby grassland	87.79	0.01	87.80	±11.85	7.80%
Mixed barren land	9.81	0.05	9.86	±3.75	0.88%
<i>Total</i>	<i>1108.27</i>	<i>17.59</i>	<i>1125.86</i>		<i>100.00%</i>

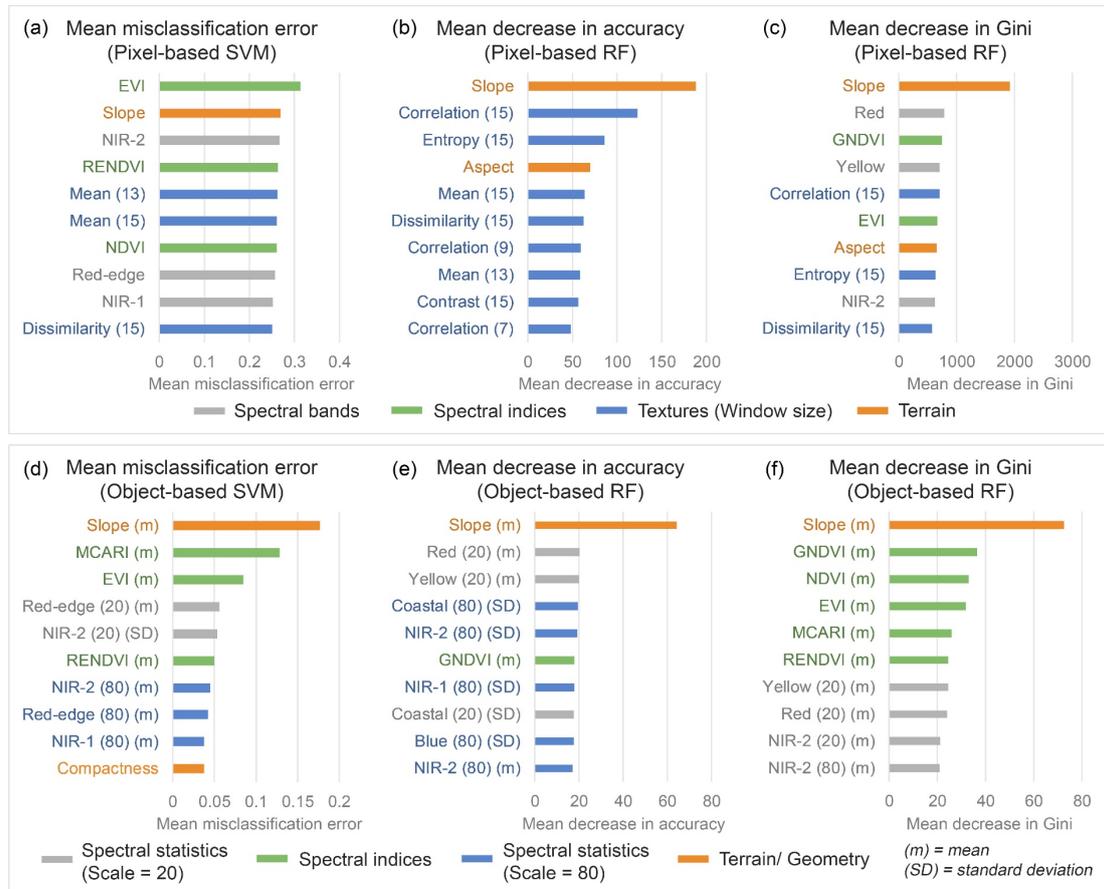


Figure S6. Top 10 classification variables with the highest importance using different evaluation metrics; (a) Mean classification error in pixel-based SVM model; (b) Mean decrease in accuracy in pixel-based RF model; (c) Mean decrease in Gini in pixel-based RF model; (d) Mean classification error in object-based SVM model; (e) Mean decrease in accuracy in object-based RF model; (f) Mean decrease in Gini in object-based RF model.

For SVM models, the values were estimated by computing the change in performance from permuting the values of a feature and comparing that to the predictions made on the unmuted data. The changes in performance were measured using mean classification errors. Larger value for a variable represents larger introduced error when the variable is permuted and hence higher importance of the variable in the model. The values were computed using *generateFeatureImportanceData* function in *mlr* package in R.

For RF models, mean decrease in accuracy was computed by recording the prediction accuracy on the out-of-bag portion of the data and comparing the difference when each variable was permuted. The value expresses the amount of accuracy the model loses by excluding the variable and larger value indicates higher importance of the variable. Mean decrease in Gini refers to total decrease in node impurities, measured by the Gini index, from splitting on the variable. It measures the contribution of each variable to the homogeneity of the nodes in the random forest and larger value indicates higher importance of the variable in the model. Both metrics were computed using *importance* function in *randomForest* package in R.

The figure illustrated that the combinations of spectral, textural, topographic and geometric variables adopted in this study could contribute to the identification of different habitats. In pixel-based approach, while EVI and slope were the most important variables in SVM and RF models respectively, the importance was followed by other spectral bands, indices and a variety of texture variables obtained with varied window sizes. In object-based approach, spectral indices like MCARI and GNDVI, as well as mean spectral statistics calculated from various bands using both segmentation scales were important in the models. The contrasts in variable importance scores also suggested the intrinsic difference of the implemented algorithms and could facilitate selection of suitable classification variables from different aspects.