

## Article

# Snow Coverage Mapping by Learning from Sentinel-2 Satellite Multispectral Images via Machine Learning

Yucheng Wang <sup>1</sup>, Jinya Su <sup>1,\*</sup>, Xiaojun Zhai <sup>1</sup>, Fanlin Meng <sup>2</sup> and Cunjia Liu <sup>3</sup>

<sup>1</sup> School of Computer Science and Electronic Engineering, University of Essex, Colchester, CO4 3SQ, U.K.; yw19282@essex.ac.uk(Y.W); j.su@essex.ac.uk(J.S); xzhai@essex.ac.uk(X.Z)

<sup>2</sup> Department of Mathematical Sciences, University of Essex, Colchester, CO4 3SQ, U.K.; fanlin.meng@essex.ac.uk(F.M)

<sup>3</sup> Department of Aeronautical and Automotive Engineering, Loughborough University, Loughborough LE11 3TU, U.K.; c.liu5@lboro.ac.uk

\* Correspondence: j.su@essex.ac.uk or sujinya1990@163.com (J.Su)

**Citation:** Wang, Y.; Su, J.; Zhai, X.; Meng, F.; Liu, C. Wang, et al.. *Remote Sens.* **2021**, *1*, 0. <https://doi.org/>

Received:

Accepted:

Published:

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Copyright:** © 2022 by the authors. Submitted to *Remote Sens.* for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** Snow coverage mapping plays a vital role not only in studying hydrology and climatology, but also in investigating crop disease overwintering for smart agriculture management. This work investigates snow coverage mapping by learning from Sentinel-2 satellite multispectral images via machine learning methods. To this end, the largest dataset for snow coverage mapping (to our best knowledge) with three typical classes (snow, cloud and background) is first collected and labelled via the semi-automatic classification plugin in QGIS. Then both random forest based conventional machine learning and U-Net based deep learning are applied to the semantic segmentation challenge in this work. The effects of various input band combinations are also investigated so that the most suitable one can be identified. Experimental results show that (1) both conventional machine learning and advanced deep learning methods significantly outperform the existing rule-based Sen2Cor product for snow mapping; (2) U-Net generally outperforms the random forest since both spectral and spatial information is incorporated in U-Net via convolution operations; (3) the best spectral band combination for U-Net is B2, B11, B4 and B9. It is concluded that U-Net based deep learning classifier with four informative spectral bands is suitable for snow coverage mapping.

**Keywords:** Snow coverage; Sentinel-2 satellite; Remote sensing; Multispectral image; Random Forest; U-Net; Semantic segmentation

## 1. Introduction

Remote sensing technology (e.g. satellite-based, unmanned aerial vehicle (UAV)-based, ground-based) is able to provide a nondestructive way of spatially and temporally monitoring the targets of interest remotely, which is drawing ever-increasing attention recently with the rapid development of autonomous system innovation, sensing technology and image processing algorithms, especially the convolutional neural network based deep learning approaches. Taking remote sensing in geosciences as an example, UAV multispectral images are used in [1] for bulk drag prediction of Riparian Arundo donax stands so that their impact on vegetated drainage channel can be assessed. Terrestrial laser scanning is also used in [2] to investigate physically-based characterization of mixed floodplain vegetation. Meanwhile, the interest in agricultural applications of remote sensing technology has also been exponentially growing since 2013 [3], where the main applications of remote sensing in agriculture include phenotyping, land-use monitoring, crop yield forecasting, precision farming and the provision of ecosystem services [3–6].

Crop disease monitoring and more importantly the accurate early forecasting of crop diseases outbreak by making use of the remote sensing data have also attracted

much attention in recent years [6–9]. In particular, temperature and humidity are the two most important environmental factors determining the activities of pathogenic microorganisms. A good example is the wheat stripe rust, one of the most destructive diseases on wheat in China, whose outbreak potential is largely determined by temperature. On the one hand, snow covering in winter means freezing low temperature, however, a study has also shown that winter snow covering could also weaken the negative effects of extreme cold temperature on rust survival [10]. A crop disease outbreak forecasting model with the consideration of snow coverage and duration information would no doubt generate better prediction performance. However, it is still a challenge to obtain accurate snow coverage and duration data for the use in the forecasting model [8]. This is because manually obtaining these information for various areas of interest (mountainous areas in particular) would pose serious logistics and safety issues for agronomists.

In the last decades, many remote sensing satellites with various sensors (including multispectral sensors) have been launched to the Earth's orbits, thus estimating the snow coverage from satellite images turns out to be the most promising solution for large-scale applications. Especially, some of these satellite data are publicly accessible and completely free, for example, Sentinel-2 satellites are able to provide multispectral imagery with a 10-meter resolution (the one with the highest resolution among freely available satellites) and a 5-day global revisit frequency for spatial-temporal land surface monitoring [11]. One of the most challenging parts of snow coverage mapping via satellite imagery is to distinguish snow from cloud. The major issue is that snow and cloud share very similar appearance and color distribution, a result, manually separating snow pixels from cloud pixels requires expert knowledge and is still a tedious and time-consuming process. This similarity (between snow and cloud) also poses challenges for image classification algorithms, particularly for color images.

Currently, there are several empirical thresholds test-based tools available to classify snow and cloud, such as Fmask[12], ATCOR [13] and Sen2Cor [14]. Despite the great potentials of accurate snow coverage mapping, there was surprisingly little literature specially focused on snow coverage segmentation via satellite images. In contrast, there were much more literature studying the segmentation of cloud, land cover scenes [15–18]. Zhu and Woodcock reported an empirical thresholds test-based method called Tmask that utilizes multi-temporal Landsat images to classify cloud, cloud shadow and snow, and they demonstrated that the multi-temporal mask method significantly outperformed the Fmask approach which only uses single-date data in discriminating snow and cloud [19]. Zhan et al. reported a fully convolutional neural network to classify snow and cloud in pixel levels for the Gaofen #1 satellite imagery, with an overall precision of 92.4% in their own dataset [20]. However, the Gaofen #1 satellite imagery is unfortunately not publicly accessible and their manually labeled dataset is not publicly available, either. More often, within the snow classification-related studies, snow is only listed as one of the many sub-classes and accounts of very limited representation [17,21,22]. Although the empirical thresholds test-based methods such as Fmask[12], ATCOR [13] and Sen2Cor [14] are able to provide rough snow prediction in most scenarios, studies that focused on improving snow coverage segmentation performance by making use of machine (deep) learning algorithms are still missing. It should be highlighted that conventional machine learning methods and the recent deep learning methods (in particular) have made significant breakthroughs with appreciable performance in various applications including agriculture and is therefore worth investigation for snow coverage mapping as well [23,24].

Therefore, in this study, we firstly carefully collected 40 Sentinel-2 satellite images with scenes including representative snow, cloud and background that are distributed across almost all continents. Each pixel of the 40 scenes are labeled into three classes including snow, cloud and background via the semi-automatic classification plugin in QGIS, thus they become the largest publicly available satellite imagery dataset dedicated for snow coverage mapping. We then compared the reflectance distributions of the three

89 classes across the twelve Sentinel-2 spectral bands and found out the most informative  
 90 band combination that was able to distinguish snow, cloud and background. Lastly, we  
 91 compared the classification performance of three representative algorithms—Sen2Cor  
 92 for thresholds test-based method, random forest for traditional machine learning, and  
 93 U-Net for deep learning. We showed that U-Net model with the four informative bands  
 94 (including B2, B11, B4 and B9) as inputs gave the best classification performance on our  
 95 test dataset.

## 96 2. Materials and Methods

97 To make the work readable, the entire workflow is illustrated in Figure 1.

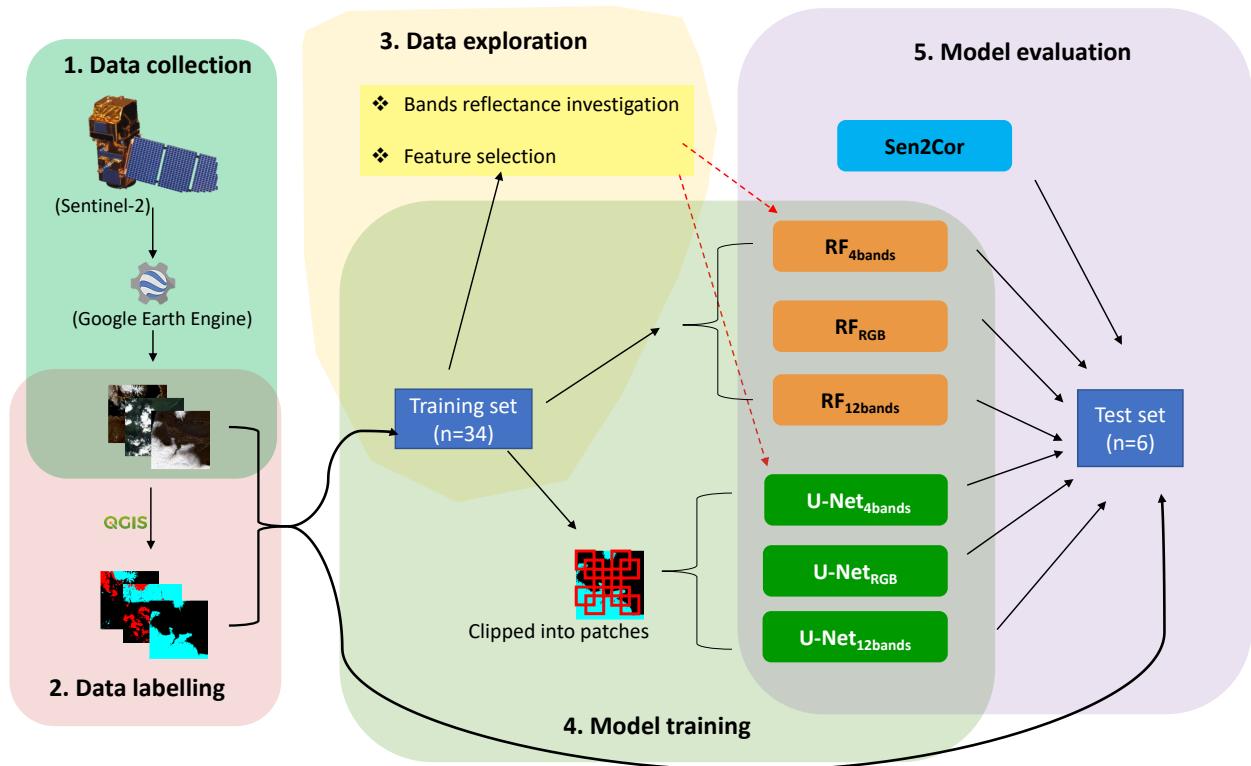


Figure 1. The entire workflow is divided into data collection, data labelling, data exploration, model training and model evaluation.

### 98 2.1. Satellite image collection

99 Sentinel-2 satellite images are freely accessible from several platforms, such as the  
 100 Copernicus Open Access Hub [25], USGS EROS Archive [26] and Google Earth Engine  
 101 [27] among others. In this study, all Sentinel-2 satellite images were directly downloaded  
 102 from the Google Earth Engine via some basic scripts in Java. Our main purpose is  
 103 to get snow mapping on the earth's surface, therefore, we focused on the corrected  
 104 Sentinel-2 Level-2A product instead of the Level-1C product, as the Level-2A provides  
 105 Orthoimage Bottom Of Atmosphere (BOA) corrected reflectance products. Moreover,  
 106 the Level-2A product has itself included a scene classification map, including cloud  
 107 and snow probabilities at 60 meters resolution, which are derived from the Sen2Cor  
 108 algorithm. However, it should also be noted that Level-2A products are only available  
 109 after December 2018, although a separate program is available to generate Level-2A  
 110 products from Level-1C product.

111 There are a total of 12 spectral bands for Sentinel-2 Level-2A product, which include  
 112 B1(Aerosols), B2(Blue), B3(Green), B4(Red), B5(Red Edge 1), B6(Red Edge 2), B7(Red  
 113 Edge 3), B8(NIR), B8A(Red Edge 4), B9(Water vapor), B11(SWIR 1) and B12(SWIR 2).  
 114 The cirrus B10 is omitted as it does not contain surface information. Of the 12 available

115 spectral bands, B2, B3, B4 and B8 are all in 10 meters resolution, the resolutions of B5, B6,  
116 B7, B8A, B9 and B12 are 20 meters, and the remaining two bands including B1 and B19  
117 have 60 meters resolution. Within all our downstream analyses, all spectral bands with  
118 resolutions different from 10 meters were re-sampled into 10 meters to achieve an equal  
119 spatial resolution across all spectral bands.

120 During our scene collection, we specifically choose scenes that include human  
121 identifiable snow, cloud, or both snow and cloud. It is important to ensure that the  
122 snow and cloud scenarios are human identifiable, as we are doing a supervised machine  
123 learning task and our next step data annotation is to label each pixel into three classes. To  
124 ensure that the collected dataset include a large diversity of scenes (to be representative),  
125 we selected the scenes to cover different continents, years, months and land cover classes.  
126 Lastly, we only kept a representative 1000 pixels  $\times$  1000 pixels region for each scene, this  
127 is on the one hand to reduce the redundant contents of a whole product, on the other  
128 hand, to greatly reduce the amount of times needed for the following data annotation  
129 step.

### 130 2.2. Data annotation

131 Upon downloading the representative images, the next step is to manually label  
132 them into different classes for the machine/deep learning model construction. The data  
133 annotation step involves manually labeling every pixel into one of the three classes  
134 (i.e. snow, cloud and background) by human experts. As the multi-spectral satellite  
135 images are not readily human-readable, we first extract the B4, B3 and B2 bands and  
136 re-scale them into the three channels of a typical RGB image (i.e. false-color RGB image).  
137 However, it is obvious that snow and cloud share very similar colors (i.e. close to  
138 white) and texture across many scenes, thus it is almost impossible to distinguish them  
139 especially when there are overlaps between snow and cloud. B12(SWIR 2) is known  
140 to have a relatively better separation between snow and cloud than other bands, thus  
141 we also created a false-color image, with B2, B3 and B12 as the R, G, B channels, for  
142 each scene. Then all the downstream labeling processes are performed on the false-color  
143 images.

144 The pixel-level labeling process was performed in QGIS platform (Version: 3.18.2)  
145 [28]. Recently, Luca Congedo developed a Semi-Automatic Classification Plugin for  
146 QGIS, which is reported to be able to ease and automate the phases of land cover  
147 classification [29]. Therefore, our labeling processes were completed with the help of the  
148 Semi-Automatic Classification Plugin (Version: 7.9) [29]. Specifically, for each image, we  
149 first select and define several representative regions for each target class, then we use the  
150 Minimum Distance algorithm of this plugin to group all other pixels into the pre-defined  
151 classes. All final generated classification masks were carefully checked by two human  
152 experts to make sure the label for snow and cloud is correct as far as possible.

### 153 2.3. Sen2Cor cloud mask and snow mask

154 The Sentinel-2 Level-2A product itself includes a cloud confidence mask and snow  
155 confidence mask, which are both derived from the Sen2Cor system. The algorithm  
156 used by Sen2Cor to detect snow or cloud is based on a series of threshold tests that  
157 use top-of-atmosphere reflectance from the spectral bands as input, the thresholds are  
158 also applied on band ratios and several spectral indices, such as Normalized difference  
159 vegetation index (NDVI) and Normalized Difference Snow Index (NDSI). In addition,  
160 a level of confidence is associated with each of these thresholds tests, the final output  
161 of the series of threshold tests are a probabilistic (0-100%) snow mask quality map and  
162 a cloud mask quality map [30]. In our study, the snow confidence mask and cloud  
163 confidence mask of each scene were directly downloaded from Google Earth Engine  
164 along with its Level-2A spectral bands data. For a better visualization of the Sen2Cor  
165 classification performance, for each satellite scene, we put the cloud confidence mask,

166 snow confidence mask and snow confidence mask into the three channels of a color  
167 image to generate the final Sen2Cor classification mask.

#### 168 2.4. Random forest with Bayesian hyperparameter optimization

169 The ‘sklearn.ensemble.RandomForestClassifier’ function in sklearn library (Version:  
170 0.24.2) [31] in Python (Version: 3.8.0) is used to construct the Random Forest (RF)  
171 models so as to evaluate the performance of traditional machine learning algorithms in  
172 classifying snow and cloud with the inputs of independent BOA corrected reluctance  
173 data from different spectral bands combinations.

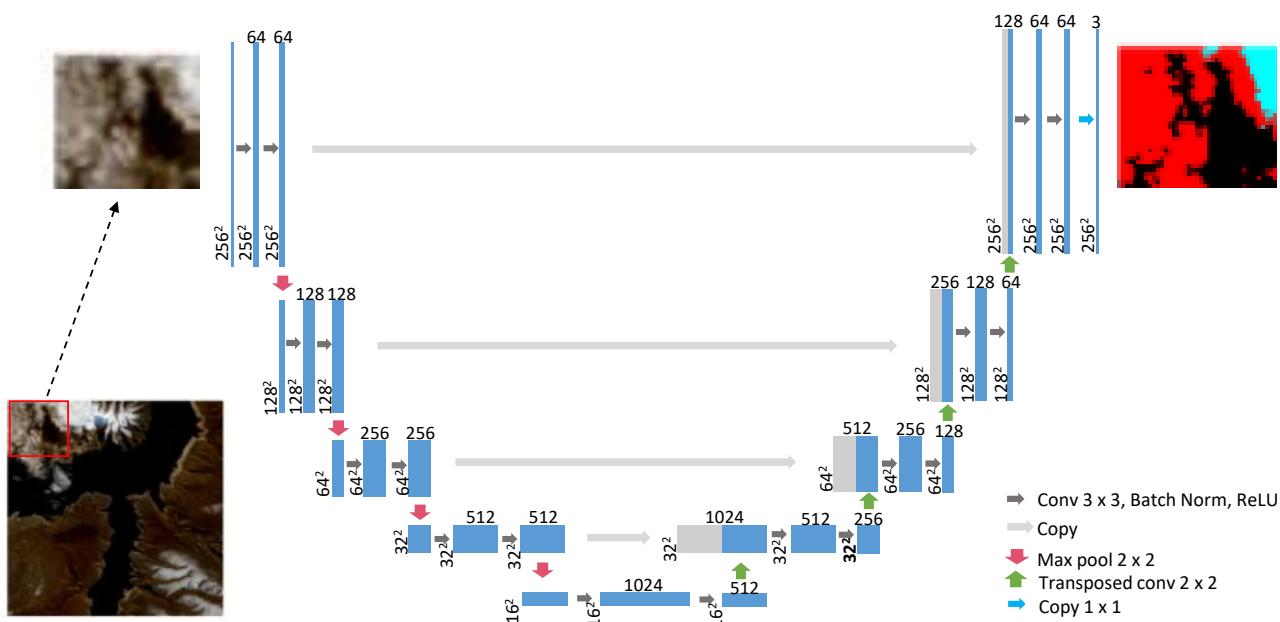
174 RF is a decision tree-based algorithm, which has been widely applied in crop disease  
175 detection [32]. To improve the prediction accuracy and control the problem of overfitting,  
176 we need to fine-tune several key hyper-parameters for the training of each RF model [33].  
177 In this study, we mainly fine-tuned three hyper-parameters: the number of trees in the  
178 forest, the maximum depth of the tree and the minimum number of samples required  
179 to split an internal node. Instead of using random or grid search for the optimal hyper-  
180 parameter combination, we applied the Bayesian Optimization [34] to find the optimal  
181 parameter combination for each RF model. Bayesian optimization enables finding out  
182 the optimal parameter combination in as few iterations as possible, which works by  
183 sequentially constructing a posterior distribution of functions (Gaussian process) that  
184 best describes the function you want to optimize. Here we used the average of five-  
185 fold cross-validation scores, which is resulted from training random forest model with  
186 weighted F1 score as loss function, as the optimization function of Bayesian optimization.  
187 After the Bayesian optimization, a random forest model with the optical parameters  
188 setting is trained.

189 We then applied both forward sequential feature selection (FSFS) and backward  
190 sequential feature selection (BSFS) to rank the importance of each spectral band and  
191 more importantly to find out the optimal bands combination which has fewer bands and  
192 at the same time are able to capture the most informative features. FSFS sequentially  
193 adds features and BSFS sequentially removes features to form a feature subset in a  
194 greedy fashion. FSFS starts with zero features, at each stage, it chooses the best feature  
195 to add based on the cross-validation score of an estimator (RF classifier in this study).  
196 In contrast, BSFS starts with full features and sequentially selects the least important  
197 features to be removed at each stage.

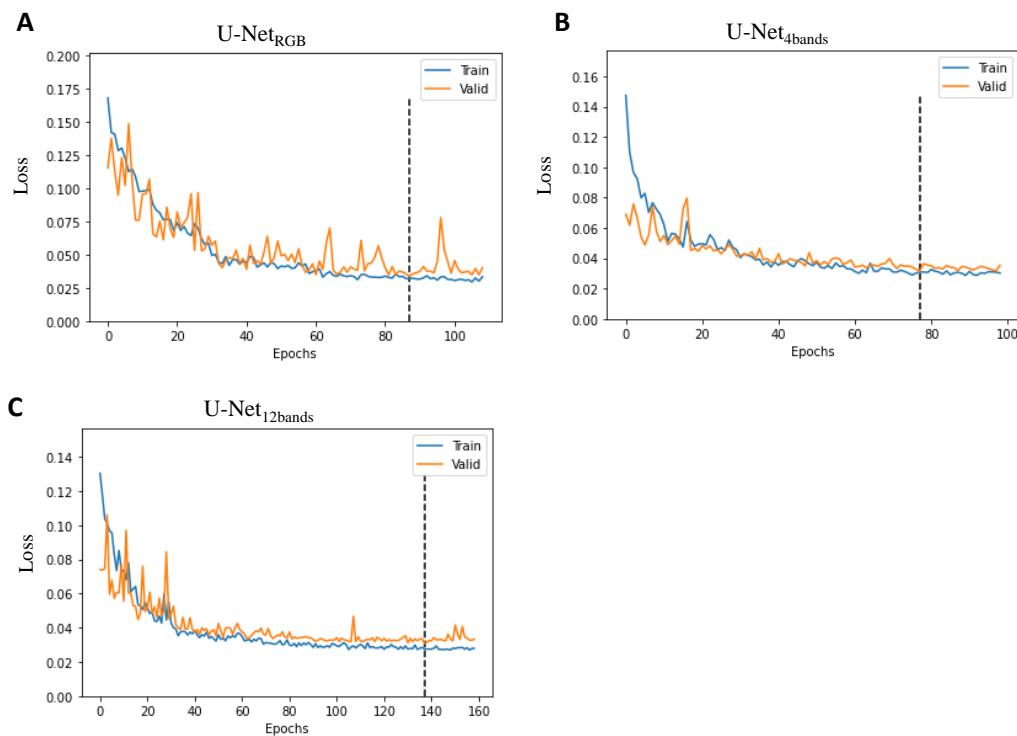
#### 198 2.5. U-Net training

199 U-Net is a convolutional network architecture for fast and precise segmentation  
200 of images [35], which has been applied for yellow rust disease mapping in [5,33]. In  
201 this study, U-Net is used as the representative deep learning model for satellite image  
202 classification. It is noted that our collected satellite images are in the size of  $1000 \times 1000$   
203 pixels. However, to make our deep learning models less rely on large-size images and  
204 also to greatly increase the samples size of training dataset, we set the model input with  
205 width of 256, height of 256 and channels of N, where N depends on the used bands  
206 combination in different models. For each satellite image in the training dataset, we  
207 clipped it into small patches in a sliding window way—with a window size of  $256 \times 256$   
208 pixels and a sliding step of 128 pixels. As a result, each satellite image will yield around  
209 49 patches of size  $256 \times 256 \times N$ .

210 The details of the U-Net network architecture are shown in Figure 2. It consists  
211 of an encoding path (left side) and a decoding path (right side). Every block of the  
212 encoding path consists of two repeated  $3 \times 3$  unpadding convolutions, each followed  
213 by a batch normalization (BN) layer and a rectified linear unit (ReLU), then a  $2 \times 2$   
214 max pooling operation with stride 2 is applied for downsampling. Each block of the  
215 decoding path includes a  $2 \times 2$  transpose convolution for feature upsampling, followed  
216 by a concatenation with the corresponding feature map from the encoding path, then  
217 two  $3 \times 3$  convolutions, each followed by a BN and a ReLU. The final layer is a  $1 \times 1$   
218 convolution to map each pixel in the input to the desired number of classes.



**Figure 2.** U-Net architecture used in this study. The blue boxes represent different multi-channel feature maps, with the numbers on the top and left edge of the box indicate the number of channels and the feature size (width and height) separately. Each white box represent a copied feature map. The arrows with different colors denote different operations. The number of channels is denoted on the top of the box.



**Figure 3.** Loss curves for training data (blue) and validation data (orange) in training process of (A) U-Net<sub>RGB</sub>, (B) U-Net<sub>4bands</sub> and (C) U-Net<sub>12bands</sub>. The dashed line indicates the epoch with smallest validation loss and the loss in the Y-axis represents the weighted cross-entropy.

219 The U-Net model is constructed and trained based on the Pytorch deep learning  
 220 framework (Version: 1.7.1) [36]. For the model training, we take the patches located in  
 221 the first column or first row of the generated patch matrix of each training satellite image  
 222 into the validation set and the remaining patches excluding those that have overlap with  
 223 validation patches are selected as the training data, the ratio of the number of validation  
 224 patches to training patches is about 19.1%. The model stops training until the loss of the  
 225 validation data do not decrease after 20 epochs. . To train the U-Net models, we used  
 226 the weighted cross-entropy loss as the loss function, stochastic gradient descent as the  
 227 optimizer with learning rate of 0.01 and momentum ratio of 0.9. The input batch size  
 228 is set to be 4. The loss curves of the training processes for U-Net with different input  
 229 bands are shown in Figure 3.

230 2.6. Evaluation matrix

To systematically compare the classification performance of the different models, we have used the following evaluation matrices including preision, recall, F1 score and Intersection Over Union(IoU) and Accuracy

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$F1\_score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (3)$$

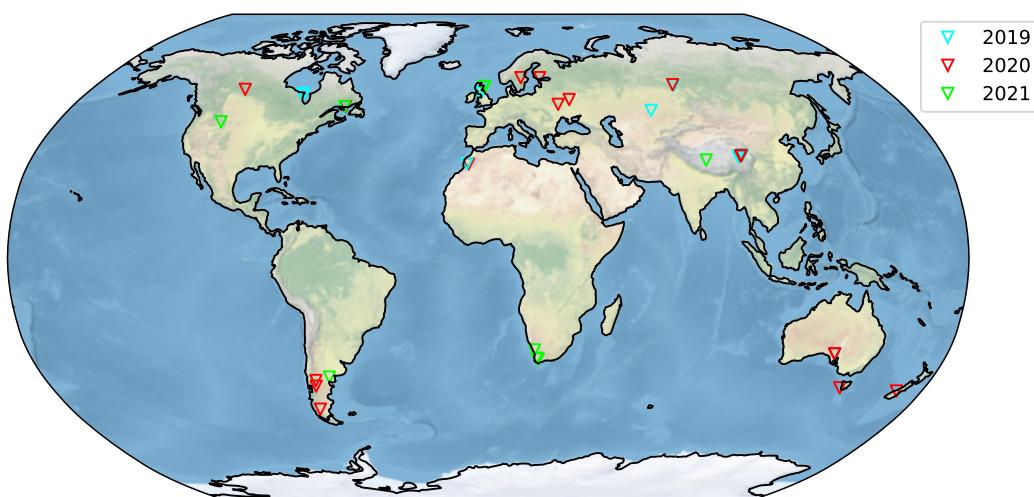
$$Intersection\ over\ Union(IoU) = \frac{Area\ of\ Overlap}{Area\ of\ Union} \quad (4)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

231 where TP represents true positive numbers, FP represents false positive numbers and  
 232 FN represents false negative numbers.

$$Loss(\text{weighted cross entropy loss}) = - \sum_{c=1}^M w_c y_{o,c} \log(p_{o,c}) \quad (6)$$

233 where  $M = 3$  is the number of classes (snow, cloud and background),  $w_c$  is the weight  
 234 value for class  $c$  which is defined as the ratio between the total number of pixels in  
 235 the training set and the number of pixels in class  $c$ ,  $y_{o,c}$  is binary indicator (0 or 1) if  
 236 class label  $c$  is the correct classification for observation  $o$ ,  $y_{o,c}$  is predicted probability of  
 237 observation  $o$  is of class  $c$ .



**Figure 4.** Geographical distribution of the 40 selected sites denoted by empty triangles, with different colours representing scenes obtained from different years, i.e. cyan, red and green denotes scenes dated from the years in 2019, 2020 and 2021.

<sup>238</sup> **3. Results**

<sup>239</sup> *3.1. Largest snow mapping satellite image dataset*

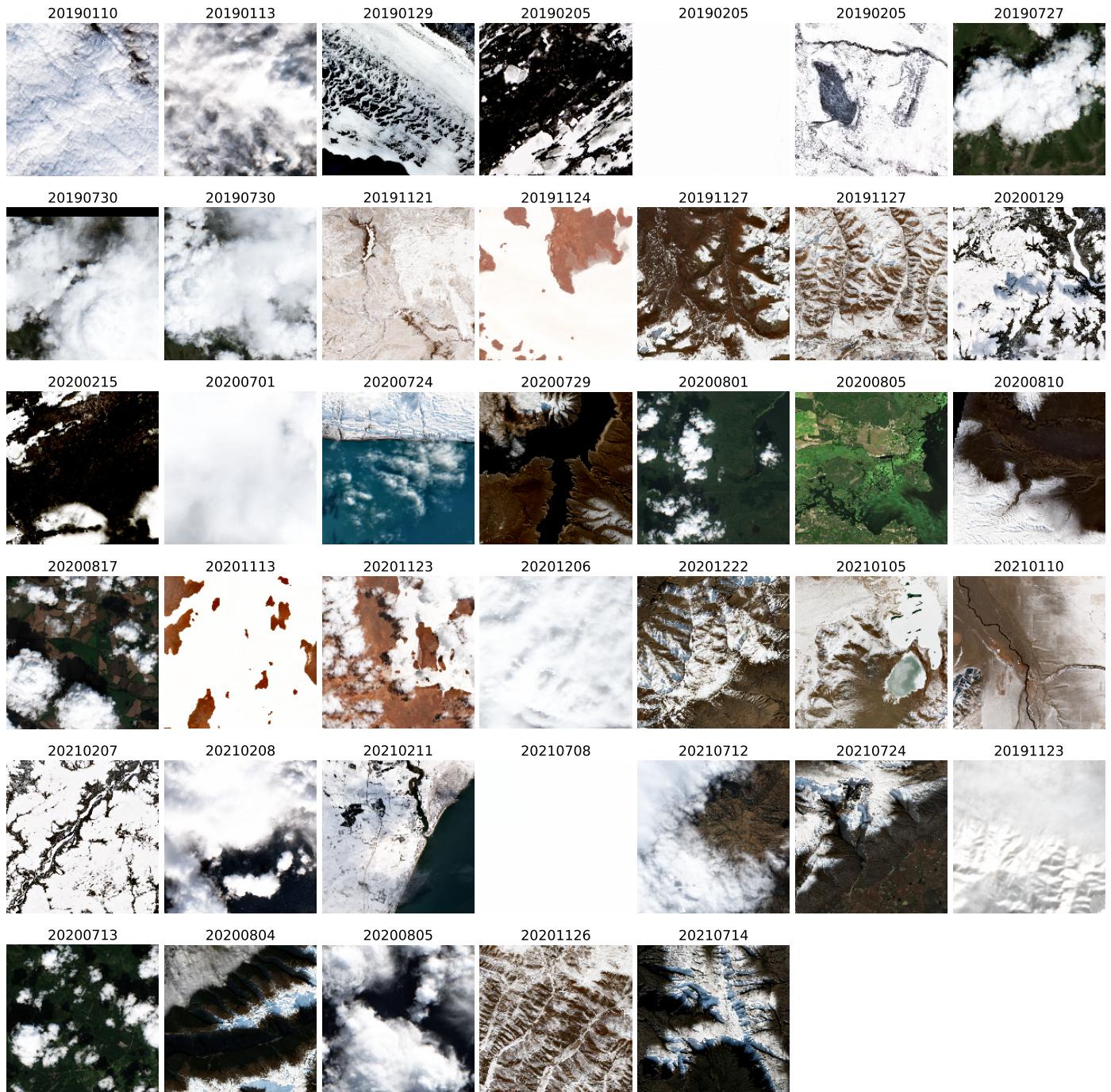
<sup>240</sup> To validate and compare different methods in classifying snow and cloud for  
<sup>241</sup> satellite images, we carefully searched and selected 40 Sentinel-2 L2A scenes across the  
<sup>242</sup> globe as displayed in Figure 4. In addition, the details about their product ID, coordinates  
<sup>243</sup> and timing are listed in Table 1.

**Table 1.** Product IDs and coordinates of the 40 collected scenes.

Product ID	X coordinate	Y coordinate	Date
20190110T112439_20190110T112436_T30UVG	-2.96	55.25	2019-01-10
20190113T113429_20190113T113432_T30UVG	-4.09	55.17	2019-01-13
20190129T151649_20190129T151651_T20UNA	-62.2	50.16	2019-01-29
20190205T055031_20190205T055025_T45VUE	85.08	57.71	2019-02-05
20190205T164459_20190205T164614_T17ULA	-82.53	54.86	2019-02-05
20190205T164459_20190205T164614_T17ULA	-83.13	54.47	2019-02-05
20190727T035549_20190727T040012_T47SPT	100.5	33.99	2019-07-27
20190730T040549_20190730T041756_T47SMS	98.79	33.4	2019-07-30
20190730T040549_20190730T041756_T47SMT	98.77	33.68	2019-07-30
20191121T062151_20191121T062148_T42UWV	69.8	48.97	2019-11-21
20191124T005709_20191124T010446_T53HNE	136.05	-31.79	2019-11-24
20191127T041109_20191127T041653_T47SMS	98.95	33.38	2019-11-27
20191127T041109_20191127T041653_T47SPT	100.16	34.18	2019-11-27
20200129T151641_20200129T151643_T20UNA	-61.71	50.32	2020-01-29
20200215T054929_20200215T054925_T45VUE	84.81	57.94	2020-02-15
20200701T095031_20200701T095034_T34VFN	24.22	60.46	2020-07-01
20200724T142739_20200724T143750_T18FXK	-72.5	-50.18	2020-07-24
20200729T142741_20200729T143445_T19GCR	-70.2	-40.51	2020-07-29
20200801T182919_20200801T183807_T12VWH	-110.14	56.28	2020-08-01
20200805T085601_20200805T085807_T35UQS	30.37	51.06	2020-08-05
20200810T141739_20200810T142950_T19GCP	-70.8	-42.4	2020-08-10
20200817T084559_20200817T085235_T36UXD	35.63	52.73	2020-08-17
20201113T005711_20201113T005712_T53JNF	135.86	-31.65	2020-11-13
20201123T005711_20201123T010434_T53HNE	135.79	-31.71	2020-11-23
20201206T041141_20201206T041138_T47SNT	99.74	33.72	2020-12-06
20201222T111501_20201222T111456_T29RPQ	-7.74	31.1	2020-12-22
20210105T050209_20210105T050811_T45SUR	85.74	32.43	2021-01-05
20210110T182731_20210110T182953_T12TVR	-111.66	45.12	2021-01-10
20210207T151649_20210207T151817_T20UNA	-62.05	50.29	2021-02-07
20210208T112129_20210208T112318_T30VWJ	-1.57	57.14	2021-02-08
20210211T113319_20210211T113447_T30VWJ	-2.02	57.31	2021-02-11
20210708T141051_20210708T142222_T20HLB	-64.31	-39.38	2021-07-08
20210712T082611_20210712T084900_T34JBM	18.11	-30.39	2021-07-12
20210724T081609_20210724T083856_T34HCJ	19.67	-33.41	2021-07-24
20191123T111259_20191123T112151_T29RNQ	-8.3	31.22	2019-11-23
20200713T103031_20200713T103026_T33VWH	15.31	60.46	2020-07-13
20200804T223709_20200804T223712_T59GLM	169.47	-44.02	2020-08-04
20200805T001109_20200805T001647_T55GCN	145.04	-42.94	2020-08-05
20201126T041121_20201126T041842_T47SNT	100.09	33.93	2020-11-26
20210714T081609_20210714T083805_T34HCJ	19.16	-33.13	2021-07-14

<sup>244</sup> The 40 sites have been chosen in order to ensure the scene diversity. In particular,  
<sup>245</sup> the 40 sites are distributed in six continents but except Antarctica. With the constant high  
<sup>246</sup> temperature in the low latitudes, our selected snow and cloud scenes are all distributed  
<sup>247</sup> in the middle and high latitude areas. Since the Sentinel-2 Level-2A products are only  
<sup>248</sup> available since December, 2018, our collected scenes are all dated from the last three

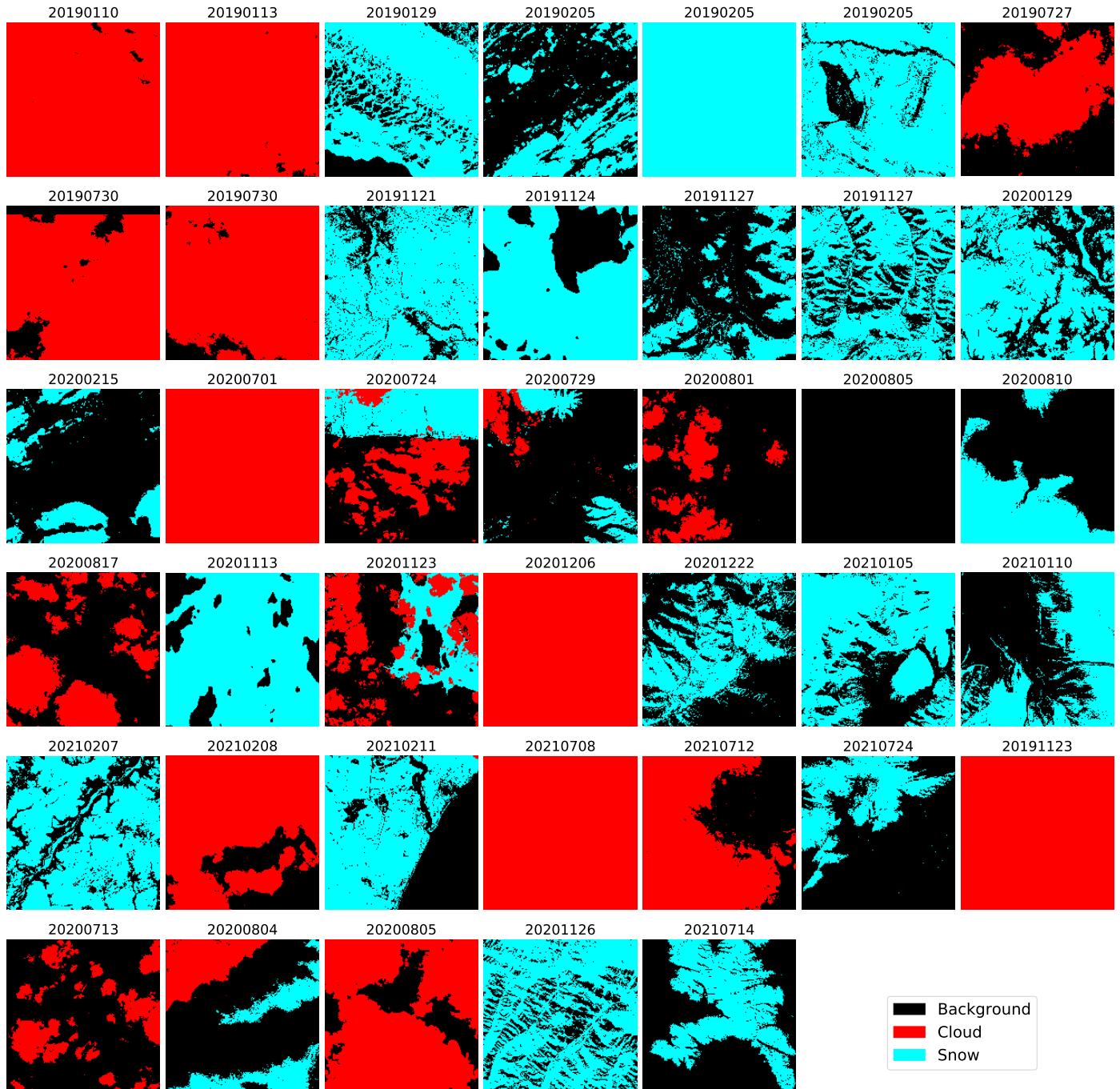
<sup>249</sup> years period, including 2019, 2020 and 2021. For each scene, all 12 atmospheric corrected  
<sup>250</sup> spectral bands including B1, B2, B3, B4, B5, B6, B7, B8, B8A, B9, B11 and B12 are collected,  
<sup>251</sup> also the cloud confidence mask and the snow confidence mask which are derived from  
<sup>252</sup> the Sen2Cor algorithm are also downloaded along with the spectral bands by the Google  
<sup>253</sup> Earth Engine. Each band of the scene is re-sampled into 10 meters resolution. For each  
<sup>254</sup> scene product, we only kept a representative region, with the sizes of width and height  
<sup>255</sup> are both 1000 pixels, which contains human identifiable snow or cloud.



**Figure 5.** Visualisation of all 40 scenes via RGB bands, with the above numbers being the scene captured date.

<sup>256</sup> Every pixel of the all 40 collected satellite images were labelled into three classes  
<sup>257</sup> including snow, cloud and background by using the Semi-Automatic Classification  
<sup>258</sup> Plugin [29] in QGIS. We took six representative scenes as the test dataset, their false-color

259 RGB images and classification masks are shown in Figure 12. The remaining 34 scenes  
 260 were put into the training dataset, their RGB images and classification masks are shown  
 261 in Figure 5 and 6, respectively.



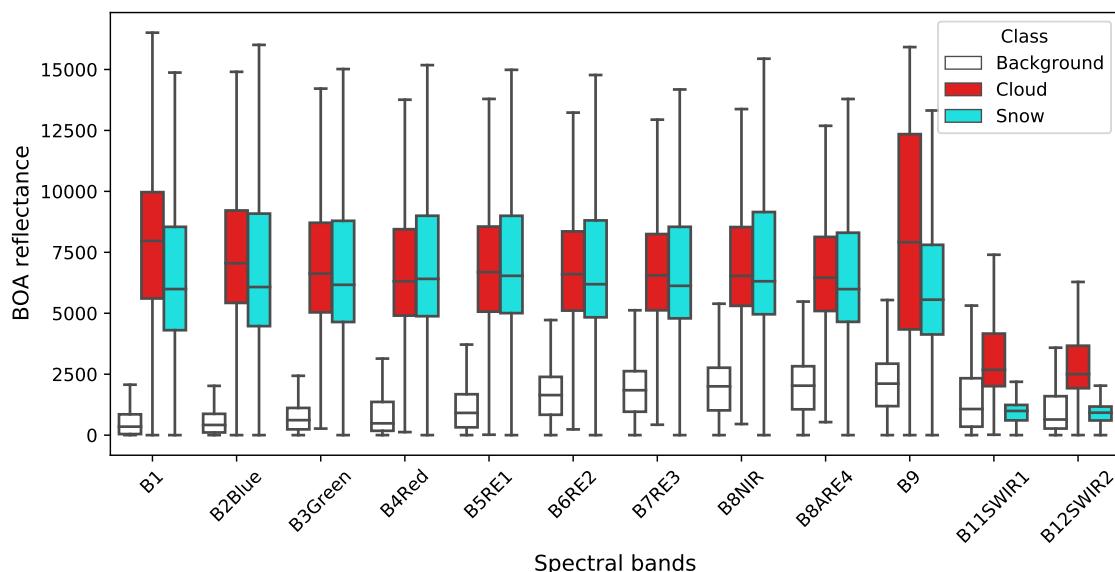
**Figure 6.** Labelled classification masks of all 40 collected scenes. The three target classes are represented by different colors: black denotes background, red denotes cloud and cyan denotes snow.

### 262 3.2. Spectral band comparison

263 Snow and cloud are both white bright objects seen from the satellite RGB images,  
 264 and they are often indistinguishable in most scenarios by only looking at the RGB images.  
 265 We first compared the reflectance distributions of snow, cloud and background in the  
 266 twelve spectral bands from the Sentinel-2 L2A product. From the boxplots in Figure 7,

we could first observe that the background pixels have relative low reflectance values across all the twelve spectral bands, and the median reflectance values are all less than 2500. Snow and cloud showed similar and relatively high reflectance values (median reflectance values are greater than 5000) in the first ten spectral bands, however, they also both have high reflectance variations in these ten bands.

With regard to snow and cloud, B12 and B11 are the top two bands that best separate snow and cloud, with a median reflectance around 950 in snow, compared to that around 2500 in the cloud. This is in line with our expectations, as B12 and B11 are both designed to measure short waves infrared (SWIR) wavelengths, and they are often used to differentiate between snow and cloud. However, the distribution of background is very similar to snow in B11 and B12 although with larger fluctuations. B9, B1 and B2 are the three next bands that have a relatively larger distribution difference between snow and cloud. Interestingly, even though snow and cloud have very similar reflectance distributions in the first 10 bands, the cloud has a slightly higher median value than snow in nine out of the ten bands (except B4). In summary, there are several spectral bands that have good separation between any two of the three classes, however, there is no single band that clearly separates the three classes simultaneously.

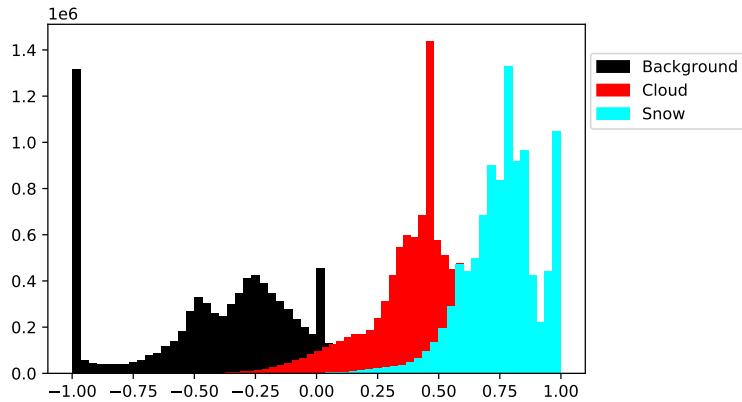


**Figure 7.** Boxplots comparing the bottom of atmosphere corrected reflectance of twelve spectral bands from Sentinel-2 L2A products for background (white), cloud (red) and snow (cyan). Note: the outliers of each boxplot are not displayed.

The Normalized Difference Snow Index (NDSI) has been suggested to be useful in estimating fractional snow cover [37,38], it measures the relative magnitude of the reflectance difference between the visible green band and SWIR band. Here we also compared the NDSI distribution of snow, cloud and background in the training dataset, where the results are displayed in Figure 8. Our results showed that despite there being three major different spikes representing the three classes respectively, the huge overlaps between the snow spike and the cloud suggest that NDSI (though being the best index for snow mapping) is not a very accurate index to distinguish snow and cloud.

### 3.3. Optimal band combination

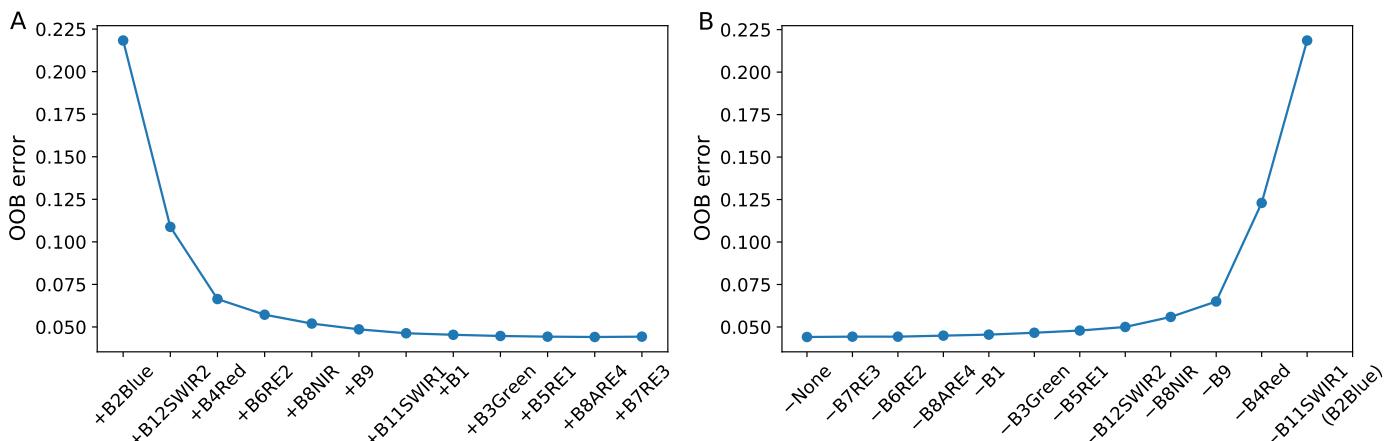
Our results in the previous section have demonstrated that no single spectral band (or index) is able to give clean separations between the three classes, i.e. snow, cloud and background. However, combining several bands is very promising. For example, the background pixels have clear separations with snow and cloud across the first ten bands, and the reflectance distribution of cloud is largely different from that in background and the snow within B12 or B11. Among the twelve spectral bands of the Sentinel-2



**Figure 8.** NDSI distribution of snow (cyan), cloud (red) and background (black) pixels, where the NDSI is defined as  $\text{NDSI} = (\text{B3} - \text{B12}) / (\text{B3} + \text{B12})$

299 Level-2A product, some bands may capture similar features and thus may be redundant  
 300 when used to distinguish the three classes. To find out the optimal bands combination  
 301 that captures the most useful information to discriminate the three classes and at the  
 302 same time has fewer (saving computation resources) bands included, we applied both  
 303 forward sequential feature selection (FSFS) and backward sequential feature selection  
 304 (BSFS) (please refer to methods section for details), where their results are displayed in  
 305 Figures 9.

306 As shown in Figure 9, B2 (Blue) band is ranked as the most important band by both  
 307 forward and backward sequential feature selection. B12 and B11 are two SWIR bands,  
 308 and they are listed as the second most important bands by FSFS and BSFS, respectively.  
 309 However, the band combination of B2 and B12 slightly outperforms the B2 and B11  
 310 combination when used as input for constructing models to separate the three classes  
 311 (OOB errors 0.109 vs. 0.123). FSFS and BSFS both identified B4 (Red) as the third  
 312 most important band, and again the combination of B2, B11 and B4 identified by BSFS  
 313 demonstrated as the best three-bands combination. The sequentially adding more bands  
 314 into the model input subset gives minimal improvements, especially when the top four  
 315 bands have already been included. As a result, we take the combination of B2, B11, B4  
 316 and B9 as the most informative band set of Sentinel-2 Level-2A products for separating  
 317 snow, cloud and background. It should be noted, although we re-sampled each band  
 318 into the highest 10 meters resolution, the original resolutions for B2 and B4 are 10 meters,  
 319 B11 is 20 meters and B9 is 60 meters.



**Figure 9.** Feature selection. (A) Forward sequential feature selection, where the tick name of the x-axis means sequentially adding the specified bands into the inputs of the model. (B) Backward sequential feature selection, where the tick name of the x-axis means sequentially removing the specified bands from the inputs of the model.

320     3.4. Performance comparison for RF models with various band combinations

321       We trained three RF models, each with a different band combination as input, and  
322       compared their performance in classifying snow, cloud and background. The three-  
323       bands combinations are RGB bands (B4, B3 and B2), the informative four bands (B2, B11,  
324       B4 and B9) and all twelve bands (B1, B2, B3, B4, B5, B6, B7, B8, B8A, B9, B11 and B12).  
325       The hyper-parameters of each random forest model were optimised independently by  
326       Bayesian optimization to achieve each model's best performance.

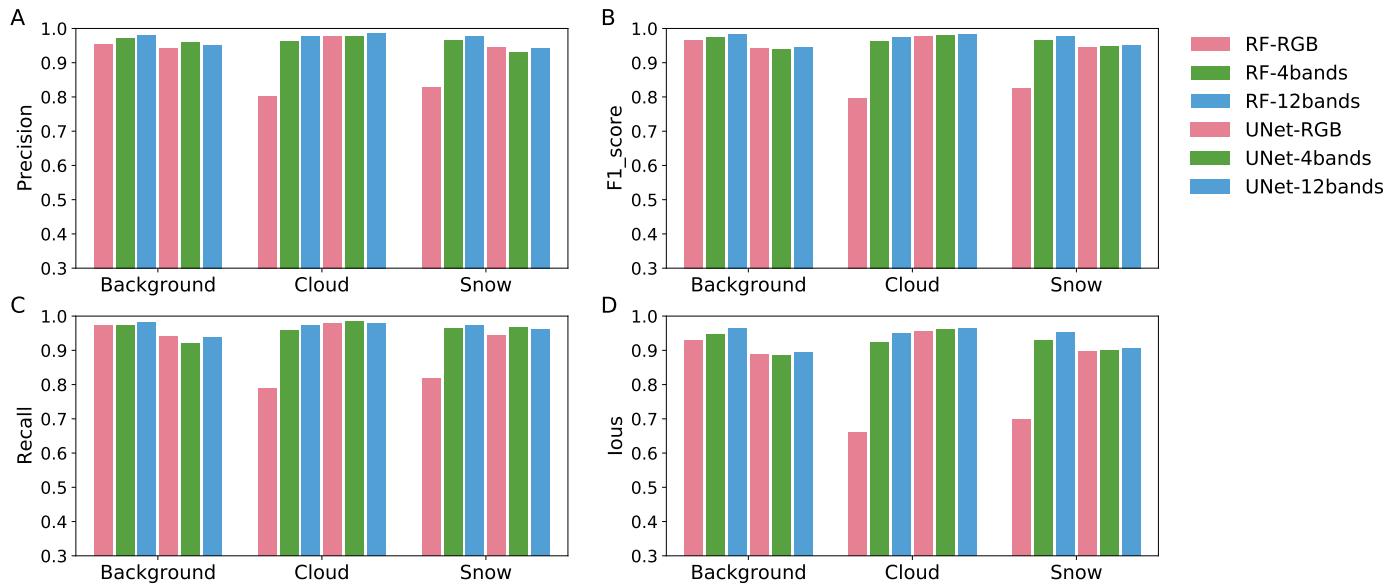
327       The comparison of four evaluation scores (precision, F1 score, recall and IoU)  
328       between the three RF models is demonstrated in Figures 10 (on training dataset) and  
329       11 (on test dataset). The three RF models showed very close and good performance (all  
330       above 0.86) in predicting background pixels across all four evaluation scores. This is in  
331       line with our previous observation that background pixels have distinct BOA reflectance  
332       distribution than cloud and snow in each of the first ten bands (Figure 7). The most  
333       apparent difference between the three models is that the RF model trained on RGB  
334       bands ( $RF_{RGB}$ ) exhibited very poor performance in predicting both cloud and snow.  
335       For example, the IoU of  $RF_{RGB}$  in predicting cloud and snow are both below 0.35, and  
336       F1 scores are both only around 0.5. These results demonstrate the RGB bands (i.e. B4,  
337       B3 and B2) do not contain enough information to discriminate snow from cloud, this  
338       finding is also reflected by that snow and cloud share similar reflectance distribution  
339       patterns across the three bands.

340       The RF model trained on the four informative bands ( $RF_{4bands}$ ) and the RF model  
341       trained on all 12 bands ( $RF_{12bands}$ ) exhibited very close performance in predicting all  
342       three classes, and they both significantly outperform the  $RF_{RGB}$  model in predicting both  
343       cloud and snow. The four bands input (B2, B11, B4 and B9) for  $RF_{4bands}$  was selected by  
344       BSFS to maximize the informative features and simultaneously minimise the number  
345       of bands. The previous section (Figure 9) has demonstrated that the top 4 important  
346       bands combined accounted for almost all the informative features in classifying the  
347       three classes, this explains that  $RF_{4bands}$  and  $RF_{12bands}$  showed close performance and  
348       are much better than  $RF_{RGB}$ . A striking finding is that  $RF_{4bands}$  is even marginally better  
349       than  $RF_{12bands}$  in classifying the three classes based on all four evaluation scores, except  
350       for the precision of cloud and recall of snow. This may be due to that the inclusion of  
351       more similar or non-relevant bands may make the machine learning model at higher  
352       risk of overfitting on the training dataset, and this point is supported by the finding that  
353        $RF_{12bands}$  slightly outperformed  $RF_{4bands}$  in classifying all three classes across the four  
354       evaluation scores in the training dataset (please refer to Figure 10).

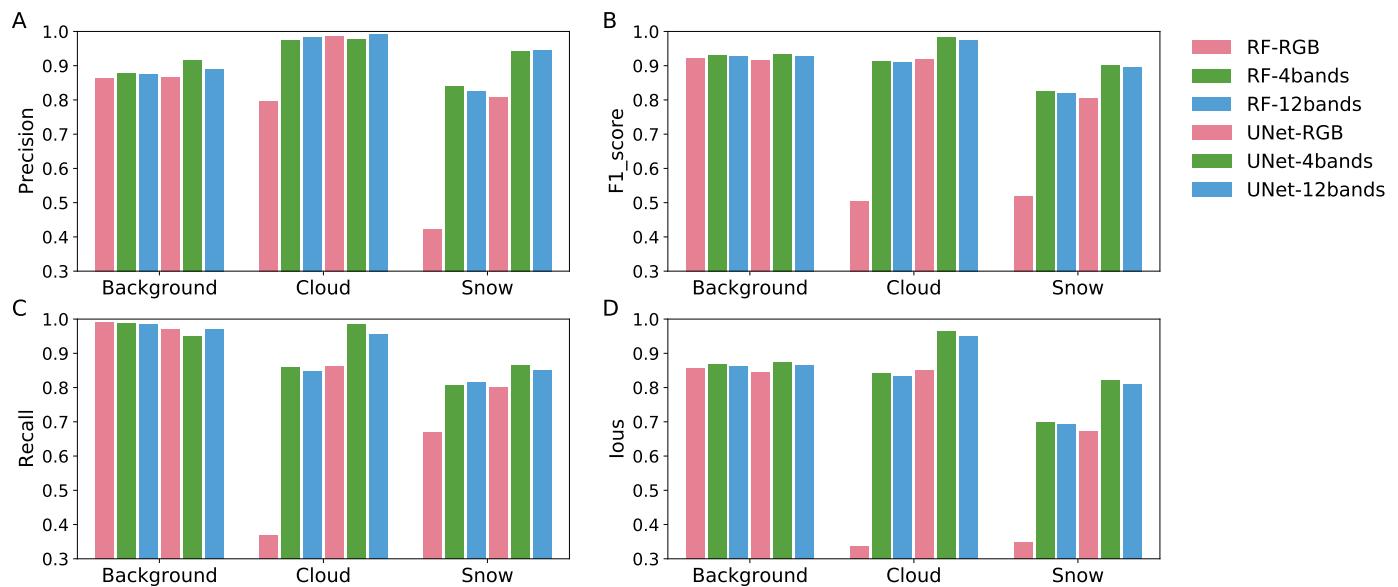
355     3.5. Performance comparison for U-Net with various band combinations

356       In addition to the RF models, we also trained three U-Net models with RGB bands,  
357       informative four bands and all 12 bands as inputs. Except for the input layer, all other  
358       layers of the U-Net structure are the same for the three U-Net models. We then compared  
359       their classification performances based on the four evaluation metrics.

360       Similar to the RF models, the three U-Net models achieved close and good perfor-  
361       mance (all above 0.84) in classifying background pixels according to all four evaluation  
362       scores. The U-Net model with informative four bands as inputs ( $U\text{-Net}_{4bands}$ ) and the  
363       U-Net model with all bands as inputs ( $U\text{-Net}_{12bands}$  model) also exhibited close perfor-  
364       mance in predicting snow pixels and cloud pixels according to the four scores, even  
365       though the  $U\text{-Net}_{4bands}$  model slightly and consistently outperformed the  $U\text{-Net}_{12bands}$   
366       model. The U-Net model with RGB bands as inputs ( $U\text{-Net}_{RGB}$ ) apparently fell behind  
367        $U\text{-Net}_{4bands}$  and  $U\text{-Net}_{12bands}$  in classifying snow and cloud in almost all evaluation  
368       scores, except that the three models all achieved nearly perfect scores (all are greater  
369       than 0.987) on precision for cloud.



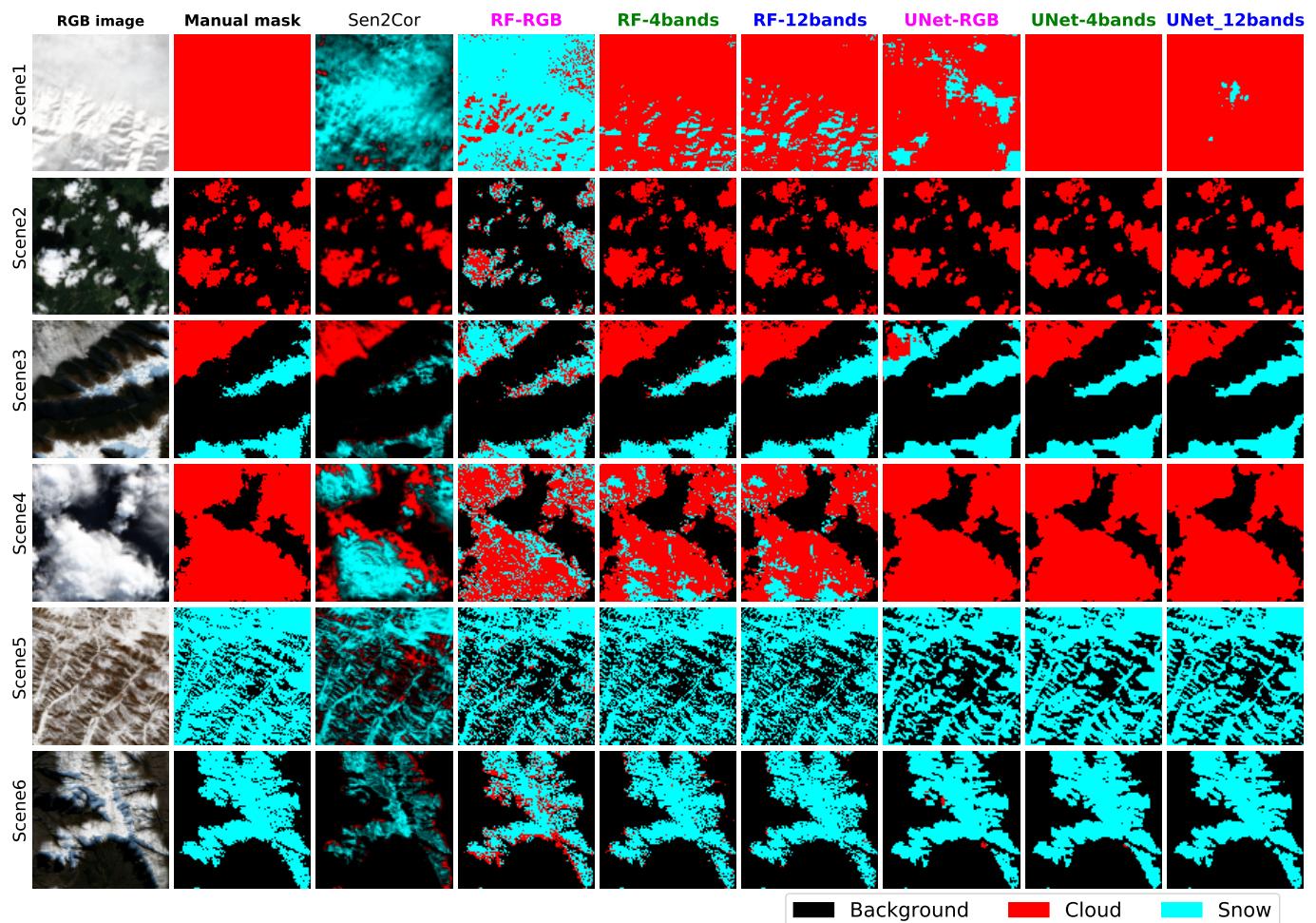
**Figure 10.** Classification performance comparisons for different models applied in training dataset images ( $n=34$ ) based on (A) precision, (B) F1 score, (C) recall and (D) IoU. The bars with three different colours, i.e. violet, green and blue, represent models with input subset made up of RGB bands, informative four bands and all 12 bands, respectively. The bar without texture denotes random forest model, while the bar with diagonal texture symbols U-Net model. Note: the evaluation was performed on image level, therefore the validation dataset pathes are also included.



**Figure 11.** Classification performance comparisons for different models applied in testing dataset images ( $n=6$ ) based on (A) precision, (B) F1 score, (C) recall and (D) IoU. The bars with three different colours, i.e. violet, green and blue, represent models with input subset made up of RGB bands, informative four bands and all 12 bands respectively. The bar without texture denotes random forest model, while the bar with diagonal texture symbols U-Net model.

### 3.6. Comparison between Sen2Cor, RF and U-Net models

We then compared the classification performance of Sen2Cor, RF and U-Net models. In terms of overall accuracy, Sen2Cor gave very poor classification results (only 58.06%) on our selected six test scenes, apparently falling far behind other models (Table 2). The poor classification performance of Sen2Cor is also reflected in its generated classification masks which are listed in the third column of Figure 12. The Sen2Cor mis-classified almost all cloud pixels in the first scene and the majority of cloud pixels in the third



**Figure 12.** Visual comparisons of the classification performance on six independent scenes for different methods. Each row represents an independent test scene, each column represents a different method. Except for the plots in the first column, the three target classes are represented by different colours, with black denotes background, red denotes cloud and cyan denotes snow.

377 scene into snow pixels, it also mis-classified many snow pixels, which are mainly located  
 378 at the boundaries between snow and background, in the last two scenes into clouds.

**Table 2.** Overall accuracy of different models.

Overall accuracy (%)	Sen2Cor	RF	U-Net
RGB	-	69.06%	87.72%
Informative 4 bands	-	90.03%	93.89%
All 12 bands	58.06%	89.65%	93.21%

379 The three U-Net models all clearly outperformed their corresponding RF models.  
 380 The greatest improvement came from the comparison between  $RF_{RGB}$  and  $U\text{-Net}_{RGB}$ .  
 381 The overall accuracy for  $RF_{RGB}$  is 69.06%, while it significantly increased to 87.72% by  
 382  $U\text{-Net}_{RGB}$  which is even closer to the performance of  $RF_{12bands}$  (89.65%). Like Sen2Cor,  
 383  $RF_{RGB}$  misclassified almost all cloud pixels in the first scene into snows, however,  $U\text{-Net}_{RGB}$   
 384 managed to correctly classify around 80% of the cloud pixels in the first scene.  
 385 Both  $RF_{RGB}$  and  $U\text{-Net}_{RGB}$  tend to predict all snow and cloud pixels in the third scene  
 386 as snow pixels. Furthermore,  $RF_{RGB}$  misclassified a lot of cloud pixels in the second

387 scene and the fourth scene into snows and predicted many snow pixels in the last scene  
388 as cloud pixels, while U-Net<sub>RGB</sub> does not have such issues (Figure 12). The above  
389 results indicate that the pure pixel-level reflectance from the RGB bands does not contain  
390 enough informative features to discriminate snow from cloud, the additional addition  
391 of spatial information as employed by U-Net model greatly improved the classification  
392 results.

393 Even though the overall accuracy of RF<sub>4bands</sub> and RF<sub>12bands</sub> reached to around 90%,  
394 which is a huge improvement over RF<sub>RGB</sub>, they still both misclassified many cloud pixels  
395 into snows in the first and fourth scene. U-Net<sub>4bands</sub> and U-Net<sub>12bands</sub> further increased  
396 the overall accuracy to above 93%, and they with U-Net<sub>RGB</sub> all avoided such “cloud  
397 to snow” misclassification issues in the fourth scene, U-Net<sub>4bands</sub> even further uniquely  
398 correctly classified all cloud pixels in the first scene (Figure 12). U-Net<sub>4bands</sub> achieved the  
399 highest overall accuracy of 93.89% and combining with its outstanding score in the other  
400 four evaluation matrices (Figure 11), makes it to be the best model among the six models  
401 we have studied to do snow mapping for Sentinel-2 imagery.

#### 402 4. Discussion

403 Snow coverage information is important for a wide range of applications, especially  
404 in agriculture, accurate snow mapping could be a vital factor for developing models  
405 to predict next-year disease development. However, an accurate snow mapping from  
406 satellite images is still a challenging task, as cloud and snow share similar spectral  
407 reflectance distribution (visible spectral bands in particular), and therefore it is not easily  
408 distinguishable. To our best knowledge, there is no large annotated satellite image  
409 datasets especially for the task of snow mapping that are currently publicly available.  
410 Although Hollstein et al. [17] manually labelled dozens of small polygonal regions from  
411 scenes of Sentinel-2 Level-1C products across the globe at 20 meters resolution, the small  
412 isolated irregular polygons are not useful enough to train convolutional neural network-  
413 based models. Baetens et al. [21] annotated 32 scenes of Sentinel-2 level-2A products in  
414 10 locations at 60 meters resolution, however, they were mainly focused on generating  
415 cloud masks and snow has very limited representation. As a result, we carefully collected  
416 and labelled 40 scenes of Sentinel-2 level-2A products at the highest 10 meters resolution,  
417 which includes a wide representative of snow, cloud and background landscape across  
418 the globe. The proposed database would one the one hand be used to evaluate the  
419 performance of different snow prediction tools, and on the other hand enables the future  
420 development of more advanced algorithms for snow coverage mapping.

421 Thresholds tests-based tools (such as Sen2Cor) could be used to make fast and  
422 rough estimations of snow or cloud. However, our results have demonstrated that  
423 they can be misleading under some circumstances. In particular, Sen2Cor tends to  
424 mis-classify the cloud under near-freezing environment temperature into snow, such as  
425 the case in the first and fourth scene of the test dataset (Figure 7). The thin layer of snow  
426 located in the junction between snow and background are also often mis-classified to be  
427 cloud. Thus an accurate snow coverage mapping requires much better snow and cloud  
428 classification tools.

429 The Sentinel-2 level-2A product includes 12 BOA corrected reflectance bands. Our  
430 results show that no single band is able to provide clean separations between snow,  
431 cloud and background, each of the twelve bands may contain redundant or unique  
432 features that are useful to classify the three classes. Including too many features, such as  
433 including all of the twelve bands, may easily lead to overfitting on training data for most  
434 machine learning and deep learning algorithms, especially when the training dataset is  
435 not large enough. Thus, identifying the optimal band combination that contains most of  
436 the informative features while also contains a few bands is essential. Our forward feature  
437 selection and backward feature selection both agreed B2 (blue) is the most important  
438 band to separate the three classes. The combination of B2, B11, B4 and B9 reserves  
439 almost all informative features among all twelve bands for separating the three classes.

440 So our results provide guidance for selecting bands for the following studies aimed at  
441 developing better snow mapping tools.

442 Random Forest as the representative traditional machine learning algorithm pro-  
443 vides much better classification performance than the thresholds test-based method  
444 Sen2Cor, especially when feeding the RF model with the four informative bands or all  
445 twelve bands. However, all the three RF models have the issue of “salt-and-pepper”  
446 noise on their classification masks [33], this issue does not only reflect the high variance  
447 of spectral reluctance values of each band within the three classes but also reflects the  
448 limitations of the traditional machine learning algorithms. Traditional machine learning  
449 algorithms, such as RF, only use pixel-level information, i.e. the reflectance values of  
450 each band for the same pixel, to make class predictions. They failed to make use of the  
451 information from the surrounding pixels or the broad spatial information. In contrast,  
452 the convolutional neural networks-based deep learning models such as U-Net, exploit  
453 the surrounding pixels information by convolution operations and take advantage of  
454 the broad spatial information by repeated convolution and pooling operations. There-  
455 fore, the U-Net models all bypass the “salt-and-pepper” issue and give even better  
456 classification performance than the RF models.

457 The important role of the spatial information in distinguishing snow and cloud is  
458 further highlighted when comparing the classification performance between U-Net<sub>RGB</sub>  
459 and RF<sub>RGB</sub>. The large overlaps of the reflectance distribution of the RGB bands between  
460 snow and cloud and the poor classification performance of RF<sub>RGB</sub> demonstrate that  
461 pixel-level information of only RGB bands contains very limited features that can be  
462 used to separate the three classes. In contrast, U-Net<sub>RGB</sub>, also only fed by RGB bands but  
463 incorporating spatial information by the neural networks, is able to achieve significant  
464 improvements in classification performance than RF<sub>RGB</sub>. Then it raises an interesting  
465 open question for future studies, i.e., with a larger training dataset and improved neural  
466 network algorithms, is it possible to build satisfactory models with inputs of only RGB  
467 bands?

468 In terms of practical applications, although we have demonstrated that the U-  
469 Net model fed with the four informative bands (RF<sub>4bands</sub>) achieved the best prediction  
470 performance in our test dataset, and is much better than Sen2Cor and RF models. We  
471 should acknowledge that the efficient execution of deep learning models often requires  
472 advanced hardware (e.g. GPU) and higher computation demands, thus making it less  
473 convenient to implement than the thresholds test-based methods. However, with the  
474 technology development and algorithms evolution, the application of deep learning  
475 models in satellite images would be the mainstream in the future.

## 476 5. Conclusions and future work

477 This work investigates the problem of snow coverage mapping by learning from  
478 Sentinel-2 multi-spectral satellite images via conventional machine and recent deep  
479 learning methods. To this end, the largest (to our best knowledge) satellite image dataset  
480 for snow coverage mapping is firstly collected by downloading sentinel-2 satellite  
481 images at different locations and times, followed by manually data labelling via a semi-  
482 automatic data labelling tool in QGIS. Then both random forest based conventional  
483 machine learning approach and U-Net based deep learning approach are applied to  
484 the labelled dataset so that their performance can be compared. In addition, different  
485 band inputs are also compared including RGB three-band image, selected four bands  
486 via feature selection and full multispectral bands. It is shown that (1) both conventional  
487 machine learning and recent deep learning methods significantly outperform the existing  
488 rule-based Sen2Cor product for snow mapping; (2) U-Net generally outperforms the  
489 random forest since both spectral and spatial information is incorporated in U-Net; (3)  
490 the best spectral band combination for snow coverage mapping is B2, B11, B4 and B9,  
491 even outperforming all spectral band combinations.

Although the results in this study are very encouraging, there is still much room for further development. For example, 1) in terms of data source, more labelled images from different locations and under diverse background conditions are required to generate a more representative dataset; 2) in terms of algorithm, the representative machine learning algorithm (e.g. random forest, U-Net) are compared to obtain a baseline performance in this study, more advanced deep learning algorithms should be further considered/developed in order to further improve the performance; 3) in terms of practical application, a supervised learning approach is adopted in this study, semi-supervised or even unsupervised algorithm should also be exploited so that the workload on data labelling can be significantly reduced.

**Author Contributions:** “Conceptualization, Y.Wang, J.Su and C.Liu; methodology, Y.Wang and J.Su; software, Y.Wang; validation, Y.Wang; formal analysis, Y.Wang; investigation, Y.Wang and J.Su; resources, J.Su and X.Zhai; data curation, Y.Wang and J.Su; writing—original draft preparation, Y.Wang; writing—review and editing, Y.Wang, J.Su, F.Meng and X.Zhai; visualization, Y.Wang; supervision, J.Su and X.Zhai; project administration, J.Su; funding acquisition, J.Su and C.Liu. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by UK Science and Technology Facilities Council (STFC) under Newton fund with Grant ST/V00137X/1.

**Data Availability Statement:** The dataset in this study will be openly shared upon publication at <https://github.com/yiluyucheng/SnowCoverage>

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Lama, G.F.C.; Crimaldi, M.; Pasquino, V.; Padulano, R.; Chirico, G.B. Bulk Drag Predictions of Riparian Arundo donax Stands through UAV-Acquired Multispectral Images. *Water* **2021**, *13*, 1333.
2. Jalonens, J.; Järvelä, J.; Virtanen, J.P.; Vaaja, M.; Kurkela, M.; Hyypä, H. Determining characteristic vegetation areas by terrestrial laser scanning for floodplain flow modeling. *Water* **2015**, *7*, 420–437.
3. Weiss, M.; Jacob, F.; Duveiller, G. Remote sensing for agricultural applications: A meta-review. *Remote Sensing of Environment* **2020**, *236*, 111402. doi:10.1016/j.rse.2019.111402.
4. Zhang, J.; Huang, Y.; Pu, R.; Gonzalez-Moreno, P.; Yuan, L.; Wu, K.; Huang, W. Monitoring plant diseases and pests through remote sensing technology: A review. *Computers and Electronics in Agriculture* **2019**, *165*, 104943.
5. Zhang, T.; Xu, Z.; Su, J.; Yang, Z.; Liu, C.; Chen, W.H.; Li, J. Ir-UNet: Irregular Segmentation U-Shape Network for Wheat Yellow Rust Detection by UAV Multispectral Imagery. *Remote Sensing* **2021**, *13*, 3892.
6. Su, J.; Liu, C.; Coombes, M.; Hu, X.; Wang, C.; Xu, X.; Li, Q.; Guo, L.; Chen, W.H. Wheat yellow rust monitoring by learning from multispectral UAV aerial imagery. *Computers and Electronics in Agriculture* **2018**, *155*, 157–166. doi:10.1016/j.compag.2018.10.017.
7. Zhang, J.; Huang, Y.; Pu, R.; Gonzalez-Moreno, P.; Yuan, L.; Wu, K.; Huang, W. Monitoring plant diseases and pests through remote sensing technology: A review. *Computers and Electronics in Agriculture* **2019**, *165*, 104943. doi:10.1016/j.compag.2019.104943.
8. Dong, Y.; Xu, F.; Liu, L.; Du, X.; Ren, B.; Guo, A.; Geng, Y.; Ruan, C.; Ye, H.; Huang, W.; Zhu, Y. Automatic System for Crop Pest and Disease Dynamic Monitoring and Early Forecasting. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2020**, *13*, 4410–4418. doi:10.1109/JSTARS.2020.3013340.
9. Hu, X.; Cao, S.; Cornelius, A.; Xu, X. Predicting overwintering of wheat stripe rust in central and northwestern China. *Plant Disease* **2020**, *104*, 44–51. doi:10.1094/PDIS-06-19-1148-RE.
10. Sharma-Poudyal, D.; Chen, X.; Rupp, R.A. Potential oversummering and overwintering regions for the wheat stripe rust pathogen in the contiguous United States. *International Journal of Biometeorology* **2014**, *58*, 987–997. doi:10.1007/s00484-013-0683-6.
11. Drusch, M.; Del Bello, U.; Carlier, S.; Colin, O.; Fernandez, V.; Gascon, F.; Hoersch, B.; Isola, C.; Laberinti, P.; Martimort, P.; others. Sentinel-2: ESA’s optical high-resolution mission for GMES operational services. *Remote sensing of Environment* **2012**, *120*, 25–36.
12. Zhu, Z.; Wang, S.; Woodcock, C.E. Improvement and expansion of the Fmask algorithm: Cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images. *Remote Sensing of Environment* **2015**, *159*, 269–277. doi:10.1016/j.rse.2014.12.014.
13. Richter, R.; Schläpfer, D. Atmospheric and topographic correction (ATCOR theoretical background document). *DLR IB* **2019**, pp. 564–03.
14. Louis, J.; Debaecker, V.; Pflug, B.; Main-Knorn, M.; Bieniarz, J.; Mueller-Wilm, U.; Cadau, E.; Gascon, F. Sentinel-2 Sen2Cor: L2A processor for users. Proceedings Living Planet Symposium 2016. Spacebooks Online, 2016, pp. 1–8.
15. Li, Z.; Shen, H.; Cheng, Q.; Liu, Y.; You, S.; He, Z. Deep learning based cloud detection for medium and high resolution remote sensing images of different sensors. *ISPRS Journal of Photogrammetry and Remote Sensing* **2019**, *150*, 197–212, [1810.05801]. doi:10.1016/j.isprsjprs.2019.02.017.

16. Shendryk, Y.; Rist, Y.; Ticehurst, C.; Thorburn, P. Deep learning for multi-modal classification of cloud, shadow and land cover scenes in PlanetScope and Sentinel-2 imagery. *ISPRS Journal of Photogrammetry and Remote Sensing* **2019**, *157*, 124–136. doi:10.1016/j.isprsjprs.2019.08.018.
17. Hollstein, A.; Segl, K.; Guanter, L.; Brell, M.; Enesco, M. Ready-to-use methods for the detection of clouds, cirrus, snow, shadow, water and clear sky pixels in Sentinel-2 MSI images. *Remote Sensing* **2016**, *8*, 1–18. doi:10.3390/rs8080666.
18. Helber, P.; Bischke, B.; Dengel, A.; Borth, D. EuroSAT: A novel dataset and deep learning benchmark for land use and land cover classification. *arXiv* **2017**, *12*, 2217–2226.
19. Zhu, Z.; Woodcock, C.E. Automated cloud, cloud shadow, and snow detection in multitemporal Landsat data: An algorithm designed specifically for monitoring land cover change. *Remote Sensing of Environment* **2014**, *152*, 217–234. doi:10.1016/j.rse.2014.06.012.
20. Zhan, Y.; Wang, J.; Shi, J.; Cheng, G.; Yao, L.; Sun, W. Distinguishing Cloud and Snow in Satellite Images via Deep Convolutional Network. *IEEE Geoscience and Remote Sensing Letters* **2017**, *14*, 1785–1789. doi:10.1109/LGRS.2017.2735801.
21. Baetens, L.; Desjardins, C.; Hagolle, O. Validation of copernicus Sentinel-2 cloud masks obtained from MAJA, Sen2Cor, and FMask processors using reference cloud masks generated with a supervised active learning procedure. *Remote Sensing* **2019**, *11*. doi:10.3390/rs11040433.
22. Zekoll, V.; Main-Knorn, M.; Louis, J.; Frantz, D.; Richter, R.; Pflug, B. Comparison of masking algorithms for sentinel-2 imagery. *Remote Sensing* **2021**, *13*, 1–21. doi:10.3390/rs13010137.
23. Ma, L.; Liu, Y.; Zhang, X.; Ye, Y.; Yin, G.; Johnson, B.A. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS journal of photogrammetry and remote sensing* **2019**, *152*, 166–177.
24. Sadeghifar, T.; Lama, G.; Sihag, P.; Bayram, A.; Kisi, O. Wave height predictions in complex sea flows through soft-computing models: Case study of Persian Gulf. *Ocean Engineering* **2022**, *245*, 110467.
25. The Copernicus Open Access Hub. <https://scihub.copernicus.eu/>. Accessed: 2021-12-20.
26. USGS EROS Archive. <https://www.usgs.gov/centers/eros/science/usgs-eros-archive-sentinel-2>. Accessed: 2021-12-20.
27. Google Earth Engine. <https://code.earthengine.google.com/>. Accessed: 2021-12-20.
28. QGIS Development Team. *QGIS Geographic Information System*. QGIS Association, 2021.
29. Congedo, L. Semi-Automatic Classification Plugin: A Python tool for the download and processing of remote sensing images in QGIS. *Journal of Open Source Software* **2021**, *6*, 3172. doi:10.21105/joss.03172.
30. Müller-Wilm, U. Sen2Cor Configuration and User Manual. *Telespazio VEGA Deutschland GmbH* **2017**, p. 53. doi:S2-PDGS-MPC-L2A-SUM-V2.9.
31. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12*, 2825–2830.
32. Su, J.; Yi, D.; Coombes, M.; Liu, C.; Zhai, X.; McDonald-Maier, K.; Chen, W.H. Spectral analysis and mapping of blackgrass weed by leveraging machine learning and UAV multispectral imagery. *Computers and Electronics in Agriculture* **2022**, *192*, 106621.
33. Su, J.; Yi, D.; Su, B.; Mi, Z.; Liu, C.; Hu, X.; Xu, X.; Guo, L.; Chen, W.H. Aerial visual perception in smart farming: Field study of wheat yellow rust monitoring. *IEEE transactions on industrial informatics* **2020**, *17*, 2242–2249.
34. Nogueira, F. Bayesian Optimization: Open source constrained global optimization tool for Python, 2014–.
35. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015; Navab, N.; Hornegger, J.; Wells, W.M.; Frangi, A.F., Eds.; Springer International Publishing: Cham, 2015; pp. 234–241.
36. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; Desmaison, A.; Kopf, A.; Yang, E.; DeVito, Z.; Raison, M.; Tejani, A.; Chilamkurthy, S.; Steiner, B.; Fang, L.; Bai, J.; Chintala, S. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems 32*; Wallach, H.; Larochelle, H.; Beygelzimer, A.; d'Alché-Buc, F.; Fox, E.; Garnett, R., Eds.; Curran Associates, Inc., 2019; pp. 8024–8035.
37. Salomonson, V.V.; Appel, I. Estimating fractional snow cover from MODIS using the normalized difference snow index. *Remote sensing of environment* **2004**, *89*, 351–360.
38. Salomonson, V.V.; Appel, I. Development of the Aqua MODIS NDSI fractional snow cover algorithm and validation results. *IEEE Transactions on geoscience and remote sensing* **2006**, *44*, 1747–1756.