Performance analysis of a U-Net landslide detection model

ITAHISA GONZÁLEZ ÁLVAREZ[1]*, KATHRYN LEEMING[1], ALESSANDRO NOVELLINO[1], SOPHIE TAYLOR[1]

[1] British Geological Survey

British Geological Survey

INTRODUCTION

We use satellite images with labelled landslide masks from known events to train a Machine Learning algorithm to automatically identify areas where landslides have taken place. These masks are time-consuming to create, resulting in a small initial training set.

U-Nets are image segmentation algorithms, a type of classifier that assigns a label to each individual pixel in an image. These models have been shown to be of particular interest when working with small training datasets, especially when combined with data augmentation techniques.

Here, we analyse differences in the performance (defined as the ability of the algorithm to correctly predict the presence of landslides in a satellite image) of our algorithm as a result of varying inputs.

U-NET MODEL

U-Nets, a type of deep convolutional neural network (CNN), were first introduced by Ronneberger et al. (2015)[1]. They are of particular interest due to their short training times and their ability to be trained end-to-end using small training datasets. Data augmentation techniques can further help the model learn the invariance and robustness to be expected in the data.

This network consists of a contracting path, with the usual architecture of a CNN (left side, Fig. 1), and an expansive path (right side). The number of feature channels doubles at each downsampling step, and is halved again at each step of the expansive path. As a result, each pixel of the input image, which can be single, or multi-band, is assigned to a class or category.

This design won the ISBI challenge for segmentation of neuronal structures in electron microscopic stack in 2015.

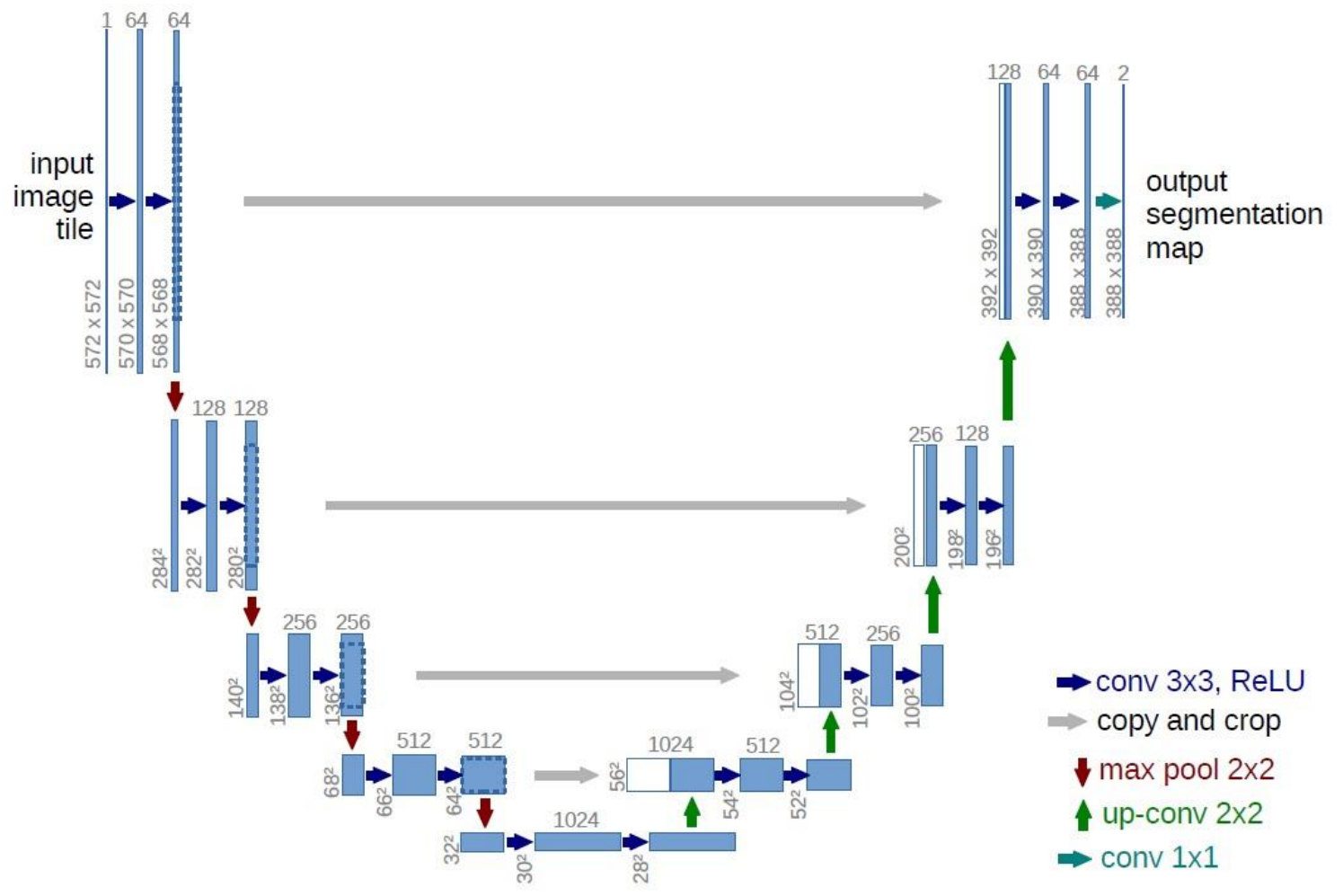


Fig. 1. U-Net architecture, as represented by Ronneberger et al. (2015). Each blue box represents a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations.

DATA

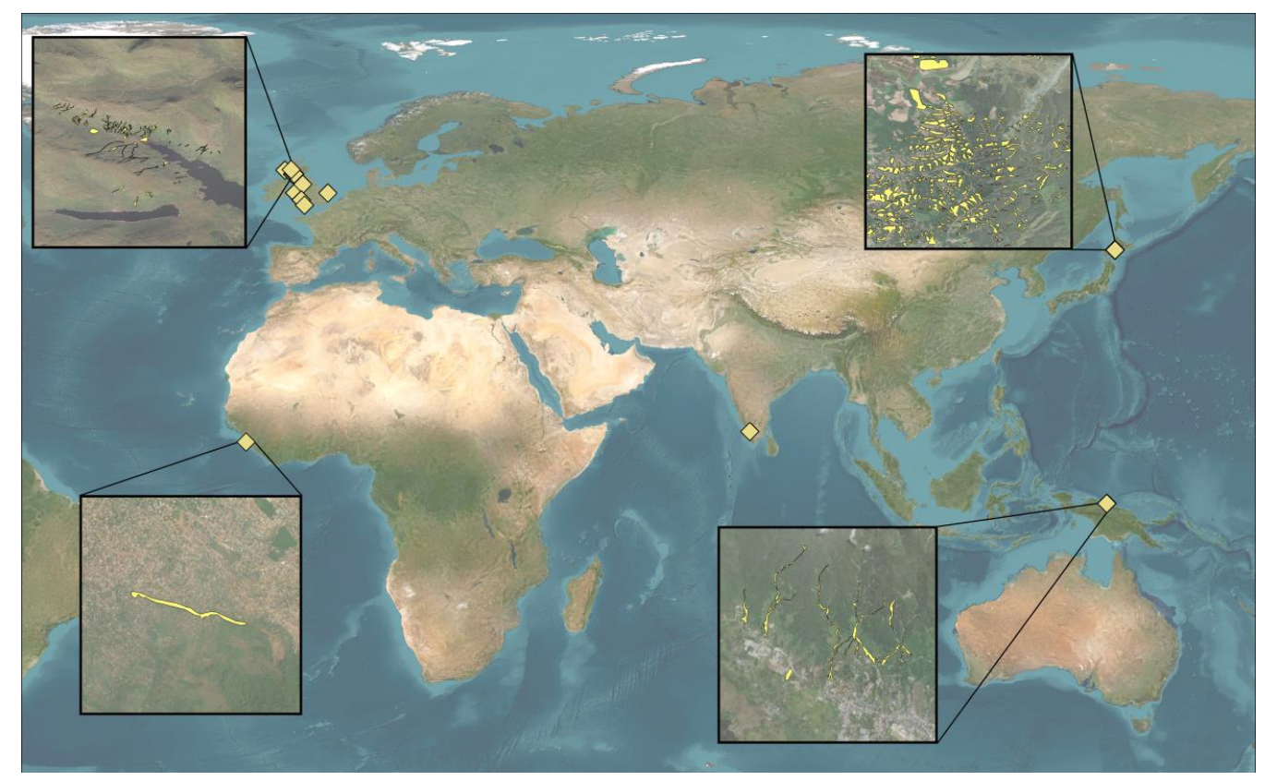


Fig. 2. Location of the mapped landslides included in our datasets. Although some locations contain a single landslide (marked in yellow for Freetown as an example, bottom left inset), datasets for most of them contain multiple landslides (rest of insets). For basemap source see [4].

We collected Sentinel 2 satellite imagery[2] from 13 locations (Figure 2) in which landslides were triggered by heavy rainfall or earthquakes. Over 700 landslides were manually mapped in these images. The polygons were then used to create the ground truth masks required to train our U-Net algorithm (Figure 3).

Our input data consists of RGB, NIR and SWIR satellite band images, as well as Digital Terrain Models (DTMs[3]). These images were then cropped into 160x160 pixels tiles. All tiles from two of our locations were completely excluded from the training/validation process, to prevent the algorithm from learning from neighbouring images.

Our dataset contains much fewer landslide pixels than non landslide ones. To address this imbalance issue, which affects the ability of the algorithm to produce accurate predictions, we introduced pixel weights. These can be all set to the same value (uniform), manually defined, or estimated based on the number of pixel types and number of pixels of each type in the dataset.

Finally, we applied random flips, rotations, translations and zooms to both satellite images and ground truth masks, as well as random variations of brightness, contrast and noise levels to satellite images (Fig. 3). Tiles with landslides were augmented more times than those without, to reduce bias towards non-landslide images and pixels.

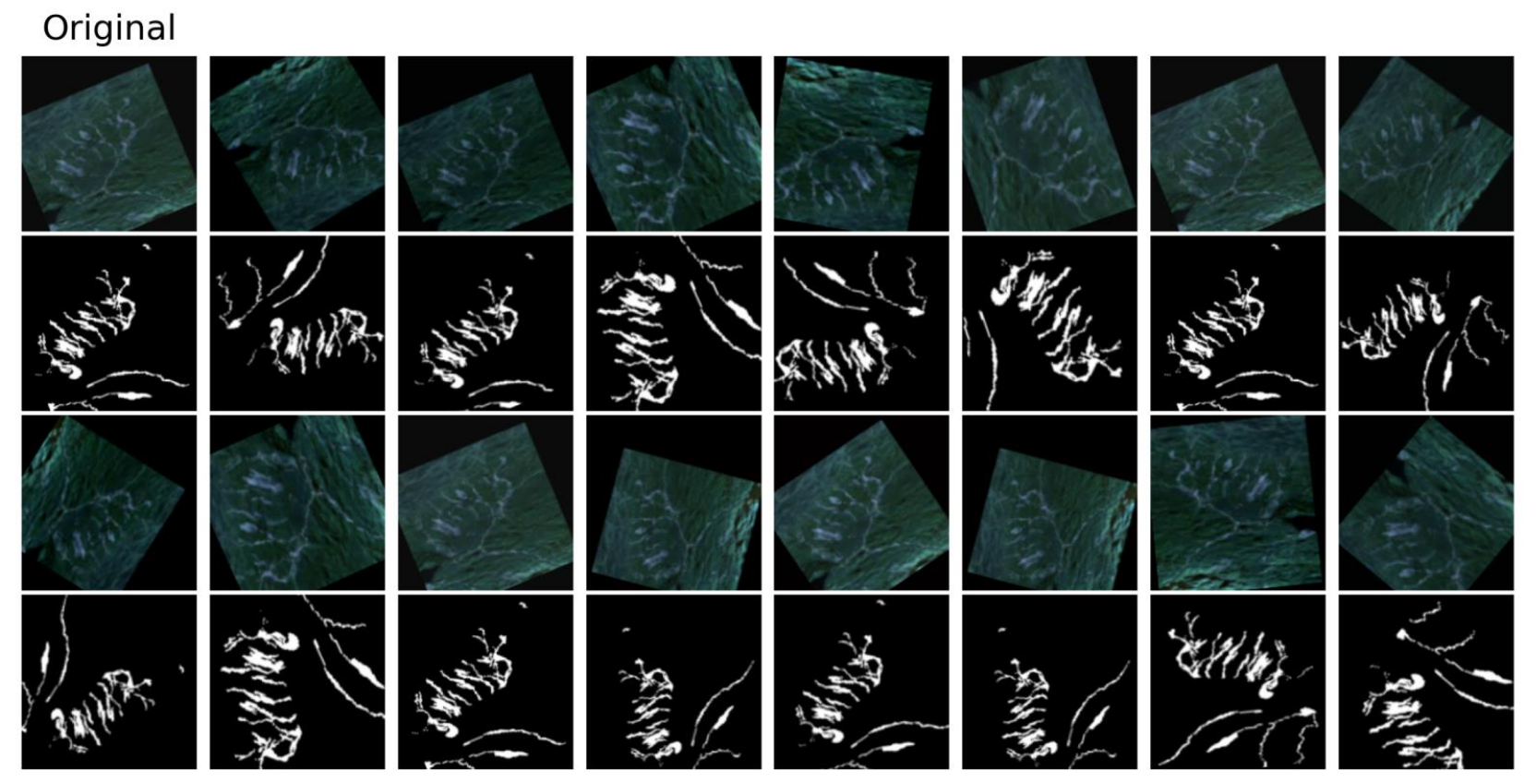
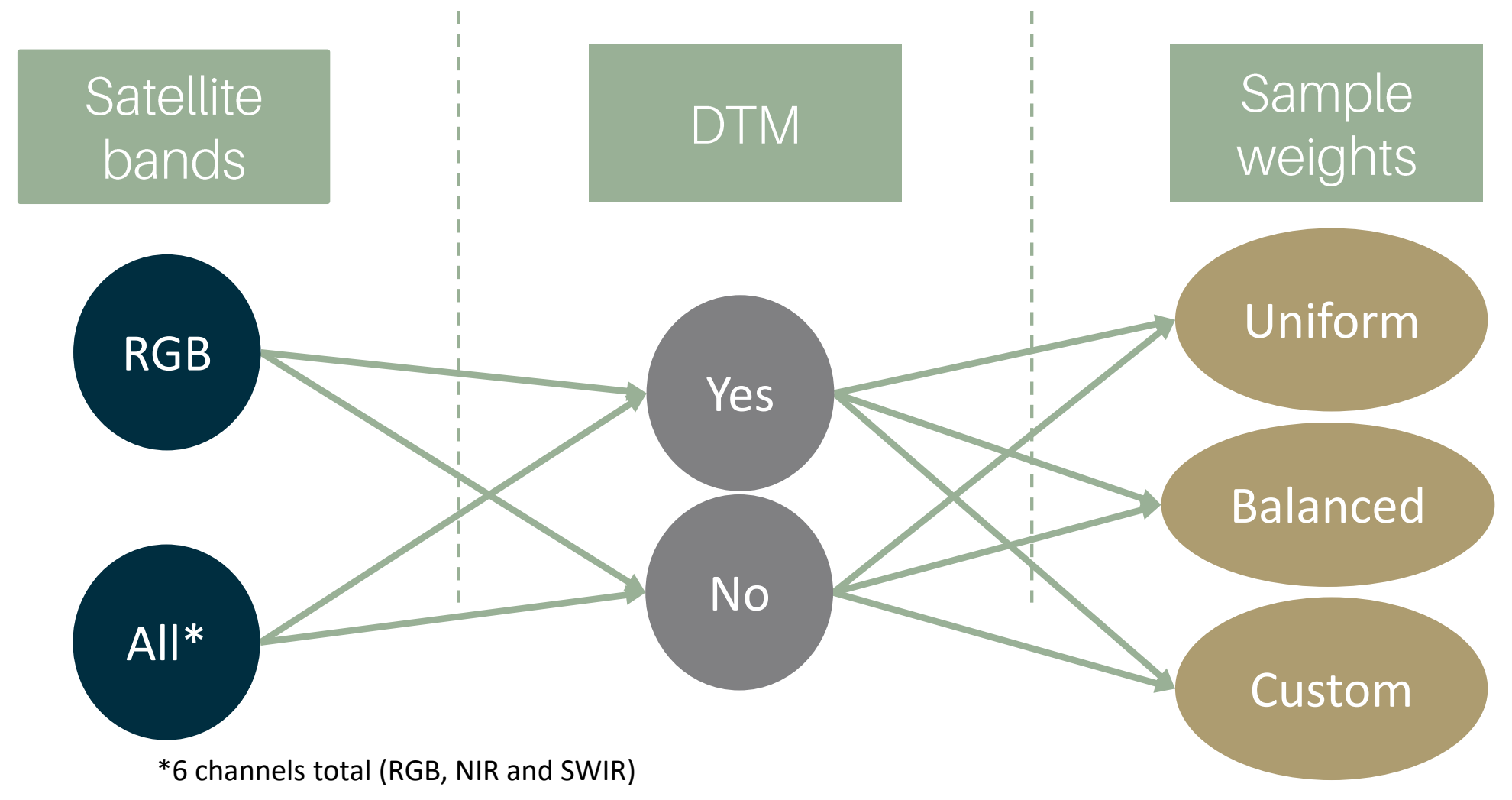


Fig. 3. Example of some data augmentations applied to one of the tiles and masks from Glengyle.

RESULTS

Input data selection

We tested the performance of our U-Net algorithm by combining different types and numbers of layers in our inputs. The diagram below illustrates how we built 16 different sets of inputs, which we then used to train and test our model.



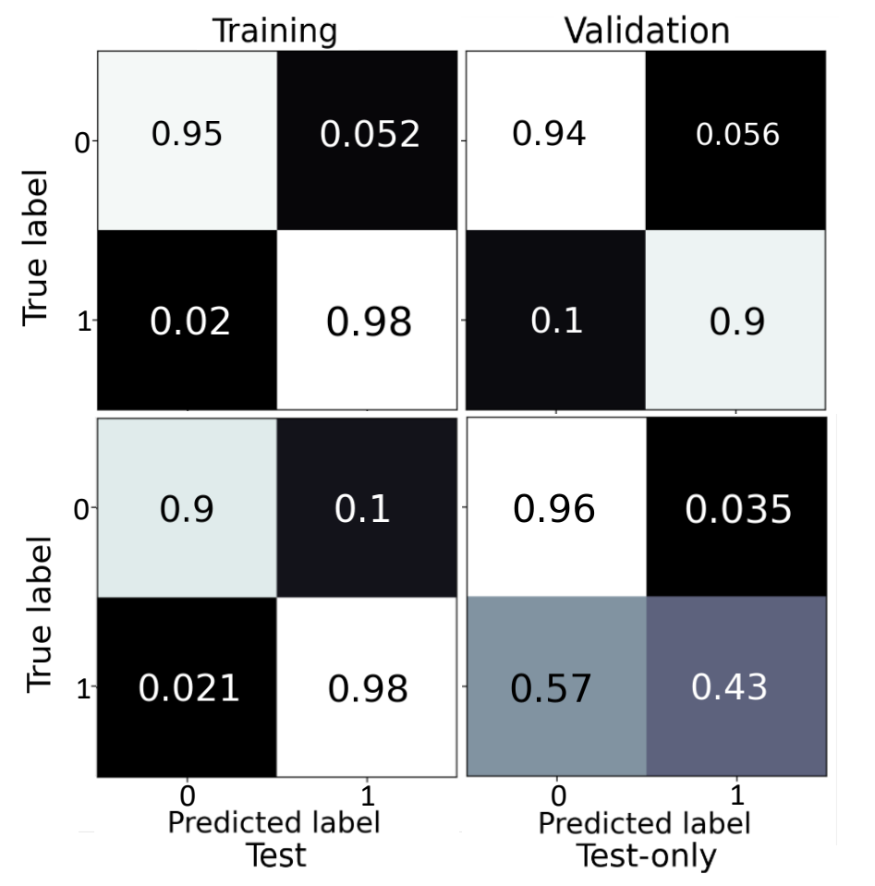
*6 channels total (RGB, NIR and SWIR)

Algorithm performance evaluation

We quantify the accuracy of the predictions by computing some of the most widely used metrics for Machine Learning algorithms:

- F1 Score (harmonic mean of the precision and recall).
- IoU (Intersection over Union, overlap between ground truth and predictions).
- Precision (number of true positives over the sum of true and false positives),
- Recall (number of true positives over the sum of true positives and false negatives)
- Accuracy (fraction of correct predictions).

We also produced the corresponding confusion matrices, which contain the proportion of pixels accurately/incorrectly assigned to either the positive (landslide) or negative (no landslide) class (Fig. 4).



Training	0	1	Validation	0	1
0	0.95	0.052	0	0.94	0.056
1	0.02	0.98	1	0.1	0.9

Test	0	1	Test-only	0	1
0	0.9	0.1	0	0.96	0.035
1	0.021	0.98	1	0.57	0.43

Fig. 4. Confusion matrices for our training, validation, test, and test-only datasets, and our best performing model (Fig. 6), which took RGB bands and DTM as inputs, and balanced weights. Top left corner is for true positives, bottom right is true negatives, top right is false negatives and bottom left true positives.

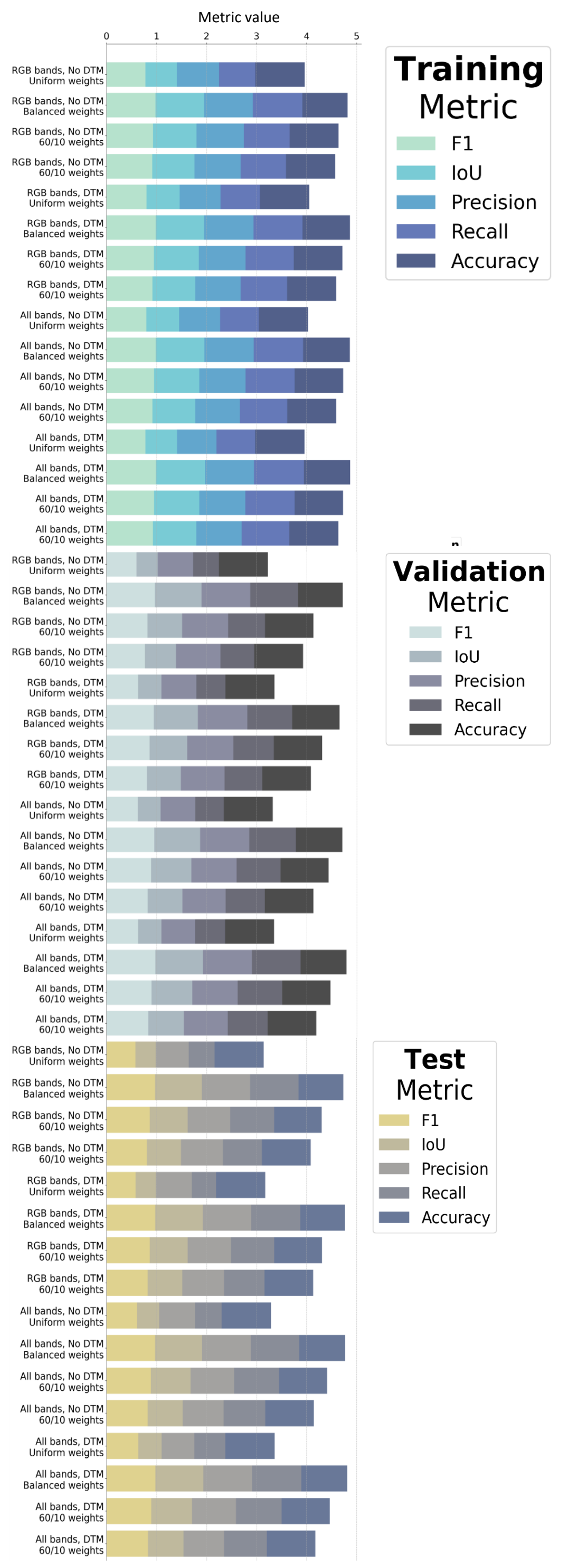


Fig. 5. Cumulative bar plots of the metrics obtained for our training, validation and testing datasets. In all cases, our model was trained for 25 epochs.

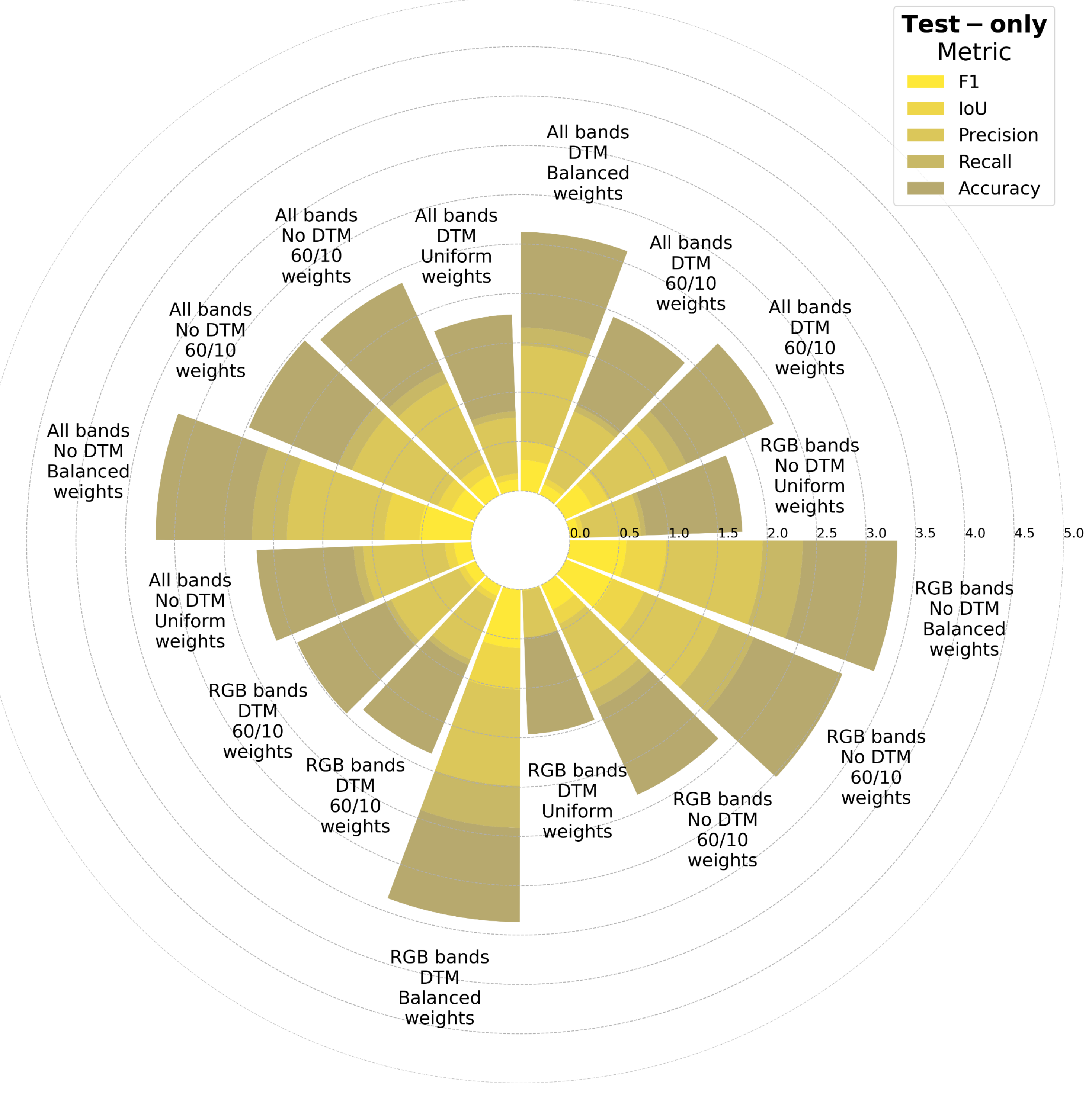


Fig. 6. Cumulative bar plot of the metrics obtained for our TEST-ONLY dataset, which includes only images from the two locations we excluded from our training, validation and test datasets. All models were trained for 25 epochs. From the inside out, these metrics correspond to the F1 Score, IoU (Intersection over Union), precision, recall and accuracy. All these metrics vary from 0 to 1, with higher values corresponding to more accurate predictions.

PREDICTIONS

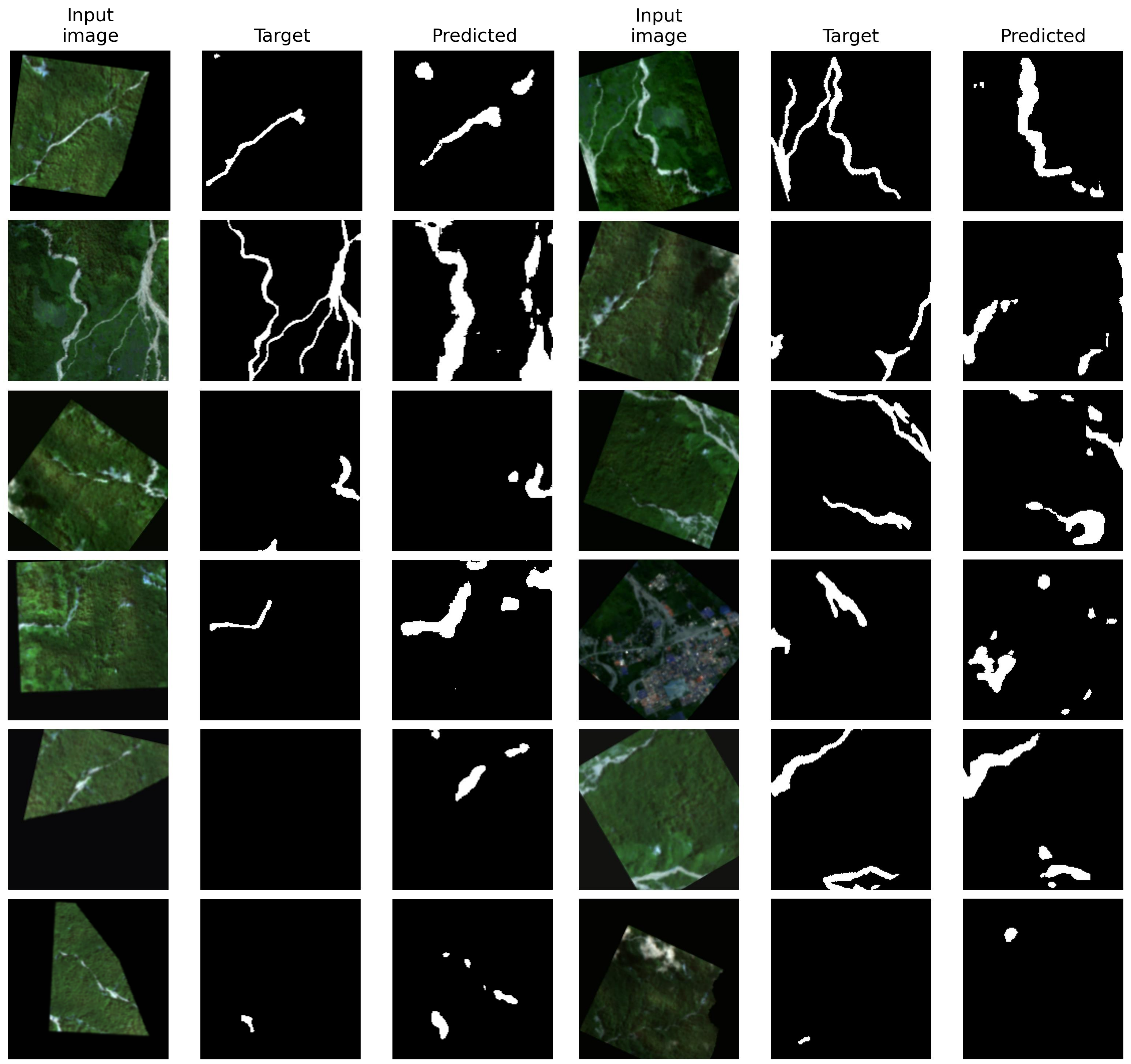


Fig. 7. Predictions obtained for a set of tiles from our TEST-ONLY dataset, which contains images from Askival (Scotland) and Sentani (Indonesia). Landslides in these locations were triggered by rainfall. Our U-Net algorithm has not seen any of these tiles, nor any of their neighbouring tiles, during the training process. Random data augmentations (flips, rotations, translations, zooms, and changes of brightness, contrast and noise levels) have been applied to both images and masks to increase the size of our dataset and help ensure the algorithm can accurately predict landslides for a wide range of settings. Left columns contain the original (augmented) tiles, central columns the ground truth, or target, and right columns the predictions obtained from our U-Net model. White colour in masks and predictions represent landslide pixels, and black marks non-landslide pixels.

FUTURE WORK

Future improvements will include:

- Increasing our dataset to include a wider variety of landscapes and regions.
- Fine tuning of the sample weights to improve the quality of the predictions.
- Increasing the resolution of the satellite imagery or digital terrain model.
- Adding geomorphological features such as slope units, DTM derivatives, different satellite imagery/information (other non RGB bands, radar, InSAR) , etc. to inputs.
- Benchmarking our algorithm against other publicly available datasets, such as the one used for the Landslide4Sense competition.

*Corresponding author: ITAHISA GONZALEZ ALVAREZ itaga@bgs.ac.uk

References:

[1] Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation: arXiv:1505.04597.
[2] Modified Copernicus Sentinel data [2023] processed by Sentinel Hub.
[3] Digital terrain model (DTM) from NASA Shuttle Radar Topography Mission (SRTM)(2013). Shuttle Radar Topography Mission (SRTM) Global. Distributed by OpenTopography. https://doi.org/10.5069/G9445JDF.).
[4] Esri, Maxar, Earthstar Geographics, and the GIS User Community

www.bgs.ac.uk