

Imperial College London
Department of Earth Science and Engineering
MSc in Environmental Data Science and Machine Learning

Independent Research Project
Final Report

Combining Landsat-8, Sentinel-1 and
NASADEM using Multi-view Learning to
Classify Water Bodies Globally

by

Kandanai Leenutaphong

Email: kandanai.leenutaphong21@imperial.ac.uk

GitHub username: kl-121

Repository: <https://github.com/ese-msc-2021/irp-kl121>

Supervisors:

Dr. Yves Plancherel

Dr. Gareth Roberts

28th August 2022

Abstract

A multi-view deep learning approach with U-Net cores is developed and used to fuse information from different satellites (Landsat-8, L8SR; Sentinel-1, S1) and a digital elevation model (NASDEM) to identify surface water bodies globally. The algorithm outperforms previously published schemes, including deep learning and threshold-based methods. The best model was produced using an experimental design consisting of 154 experiments combining different bands from Landsat-8, Sentinel-1 and NASADEM with different U-Net structures. The best algorithm can detect water bodies in hilly and high elevation terrains, such as Tibet, with a high degree of accuracy, which other schemes struggle to achieve. Importantly, because of this multi-view algorithm combines data from passive (L8SR) and active (S1) radiation sources, the overall algorithm is much less sensitive to cloud masking than schemes relying on L8SR alone. The global average F1 score is 0.94, with regional predictions of 0.95, exceeding the average feature stack of U-Net performance. To demonstrate the value of the final product, the algorithm was applied to map the extent of flooding in Thailand in 2016, demonstrating damages to built-ups, crop land and human activities. Other applications such as the flood event in Bangladesh and the drought caused by the European heatwave in 2022, are made freely available in the final product via a React-build online platform called 'WaterOracle', leveraging the Google Cloud Platform.

Acknowledgements

Firstly, I would like to thank my supervisor Dr. Yves Plancherel, whose support, guidance, and expertise led me to complete the project. In addition, I would like to thank Dr. Roberts Gareth, a hydrological expert, for giving excellent insights about water and data sources. In addition, I would also like to thank Dr. Marijan Beg for providing lots of helpful information about report writing. I would like to thank my tutor Dr. Rossella Arcucci, my family and Pinnareen, for encouraging and giving advice about time management during a stressful period. Lastly, I would also like to thank the Endian groups for giving weekly ideas and help.

1 Introduction

Water is an essential natural resource (Tyson Brown 2022), but water can also negatively impact human lives through natural hazards. The most common natural hazards caused by water include flooding and droughts, affecting three billion people from 1998-2017 (WHO 2022a,b). The creation of near-real-time water maps can aid the government and emergency service in preparing for flooding or drought events (Verbeiren et al. 2018). The flood inundated maps need the highest possible accuracy as a small error can lead to loss of lives and property insurance companies using the maps can suffer from enormous financial loss (Kousky & Kunreuther 2009).

Due to extensive global coverage and frequent acquisition of satellites, remote sensing is often used to create water extent maps (Yang et al. 2013). The most common satellite mission which are used for land and sea monitoring are Landsat-8 and Sentinel-1 (Spoto et al. 2012, Acharya & Yang 2015, Attema et al. 2009). The optical sensor in Landsat-8 can be used to detect water in any cloud-free area and the performance is less affected by the terrain (Özelkan 2020). The Synthetic Aperture Radar (SAR) in Sentinel-1 enables detection of water under the cloud, but careful terrain correction needs to be done using Digital Elevation Model (DEM) such as Shuttle Radar Topography Mission (SRTM) (Kseňák et al. 2022, Gulácsi & Kovács 2020, Song et al. 2007). Both sensors are useful to create water extent map, especially SAR due to the higher occurrence of cumulonimbus clouds during flood events (Cotton et al. 2011).

A technique to detect water using the satellite data is setting a threshold to the water indices in the multi-spectral images (Özelkan 2020) or the SAR imagery (Kseňák et al. 2022, Gulácsi & Kovács 2020). The process of selecting the threshold are affected by geographical location and the accuracy of threshold methods could be improved further. Machine learning models proved to predict water bodies more accurately than the threshold selection method. K-means clustering (Sobiech & Dierking 2013) was used to predict water bodies without terrain correction and threshold selection. Convolutional Neural Networks (CNN) are more superior to simple machine learning methods because of the ability of CNN to capture spatial information (Gu et al. 2018). Li et al. (2021) trained CNN models on Landsat-8 and Mayer et al. (2021) trained the model on Sentinel-1 and both exceeded the performance of simple machine learning and threshold methods.

Many studies train classification models using the data from one satellite (Mayer et al. 2021, Li et al. 2021), but multi-sensor data are of interest to many researchers as it can improve the ability for models to predict water bodies (Bai et al. 2021). As an example, Tang et al. (2022) combined the data from Landsat-8 and Sentinel-1 to detect water bodies with random forests, but all features were stacked so it was unclear whether combining all the features would be better than using individual features. Konapala et al. (2021) combined the different features of the Sentinel-1, Sentinel-2 and DEM models in 32 experiments to study each feature's effect and find the optimal combination of features. Konapala et al. (2021) discovered that the F1 score of using just the Sentinel-2 satellite was similar to combining Sentinel-1, Sentinel-2 and Digital Elevation Model (DEM). At this point, it is uncertain whether the multi-sensor data will give an improvement in predictions.

All attempts to train water classification models concatenate the features into one view, but the feature could be split into multiple views before training to improve the generalization performance (Zhao et al. 2017). The process of fusing features in multiple views is called multi-view learning and it has been applied to various fields to predict the future frames in videos (Xiong et al. 2021) and classifying a sleep position. (Piriyajitakonkij et al. 2020). The multi-view learning has also been applied for landcover classification fusing sentinel-2, aerial RGB image and ground perspective image (Machado et al. 2020) giving enhanced accuracy. However, no research has attempted to apply multi-view learning to the water classification problem by fusing Sentinel-1, Landsat-8 and NASADEM (Zhao et al. 2017) with U-Net. By fusing the satellites, the model is expected to best utilize the ability of Sentinel-1 to detect water under the cloud contaminated view and Landsat-8 to detect water regardless of the terrain.

Deep learning neural networks are a powerful tool to map water extent accurately, but there are two main challenges in training the model on a large scale. The main challenges include getting labeling data and the requirement of computational resources. Labeling data is time-consuming, and studies annotated data with data-labeling tools such as Label box (Labelbox 2022) or directly in Google Earth Engine (GEE). Otherwise, a more straightforward method would involve using existing labeling data. The most commonly used labeling data are the JRC Monthly Water History and the OpenStreetMap (Tang et al. 2022, JRC/Google 2022, Pekel et al. 2016, Li et al. 2021). To train deep learning models on a large scale researchers have used accessible services like Google Cloud Platform (GCP) and Google Collaboratory (GC) (Google 2022a). GCP allows users to store large geographical images and GC provides a large GPU to train deep learning models (Mutanga & Kumar 2019, Gorelick et al. 2017, Mayer et al. 2021). GCP and GC is expensive if the limit is exceeded (Google 2022c), so the data size and cost must be balanced throughout the training process.

These deep learning methods require a large amount of data to train and are often hard to reproduce. Government, emergency service interested in using the models to create water maps have to retrain complex neural network model to predict water locations. The model created can be deployed to Google AI platform to create an app for anyone to generate real-time predictions of water bodies, similar to Dynamic World by Brown et al. (2022).

By combining multi-view learning and CNN methods, we aim to invent a new method to improve the current accuracy of published work. The task is broken down into five areas. (1) Designing the architecture of multi-view learning with U-Net to combine multi-sensor data. (2) Demonstrating the development strategies and packages to permit anyone to use the new algorithm. (3) Finding the optimal combinations of Sentinel-1, Landsat-8 and NASADEM which generates the best prediction. (4) Validating the best model against difficult situations such as high elevation terrain, regions of high cloud cover, fine details and comparing results with published work. (5) Developing a web application to showcase applications of the best model.

2 Methodology

In this study, we built several U-Net models with multiple views. These models were trialed against different combinations of features from multi-sensor data similar to Konapala et al. (2021). Alongside this, NDWI was used to benchmark the new model. Once the method is finalized, we demonstrated our best model to predict a flooding event in southern Thailand from December 2016 to early 2017.

2.1 Study Area

Since we aimed to create the most accurate water prediction model, the choice of training data is important. Two training areas were investigated in Thailand and worldwide as shown in Figure 1. By comparing the performances of models trained in Thailand and globally, we hoped to learn whether training locally or globally significantly affected model performances. Therefore, we would know which training data to use with a particular flooding event.

Thailand was chosen as a study area because flooding is common due to the influence of seasonal monsoon rain. The training data in Thailand were selected randomly, from Sinakharin lake, Pasak Chonlasit lake and Sirindhorn lake, each containing 240 patches of 256 x 256 pixels as seen in Figure 1. Global training data was selected in ten countries ie. Thailand, China, Ghana, Brazil, Mexico, Pakistan, Egypt, Cambodia, India and Bangladesh, with 72 patches each. The number of samples chosen in Thailand and globally was kept at 720 in two study areas for a fair experiment. The global training data selected aimed to improve the generalization performance by including countries with water management issues, ordinary water surfaces and countries with hilly terrain.

Ten different testing data were chosen globally and the other ten were selected in Thailand as described in Figure 1. Model predictions on these test data allowed measures of the model's performance locally and globally. Selecting unseen test samples was done over the K-fold cross-validation method (Rodriguez et al. 2009) to avoid multiple training for one experiment leading to extra financial costs.

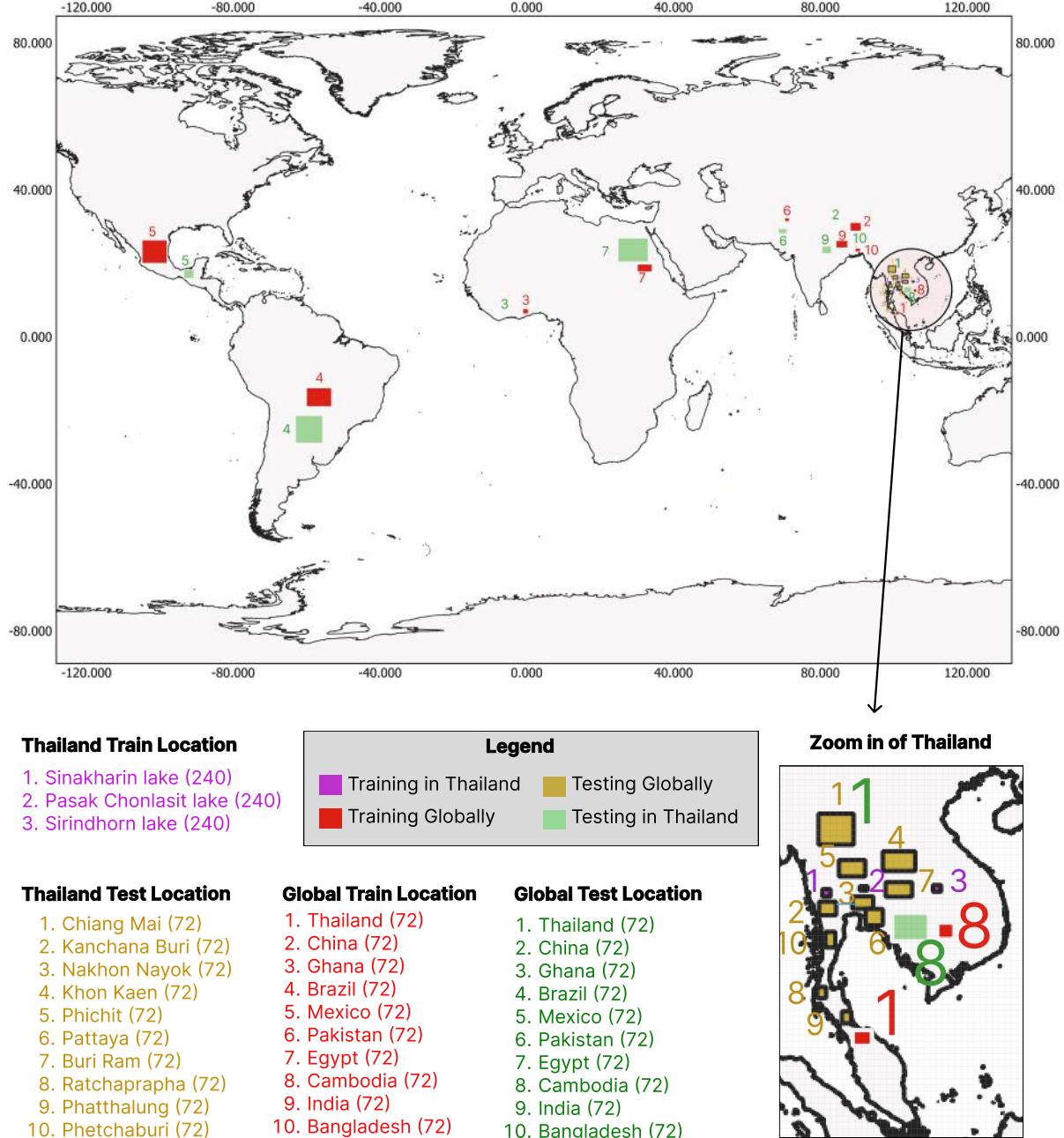


Figure 1: The location of training and testing points in Thailand and globally from January to February 2018.

2.2 Data

In this study, the labeling data we used is the JRC monthly water history data. As described briefly in Section 1, we trained the models with the Copernicus Sentinel-1 data, Landsat-8, courtesy of the U.S. Geological Survey and NASADEM digital elevation model.

The JRC monthly water history data was used because it provided extensive spatial coverage of multiple countries and the monthly data interval allowed for testing against temporal water evolution. JRC monthly data was produced by an expert and is available in the GEE repository. Some pixels in JRC data contain no-data due to shadows and clouds and these points are removed before training the models (Pekel et al. 2016). JRC monthly data is extracted from January and February 2018 to minimize cloud cover and no-data regions near South East Asia. Likewise, the training data were chosen to match the same period.

The 10m Copernicus Sentinel-1 data contained SAR and was used due to its ability to penetrate clouds as described in Section 1. The Sentinel-1 data was preprocessed using Shuttle Radar Topography Mission (STRM) 30m and ASTER data with the Sentinel-1 toolbox before being uploaded to GEE. The bands considered were the vertical transmitting with vertical receive (VV), horizontal receive (VH) and the approximate incidence angle from the ellipsoid (angle). The features were resampled to 30m to match the JRC monthly data.

Landsat-8 was used instead of Sentinel-2 because it contains more historical data than Sentinel-2 as it was launched in 2013. Landsat-8 has more spatial coverage in South East Asia than Sentinel-2 near Thailand in 2018. Using Landsat-8 avoids downscaling to match the labeling data because bands of Landsat-8 already have a 30m resolution. The Landsat-8 data is available in GEE and the training locations were chosen to minimize the cloud covers. Clouds in Landsat-8 were masked after importing the data from GEE. The bands considered were the optical bands from B2 - B7.

Terrain correction of Sentinel-1 can improve the measurement in hilly and high-elevation areas. Still, according to (Markert et al. 2020), this does not give a substantial improvement. NASADEM data was added as an additional predictor to improve the Sentinel-1 performance (Konapala et al. 2021). NASADEM is an improved elevation product created by reprocessing STRM and ASTER elevation products and is available in the GEE catalog. NASADEM provides elevation information at a 30m resolution from February 2000 (Crippen et al. 2016). The features in NASADEM include the elevation, slope and aspect.

2.3 Model Development

Three models were designed to be trained and compared for the Feature Stack U-Net abbreviated FS U-Net, Multiview learning with U-Net and two inputs (M2 U-Net) and Multi-view learning with U-Net and three main cores (M3 U-Net).

Each model was trained with ten epochs to balance good prediction with computational costs. The models used the categorical cross entropy (CC) and adam optimizer with learning rate of 0.001.

2.3.1 The Feature Stack U-Net

U-Net is one of the best image segmentation neural networks that won the ISBI cell tracking challenge in 2015 (Ronneberger et al. 2015). The structure of the U-Net in Figure 2a was adapted from Ronneberger et al. (2015) and Google (2022c), which is referred to as the Feature Stack (FS) U-Net. The input data is 256 x 256 pixels because this dimension is widely used in training convolutional network models in remote sensing (Waldner & Diakogiannis 2020). The notable difference between FS U-Net and Ronneberger (2015) is an additional max pooling and up convolution, structure to compensate for the half size of the input image (2015).

The FS U-Net architecture consists of two main parts - the downsampling and the upsampling. The downsampling section involves passing the tensor through the max pool layer, which halves the dimension of the tensor in every call, allowing the model to understand what is present in the image (Lamba 2019). The upsampling uses the convolutional block to double the dimension of the pixels, where the model gains the ability to capture where the location of classification happens. In addition, the skip connection enables each encoded information to be passed to every decoder, allowing a more

accurate prediction of the segmented image. The final layer is a softmax to determine whether a pixel is more likely to be classified as water or non-water.

However, the FS U-Net involves combining the features into one view and inputting this into the U-Net may lead to an over-fitting problem.

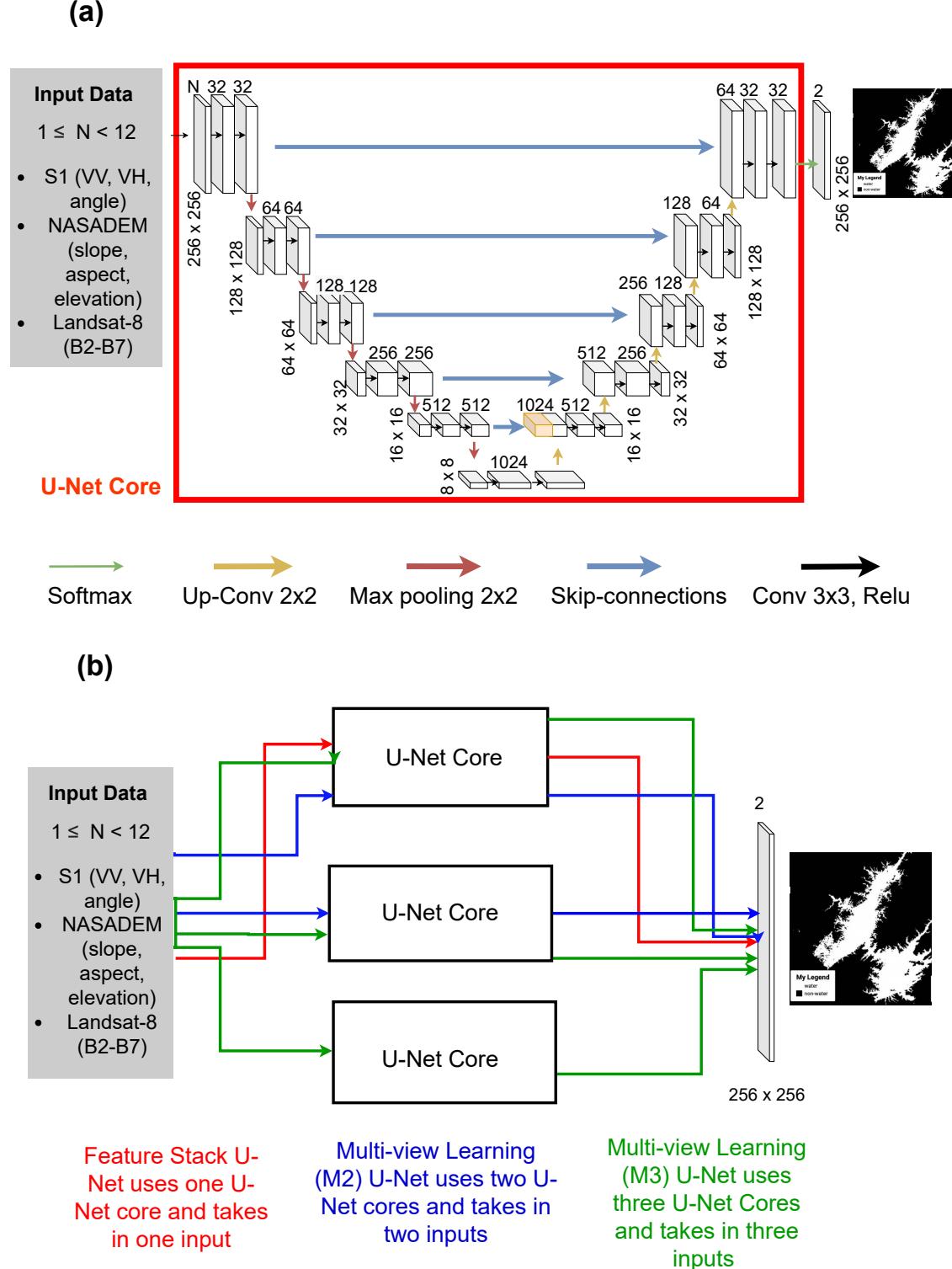


Figure 2: (a) U-Net architecture for water segmentation problem modified for taking in 256×256 pixels of input with N features. (b) Feature stack U-Net, Multi-view learning with two and three input architecture for water segmentation problems.

2.3.2 Multi-view learning and U-Net neural network classification

Multi-view learning is a data fusion process known for improving the generalization and reducing the effect of over-fitting in a smaller data. The proposed multi-view learning U-Net for water classification was modified from Piriyanitakonkij et al. (2020) using FS U-Net as the core structure, as shown in Figure 2a. Figure 2a illustrates the difference in core structures between FS U-Net, M2 U-Net and M3 U-Net. Multiple input layers go through each FS U-Net core in multi-view U-Net adopted. Then, the penultimate layers for each U-Net are concatenated. This step is crucial because the layers must be combined for the softmax layer to be able to classify water bodies. M2 U-Net is annotated with blue arrows in Figure 2b and has two inputs and FS U-Net cores. Meanwhile, M3 U-Net has three input layers and three FS U-Net cores and is annotated with green arrows.

2.4 Experimental Setup

Joining all the features from multi-sensor data may not give the best image segmentation model so different combinations of features were experimented to determine the best combination. As seen in Figure 3, the multi-view U-Net can have up to three inputs and 12 features. The features were divided into six smaller features as shown in Table 4. Each of these six features has a U-Net core and was paired either by concatenation or multiple views. 77 promising feature combinations were chosen due to the limitations in computational resources. In Section, 2.1 two locations were considered when training. The experiment that trained in Thailand is described as TH_exp and the global experiment is denoted by global_exp. The benchmark NDWI value was also established in those regions. Full experiment details are described in Appendix Table 4 and the symbols used to denote a combination are explained in Table 1. Once the best model was obtained, the model in TH_exp and Global_exp was tuned between CC and combo loss between CC and DICE (CC-DICE) (Jadon 2020).

Table 1: The symbols are used to describe the models and their connection. The model names are L8SR, S1, S1A, sl, el, as and the two connection symbols are: -, +

Symbol	Definition
L8SR	U-Net with optical bands of Landsat-8 including the B2, B3, B4, B5, B6
S1	U-Net with VV, VH bands from Sentinel-1 resampled at 30m
S1A	U-Net with VV, VH and angle bands from Sentinel-1 resampled at 30m
sl	U-Net with slope band from NASADEM
el	U-Net with elevation band from NASADEM
as	U-Net with aspect band from NASADEM
-	The feature is joined by feature stacking in the same U-Net as described in Section 2.3.1
+	The feature is joined in multiple U-Nets as described in Section 2.3.2

2.5 Establishing a baseline with NDWI

As described in Section 2.4, NDWI was used as a baseline result to compare the different variations of U-Net models. NDWI can be calculated from the Landsat-8 bands 3 and 5 and the value range varies from -1 to 1 (Özelkan 2020). A higher value indicates the presence of water and lower means vegetation.

$$\text{NDWI} = \frac{B3 - B5}{B3 + B5}$$

Ideally, a positive NDWI would indicate the presence of water, but the best threshold can vary anywhere near 0 (Ji et al. 2019). Hence, the NDWI threshold was varied from 0 to 0.2 and the average F1 score was calculated to establish a baseline.

2.6 Model Evaluation

Accuracy is often used for classification as an evaluation metric, but it does not factor into the performance of each class. High accuracy may be misleading in imbalance classes. Hence, the F1 score was used because it looks at individual classes. The F1 score is composed of the recall and precision metrics. The recall is the ability of a classifier to detect that water is present while precision measures how well the classifier does not mislabel that water is present.

These two metrics can be combined to average the information from both metrics with the F1 score. We use the macro average in Scikit-learn to give equal weights to each class (Pedregosa et al. 2011).

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

2.7 Development Strategy

The project was developed as an extension to the workflow provided by Google (2022c) and used three Google services including the GC, Google Cloud Bucket (GCB) and Google App Engine (GAE). Details and the workflow are shown in Figure 3.

GC Pro+ was used because it provides background execution for up to four concurrent sessions and a high GPU RAM, including P100, Nvidia T4 and V100 (Google 2022b). Eleven notebooks were used to utilize the concurrency ability of GC Pro+, allowing multiple sessions to run simultaneously independent of the other notebooks. TrainingModels(_global).ipynb and metrics_assessment(_global).ipynb used the GPU and background execution to make the training and evaluation of 154 models more efficient. The remaining notebooks use a standard CPU to stay within the computational limits of GC Pro+.

The GCB was used to store training, testing data and the weights of the models (Google 2022a). Training the models has led to unavoidable storage costs of averagely 300MB-1GB per experiment even when the training patches was reduced to 720 compared to the default 2000 patches in Google (2022c). The GCB service was used as it provides the ability to export predicted images to GEE to create an interactive visualizer using GAE. GAE apps were created and connected with the MERN stack (Mongo, Express, React, and Node.js) to build a full-stack web application (Subramanian 2017) to showcase the application of our deep learning models.

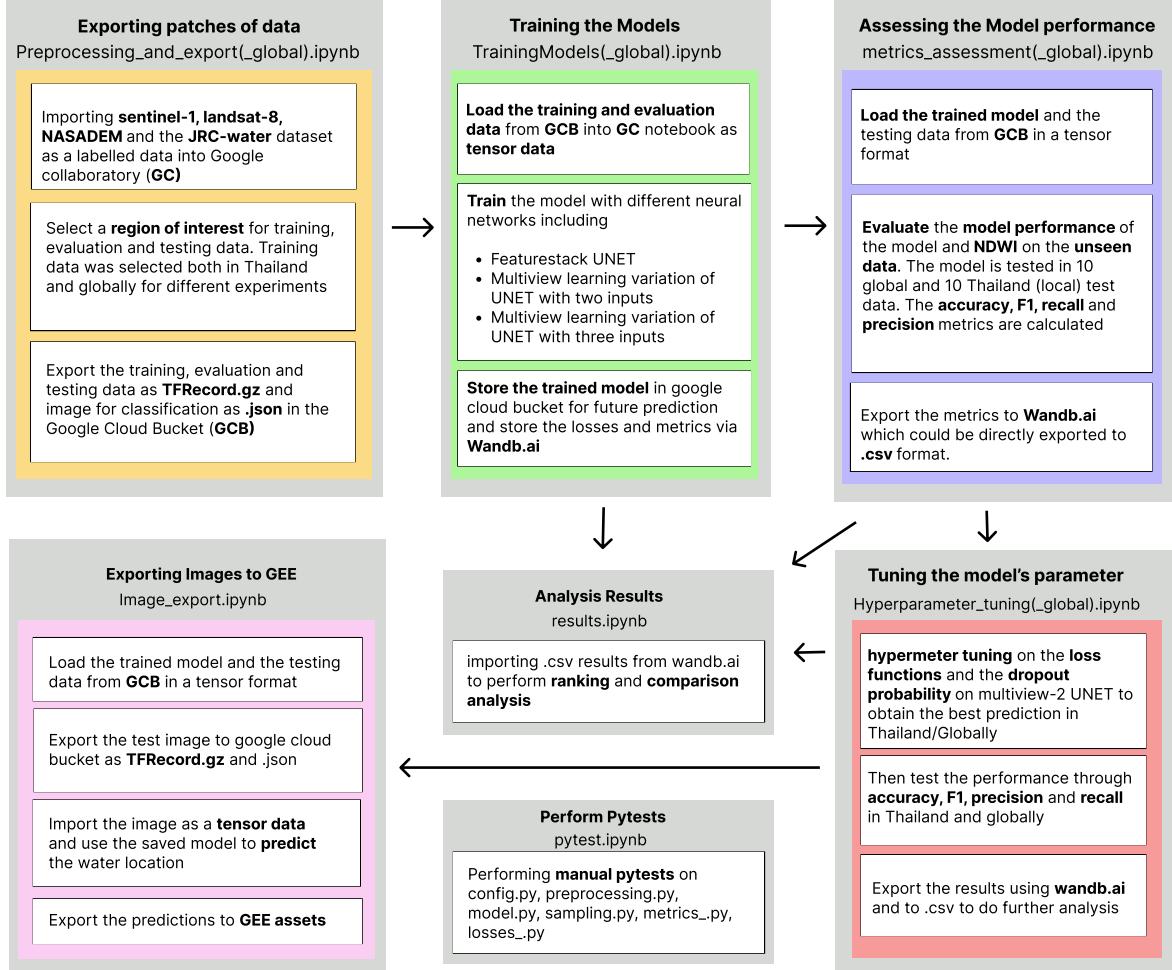


Figure 3: The development methodology using 11 GC notebooks. The experiments rely on training locally and globally so the notebook is split into _Thailand and _global, respectively

2.8 Code metadata

The experiments described in Section 2.4 relied on TensorFlow (2.8.2) to generate variations of the U-Net models and earthengine-api (0.1.317) to preprocess the dataset conveniently. To keep track of all the experiments Wandb (0.12.21) and pandas (1.3.5) provided easy integration with an online database to store model results in case the connection got interrupted. All the tasks were made more convenient through the custom-made ‘tools’ packages in python (3.7.13) GC with integrated testing using Sphinx (5.1.1) and Pytest (3.6.4).

The main functions are inside the tools folder, while the brief functions of each script are:

- **config.py** : Track and store variables belonging to each experiments.
- **metrics_.py** : Evaluation of the F1 score, recall and precision on models with Sklearn (0.0) and Numpy (1.21.6).
- **model.py**: Initializes the FS, M2, M3 U-Net models.
- **preprocessing.py** : Preprocessing Landsat-8 data and importing tfrecord.gz from GCB into GC.
- **sampling.py** : Exporting training, evaluation and testing data from GC to GCB as tfrecord.gz

- **Losses.py** : Initialize custom loss functions for hyperparameter tuning.
- **images.py** : Generating predictions as images from GC to GEE

Further instructions are in the README file in the GitHub repository accessed at <https://github.com/ese-msc-2021/irp-kl121/> and the details about the functions of the custom package is provided in waterclassification.pdf in the docs folder.

3 Results

The experiments carried out in this study aim to address two key aspects of the research. The first is concerned with finding the best combination between the features described in 2.4 and comparing the results to other published methods. The second explores and assesses the application of the model performance in different situations.

3.1 NDWI baseline comparison

As mentioned in Section 2.5, the NDWI is used as a benchmark, and the thresholds were 0, 0.1 and 0.2. The NDWI was applied to the ten global testing samples as shown in 1. The respective F1 scores are 0.90, 0.89, 0.86 and the average of 0.88 is used as the baseline.

3.2 FS U-Net results

The FS U-Net combinations in TH_exp and Global_exp were performed on 32 different models and the results are in Table 2. The most robust model for the FS U-Net is L8SR because regardless of training or testing in Thailand or globally, L8SR gives an F1 score varying from 0.90 to 0.92 which exceeds the baseline result. This coincides with the findings in Konapala et al. (2021) that L8SR gives a better prediction than S1 with a single view.

At this point, there seems to be little benefit to adding S1 or elevation bands to the model as L8SR_S1A.sl has worse predictions than L8SR which is similar to the findings in (Konapala et al. 2021). Nevertheless, the best combination must include S1 because S1 can penetrate clouds, which will be explored more in the multi-view results. Table 2 shows that the F1 score of S1 is improved by adding slope from 0.62-0.84 to 0.8-0.9. We refer to S1_sl or S1A_sl as the SentSlope. The finding that slope pairs well with S1 than elevation also coincide with Song et al. (2007). The choice of training data does not seem to affect the performance of L8SR, but it affects SentSlope. The performance of SentSlope in Thailand is improved by training in Thailand compared to globally as shown in 10FS in Table 2. This may illustrate that terrain information is important for SentSlope and to give SentSlope optimal results, the models should be trained in the same geographical area as where it is predicted.

Table 2: The feature Stack Results for TH_EXP and Global_EXP.

Experiment ID	model name	F1_ev ¹	F1_ev ²	F1_loc ³	F1_loc ⁴
		TH	Global	TH	Global
1FS	S1A_el_sl_as	0.58	0.66	0.69	0.73
2FS	S1A_el	0.66	0.62	0.85	0.64
3FS	S1A_sl	0.78	0.48	0.92	0.48
4FS	S1A_as	0.75	0.78	0.80	0.81
5FS	S1A_sl_as	0.69	0.77	0.78	0.79
6FS	S1A_el_sl	0.67	0.70	0.81	0.84
7FS	S1A_el_as	0.60	0.70	0.77	0.79
8FS	S1_el_sl_as	0.59	0.69	0.70	0.77
9FS	S1_el	0.54	0.63	0.70	0.65
10FS	S1_sl	0.80	0.81	0.90	0.86
11FS	S1_as	0.77	0.77	0.82	0.81
12FS	S1_sl_as	0.78	0.77	0.81	0.79
13FS	S1_el_sl	0.56	0.64	0.70	0.63
14FS	S1_el_as	0.59	0.59	0.70	0.61
15FS	L8SR_el_sl_as	0.61	0.67	0.71	0.62
16FS	L8SR_el	0.49	0.50	0.49	0.54
17FS	L8SR_sl	0.79	0.85	0.79	0.83
18FS	L8SR_as	0.78	0.77	0.79	0.77
19FS	L8SR_sl_as	0.79	0.77	0.81	0.77
20FS	L8SR_el_sl	0.62	0.50	0.67	0.55
21FS	L8SR_el_as	0.55	0.70	0.67	0.70
22FS	L8SR_S1_el	0.56	0.67	0.67	0.70
23FS	L8SR_S1_sl	0.80	0.89	0.84	0.88
24FS	L8SR_S1_sl_el_as	0.64	0.68	0.72	0.76
25FS	L8SR_S1A_el	0.54	0.65	0.73	0.71
26FS	L8SR_S1A_sl	0.86	0.89	0.91	0.88
27FS	L8SR_S1A_sl_el_as	0.59	0.70	0.71	0.78
28FS	L8SR	0.90	0.91	0.92	0.92
29FS	S1	0.75	0.62	0.84	0.68
30FS	S1A	0.78	0.49	0.89	0.50
31FS	L8SR_S1	0.57	0.90	0.57	0.88
32FS	L8SR_S1A	0.83	0.88	0.91	0.90

¹ F1_ev TH is calculated when the model is trained in Thailand and tested globally.

² F1_ev Global is calculated when the model is trained globally and tested globally.

³ F1_loc TH is calculated when the model is trained and tested in Thailand.

⁴ F1_loc Global is calculated when the model is trained globally but tested in Thailand.

3.3 M2 and M3 Multi-view results

The multi-view U-Net combinations in TH_exp and Global_exp were performed on 45 different models and the results are in Table 3. The Krushal-wallis test was applied on the combinations with the same feature but with different connection in Table 2 and 3. The statistic test showed that multi-view learning gives a higher F1 score than U-Net with one view (Feir & Toothaker 1974). Notably, the features of Landsat-8 joined in multiple views with SentSlope data was superior to stacking the features.

Table 3: The results for M2 and M3 U-Net in TH_EXP and Global_EXP

Experiment ID	Model name	F1_ev ¹	F1_ev ²	F1_loc ³	F1_loc ⁴
		TH	Global	TH	Global
1M2	S1A+(el_sl_as)	0.76	0.74	0.91	0.82
2M2	S1A+el	0.72	0.59	0.91	0.62
3M2	S1A+sl	0.78	0.76	0.90	0.80
4M2	S1A+as	0.77	0.76	0.97	0.95
5M2	S1A+(sl_as)	0.79	0.76	0.90	0.78
6M2	S1A+(el_sl)	0.73	0.68	0.91	0.74
7M2	S1A+(el_as)	0.74	0.76	0.90	0.87
8M2	S1+(el_sl_as)	0.70	0.69	0.86	0.73
9M2	S1+el	0.71	0.49	0.82	0.49
10M2	S1+sl	0.78	0.71	0.85	0.77
11M2	S1+as	0.81	0.80	0.87	0.85
12M2	S1+(sl_as)	0.81	0.79	0.87	0.85
13M2	S1+(el_sl)	0.70	0.50	0.81	0.52
14M2	S1+(el_as)	0.69	0.76	0.80	0.78
15M2	L8SR+(el_sl_as)	0.83	0.90	0.82	0.90
16M2	L8SR+el	0.87	0.92	0.90	0.91
17M2	L8SR+sl	0.90	0.91	0.91	0.92
18M2	L8SR+as	0.89	0.93	0.93	0.91
19M2	L8SR+(sl_as)	0.86	0.91	0.87	0.91
20M2	L8SR+(el_sl)	0.87	0.90	0.92	0.89
21M2	L8SR+(el_as)	0.85	0.90	0.91	0.90
22M2	L8SR+(S1_as)	0.92	0.90	0.93	0.89
23M2	L8SR+(S1_el)	0.82	0.92	0.86	0.91
24M2	L8SR+(S1_sl)	0.92	0.93	0.94	0.92
25M2	L8SR+(S1_sl_as)	0.90	0.86	0.90	0.84
26M2	L8SR+(S1_el_sl)	0.80	0.92	0.90	0.90
27M2	L8SR+(S1_el_as)	0.82	0.91	0.89	0.90
28M2	L8SR+(S1_sl_el_as)	0.79	0.90	0.84	0.89
29M2	L8SR+(S1A_as)	0.90	0.87	0.91	0.85
30M2	L8SR+(S1A_el)	0.85	0.91	0.86	0.90
31M2	L8SR+(S1A_sl)	0.93	0.93	0.95	0.93
32M2	L8SR+(S1A_sl_as)	0.87	0.92	0.93	0.92
33M2	L8SR+(S1A_el_sl)	0.79	0.88	0.81	0.88
34M2	L8SR+(S1A_el_as)	0.83	0.91	0.87	0.88
35M2	L8SR+(S1A_sl_el_as)	0.79	0.87	0.84	0.88
36M2	L8SR+S1	0.91	0.88	0.94	0.89
37M2	L8SR+S1A	0.91	0.93	0.92	0.92
1M3	L8SR+S1+as	0.87	0.92	0.93	0.89
2M3	L8SR+S1+el	0.61	0.93	0.70	0.92
3M3	L8SR+S1+sl	0.92	0.93	0.95	0.91
4M3	L8SR+S1+(sl_el_as)	0.85	0.93	0.94	0.92
5M3	L8SR+S1A+as	0.91	0.93	0.92	0.92
6M3	L8SR+S1A+el	0.63	0.93	0.80	0.90
7M3	L8SR+S1A+sl	0.93	0.93	0.94	0.92
8M3	L8SR+S1A+(sl_el_as)	0.93	0.93	0.92	0.92

¹ F1_ev TH is calculated when the model is trained in Thailand and tested globally.

² F1_ev Global is calculated when the model is trained globally and tested globally.

³ F1_loc TH is calculated when the model is trained and tested in Thailand.

⁴ F1_loc Global is calculated when the model is trained globally but tested in Thailand.

Among all the multi-view combinations, the most robust models are L8SR + (S1A_sl) and L8SR + S1A + sl, reaching an F1 score of 0.92-0.95. The performance of the L8SR + (S1A_sl) model is improved when tested in the same geographical area as trained which is similar to the behaviour of SentSlope in Section 3.2. This model structure exceeds the L8SR performance not only with the F1 score but by spatial coverage. In regions with cloud masking, L8SR cannot be used. The L8SR + S1A + sl model or L8SR + (S1A_sl) can be used and more emphasis is put on the SentSlope, allowing improved prediction in cloudy areas.

Finally, we can conclude that L8SR + (S1A_sl) is a better model than L8SR + S1A + sl because L8SR + (S1A_sl) is a simpler model by the principle of parsimony (Vandekerckhove et al. 2015). This model is named the WaterNet.

3.4 Loss function tuning results

The WaterNet was tuned between CC and CC-DICE loss function. The CC-DICE loss function for model trained globally give a higher F1 score than CC. For predictions in Thailand we will use CC loss function with F1 score of 0.95 and for global predictions we will use CC-DICE with F1 score of 0.94

3.5 Model Application

The results from the previous section demonstrate that the WaterNet can reliably locate surface water activity from Landsat-8, Sentinel-1 and NASADEM. The WaterNet is tested under challenging situations: high-elevation mountain, flood regions with a high cloud cover percentage and coastline detection.

3.5.1 High elevation terrain

The model was validated against Tibet at high elevation to test S1 for a similar backscattering coefficient. The WaterNet model used was trained globally because Tibet is outside of Thailand. Five different combinations of models including S1, S1_sl, L8SR, L8SR+(S1_sl) and L8SR + S1 + sl are compared in Figure 4.

S1 was the worst performing failing to detect a lake in 4 (h) which is expected due to the similar backscattering effect. When slope was added the largest lake was detected and the middle pond became correctly labeled. This illustrates that adding slope to S1 can improve water detection in high elevation areas. L8SR mispredicts a few points in the middle lake in some areas of L8SR as seen in the diagram, but L8SR can capture smaller lakes better than S1_sl 4 (j). The multi-view combination between L8SR, S1 and slope captures most water details including the cloudy region illustrating that L8SR + (S1_sl) and L8SR + S1 + sl in 4 (k) and (l) is the dominant model.

One concern is whether the no-data part in Figure 4 (a) is correctly predicted. Brown et al. (2022)'s landcover image was studied from April to July 2018 in 4 (b) and it turns out that the no-data part represents ice. The result from our model is consistent with the Dynamic World prediction. We conclude that WaterNet performs well in difficult terrain and can distinguish between water and ice.

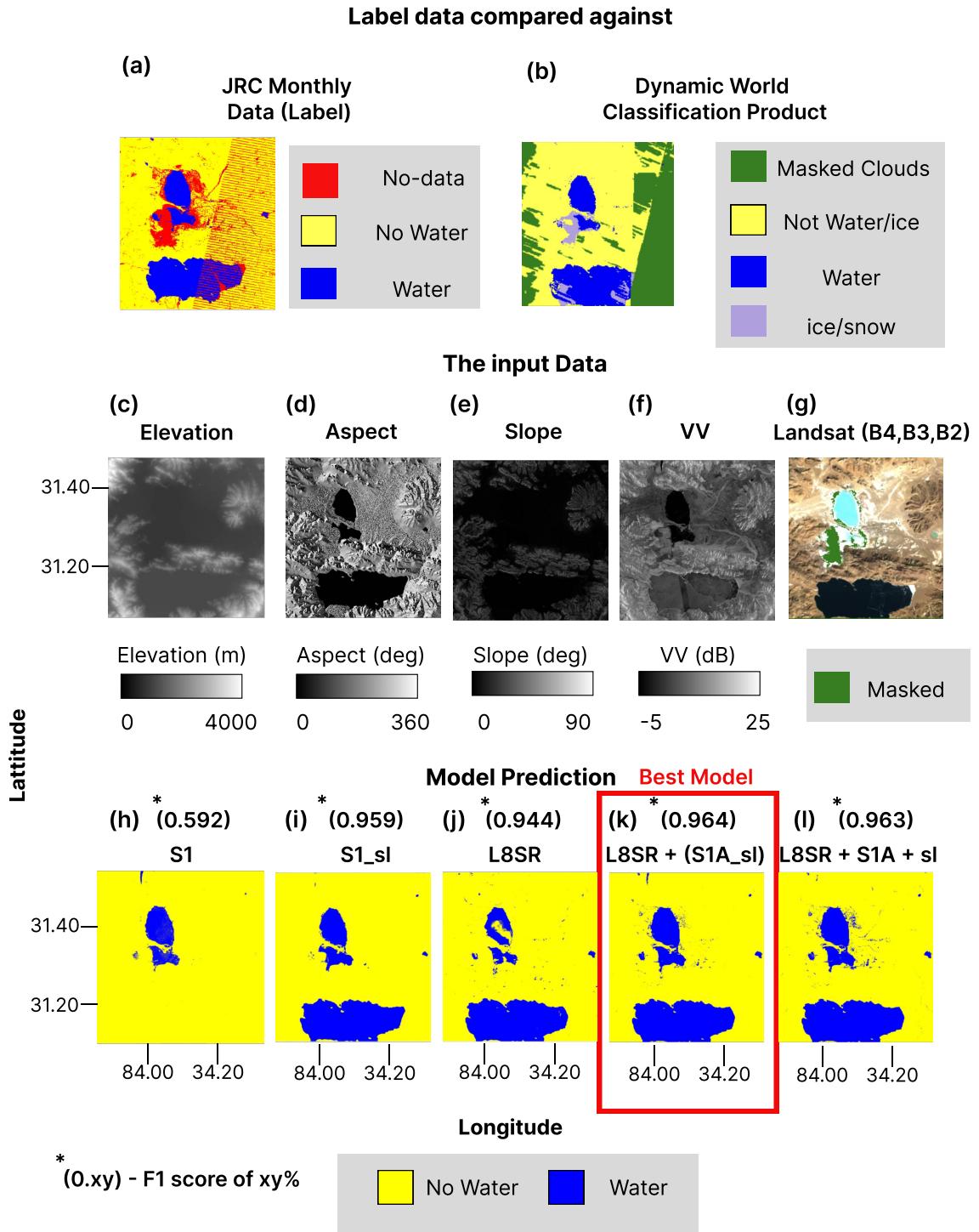


Figure 4: The verification, input and prediction images of S1, S1_sl, L8SR, L8SR+(S1_sl) and L8SR + S1 + sl. The experiment is in Tibet from January to February 2018

3.5.2 Spatio-temporal performance of WaterNet

As described earlier, the flood inundated areas in southern Thailand are mapped monthly from November 2016 to April 2017 in Figure 5. The performance of WaterNet is compared with L8SR + S1A + sl M3 U-Net, L8SR U-Net, Sl_sl U-Net and NDWI all trained in Thailand.

The first column shows the cloud NDWI imagery of the scene. From November 2016 to February 2017, the clouds covered most of the areas, and the NDWI could not give a good indication of water location, contrary to the high F1 score. This is because the F1 value for NDWI was calculated using the unmasked scenes. The second column is the JRC label data. The column L8SR + (S1A_sl) gives the best prediction indicating that the WaterNet gives robust results regardless of cloud cover. The L8SR + S1A + sl gave similar results to WaterNet. The remaining columns are other models. Notably, the L8SR model performs well in the cloud-free period and the SentSlope model performs worse than L8SR which is consistent with the results in Section 3.2.

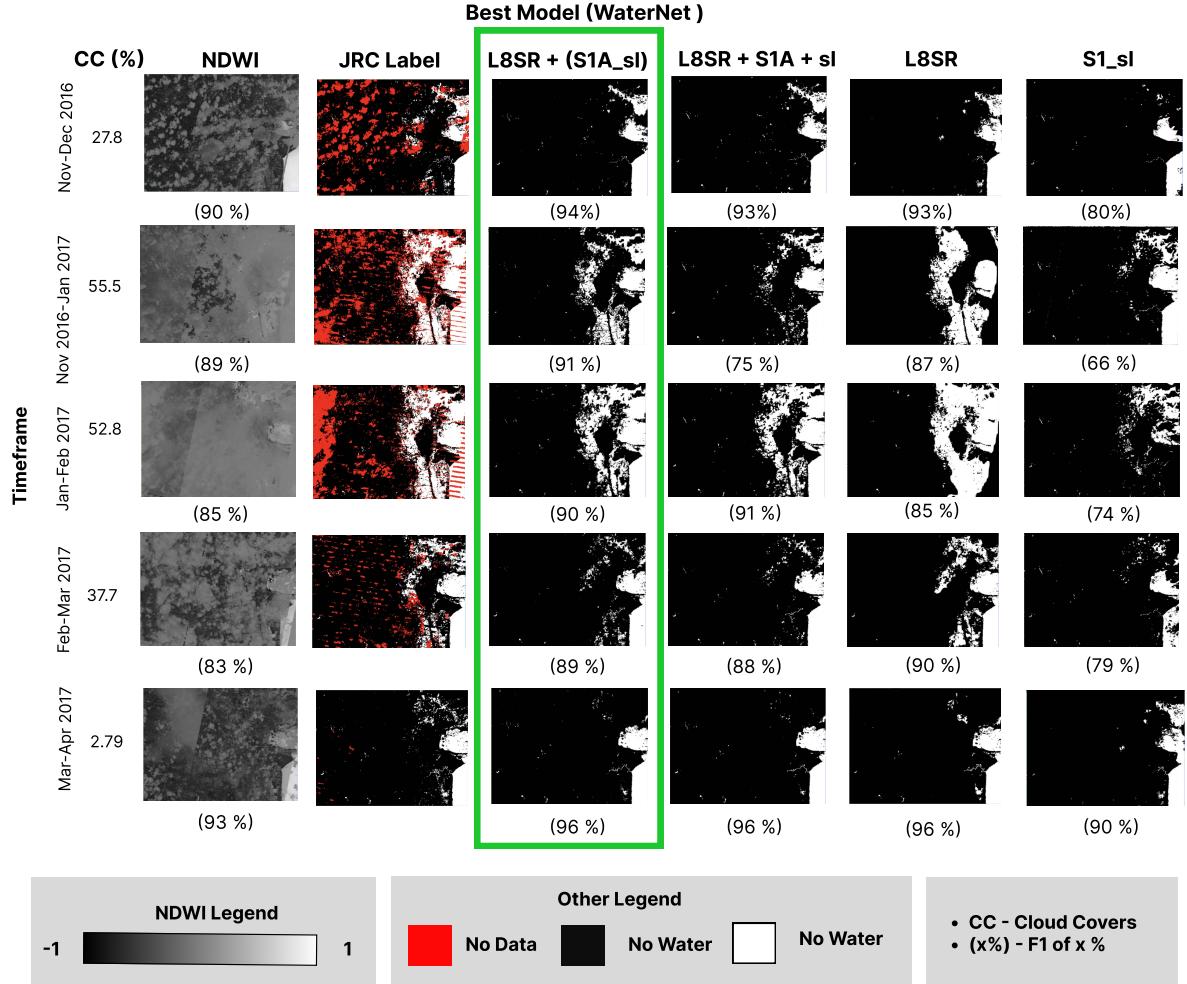


Figure 5: Comparison of the WaterNet, S1_sl, L8SR, L8SR + S1 + sl and NDWI in the southern part of Thailand (Nakhorn Si Thammarat) from December 2016 to early 2017.

3.5.3 Cloud Cover Analysis

More points are sampled monthly in the same region as Section 3.5.2 to understand the effect of cloud cover on the models. The t-test tests the relationship between cloud cover and the F1 score.

Null Hypothesis: The gradient fitting the model is 0

Alternative Hypothesis: The gradient fitting the model is not 0

Figure 6 shows S1_sl is the most robust to cloud cover with a p-value of 73 %, accepting the null hypothesis, hence, insensitive to cloud cover. However, all the other models have a p-value lower than 5 % so they are all affected by cloud cover statistically. However, the L8SR + (S1A_sl) model

has a much lower gradient magnitude than L8SR or NDWI, indicating its improvement in detecting under clouds.

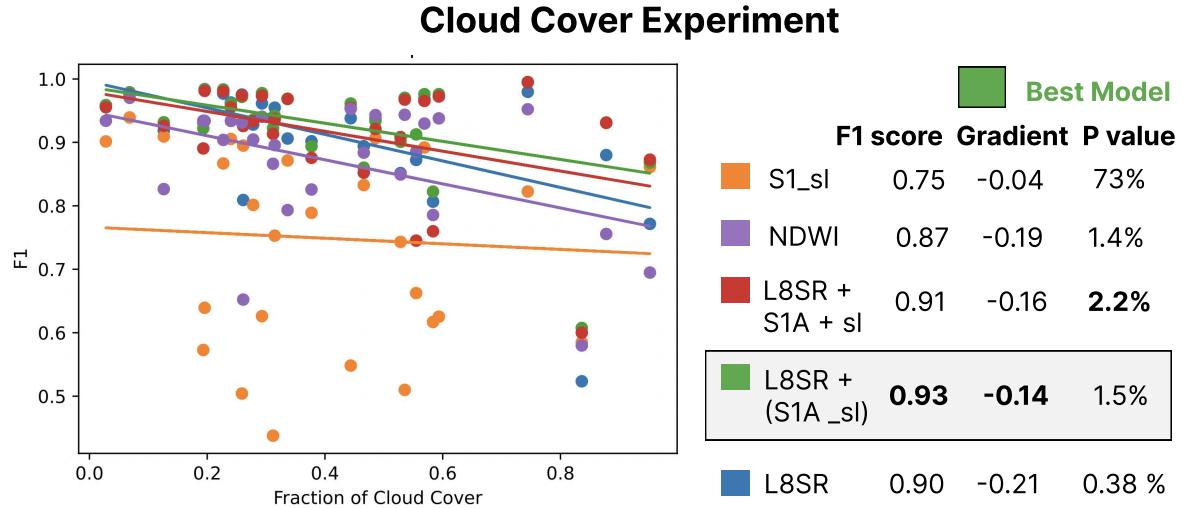


Figure 6: Comparing the ability of the different methods to predicting under cloud covers from the flood region in Section 3.5.2 in the 28 months from November 2016 to March 2019

3.6 Coastline detection

The ability WaterNet to detect small water details is studied by assessing the predictive ability of WaterNet in coastline areas such as Iceland and Israel are studied in Figure 7. Catherine Seale (2019) released a benchmark dataset for coastline data created by hand labeling the 10m Sentinel-2 data in the cloud-free period during 2019 called the Sentinel-2 Water Edges Dataset (SWED). The original SWED at 10m is shown in the first row of Figure 7 before resampling to 30m in the second row. The SWED is resampled to give a fair comparison to our 30m WaterNet predicted over the whole year in 2019 in the last column. The JRC monthly labeling data was chosen in a month with minimal no-data in 2019 to verify the WaterNet performance. The WaterNet shown in the last row cannot capture small details at 10m, but captures most details at 30m in SWED and JRC Monthly labels. With a higher resolution labeling data we can expect WaterNet to capture finer water bodies detail.

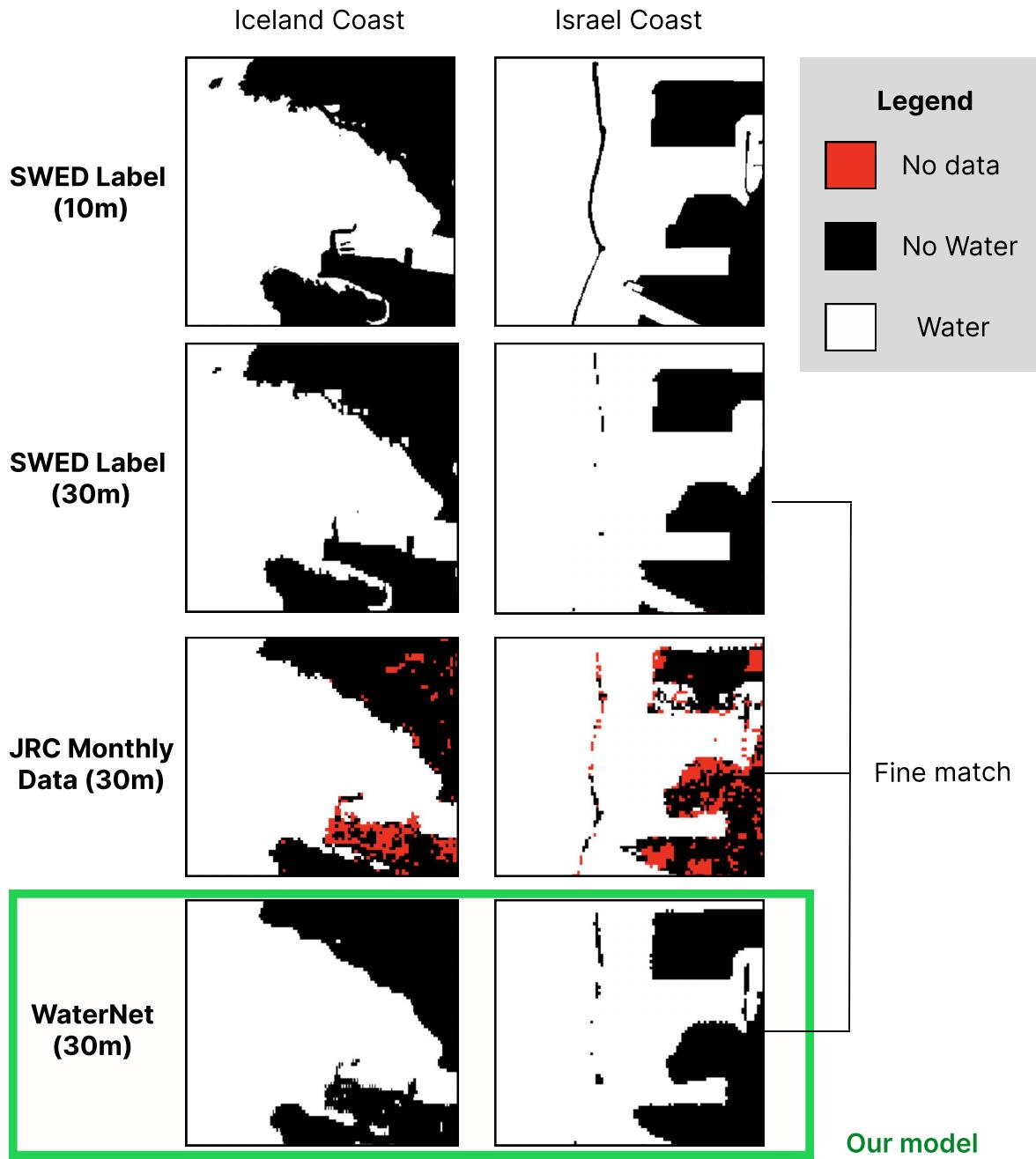


Figure 7: Coastline detection in Iceland and Israel comparing the SWED at 30m and 10m, JRC monthly data and WaterNet in 2019

4 Discussion

4.1 Multi-view learning using Sentinel-1, Landsat-8 and NASADEM

The results from the series of accuracy assessments show that joining the features through multiple views give a better prediction than joining features through a single view from Section 3.3. Among the multi-view predictions, we discover the WaterNet, a combination between L8SR and SentSlope, to be the most effective model for detecting water bodies. The proposed WaterNet outperforms U-Net, the baseline NDWI and exceeds all the performance in Konapala et al. (2021). The WaterNet

is less affected by cloud cover because when SentSlope is added as a multiple view, SentSlope can compensate for the cloud masking regions as seen in Figure 5 and 6. Because of this, WaterNet has a high average F1 score of 0.94-0.95 (Table 2, Table 3).

4.2 Applications for WaterNet and reusing the model

Aside from WaterNet being applied to Southern Thailand as described in Section 3.5.2, the WaterNet was also applied to map the recent Bangladesh flood in 2022 and the drought events caused by the European heatwave in 2022. Many more applications of WaterNet is accessible to anyone at <https://geeiimperial.herokuapp.com/> and were created by integrating GAE and web applications (React and node.js).

For governments, insurance companies interested to use WaterNet, a simple framework was setup. The framework requires small budget but for anyone reusing the algorithm the WaterNet have to be retrained and predictions are generated from GC notebooks. A simple integration with GAE and web application was demonstrated but this can be extended using the Google AI platform similar to Brown et al. (2022). Future work should secure funds to launching models in Google AI platform and create web application. Using this workflow government, insurance companies can assess a real-time global map without coding.

4.3 Limitations on the model and training methods

More training samples could be selected to improve the performance of SentSlope. Mayer et al. (2021) has trained the Sentinel-1 U-Net model locally in Cambodia reaching an F1 score of 0.97. To reach an F1 score of 0.97, Mayer et al. (2021) trained 8843 patches of 256 x 256 pixels. On the other hand, 720 patches was used in our experiment as it was not financially and computationally possible to train as many patches of data as Mayer et al. (2021) to train 154 combinations of models. An optimal time-to-cost ratio had to be chosen. Training data should also be selected in the same geographical region rather than everywhere on the globe as the F1 score of SentSlope (a combination of WaterNet) is improved up to 0.90 from 0.86 described in Section 3.2.

WaterNet's prediction can be further improved by tuning the hyperparameters, such as the learning rate and the loss functions. This work briefly looked at the loss functions of CC and CC-DICE, but other loss functions should be explored (Jadon 2020, Bai et al. 2021, Seale et al. 2022). An example of a hyperparameter one could look at is the Sobel-Edge loss where the Sobel-Edge loss was able to detect fine-scale water detail near the coastline better than CC in coastline benchmark data (Catherine Seale 2019). Future work should investigate the loss functions and other hyperparameters in more depth.

There are limitations on the labeling data and training data we chose. First, the JRC label contains no-data and cloud masking region are likely to coincide so there is likely to be an overestimate of the F1 score for combinations relying on Landsat-8 in Table 2. Secondly, the WaterNet can only predict at 30m, because the JRC monthly water data gives a resolution of 30m as described in Section 3.6. WaterNet's resolution can improve with a higher resolution labeling and training data. Models can be trained with Sentinel-2, Planet Scope or Rapid Eye (Planet 2022) to match the increase in resolution. Lastly, the Landsat-8 collection 1 was used to train the models, but the mission ceased in December 2021. For predictions in 2022, collection 2 was used to predict water bodies but this is unsupported (ESA 2021). Collection 2 should be used to train the models in the future.

5 Conclusions

In this study, we developed a multi-view learning approach with U-Net cores to create water classification maps by fusing the Sentinel-1, Landsat-8 and NASADEM imagery on the JRC monthly

data. By exploring diverse bands of L8SR and S1 satellites with NASADEM the best model called WaterNet was obtained. The comparison experiments demonstrate that the WaterNet could detect water bodies in various terrain in hilly and high elevation regions, a task that has proved problematic in the past (Song et al. 2007). The WaterNet performed well to map the flooding event over five months in Southern Thailand 2016, showing robustness under cloud cover of 55 %. WaterNet is able to detect fine water details between coastline well up to 30m. The workflow can be used by governments, emergency services and researchers to easily evaluate water bodies in an area of one million hectares. The application of the model was deployed to the freely available React-built web application called WaterOracle. The extension of our approach is to incorporate additional training data with higher resolution and hyperparameter tuning is left for future work.

References

- Acharya, T. D. & Yang, I. (2015), 'Exploring landsat 8', *International Journal of IT, Engineering and Applied Sciences Research (IJIEASR)* **4**(4), 4–10.
URL: https://www.researchgate.net/profile/Tri-Acharya/publication/311901147_Exploring_Landsat_8/links/589c0de6458515e5f4549e58/ExploringLandsat8.pdf
- Attema, E., Davidson, M., Snoeij, P., Rommen, B. & Flouri, N. (2009), Sentinel-1 mission overview, in '2009 IEEE International Geoscience and Remote Sensing Symposium', Vol. 1, IEEE, pp. I–36.
URL: <https://ieeexplore.ieee.org/abstract/document/5416921>
- Bai, Y., Wu, W., Yang, Z., Yu, J., Zhao, B., Liu, X., Yang, H., Mas, E. & Koshimura, S. (2021), 'Enhancement of detecting permanent water and temporary water in flood disasters by fusing sentinel-1 and sentinel-2 imagery using deep learning algorithms: Demonstration of sen1floods11 benchmark datasets', *Remote Sensing* **13**(11), 2220.
- Brown, C. F., Brumby, S. P., Guzder-Williams, B., Birch, T., Hyde, S. B., Mazzariello, J., Czerwinski, W., Pasquarella, V. J., Haertel, R., Illyushchenko, S. et al. (2022), 'Dynamic world, near real-time global 10 m land use land cover mapping', *Scientific Data* **9**(1), 1–17.
URL: <https://www.nature.com/articles/s41597-022-01307-4>
- Catherine Seale, Thomas Redfern, P. C. (2019), 'Sentinel-2 water edges dataset (swed)', <https://openmldata.ukho.gov.uk/>, note = "[Online; accessed 23-Aug-2022]" .
- Cotton, W. R., Bryan, G. & van den Heever, S. C. (2011), Cumulonimbus clouds and severe convective storms, in 'International geophysics', Vol. 99, Elsevier, pp. 315–454.
- Crippen, R., Buckley, S., Belz, E., Gurrola, E., Hensley, S., Kobrick, M., Lavalle, M., Martin, J., Neumann, M., Nguyen, Q. et al. (2016), 'Nasadem global elevation model: Methods and progress'.
- ESA (2021), 'New landsat-8 collection-2 level 1 and level 2 data available from january', <https://earth.esa.int/eogateway/news/new-landsat-8-collection-2-level-1-and-level-2-data-available-from-january>. [Online; accessed 20-Aug-2022].
- Feir, B. J. & Toothaker, L. E. (1974), 'The anova f-test versus the kruskal-wallis test: A robustness study.'.
URL: <https://eric.ed.gov/?id=ED091423>
- Google (2022a), 'Ai platform training pricing', <https://cloud.google.com/ai-platform/training/pricing>. [Online; accessed 24-Jun-2022].
- Google (2022b), 'Choose the colab plan that's right for you', <https://colab.research.google.com/signup>, note = "[Online; accessed 2-Aug-2022]" .
- Google (2022c), 'Tensorflow example workflows', https://developers.google.com/earth-engine/guides/tf_examples. [Online; accessed 24-Jun-2022].
- Gorelick, N., Hancher, M., Dixon, M., Illyushchenko, S., Thau, D. & Moore, R. (2017), 'Google earth engine: Planetary-scale geospatial analysis for everyone', *Remote Sensing of Environment* .
URL: <https://doi.org/10.1016/j.rse.2017.06.031>
- Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroud, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J. et al. (2018), 'Recent advances in convolutional neural networks', *Pattern recognition* **77**, 354–377.
URL: <https://www.sciencedirect.com/science/article/pii/S0031320317304120>
- Gulács, A. & Kovács, F. (2020), 'Sentinel-1-imagery-based high-resolution water cover detection on wetlands, aided by google earth engine', *Remote Sensing* **12**(10), 1614.
URL: <https://www.mdpi.com/2072-4292/12/10/1614/pdf?version=1589809132>

- Jadon, S. (2020), A survey of loss functions for semantic segmentation, *in* '2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)', IEEE, pp. 1–7.
- Ji, R., Yu, W., Feng, R., Wu, J. & Zhang, Y. (2019), The threshold determination methods of water body information extraction using gf-1 satellite image, *in* 'IOP Conference Series: Materials Science and Engineering', Vol. 592, IOP Publishing, p. 012088.
- JRC/Google, E. (2022), 'Jrc monthly water history, v1.3', https://developers.google.com/earth-engine/datasets/catalog/JRC_GSW1_3_MonthlyHistory#citations. [Online; accessed 26-Jun-2022].
- Konapala, G., Kumar, S. V. & Ahmad, S. K. (2021), 'Exploring sentinel-1 and sentinel-2 diversity for flood inundation mapping using deep learning', *ISPRS Journal of Photogrammetry and Remote Sensing* **180**, 163–173.
- URL:** <https://ui.adsabs.harvard.edu/abs/2021JPRS..180..163K/abstract>
- Kousky, C. & Kunreuther, H. (2009), 'Improving flood insurance and flood risk management: Insights from st. louis, missouri'.
- Kseňák, L., Pukanská, K., Bartoš, K. & Blišťan, P. (2022), 'Assessment of the usability of sar and optical satellite data for monitoring spatio-temporal changes in surface water: Bodrog river case study', *Water* **14**(3), 299.
- URL:** <https://www.mdpi.com/2073-4441/14/3/299/htm>
- Labelbox (2022), 'Labelbox', <https://labelbox.com>. [Online; accessed 1-Jul-2022].
- Lamba, H. (2019), 'Understanding semantic segmentation with unet', <https://towardsdatascience.com/understanding-semantic-segmentation-with-unet-6be4f42d4b47>. [Online; accessed 31-Jul-2022].
- Li, H., Zech, J., Ludwig, C., Fendrich, S., Shapiro, A., Schultz, M. & Zipf, A. (2021), 'Automatic mapping of national surface water with openstreetmap and sentinel-2 msi data using deep learning', *International Journal of Applied Earth Observation and Geoinformation* **104**, 102571.
- URL:** <https://www.sciencedirect.com/science/article/pii/S0303243421002786>
- Machado, G., Ferreira, E., Nogueira, K., Oliveira, H., Brito, M., Gama, P. H. T. & dos Santos, J. A. (2020), 'Airound and cv-brct: Novel multiview datasets for scene classification', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **14**, 488–503.
- Markert, K. N., Markert, A. M., Mayer, T., Nauman, C., Haag, A., Poortinga, A., Bhandari, B., Thwal, N. S., Kunlamai, T., Chishtie, F. et al. (2020), 'Comparing sentinel-1 surface water mapping algorithms and radiometric terrain correction processing in southeast asia utilizing google earth engine', *Remote Sensing* **12**(15), 2469.
- Mayer, T., Poortinga, A., Bhandari, B., Nicolau, A. P., Markert, K., Thwal, N. S., Markert, A., Haag, A., Kilbride, J., Chishtie, F. et al. (2021), 'Deep learning approach for sentinel-1 surface water mapping leveraging google earth engine', *ISPRS Open Journal of Photogrammetry and Remote Sensing* **2**, 100005.
- URL:** <https://www.sciencedirect.com/science/article/pii/S2667393221000053>
- Mutanga, O. & Kumar, L. (2019), 'Google earth engine applications'.
- URL:** <https://www.mdpi.com/2072-4292/11/5/591/htm>
- Özelkan, E. (2020), 'Water body detection analysis using ndwi indices derived from landsat-8 oli', *Polish Journal of Environmental Studies* **29**(2), 1759–1769.
- 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. (2011), 'Scikit-learn: Machine learning in Python', *Journal of Machine Learning Research* **12**, 2825–2830.
- Pekel, J.-F., Cottam, A., Gorelick, N. & Belward, A. S. (2016), 'High-resolution mapping of global surface water and its long-term changes', *Nature* **540**(7633), 418–422.
URL: <https://www.nature.com/articles/nature20584?ref=https://githubhelp.com>
- Piriyajitakonkij, M., Warin, P., Lakhan, P., Leelaarporn, P., Kumchaiseemak, N., Suwananakorn, S., Pianpanit, T., Niparnan, N., Mukhopadhyay, S. C. & Wilaiprasitporn, T. (2020), 'Sleepposenet: multi-view learning for sleep postural transition recognition using uwb', *IEEE Journal of Biomedical and Health Informatics* **25**(4), 1305–1314.
- Planet (2022), 'Planet and imagery archive', <https://www.planet.com/products/planet-imagery/>, note = "[Online; accessed 20-Aug-2022]" .
- Rodriguez, J. D., Perez, A. & Lozano, J. A. (2009), 'Sensitivity analysis of k-fold cross validation in prediction error estimation', *IEEE transactions on pattern analysis and machine intelligence* **32**(3), 569–575.
- Ronneberger, O., Fischer, P. & Brox, T. (2015), U-net: Convolutional networks for biomedical image segmentation, in 'International Conference on Medical image computing and computer-assisted intervention', Springer, pp. 234–241.
URL: <https://arxiv.org/abs/1505.04597>
- Seale, C., Redfern, T., Chatfield, P., Luo, C. & Dempsey, K. (2022), 'Coastline detection in satellite imagery: A deep learning approach on new benchmark data', *Remote Sensing of Environment* **278**, 113044.
- Sobiech, J. & Dierking, W. (2013), 'Observing lake-and river-ice decay with sar: advantages and limitations of the unsupervised k-means classification approach', *Annals of Glaciology* **54**(62), 65–72.
- Song, Y.-S., Sohn, H.-G. & Park, C.-H. (2007), 'Efficient water area classification using radarsat-1 sar imagery in a high relief mountainous environment', *Photogrammetric Engineering & Remote Sensing* **73**(3), 285–296.
URL: <https://www.ingentaconnect.com/content/asprs/pers/2007/00000073/00000003/art00005>
- Spoto, F., Sy, O., Laberinti, P., Martimort, P., Fernandez, V., Colin, O., Hoersch, B. & Meygret, A. (2012), Overview of sentinel-2, in '2012 IEEE international geoscience and remote sensing symposium', IEEE, pp. 1707–1710.
URL: <https://ieeexplore.ieee.org/abstract/document/6351195>
- Subramanian, V. (2017), *Pro MERN Stack*, Springer.
- Tang, H., Lu, S., Ali Baig, M. H., Li, M., Fang, C. & Wang, Y. (2022), 'Large-scale surface water mapping based on landsat and sentinel-1 images', *Water* **14**(9), 1454.
URL: <https://www.mdpi.com/2073-4441/14/9/1454>
- Tyson Brown, N. G. S. (2022), 'Surface water', <https://education.nationalgeographic.org/resource/surface-water>. [Online; accessed 26-Jun-2022].
- Vandekerckhove, J., Matzke, D. & Wagenmakers, E.-J. (2015), 'Model comparison and the principle', *The Oxford handbook of computational and mathematical psychology* **300**.
- Verbeiren, B., Seyoum, S. D., Lubbad, I., Xin, T., Veldhuis, M.-C. t., Onof, C., Wang, L.-P., Ochoa-Rodriguez, S., Veeckman, C., Boonen, M. et al. (2018), Floodcitisense: Early warning service for urban pluvial floods for and by citizens and city authorities, in 'International Conference on Urban Drainage Modelling', Springer, pp. 660–664.

- Waldner, F. & Diakogiannis, F. I. (2020), 'Deep learning on edge: Extracting field boundaries from satellite images with a convolutional neural network', *Remote Sensing of Environment* **245**, 111741.
- WHO (2022a), 'Drought', https://www.who.int/health-topics/drought#tab=tab_2, note = "[Online; accessed 17-Aug-2022]".
- WHO (2022b), 'Flood', https://www.who.int/health-topics/floods#tab=tab_1, note = "[Online; accessed 17-Aug-2022]".
- Xiong, B., Fan, H., Grauman, K. & Feichtenhofer, C. (2021), Multiview pseudo-labeling for semi-supervised learning from video, in 'Proceedings of the IEEE/CVF International Conference on Computer Vision', pp. 7209–7219.
- Yang, J., Gong, P., Fu, R., Zhang, M., Chen, J., Liang, S., Xu, B., Shi, J. & Dickinson, R. (2013), 'The role of satellite remote sensing in climate change studies', *Nature climate change* **3**(10), 875–883.
- Zhao, J., Xie, X., Xu, X. & Sun, S. (2017), 'Multi-view learning overview: Recent progress and new challenges', *Information Fusion* **38**, 43–54.
- URL:** <https://www.sciencedirect.com/science/article/pii/S1566253516302032>

6 Appendix

6.1 Appendix.A

Table 4: The 77 experimental ID and their descriptions including FS U-Net, M2 U-Net and M3 U-Net.

Experiment ID	Name	Description
1FS	S1A_el_sl_as	VV, VH, angle of S1 resampled and el, sl, as from NASADEM
2FS	S1A_el	VV, VH, angle of S1 resampled and elevation band from NASADEM
3FS	S1A_sl	VV, VH, angle of S1 resampled and slope band from NASADEM
4FS	S1A_as	VV, VH, angle of S1 resampled and aspect band from NASADEM
5FS	S1A_sl_as	VV, VH angle of S1 resampled and elevation and slope band from NASADEM
6FS	S1A_el_sl	VV, VH, angle of S1 resampled and elevation and slope band from NASADEM
7FS	S1A_el_as	VV, VH, angle of S1 resampled and elevation and aspect band from NASADEM
8FS	S1_el_sl_as	VV, VH of S1 resampled and elevation, slope and aspect band from NASADEM
9FS	S1_el	VV, VH of S1 resampled and elevation, band from NASADEM
10FS	S1_sl	VV, VH, of S1 resampled and slope band from NASADEM
11FS	S1_as	VV, VH of S1 resampled and aspect from NASADEM
12FS	S1_sl_as	VV, VH of S1 resampled and slope and aspect band from NASADEM
13FS	S1_el_sl	VV, VH, of S1 resampled and elevation and slope band from NASADEM
14FS	S1_el_as	VV, VH of S1 resampled and elevation and aspect band from NASADEM
15FS	L8SR_el_sl_as	optical Bands from L8SR and elevation and slope band from NASADEM
16FS	L8SR_el	Optical Bands from L8SR and elevation from NASADEM
17FS	L8SR_sl	Optical Bands from L8SR and slope from NASADEM
18FS	L8SR_as	Optical Bands from L8SR and aspect from NASADEM
19FS	L8SR_sl_as	Optical Bands from L8SR and slope and aspect from NASADEM
20FS	L8SR_el_sl	Optical Bands from L8SR and elevation and slope from NASADEM
21FS	L8SR_el_as	Optical Bands from L8SR and elevant and aspect from NASADEM
22FS	L8SR_S1_el	Optical Bands from L8SR, VV and VH bands from S1 and slope and aspect from NASADEM
23FS	L8SR_S1_sl	Optical Bands from L8SR, VV and VH bands from S1 and slope from NASADEM
24FS	L8SR_S1_sl_el_as	Optical Bands from L8SR, VV and VH bands from S1 and elevation and slope from NASADEM
25FS	L8SR_S1A_el	Optical Bands from L8SR, VV, VH and angle bands from S1 and elevation from NASADEM

Continued on next page

Table 4 – continued from previous page

Experiment ID	Model Name	Description
26FS	L8SR_S1A.sl	Optical Bands from L8SR, VV, VH and angle bands from S1 and slope from NASADEM
27FS	L8SR_S1A.sl_el_as	Optical Bands from L8SR, VV, VH and angle bands from S1 and slope elevation from NASADEM
28FS	L8SR	Optical Bands from L8SR
29FS	S1	VV, VH bands from S1
30FS	S1A	VV, VH and angle bands from S1 resampled
31FS	L8SR_S1	Optical Bands from L8SR and VV, VH bands from S1 resampled
32FS	L8SR_S1A	Optical Bands from L8SR and VV, VH, angle bands from S1 resampled
1M2	S1A+(el_sl_as)	VV, VH, angle bands from S1 resampled and elevation, slope and aspect from NASADEM joined in two views
2M2	S1A+el	VV, VH, angle bands from S1 resampled and elevation from NASADEM joined in two views
3M2	S1A+sl	VV, VH, angle bands from S1 resampled and slope from NASADEM joined in two views
4M2	S1A+as	VV, VH, angle bands from S1 resampled and aspect from NASADEM joined in two views
5M2	S1A+(sl_as)	VV, VH, angle bands from S1 resampled and slope, aspect from NASADEM joined in two views
6M2	S1A+(el_sl)	VV, VH, angle bands from S1 resampled and elevation, slope from NASADEM joined in two views
7M2	S1A+(el_as)	VV, VH, angle bands from S1 resampled and elevation, aspect from NASADEM joined in two views
8M2	S1+(el_sl_as)	VV, VH bands from S1 resampled and elevation, slope and aspect from NASADEM joined in two views
9M2	S1+el	VV, VH bands from S1 resampled and elevation from NASADEM in two views
10M2	S1+sl	VV, VH bands from S1 resampled and slope from NASADEM in two views
11M2	S1+as	VV, VH bands from S1 resampled and aspect from NASADEM in two views
12M2	S1+(sl_as)	VV, VH bands from S1 resampled and slope and aspect from NASADEM in two views
13M2	S1+(el_sl)	VV, VH bands from S1 resampled and elevation and slope from NASADEM in two views
14M2	S1+(el_as)	VV, VH bands from S1 resampled and elevation and aspect from NASADEM in two views
15M2	L8SR+(el_sl_as)	Optical Bands from L8SR and elevation, slope and aspect band from NASADEM in two views
16M2	L8SR+el	Optical Bands from L8SR and elevation band from NASADEM in two views
17M2	L8SR+sl	Optical Bands from L8SR and slope band from NASADEM in two views
18M2	L8SR+as	Optical Bands from L8SR and aspect band from NASADEM in two views
19M2	L8SR+(sl_as)	Optical Bands from L8SR and slope band and aspect from NASADEM in two views
20M2	L8SR+(el_sl)	Optical Bands from L8SR and elevation band and slope from NASADEM in two views

Continued on next page

Table 4 – continued from previous page

Experiment ID	Model Name	Description
21M2	L8SR+(el_as)	Optical Bands from L8SR and elevation band and aspect from NASADEM in two views
22M2	L8SR+(S1_as)	Optical Bands from L8SR and VV, VH of S1 resampled and aspect band from NASADEM in two views
23M2	L8SR+(S1_el)	Optical Bands from L8SR and VV, VH of S1 resampled and elevation band from NASADEM in two views
24M2	L8SR+(S1_sl)	Optical Bands from L8SR and VV, VH of S1 resampled and slope band from NASADEM in two views
25M2	L8SR+(S1_sl_as)	Optical Bands from L8SR and VV, VH of S1 resampled and slope and aspect band from NASADEM in two views
26M2	L8SR+(S1_el_sl)	Optical Bands from L8SR and VV, VH of S1 resampled and elevation and slope band from NASADEM in two views
27M2	L8SR+(S1_el_as)	Optical Bands from L8SR and VV, VH of S1 resampled and slope and aspect band from NASADEM in two views
28M2	L8SR+(S1_sl_el_as)	Optical Bands from L8SR and VV, VH of S1 resampled and slope, elevation and aspect band from NASADEM in two views
29M2	L8SR+(S1A_as)	Optical Bands from L8SR and VV, VH, angle of S1 resampled and aspect band from NASADEM in two views
30M2	L8SR+(S1A_el)	Optical Bands from L8SR and VV, VH, angle of S1 resampled and elevation band from NASADEM in two views
31M2	L8SR+(S1A_sl)	Optical Bands from L8SR and VV, VH, angle of S1 resampled and slope band from NASADEM in two views
32M2	L8SR+(S1A_sl_as)	Optical Bands from L8SR and VV, VH, angle of S1 resampled and slope and aspect band from NASADEM in two views
33M2	L8SR+(S1A_el_sl)	Optical Bands from L8SR and VV, VH, angle of S1 resampled and elevation and slope band from NASADEM in two views
34M2	L8SR+(S1A_el_as)	Optical Bands from L8SR and VV, VH, angle of S1 resampled and elevation and aspect band from NASADEM in two views
35M2	L8SR+(S1A_sl_el_as)	Optical Bands from L8SR and VV, VH, angle of S1 resampled and slope and elevation and aspect band from NASADEM
36M2	L8SR+S1	Optical Bands from L8SR and VV, VH of S1 in two views
37M2	L8SR+S1A	resampled and slope and elevation Optical Bands from L8SR and VV, VH, angle of S1 resampled and slope and elevation in two views
1M3	L8SR+S1+as	Optical Bands from L8SR and VV, VH of S1 resampled and aspect from NASADEM in three views
2M3	L8SR+S1+el	Optical Bands from L8SR and VV, VH of S1 resampled and elevation from NASADEM

Continued on next page

Table 4 – continued from previous page

Experiment ID	Model Name	Description
3M3	L8SR+S1+sl	Optical Bands from L8SR and VV, VH of S1 resampled and slope from NASADEM in three views
4M3	L8SR+S1+(sl_el_as)	Optical Bands from L8SR and VV, VH of S1 resampled and slope, elevation, aspect from NASADEM in three views
5M3	L8SR+S1A+as	Optical Bands from L8SR and VV, VH, angle of S1 resampled and aspect from NASADEM in three views
6M3	L8SR+S1A+el	Optical Bands from L8SR and VV, VH, angle of S1 resampled and elevation from NASADEM in three views
7M3	L8SR+S1A+sl	Optical Bands from L8SR and VV, VH, angle of S1 resampled and slope from NASADEM in three views
8M3	L8SR+S1A+(sl_el_as)	Optical Bands from L8SR and VV, VH, angle of S1 resampled and slope, elevation and aspect from NASADEM in three views