

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

31st 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 an F1 score of 0.96, which other schemes struggle to achieve. Importantly, because the multi-view algorithm combines data from passive (L8SR) and active (S1) radiation sources, the 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, showing damages to infrastructures, cropland and human activities. The algorithm can also monitor other water issues, such as the flood event in Bangladesh and the drought caused by the European heatwave in 2022. These applications are made freely available 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 negatively impact human lives through natural hazards. The two most common natural hazards caused by water include flooding and droughts, affecting three billion people from 1998-2017 (WHO 2022a,b). Creating a near-real-time water map has many applications such as aiding the government and emergency services in preparing for flooding or drought events (Verbeiren et al. 2018). The flood inundated maps need the highest possible accuracy to prevent loss of lives and enormous financial loss from property insurance companies (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 the 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 maps, especially SAR due to the higher occurrence of cumulonimbus clouds during flood events (Cotton et al. 2011).

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) is a common method to detect water, but the method is affected by geographical location and time (Bai et al. 2021). Machine learning models proved to predict water bodies more accurately than the threshold selection method on top of not having to select the threshold (Bangira et al. 2019). Convolutional Neural Networks (CNN) trained on either Landsat-8 or Sentinel-1 exceeded the performance of simple machine learning and threshold methods (Li et al. 2021, Mayer et al. 2021). CNN is one of the most suitable for image segmentation problems and among many CNN, U-Net is one of the most accurate and commonly used (Ronneberger et al. 2015). Despite the accuracy of CNN, a large amount of labeling data is needed and is acquired either by manual labeling tools (Labelbox 2022) or pre-existing label data (Tang et al. 2022, JRC/Google 2022, Pekel et al. 2016, Li et al. 2021).

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 they can extend CNN ability to predict water bodies (Bai et al. 2021). Konapala et al. (2021) explored 32 combinations of Sentinel-1, Sentinel-2 and DEM bands with U-Net to understand the connection between 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). Konapala et al. (2021)'s result indicates that there may not be benefits to concatenating features. However, a data fusion technique called multi-view learning (Zhao et al. 2017) has proven to outperform the U-Net in detecting water bodies (Machado et al. 2020, Muñoz et al. 2021).

Even though the fusion of satellites has improved the ability of CNN to classify water bodies, the architecture could be improved and made more complex with U-Net (Muñoz et al. 2021). Most attempts to explore the best way to connect the satellite features have either concatenated or fused the features, but none have experimented with using both connection methods in one model (Konapala et al. 2021, Rudner et al. 2019). As a result, more combinations to connect the features are left to be explored. Most papers using the multi-view technique emphasized how accurate data fusion techniques were in predicting water bodies, but no papers have quantified the sensitivity of fusion techniques to detect water under clouds and in various terrains. Attempts to create global water classification maps have trained models in a location, but no work has trained the model in global and local regions and compared the pros and cons (Mayer et al. 2021, Konapala et al. 2021).

This study aims to develop a multi-view learning algorithm that leverages multi-sensor data to identify surface water bodies on Earth. The task is broken down into four areas. (1) Designing the architecture of multi-view learning with U-Net to combine multi-sensor data. (2) Finding the optimal combinations of Sentinel-1, Landsat-8 and NASADEM and discovering the training location which generates the best prediction. (3) 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. (4) Developing a web application to showcase applications of the best model.

2 Methodology

In this study, several U-Net models with multiple views were built. 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. The best combination was applied to predict a highly elevated area and a flooding event in southern Thailand.

2.1 Study Area

Since the aim was 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. Comparing the performances of models trained in Thailand and globally will result in an understanding of the training location to use.

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 with 240 patches of 256 x 256 pixels (Figure 1). Global training data was selected in ten countries i.e. Thailand, China, Ghana, Brazil, Mexico, Pakistan, Egypt, Cambodia, India and Bangladesh, with 72 patches each (Figure 1). 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. K-fold cross-validation method (Rodriguez et al. 2009) would be more reliable than testing unseen data but was avoided due to extra financial costs.

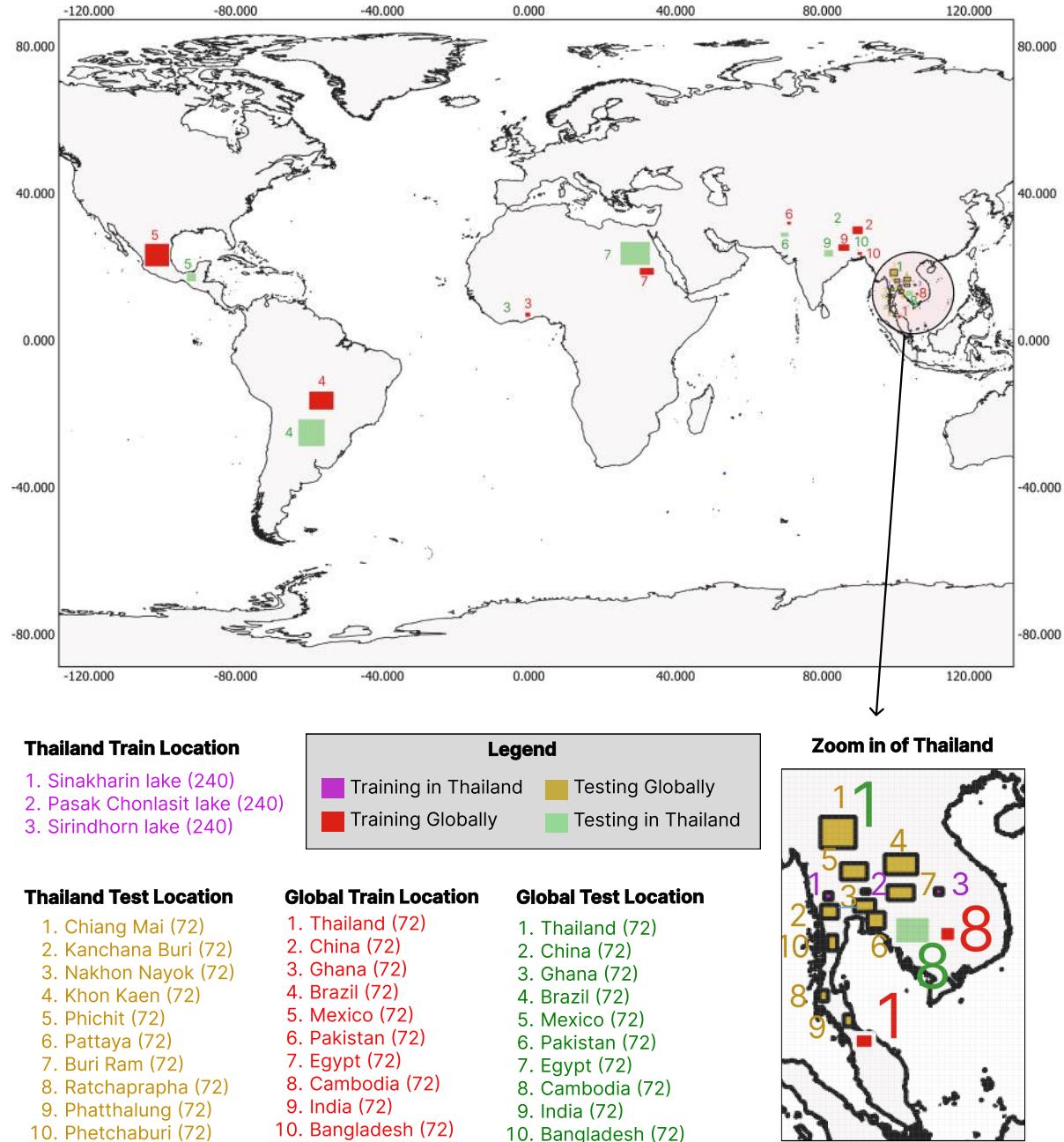


Figure 1: The location of training and testing points in Thailand and globally from January to February 2018.

2.2 Data

This study used the JRC monthly water history as labeling data. The models were trained 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 every month, allowing testing against temporal water evolution. The 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). The 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 containing SAR was used because of its ability to penetrate clouds. 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 Sentinel-1 was imported without further preprocessing, which included the vertical transmitting with vertical receive (VV), horizontal receive (VH) and the approximate incidence angle from the ellipsoid (angle) bands. 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 slope and aspect were calculated from NASADEM using the earthengine-api (Gorelick et al. 2017).

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).

The models used the categorical cross entropy (CC) and each model was trained with ten epochs to balance good prediction with computational costs.

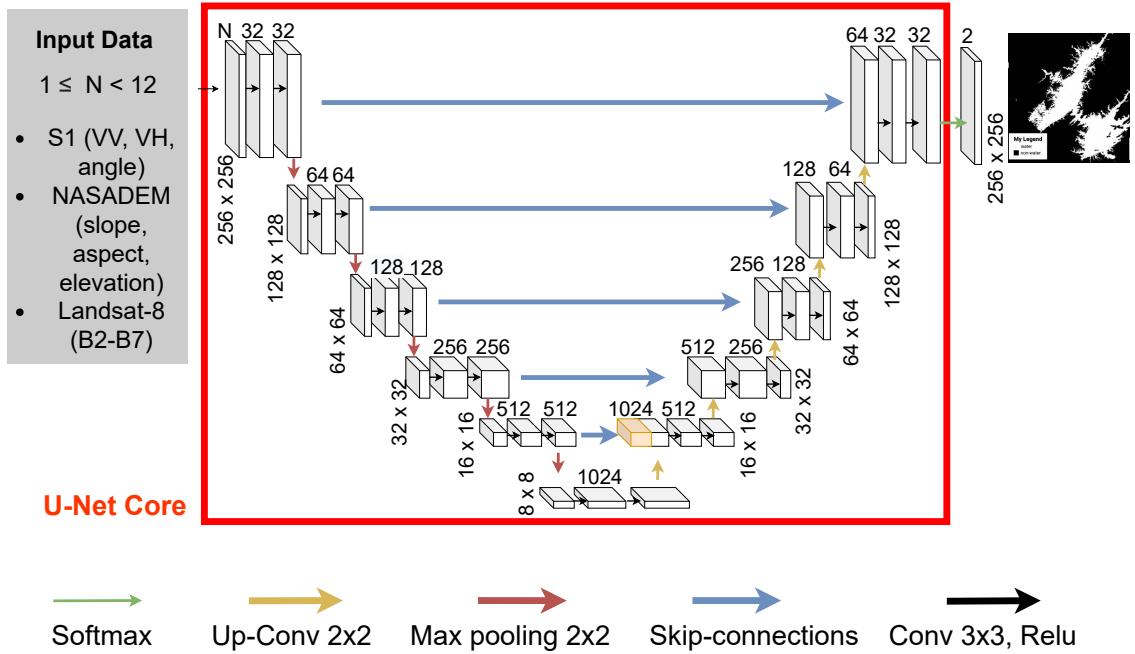
2.3.1 The Feature Stack U-Net architecture

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, called the Feature Stack (FS) U-Net, was adapted from Ronneberger et al. (2015) and Google (2022c). 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 in Figure 2a and Ronneberger et al. (2015) is an additional max pooling and up convolution structure to compensate for the half size of the input image.

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.

(a)



(b)

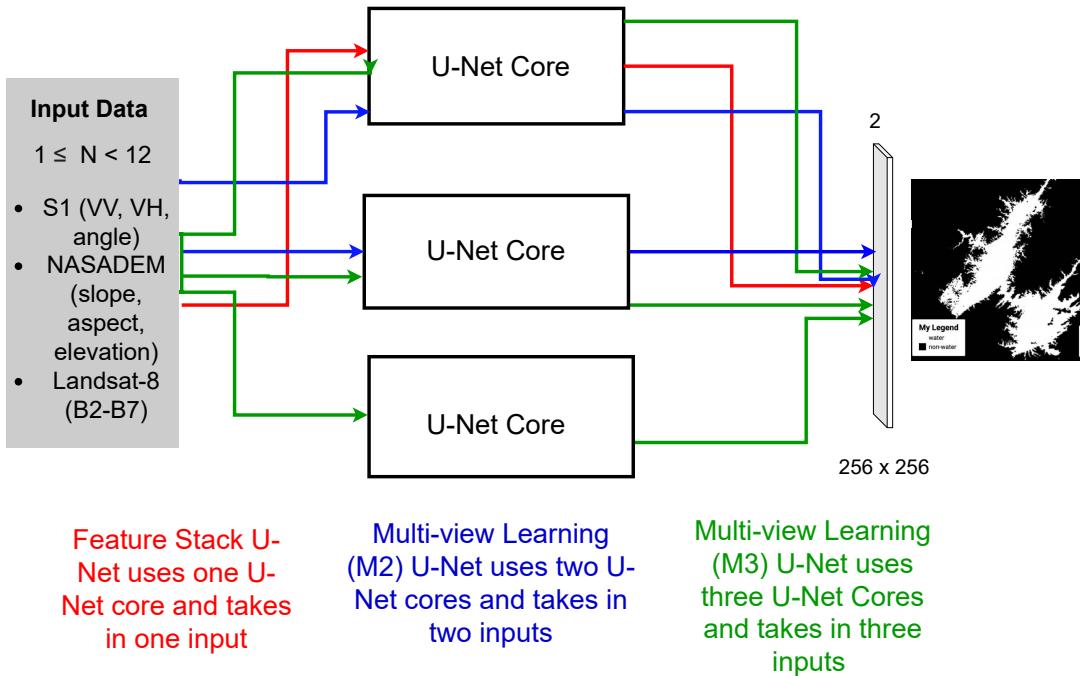


Figure 2: (a) U-Net architecture for water segmentation problem modified for taking in 256×256 pixels of input with N features. (b) Comparison between the architecture of Feature stack U-Net and Multi-view learning with two and three input U-Net cores for water segmentation problem.

2.3.2 The multi-view learning with U-Net architecture compared with feature stack U-Net

Multi-view learning is a data fusion process known for improving generalization and reducing the over-fitting problem. 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 2b 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 design

There are many possible ways to combine features from multi-sensor data so different combinations of features were experimented to determine the best. As seen in Figure 2, the multi-view U-Net can have up to three inputs and 12 features and the features were divided into six smaller features as shown in Table 1. Each of these six features has a U-Net core and was paired either by concatenation or multiple views. 77 feature combinations were selected in Appendix Table 4 which is 35 more than Konapala et al. (2021) and 72 more than Rudner et al. (2019). For each combination, the model was trained in Thailand and globally as described in Section 2.1. The best model was tuned between CC and combo loss with dice (CC-DICE) to improve the model fitting with imbalance data (Jadon 2020, Mayer et al. 2021).

Table 1: The symbols are used to describe the features used in a U-Net core and how these cores are joint. The feature and the U-Net core names are L8SR, S1, S1A, sl, el, as and the two connection symbols are either _ or +

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.

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

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 imbalanced 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 as seen in equation 1. The macro average was used in Scikit-learn to give equal weights to each class (Pedregosa et al. 2011).

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (1)$$

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 to provide 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 model's weights, (Google 2022a). Training the models had led to unavoidable storage costs of an average of 300MB-1GB per experiment even when the training patches were 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 with the MERN stack (Mongo, Express, React, and Node.js) to build a full-stack web application (Subramanian 2017) showing applications of the models.

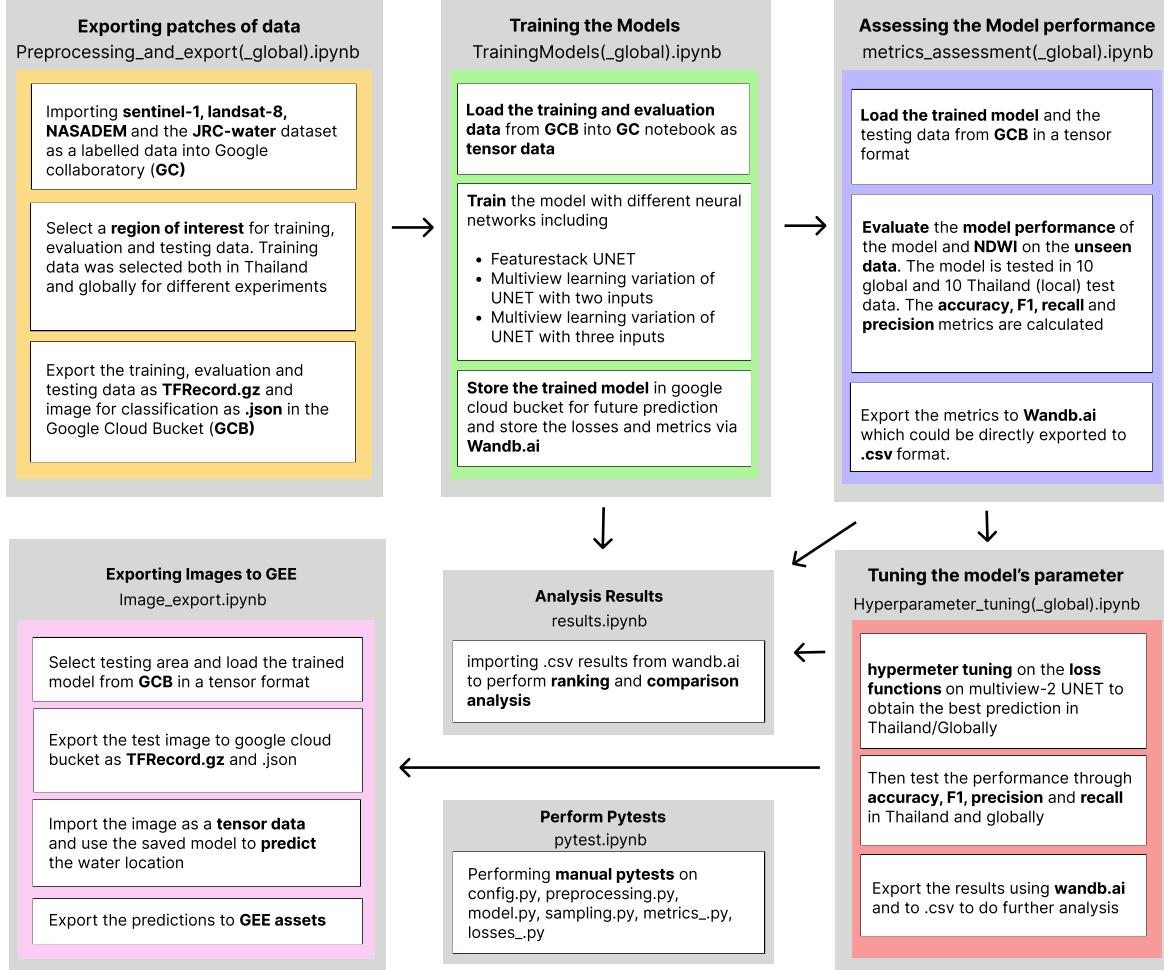


Figure 3: The development methodology using eleven Google Collaboratory notebooks to carry out the experiments. The experiments rely on training locally and globally, so the notebook is split into _Thailand and _global, respectively.

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 experiment.
- **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 are 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 Methods comparison

3.1.1 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 Figure 1. The respective F1 scores are 0.90, 0.89, 0.86 and the average of 0.88 is used as the baseline.

3.1.2 Comparison between different combinations and of Feature Stack U-Net

The FS U-Net combinations trained in Thailand and globally 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, adding S1 or elevation bands to the model seems to be no benefit 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 best pairing of S1 with NASADEM is slope, increasing the F1 score from 0.62-0.84 to 0.8-0.9, agreeing with Song et al. (2007). S1_sl or S1A_sl is denoted as SentSlope. As expected, the choice of training location is important for SentSlope, but not L8SR. Training S1_sl in Thailand gives a higher F1 score than training S1_sl globally in different terrain as shown in 10FS in Table 2. To give SentSlope optimal results, the models should be trained in the same geographical area as where it is predicted.

Table 2: The F1 scores of 32 feature stack combinations that were trained in Thailand and globally.

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.

3.1.3 Comparison between different combinations of multi-view U-Net

The multi-view U-Net combinations trained in Thailand and globally were performed on 45 different models and the results are in Table 3. The Krushal-Wallis test was applied to the combinations with the same feature but different connections (Table 2 and 3). The p-value for Krushal-Wallis is 0.004 indicating that the multi-view learning gives a higher F1 score than U-Net with one view (Feir & Toothaker 1974). Notably, the feature of Landsat-8 joined in multiple views with SentSlope data was superior to stacking the features.

Table 3: The F1 scores of 45 multi-view with U-Net combination that was trained in Thailand and globally.

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 model is L8SR + (S1A_sl), reaching an F1 score of 0.93-0.95. 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 3.1.2. 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) can be used and more emphasis is put on the SentSlope. The discovery that L8SR + (S1A_sl) is the best model shows that utilizing both concatenation and multiple views is better than considering one alone like in Rudner et al. (2019), Muñoz et al. (2021). This may also agree with Vandekerckhove et al. (2015) that using a simpler model is better which is why L8SR + (S1A_sl) outperformed L8SR + S1A + sl. The best model, L8SR + (S1A_sl), is named the WaterNet.

As described in Section 2.4, the WaterNet was tuned between CC and CC-DICE loss function. The CC-DICE loss function for WaterNet trained globally gives a higher F1 score than CC. CC loss is used for predictions in Thailand with F1 score of 0.95 and CC-DICE is used for global predictions with F1 score of 0.94.

3.2 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 and flood regions with a high cloud cover percentage.

3.2.1 High elevation terrain

The model was validated in Tibet, where an area of size 400,000 hectares was selected to assess the performance. This highly elevated area was chosen to validate the effect of similar backscattering coefficient in S1 as described in Song et al. (2007). 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 were compared in Figure 4.

S1 was the worst performing model 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 due to the partial cloud cover, 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.

Some parts in the training data have no data and were removed but the gap inside the training data may make it difficult for the model to capture spatial information. To understand how the model performs in no-data region, Figure 4 (a) was studied in the no-data parts. Brown et al. (2022)'s landcover image was used as a second verification to understand the classification in the no-data from January to July 2018 in 4 (b). The Brown et al. (2022)'s prediction shows that ice is present in the no-data region which agrees with WaterNet's prediction. Overall, WaterNet performs well in difficult terrain and can predict well in the region of no-data.

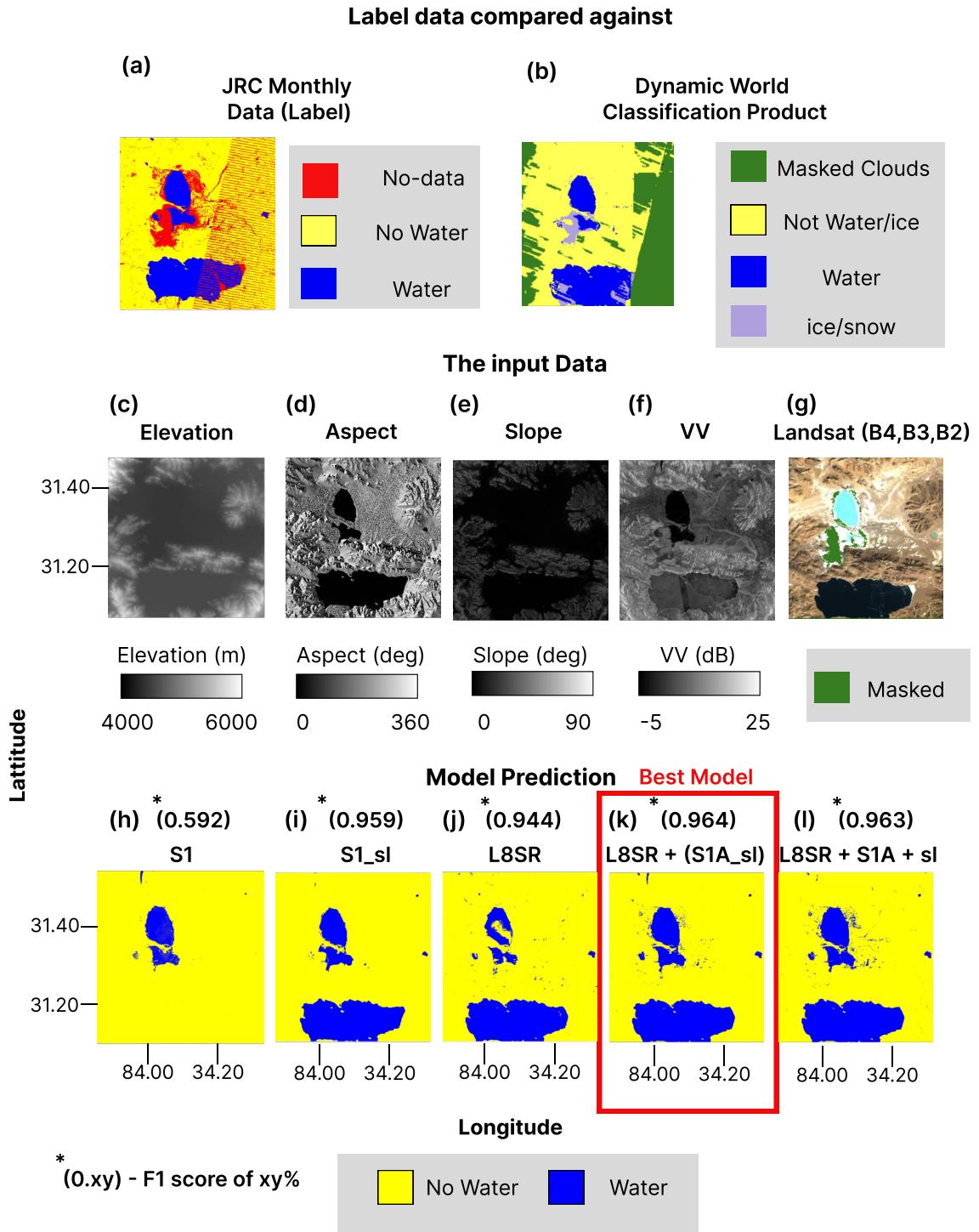


Figure 4: Comparing the performance of S1, S1_sl, L8SR, L8SR+(S1_sl) and L8SR + S1 + sl in a hilly terrain of high elevation in Tibet from January to February 2018 by looking at the assumed ground truth, input data and model prediction

3.2.2 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, S1_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. F1 value for NDWI was high because only the cloud-free areas were used for testing. The second column is the JRC label data and noticeably, there are more no-data regions during cloudy scenes agreeing with the data labeling procedure from (Pekel et al. 2016). The L8SR + (S1A_sl) column 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 L8SR column performs well during the cloud-free period and the SentSlope model consistently has a lower F1 score than L8SR, agreeing with the results in Section 3.1.2.

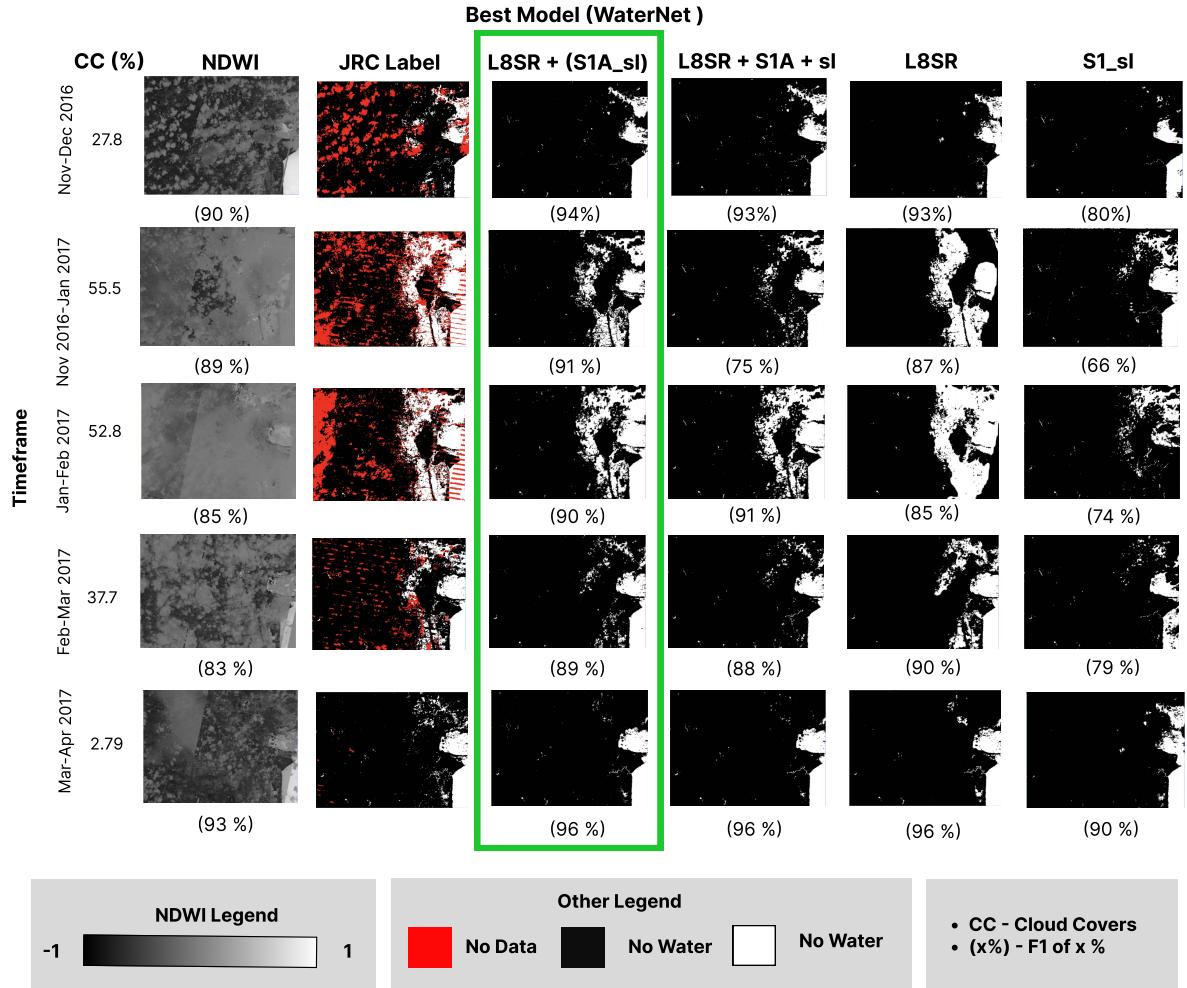


Figure 5: Comparing the temporal performance of WaterNet, S1_sl, L8SR, L8SR + S1 + sl and NDWI in the southern part of Thailand (Nakhon Si Thammarat) from December 2016 to early 2017 with varying cloud covers.

3.2.3 Cloud Cover Analysis

To understand the effect of cloud cover on the models, more points are sampled monthly in the same region as Section 3.2.2. The t-test tests the relationship between cloud cover and the F1 score. The null hypothesis is defined as cloud cover having no relationship with F1 score (gradient is 0) and the alternative hypothesis indicates having a relationship.

Figure 6 shows S1_sl is the most robust to cloud cover with a p-value of 73 %, not rejecting 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 water under the clouds. Also, the average F1 score of the L8SR + (S1A_sl) is 0.93 despite the cloud cover percentage varying from 0 to almost 100 %.

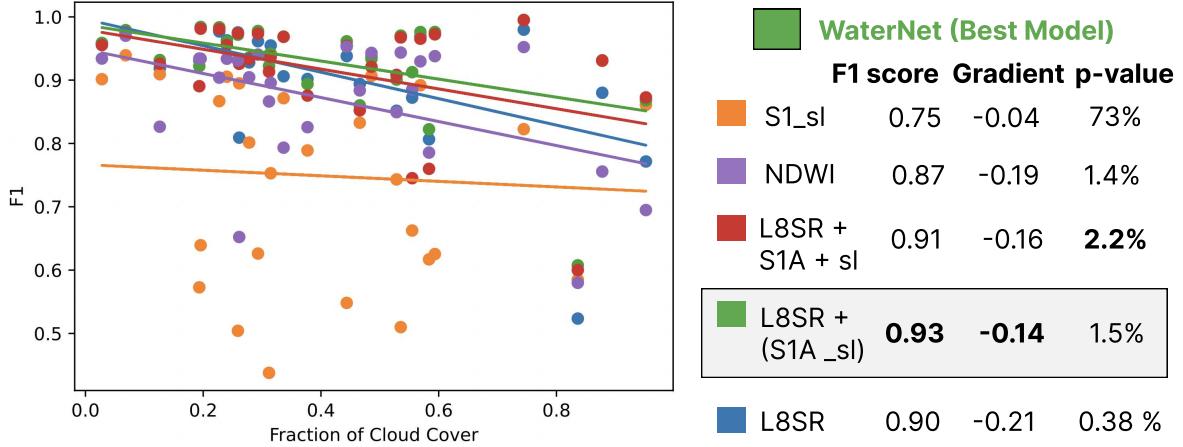


Figure 6: Cloud cover sensitivity analysis using the gradient, F1 score and p-value. 28 months from November 2016 to March 2019 were analyzed in the same area as in Section 3.2.2.

4 Discussion

4.1 Overall performance of WaterNet

Among the 154 experiments in Section 3.1, WaterNet, a fusion between L8SR and SentSlope, is the most effective model for detecting water bodies having a high average F1 score of 0.94-0.95 (Table 2, Table 3). Rudner et al. (2019) and Muñoz et al. (2021) explored the fusion between satellite features and proposed that three views give the best result. A series of assessments show that joining the features in two views, using both + and - connection is superior to using three views according to Section 3.1.3. In particular, the WaterNet outperformed L8SR + S1A + sl in terms of a higher F1 score and greater cloud insensitivity as shown in Figure 5 and 6. The WaterNet also outperforms U-Net, the baseline NDWI and exceeds all the performance in Konapala et al. (2021). It was also shown that WaterNet performs well in elevated areas (Figure 4), overcoming the backscattering problem mentioned by Song et al. (2007).

4.2 WaterNet: Web Applications

The WaterNet was also applied to other applications such as the recent Bangladesh flood in 2022 and the drought events caused by the European heatwave in 2022. More applications of WaterNet are accessible at <https://geeiimperial.herokuapp.com/> and were created by integrating GAE and React-built web applications. Creating an app similar to Brown et al. (2022) to enable near-real-time prediction is adaptable from the current development but the running cost of deploying the model to GCP prevented completion. However, future work should secure funds to create an app like Brown et al. (2022).

Users interested in using the algorithm can generate predictions from GC notebooks. The current algorithm can predict water bodies near-real-time by using Landsat-8 collection 1 and 2. For future works, ESA (2021) recommends that Landsat-8 collection 2 should be used to train the models

because Landsat-8 collection 1 mission ceased after December 2021.

4.3 Influence of resolution on WaterNet performance

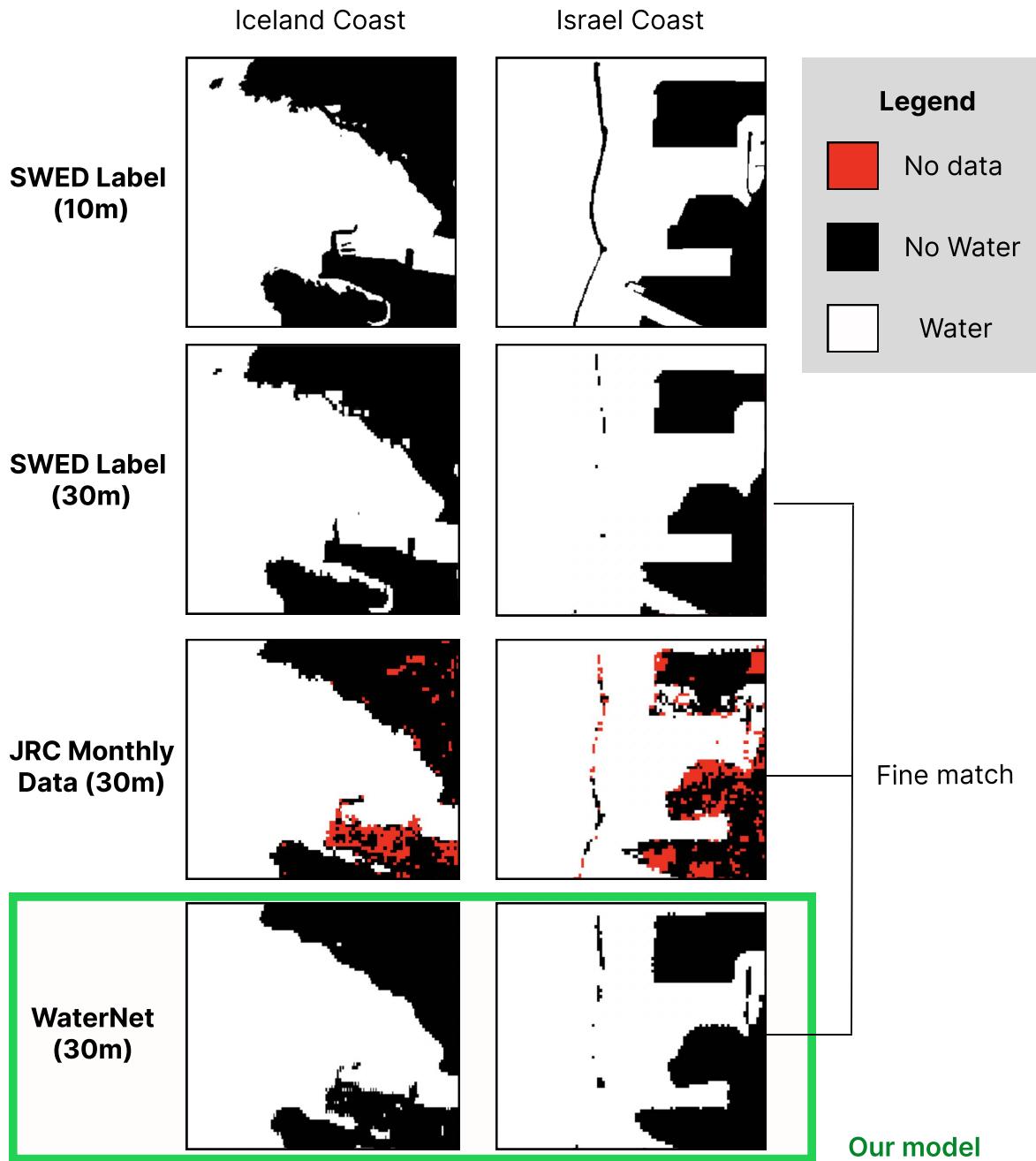


Figure 7: Testing the resolution of WaterNet by detecting a coastline in Iceland and Israel comparing the SWED at 30m and 10m, JRC monthly data and WaterNet in 2019

WaterNet was shown to detect large rivers and lakes well, but the ability to detect small water details is crucial for applications such as drought detection in small rivers, i.e. with a width smaller than the 30m resolution available in the data products used here. WaterNet's ability to detect small water details was studied by assessing the predictive ability of WaterNet in coastline areas such as Iceland and Israel during 2019 in Figure 7. The 10m Sentinel-2 Water Edges Dataset (SWED) (Seale 2019)

created by labeling Sentinel-2 data in the cloud-free period during 2019 is shown in the first row of Figure 7. The dataset was resampled to 30m to give a fair comparison to the 30m resolution WaterNet as seen in the second row of Figure 7. WaterNet was trained the whole year in 2019 and is shown on the last row. JRC monthly label data was chosen in a month with least cloud cover and no-data in 2019 as a second verification on the WaterNet performance. The WaterNet cannot capture small details at 10m, but captures most details at 30m in SWED and JRC Monthly labels. Although WaterNet can detect details up to 30m, the resolution of WaterNet can be improved with a higher resolution of labeling and training data. Future work should train the model with either Sentinel-2, Planet Scope or Rapid Eye (Planet 2022) to match the increasing resolution of labeling data.

4.4 Limitations and improvement of the 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. Only 720 patches were used in Section 2.1 because of the financial and computational aspects of performing 154 experiments.

Extending the coverage of training data to cover all the points on the globe may improve the model's performance. It was shown that training the model in the same geographical region rather than a few points around the globe will improve the F1 score as described in Section 3.1.2. Training data in the same area helped SentSlope, a combination of the WaterNet, better predict water bodies in the same terrain.

Further tuning the hyperparameters, such as the learning rate and the loss functions would improve WaterNet's performance. This work briefly looked at the loss functions of CC and CC-DICE, but other loss functions such as the Sobel-Edge loss or Focal loss should be explored (Jadon 2020, Bai et al. 2021, Seale et al. 2022, Seale 2019). Especially, Sobel-Edge loss was shown to detect fine details better than CC (Seale et al. 2022).

5 Conclusions

Multi-view learning approach with U-Net cores was developed to create water classification maps by fusing the Sentinel-1, Landsat-8 and NASADEM imagery on the JRC monthly data. By exploring different bands of L8SR and S1 satellites with NASADEM the best model consisting of two U-Net cores named the 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 in mapping 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. Extension to the approach should incorporate additional training data with higher resolution and hyperparameter tuning is left for future work.

Overall, governments, emergency services and researchers can use the methods developed to quickly evaluate water bodies in an area of one million hectares in events such as flooding or drought. Creating an app similar to Brown et al. (2022) is technically feasible by leveraging the code architecture developed here, but running costs prevented the completion. The application of the model was instead deployed to the freely available React-built web application called WaterOracle, using Google Cloud Platform.

References

- Acharya, T. D. & Yang, I. (2015), 'Exploring landsat 8', *International Journal of IT, Engineering and Applied Sciences Research (IJIEASR)* **4**(4), 4–10.
- 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: <http://doi.org/10.1109/IGARSS.2009.5416921>
- 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.
URL: <https://doi.org/10.3390/rs13112220>
- Bangira, T., Alfieri, S. M., Menenti, M. & Van Niekerk, A. (2019), 'Comparing thresholding with machine learning classifiers for mapping complex water', *Remote Sensing* **11**(11), 1351.
URL: <https://doi.org/10.3390/rs11111351>
- 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://doi.org/10.1038/s41597-022-01307-4>
- 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.
URL: [https://doi.org/10.1016/S0074-6142\(10\)09914-6](https://doi.org/10.1016/S0074-6142(10)09914-6)
- 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'.
URL: <https://doi.org/10.5194/isprs-archives-XLI-B4-125-2016>
- 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>
- Gulácsi, 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://doi.org/10.3390/rs12101614>
- 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.
URL: <https://doi.org/10.48550/arXiv.2006.14822>
- 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.
URL: <https://doi.org/10.1088/1757-899x/592/1/012088>
- 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://doi.org/10.1016/j.isprsjprs.2021.08.016>
- Kousky, C. & Kunreuther, H. (2009), 'Improving flood insurance and flood risk management: Insights from st. louis, missouri'.
URL: <http://doi.org/10.2139/ssrn.1352466>
- Kseňák, L., Pukanská, K., Bartoš, K. & Blišt'an, 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://doi.org/10.3390/w14030299>
- 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://doi.org/10.1016/j.jag.2021.102571>
- 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.
URL: <https://doi.org/10.1109/JSTARS.2020.3033424>
- 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.
URL: <https://doi.org/10.3390/rs12152469>
- 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://doi.org/10.1016/j.pholo.2021.100005>
- Muñoz, D. F., Muñoz, P., Moftakhari, H. & Moradkhani, H. (2021), 'From local to regional compound flood mapping with deep learning and data fusion techniques', *Science of the Total Environment*

782, 146927.

URL: <https://doi.org/10.1016/j.scitotenv.2021.146927>

Özelkan, E. (2020), 'Water body detection analysis using ndwi indices derived from landsat-8 oli', *Polish Journal of Environmental Studies* **29**(2), 1759–1769.

URL: <https://doi.org/10.15244/pjoes/110447>

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://doi.org/10.1038/nature20584>

Piriyajitakonkij, M., Warin, P., Lakhan, P., Leelaarporn, P., Kumchaiseemak, N., Suwajanakorn, 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.

URL: <https://doi.org/10.1109/JBHI.2020.3025900>

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.

URL: <https://doi.org/10.1109/TPAMI.2009.187>

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://doi.org/10.48550/arXiv.1505.04597>

Rudner, T. G., Rußwurm, M., Fil, J., Pelich, R., Bischke, B., Kopačková, V. & Biliński, P. (2019), Multi3net: segmenting flooded buildings via fusion of multiresolution, multisensor, and multitemporal satellite imagery, in 'Proceedings of the AAAI Conference on Artificial Intelligence', Vol. 33, pp. 702–709.

URL: <https://doi.org/10.48550/arXiv.1812.01756>

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.

URL: <https://doi.org/10.1016/j.rse.2022.113044>

Seale, Redfern, C. (2019), 'Sentinel-2 water edges dataset (swed)', <https://openmldata.ukho.gov.uk/>, note = "[Online; accessed 23-Aug-2022]" .

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://doi.org/10.14358/PERS.73.3.285>

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://doi.org/10.1109/IGARSS.2012.6351195>
- 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://doi.org/10.3390/w14091454>
- 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.
URL: https://doi.org/10.1007/978-3-319-99867-1_114
- 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.
URL: <https://doi.org/10.1016/j.rse.2020.111741>
- 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]".
- 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.
URL: <https://doi.org/10.1038/nclimate1908>
- 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://doi.org/10.1016/j.inffus.2017.02.007>

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