UNIVERSITY OF THE PELOPONNESE & NCSR "DEMOCRITOS"
MSC PROGRAMME IN DATA SCIENCE

Flood Mapping Using Satellite Images

by

Konstantinos Fokeas

A thesis submitted in partial fulfillment
of the requirements for the MSc
in Data Science

Supervisor: Eleni Charou
Researcher (C)

Co-supervisors: Theodoros Giannakopoulos, Anastasia Krithara
Researcher (B), Researcher (D)

Athens, January 2023

Flood Mapping Using Satellite Images

Konstantinos Fokeas

MSc. Thesis, MSc. Programme in Data Science

University of the Peloponnese & NCSR “Democritos”, January 2023

UNIVERSITY OF THE PELOPONNESE & NCSR "DEMOCRITOS"
MSC PROGRAMME IN DATA SCIENCE

Flood Mapping
Using Satellite Images

by

Konstantinos Fokeas

A thesis submitted in partial fulfillment
of the requirements for the MSc
in Data Science

Supervisor: Eleni Charou
Researcher (C)

Co-supervisors: Theodoros Giannakopoulos, Anastasia Krithara
Researcher (B), Researcher (D)

Approved by the examination committee on January, 2023.

(Signature)	(Signature)	(Signature)
.....	.....	.....
Eleni Charou Researcher (C)	Theodoros Giannakopoulos Researcher (B)	Anastasia Krithara Researcher (D)

Athens, January 2023

UNIVERSITY OF THE PELOPONNESE & NCSR "DEMOCRITOS"
MSC PROGRAMME IN DATA SCIENCE

Declaration of Authorship

(1) I declare that this thesis has been composed solely by myself and that it has not been submitted, in whole or in part, in any previous application for a degree. Except where states otherwise by reference or acknowledgment, the work presented is entirely my own.

(2) I confirm that this thesis presented for the degree of Master of Science in Data Science, has

   (i) been composed entirely by myself

   (ii) been solely the result of my own work

   (iii) not been submitted for any other degree or professional qualification

(3) I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or processional qualification except as specified.

(Signature)

.....

Konstantinos Fokeas

Athens, January 2023

Acknowledgments

Words cannot express my gratitude to my supervisor Eleni Charou for her invaluable patience and feedback. I also could not have undertaken this journey without my defense committee, who generously provided knowledge and expertise. Additionally, this endeavor would not have been possible without the generous support of Ioannis Vernikos and George Gianopoulos. Lastly, I would be remiss in not mentioning my family, especially my parents (Panagiotis & Eleni), my brother (Sotiris), and my life partner (Nicole). Their belief in me has kept my spirits and motivation high during this process.

Περίληψη

Σκοπός της παρούσας διπλωματικής εργασίας είναι η μελέτη του αντικειμένου της χαρτογράφησης πλημμυρών με τη χρήση εικόνων που λαμβάνονται από δορυφόρους και αλγόριθμους μηχανικής μάθησης. Από την έναρξη του προγράμματος Κοπέρνικος του Ευρωπαϊκού Οργανισμού Διαστήματος (ESA) και των δορυφόρων Sentinel, ένας τεράστιος αριθμός ελεύθερα διαθέσιμων εικόνων λαμβάνεται καθημερινά, επεκτείνοντας τις πιθανές εφαρμογές. Με την εκτόξευση των Sentinel 1 και Sentinel 2 οι οποίοι εποπτεύουν τον πλανήτη Γη με πρωτοφανή συχνότητα και χωρική ανάλυση, οι επιστήμονες και οι μηχανικοί μπορούν πλέον να αναπτύξουν εργαλεία για να κατανοήσουν τις διαδικασίες της Γης και να λάβουν πιο τεκμηριωμένες αποφάσεις. Μία από αυτές τις γήινες διαδικασίες είναι οι πλημμύρες, μια από τις πιο καταστροφικές φυσικές καταστροφές που επηρεάζουν πολλούς ανθρώπους κάθε χρόνο, προκαλώντας θανάτους, ζημιές και απώλειες περιουσιών. Προκειμένου να μετριαστούν οι επιπτώσεις των πλημμυρών απαιτείται λήψη κρίσιμων αποφάσεων η οποία μπορεί να υποστηριχθεί από την χρήση δορυφορικών εικόνων και μεθόδων μηχανικής μάθησης. Η παρούσα μελέτη εξετάζει την απόδοση τριών διαφορετικών μεθόδων μηχανικής εκμάθησης στον εντοπισμό πλημμυρισμένων εκτάσεων σε επίπεδο εικονοστοιχείου. Ειδικότερα, εξετάζεται η μέθοδος βαθιάς μάθησης βασισμένη στην αρχιτεκτονική UNET, η μάθηση μέσω της τεχνικής μεταφοράς γνώσης καθώς και μία παραδοσιακή μέθοδος βασισμένη στα δέντρα αποφάσεων. Τα πειράματα περιλαμβάνουν εκπαίδευση μοντέλων είτε μέσω *αυστηρής* είτε μέσω *ασθενούς* (weakly) επίβλεψης καθώς και πολυτροπικούς χώρους χαρακτηριστηκών (multimodal feature space). Τέλος οι τεχνικές μηχανικής μάθησης συγκρίνονται ως προς την απόδοση με μία τεχνική βασιζόμενη στην τμηματοποίηση του ιστογράμματος της εικόνας, η επονομαζόμενη ώς μοντέλο βάσης (baseline model).

Λέξεις-κλειδιά: δορυφορική τηλεπισκόπηση, χαρτογράφηση πλημμυρών, αντιμε-

τώπιση καταστροφών, δορυφορικές εικόνες, εποπτευόμενη σημασιολογική κατάτμηση.

Abstract

The aim of this thesis is to study the subject of flood mapping utilizing images captured from satellites and machine learning algorithms. Since the rise of ESA's Copernicus program and the consecutive Sentinel satellites a vast number of freely available images are captured every day expanding the potential applications. With the launch of Sentinel 1 and Sentinel 2 sensing the planet Earth in an unprecedented frequency and spatial resolution, scientists and engineers can now develop tools in order to understand the processes of the Earth and make more informed decisions. Floods are one of the most devastating natural disaster affecting many people each year, causing a lot of deaths, infrastructure damages and loss of properties. In order to mitigate the effects of floods on people. critical decision making is needed, which can be assisted by satellite images and machine learning methods. This study examines the performance of three different machine learning methods in identifying pixels in satellite images containing flooded areas. More specifically, the three tested methods are based on deep learning architecture, transfer learning and traditional swallow learning pixel based semantic segmentation, consequently. In particular, the deep learning method based on the UNET architecture, transfer learning using as backbone the VGG16 network and a traditional method based on decision trees. Experiments involve training models either through strict or through weak supervision as well as multimodal feature spaces, combining sentinel 1 and sentinel 2. Finally, the machine learning techniques are compared in terms of performance with a technique based on the segmentation of the histogram of the image, called as baseline model.

Keywords: remote sensing, flood mapping, disaster response, satellite images, supervised semantic segmentation.

Contents

List of Tables

List of Figures

List of Abbreviations

RS	Remote Sensing
ML	Machine Learning
RF	Random Forest
EO	Earth Observation
SAR	Synthetic Aperture Radar
NIR	Near Infrared
SWIR	Short Wave Infrared
S1	Sentinel 1
S2	Sentinel 2
DL	Deep Learning
TL	Transfer Learning

Chapter 1

Introduction

The identification of permanent and temporal water segments in a flooded area has mainly relied on change detection methods utilizing multitemporal imageries. The problem that still remains unsolved is the ability to identify and distinguish water type in flood events using only a single post-disaster remote sensing image. Due to the high availability of satellite images, in the last decade, a lot of research has been made in flood mapping. In general the problem of distinguishing water from non water elements is mainly based on histogram thresholding approaches which are highly affected by the geography, time and atmospheric conditions at the time the images were captured. Therefore, the generalization ability in threshold based methods is greatly limited. One of the most challenging tasks in flood disasters is to distinguish permanent water from temporal water. The identification of temporary water relies on multi-temporal change detection methods which require at least one pre-event image, which is a significant limitation [1].

Rapid response to natural hazards is crucial in mitigation actions for life and property losses. Emergency response teams require timely and accurate data in order to form critical decisions. Satellite imagery offers a significant amount of information for regions affected by a disaster in terms of rapid mapping. Currently, much of the flood analysis is manual or semi-automated, and carried out by experts from a range of organizations [4].

Some interesting facts about floods on a global scale can be summarized in the

following list [6].

- Floods cause more than 40$ billion per year in damages worldwide.
- 40% of the world's population lives close to coasts.
- Flooding events are on the rise due to climate change, increasing sea levels and extreme weather events.
- Flood level estimation needs to be done remotely as physical access to flooded areas is limited.
- Deploying instruments in potential flood zones can be dangerous.

1.1 Problem Description

Flooding alone accounted for 47% of all weather-related disasters during the decade 1995-2015, affecting 2.3 billion people, the majority of whom (95%) live in Asia. Since 1995, floods have accounted for 47% of all weather-related disasters (Figure 1.1), affecting 2.3 billion people. The number of floods per year rose to an average of 171 in the period 2005-2014, compared to the annual average of 121 incidents during the last decade [7].

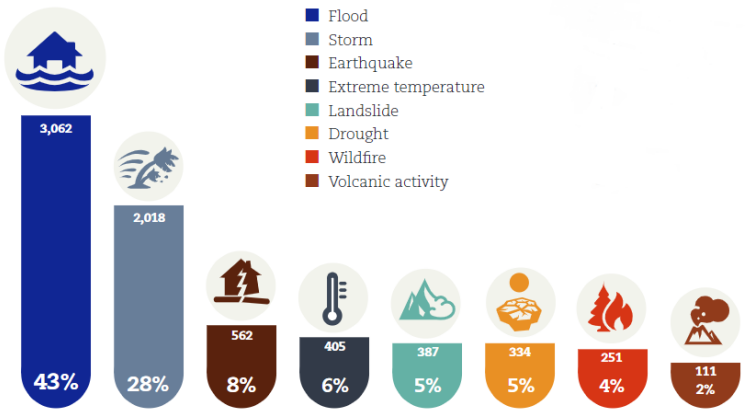

Figure 1.1: Weather-related disasters

The nature of disastrous floods has also changed in recent years, with flash floods, acute riverine and coastal flooding increasingly frequent. In addition, urbanization has significantly increased flood run-offs, while recurrent flooding of agricultural land, particularly in Asia, has taken a heavy toll in terms of lost production, food shortages and rural under-nutrition.

Many authors and researchers over the years have proposed various types of classification of flood types. In the following section the most prominent and most comprehensive list of six flood types is illustrated [8].

- **Flash floods** are one of the most often occurring in urban areas and one of the most catastrophic for human lifes and infrastructures and properties. They are caused by heavy rainfall or rapid snow thaw. This type of flood can occur with little to no notice and even though the relatively small area they cover can drift large objects like cars and trees.
- **Coastal floods** are caused exclusively by strong winds directed towards the coast during high tide. Coastal areas within lower elevation are the most affected. The following picture illustrates such an example where coastal settlements are devastated by this type of flood.
- **River floods**. The distinct characteristic of river floods is the gradual riverbank overflow caused by extensive rainfall over a period of time. The riparian areas covered by this type of flood depend on the size of the river and the amount of excessive rainfall. Floods of this type mainly because of their slow evolution rarely result in loss of human lives but can cause immense economic losses.
- **Urban floods** occur when the drainage system of a city fails to handle the excessive water coming from heavy rain. Additionally, the lack of natural drainage in urban areas due to impermeable materials can also contribute to urban floods. Although water levels can be just a few centimeters higher from the ground, this type of flood can cause major damages in the structures.
- **Pluvial flooding** is very similar to urban flooding but it occurs mainly in rural areas affecting mostly agricultural activities and properties. This type of floods form in flat areas where the terrain can't absorb the excessive water from rain, causing puddles and ponds.
- **Dam and dyke breach floods** are caused by structural failures due to extreme events and insufficient management.

To mitigate the impact of floods on human lives and properties, both preventive and emergency measures are required (European Union, 2007). Preventive measures include policy measures in order to reduce the possibility of a flood event while

emergency require operations carried out before, during and after the flood event. Both of them can be determined by maps that indicate potential hazards, the extent of the flood and areas that are in danger. There are three main flood maps used for dealing with such measures focusing on the spatial variability of floods, namely flood susceptibility, flood inundation, and flood hazard maps [9]. The following images and paragraphs illustrate an example of these kinds of maps [8] [9].

Flood susceptibility maps (Figure 1.2) determine the tendency of flooding in a certain area based on its physical characteristics. Particularly, flood susceptibility mapping considers the topographical, geographical and meteorological features and the correlation between the spatial distribution of past flood events. This is done with multivariate analysis and multi-criteria decision analysis and the product is a qualitative map.

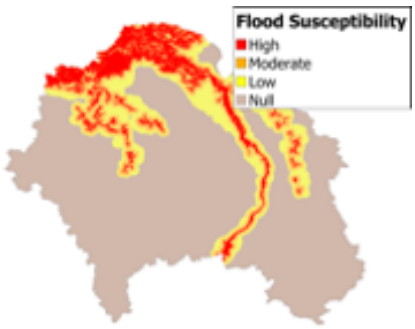

Figure 1.2: Flood Susceptibility Map

Flood inundation maps (Figure 1.3) determine the extent of a flood during or after the event. This type of maps generally include two thematic classes namely flooded and non-flooded areas. It is used for post-flood evacuation, urban planning and damage assessment. Remotely sensed images are fed into a statistical model, mostly based on a threshold value, in order to define the pixels including flooded areas.

Flood hazard maps (Figure 1.4) measure the water depth and extent across a flooded area. Flood hazard maps are carried out by numerical models, which simulate flood events by discretizing the governing equations and the computational domain. They are particularly useful for exceptional events such as tsunamis and dam breaks. However, they are computationally demanding and thus less used than

Figure 1.3: Flood Inundation Map

the other models.

Figure 1.4: Flood Hazard Map

The current study is focused on inundation maps utilizing remote sensing satellite images.

1.2 Thesis Objectives and Contributions

The rest of the master thesis is organized as follows. In Chapter 2, the literature review that took place is listed with the most prominent publications in the field of flood mapping, while also the basic principles of satellite remote sensing are demonstrated and explained. In Chapter 3, the used dataset is defined and further explained along with the theoretic basis of the three exploited machine learning architectures namely U-NET, Random Forest and Transfer Learning, along with the baseline model. In Chapter 4, the experimental results are demonstrated starting with the necessary preprocessing steps applied on the initial dataset. Finally in Chapter 5 the conclusion of the study are critically discussed giving insights for future experimentation.

The main objectives of this master thesis can be summarized as follows:

1. Test three different machine learning, namely U-NET, Random Forest and Transfer Learning, approaches on their efficiency to detect and distinct flooded pixels from satellite images.
2. Compare machine learning approaches against a baseline model based on histogram segmentation.
3. Examine and compare the effectiveness of optical versus radar images on flood mapping.
4. Examine the performance of models trained on weakly labeled images and semi-supervised environment.
5. Examine and compare multi modal feature spaces against single or uni modal feature spaces, combining sentinel-1 and sentinel-2.

Chapter 2

Background on Remote Sensing and Flood Mapping

2.1 Literature Review

Many papers have been published in the domain of flood mapping using satellite images, but in this thesis we are concentrating on studies utilizing machine learning techniques, mostly based on deep learning architectures. In the following section five summaries of research papers from the international literature are presented.

2.1.1 Flood Detection in Time Series of Optical and SAR Images C. Rambour, et al.

In this study it is recognized a research gap in the application of deep learning architectures on SAR images mainly due to the lack of labeled data. To tackle this issue they introduce a new dataset named SEN12-FLOOD composed of co-registered optical Sentinel 2 and SAR Sentinel 1 images in the form of time series. The study areas correspond to African, Iranian and Australian city centers with or without a flood event, occurring during the sensing period. Each image has a binary label specifying whether a flood event is visible or not. The labels have been provided by the MediaEval 2019 dataset and were obtained from Copernicus Emergency Management Service. The Sentinel 1 images were acquired in Interferometric Wide Swath (IW)

mode at polarization VV and VH and delivered in a spatial resolution of 10x10 meters. Initial products were processed before the analysis, including radiometric calibration and terrain corrections.

The dataset is composed of 412 time series with 4 to 20 optical images and 10 to 58 SAR im- ages in each sequence. The period of acquisition goes from December 2018 to May 2019, while a flood event is occuring in 40% of the optical Sentinel 2 images and in 47% of the SAR Sentinel 1 images. The network that was used is the ResNet-50 which is designed to process only RGB images thus, the first convolutional layer had to be modified to take into account the correct number of bands for multispectral and sar images. One of the experiments was to consider only the spatial and spectral features of the dataset with the optical images to illustrate the most promising results. In order to take into account the temporal dimension, the authors proposed to feed the extracted features from ResNet into a Gated Recurrent Unit (GRU). For multimodal classification the features from SAR and RGB ResNets are concatenated and fed to the GRU layer. The output of this process is a sequence of binary labels indicating whether there is flood or not for every time frame. The architecture can be seen in the following figure.

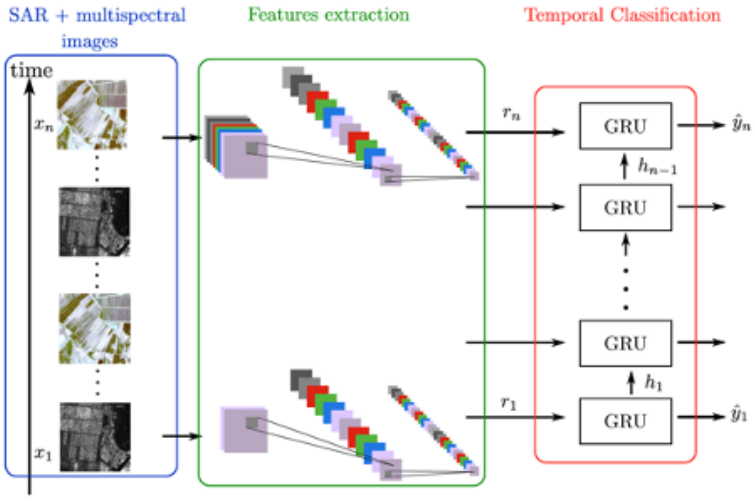

Figure 2.1: The proposed architecture.

The major findings is that deep learning techniques perform well in this type of task given both optical and SAR images. Furthermore, the implementation of GRUs in order to capture the temporal variations is a key parameter in modeling

flood events, leading in a significant error reduction, using both modalities in a complementary way. The behavior of the flood may differ greatly from one area to another, while open water areas appear clearly in SAR images, flooded vegetation or soaked ground areas are harder to discriminate from dry areas. More specifically, It appears that on open water areas the detection using sar data is close or even better than with optical images, whereas when the area is occluded by vegetation, optical images offer better results. The accuracy achieved by the proposed architecture can be seen in the table below, where in the first row is illustrated the flood detection task on each image while in the second row accuracy is given by a recurrent network on the sequence of image features.

Model	Data	Accuracy
Resnet-50	RGB	0.885
	Multispectral	0.793
	SAR	0.753
Resnet-50 + GRU	RGB	0.930
	SAR	0.875
	SAR+RGB	**0.957**

Figure 2.2: Accuracies achieved

Finally the authors are mentioning that future efforts should be given in designing a network, like attentional models, which will be able to learn specific behaviors of the input data given that depending on the type of land cover different results are achieved using optical or sar images.

2.1.2 Sen1Floods11:a georeferenced dataset to train and test deep learning flood algorithms for Sentinel-1 D. Bonafilia, et al.

The Sen1Flood11 dataset is introduced which includes Sentinel 1 and Sentinel 2 imagery along with the ground truth labels for permanent and flooded water. The dataset consists of 4,831 of 512x512 patches of images covering in total an area equivalent of 120,406 Km2 spanning all 14 biomes and 6 continents of the world

across 11 distinct flood events.

Results of fully convolutional neural networks on classifying permanent, flood and total surface water on four subsets of the Sen1Floods11 dataset. More specifically the dataset is consisted of i) 446 hand labeled patches of surface water from flood events, ii) 814 patches of permanent water from JRC used as ground truth labels, iii) 4,385 patches of surface flood water from Sentinel-2 images, iv) 4,385 patches of of surface flood water classified from Sentinel 1 images. The results are compared to a common approach of thresholding radar backscatter to segment water surfaces.

The contribution of this study can be manifold with the dataset serving a benchmark for future studies and the examination of four research questions dealing with the improvement of flood detection and the operationalization of CNNs for global flood mapping. The research questions are:

1. Do we need hand-labeled training data to train CNNs to detect flood water or can we use weakly supervised training data derived from re- mote sensing water detection algorithms?
2. Which imagery sources and algorithms provide the best labels for weakly supervised training?
3. What is the impact on model performance when a CNN is trained on permanent water data only as compared to training data that included flood events?
4. Do CNNs identify flood and/or permanent water in radar data more accurately than conventional remote sensing methods such as backscatter thresholding?

2.1.2.1 Sampling Permanent Water Data

The permanent water raster images came from the Surface Water Dataset released by the European Commission's Joint Research Center. This dataset includes monthly observations of surface water at 30 meter spatial resolution using Landsat acquisitions. From this dataset was extracted only the water, non-water samples from the transition layer which identifies as permanent water the pixels that were observed to have water presence at both the beginning (1984) and the end (2018) of the study period. The non-water label was adopted for pixels that were never observed as

water.

2.1.2.2 Sampling Flood Event Data

The flood events that the SEN1floods11 dataset contains were identified from a global database of flood events from the Dartmouth Flood Observatory. The events were selected using the criterion of Sentinel-1 and Sentinel - 2 imagery acquired on the same day or within 2 days difference. In total, 5 events had coincident imagery, 4 had imagery within 1 day difference and 2 had imagery with 2 days time difference.

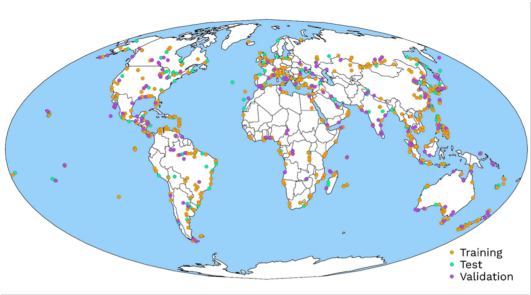

Figure 2.3: Locations from where flood event data was sampled

2.1.2.3 Data Pre-processing

Reference flood maps were created from Sentinel - 1 and Sentinel - 2 respectively. Reference Sentinel 1 flood maps were derived from VH band dividing the images into 1 Km x 1Km high variance grids and for each of them the histogram was extracted. In a later stage the OTSU thresholding algorithm was utilized on the histogram resulting in a binary flood map. Reference Sentinel 2 flood maps were created utilizing the Normalized Difference Vegetation Index (NDVI) and the Modified Normalized Difference Water Index (MNDWI). A photointerper expert defined a threshold of 0.2 and 0.3 for NDVI and MNDWI respectively. Regarding the clouds and cloud shadows the first were identified by a threshold on the blue band reflectance and the latter were removed exploiting information about the potential cloud heights, the solar azimuth angle and solar zenith angle.

For each flood event a subset of the image was selected in order to sample regions predominantly affected by the flood. The resulting images were further divided into

512 x 512 pixel non overlapping patches. The patches covered with clouds were removed and the area of Sentinel 2 water was calculated for each patch. A stratified sample of 336 patches were selected for hand labeling, while the rest 4.385 patches were exported for training, validation and testing.

The aforementioned hand labeling took place within Google Earth Engine utilizing a custom GUI for trained remote sensing analysts to label the water areas. Those analysts had access to Sentinel -1 VH band, two false color composites from Sentinel 2 and the reference water classification from Sentinel 2.

2.1.2.4 Training, Validation, and Testing Data

All of the hand labeled chips of Bolivia were held out in order to evaluate the performance of the trained models on flood events never seen before. Consequently Sentinel 2 reference maps for Bolivia were also withheld from training and validation sets, but Sentinel 1 based flood maps for weakly supervised training data for Bolivia were included. Apart from the Bolivia data special treatment the hand labeled data were splitted into training, validation, testing with a random 60-20-20 split while the non-hand labeled Sentinel 1 and Sentinel 2 data were used specifically for weakly supervised training.

2.1.2.5 Convolutional Neural Network Models

Having defined all of the training sets four models were built based on Fully Convolutional Neural Networks (FCNN). More specifically, one model was built based on weakly supervised training data using Sentinel 1 based flood classification as labels, one model based on weakly supervised Sentinel 2 flood maps, one model trained using hand-labeled flood classification maps. Lastly the fourth model trained on the JRC permanent water dataset which is produced from Landsat 8 data. In total four fully connected neural networks were trained and tested on each of the training datasets described before and compared to a backscatter thresholding algorithm. The backscatter thresholding model used Otsu thresholding on the VH band.In order to assess the transferability of the proposed architecture on permanent water

detection to flood water detection an evaluation of each model to identify permanent water, flood water, and total surface water took place.

2.1.2.6 Accuracy Assessment

All models were trained on PyTorch and for the prediction of water in each pixel used a fully convolutional network with a Resnet50 backbone. No extensive hyperparameter tuning took place but data augmentation with random image crops, horizontal and vertical flips was applied. Intersection over union was the evaluation metric with the final results being illustrated on the following tables.

Dataset	PW	FW	AW
Sentinel-1 Weak	.2872	.2422	.0392
Sentinel-2 Weak	.3818	.3389	.4084
Hand-Labeled	.2570	.2421	.3125
Permanent Water	.3391	.1693	.2452
Otsu Threshold-VH	.4571	.2850	.3591

Performance on the hand-labeled test set of 10 flood events (all besides Bolivia) of models trained on each dataset in terms of Mean IOU for the water class. Results shown on permanent water (.PW), flooded water (FW) and all water (AW)

Dataset	PW	FW	AW
Sentinel-1 Weak	.2506	.3296	.3871
Sentinel-2 Weak	.1946	.2738	.3160
Hand-Labeled	.2300	.2905	.3524
Permanent Water	.2881	.2684	.3422
Otsu Threshold-VH	.2859	.3239	.3862

Table 2: Performance on the hand-labeled test set of the flood event in Bolivia of models trained on each dataset in terms of Mean IOU for water class. Results shown on permanent water (PW), flooded water (FW) and all water (AW)

2.1.2.7 Discussion and Conclusion

Hand-labeled training data is not necessary to train FCNNs to detect flood water. Sentinel-2 provides better automatic labels for Sentinel-1 based flood detection. FCNNs trained on flood water perform better than those trained on permanent water alone. FCNNs outperform thresholding algorithms to identify flooded but not permanent water.

The writers claim that accuracy gains could be achieved by different data augmentation schemes and extended hyperparameter tuning. Fully connected convolutional neural networks could be compared by other machine learning algorithms and improved thresholding methods could be tested. The validation dataset does not include any urban flood event mainly because Sentinel 1 algorithms are not optimal for mapping floods in urban areas. The writers encourage others to expand the dataset to include urban flood events and radar information such as interferometry and change in coherence which has shown to have promising results in mapping urban floods.

2.1.3 Urban flood mapping with an active self-learning convolutional neural network based on TerraSAR-X intensity and interferometric coherence, Yu Li, et al.

In this study the roles of SAR intensity and interferometric coherence are being assessed in urban flood detection, using multi-temporal TerraSAR-X data. The demonstrated method in this paper is based mainly in an active self-learning convolutional neural network framework, which is independent of the effect of limited annotated data. The study area is located in the city of Houston in the US state of Texas. Houston was affected by floods associated with heavy rainfall that accompanied Hurricane Harvey in August 2017. The city represents a typical urban landscape that is mainly covered by dense housing and apartments, as well as commercial and industrial areas with schools, warehouses, stadiums, parks and parking lots.

The main idea of this study relies on the properties of SAR images in urban environments. The backscatter of synthetic aperture radar in urban areas consists of specular reflection, surface backscatter, as well as single, double and triple light reflections (bounces). Due to the different backscatter mechanisms, flooded urban areas may appear differently in SAR-intensity images. In general, flooded water can appear in either a darker or a lighter color tone, depending on the difference of the backscattered energy between flooded and non-flooded surfaces. Figure 2.4 illustrates the intensity and coherence variation for different types of covered surfaces under flooded or non-flooded conditions in TerraSAR-X data and related visual reference data. Flood mapping in complex urban areas based on SAR intensity alone is a major challenge. The interferometric coherence, which indicates the correlation of two complex observations (amplitude and phase information), provides additional information for urban flood mapping as an urban settlement can generally be considered as a fixed target characterized by high coherence. Cohesion variation makes flooded built-up areas distinguishable from non-flooded ones.

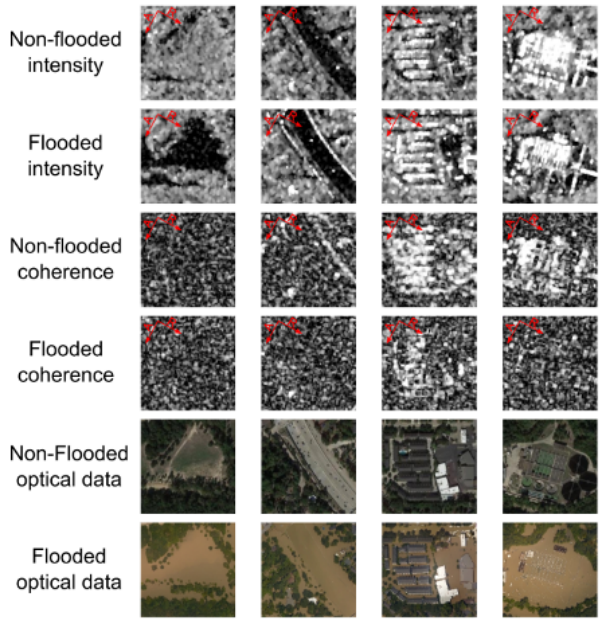

Figure 2.4: Intensity and coherence variation for different types of covered surfaces under flooded or non-flooded conditions in TerraSAR-X data and related visual reference data.

Concerning the above figure 2.4 the non-flood visual data retrieved from Google Earth and the flood optical data is sourced from NOAA (National Oceanic and Atmospheric Administration): (a) meadow (b) roads (c) low floodplain buildings (

d) buildings with a high level of flood water.

The proposed framework aims to process both labeled and unlabeled specimens through the integration of active learning and self-learning at the perceptual level under the model of deep CNN time synthesis. More specifically, the proposed framework operates repeatedly based on 2 steps: a) training and retraining of CNN time synthesis b) informational updating of unlabeled samples by placing pseudo-labels on the training data.

First, a deep CNN model, called a "student," is trained with the training samples bearing the original labels. At the same time, the creation of a "teacher" model is achieved by synthesizing the parameters of the "student" model during the training steps.

Unlabeled data samples, which contain information, are then checked or selected following disagreements between the "student" and "teacher" models. Assuming that the adjacent spatial samples belong to the same class, the selected samples are filtered and self-labeled via a multi-scale spatial constraint. Additionally, consistency regularization is introduced to compress errors in pseudo-labels.

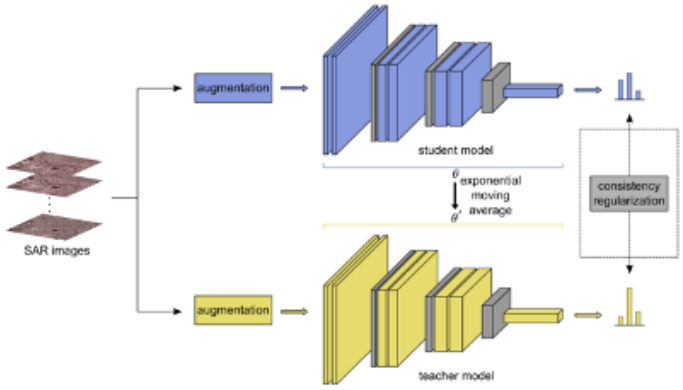

Figure 2.5: The structure of the CNN temporal composition model.

In the present study, four data sets were taken: TerraSAR-X HH-polarized Stripmap: 1 image before the flood event (August 10, 2017), 1 image shortly after the flood (September 1, 2017) and 2 images well after the flood (October 26, 17 and 28 November 2017). The spatial resolution (m) of TerraSAR-X images is 1.2 x 3.3 (range x azimuth). Additionally all experimental results were extracted based on very high spatial optical data of approximately 35 cm, taken on 30 and

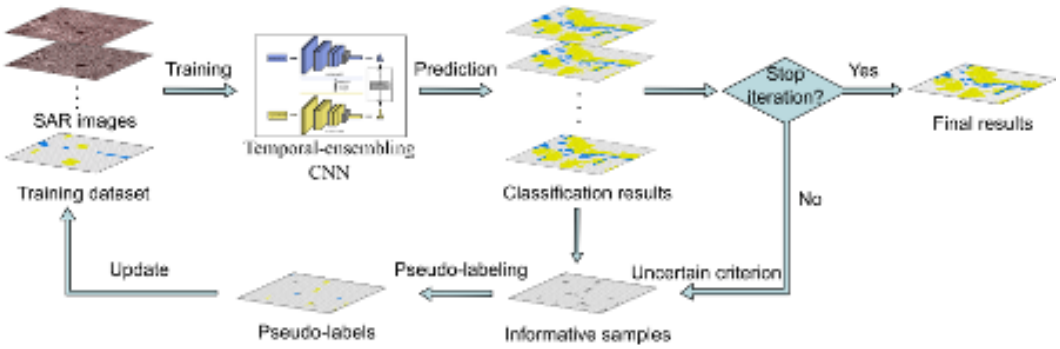

Figure 2.6: The context of active self-learning.

31 August 2017 using a Trimble Digital Sensor (DSS) gas system provided by the NOAA Remote Sensing Department.

The raw TerraSAR-X data were calibrated and transformed into decibels (dB) and a 5x5 Lee Sigma filter was applied to each image to reduce speckle noise. All intensity and coherence data were co-registered and geocoded in the WGS1984 UTM 15N zone with a pixel distance of 1.25 m and a size of 4,800 x 6,400 pixels. The pre-processed images were splitted into non-overlapping patches 32 x 32 each. The patches were classified into 3 classes: Open flooded areas absence of buildings- (FO class) Flooded areas with buildings (FB class) Non-flooded areas (NF class) A total of 30,000 image patches were created a) 1,130 class FO b) 2,500 class FB c) 26,370 class NF.

The main conclusions of this study can be summarized as follows: Multi-temporal intensity plays the most important role in urban flood mapping and makes it possible to outline the exact pattern of distribution of flooded areas. Adding multi-temporal coherence to the multi-temporal intensity can significantly improve classification accuracy, as it makes flooded areas indistinguishable from non-flooded areas. The proposed methodology framework is generalized and could be used in other classification applications with different types of remote sensing data and not necessarily SAR data.

2.1.4 Fully Convolutional Neural Network for Rapid Flood Segmentation in Synthetic Aperture Radar Imagery, Nemni E, et al.

The main scope of this research is to design a Convolutional Neural Network (CNN) based method which can distinguish the flooded pixels in Sentinel -1 SAR imagery without the need for optical data and requiring minimal data preprocessing. The methodology does not require any additional pre-processing apart from orthorectification using a digital elevation model (DEM).

A variety of CNN architectures are tested while the training datasets are generated using a combination of a classical histogram based method in combination with manual cleaning and visual inspection.

The current study focuses on the general water/flood detection by implementing a method with high generalization capacity across different eco-systems and countries, with most of them covering urban land. In addition, a simple linear baseline model is deployed against which compare the more complex machine learning approaches.

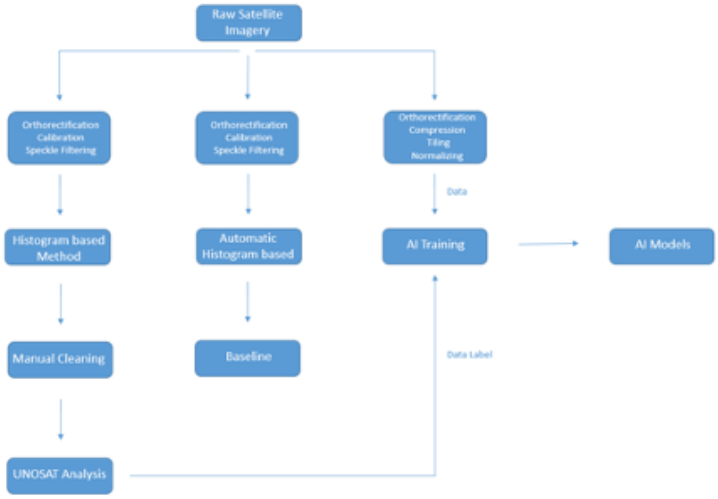

Figure 2.7: Methodological framework. Overview of our general workflow

This study is based on the UNOSAT Flood dataset which has been created using Copernicus Sentinel-1 satellite imagery in Interferometric Wide Swath and Ground Range Detected resolution at 10 meters, along with the corresponding ground truth

flood vectors stored in shapefiles. More specifically, the dataset is composed of VV polarized SAR imagery with their corresponding flood extent boundaries. These boundaries were created in the context of preexisting UNOSAT analyses using a histogram based approach followed by an extensive manual cleaning and noise reduction. These ground truth flood maps were used for training and validating the proposed machine learning architectures.

Location	Event Date	Image Size	Number of Tiles
Myanmar	11 August 2015	21,486 × 30,312	3467
Myanmar	08 May 2016	21,601 × 29,409	3666
Bangladesh	12 August 2017	21,486 × 31,111	6990
Somalia	01 May 2018	22,086 × 28,223	1977
Ethiopia	07 May 2018	21,953 × 28368	4087
Mozambique	13 March 2019	29,657 × 26,488	6282
Mozambique	20 March 2019	22,618 × 16,159	1021
Vietnam	06 September 2019	18,999 × 22,267	2227
Thailand	11 September 2019	21,729 × 29,362	6217
Thailand	11 September 2019	21,687 × 29,559	5811
Cambodia	23 September 2019	21,812 × 29,055	4701
Vietnam	28 September 2019	24,453 × 29,041	5258
Cambodia	31 September 2019	21,698 × 28,768	3169
Mozambique	03 December 2019	20,382 × 15,915	2066
Mozambique	20 January 2020	20,292 × 26,194	1189
Total			58,128

Figure 2.8: Location, event date and image size for each analysis in the UNOSAT Flood Dataset.

Since feeding a neural network with the entire satellite image is computationally impossible, the images were splitted into smaller tiles of size 256x256 pixels, along with their corresponding labels. Even though different tile sizes were tested the best performance was observed with tiles sized 256x256 pixels. The initial dataset is reported to be highly imbalanced with the water pixels accounting for the 6%, which can have significant deterioration in the model performance. In mitigating this problem the authors excluded all the tiles that contained only background pixels. A ratio of 50:50 between classes in the pixel level was not possible, however with the aforementioned under sampling the ratio was increased from 6% to 16% while also speeding up the training process.

The performance of the well used U-Net model, and an alternative model, named XNet is assessed.

2.1.4.1 U-NET

A U-NET architecture consists of two major parts namely an encoder and a decoder. The encoder consists of a stack of the convolutional layers followed by rectified linear unit (ReLU) activation functions and max-pooling layers for downsampling. The decoder consists of convolutional and upsampling layers. U-NET is a Fully Convolutional Neural Network whilst it doesn't contain any dense layers and therefore can accept image patches of any size.

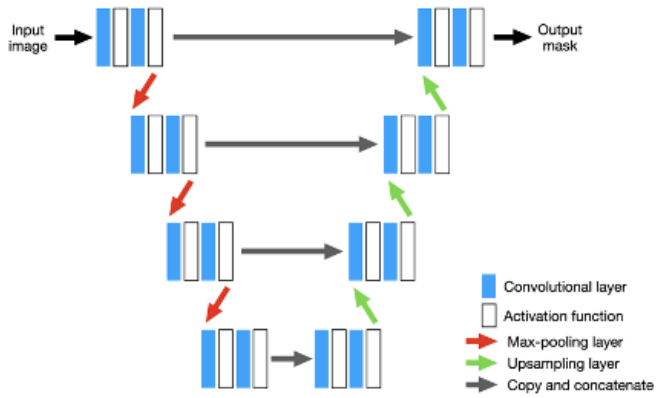

Figure 2.9: A U-Net architecture using 3x3 convolutional layers and ReLU activation functions.

2.1.4.2 XNET

The X-NET architecture is almost identical to U-NET, however, instead of following the encoder-decoder structure this architecture consists of a symmetric encoder-decoder-encoder-decoder structure. XNet is designed to be sensitive to boundary level detail, particularly around small structures, while still achieving strong performance on large scale structures.

2.1.4.3 Transfer Learning

In the context of transfer learning, the authors are testing the performance of a U-net architecture in which the encoding/downsampling stage is replaced by a ResNet model, which has been previously trained on the imageNet dataset. The classification, dense and flattening layers from ResNet were removed and only the con-

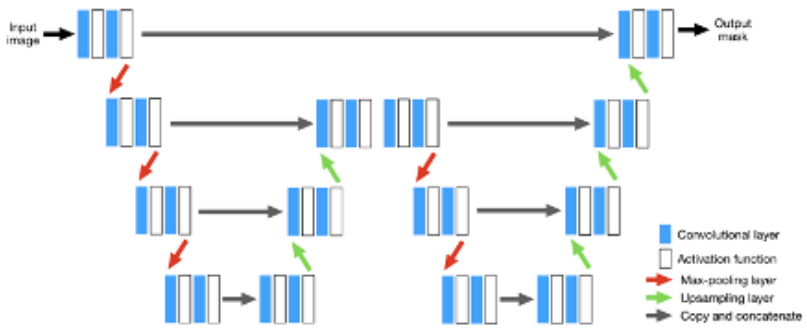

Figure 2.10: XNet architecture using 3x3 convolutional layers and ReLU activation functions.

volutional layers remained. By making this alteration the ResNet architecture is responsible only for feature encoding similar to the encoding layers of the U-net architecture. The encoding stage can potentially contain useful information for high feature extraction such as edges and general shape recognition.

2.1.4.4 Comparison Algorithms

All the aforementioned proposed algorithms are compared against a baseline model which is based on a semi automatic classification based on histogram thresholding. The performance assessment is based on the minimization of the mean squared-error between the labels and the output of the model. Against the baseline model are tested a number of different neural network architectures with different hyper-parameter values and compared their performance in terms of mean squared-error and the time needed in train and test phase. Input tiles (patches) are normalized prior to training step by mean subtraction and min/max scaling. Both XNet and U-Net architectures are implemented using Keras with Tensorflow backend, while the transfer learning approach was achieved by training a U-Net architecture with a ResNet-34 backbone performed within fastai python library. Furthermore, all deep learning models were trained using early stopping and adam optimization with a binary cross-entropy loss function.

2.1.4.5 Neural Network Hyper-Parameter Tuning

Hyper-parameter tuning was carried out by manually altering the relevant parameters such as batch size, number of complete passes, the use of weighted loss function, filter depth and the training epochs. Performance was measured using the precision, recall, CSI and F1 statistics. For the case of XNet and U-Net models the learning rate was set to 10^5 and varied by an order of magnitude with no immediate effect on performance. For the U-Net-Resnet transfer learning approach the fastai 'learning rate finder' tool was utilized in choosing the learning rate for each training phase. Additionally different ResNet depths were tested although no significant difference was observed. The XNet and UNet models were fixed to use a kernel size of 3x3 with a stride of 1.

2.1.4.6 Experimental Results

The experimental results are demonstrated using plots and tables comparing the different methodologies on an unseen test set after hyper-parameter tuning. The test set is made of 5813 image patches across the different locations. The following figure 2.11 plots the best performing model for each of the three different architectures.

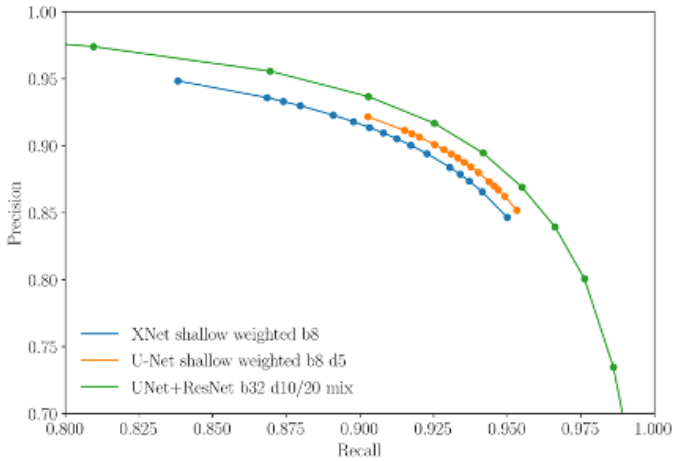

Figure 2.11: Precision-recall curves of the best XNet, U-Net and U-Net+ResNet models after hyper-parameter tuning.

While the models do not significantly differ in performance, the results presented in Figure 9 show that when using the U-Net+ResNet model, the choice of probability

threshold can have a more significant effect on precision/recall statistics than in the case of the XNet and U-Net architectures. Example outputs of the best performing U-Net+ResNet architecture can be seen in Figure 2.12, where can be seen the neural network's ability to detect the flood area with little difference in comparison to the labeled data. The baseline (third column in Figure 2.12) was generated using the automatic threshold based method.

Table 2.1: Overall quantitative comparison.

Model	Accuracy	Precision	Recall	Critical Success Index	F1/Dice
Baseline	91%	62%	84%	0.55	0.71
XNet	97%	91%	91%	0.81	0.91
U-Net	97%	91%	92%	0.83	0.91
U-Net+ResNet	97%	91%	92%	0.77	0.92

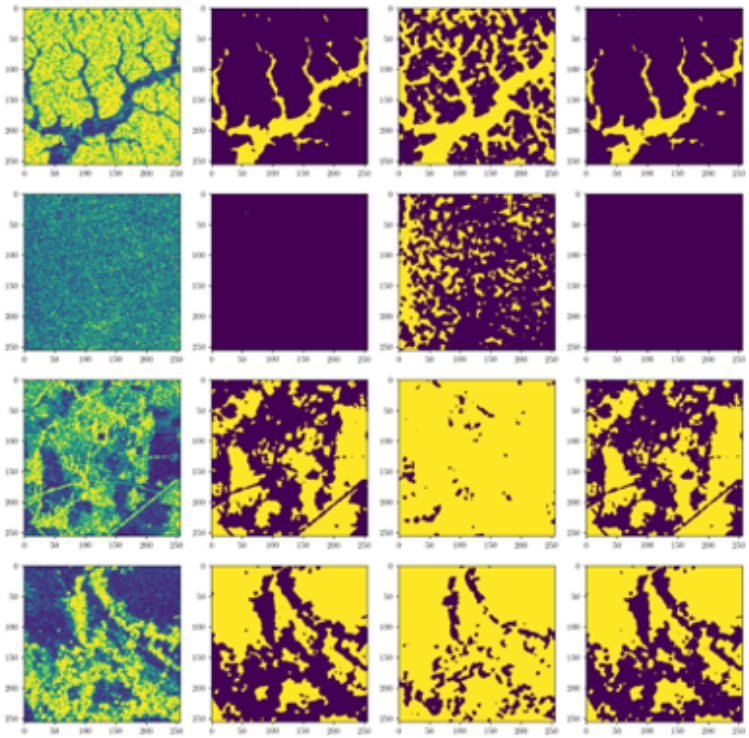

Figure 2.12: From left to right: raw SAR tiles displayed followed by tiles of different analyses corresponding to classical histogram based, baseline and neural network predictions.

2.1.4.7 Sagaing Region—18 July 2019

The authors of this study tested the U-Net-ResNet model on a complete unseen location, in order to further explore the generalizability of the trained mode. The new location is the Sagaing Region, Myanmar and the flood that took place on 18 july 2019. The image was acquired from Copernicus open access hub and results compared to the UNOSAT's analysis. The tests were performed to determine results on both overall water extent detection as well as only on the flooded areas. In order to subtract the permanent water from the model's predictions the Global Surface Water Dataset was used. The output of the model can be seen on the following figure 2.13.

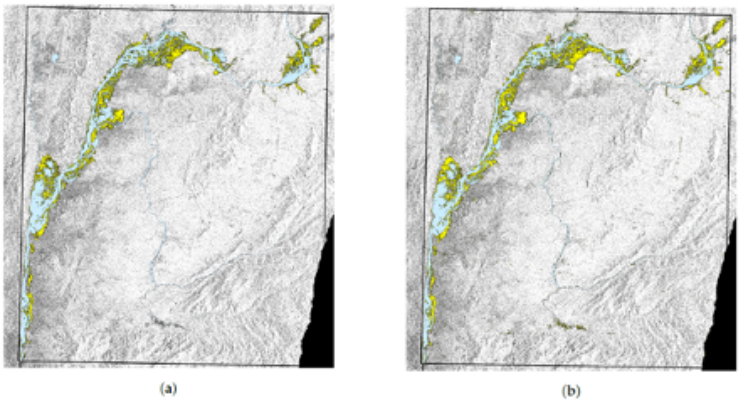

Figure 2.13: (a) shows the classically generated label in yellow. (b) shows the machine learning prediction in yellow. Neither the label nor the prediction include any permanent water.

In this certain experiment the authors try to eliminate the border artifacts between classes on segmented images. They mitigate this issue by tiling the image with a stride less than the tile dimensions thereby ensuring overlap between consecutive patches. Furthermore the tiles are clipped to remove the exterior and in the overlapping regions the average value is taken into account. The results of this experiment are illustrated in the Table 2.2 where it is noticed that performance on flood only regions is slightly lower than the previously presented experiments. Another important note is that the model performance well on detecting no flooded areas, given that most of the image is covered by this class (background).

Table 2.2: Quantitative comparison over Sagaing Region, Maynmar-18 July 2019.

Model	Class	Accuracy	Precision	Recall	CS Index	F1/Dice
U-Net+ResNet	Water	99%	93%	97%	0.90	0.95
U-Net+ResNet	Flood	99%	82%	97%	0.90	0.89

2.1.4.8 Conclusions

Many organizations dealing with issues relevant with flood mapping and disaster management are currently spending many hours using a variety of manual and semi-manual image processing techniques producing highly detailed maps of flood extent boundaries. The current study illustrates a machine learning based approach with a goal to automate and increase the speed required to map floods, while achieving significant accuracy metrics across a broad range of environmental conditions and topography. Utilizing SAR imagery and convolutional neural networks significant performances are reported outperforming previous studies of SAR flood segmentation as well as methods based on optical satellite images. In addition the current reported approach does not require extensive pre-processing before the image is fed onto the network hence spending less time on that stage. The authors as a future research on the topic suggest to train a model on data supplemented by ground truth data generated by field surveys, which could lead into high thematic accuracy

2.1.5 OmbriaNet - Supervised flood mapping via convolutional neural networks using multitemporal Sentinel-1 and Sentinel-2 data fusion (Georgios I. Drakonakis, et al.

This study proposes a new deep neural network architecture that is able to detect and distinguish permanent from flooded water by exploiting the temporal differences among these types of water (permanent and flooded), while using multimodal data by different sensors. In order to validate the proposed architecture a new dataset was initiated named OMBRIA which consists of a total 3.376 images, SAR from Sentinel 1 and multispectral optical from Sentinel 2, accompanied with ground truth binary

images exported by the Emergency Management Service of ESA. The aforementioned data covers 23 flood events spanning around the world, from 2017 to 2021. The main source for labeled datasets is the Emergency Management Service of the Copernicus program (CEMS) which provides mapping data packages and products for natural disasters like floods and earthquakes. From CEMS, 20 flood events were selected ranging from 2017 to 2020 Table 2.3. These data are offered as vector files so they had to be converted into raster images in order to be used as ground truth. Sentinel-1 images in Level-1 Ground Range Detected (GRD) and with VV polarization were used. The pixel values represent the detected amplitude, while the phase information is lost. In order to reduce the speckle effect, a morphological filter with a median value structure of size 30 x 30 m was applied. The Sentinel-2 bands used are band 3 (green-0.560 µm), band 8 (near-infrared-NIR-0.842 µm) and band 11 (short-wave infrared-SWIR-1.610 µm). Bands 3 and 8 have a spatial resolution of 10 m while band 11 has a spatial resolution of 20 m. For this application Sentinel 2 Level-2A was selected, which is an atmospherically corrected surface reflectance product.

The images from each flood event were separated into non-overlapping tiles of size 256 x 256 pixels. In total, 844 image tiles were obtained for each time stamp, i.e. before and after the event, and for each type of satellite data, i.e. Sentinel-1 and Sentinel -2, figure 2.14. Since the flood events are mainly caused by rainfalls the probability that the area is covered with clouds is very high. Given that, Sentinel 2 images with cloud cover had also to be included. The coordinate system that Sentinel images are offered is the World Geodetic System 1984 (EPSG: 4326) while the coordinate system of CEMS is the cartographic projection UTM (Universal Transverse Mercator). This means that the images had to be reprojected in a common cartographic system and co-register them with the flood data.

The proposed architecture is based on U-Net with further developments and alterations in order to exploit multimodal and multitemporal data. The limitation of traditional U-Net is that makes the network incapable of distinguishing permanent (lakes, rivers etc) from flooded water. To overcome this inability of U-Net they introduce a new model called OmbriaNet. The core of this new model is that there

Table 2.3: Flood events used in OMBRIA data as Emergency Management Service Rapid Mapping Activations.

EMS ID	Country	Date 1	Date 2	UTM Zone
271	Greece	01/05/2017	28/02/2018	34 N
273	Albania	01/05/2017	11/03/2018	34 N
275	Croatia	01/05/2017	22/03/2018	33 N
279	Spain	01/05/2017	15/04/2018	30 N
324	France	01/05/2018	10/16/2018	31 N
342	Australia	15/04/2018	13/02/2019	54 S
388	Spain	01/05/2019	14/09/2019	30 N
416	France	01/05/2019	15/12/2019	30 N
417	Portugal	01/05/2019	12/23/2019	29 N
419	Iran	01/05/2019	13/01/2020	41 N
422	Spain	01/05/2019	26/01/2020	31 N
424	Madagascar	01/05/2019	29/01/2020	39 S
429	Ireland	01/05/2019	23/02/2020	29 N
441	Finland	01/05/2019	04/06/2020	34 N
465	Greece	01/05/2020	20/09/2020	31 N
466	Niger	01/05/2020	27/09/2020	32 N
468	Italy	01/05/2020	10/10/2020	32 N
470	Togo	01/05/2020	17/10/2020	31 N
482	Honduras	01/05/2020	11/22/2020	17 N
492	France	01/05/2020	02/01/2021	30 N
501	Albania	01/05/2020	15/02/2021	35 N
507	Timor	01/05/2020	06/04/2021	51 S
514	Guyana	01/05/2020	06/06/2021	21 S

are three sources of water with the first accounting for water bodies such as oceans and rivers, the second accounting for temporal streams of water and the third one being the flood water. Additionally, temporal streams and flood water present a periodicity which can be captured if the Deep Learning model is fed with the same area of interest in two different chronological moments, pre-event and after the event. In the present study the problem is formulated as pixel based semantic segmentation. Each pixel of the image can be classified as "water" or "non-water" which in a computer language the value 1 means water (or flood) and the value 0 means non-water. More specifically, the proposed architectures are listed below.

1. U-Net Basic Architecture: The modern and well-known image segmentation neu-

Figure 2.14: Map of Emergency Management Service Activations for flood events.

ral network called U-Net is used as the basic model. Two more prototype deep learning architectures, multimodal and multitemporal, are developed, which are specifically designed for separating flooded areas from the rest, as a change detection problem. In the present study, modifications are applied to the basic U-Net to perform inundation mapping, which serves as a baseline evaluation for the experiments. Ultimately U-Net proves to be incapable of separating flooded areas from areas that have a permanent presence of water such as lakes, rivers and oceans.

2. Bitemporal OmbriaNet: Generating feature maps from multitemporal images improves change detection accuracy as the network detects modifications based on existing temporal information. The bitemporal OmbriaNet accepts as input two images of the same geographic area taken at different times, before and after the flood event. Figure 2.16 illustrates this particular network.
3. Multimodal OmbriaNet: Multimodal OmbriaNet is an improved version of bitemporal OmbriaNet taking advantage of multimodal information. This network accepts as input four images, two Sentinel-1 images and two Sentinel-2 images, which were taken before and after the flood event and depict the same area. This architecture is demonstrated in figure 2.17.

The data were pre-processed on the Google Earth Engine platform, which is a cloud based and widely used for geospatial science data. Sentinel-1 and Sentinel-2 datasets were accessed through the aforementioned platform, for two different timestamps. The first timestamp spans from 1st May to 31st May of the same year before the flood event. The cloud cover was set to range between 10%-30%

and the values of the final image was the average for all available images from the aforementioned time range. The second timestamp starts on the date the flood event occurred and spans 15 days after the event. The same cloud cover threshold as before is applied, but now the first available pixels are selected as input without an intensity averaging procedure. A data augmentation practice is adopted, where for each patch of input image a left-right flip, horizontal-vertical flip, shearing and also random rotations are applied. In this way, the size of the data set was increased by a factor of 2 leading to a more balanced dataset. Lastly, the data were splitted into 80% for training, 10% for validation while the remaining 10% was used for testing. All pre-processing steps are illustrated in flowchart form in figure 2.18. Figure 2.19 shows results for a selected image sample from a flood event in France.

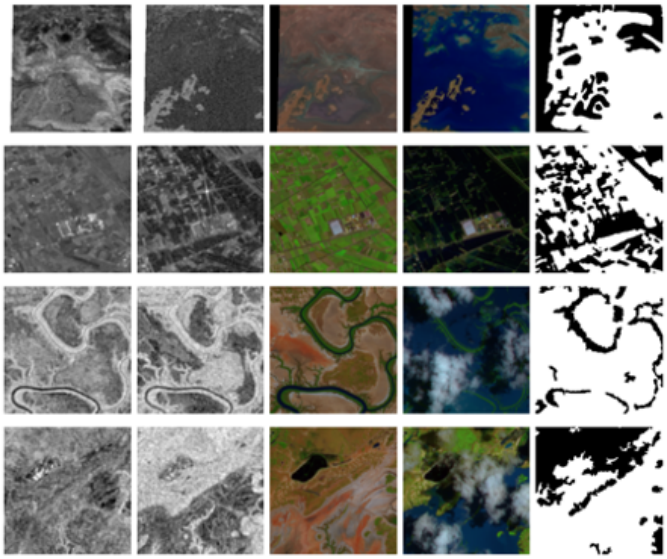

Figure 2.15: Samples of the OMBRIA data. From left to right: Sentinel-1 before the flood event, Sentinel-1 after the flood event, Sentinel-2 before the event, Sentinel-2 after the event, Ground truth where white color indicates flood.

2.1.5.1 Conclusions

In conclusion the proposed architecture for flood mapping from satellite images is named as OmbriaNet. This architecture is capable of exploiting features from multimodal and multi temporal satellite images for pixel based semantic segmentation under real situations. Modern cloud platforms such as Google Earth Engine proved to be useful in generating data for satellite remote sensing applications while Coper-

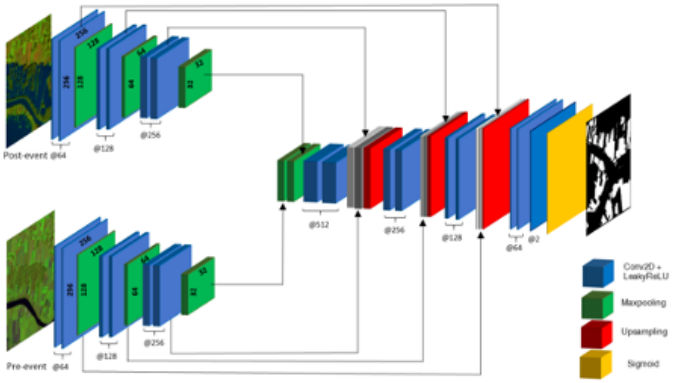

Figure 2.16: The architecture of the bitemporal OmbriaNet.

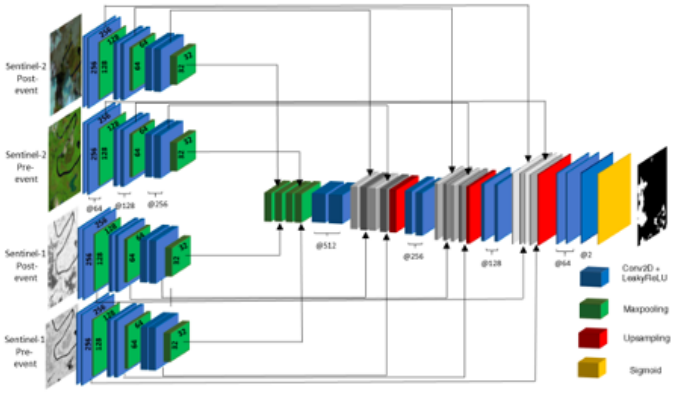

Figure 2.17: The architecture of the multimodal OmbriaNet.

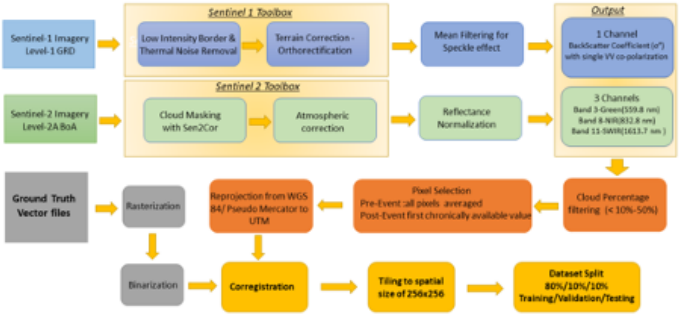

Figure 2.18: Flowchart of the pre-processing of the data in the Google Earth Engine platform.

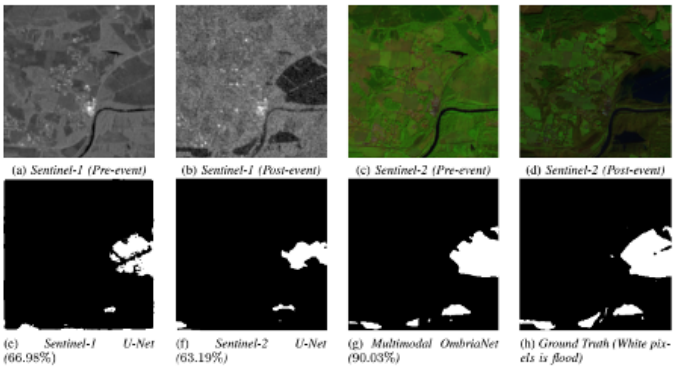

Figure 2.19: Comparisons for a selected image sample from the ID492 flood event in France. The numerical measure Intersection over Union (IoU) is given.

nicus Services provide critical ground truth annotated data of high quality. As future endeavors the authors plan to expand the number of samples, include more spectral bands and also to experiment with the very high resolution images captured from Unmanned Aerial Vehicles.

2.2 Literature Review - Conclusions

Most of the studies related to flood mapping pursue to solve the problem with binary pixel-based supervised semantic segmentation approaches or less often as instance based where images are classified depending on whether or not they contain flood or not. From literature it is apparent that most of the studies use SAR images due to their nature to penetrate clouds making them more appropriate for studying floods since floods are mainly caused by heavy rainfalls. Some studies use sar or optical bands in a single modal feature spaces and others use both of them as multi-modal.

In the concept of choosing the appropriate spectral bands for flood identification the VH band is the most informative and highly correlated with the target value while the near infrared and the short wave infrared optical bands have also been extensively used, with great results. As far as synthetic spectral indices the most prominent are the NDVI and NDWI/MNDWI.

In terms of choosing machine learning models the fully convolutional neural networks and especially UNET architecture seem to be very popular and also very robust in different environmental conditions, leading to high generalization. The

shallow machine learning methods are still producing great results, more explainable and challenge the performance of deep learning methods. On the other hand old methods based on histogram thresholding require human intervention since the optimal or a global threshold value is not easy to be found-calculated. In the same concept as baseline model the simple histogram segmentation using automatic algorithms like Otsu thresholding on VH band or NDWI index is the most prominent. Lastly, from literature review it became evident that the most challenging issue is to separate permanent water from flooded water, without using pre-event images.

2.3 Satellite Remote Sensing

Remote sensing (RS) can be defined as the technology used to acquire physical data about an object by detecting energy reflected or emitted by that object when the distance between the object and the sensor is much greater than any linear dimension of the sensor. Earth observation satellites are equipped with instruments operating in wavelengths extending from the visible to microwave range.

Satellite remote sensing data can be described with the following four different types of resolution [10]:

- **Spatial resolution**. It is the size of a pixel projected in the ground. It indicates the size of the smallest object from which a sensor can retrieve information. High spatial resolution translates into a higher amount of information.
- **Spectral resolution**. It is the number of spectral bands and also the range of wavelengths each band is sensitive to radiance. Today's technology offers hyperspectral sensors with hundreds of spectral bands.
- **Temporal resolution**. Refers to the frequency of revisit above a specific geographic area. For example Sentinel 3 scans the same area of the earth every 6 days approximately.
- **Radiometric resolution**. Corresponds to the sensitivity to the magnitude of the electromagnetic energy of the sensor. Most of the sensors store the radiance in a 16 bit color depth.

The scientific field of Earth Observation (EO) uses remote sensing techniques

and methods to gather information about Earth's physical, chemical and biological systems. The data are gathered from satellites carrying imaging devices which can be grouped in two types depending on the source of energy. Passive remote sensing utilizes the natural source of energy from the sun, while active remote sensing exploits controlled energy sources that beam at a specific section of the electromagnetic section [11].

In the study of flood mapping traditional approaches exploit the capacity of water to absorb the light at certain wavelengths formulating this behavior in spectral indices like the Normalized Difference Water Index (NDWI) and its improved version modified NDWI. These indices are suitable in separating water areas from background information irrespectively of the landscape.

Concerning classification algorithms, since early 2015 the advent of deep learning architectures [12] in conjunction with the massive stream of freely available satellite images lead to superior accuracies than the older traditional machine learning methods like Support Vector Machines and Random Forests. Although most of current work is focused on land classification applications, there is an increased interest in water oriented applications, such as water detection.

2.3.1 Active Remote Sensing

SAR images are sensitive to the geometrical attributes of the backscattering elements. Smooth surfaces such as roads and water bodies tend to backscatter most of the transmitted electromagnetic wave away from the direction of the sensor, resulting in dark pixels in the image. The polarization of the wave is also affected by the presence of water or other similar plane surfaces. SAR sensors due to their ability to penetrate clouds and being independent of weather conditions play an important role for flood detection projects. More specifically sensors like the one onboard Sentinel 1 have been used to map inundation by exploiting the relatively lower backscattered values of water from other ground features. Water is identified by applying a threshold value on a single image, the difference in backscatter values between two images, or using the variance of backscatter in a time series of

images. Some case studies have approximated the flood damage caused by flood using the loss of interferometric signal coherence between two time periods The detection of water bodies in SAR imagery is largely dependent on the fact that water bodies appear as smooth surfaces with a well defined intensity of backscatter. Even though water surfaces are well distinct from the background, a universal threshold for backscatter intensity does not exist, because of the effects of topography and shadows. For instance the backscattered signal can be significantly affected by the buildings which causes the effect of double-bounce scattering. Other urban elements like bridges and cars behave like dihedral reflectors, which reflect much of the incident radiation back to the receiver. As a result urban areas appear to be bright in SAR imagery. Therefore, the detection of floods in urban areas is a challenging task since they are not visible either due to the aforementioned dihedral phenomenon or due to the low spatial resolution [13].

SAR imagery is widely used for flood mapping due to its weather independent imaging capabilities. Radar systems with long wavelengths, like L and P bands, can penetrate the canopy and provide information about the inundation state beneath vegetation. Unlike optical sensors, which detect geochemical properties of earth's elements, radar data characterize the geophysical features from the different backscattering mechanisms in various land types, which potentially allows classifying different states of the ground (flooded urban areas etc). In urban environments the buildings, cars or bridges and other elements of urban equipment may behave like dihedral reflectors, which reflect most of the radiation back to the receiver. As a result urban areas on radar imagery appear brighter than expected. Therefore, detecting flooded areas in urban environments is not an easy task since the dihedral angle phenomenon is present [13]. An example of this phenomenon is illustrated in the following Figure 2.20.

2.3.2 Passive Remote Sensing

Passive remotely sensed data is well known as optical imagery with the basic characteristics being acquired only during the day as it depends on the reflections of the sunlight from objects on the earth surface. Another peculiarity is that clouds

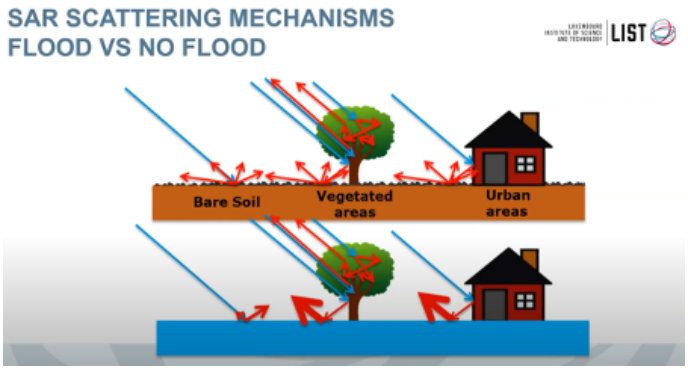

Figure 2.20: SAR scattering mechanisms on flooded and non flooded areas.

can create serious obstructions of the reflected light to reach the sensor. Data from these satellites is both free and commercial where the cost is significantly high but the spatial resolution is also higher than the free ones. Through the years optical images have been used in the development of algorithms for land use - land cover change, crop mapping, disaster monitor etc [10].

One of the most studied optical sensors for inundation mapping is MODIS (Moderate Resolution Imaging Spectrometer). The products of this satellite are characterized by the moderate spatial resolution of 250 meters and the high temporal resolution of 1 to 2 days. The highly absorptive capabilities of water in short wave infrared (SWIR band) relative to other objects or the use of the near infrared (NIR band) have been used to map inundated areas at a daily step. Similar approaches have been tested with medium resolution (30 to 10 meters) sensors such as Landsat and Sentinel 2 using band thresholding or calculating normalized indexes like NDWI. Both Landsat and Sentinel 2 suffer from misclassifications (false positives) of water and cloud shadows, since both of them have low reflectance values in the SWIR and NIR part of the spectrum. Cloud shadows have a similar spectral signature of floods so before any analysis is crucial to mask out these pixels [1].

Spectral indices are images which result from mathematical operations between individual spectral bands of the same image or temporal different images. These mathematical operations can be simple as subtractions or complex like rations. Rations between spectral bands are based on the spectral properties of the ground materials such as absorption and reflection at different wavelengths. Absorption depends on the molecular structure of the surface under observation while reflection

depends on the geometry of the surface - target [14].

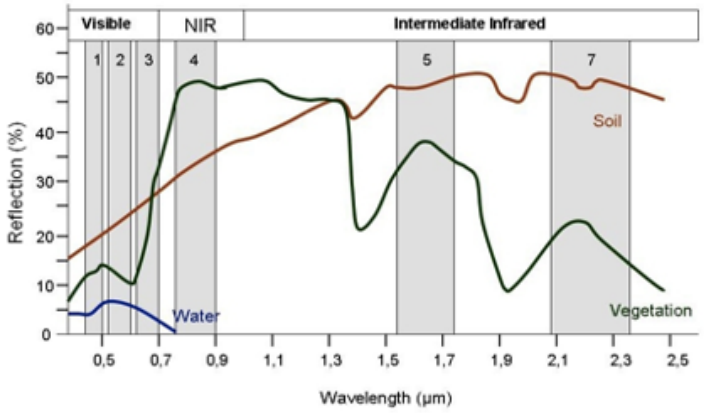

Figure 2.21: Spectral Signatures

Spectral indices are mainly based on the fact that different parts of the spectrum reflect differently on different materials. An example is illustrated in the Figure 2.21. In remote sensing, the most well-known indicators are the vegetation indicators. Since the 1960s, they have been applied to monitor biomass and other biophysical parameters of vegetation at both global and local scales [14].

Vegetation indices are based on the interaction of electromagnetic radiation with plant leaves. The leaves contain special pigments such as chlorophylls, carotenes and xanthophylls. These pigments are contained in plant cell organelles called chloroplasts. Chlorophylls are used to absorb light energy to carry out the function of photosynthesis. During photosynthesis, plants using energy (from specific wavelengths of electromagnetic radiation), produce the components necessary for their nutrition such as carbon dioxide and water. These indices are formed by combinations of spectral channels in such a way as to give a value which expresses the amount of healthy vegetation in a pixel. High values of vegetation indices indicate a high ground cover of healthy vegetation. The simplest form of these indices is the ratio between two spectral channels. The result of this ratio is a new black and white image where each pixel represents the division of the pixel brightness of the two original images. Thus, the areas covered by vegetation are rendered with lighter shades of gray to white, due to the high reflectance it presents in the near infrared and the corresponding low in the visible, while the water masses appear in black

[14].

Normalized Difference Vegetation Index (NDVI)

The Normalized Difference Vegetation Index (NDVI) was created with the aim of separating vegetation from soil brightness using Landsat MSS satellite data. Among the advantages of the index is the minimization of topographical effects. It is also almost invariant to different conditions because of the normalized values. The range of values is from -1 to +1 with 0 expressing the absence of vegetation while negative values describe land covers such as water, man-made structures, etc. More specifically, values close to zero (-0.1 to 0.1) generally correspond to barren areas of rock, sand, or snow. Low, positive values represent shrub and grassland (approximately 0.2 to 0.4), while high values indicate temperate and tropical rainforests (values approaching 1). The disadvantages of the index are that it shows saturation at very high concentrations of vegetation and overestimation at low vegetation concentrations due to soil reflectivity. Finally, atmospheric conditions, such as thin clouds, can potentially affect NDVI values [15]. The formula for retrieving NDVI values is the following:

$$NDVI = (NIR - RED)/(NIR + RED) \tag{2.1}$$

Normalized Difference Water Index (NDWI)

Another important class of spectral indices are the water indices for the study of water bodies as well as droughts. In the current study we are going to focus on the most broadly used ones, namely NDWI and modified NDWI. The first one is used to monitor changes related to water bodies. As water bodies strongly absorb light in the visible to infrared electromagnetic spectrum, NDWI uses green and near infrared bands to highlight water bodies. It is sensitive to built-up land and can result in overestimation of water bodies. Index values greater than 0.5 usually correspond to water bodies, while vegetation usually corresponds to much smaller values and built-up areas to values between zero and 0.2. This index can be used as a complementary to NDVI since it is sensitive to changes in water content of

vegetation canopies [16]. The NDWI results from the following equation:

$$NDWI = (NIR–SWIR)/(NIR + SWIR) \tag{2.2}$$

The Modified Normalized Difference Water Index (MNDWI) uses green and SWIR bands for the enhancement of open water features. It also diminishes built-up area features that are often correlated with open water in other indices. The modified NDWI (MNDWI) can enhance open water features while efficiently suppressing and even removing built-up land noise as well as vegetation and soil noise. The enhanced water information using the NDWI is often mixed with built-up land noise and the area of extracted water is thus overestimated. Accordingly, the MNDWI is more suitable for enhancing and extracting water information for a water region with a background dominated by built-up land areas [17]. The MNDWI results from the following equation:

$$MNDWI = (Green - SWIR)/(Green + SWIR) \tag{2.3}$$

2.3.3 Sentinel 2

Sentinel 2 is a European Copernicus program consisting of two polar-orbiting satellites with a wide swath width of 290 Km and high revisit time of 5 days at the equator. Each satellite carries an optical instrument payload with 13 spectral bands. The Sentinel 2 program provides continuity of SPOT and LANDSAT optical images archive, contributing to applications such as land management, agriculture and forestry, disaster control, humanitarian relief operations, risk mapping and border security [18].

The following table 2.4 illustrates the main technical characteristics of Sentinel 2 spectral bands.

Sentinel 2 products are served to the users as granules of fixed sizes as a minimum indivisible partition containing all 13 spectral bands. Granules are also called tiles with a spatial coverage of 100 x 100 km2 projected in UTM /WGS84 cartographic projection. Sentinel 2 products are offered in two levels of processing, namely Level

Table 2.4: Sentinel 2 Spectral Bands

Band	Name	Spatial Resolution
2	Blue	10 m
3	Green	10 m
4	Red	10 m
8	NIR	10 m
5	Vegetation Red Edge	20 m
6	Vegetation Red Edge	20 m
7	Vegetation Red Edge	20 m
8a	Vegetation Red Edge	20 m
11	SWIR	20 m
12	SWIR	20 m
1	Coastal aerosol	60 m
9	Water vapour	60 m
10	SWIR - Cirrus	60 m

1C and Level 2A. The main difference is that the latter is corrected from atmospheric effects and it is recommended to be used by non expert users. The Level 2A products are not offered systematically and in many cases the user should first download the Level 1C products and then convert them into Level 2A using the dedicated software named Sen2Cor. Level-1C product provides orthorectified Top-Of-Atmosphere (TOA) reflectance, with sub-pixel multispectral registration. Cloud and land/water masks are included in the product. Level-2A product provides orthorectified Bottom-Of-Atmosphere (BOA) reflectance, with sub-pixel multispectral registration. A Scene Classification map (cloud, cloud shadows, vegetation, soils/deserts, water, snow, etc.) is included in the product [18].

2.3.4 Sentinel 1

A single Sentinel-1 satellite is able to map the entire world once every 12 days. The two-satellite constellation offers a 6 day exact repeat cycle. The constellation have a repeat frequency (ascending/descending) of 3 days at the equator, less than 1 day at the Arctic. The trajectory of the constellation is in a near-polar, sun-synchronous orbit with a 12 day repeat cycle and 175 orbits per cycle for a single satellite. Both Sentinel-1A and Sentinel-1B share the same orbit plane with a 180° orbital phasing

difference, with both satellites operating, the repeat cycle is six days. The Sentinel-1 mission comprises a constellation of two polar-orbiting satellites, operating day and night performing C-band synthetic aperture radar imaging, enabling them to acquire imagery regardless of the weather. The Sentinel-1 mission comprises a constellation of two polar-orbiting satellites, operating day and night performing C-band synthetic aperture radar imaging, enabling them to acquire imagery regardless of the weather [19].

Sentinel data products follow an open access policy to all users including the general public with data being delivered within 24 hours of reception. All products are distributed in one of the four types namely SLC, GRD, OCN and RAW along with different sensor modes. For the current study level 1 GRD products were selected which contain only intensity and the amplitude of the backscattered wavelength while the phase value is emitted.

2.3.5 Ground Truth - Data Sources

This subsection tries to demonstrate current operational web services that offer maps, raw data and background knowledge on flood mapping using remorse sensing technologies. More specifically, four services are described namely, CEMS, UNOSAT Flood Portal, Global Flood Database, JRC surface water dataset.

Copernicus Emergency Management Service CEMS

CEMS is implemented by the European Commision as part of the Copernicus Programme and it has two main divisions accounting for "On Demand Mapping" and "Early Warning & Monitoring". The first one provides on-demand information for emergency situations that are caused from natural or man-made disasters, while the later offers information on observational and forecast level about floods, droughts and forest fires. The Copernicus Emergency Management Service (Copernicus EMS) provides all actors involved in the management of natural disasters, man-made emergency situations, and humanitarian crises with timely and accurate geo-spatial information derived from satellite remote sensing and completed by available in situ or open data sources [20].

UNOSAT Flood Portal

satellite-derived flood data in GIS vector format. The portal includes data for selected flood events occurring since 2007, for which UNOSAT did satellite image analysis. You can find and freely load the flood data into online maps, ArcGIS and other GIS systems, such as Google Earth, to combine with your own data or do additional analysis for example in support of disaster risk reduction [4].

Global Flood Database

A research project funded by Google Earth Outreach in a collaboration of various universities, institutes and companies while the main partners were Clod to Street and The Flood Observatory (DFO). The flood maps in this database were created using MODIS optical satellite images which offers two images on a daily basis for the entire earth, since 2022 with the launch of the second twin payload named Aqua. The used spectral bands are the Band 1 and Band 2 served in 250 m spatial resolution but also the SWIR Band pansharpened to 250 m resolution. The flood maps hosted on this database illustrate selected major flood events recorded by the DFO Flood Observatory since the launch of MODIS satellites. The user is able to draw a polygon on a map in order to select the area of interest while also the date of the event can be selected. The user can see basic descriptive statistics concerning the selected flood event and can also download raster images demonstrating the flood extent and the permanent water in the area [1].

Joint Research Center Data Catalog

The European Commission's Joint Research Center initiated the development of a water database within the framework of the Copernicus Programme. The database covers the temporal distribution of water surfaces at a global scale for almost the entire last 4 decades, providing significant statistics on the extent and change of water bodies. The observations produced from Landsat imagery, supporting applications like water resources management, climate modeling and vital information for decision making [21].

Chapter 3

Dataset and Methodology

3.1 Dataset

The dataset used is named as Sen1Floods11 and it is comprised with Sentinel 1 & 2 images with the corresponding ground truth masks. It is publicly available for downloading (approximately 14 GB in size) and accompanies the publication "D. Bonafilia, et al., "Sen1Floods11: a georeferenced dataset to train and test deep learning flood algorithms for Sentinel-1,". The dataset provides global coverage of 4,831 image tiles of 512 x 512 pixels across 11 distinct flood events, covering 120,406 sq km. In terms of organization the dataset is comprised by two main folders covering flood events and permanent water image patches, respectively.

In this study we are only interested in using images included on the flood events folder and excluding the permanent water images. The flood events folder is further splitted into 2 subfolders named as Hand Labeled and Weakly Labeled. The first one includes image patches which have been annotated accurately by photointerpretation while the latter one contains images annotated automatically by thresholding procedures without human intervention, thus low accuracy. The hand labeled folder contains image patches from sentinel 1 and sentinel 2 satellites along with their respective label patches, while the weakly labeled folder contains image tiles created by sentinel-1 only. A critical distinction between image tiles is that the ones created by optical sentinel-2 bands comprises of three thematic classes accounting for flood,

Table 3.1: Sen1Foolds11

"ID"	Country	S2 Date	S1 Date	Days Apart
1	Bolivia	2/15/18	2/15/18	0
2	Ghana	9/19/18	9/18/18	1
3	India	8/12/16	8/12/16	0
4	Mekong	8/4/18	8/5/18	1
5	Nigeria	9/20/18	9/21/18	1
6	Pakistan	6/28/17	6/28/17	0
7	Paraguay	10/31/18	10/31/18	0
8	Somalia	5/5/18	5/7/18	2
9	Spain	9/18/19	9/18/19	0
10	Sri Lanka	5/28/17	5/30/17	2
11	USA	5/22/19	5/22/19	0

non-flood and clouds, while image tiles from sar sentinel-1 accounts only for flood, non-flood.

Each tile follows the naming scheme EVENT_CHIPID_LAYER.tif (e.g. Bolivia_103757_S2Hand.tif). Tile IDs are unique, and not shared between events. Events are named by country and further information on each event (including dates) can be found in the event metadata below. Each layer has a separate GeoTIFF, and can contain multiple bands in a stacked GeoTIFF. All images are projected to WGS 84 (EPSG:4326) at 10 m ground resolution. It should also be noted that weakly labeled patches don't overlap with the hand labeled patches.

As can be seen from Table 3.1 the different acquisitions from Sentinel 2 and Sentinel 1 are not many days apart with the maximum being 2 days. In a post flood flood two days difference are considered.

In the figure 3.1 are pointed out the locations of the extracted flood events, showing the geographic disparity among the dataset, covering every continent apart from Australia.

3.1.1 Hand Labeled

The Hand Labeled subfolder contains one folder S1Hand which consists of Sentinel 1 image patches with two polarization bands (VH & VV) and another one called

Figure 3.1: Spatial Distribution of Dataset, [1]

Table 3.2: Hand Labeled

	LabelHand	**S1OtsuLabelHand**	**S1Hand**	**S2Hand**
Bolivia	15	15	15	15
Ghana	53	53	53	53
India	68	68	68	68
Mekong	30	30	30	30
Nigeria	18	18	18	18
Pakistan	28	28	28	28
Paraguay	67	67	67	67
Somalia	26	26	26	26
Spain	30	30	30	30
Sri-Lanka	42	42	42	42
USA	69	69	69	69
SUM	446	446	446	446

S2Hand which includes Sentinel 2 image patches with 13 spectral bands. It should be noticed that in order to achieve homogeneity among spectral bands an upsample method has been applied, so every bands has 10 meter spatial resolution. The size of the patches is 512x512 within the coordinate system EPSG:4326 - WGS 84 - Geographic. The rest folders are the corresponding ground truth mask, each one being created with a different method. Numerical information is plotted in Table 3.2 where the total number per area is illustrated. As can be seen the number of patches per country is not equal but remains balanced.

Table 3.3: Weakly Labeled

	S1OtsuLabelWeak	**S2IndexLabelWeak**	**S1Weak**
Bolivia	224	224	224
Colombia	534	534	534
Ghana	181	181	181
India	467	467	467
Mekong	1353	1353	1353
Nigeria	109	109	109
Pakistan	249	249	249
Paraguay	316	316	316
Somalia	129	129	129
Spain	146	146	146
Sri-Lanka	190	190	190
USA	486	486	486
SUM	4384	4384	4384

3.1.2 Weakly Labeled

The folder with the weakly labeled images is almost similar with the aforementioned hand labeled, with the only distinction of not including sentinel 2 images. Apart from that the total number of patches is significantly higher but also an additional country is included named Colombia. For the current study it was not computationally possible to manage all these images, thus a method to eliminate most of the patches is described in the following chapter 4.

3.2 Methodology

Experiments were splited into four parts, with each one based on a different semantic segmentation scheme. The first one is based on a Fully Convolutional Neural Network called U-NET, the second approach is based on a Random Forest and a set of hand crafted features while the third one is based on the concept of Transfer Learning using as a backbone the VGG16 model. Lastly, a baseline model based on a histogram thresholding approach was designed.

3.2.1 U-Net

U-Net is a convolutional neural network that was developed for biomedical image segmentation. The network is based on the fully convolutional network and its architecture was modified and extended to work with fewer training images and to yield more precise segmentations. The network consists of a contracting path (convolution) and an expansive path (deconvolution), which gives it the u-shaped architecture. The contracting path is a typical convolutional network that consists of repeated application of convolutions, each followed by a rectified linear unit (ReLU) and a max pooling operation. During the contraction, the spatial information is reduced while feature information is increased. The expansive pathway combines the feature and spatial information through a sequence of up-convolutions and concatenations with high-resolution features from the contracting path. The basic concept of U-NET is illustrated in the Figure 3.2.

The contracting module consists of a series of convolutional layers for feature extraction, along with max-pooling layers which downsample the input. The decoder is applied after feature extraction and performs upsampling to generate a segmentation mask of equal dimension to the input. The expansive also consists of further convolutional layers which allows for additional feature extraction and thus produces a dense feature map [2].

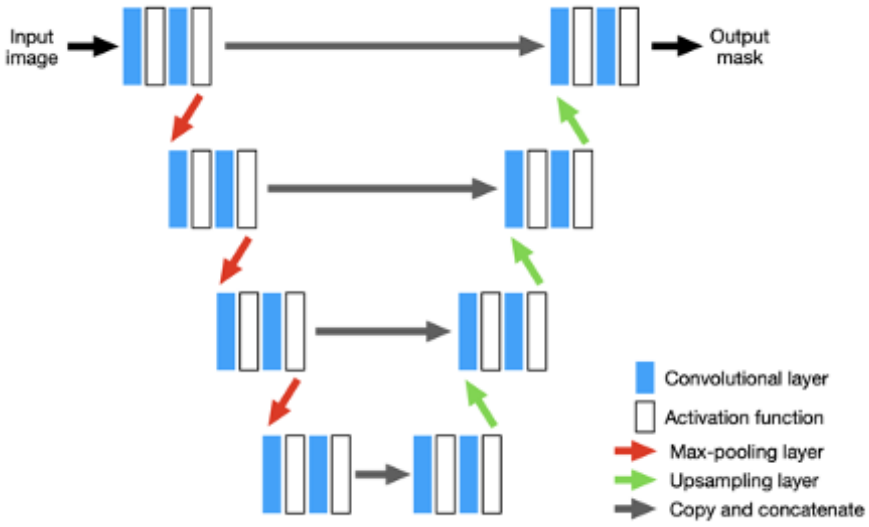

Figure 3.2: The Architecture of UNET, [2]

3.2.2 Transfer Learning

A CNN can be divided into two main parts, as can be seen in Figure 3.3, accounting for feature learning and classification respectively. Transfer learning is a machine learning approach where a pre-trained deep learning model is used as a starting point. Pre-trained models on different datasets are leveraging previously learnt features, which often helps with the generalizability, speeds up the training and developing time [3].

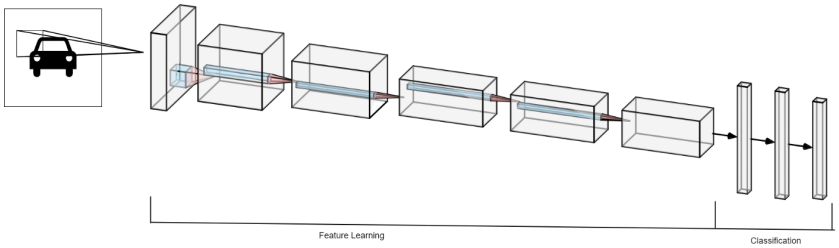

Figure 3.3: The two main parts of a CNN architecture [3].

Deep neural networks extract relevant information in a hierarchical approach where the first layers detect high-level features such as corners and edges while the later layers detect domain specific features. Due to this trait, deep neural networks are highly suited for transfer learning. A common approach is to alter the architecture of the model such that part of it inherits the design of another pre-trained model and then the weights from the pre-trained model are then imported into the selected architecture. This method can also help in model performance and in training speed [4].

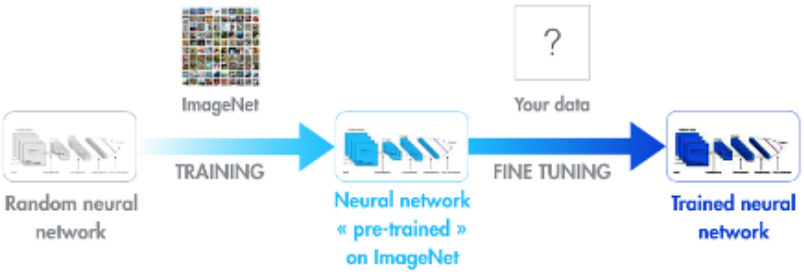

Figure 3.4: The Concept of Transfer Learning, [4]

In the current study the VGG16 architecture pretrained on the Imagenet publicly available data is being used, while the Sen1Floods11 dataset is used for fine tuning. The main concept of this approach is demonstrated in Figure 3.4. Lastly, the classification part is handled by a random forest. VGG Net is the name of a pre-trained convolutional neural network (CNN) invented by Simonyan and Zisserman from Visual Geometry Group (VGG) at University of Oxford in 2014. VGG Net has learned to extract the features (feature extractor) that can distinguish the objects and is used to classify unseen objects. VGG was invented with the purpose of enhancing classification accuracy by increasing the depth of the CNNs. VGG 16 and VGG 19, having 16 and 19 weight layers, respectively, have been used for object recognition. VGG Net takes input of 224×224 RGB images and passes them through a stack of convolutional layers with the fixed filter size of 3×3 and the stride of 1. There are five max pooling filters embedded between convolutional layers in order to down-sample the input representation (image, hidden-layer output matrix, etc.). The stack of convolutional layers are followed by 3 fully connected layers, having 4096, 4096 and 1000 channels, respectively. The last layer is a soft-max layer [3]. The Figure 3.5 shows the VGG16 network structure.

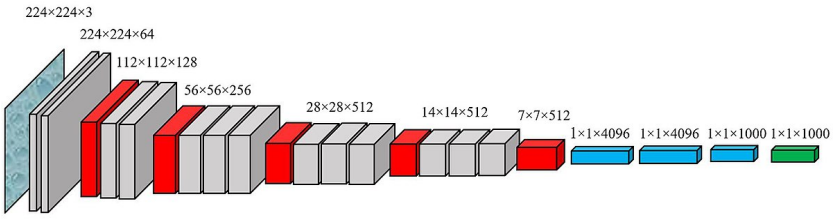

Figure 3.5: VGG Architecture

3.2.3 Random Forest

A random forest (RF) classifier is an ensemble classifier that produces multiple decision trees, using a randomly selected subset of training samples and variables. Random forests train a large number of strong decision trees and combine their predictions through bagging. Two sources of 'randomness': Each tree is only allowed to choose from a random subset of features to split on (leading to feature selection) Each tree is only trained on a random subset of observations. In practice, RF

tend to perform very well right out of the box and almost always gets good results while are also simple with not many complicated parameters to tune [5]. A general architecture is illustrated on Figure 3.6 bellow:

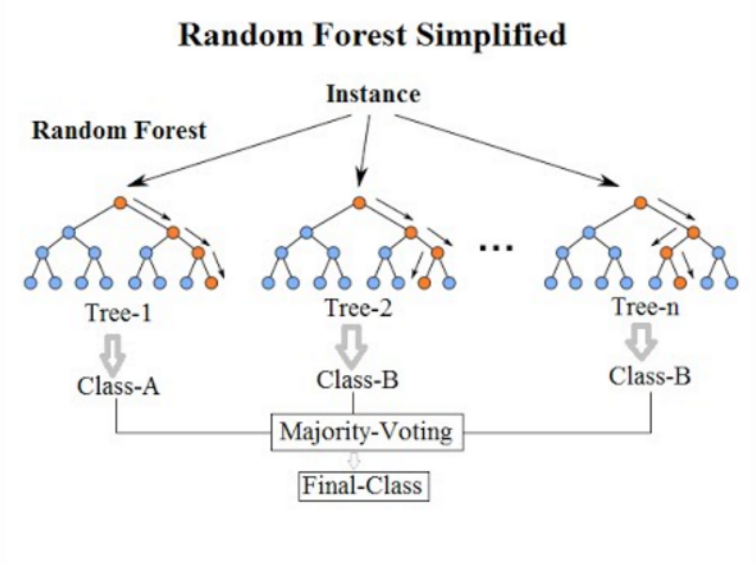

Figure 3.6: The Concept of Random Forest, [5]

For the current study a set of hand crafted features were utilized in order to train a random forest classifier. These features were the previously mentioned spectral indices NDVI and NDWI but also kernel based features like variance and median.

3.2.4 Baseline Model

As baseline model without training phase and feature construction we came up with the idea of thresholding image histogram and then segmenting the image into classes. To achieve this we incorporated the Otsu thresholding method along with certain spectral bands.

The Otsu threshold method is an unsupervised classification method for single band images. With this method, a threshold is automatically calculated which divides the histogram of a greyscaled image into two categories according to the tones of gray that each object in the image has. The goal is to find the value that will separate each object from the rest of the image. The Otsu method involves the selection of a threshold only by editing the histogram of a single band image, without the need to know anything about the image a priori. [22].

For the baseline model assessment two individual experiments were designed. The first thresholds the hand labeled sar images with VH band, while the second one threshold the NDWI index, which were described earlier in the text. It should be noted that since the threshold method can only work with binary classifications, for the optical images only the ones without clouds were used. Concretely less optical images were used, accounting for 337 image patches instead of initial 577 patches.

3.3 Programming Environment

In data science domain the most prominent programming languages are Python and R [23]. Python, however, is a general purpose programming language, while R is generally limited to statistical computing. Furthermore, one of the major benefits of using Python is the large number of libraries, in form of modules, that are available for free. More specifically, A great combination on using Python is the web-based interactive development environment, named as Jupyter Notebook, in which you can present and execute code, including descriptive text and visualizations in a single document [24]. A great web environment to use Jupyter notebooks is the Google Colab which allows anybody to write and execute python code through the browser, and is especially well suited for machine learning and data analysis. Google Colab requires no previous setup while provides access free of charge computing resources including GPU, even though resources are not unlimited. Notebooks are stored in Google Drive or can be loaded from Github [25].

Given the aforementioned information for this study Python and Google Colab were used. The most important python modules included keras for deep learning machine learning approaches, scikit-learn for random forest algorithm and image preprocessing while also numpy and pandas for general purpose data engineering.

Chapter 4

Experimental Results

4.1 Data Pre-Processing

As mentioned in the previous chapter the number of images are too many to be managed by a personal computer thus there was a need to eliminate a number of images. In this chapter is described the process to achieve the aforementioned elimination but also the evaluation metrics applied to measure the performance of each machine learning model and the final results from each designed experiment.

4.1.0.1 Hand Labeled

After visually checking the dataset with manually loading image patches on a free and open Geographic Information System software called QGIS, we noticed that many images contain corroded pixels with no information or the number with flooded pixels is significant lower than the background pixels. Additionally it was noticed that a large number of sentinel 2 images are heavily or totally covered with clouds. Bellow (Figure 4.2) is an illustration of a sentinel 2 image tile blocked with clouds, the corresponding sentinel 1 tile and the respective ground truth.

The initial image tiles of 512x512 size were splited into patches of 128x128, so from each itinial image 16 patches were created. The splitting process in a google colab environment took 8 to 10 hours to complete.

Another critical issue was the imbalance between the number of flooded pixels

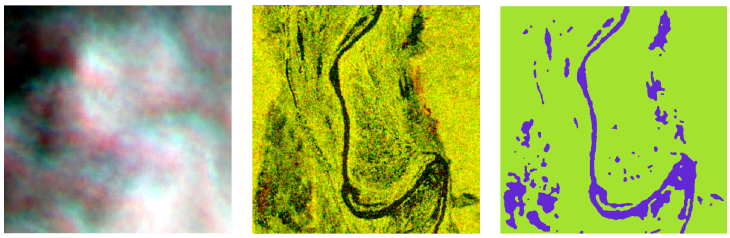

Figure 4.1: From Left to right: A cloudy sentinel 2 image, a sentinel 1 image and a labeled patch based on sentinel 1.

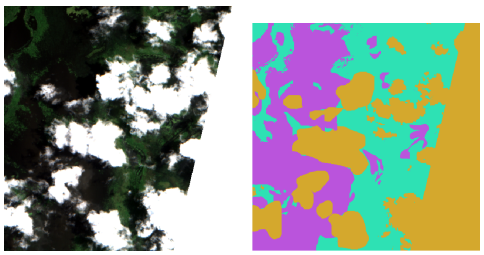

Figure 4.2: From Left to right: A partially cloudy sentinel 2 patch with background noise and the corresponding ground truth based on sentinel 2.

and the background pixels. In order to overcome all these challenges and create a coherent multimodal dataset, we eliminated patches completely covered with clouds, with no flooded pixels or corroded pixels but also the patches with unbalanced number of flooded pixels and background pixels. The remaining number of patches per geographic area is illustrated in Table 4.1. with a total number of images of 577.

Table 4.1: Number of hand labeled patches per area after pre-processing.

Hand Labeled		
1	Bolivia	24
2	Ghana	15
3	India	36
4	Mekong	52
5	Nigeria	33
6	Pakistan	101
7	Paraguay	138
8	Somalia	90
9	Spain	35
10	Sri-Lanka	19
11	USA	34
	SUM	577

4.1.0.2 Weakly Labeled

The initial total number of images were 4384. Each image was splitted into 16 patches of a size 128x128 pixels, resulting in 70144 patches in total. From these we remove the patches having at least one cropped pixel labeled as (-1), patches were the number of flooded pixels were more than 50% than the background pixels and patches with with background pixels more than 50% of the flooded pixels, resulting in a dataset comprised of 6835 patches. Since the number of patches were still very high and not easy to handle, only the first 50 patches from each geographic area were kept, resulting in 600 patches in total as shown in the Table 4.2.

Table 4.2: Number of weakly labeled patches per area after pre-processing.

Weakly Labeled	Country	
1	Bolivia	50
2	Colombia	50
3	Ghana	50
4	India	50
5	Mekong	50
6	Nigeria	50
7	Pakistan	50
8	Paraguay	50
9	Somalia	50
10	Spain	50
11	Sri-Lanka	50
12	USA	50
	SUM	600

4.2 Evaluation Metrics

Two evaluation metrics were used to asses the performance of each model namely Intersection over Union or Jaccard Index and the accuracy index.

4.2.0.1 Intersection over Union (IoU) (Jaccard index)

IoU is the most frequently used metric for image segmentation. It stands as the ratio between the intersection and the union of two sets. In our formulation, it

represents the prediction and the ground truth. It is formulated as the number of true positives over the sum of true positives, false negatives and false positives. It is computed in a per-class basis and averaged [26]. For the binary problem is written as: Similarity of two sets U and V.

Jaccard(U,V) =

$$\frac{|U \cap V|}{|U \cup V|}$$

4.2.0.2 Accuracy

Accuracy is an evaluation metric that allows you to measure the total number of predictions a model gets right. The formula for accuracy is the following:

Accuracy =

$$\frac{TP + TN}{TP + TN + FP + FN}$$

Accuracy will answer the question, what percent of the models predictions were correct? Looking at True Positives and True Negatives.

4.2.0.3 Recall

Recall is the ratio tp / (tp + fn) where tp is the number of true positives and fn the number of false negatives. The recall is intuitively the ability of the classifier to find all the positive samples [27].

Recall =

$$\frac{TP}{TP + FN}$$

4.2.0.4 Precision

Precision is the ratio tp / (tp + fp) where tp is the number of true positives and fp the number of false positives. The precision represents the ability of the classifier

not to label as positive a sample that is negative [27].

Precision =

$$\frac{TP}{TP + FP}$$

4.2.0.5 F1 Score

The F1 score can be interpreted as a harmonic mean of the precision and recall [27]. The formula for the F1 score is:

F1 score =

$$2 * \frac{Precision * Recall}{Precision + Recall}$$

4.3 Results

For every experiment the dataset was splitted into 70% for training and 30% for testing, apart from baseline model for which the reported performance is accounts for all available images. All reported performance numbers are the result of averaging 5 consecutive executions, apart from the baseline model.

Experiments were splitted into three main parts, with each one being based on a different semantic segmentation scheme. The first one is a U-NET fully concolutional neural network, the second one is based on a Random Forest architecture and a set of hand crafted features, while the last one is based on the concept of transfer learning using as backbone the a pretrained VGG16 on Imagenet for feature extraction and a Random Forest for classification.

For all three approaches there are two sections namely single modal and multi modal respectively. The first one makes use of raw spectral bands only from one sensor which is either sentinel 1 bands (VV and VH) or sentinel 2 bands (Red, Green, Blue, Near Infrared and Shortwave Infrared). The latter attempts to combine the spectral bands in a single multimodal feature space. Apart from the raw spectral bands synthetic features are also incorporated, as described in the Table 4.3.

Table 4.3: The Feature Space as a list.

Name	Discription	Sensor
Blue	Blue spectral band	S2
Green	Green spectral band	S2
Red	Red spectral band	S2
NIR	Near Infrared spectral band	S2
SWIR	Sort wave infrared spectral band	S2
NDVI	Normalized Difference Vegetation Index	S2
MNDWI	Modified Normilized Difference Water Indes	S2
VV	Vertical Transmition and Vertical Reception	S1
VH	Vertical Transmition and Horizontal Reception	S1
VH/VV	The ratio between VV and VH	S1
Median	Median filter with a kernel size of three, based on VH and NIR	S1/S2
Variance	Variance filter with a kernel size of three, based on VH and NIR	S1/S2
Roberts	The Roberts' Cross edge map based on NIR and VH	S1/S2

The features NDVI and MNDWI were explained extensively on section 2.3.2 while the mechanisms of SAR images on section 2.3.1. The Green, Red, NIR and SWIR features are not used individually but as part of NDVI and MNDWI. The features based on kernels (Median, Edge etc.) were mainly incorporated to capture the spatial information encapsulated in satellite images, while this type of features have been utilized by many land cover mapping campaigns like the map product named S2CLC-2017 [28]. This type of spatial features are created from deep learning models while for shallow architectures an extra effort is needed. Lastly, the ratio between VV and VH was insipred by Copernicus Land Monitoring Services technical documentations [29]. The Figure 4.3 illustrates the Pearson correlation between features and the target value (flood, no flood).

As can be seen the features with highest correlation with the target value are the SAR bands. More specifically, the Median-s3-VH feature has the highest correlation, with the VH and VV to be following. On the other hand, features based on raw optical bands, apart from near infrared, do not appear to be highly correlated with the target value. However the synthetic features accounted as NDVI and NDWI are more than 50% correlation with the label. As far as correlation between features there is need some high correlations reaching the number 80% but these in the

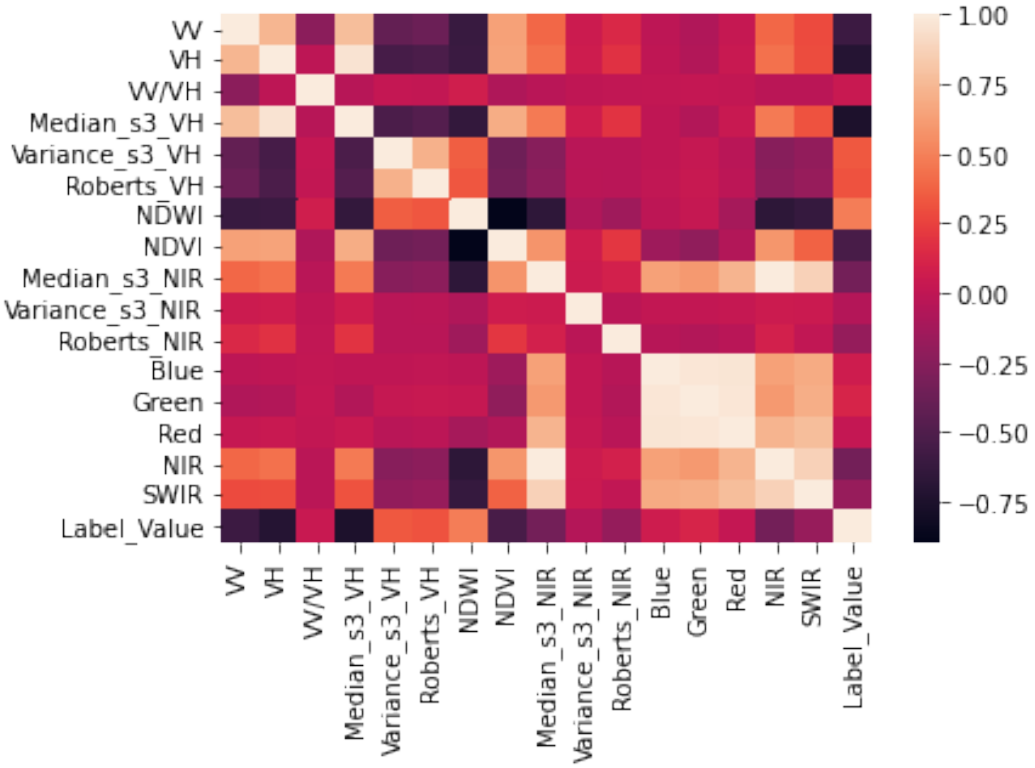

Figure 4.3: Correlation matrix between features

experiments are not used simultaneously.

4.3.1 U-NET

The UNET architecture was thoroughly described in the previous chapter 3. As an optimizer was used the Adam algorithm instead of the classical stochastic gradient descent procedure to update network's weights. Additionally, as a cost function the binary cross entropy function was used for binary classification while the categorical cross entropy for multi-label classification. In the same concept for binary segmentation the last layer had as an activation function the sigmoid while for multi labeled segmentation the softmax. Furthermore a 10% of the training set was kept for validation during the training procedure, while an early stop was applied based on validation loss, with 200 epochs. The number of total epochs was choosen based on the computational limitations of google colab. The total trainable parameters reached the number of 1,941,681.

4.3.1.1 Single-Modal - UNET

The results presented on Table 4.4 are separate experiments with radar and optical bands along with the respective labels. More specifically from sentinel 1 were extracted the two available bands meaning the VV and VH, while from sentinel 2 were extracted the red, green, blue, near infrared and shortwave infrared. It should also be mentioned that sentinel 2 comes with three classes including clouds while sentinel 1 only two classes since radar penetrates clouds.

Table 4.4: UNET Single Modal Hand Labeled

Hand Labeled					
Source & Labels	**IoU**	**Acc**	**F1**	**Prec**	**Recall**
S1Hand & S1OtsuLabelHand	**0.89**	**0.94**	0.94	0.94	0.94
S2Hand & LabelHand	0.47	0.72	0.72	0.72	0.72

Based on Table 4.4 Sentinel 1 radar bands gave superior results than Sentinel 2 optical bands, in terms of both evaluation metrics (Accuracy and Intersection Over Union). However the algorithm fed with features from optical bands had to predict apart from water/ non-water pixels the ones with clouds which is not an easy task.

In the Figure 4.4 are illustrated two examples from single modal experiments subject to Radar and Optical images, respectively. From left to right the first image indicates the binary ground truth where white represent water and black indicates background pixels, mainly land cover. The second black and white image represents the predicted values from trained U-NET model. The UNET gave good predictions apart from the areas where small water parts are close to each other. On the other hand the trained on optical bands, UNET model, did not achieved an equivalent performance where failed to recognize most of the cloudy pixels (green colour) and small water parts (red colour) from the background.

Table 4.5: UNET Single Modal Weakly Labeled

Weakly Labeled					
Source & Labels	**IoU**	**Acc**	**F1**	**Prec**	**Recall**
S1Hand & S1OtsuLabelWeak	0.81	0.87	0.87	0.87	0.87

Weakly labeled image patches showed significant results even though only two

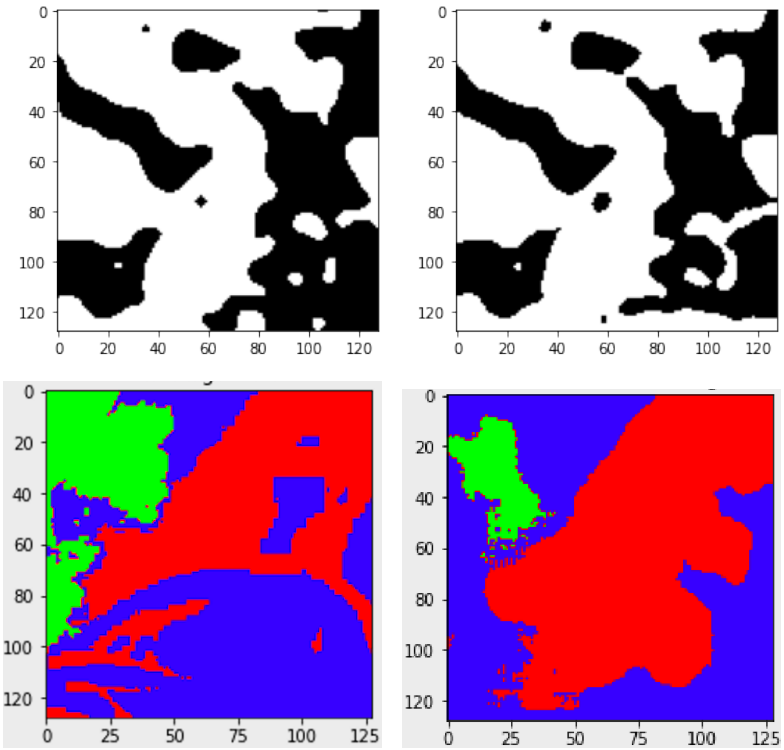

Figure 4.4: An example of a ground truth mask and the corresponding prediction based on UNET and sentinel 1 (left) and sentinel 2 (right).

bands were used, Table 4.5.

Table 4.6: UNET Single Modal Weakly Supervised

Weakly Supervised						
Trained On	Tested on	**IoU**	**Acc**	**F1**	**Prec**	**Recall**
S1Hand & S1OtsuLabelWeak	S1OtsuLabelHand	0.77	0.86	0.86	0.86	0.86

The weakly supervised experimental results presented in Table 4.6 shows that deep learning is capable to achieve considerable accuracies without much effort on labeling.

4.3.1.2 Multi-Modal - UNET

The multi modal results illustrated in Table 4.7 imply that the combination of sentinel 1 and sentinel 2 bands do not improve the accuracy in detecting flooded pixels. Part of the reason could be the difference in acquisition date between sentinel 1 and sentinel 2.

Table 4.7: UNET Multi Modal Hand Labeled

Hand Labeled					
Source & Labels	**IoU**	**Acc**	**F1**	**Prec**	**Recall**
S1Hand - S2Hand & S1OtsuLabelHand	**0.72**	**0.82**	0.82	0.82	0.82
S1Hand - S2Hand & LabelHand	0.42	0.71	0.71	0.71	0.71

Finally, the best results using U-NET were given by single modal sentinel 1 bands along with the hand labeled ground truths.

4.3.2 Random Forest

For this set of experiments a various hand crafted features were utilized from both sentinel 1 and sentinel 2 raw spectral bands. More specifically from optical bands were constructed the NDVI and NDWI while from sentinel 1 the devision between VV and VH. Apart from these futures three more kernel based features were constructed based on VH and NIR bands respectively. Those are the median filter with and variance filters with kernel size of 3 and the roberts edge detection filter. It should be noted that this type of models are note fed by batches of 2D images but with flattening the images into 1D vectors.

4.3.2.1 Single-Modal - Random Forest

Single modal hand labeled experiments showed satisfactory results with the Sentinel-2 giving superior results probably because of the higher number of features utilized compared to Sentinel-1. For Sentinel-1 use 6 features VV, VH, VV/VH, Median, Variance, Rodert Edge and for Sentinel-2 7 features NDVI, NDWI, Median, Variance, Robert edge, Blue, Green.

Table 4.8: Random Forest Single Modal Hand Labeled

Hand Labeled					
Source & Labels	**IoU**	**Acc**	**F1**	**Prec**	**Recall**
S1Hand & S1OtsuLabelHand	0.79	0.89	0.89	0.89	0.89
S2Hand & LabelHand	**0.87**	**0.93**	0.93	0.93	0.93

In the Table 4.8 are listed the quantitative evaluation showing the superiority of

sentinel-2 bands in segmenting flooded areas.

Table 4.9: Random Forest Single Modal Weakly Labeled

Weakly Labeled					
Source & Labels	**IoU**	**Acc**	**F1**	**Prec**	**Recall**
S1Weak & S1OtsuLabelWeak	0.81	0.90	0.90	0.90	0.90

In the Table 4.9 the quantitative evaluation report is shown based on weakly labeled dataset, showing a slight increase in the performance compared to previously noted hand labeled dataset.

Table 4.10: Random Forest Single Modal Weakly Supervised

Weakly Supervised			
Trained	Tested	**IoU**	**Acc**
S1Weak & S1OtsuLabelWeak	S1Hand & S1OtsuLabelHand	0.77	0.88
		Precision	**Recall**
		0.88	0.88
		F1 score	
		0.88	

Weakly supervised (Table 4.10) experiments gave promising results, even though slightly decreased performance than the ones based on hand labeled ground truths.

As can be seen from Figure 4.5 the Sentinel-1 gave good quality results despite some salt and pepper noise in the land. On the other hand experiment based on Sentinel-2 managed to predict most of the land area but failed to distinguish clouds from water. Overall the Sentinel-1 bands gave superior results in quantitative as qualitative evaluation as well. For sentinel 1 black color indicates land and white the flooded area. For sentinel 2 blue indicates land, white the flooded pixels and red the clouds.

4.3.2.2 Multi-Modal - Random Forest

Multi modal illustrated in Table 4.11 uses all the available features from sentinel 1 and sentinel 2 spectral bands. The distinction between the two experiments bellow is that the first one is a binary classification while the second has additionally one class concerning the clouds.

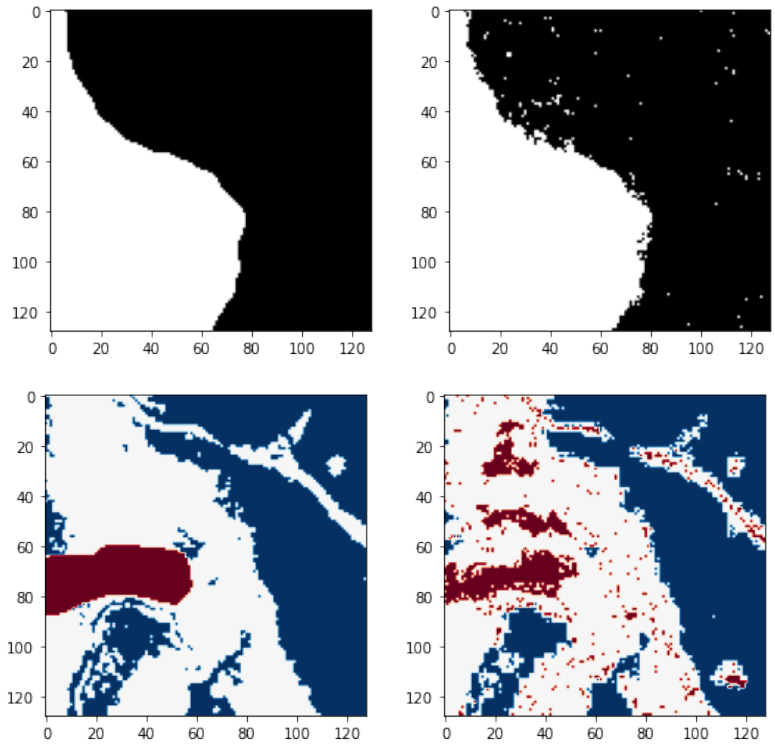

Figure 4.5: An example of a ground truth mask and the corresponding prediction based on RF and sentinel 1 (left) and sentinel 2 (right).

Table 4.11: Random Forest MultiModal Hand Labeled

Hand Labeled					
Source & Labels	**IoU**	**Acc**	**F1**	**Prec**	**Recall**
S1Hand - S2Hand & S1OtsuLabelHand	0.84	0.92	0.92	0.92	0.92
S1Hand - S2Hand & LabelHand	**0.87**	**0.93**	0.93	0.93	0.93

Even though the second experiment dealt with the detection of clouds apart of the flooded pixels, gave superior accuracies. In conclusion, the best results using RF were given by sentinel 2 bands.

Since Random Forest works with hand crafted features a feature importance ranking is possible to be informative. In the Figure 4.6 it can be seen that optical band are more informative in detecting flooded pixels than the sar bands. More specifically the NDVI is while the NDWI cames second as the literature suggest. In the literature there is also strong evidence that VH band is the most significant in separating flood from background pixels.

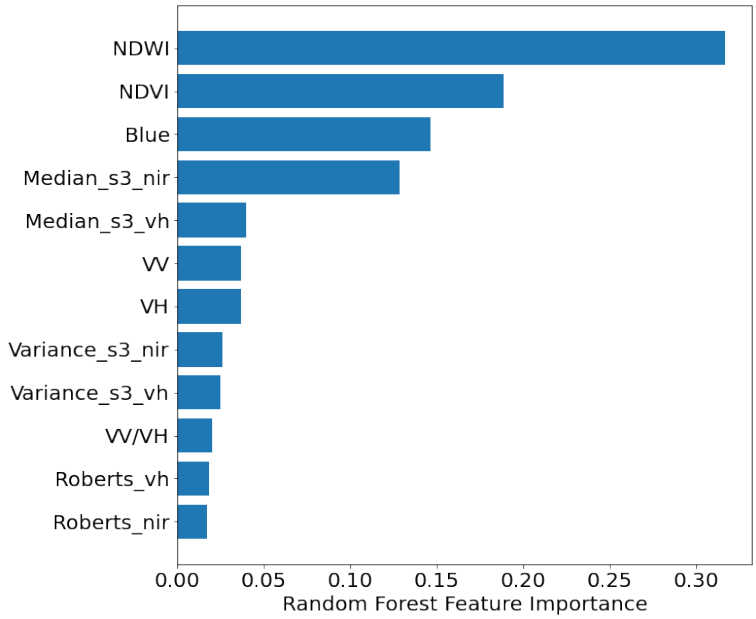

Figure 4.6: Feature importance for multi-modal RF

4.3.3 Transfer Learning

In the current study the VGG16 architecture pretrained on the Imagenet publicly available data is being used, while the Sen1Floods11 dataset is used for fine tuning. For these experiments only first two convolutional layers from VGG16 were used as feature extractors. A summary of the new constructed model can be seen in the Figure 4.7.

```
Model: "model"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 input_1 (InputLayer)        [(None, 128, 128, 3)]     0

 block1_conv1 (Conv2D)       (None, 128, 128, 64)      1792

 block1_conv2 (Conv2D)       (None, 128, 128, 64)      36928

=================================================================
Total params: 38,720
Trainable params: 0
Non-trainable params: 38,720
_________________________________________________________________
```

Figure 4.7: The model summary of the VGG16 used.

4.3.3.1 Single-Modal - Transfer Learning

For single modal experiments and in order to be compiled with the VGG16 architecture which takes as input only three bands we had to eliminate some of the available information. For sentinel 1 experiments the VV, VH and VH/VV were used while for sentinel 2 experiments the Red, Near Infrared and the Shortwave Infrared bands we used mainly because in literature these are the most common spectral bands utilized to detect water.

Table 4.12: Transfer Learning Single Modal Hand Labeled

Hand Labeled					
Source & Labels	**IoU**	**Acc**	**F1**	**Prec**	**Recall**
S1Hand & S1OtsuLabelHand	**0.84**	**0.92**	0.92	0.92	0.92
S2Hand & LabelHand	0.47	0.65	0.65	0.65	0.65

For this set of experiments the best results were achieved by sentinel 1 bands while sentinel 2 bands illustrated quite poor results, mainly because of the cloud coverage and many false positives as shown in the Figure 4.6.

The Figure 4.8 gives an example of two classification maps, one based one sentinel 1 (on the left) and another one based on sentinel-2 (on the right). The transfer learning approach based on sar bands achieves visually an excellent result with no apparent mistakes. On the other hand sentinel 2 bands achieved a significant result apart from the salty noise in the center, which is mainly because of the bad quality of the ground truth. For sentintel 1 white color indicated flooded areas while black pixels are the surrounding area. For sentinel 2 green color indicates the flooded pixels while the blue the background and red the clouds.

Table 4.13: Transfer Learning Single Modal Weakly Labeled

Weakly Labeled					
Source & Labels	**IoU**	**Acc**	**F1**	**Prec**	**Recall**
S1Weak & S1OtsuLabelWeak	0.86	0.92	0.92	0.92	0.92

The experiments based on weakly labeled (Table 4.13) image gave superior accuracies than the hand labeled images, mainly in terms of IoU. Single modal weakly

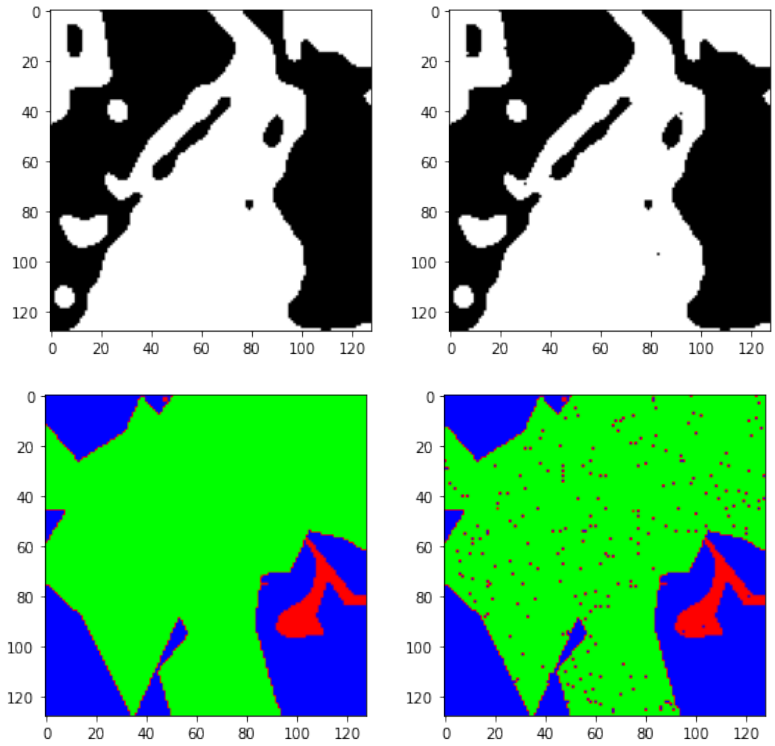

Figure 4.8: An example of a ground truth mask and the corresponding prediction based on Transfer Learning - VGG16 and sentinel 1 (left) and sentinel 2 (right).

supervised (Table 4.14) achieved significant performance equivalent of hand labeled experiments.

Table 4.14: Transfer Learning Single Modal Weakly Supervised

Weakly Supervised			
Trained	Tested	**IoU**	**Acc**
S1Hand & S1OtsuLabelWeak	S1Hand & S1OtsuLabelHand	0.83	0.91
		Precision	**Recall**
		0.91	0.91
			F1 score
			0.91

4.3.3.2 Multi-Modal - Transfer Learning

For the multi modal experiments the VH band was used from sentinel 1 while from sentinel 2 we incorporated only the Red and Near Infrared bands.

Mutimodal experiments were not as well as expected but if instead raw spectral bands we used hand crafted features like the ones used in Random Forest, the results

Table 4.15: Transfer Learning Multi Modal Hand Labeled

Hand Labeled					
Source & Labels	**IoU**	**Acc**	**F1**	**Prec**	**Recall**
S1Hand - S2Hand & S1OtsuLabelHand	**0.73**	**0.85**	0.85	0.85	0.85
S1Hand - S2Hand & LabelHand	0.55	0.71	0.71	0.71	0.71

might be better. Lastly, the best results using Transfer Learning were given by single modal sentinel 1 weakly labeled.

4.3.4 Baseline Model

For this section of experiments only the hand labeled images were used. Since the baseline model does not make use of training phases the weakly labeled images do not make sense to use them. Additionally, it should be noted that for sentinel 2 images only the ones without clouds were selected, since no thresholding algorithm could detect them from spectral bands information.

The following tables 4.16, 4.17 present the performance of each experiment. It can be seen that sentinel 1 gave significantly better results compared with sentinel 2, in terms of accuracy and IoU. Overall both of them illustrate a general good quality performance and they could be used for first stage flood disaster assessments.

Table 4.16: Baseline approach based on VH band

Hand Labeled					
Source & Labels	**IoU**	**Acc**	**F1**	**Pre**	**Recall**
S1Hand & S1OtsuLabelHand	**0.73**	**0.86**	0.86	0.86	0.86

Table 4.17: Baseline approach based on NDWI spectral index

Hand Labeled					
Source & Labels	**IoU**	**Acc**	**F1**	**Pre**	**Recall**
S2Hand & LabelHand	**0.66**	**0.81**	0.81	0.81	0.81

The Figure 4.9 demonstrates two predicted images based on Otsu thresholding approach. The first row presents an example from thresholding a VH band and the second row an example from thresholding a NDWI image. Both of them indicate sufficient quality for a first estimation of the flood event but the results are quite

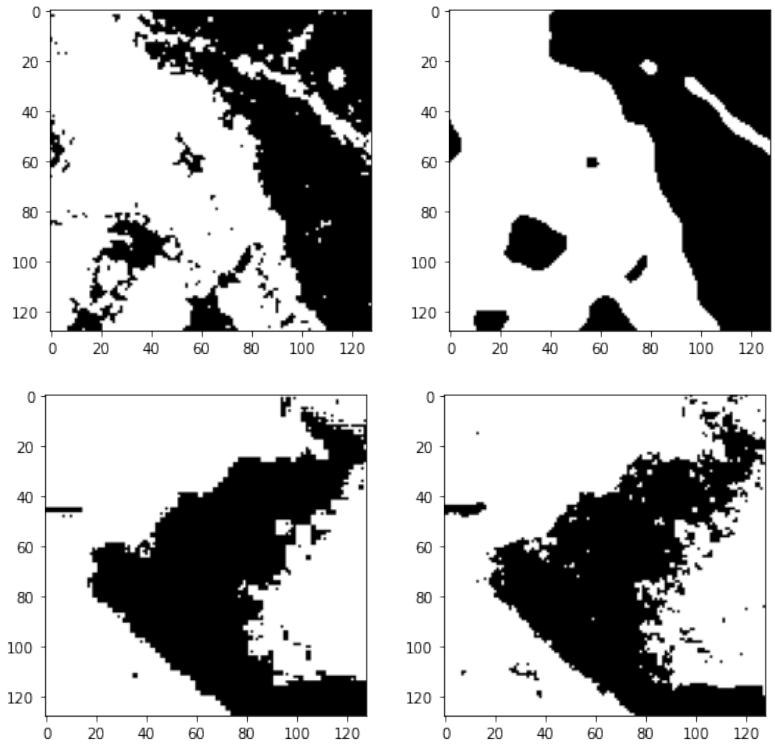

Figure 4.9: An example of a predicted image and the corresponding ground truth mask based on Otsu thresholding. The first row illustrates the results from sentinel 1 while the second row the results from sentinel 2. The white color indicates areas with flood and the black pixels the background

noise. Inspecting visualy the classification maps, quantitatively, the best result is given by the NDWI index rather than VH band. SAR based segmentation failed to capture small land areas surrounded by water and also suffers from salt and pepper kind of noise. On the contrary optical based segmentation illustrates less noisy results. The white color indicates areas with flood and the black pixels the background.

Chapter 5

Conclusions and Future Work

5.1 Conclusions

Even though the initial dataset is very diverse and large, covering large geographic areas and plenty of flood events it should be noted that it is very difficult to manage and contains many faulty tiles and pixels. The trial of experiments included images with the original tile size of 512x512 pixels but the training phase was disappointingly long and thus restricting the experiments. Thus measures had to be taken so to limit the required computational power and time. This was achieved with cropping the tiles into 126x126 pixels and eliminating tiles with unbalanced ground truth and faulty pixels. The experimental results are illustrated aggregated in a comparative way in the Tables 5.1 and 5.2.

Answering the question about which sensor achieved better results the answer is Sentinel-1 synthetic aperture radar, which is apparent from the three single hand labeled experiments from all three classification schemes. More specifically, UNET and Transfer Learning approaches gave better results when fed with SAR data while Random Forest when fed with optical bands.

Comparing the results from multimodal and single modal experiments there is no clear answer on which approach is better, since Transfer Learning gave equivalent results for both feature spaces.

For the weakly supervised part of experiments we only have available sentinel 1

Table 5.1: Single Modal Aggregated Results

	UNET		**TL**		**RF**	
Hand Labeled	**IOU**	**Acc**	**IOU**	**Acc**	**IOU**	**Acc**
S1Hand&S1OtsuLabelHand	**0.89**	**0.94**	0.84	0.92	0.79	0.89
S2Hand&LabelHand	0.47	0.72	0.47	0.65	**0.87**	**0.93**
Weakly Labeled						
S1Weak&S1OtsuLabelWeak	0.81	0.87	**0.86**	**0.92**	0.81	0.90
Weakly Supervised						
Trained On=S1Weak&S1OtsuLabelWeak	0.77	0.86	**0.83**	**0.91**	0.77	0.88
Tested On=S1Hand&S1OtsuLabelHand						
Baseline	–	–	–	–	–	–
S1Hand(VH) & S1OtsuLabelHand	0.73	0.86				
S2Hand(NDWI) & LabelHand	0.66	0.81				

Table 5.2: Multi Modal Aggregated Results

	UNET		**TL**		**RF**	
Hand Labeled	**IOU**	**Acc**	**IOU**	**Acc**	**IOU**	**Acc**
S1Hand - S2Hand & S1OtsuLabelHand	0.72	0.82	0.73	0.85	**0.84**	**0.92**
S1Hand - S2Hand & LabelHand	0.42	0.71	0.55	0.71	**0.87**	**0.93**

images. Comparing the results from weakly supervised trained models against the ones with hand labeled ground truth data, we see great results from both sides. It could be argued that weakly labeled ground truth can achieve equivalent results with the strictly hand labeled ones.

Even though deep learning can perform high accuracies the casual way of hand crafted features with shallow architectures like RF can also achieve high scores. Comparing these two classification schemes given the current results, we have no clear winner but in general swallow architecture achieves superior results on multi-modal datasets compared to deep learning.

Last but not least, the developed baseline model failed to provide better performance for every deep and machine learning models in terms of hand labeled sentinel 1 data. On the other hand baseline surpassed the performance of deep learning in terms of using hand labeled sentinel 2 images, but this might be because of the exclusion of cloudy pixels.

5.2 Future Work

The dataset in its original form is quite large but given the computational power it should be tried to run experiments including the entire dataset apart from the faulty tiles as described in chapter 4. Overall the future endeavours could be summarized in the following list, without order of significance.

- Sentinel's constellation offers an enormous amount of data but with low spatial resolution, which hinders the accuracy of the flood mapping tasks. Inline with that assertion could be examined the high spatial resolution satellite images offered by ESA's third party contributing mission where a researcher can apply for free access to a large amount of data ownership of private companies. Apart from this program other private companies offer similar educational accounts like the ones advertised by Planet and ICEYE. The high spatial resolution of less than one meter (GSD) could improve the accuracies of the models and produce better maps.
- Apart from pre-trained VGG16 could be examined the utilization of other architectures like Xception, ResNET or EfficientNet. Keras python module offers plenty of pre-trained models.
- Experimentation with other hand crafted features could lead to superior results. In literature there is evidence that geomorphological features like elevation, slope and aspect can potentially help in the identification of flooded areas. These features could be calculated from a digital elevation model.
- Expand the dataset with extra flood events in different geographic regions utilizing the ground truth layers available on the public databases described in section 2.2.2. This way the models could be valuated on unseen images and lead to unbiased results and generalized models.
- Expand the labels to include cloud shadows as a thematic class. It was noticed that most the cloudy images are also suffering from cloud shadows, causing changes in spectral responses in unpredicted way.

References

[1] Derrick Bonafilia, Beth Tellman, Tyler Anderson, and Erica Issenberg. Sen1floods11: A georeferenced dataset to train and test deep learning flood algorithms for sentinel-1. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 210–211, 2020.

[2] Huimin Huang, Lanfen Lin, Ruofeng Tong, Hongjie Hu, Qiaowei Zhang, Yutaro Iwamoto, Xianhua Han, Yen-Wei Chen, and Jian Wu. Unet 3+: A full-scale connected unet for medical image segmentation. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1055–1059. IEEE, 2020.

[3] Srikanth Tammina. Transfer learning using vgg-16 with deep convolutional neural network for classifying images. *International Journal of Scientific and Research Publications (IJSRP)*, 9(10):143–150, 2019.

[4] Edoardo Nemni, Joseph Bullock, Samir Belabbes, and Lars Bromley. Fully convolutional neural network for rapid flood segmentation in synthetic aperture radar imagery. *Remote Sensing*, 12(16):2532, 2020.

[5] Mariana Belgiu and Lucian Drăguţ. Random forest in remote sensing: A review of applications and future directions. *ISPRS journal of photogrammetry and remote sensing*, 114:24–31, 2016.

[6] N Douben and RMW Ratnayake. Characteristic data on river floods and flooding; facts and figures. In *Floods, from Defence to Management: Symposium Proceedings of the 3rd International Symposium on Flood Defence, Book, Section vols., J. van Alphen, E. van Beek, and M. Taal, Eds*, pages 19–35, 2006.

[7] CRED UNISDR et al. The human cost of natural disasters: A global perspective. 2015.

[8] Roberto Bentivoglio, Elvin Isufi, Sebastian Nicolaas Jonkman, and Riccardo Taormina. Deep learning methods for flood mapping: A review of existing applications and future research directions. *Hydrology and Earth System Sciences Discussions*, pages 1–43, 2021.

[9] Sebastiaan N Jonkman. Global perspectives on loss of human life caused by floods. *Natural hazards*, 34(2):151–175, 2005.

[10] George Joseph. *Fundamentals of remote sensing*. Universities Press, 2005.

[11] Katherine Anderson, Barbara Ryan, William Sonntag, Argyro Kavvada, and Lawrence Friedl. Earth observation in service of the 2030 agenda for sustainable development. *Geo-spatial Information Science*, 20(2):77–96, 2017.

[12] Xiao Xiang Zhu, Devis Tuia, Lichao Mou, Gui-Song Xia, Liangpei Zhang, Feng Xu, and Friedrich Fraundorfer. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4):8–36, 2017.

[13] Ian G Cumming and Frank H Wong. Digital processing of synthetic aperture radar data. *Artech house*, 1(3):108–110, 2005.

[14] Yichun Xie, Zongyao Sha, and Mei Yu. Remote sensing imagery in vegetation mapping: a review. *Journal of plant ecology*, 1(1):9–23, 2008.

[15] Nathalie Pettorelli. *The normalized difference vegetation index*. Oxford University Press, 2013.

[16] Stuart K McFeeters. The use of the normalized difference water index (ndwi) in the delineation of open water features. *International journal of remote sensing*, 17(7):1425–1432, 1996.

[17] Hanqiu Xu. Modification of normalised difference water index (ndwi) to enhance open water features in remotely sensed imagery. *International journal of remote sensing*, 27(14):3025–3033, 2006.

[18] Matthias Drusch, Umberto Del Bello, Sébastien Carlier, Olivier Colin, Veronica Fernandez, Ferran Gascon, Bianca Hoersch, Claudia Isola, Paolo Laberinti, Philippe Martimort, et al. Sentinel-2: Esa's optical high-resolution mission for gmes operational services. *Remote sensing of Environment*, 120:25–36, 2012.

[19] Dirk Geudtner, Ramón Torres, Paul Snoeij, Malcolm Davidson, and Björn Rommen. Sentinel-1 system capabilities and applications. In *2014 IEEE Geoscience and Remote Sensing Symposium*, pages 1457–1460. IEEE, 2014.

[20] Patrick Matgen, Sandro Martinis, Wolfgang Wagner, Vahid Freeman, Peter Zeil, and Niall McCormick. Feasibility assessment of an automated, global, satellite-based flood-monitoring product for the copernicus emergency management service. *Luxembourg: Publications Office of the European Union*, 2020.

[21] Jose I Barredo, A De Roo, and C Lavalle. Flood risk mapping at european scale. *Water science and technology*, 56(4):11–17, 2007.

[22] Xiangyang Xu, Shengzhou Xu, Lianghai Jin, and Enmin Song. Characteristic analysis of otsu threshold and its applications. *Pattern recognition letters*, 32(7):956–961, 2011.

[23] Robert J Brunner and Edward J Kim. Teaching data science. *Procedia Computer Science*, 80:1947–1956, 2016.

[24] Thomas Kluyver, Benjamin Ragan-Kelley, Fernando Pérez, Brian E Granger, Matthias Bussonnier, Jonathan Frederic, Kyle Kelley, Jessica B Hamrick, Jason Grout, Sylvain Corlay, et al. *Jupyter Notebooks-a publishing format for reproducible computational workflows.*, volume 2016. 2016.

[25] Ekaba Bisong. *Building machine learning and deep learning models on Google cloud platform: A comprehensive guide for beginners*. Apress, 2019.

[26] Hamid Rezatofighi, Nathan Tsoi, JunYoung Gwak, Amir Sadeghian, Ian Reid, and Silvio Savarese. Generalized intersection over union: A metric and a loss for bounding box regression. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 658–666, 2019.

[27] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos,

D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[28] Ewa Kukawska, Stanisław Lewiński, Michał Krupiński, Radosław Malinowski, Artur Nowakowski, Marcin Rybicki, and Andrzej Kotarba. Multitemporal sentinel-2 data-remarks and observations. In *2017 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp)*, pages 1–4. IEEE, 2017.

[29] EU Copernicus. Copernicus land monitoring service. 2016.