Sentinel-1 Flood Delineation with Supervised Machine Learning

Sentinel-1 Flood Delineation with Supervised Machine Learning

Giulio Palomba
LINKS Foundation - DSISA dept.
giulio.palomba@linksfoundation.com

Alessandro Farasin
Politecnico di Torino - DAUIN dept.*
and LINKS Foundation - DSISA dept.†
alessandro.farasin@polito.it

Claudio Rossi
LINKS Foundation - DSISA dept.
claudio.rossi@linksfoundation.com

ABSTRACT

Floods are one of the major natural hazards in terms of affected people and economic damages. The increasing and often uncontrolled urban sprawl together with climate change effects will make future floods more frequent and impacting. An accurate flood mapping is of paramount importance in order to update hazard and risk maps and to plan prevention measures. In this paper, we propose the use of a supervised machine learning approach for flood delineation from satellite data. We train and evaluate the proposed algorithm using Sentinel-1 acquisition and certified flood delineation maps produced by the Copernicus Emergency Management Service across different geographical regions in Europe, achieving increased performances against previously proposed supervised machine learning approaches for flood mapping.

Keywords

Floods, Mapping, Deep Learning, Copernicus EMS, Sentinel-1, SAR.

INTRODUCTION

According to the Organization for Economic Cooperation and Development (OECD 2016), floods cause more than $40 billion in damage worldwide every year. For instance, in the U.S., losses average close to $8 billion a year. Death tolls have increased in recent decades to more than 5000 people a year (WMFD 2018), with extreme disasters such as in China's Yellow River Valley, where some of the world's worst floods have killed millions of people.

Population growth and uncontrolled urban sprawl have led to an increase in built-up areas in flood-prone regions, increasing the assets at risk and subsequently the potential adverse consequences of future flood events. Also, the capacity of rain absorption in urban areas is lower compared to natural areas, due to both the higher water runoff and the inefficiencies of the sewage system. With the growing urbanization rate, the flood severity in urban areas is likely to increase. Moreover, climate change is expected to exacerbate future flood risks as a result of the increased extreme precipitation frequency, the intensity of cyclones and the rise of the sea levels (IPCC 2012).

Immediate impacts of flooding include loss of human lives, damages to property, destruction of crops, loss of livestock, and deterioration of health conditions owing to waterborne diseases like typhoid, hepatitis A and cholera. Damage to roads and infrastructure can also cause long-term impacts, such as disruptions of clean water supplies, wastewater treatment, key infrastructures (electricity, transport communication), education and healthcare systems. When floodwaters recede, affected areas are often blanketed in silt and mud, requiring recovery activities. The water and landscape can be contaminated with hazardous materials such as sharp debris, pesticides, fuel, and untreated sewage. Furthermore, floods can also traumatize victims and their families for long periods: the loss of loved

*Dipartimento di Automatica e Informatica (DAUIN), Corso Duca degli Abruzzi, 24, 10129 Torino, Italy
†Department of Data Science for Industrial and Societal Applications (DSISA), Via Pier Carlo Boggio, 61, 10138 Torino, Italy

ones, the displacement from home, the loss of property, post-traumatic illnesses and the disruption from the social environment can produce psychological impacts that can be long-lasting.

To avoid all these tragic consequences, the adoption of countermeasures aimed at reducing the devastating effect of floods is fundamental: analyzing past catastrophes dynamic and extension, tracking risk maps and applying flood management methods or hazard risk reduction policies (as strategic retreat or resilience compliant buildings) where needed, can save thousands of lives. Moreover, responders deployed in the field in first rescue missions or floods following operations are exposed to potential dangers include electrical hazards, drowning, and exposure to hazardous materials: a near real-time mapping of the flooded area can be a huge help to volunteers employed on field, supplying an overview of the scenario in which they are moving, warning in case of dangerous situations and sites.

Given the aforementioned reasons, the capability to timely and accurately map the extension of flooded areas is of paramount importance for two main reasons: first, for the creation and update of flood hazard and flood risk maps, required to plan prevention actions aimed to reduce the impacts of upcoming emergencies; second, for the creation of a near real-time mapping service, which could be used during the emergency response phase to provide additional information to first responders.

The best way to implement a worldwide and accurate flood mapping system is to exploit the data provided by satellite networks. Among the most recent satellites are the one provided by the Copernicus Programme (Programme 2019), which is the Earth Observation initiative promoted by the European Commission (EC) in partnership with the European Space Agency (ESA). The Copernicus Programme aims to gather accurate, timely and easily accessible information to improve the management of the environment, understand and mitigate the effects of climate change and ensure civil security. It is composed of two main parts:

- The space component (Space 2019): it coordinates the delivery of satellite data generated by the Sentinel family. For the purposes of this work, we highlight two missions: the Sentinel-1 mission (Sentinel-1 2019), that consists of two twin satellites featuring a Synthetic Aperture Radar (SAR) instrument in order to provide all-weather, day and night radar images, and the Sentinel-2 mission (Sentinel-2 2019), which is designed to deliver high-resolution optical images for land services. Copernicus provides free access to all satellite data, offering great possibilities to researchers and business companies likewise.
- The service component (Services 2019): it provides services for a range of different applications such as air-quality forecasting, flood warnings, early detection of drought and desertification, oil-spill detection and drift prediction, sea-water quality, crop analysis, forest monitoring and many other services to manage and protect the environment and its natural resources and ensure civil security. The Copernicus Emergency Management System (EMS) (EMS 2019) is the service intended to map, gather and provide information for emergency management covering several hazards such as flood, earthquake, fire, etc.

The purpose of this work is to apply a traditional Machine Learning algorithm (Random Forest) and an adapted Deep Learning model (U-Net) to the problem of flood delineation from satellite data and to evaluate its performances against several validated flood maps in different geographical regions. The objective of these models will be to create automatic and near real-time mappings of the extension of floods striking worldwide, starting from data gathered by Copernicus Sentinel-1 satellites.

This is made possible by the exploitation of Machine Learning and Deep Learning models trained to learn how Delineation Map supplied by EMS are crafted with respect to satellite data and being able to autonomously (without the need of human intervention) produce mappings of never seen before satellite data of regions affected (or not) by floods, when triggered. These models will contribute to save time (output mappings are available few seconds after the input of the correspondent satellite image) and workforce for the maps crafting, being very useful in the Response and Recovery phase, providing to on-field responders precious information to operate in the first hours post-hazard, for crisis management and reaction. Moreover, an automatic detection and delineation service could lead to the gathering of a large, heterogeneous data-set, exploitable for risk analysis and elaboration of risk map in the Preparedness phase, other than damage estimation in the Recovery phase.

Moreover, we compare the performances of our model with recent works that proposed supervised machine learning approaches as well as with state-of-the-art classifiers, obtaining an improvement of the mapping accuracy of about 10% over our dataset.

This work is organized as follows: first, we review related works, highlighting the novelties of our study, then, we describe the Data Sources exploited to obtain the dataset used for the flood mapping task. Next, we describe the supervised machine learning model used and how they were trained and evaluated. Finally, we present the results obtained and outline the conclusions as well as other possible future works.

RELATED WORKS

In the last 20 years, several works have been proposed to map floods extent from satellite acquisitions. These studies present solutions differing in terms of algorithms and data sources used. Among the different satellite instruments, the most convenient one for the task of flood mapping is the SAR because it operates at wavelengths not impeded by cloud cover or lack of illumination. Hence, it can observe the Earth's surface at any time of the day or night, regardless of weather and environmental conditions, situations in which optical instruments, such as the Sentinel-2, are often not very effective. For this reason, the majority of the flood mapping literature revolves around the use of SAR data, where the most used satellite networks are RADARSAT, TerraSAR-X, COSMO-SkyMed and also Sentinel-1, with Sentinel-1 being the most convenient option given the spatial resolution offered and the free data availability. In general, most of these works aim to create an automatic near real-time system for flood detection, using different techniques to reach this target.

Early approaches are mainly based on data pre-processing, masking and thresholding (Voormansik et al. 2013, Martinis, Twele, et al. 2009, Lu et al. 2014, Ali et al. 2018), while other works use a Fuzzy Logic approach (Twele et al. 2016, Pulvirenti et al. 2011, Martinis, Kersten, et al. 2015). With the growing development of Artificial Intelligence techniques, supervised machine learning classifiers have been used to delineate flood extent using satellite data. In particular, previous works have proposed Support Vector Machines (Ireland et al. 2015), (Benoudjit and Guida 2019), Bayesian Networks (D'Addabbo et al. 2016), Artificial Neural Networks (Kussul et al. 2011), as well as deep learning techniques including Deep Belief Networks (DBN) (Bayik et al. 2018) and Fully Convolutional Neural Networks (FCN) (Kang et al. 2018).

Our work differs from all the above for two main reasons: first, we focus on a supervised machine learning approach comparing several models, including Support Vector Machines, Random Forest and one of the most recent and effective Deep Neural Network architectures for semantic image segmentation, namely the U-Net model; second, we evaluate the performances of the proposed models using a wide set of validated flood maps, which we obtained from the Copernicus EMS.

DATA SOURCES

The dataset used to train and evaluate the machine learning models is composed of a set of satellite acquisition gathered from the Copernicus Sentinel-1 mission and a set of binary masks, delineating the flooded areas and the permanent water bodies (hydrography) that we obtain from Copernicus EMS. Next, we briefly discuss these two data sources and how we created the final dataset used to train and evaluate the flood delineation machine learning models.

Sentinel-1

As stated before, we opted for Sentinel-1 because its SAR instrument operates at microwaves that are not shielded by clouds, avoiding the so-called 'cloud coverage' problem, which heavily undermines the quality and the usability of optical acquisition such as the ones provided by Sentinel-2.

Sentinel-1 is characterized by a short revisit time, which is 3 days at the equator, less than 1 day at the Arctic and is able to cover Europe, Canada and main routes in 1-3 days. It carries a single C-band SAR instrument operating at a center frequency of 5.405 GHz (ObservationScenario 2019). Among the four Acquisition Mode supported we choose the Interferometric Wide swath (IW) mode, that combines a large swath width (250 km) with a moderate geometric resolution (5 m x 20 m), with VV-VH (transmit in Vertical polarisation, receives the returning signal in both Vertical and Horizontal polarization) or VV polarization (transmit in V polarisation, receives the returning signal in V polarisation), which is the pre-defined mode over land and coastal areas.

The Sentinel-1 instrument is able to gather different images from the same series of pulses by using its antenna to receive specific polarisations at the same time: C-SAR instrument supports operation in dual polarisation (HH+HV, VV+VH). It can transmit a signal in either horizontal (H) or vertical (V) polarisation, and then receive the returning signal in both H and V polarisations. Targets on the ground have distinctive polarisation signatures reflecting different polarisations with different intensities and converting one polarisation into another: for example, volume scatterers (e.g. forest canopy) have different polarisation properties than surface scatterers (e.g. sea surface). (*ESA Sentinel Product Overview: Polarimetry* 2019)

However, Sentinel-1 raw data has a very large dimension and thus is hard to manage. For this reason, we exploit a service that provides the access to post-processed data: Sinergise Sentinel-Hub Service (*Synergise Sentinel Hub Overview* 2019), an engine that allows the end-user to interface satellite data in a simple way in form of satellite imagery, handling the complexity of management and the processing of raw data internally, thus making

products coming from the most important satellite earth observation services accessible for fast and simple browsing, visualization and analysis through a standard Web Service (*Sentinel Hub EO Browser* 2019) or API.

Copernicus EMS

Copernicus EMS is a free of charge service offered by the European Union that allows Authorized Users (National Focal Points in the EU Member States and countries participating in the Copernicus program, as well as European Commission services and the European External Action Service (EEAS)) to request a mapping activity of an Area of Interest (AOI) by triggering an Activation that can be related to an emergency event. There are two service modes:

- The Rapid Mapping: this service level guarantees that the first map is produced within hours or days from the activation in support of emergency management activities immediately following a disaster;
- The Risk & Recovery Mapping: the service level support of activities not related to immediate response, such as prevention, preparedness, disaster risk reduction and recovery phases.

The Copernicus EMS service allows downloading for free all the maps ever produced by the service in response to EMS Activations. For each EMS Activation, a variable number of delineation maps is available, depending on the number of AOIs and the time interval specified in the initial request. Delineation maps, providing an assessment of the geographic extent of the events, are derived from satellite images acquired immediately after the disaster using the first source available, through a semi-automatic approach, where human experts have to manually fine-tune and validate the maps. (*EMS Rapid Mappings Portfolio* 2019)

DATASET CREATION

Sentinel-1 data extraction

Considering that the first Sentinel-1 satellite was launched on 3 April 2014, we consider only flood delineation maps produced after this date until October 2018, when we performed this study. We download the Sentinel-1 data over the same AOI of the selected delineation maps only if the data is gathered within a maximum period of 24 hours with respect to the date of the Copernicus maps in order to ensure that both data could be comparable.

Moreover, due to the fact that Sentinel-1 satellites acquire data in stripes while following an orbit, it can happen that some acquisition matching the aforementioned conditions is not complete, covering only a portion of the requested AOI: in such cases, we had to discard the data.

The Sentinel-1 data is downloaded from the Sentinel-Hub Service as images with a spatial resolution of 10x10m using the RGB_RATIO configuration, which maps the input bands given by the different polarizations of the SAR instrument into a false RGB image, using the VV channel for red, 2 times the value of VH channel for green, and the ratio |VV|/|VH|/100 for blue (R=VV, G=2VH, B=|VV|/|VH|/100). We use the RGB GeoTiff image format that is georeferenced and orthorectified. Depending on the requested AOIs, the downloaded GeoTIFF has a size ranging between 1000-2000 x 2000-3000 pixels.

We obtained a final dataset composed of images related to flood activations in 5 countries, namely Australia (AU), Greece (GR), Ireland (IR), Italy (IT) and the United Kingdom (UK). We report in Table 1 the composition of the dataset, displaying the country to which the maps belong, the map code, and the Activation.

Ground Truth Extraction

In order to use a supervised machine learning approach, we require ground truth masks, which we built using the vector data provided in the EMS delineation maps shapefiles. For each map we create a raster flood mask, setting all pixels belonging to flooded areas equal to 255 and 0 otherwise. We apply the same approach to generate a raster hydrography mask, whose usage is described in the next section, in order to assist the models in discriminating between flooded area and permanent water bodies.

The masks were adjusted to the size of the input images, adapting the length of latitude and longitude boundings of the AOI to its dimensions.

Table 1. Copernicus EMS maps considered in the study.

Group	Activation Code	Location Name	Activation Date
AU	EMSR184	JEMALONGCONDOBOLIN	2016-09-26
GR	EMSR122	01STRYMONAS	2015-03-31
	EMSR122	04MAVROTHALASSA	2015-03-31
IR	EMSR149	05ENNIS	2015-12-04
	EMSR149	08GORT	2015-12-04
	EMSR149	13PORTUMNA	2015-12-04
	EMSR149	02ATHLONE	2015-12-04
	EMSR149	06COROFIN	2015-12-04
	EMSR149	04CASTLECONNEL	2015-12-04
	EMSR156	02LOUGHFUNSHINAGH	2016-03-04
IT	EMSR192	04ASTI	2016-11-24
	EMSR192	10CASALEMONFERRATO	2016-11-24
	EMSR192	14ALESSANDRIA	2016-11-24
	EMSR192	13SALE	2016-11-24
UK	EMSR147	01CARLISLE	2015-12-05
	EMSR147	04KENDAL	2015-12-05
	EMSR150	01YORK	2015-12-27
	EMSR150	02SELBY	2015-12-27
	EMSR150	08LEEDS	2015-12-27

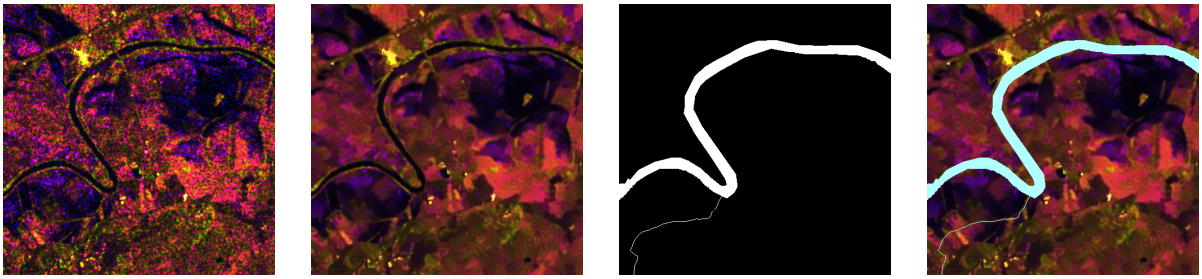
Data pre-processing

Despite the advantage given by the possibility to acquire images both at night and in bad weather conditions, a common drawback of Synthetic Aperture Radar acquisitions is that it suffers from speckle, which is a granular noise inherent to acquisition systems based on waves relevation (including SAR) and is often characterized by a smaller color distribution.

To limit these problems we apply a denoising operator called Non-Local Means (NL-means) (Buades et al. 2005), which recomputes the value of each pixel p as a weighted average of the square neighborhood of fixed size $k = 5$ centered at p where the family of weights depends on the similarity between the pixel p and the neighborhood ones. NL-means filtering takes a mean of all pixels in the image, weighted by how similar these pixels are to the target pixel in order to obtain an average value for each pixel that is the closest possible to the real value of the pixel subtracting the noise.

Another operation performed in order to evaluate possible performance improvements is the aforementioned addition of hydrography information in input to the model. In the subsection 'Ground Truth Extraction' is presented the process of crafting binary masks pointing out whether the considered pixels represent natural water bodies or not. Since in case of riverine or flash floods, flooded areas are really close to watercourse or bodies, we isolate hydrography information adding it to the input of our algorithms in order to focus the model into discriminating between a flooded area not naturally located there, and rivers, lakes or other water bodies, normally located on that slice of land (not to be classified as flooded areas). This information is fed to the models using the hydrology masks as layers overlaid above the input images.

A risk of this approach was the introduction of confusion given by adding this layer in a color already present: this could lead to an offset in the evaluation of images in which water bodies or rivers are present with respect to those where this doesn't happen. For this reason, for the hydrography layer to stack on the input images, an RGB color not present in any of these images was searched empirically and identified in [178,255,255].



(a) (b) (c) (d)

Figure 1. ***Pre-processing operations steps: (a) RGB Ratio, (b) RGB Ratio + Non Local Mean, (c) Hydrography Layer, (d) RGB Ratio + Non Local Mean + Hydrography.***

In order to make the input data, i.e. the Sentinel-1 RGB images, more homogeneous, each map is cropped into 480x480 images, using a step of 480 pixels, meaning that there is no overlap between the resulting patches. Consequently, we apply the same procedure to the corresponding flood and hydrography masks.

FLOOD DELINEATION MODELS: TRAINING AND EVALUATION

The final target of the work is to build a model able to detect flooded areas. Hence, the model not only has to classify whether the image contains a flooded area but most importantly it has to identify the flood extension, detecting which portions of the image are depicting a flood. This problem is a pixel-wise Classification task, for which the algorithm has to classify, for each pixel of the image, if it's flooded or not.

In this chapter the three models compared are explained, namely, the Stochastic Gradient Descent Classifier (SGD) presented in (Benoudjit and Guida 2019), the Random Forest (RF) and finally the U-Net.

Cross Validation

In order to evaluate the capability of each model to obtain good results on different geographical areas, we compute the model performances using a Cross-Validation approach (Stone 1974), which we achieve using a k-Fold technique, setting k to the number of groups, i.e. k=5. At each fold, we train the models on $k-1$ groups and we computed the F1-Score on the k^th one. We apply the same evaluation approach to all models, namely SGD, RF, and U-Net to have comparable results.

The 5 groups are composed by using a geographical criterion, dividing areas from Australia (AU), Greece and Macedonia (GR-MD), Italy (IT), Ireland (IR) and United Kingdom (UK), in order to generalize the results and ensure their independence with respect to different morphologies.

The number of samples for each folder is limited to 8, which is the number of images composing the group with the lowest cardinality.

Feature Construction

While U-Net takes as input an image and returns as output another picture, containing the labels for each pixel (working at picture granularity and without needing explication of the features, being a feature extractor), both the Stochastic Gradient Descent Classifier and the Random Forest take as input a series of rows, each one composed of the features of the single pixel of the picture to classify. Based on these features the model will output a classification for it, working at pixel granularity, ignoring spatial information, and needing an explicit representation of the features for the particular pixel analyzed.

Among the possible features that could be engineered for each pixel, we experimented the RGB channels value of a squared area of neighbor pixels, thus trying to provide a spatial contextualization, in the chance that flooded pixels characterized by a local pattern (for instance proximity to water mirrors, or sharp color gaps passing from land to water, or any other scheme not detectable by human eye), would be better recognized.

Taking each analyzed pixel as the center, the δ parameter, representing the number of pixels considered in every direction, is evaluated performing an iterative performance evaluation for models trained on sets characterized by increasing δ of neighboring pixels around the evaluated one. The δ value which led to the best performance appeared to be δ=3 (6x6 pixels neighbor masks). Moreover, for these two models, hydrography information was included as an additional feature when required, using a value of 0 for pixels not representing water bodies, 255 otherwise.

Evaluation Criterion

Following our previous discussion, we can consider Flood Detection as a binary classification problem. The standard way to evaluate the performance of models, in this case, is through the use of the confusion matrix, where, for this work, the 'flooded' (value=1) class is conventionally defined as positive, whilst 'not flooded' (value=0) is considered negative.

Analyzing the dataset it emerges that the number of 'not flooded' pixels is considerably higher than the number of 'flooded' pixels: precisely, the 'not flooded' pixels ratio with respect to the total pixel count is 80%. This translates, in the Machine Learning domain, in a situation of class imbalance, that can often lead to an under-fit of the model during the training phase and wrong classifications in the test phase.

For the same reason using accuracy for performance assessments is not reliable at all: considering the described imbalance, any algorithm could reach about the 80% of accuracy by just classfying all the pixels as 'not flooded', mistaking the entirety of classifications over the actual task: the detection of 'flooded' pixels.

Training the model based on this metric led to a high accuracy score, but resulted in mainly 'not flooded' classifications, definitely far from the ground truth. For this reason, the opposite of the F1-Score is adopted as the loss function, keeping the F1-Score as the assessment metric.

The F1-Score (Dice 1945) (Sorensen 1948) is defined as the harmonic mean of precision and recall:

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

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

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Considering our task, Precision points out how many pixels are classified right (TP) with respect to the totality of pixels classified as flooded by the model (TP+FP), while Recall says how many pixels are classified right (TP) with respect to the totality of flooded pixels in the satellite image (TP+FN).

In fact, since for our task the correct classification of the positive ('flooded') class is more important than the negative one ('not flooded'), and given the presence of class imbalance, while Accuracy score is inflated by a large number of true negatives (majority of 'not flooded' pixels), the F1-Score takes into account, concerning the two classes separately, how the hit classifications are relevant and how many they are with respect to the total, and is, therefore, a better choice.

Model description

Baseline: Stochastic Gradient Descent

In 2019, a solution to automate the mapping of the flood extent on SAR images using an SGD based supervised classifier was proposed by Benoudjit and Guida (Benoudjit and Guida 2019). Due to the similarity of the final task and data source used, the model defined in the work just mentioned is used as a baseline for our models and evaluated on our dataset.

The model proposed is the Stochastic Gradient Descent classifier as implemented in the Scikit-Learn library.(Classifier 2019) This estimator implements regularized linear models (by default, it fits a linear support vector machine (SVM)), with stochastic gradient descent (SGD) learning: the gradient of the loss is estimated each sample at a time and the model is updated along the way with a decreasing strength schedule (aka learning rate). For this work, the model is trained on 1000 iterations, using Hinge Loss function, with regularization term=L2 and alpha=0.0001.

Random Forest

The first model tested is a Random Forest Classifier[11] (RF), well known to be one of the most versatile Machine Learning algorithms suitable for classification.

Random Forest is one of the most popular Machine Learning algorithms, based on the training of a certain number of Decision Trees on different subsets of features over the same dataset, where each Decision Tree learns to classify new samples evaluating the feature values of some training samples which class is known. This means that the algorithm is not extracting information from data, but just analyzing and learning to classify new samples based on those yet seen: in few words, is not a feature extractor.

The number of trees used in the Random Forest Classifier is empirically chosen, by testing 50 models, with a linearly increasing number of Trees with a step of 2 between each one, starting from 2 arriving at 100 trees (the default values in the implementation proposed by the Scikit-Learn library) on the test set. Analyzing the results, optimal performances are already obtainable training a Random Forest composed by just 19 trees.

U-Net

The Deep Learning model applied is U-Net architecture: a particular instance of Convolutional Neural Network (CNN, a class of Deep Neural Networks that represent the State of the Art in the image recognition field), based on the model of Fully Convolutional Network, adapted to perform pixel-wise classification. CNNs are Deep Neural Networks, a class of Artificial Neural Network with multiple layers between the input and output layers, trained to progressively extract features of growing level from the input, passing for example from edges, lines to geometric shapes to letters or human faces in the last levels: for this reason these models are called 'Feature Extractors'. Since classic CNNs produce as output a single class label for each input image, detecting for example if there is a flood or not, they are not useful for our task, namely pixel-wise classification.

A possible solution to this problem is the one proposed by Olaf Ronneberger, Philipp Fischer, and Thomas Brox in their work 'U-Net: Convolutional Networks for Biomedical Image Segmentation' (Ronneberger et al. 2015), originally introduced for Biomedical Image Segmentation tasks.

The U-Net is a particular instance of CNN based on the Fully Convolutional Networks (FCNs) model (Long et al. 2015), consisting of two subsequential paths of layers, which form a U-shaped architecture: in the contraction path, the spatial information is reduced while feature information is increased at high-resolution level (the number of feature channels increases and input dimensions decrease); In the expansion path, on the other hand, with the concatenation between this features and the correspondent level of upsampling convolution results, a spatial localization of the features is performed (dimensionality increases and number of feature channels decreases).

The U-Net architecture used for this work is composed by 10 Convolutional Layers in the Contracting path and 8 in the Expanding path, using Rectified Linear Units (ReLUs) as activation functions, with max-pooling layers operating on a 2x2 Pool Size with Stride=2 with Batch Normalization.

The input of the network is composed of 480x480x3 images, while the output is 480x480x1 binary mask, given by a last Convolutional Layer that uses the Sigmoid function as the activation function. We trained the model over 70 epochs with Batch Size=8 for each cross-validation fold, using a standard Adam Optimizer. The original U-Net model exploits Weighted Cross-Entropy as Loss Function, requiring a particular focus on the discrimination of borders. The Weighted Cross-Entropy is in fact particularly appropriated for this purpose.

The task of this paper involves instead the detection of extensive areas characterized by a wide variety of textures and colors, in order to classify each pixel as 'flooded' or not. For this reason, a function focused on border discrimination is not the right choice: as explained in 'Evaluation Criterion', the opposite of the F1-Score is used as Loss Function.

EVALUATION RESULTS

In this section, the performance of the defined models is evaluated on the same input dataset using the same cross-validation process, and their results compared. Moreover, in order to analyze the effects of pre-processing operations on the models performances, the models are trained and tested on the dataset in three different configurations: raw False-RGB images as taken from the Sentinel-Hub Service, raw False-RGB images processed with Non-Local Mean operator, and, lastly, these one with the addition of the hydrographic information layer.

Table 2 reports the experimental results obtained, considering the different models for each of the pre-processing process operated on columns and the cross-validation folds on rows.

As it can be seen, both the pre-processing operations provide a performance boost: the Non-Local Mean Filter results in little improvements for U-Net and RF, but substantial gains for SGD. On the contrary, the addition of the hydrographic layer turns out to be a fundamental source of learning for the models proposed in this work, leading to the gain of about 8% in F1-Score. An example of the performance improvements resulting from the addition of hydrographic information is visible in Figure 2.

Table 2. F1-Score score achieved by the evaluated machine learning models.

RGB Ratio			
Group	*SGD*	*RF*	*U-Net*
AU	68,0%	69,0%	88,0%
GR-MD	67,9%	88,2%	89,0%
IR	63,4%	73,3%	74,3%
IT	66,5%	60,0%	64,5%
UK	76,0%	77,7%	82,3%
Total	**68,4%**	**73,6%**	**79,6%**
False RGB + NL-means			
Group	*SGD*	*RF*	*U-Net*
AU	78,2%	75,3%	88,9%
GR-MD	76,2%	85,3%	88,6%
IR	71,6%	72,3%	75,3%
IT	66,1%	62,1%	65,1%
UK	80,8%	77,6%	83,0%
Total	**74,6%**	**74,5%**	**80,2%**
False RGB + NL-means + Hydrography			
Group	*SGD*	*RF*	*U-Net*
AU	70,8%	88,7%	87,2%
GR-MD	75,8%	91,9%	92,6%
IR	75,9%	85,5%	85,7%
IT	74,1%	72,3%	75,1%
UK	82,1%	88,1%	89,2%
Total	**75,7%**	**85,3%**	**86,0%**

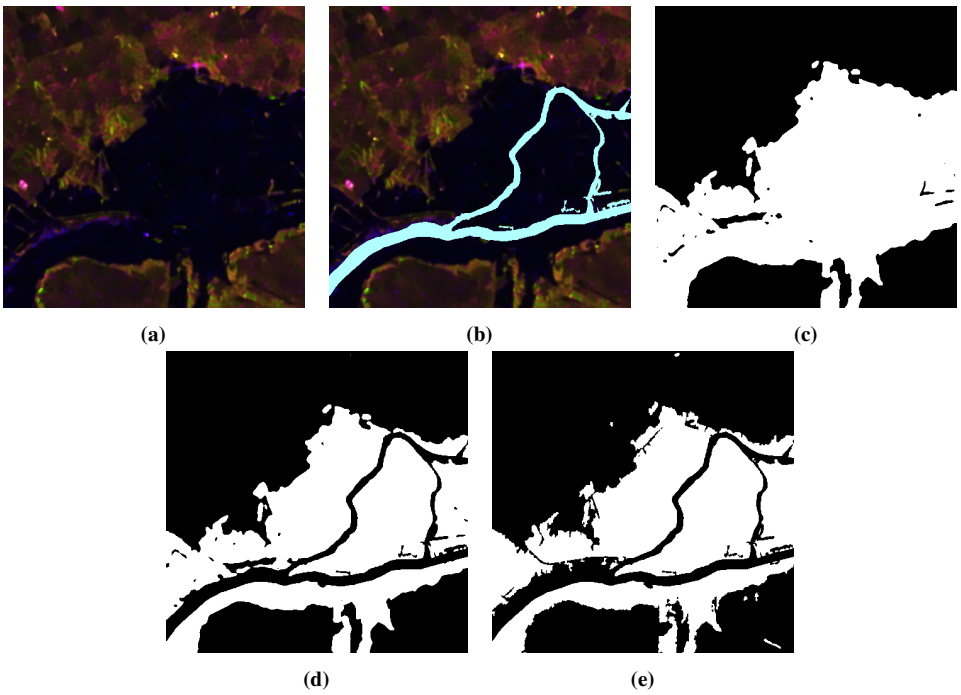
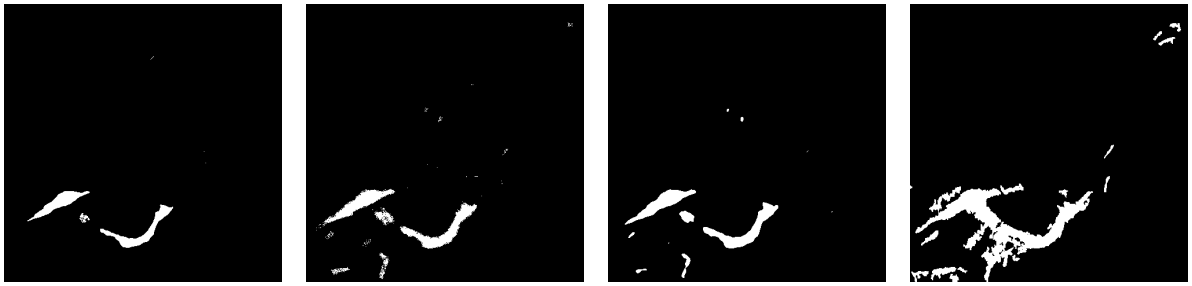


Figure 2. ***Example of performance improvement with the addition of hydrography layer for EMSR149 - 13PORTUMNA, U-Net results: (a) Denoised input without hydrography, (b) Denoised input with hydrography, (c) Classification without hydrography (F1-Score = 89,2%), (d) Classification with hydrography (F1-Score = 96,8%), (e) Ground truth.***

(a) (b) (c) (d)

Figure 3. ***Performances comparison on U-Net model best result (EMSR122 - STRYMONAS): (a) SGD (F1-Score = 90,7%), (b) RF (F1-Score = 98,4%), (c) U-Net (F1-Score = 98,6%), (d) Ground truth***



(a) (b) (c) (d)

Figure 4. ***Performances comparison on U-Net model worst result (EMSR192 - 13SALE): (a) SGD (F1-Score = 40.5%), (b) RF (F1-Score = 55,6%), (c) U-Net (F1-Score = 54,7%), (d) Ground truth***

Moreover, the U-Net is the model that obtains the best overall performances considering all the three input configurations, while RF and SGD take over on each other alternatively, depending on the cross-validation group or type of input. On average, the U-Net first, and the RF then, overcome the performances of the baseline, registering respectively, in the best-case scenario, an F1-Score of 86% and 85,3%. A visual comparison of some of the results obtained from the models is visible in Figure 3 and 4.

CONCLUSION AND FUTURE WORKS

In this paper we presented a possible solution for automatic and fast detection of flooded areas based on the ESA Copernicus Sentinel-1 mission, analyzing the State of the Art in the field, describing a methodology and two different models compared to a baseline.

Despite the fact that the purposed system could provide an important contribution in emergency management in the Response phase, it is important to remember how the creation of near real-time mapping strongly depends on the availability of satellite data acquired over the affected the areas: since Sentinel-1 mission is able to cover Europe, Canada and main routes in 1-3 days, this is obviously not always possible.

Moreover, this service could lead to the gathering of big amount of information about floods in a far less complex way, strongly supporting Recovery and Preparedness tasks as risk maps creation and damage estimation.

We described the data fetching and mask crafting processes starting from certified emergency annotations created by ESA Copernicus Emergency Management System, then listed the pre-processing operations, performed and analyzed the improvements brought by those operations, and pointed out the importance of supplying to the models the hydrography information about the affected areas.

We then analyzed the performances of the proposed models with respect to the baseline, showing how both models, U-Net in particular, reach high scores in the flood detection task, exploiting the overlay of hydrographic information layers and overcoming the baseline presented.

It is also noteworthy how, while CNNs generally need a high number of training samples in order to perform well, in this work the U-Net architecture reaches good performances exploiting a very limited set of samples. This is most certainly due to the advantage of U-Net architecture, that can be trained on fewer samples, as stated in the

work in which it is introduced, but the results reached are still beyond expectations considering the very low number of training samples used.

For this reason, planned future work is to experiment the models on a dataset composed of a higher number of samples, increasing the number of folds involved in Cross-Validation. This translates in increasing the number of Earth places (and so possible different morphologies) analyzed, to ensure the global extensibility of the method proposed.

Moreover, considering the fact that Deep Neural Networks currently represents the state of the art in the field of image recognition, the Random Forest model, indeed, showed great potential, with performances comparable to the U-Net: possible future works could focus on a feature engineering process more complex than the one used in this work, in order to improve Random Forest performances: a possible solution could be exploiting spatial concepts as proximity of the analyzed pixel to water bodies (that could help to overcome problems due to noise or color incoherence) or information coming from the Digital Elevation Model of the considered areas of interest.

The increasing number of samples to analyze and the consequential training time difficulties could be handled through the parallelization of the training on GPUs, made possible by RAPIDS, a suite of software libraries for executing end-to-end data science & analytics pipelines entirely on GPUs (Rapidsai 2019, cuML 2019).

ACKNOWLEDGEMENT

This work was partially funded by the European Commission through the SHELTER project, grant agreement n.821282.

REFERENCES

Ali, I., Freeman, V., Cao, S., and Wagner, W. (2018). "Sentinel-1 Based Near-Real Time Flood Mapping Service." In: *ISCRAM*.

Bayik, C., Abdikan, S., Ozbulak, G., Alasag, T., Aydemir, S., and Sanli, F. B. (2018). "Exploiting multi-temporal Sentinel-1 SAR data for flood extend mapping". In: *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci* 42.3, W4.

Benoudjit, A. and Guida, R. (2019). "A Novel Fully Automated Mapping of the Flood Extent on SAR Images Using a Supervised Classifier". In: *Remote Sensing* 11.7, p. 779.

Buades, A., Coll, B., and Morel, J.-M. (2005). "A non-local algorithm for image denoising". In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. Vol. 2. IEEE, pp. 60–65.

Classifier, S.-L. .-. S. (2019). *Scikit-Learn - SGD Classifier*. URL: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.SGDClassifier.html.

cuML, R. (2019). *Rapidsai cuML web site*. URL: https://github.com/rapidsai/cuml.

D'Addabbo, A., Refice, A., Pasquariello, G., Lovergine, F. P., Capolongo, D., and Manfreda, S. (2016). "A Bayesian network for flood detection combining SAR imagery and ancillary data". In: *IEEE Transactions on Geoscience and Remote Sensing* 54.6, pp. 3612–3625.

Dice, L. R. (1945). "Measures of the amount of ecologic association between species". In: *Ecology* 26.3, pp. 297–302.

EMS (2019). *Copernicus EMS*. URL: https://www.copernicus.eu/en/services/emergency.

EMS Rapid Mappings Portfolio (2019). URL: https://emergency.copernicus.eu/mapping/sites/default/files/files/CopernicusEMS-Service_Portfolio-Rapid_Mapping.pdf.

ESA Sentinel Product Overview: Polarimetry (2019). URL: https://sentinel.esa.int/web/sentinel/user-guides/sentinel-1-sar/product-overview/polarimetry.

IPCC (2012). *IPCC Managing the risk of extreme events and disasters to advance climate change adaptation*. URL: https://www.ipcc.ch/site/assets/uploads/2018/03/SREX_Full_Report-1.pdf.

Ireland, G., Volpi, M., and Petropoulos, G. (2015). "Examining the capability of supervised machine learning classifiers in extracting flooded areas from Landsat TM imagery: A case study from a Mediterranean flood". In: *Remote sensing* 7.3, pp. 3372–3399.

Kang, W., Xiang, Y., Wang, F., Wan, L., and You, H. (2018). "Flood Detection in Gaofen-3 SAR Images via Fully Convolutional Networks". In: *Sensors* 18.9, p. 2915.

Kussul, N., Shelestov, A., and Skakun, S. (2011). "Flood monitoring from SAR data". In: *Use of Satellite and In-Situ Data to Improve Sustainability*. Springer, pp. 19–29.

Long, J., Shelhamer, E., and Darrell, T. (2015). "Fully convolutional networks for semantic segmentation". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431–3440.

Lu, J., Giustarini, L., Xiong, B., Zhao, L., Jiang, Y., and Kuang, G. (2014). "Automated flood detection with improved robustness and efficiency using multi-temporal SAR data". In: *Remote sensing letters* 5.3, pp. 240–248.

Martinis, S., Kersten, J., and Twele, A. (2015). "A fully automated TerraSAR-X based flood service". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 104, pp. 203–212.

Martinis, S., Twele, A., and Voigt, S. (2009). "Towards operational near real-time flood detection using a split-based automatic thresholding procedure on high resolution TerraSAR-X data". In: *Natural Hazards and Earth System Sciences* 9.2, pp. 303–314.

ObservationScenario, S.-1. (2019). *Sentinel-1 Observation Scenario*. URL: https://sentinel.esa.int/web/sentinel/missions/sentinel-1/observation-scenario.

OECD (2016). *OECD Financial Management Of Flood Risk*. URL: https://read.oecd-ilibrary.org/finance-and-investment/financial-management-of-flood-risk_9789264257689-en#page17.

Programme, E. C. (2019). *ESA Copernicus Overview*. URL: https://www.copernicus.eu/en.

Pulvirenti, L., Pierdicca, N., Chini, M., and Guerriero, L. (2011). "An algorithm for operational flood mapping from synthetic aperture radar (SAR) data based on the fuzzy logic". In: *Natural Hazard and Earth System Sciences*.

Rapidsai (2019). *Rapidsai web site*. URL: https://rapids.ai/.

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical image computing and computer-assisted intervention*. Springer, pp. 234–241.

Sentinel Hub EO Browser (2019). URL: https://www.sentinel-hub.com/explore/eobrowser.

Sentinel-1 (2019). *Sentinel-1 Overview*. URL: https://sentinel.esa.int/web/sentinel/missions/sentinel-1.

Sentinel-2 (2019). *Sentinel-2 Overview*. URL: https://sentinel.esa.int/web/sentinel/missions/sentinel-2.

Services, E. C. (2019). *Copernicus Services Overview*. URL: https://www.copernicus.eu/en/services.

Sorensen, T. A. (1948). "A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons". In: *Biol. Skar.* 5, pp. 1–34.

Space, C. E. (2019). *ESA Copernicus Space Overview*. URL: https://sentinel.esa.int/web/sentinel/home.

Stone, M. (1974). "Cross-validatory choice and assessment of statistical predictions". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 36.2, pp. 111–133.

Synergise Sentinel Hub Overview (2019). URL: https://www.sentinel-hub.com/about.

Twele, A., Cao, W., Plank, S., and Martinis, S. (2016). "Sentinel-1-based flood mapping: a fully automated processing chain". In: *International Journal of Remote Sensing* 37.13, pp. 2990–3004.

Voormansik, K., Praks, J., Antropov, O., Jagomägi, J., and Zalite, K. (2013). "Flood mapping with TerraSAR-X in forested regions in Estonia". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7.2, pp. 562–577.

WMFD (2018). *World Mapper Floods Deaths 2001-2017*. URL: https://worldmapper.org/maps/flood-deaths-2001to2017/.