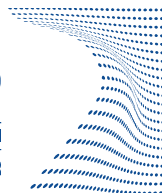




ScuDo
Scuola di Dottorato ~ Doctoral School
WHAT YOU ARE, TAKES YOU FAR



Doctoral Dissertation
Doctoral Program in Computer and Control Engineering (33.rd cycle)

Data science for geo-referenced and heterogeneous data analysis

With applications in the emergency management domain

Alessandro Farasin

* * * * *

Supervisor

Prof. Paolo Garza., Supervisor

Doctoral Examination Committee:

Prof. Annalisa Appice, Referee, Università degli Studi di Bari Aldo Moro

Prof. Giacomo Boracchi, Referee, Politecnico di Milano

Prof. Davide Cavagnino, Università degli Studi di Torino

Prof. Silvia Anna Chiusano, Politecnico di Torino

Prof. Dino Ienco, UMR Tetis - INRAE, Montpellier

Politecnico di Torino
September 02, 2021

This thesis is licensed under a Creative Commons License, Attribution - Noncommercial-NoDerivative Works 4.0 International: see www.creativecommons.org. The text may be reproduced for non-commercial purposes, provided that credit is given to the original author.

I hereby declare that, the contents and organisation of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

.....

Alessandro Farasin
Turin, September 02, 2021

Summary

Natural hazards affect every year thousands of lives, damaging cities, forests, and habitats, other than causing large economical losses. The European Commission and its Member States are involved in fighting the phenomena, cooperating and supporting several initiatives to reduce disasters impacts. Among its initiatives, the EU funds research projects to innovate and support emergency management operations.

This thesis presents several works partially carried out during two of these projects, focusing on data science approaches for the exploitation of satellite acquisitions and social media data to support emergency operations during and after wildfire and flood events.

Satellite data provide a wide and complete view of regions hit by a hazardous event. Those data are employed to localize the affected areas and to estimate their damage. In this regard, we adapted machine learning approaches and we assessed their performances in both tasks. Furthermore, we studied novel approaches, able to solve the same problems with higher performances and using less information than those adopted in the literature. Then, we operationalized the approaches through the development of a platform that is able to provide an end-to-end mapping service.

Social media data provide a large volume of information in near-real-time and can contribute to increase the general knowledge about the context of an emergency and to help coordinating emergency operations. To this end, we dealt with heterogeneous information which included textual and visual data, proposing novel challenging approaches. Firstly, we aimed to detect people potentially in danger through the evaluation of flooded sources depth and then, we wanted to detect flooded roads that could be still viable, useful for transporting emergency support to victims.

Acknowledgements

I would like to acknowledge the guidance of my supervisor, Prof. Paolo Garza, and his support during the years of my PhD. I thank him for his willingness and his insightful suggestions.

I would express my gratitude to Dr Laura López Fuentes for being part of this trip and for the chance that I had to collaborate with her. Her tenacity and precision really inspired me.

There will not be enough words to express my gratitude to Edoardo, Luca, and Mirko for being part of this journey. It has been a pleasure to research with them: without their collaboration, many research works would not be as much accurate as they are. Also, I would thank Giovanni and Giulio for the chance that I had to meet and work with them, for their support and friendship.

Thanks to all friends and colleagues from Polito and LINKS Foundation for the fruitful collaboration and the experiences made along this journey. Finally, I thank LINKS Foundation for allowing me to collaborate with its team on important research projects.

Questo documento rappresenta per me il completamento di un percorso, un percorso che non riguarda solamente l'aspetto accademico, ma riguarda anche innumerevoli esperienze di vita. Questo percorso é iniziato con una scelta che mi ha portato a lasciare un lavoro e a cambiare città per perseguire un mio desiderio profondo: inoltrarmi nel mondo della ricerca, iniziare a capire come funziona e mettermi alla prova per cercare di dare anche solamente un minimo contributo. Questa esperienza porta con sé innumerevoli sforzi, sacrifici, paure, delusioni e fallimenti, ma é anche stata ripagata con innumerevoli momenti di gioia, oltre che ad avermi permesso di conoscere meravigliose persone, tra cui brillanti ricercatori e professori, per me fonte di ispirazione.

Dedico questa tesi ai miei genitori, ai quali devo la persona che sono oggi e che ringrazio per aver sempre creduto in me.

Dedico questa tesi alla mia compagna Susanna, che ringrazio per essere stata al mio fianco e per avermi sostenuto in questo cammino.

Contents

List of Tables	XI
List of Figures	XIV
1 Introduction	1
1.1 Natural hazards: a threat for society	1
1.2 Emergency Management cycle	2
1.3 Research Contributions	3
1.4 Thesis report outline	4
2 State of the Art	5
2.1 European Copernicus programme	5
2.1.1 Early warning system	6
2.1.2 Mapping service	6
2.2 Copernicus Sentinel's missions	7
2.2.1 Sentinel-1	8
2.2.2 Sentinel-2	9
2.3 Literature review for burned areas delineation and damage severity estimation	10
2.3.1 Delineation of burned areas	11
2.3.2 Damage Severity assessment	12
2.3.3 Research contributions	13
2.4 Literature review for flooded areas delineation	14
2.4.1 Research contributions	15
2.5 Social Media data analysis during flood events	16
2.5.1 MediaEval conference: a Benchmarking Initiative for Multi-media Evaluation	16
2.5.2 Literature review for computer vision approaches applied to Twitter data about flood events	17
2.5.3 Research contributions	18

3	Post-wildfire assessment using Satellite data	19
3.1	Data acquisition and analysis	19
3.1.1	Data description and constraints	20
3.1.2	Data analysis	21
3.2	Unsupervised delineation on visible light with pre- and post-wildfire data	24
3.2.1	Problem statement	24
3.2.2	Methodology	25
3.3	Delineation assessment with Convolutional Neural Networks using post-wildfire data	29
3.3.1	Problem statement	29
3.3.2	Methodology	29
3.4	Damage severity estimation using post-wildfire data	33
3.4.1	Problem Statement	33
3.4.2	Methodology	33
3.5	Experiments	36
3.5.1	Dataset preparation	36
3.5.2	Experiments and evaluation processes	37
3.5.3	CNNs hyperparameters tuning, regularization techniques, and training process	38
3.5.4	Results on delineation problems	40
3.5.5	Results on Damage Severity estimation problem	48
3.6	Summary	55
3.7	Relevant Publications	55
4	Flood delineation assessment using Satellite data	57
4.1	Flood delineation using Sentinel-1 data	58
4.1.1	Data sources	58
4.1.2	Problem Statement	60
4.1.3	Methodology	60
4.1.4	Experiments	62
4.1.5	Results	64
4.2	Long-lasting flood event detection in cities	69
4.2.1	Dataset	69
4.2.2	Problem Statement	70
4.2.3	Methodology	70
4.2.4	Experiments	71
4.2.5	Results	72
4.3	Summary	73
4.4	Relevant publications	73

5	Rapid Mapping and Damage Assessment platform	75
5.1	The architecture: a big picture	75
5.2	Modules	77
5.2.1	Queue messaging system	77
5.2.2	Dispatcher	79
5.2.3	Controller	79
5.2.4	Geospatial Downloader	80
5.2.5	Deep Learning Platform	82
5.2.6	Cross Data Interface (xDI)	83
5.3	Technological stack	84
5.4	Performance Evaluation	85
5.5	Summary	86
6	Knowledge extraction from Social Media during flood events	89
6.1	Detection of roads and estimation of their viability in flooded areas	90
6.1.1	Problem Statement	90
6.1.2	Dataset	90
6.1.3	Methodology	93
6.1.4	Combining metadata and visual information	100
6.1.5	Evaluation and Results	101
6.2	Flood depth estimation	107
6.2.1	Problem Statement	108
6.2.2	Dataset	108
6.2.3	Methodology	109
6.2.4	Evaluation and Results	111
6.3	Summary	113
6.4	Relevant publications	114
7	Conclusions	115
A		117
A.1	Dataset	117
A.2	Dataset Land Use	119
A.3	Double-Step U-Net architecture	122
A.4	Compactness loss gradient derivation	123
	Bibliography	125

List of Tables

2.1	Sentinel-2 spectral bands description	9
2.2	Differenced Normalized Burn Ratio (dNBR) thresholds proposed by the European Forest Fire Information Service (EFFIS).	13
3.1	Data augmentation parameters.	39
3.2	Burned areas delineation results using <i>visible light</i> data. BAE is an <i>unsupervised</i> approach and leverages on in <i>pre- and post-wildfire acquisitions</i> , while CuMedVision and U-Net are supervised approaches and only leverage on post-wildfire data. (†) marks best Precision values, (★) marks best Recall values, and bold text marks best F1-Score values.	40
3.3	Burned areas delineation results using <i>all spectral bands</i> data in <i>post-wildfire acquisitions</i> . (†) marks best Precision values, (★) marks best Recall values, and bold text marks best F1-Score values.	41
3.4	Inference times of the assessed methods for the delineation task, considering input tiles of dimension 480×480 px.	42
3.5	Cross-validation performance per fold. (*) indicates the best RMSE per severity category among the three U-Net versions. (†) indicates the best RMSE per severity category, dNBR included.	50
3.6	Statistical significance between grading maps produced by the approaches shown in Table 3.5, considering different folds (shortened to the second letter) and severity levels. The Nemenyi test was performed with $\alpha = 0.05$. Check marks (✓) highlight statistical relevance (null hypothesis is rejected). Dashes (-) mark unavailable severity for the corresponding fold.	51
3.7	Inference times of the assessed methods for the damage severity estimation task, considering input tiles of dimension 480×480 px and 12 bands.	51
3.8	Average performance for severity level. (*) indicates the best RMSE per severity category among the three U-Net versions. (†) indicates the best RMSE per severity category, dNBR included.	53
4.1	Copernicus EMS delineation maps considered in the study.	59

4.2	Cross-validation results on Test case #1 (Raw data only). (†) marks best Precision values, (★) marks best Recall values, and bold text marks best F1-Score values.	65
4.3	Cross-validation results on Test case #2 (Despeckled data). (†) marks best Precision values, (★) marks best Recall values, and bold text marks best F1-Score values.	65
4.4	Cross-validation results on Test case #3 (Despeckled data + Hydrography). (†) marks best Precision values, (★) marks best Recall values, and bold text marks best F1-Score values.	66
4.5	Inference times of the assessed methods for the delineation task, considering input tiles of dimension 480×480 px and 3 channels. . .	67
4.6	Results for the subtask of "City-centered satellite sequences" of MediaEval 2019. Results refer to F1-Score metric. * Subset of the Development set, which ranges from 10% to 30% of its size.	72
5.1	Computation times of the main steps of the Deep Learning Platform, during the operationalization of the Double-Step U-Net. Also, times are compared by varying the batch size of the analyzed satellite acquisition tiles.	86
6.1	Dataset composition: for each set and class, the number of tweets is shown according to the class label. Note that only tweets presenting evidence of road are labelled for the road passability task.	91
6.2	Description of the metadata information contained in Tweets. . . .	92
6.3	F1-Scores achieved using only tweet images. Results are compared both on the official test set used in the MediaEval challenge, and on our own set. In the latter, the result of the Network ensemble of 90 models is used as reference. *Results on a subset of 50 images. . . .	104
6.4	F1-Scores achieved using only metadata. *Results on a subset of 50 images. **Results given on our own test set.	105
6.5	F1-Scores achieved using both image and metadata. *Results given on our own test set.	105
6.6	Summary of the results achieved by the proposed Double-ended network approach for the three test cases: visual information only (V), metadata information only (M), visual and metadata information (VM). The result is reported with the F1-Score metric.	106
6.7	Results of the challenge, evaluated on validation and test sets. For the visual approach, performances are also evaluated for the upper and lower branches, separately. Results refer to F1-Score metric. . .	112
A.1	Areas of Interest (AoIs) considered in this work. Each AoI reports information about the Country (ISO code), the grading map identifier for Copernicus EMS (EMSR), the coordinates of the AoI's top-left and bottom-right corners, the Pre-fire (PRE Date) and Post-fire (POST Date) Sentinel-2 acquisition dates, and the related fold. . .	118

A.2	Details on land use for the areas of interest considered in this study. This table is partial and continues in Table A.3. It reports, in hectares: Residential/Industrial areas, Arable lands, Grasslands, Forests, Heterogeneous agricultural areas, and Open spaces with little or no vegetation. For each land use type, burnt regions are reported (Burnt).	120
A.3	Land use details for the AoIs considered in this work. For the sake of space, this table is partial, and continues from Table A.2. It reports, in hectares: Pastures areas, Permanent croplands, Shrubs or herbaceous vegetation areas, Inland wetlands areas, and Woodlands. For each land use type, the areas affected by wildfire are reported (Burned).	121

List of Figures

2.1	Example of Copernicus EMS mappings regarding the wildfire event occurred in August 2017 in Torre Pedro, Spain. The official maps are available in the Copernicus Emergency Management portal, at the following link: https://emergency.copernicus.eu/mapping/list-of-components/EMSR216 . According to the specified Area of interest, (a) the Delineation map shows the extent of the burned area, while (b) the Grading map highlights the damage severity in the affected areas.	7
3.1	Map of the areas affected by wildfires in this study, divided by fold. The position of the wildfires considered in this study is determined by circles, while the color of each circle defines a specific fold: circles of the same color identify areas of interest assigned to the same fold.	21
3.2	Separability Index computed on Sentinel-2 L2A spectral bands (B01-B12) and on spectral indices used for burned areas delineation.	22
3.3	Correlation Matrix computed on burned regions included in the dataset. The spectral bands (B01–B12) refer to the post-fire Sentinel-2 acquisition, dNBR is the Delta Normalized Burn Ratio, and Ground Truth (GT), is the damage severity level used as target variable (i.e., the Copernicus EMS grading map).	24
3.4	Architecture of the BAE’s approach for segmenting burned regions. The rectangular boxes indicate the main steps of the algorithm, while in the arrows the input/output types are presented. Large boxes with dotted borders enclose the two different segmentation strategies.	26
3.5	Simplified illustration of the Self-Organizing Map training phase. The dataset is represented by the violet area. The SOM size (5, 5) is represented by the black mesh, having a neuron at each intersection. The phases are described as follows: (a) The BMU is identified: it is depicted as the yellow-circled neuron, whose radius indicates the neighbourhood function that affects the weight updates of the other neurons; (b) result of the weights update; (c) SOM’s neurons displacement after the end of the training. Illustration is from Wikipedia [135].	28

3.6	Illustration of post-wildfire acquisition in the RGB space. (a) pixel values in the raw acquisition (red dots), (b) pixel values in the pre-processed acquisition (red dots) and initialized SOM neurons (blue dots).	28
3.7	Geometrical similarities between different ground truths of (a) burned regions, (b) biological cells, the picture is from the work of O. Ronneberger et al. [127].	30
3.8	U-Net architecture. The picture is from the work of O. Ronneberger et al. [127].	30
3.9	CuMedVision architecture. The picture is from the work of H. Chen et al. [22].	32
3.10	Simplified Parallel U-Net diagram. The burned/unburned binary mask, the output of the Binary Classification U-Net, is multiplied pixel-wise with the Regression U-Net output. This operation filters out the unburned regions from the grading mask produced by the Regression U-Net. The final output is the estimate of the damage severity in the area of interest.	35
3.11	Simplified Double-Step U-Net diagram. The damage severity estimation is computed in two steps: (i) burned area delineation through the Binary Classification U-Net, and (ii) damage severity estimation by means of Regression U-Net. The Regression U-Net receives as input the Sentinel-2 L2A tile filtered with the binary segmentation mask.	35
3.12	Severity level distribution for each fold. The percentages shown in the histograms are computed considering the whole dataset. Therefore, the percentage associated to each severity level has to be considered with respect to all the other folds.	37
3.13	Burned area segmentation in a coastal area. (a1) Satellite acquisition of the burned region, realised using visible light spectrum; bands 4, 3 and 2 correspond to R, G, B channels, respectively. (a2) Ground Truth, derivated from the Copernicus EMS delineation map. White pixels represent burned regions, while black pixels represent unburned regions. (a3) BAE's prediction, using visible light data. (a4, a5) CuMedVision's and U-Net's predictions, using visible light data. (a6, a7) CuMedVision's and U-Net's predictions, using all spectral bands.	45

3.14	Burned area segmentation in a forest region, with the presence of settlements. (b1) Satellite acquisition of the burned region, realised using visible light spectrum; bands 4, 3 and 2 correspond to R, G, B channels, respectively. White pixels represent burned regions, while black pixels represent unburned regions. (b2) Ground Truth, derivated from the Copernicus EMS delineation map. (b3) BAE's prediction, using visible light data. (b4, b5) CuMedVision's and U-Net's predictions, using visible light data. (b6, b7) CuMedVision's and U-Net's predictions, using all spectral bands.	46
3.15	Burned area segmentation in an arid area, made of mountains presenting bare rocks, arable lands and shrubs. (c1) Satellite acquisition of the burned region, realised using visible light spectrum; bands 4, 3 and 2 correspond to R, G, B channels, respectively. (c2) Ground Truth, derivated from the Copernicus EMS delineation map. White pixels represent burned regions, while black pixels represent unburned regions. (c3) BAE's prediction, using visible light data. (c4, c5) CuMedVision's and U-Net's predictions, using visible light data. (c6, c7) CuMedVision's and U-Net's predictions, using all spectral bands.	47
3.16	Grading maps for the the estimation of the damage severity level. Severity levels are presented though five shades of grey, ranging from black (severity = 0) to white (severity = 4). (a1, b1) Sentinel-2 L2A acquisition; (a2, b2) Copernicus EMS grading map (GT); (a3, b3) Binary mask generated by the Binary U-Net: black and white colors indicate unburned and burned regions, respectively; (a4, b4) Thresholded dNBR, obtained from pre and post-fire acquisitions; (a5, b5) Single U-Net prediction; (a6, b6) Double-Step U-Net prediction.	54
4.1	Flow-diagram of the flood delineation process. Input data is subjected to despeckling operation, which removes noise. Cartography mask is inverted (I) and multiplied pixel-wise (\cdot) with the despeckled data, highlighting natural water sources. Finally, the model segments the flooded regions.	61
4.2	Example of how the addition of a hydrography layer improved the U-Net's performance (EMSR149 - 13PORTUMNA): (a) Despeckled data (no hydrography), (b) Despeckled data with hydrography (colored in light blue), (c) Delineation obtained from despeckled data without hydrography (F1-Score = 0.89), (d) Delineation obtained from despeckled data with hydrography (F1-Score = 0.97), (e) Ground truth.	68

4.3	Performances comparison on U-Net model best result (EMSR122 - STRYMONAS): (a) SVM (F1-Score = 0.91), (b) RF (F1-Score = 0.98), (c) U-Net (F1-Score = 0.99), (d) Ground truth	69
4.4	Performances comparison on U-Net model worst result (EMSR192 - 13SALE): (a) SVM (F1-Score = 0.40), (b) RF (F1-Score = 0.56), (c) U-Net (F1-Score = 0.55), (d) Ground truth	69
4.5	Flow-diagram of the expert system presented to the challenge.	71
5.1	Architecture of the Rapid Mapping and Damage Assessment Platform architecture.	76
5.2	Overview of the Deep Learning Platform architecture	83
5.3	Architecture of the Rapid Mapping and Damage Assessment Platform architecture, enriched with the technologies used for the implementation. Blue dashed boxes identify modules running on Docker containers.	84
6.1	Examples of images from the dataset. The first row (a–d) contains images classified as not containing Evidence of Roads (ER), while the second row (e–h) contains images classified as presenting evidence of roads and their corresponding Evidence of Road Passability (ERP). The third row (i–l) corresponds to images that were difficult to classify or wrongly classified.	93
6.2	Simplified schema of a CNN trained for the ER and for the ERP tasks. In both tasks, the network is trained by keeping the first half of its layers frozen and fine-tuning the second half.	94
6.3	Double-ended network architecture. The first part, a pre-trained CNN, is shared between the two tasks. The first half parameters are kept frozen, while the second half is fine-tuned during training. Its last layer is replaced with a Fully Connected (FC) layer, which extracts the image features. From that, two branches start one for each task. The first one solves the Evidence of Roads task, while the other one solves the Evidence of Road Passability task. At the end of each branch, a square function thresholds the prediction. To avoid the inconsistent prediction of not having evidence of any road, but having at the same time evidence of roads passability ($ER = 0$ and $ERP = 1$).	96
6.4	Correlations of selected metadata features with respect to class labels.	99
6.5	Diagram of the metadata-only approach.	100
6.6	Double-ended architecture, modified to process both visual and metadata information.	101

6.7	F1-Scores achieved through the variation of the number of ensemble models in the (a) evidence of road, and (b) evidence of passable road tasks. Three ensembling approaches are compared: (i) the majority voting strategy, (ii) the average voting strategy, and (iii) the aggregation strategy proposed at the beginning of Section 6.1.3.	103
6.8	Tweets containing informative text that helped the classifier to disambiguate the visual content for the correct prediction of the evidence of road passability task.	107
6.9	Diagram on the generation process of image crops depicting knees. .	109
6.10	Example of the output of the pose estimator algorithm. In this image, even though the legs of the person are not visible because they are below the water, the pose estimator algorithm makes an estimation of where they should be.	110
6.11	Double branched model to estimate the depth of the water by determining if the water is above or below the knee. The upper branch of the model gets as input knee crops while the lower part gets the full image.	110
A.1	Double-Step U-Net architecture.	122

Chapter 1

Introduction

This thesis covers the research activities conducted during my PhD. Most of them contributed to two European H2020 funded projects for the support of emergency management and the protection of cultural heritage against natural hazards, I-REACT (G.A. 700256) and SHELTER (G.A. 821282). In both projects, Politecnico di Torino was a member of the consortium (project partner). The research activities were carried out in collaboration with LINKS Foundation, a private research centre located in Turin.

This chapter introduces the issue of natural hazards and their impact across Europe. Then, it presents the principal steps taken to cope with an emergency and to limit the effects of a disaster, introducing how recent computer science and artificial intelligence methodologies can support the operations. Among them, the goals and ambitions of my research are clarified, also determining the borders of the studied domains. Subsequently, the main activities, contributions, and achievements performed during the PhD are summarised. Finally, a dissertation plan introduces the topics discussed in the following chapters.

1.1 Natural hazards: a threat for society

A natural hazard is a natural process or phenomenon that may cause death, injury, or other health consequences, as well as property damage, loss of livelihoods and services, social and economic upheaval, and environmental degradation.[\[109\]](#). Due to the effects of climate change, during the last decades, the impact of natural hazards increased in terms of intensity and frequency, threatening the entire world. The European Commission estimated that, between 1980 and 2017, natural hazards cost the EU more than 90,000 lives and more than €500 billion of economic losses [\[40\]](#). According to the Intergovernmental Panel on Climate Change (IPCC), extreme weather and climate events have become more frequent and intense as a

result of global warming and will continue to increase under medium and high emission scenarios. Along with human activity, global warming is having a larger role in determining wildfire regimes, with future climatic variability projected to increase the danger and intensity of wildfires in many biomes [73, 72]. Furthermore, global warming has a direct impact on precipitation: increased temperature causes higher evaporation and hence surface drying, which increases the severity and duration of droughts. Indeed, it is estimated that the water holding capacity of air increases by about 7% over 1°C warming that leads to more water vapour being retained in the atmosphere [158]. Therefore, storms are supplied with more moisture and produce more extreme precipitation events and consequently increase the risk of flood events.

The European Union and its Member States are actively involved in finding solutions to such impactful issues and take actions in several directions. The EU Civil Protection Mechanism allows the Member States and cooperating countries to share catastrophe risk information, conduct joint drills, and pool rescue troops and equipment. Moreover, through the Horizon 2020 programme, the European Union allocates funds to support research and innovation on a variety of topics, including the support to the emergency management against natural disasters [39].

1.2 Emergency Management cycle

As introduced in the previous section, natural hazards represent a real issue, that needs to be properly handled. Protection of people, property, the environment, and cultural heritage is essentially a national responsibility in the European Union. However, because disasters know no boundaries, the EU supplements, supports, and coordinates national efforts while also encouraging cross-border collaboration [41]. Contrary to popular belief, emergencies are managed both before their occurrence and after their conclusion, through the implementation of emergency management policies that aim to reduce vulnerability to hazards and help to cope with disasters. Emergency management consists of five steps:

- Prevention: actions taken to prevent an emergency. Preventive measures are designed to provide permanent protection from disasters;
- Mitigation strategy: measures aimed to reduce or eliminate the impacts and risks of hazards through proactive actions taken before the occurrence of an emergency;
- Preparedness: equipment and procedures aimed to increase a community's ability to respond when a disaster occurs. They can be used to reduce vulnerability to a disaster, to mitigate its impacts, and to respond more efficiently;

- Response: actions carried out immediately before, during, and immediately after a hazard impact, which is aimed at saving lives, reducing economic losses, and alleviating suffering. Response actions may include activating the emergency operations centre, evacuating threatened populations, opening shelters and providing mass care, emergency rescue and medical care, fire fighting, and urban search and rescue;
- Recovery: actions taken to return a community to normal or near-normal conditions, including the restoration of basic services and the repair of physical, social and economic damages. Typical recovery actions include debris cleanup, financial assistance to individuals and governments, rebuilding of roads and bridges and key facilities, and sustained mass care for displaced human and animal populations.

Satellite acquisitions and user-generated content, such as pictures and tweets, are only a subset of the potential big geo-referenced data sources available today. The proper integration of the different data sources can be profitably exploited to build accurate, descriptive and predictive models. During the PhD program, data from satellites and social media were analysed to study and research newer approaches, aiming to support the Response and Recovery phases of the emergency management cycle.

1.3 Research Contributions

During the PhD, I focused on techniques to support the phases of Response and Recovery during wildfires and floods, using either satellite or social media data. In the Response phase, my contributions concerned the monitoring of floods, intending to support the operations to enhance their knowledge about the hazard, by:

- locating affected regions delineating flooded areas, using satellite data;
- evaluating roads conditions in order to determine their viability, using social media data;
- detecting, among several people, the ones in danger, using social media data.

In the Recovery phase, my contributions concerned the census of areas affected by wildfires after their extinction. This activity is usually undertaken to evaluate the monetary impact of the hazard and to plan a proper restoration. My contributions aimed to provide models potentially applicable to any kind of ground, being independent of its morphology, and its characterization of vegetation, buildings, and biomes. The models are evaluated to be both fast and highly accurate, even in contexts where the amount of information is limited. Therefore, they involved:

- the analysis of multi-spectral data in order to assess the feasibility of the delineation activity using the information of either visible or invisible spectral bands. On one hand, we limited the available information considering only the visible spectral bands to delineate a wildfire, with the benefit of reducing the cost of the sensors needed from aerial inspections (e.g. aircraft). On the other hand, using all the spectral wavelengths may suggest higher chances to be more accurate, but at the cost of more expensive sensors;
- the evaluation of the damage severity in the affected areas, in order to identify regions mostly ruined.

1.4 Thesis report outline

This thesis deals with the studies and the analyses performed during the PhD, critically presenting the results. It is structured as follows:

- Chapter 2 presents the state of the art of the studied domains. For the satellite part, it introduces the current approaches and tools used to create maps, highlighting the areas affected by the hazard and estimating its impact. For the social media part, similar studies are presented. The limitations of literature approaches are explained and exploited to introduce the ambitions and the improvements brought by my techniques;
- Chapter 3 presents methods leveraging satellite data, assessing the performances of recent deep learning algorithms and proposing advancements to post wildfires delineation and damage severity estimation approaches;
- Chapter 4 presents ongoing flood events monitoring approaches, assessing the performances of machine learning models in different test cases, which involve preprocessing steps of satellite acquisitions and cartography maps. Moreover, a novel expert system is proposed for the detection of long-lasting floods considering a pre-defined time range;
- Chapter 5 presents the platform developed during two European projects. A general overview of the architecture is provided, then every module is explained under its functional aspect. Finally, performances are assessed to prove the practical advantages brought by the platform itself;
- Chapter 6 presents the approaches developed using social media data for the assessment of road passability, and the detection of people in danger during flood events;
- Chapter 7 concludes, summarising and reporting the principal results of the research conducted during the PhD.

Chapter 2

State of the Art

This chapter presents the literature review about the approaches for the domains of earth observation and social media analysis to support emergency management operations. The first part is related to the use of satellite data, introducing the mapping process for the delineation of regions affected by a hazard and the estimation of the damage severity. To this end, the European programme for Earth Observation is presented, with a focus on its service for the Emergency Management, which is the most acknowledged source of certified mappings for natural hazards. The chapter goes through the official mapping process, describing the methodology currently adopted. Then, the main characteristics of satellites and the data they are able to acquire are presented. Finally, recent advances from the literature are discussed.

The second part is dedicated to the use of social media data in the context of flooding events, presenting recent advances on the identification of road viability and of people in danger.

2.1 European Copernicus programme

The European Union Copernicus programme was signed in 1998 in Baveno, Italy, with the aim of monitoring and forecasting the state of the environment on the land, in the sea and in the atmosphere, based on satellite Earth Observation and in situ (non-space) sensors [42]. It is coordinated by the European Commission and implemented in partnership with the Member States, the European Space Agency (ESA), the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT), the European Centre for Medium-Range Weather Forecasts (ECMWF), EU Agencies and Mercator Océan. It allows free access to terabytes of reliable and up-to-date information to anyone interested, including large companies, scientists, civil protection and disaster risk management bodies.

With a focus on emergency management, Copernicus includes the Emergency Management Service (EMS), which provides two types of services: (i) early warning, and (ii) mapping service.

2.1.1 Early warning system

The early warning and monitoring component of the Copernicus Emergency Management Service is based on continuous observation and forecasts at European and global levels of floods, droughts and wildfires. The European Flood Awareness System (EFAS) provides overviews on ongoing and forecasted floods in Europe up to 10 days in advance. The European Forest Fire Information System (EFFIS) provides near real-time and historical information on forest fires and forest fire regimes in the European, Middle Eastern and North African regions.

2.1.2 Mapping service

The mapping service consists of either Rapid Mapping or Risk & Recovery Mapping. The Rapid mapping service can provide geospatial information within hours or days from a request in order to support response to emergency situations, during a disaster. Risk & Recovery Mapping offers geospatial information that can feed into multiple disaster risk prevention, preparedness, reduction and recovery activities.

Results of a mapping request can be delineation or grading maps. Delineation maps provide an assessment of the event extent, considering the area of the affected regions. Grading maps provide information about the damage grade, its spatial distribution and extent. The grading product is a superset of the delineation product as it contains the event type, impact extent (delineation) and the damage grading. The damage grading defined according to five intensities: “No damage”, “Negligible to slight damage”, “Moderately Damaged”, and “Highly Damaged”. Both delineation and grading maps are derived from satellite images acquired: (i) immediately after the disaster using the first source available, the emergency event, for the rapid mapping, and (ii) after the event conclusion, for recovery and restoration purposes. An example of the mappings is shown in Figure 2.1.

Both delineation and grading maps, providing an assessment of the geographic extent of the events, are derived from satellite images through semi-automatic approaches, where human experts have to manually fine-tune and validate the maps [36].

In addition to the Emergency Management Service, other services of the Copernicus programme can support risk prevention and management with relevant data on climate, land cover and its changes, land use, water cycle, the safety of infrastructure, and marine safety.

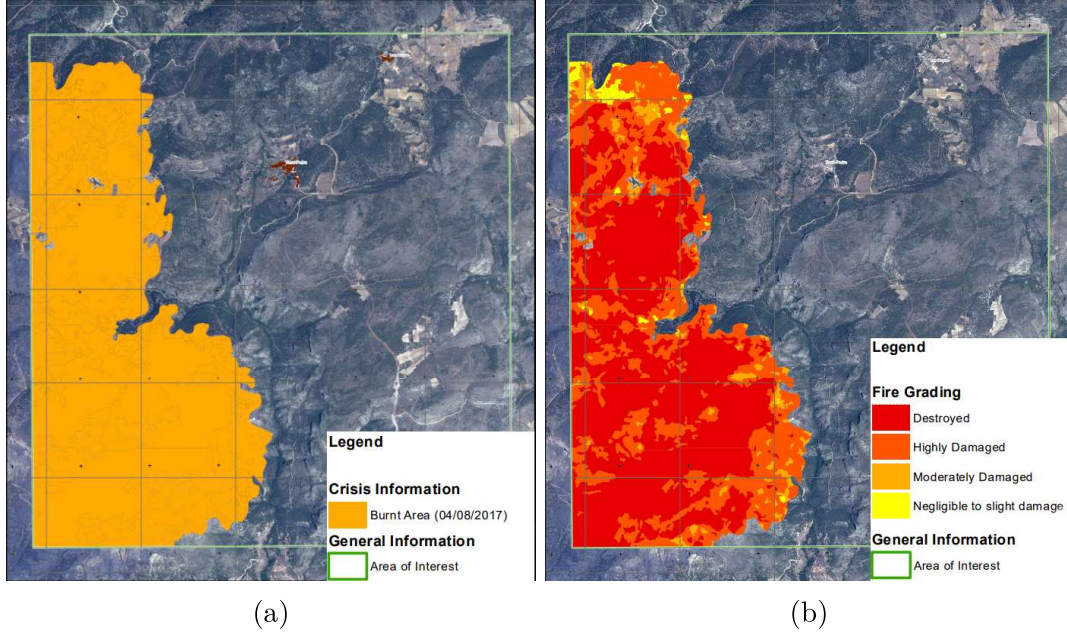


Figure 2.1: Example of Copernicus EMS mappings regarding the wildfire event occurred in August 2017 in Torre Pedro, Spain. The official maps are available in the Copernicus Emergency Management portal, at the following link: <https://emergency.copernicus.eu/mapping/list-of-components/EMSR216>. According to the specified Area of interest, (a) the Delineation map shows the extent of the burned area, while (b) the Grading map highlights the damage severity in the affected areas.

2.2 Copernicus Sentinel’s missions

The Sentinel’s missions were introduced as part of the Copernicus programme through the partnership between the European Commission and the European Space Agency (ESA). Every Sentinel mission consists of a constellation of two satellites to fulfil revisit and coverage requirements. They carry technologies specifically projected for the operational needs of the Copernicus programme. The missions are briefly introduced as follows:

- Sentinel-1: focused on land and ocean monitoring, Sentinel-1 is composed of two polar-orbiting satellites, and performs Radar imaging, enabling their acquisition regardless of the weather conditions;
- Sentinel-2: the objective of Sentinel-2 is related to land monitoring, it is composed of two polar-orbiting satellites providing high-resolution optical imagery. Vegetation, soil and coastal areas are among the monitoring objectives;

- Sentinel-3: the objective of Sentinel-3 is marine observation, and it studies sea-surface topography, sea and land surface temperature, ocean and land colour. Composed of three satellites, the mission's primary tool is a radar altimeter, but the polar-orbiting satellites carry multiple instruments, including optical imagery.
- Sentinel 4, 5, 5P: all three missions are dedicated to providing continuous monitoring of the composition of the Earth's atmosphere, focused to monitor the air quality. In addition, Sentinel-5P extends its applications to climate forcing, ozone and UV radiation monitoring.

In the context of floods and wildfires, Sentinel-1 and Sentinel-2 satellites are adopted by Copernicus EMS and the research community. Therefore, in the next section, a brief description of their characteristics is introduced.

2.2.1 Sentinel-1

The Sentinel-1 satellites are equipped with the Synthetic Aperture Radar (SAR), an instrument that operates at radio waves that are not shielded by atmospheric conditions, avoiding imagery occlusions or disturbances like clouds and fog. SAR gathers different images from the same series of pulses by using its antenna to receive specific polarisations at the same time. It can transmit a radar signal in either horizontal (H) or vertical (V) polarisation, and then receive the returning signal in both H and V polarisations, supporting operations in dual polarisation: HH+HV or VV+VH. Targets on the ground have distinctive polarisation signatures reflecting different polarisations with different intensities and converting one polarisation into another: for instance, volume scatterers (e.g. forest canopy) have different polarisation properties than surface scatterers (e.g. sea surface) [37]. It supports four acquisition modes: StripMap (SM), Interferometric Wide Swath (IW), Extra Wide Swath (EW), and Wave (WV). SM, IW and EW products are available in single (HH or VV) or dual polarisation (HH+HV or VV+VH), while WV is single polarisation only (HH or VV). SM mode acquires data at the resolution of 5m x 5m per pixel, it is only used for small islands and on request for extraordinary events such as emergency management. IW mode supports the resolution of 5m x 20m per pixel and it is largely adopted with VV+VH polarisation for land monitoring. EW mode supports the resolution of 20m x 40m, it is primarily used for wide-area coastal monitoring including ship traffic, oil spill and sea-ice monitoring. Finally, WV supports the resolution of 5m x 5m per pixel and, with VV polarisation, it is adopted for open ocean monitoring [138, 137, 136, 139]. Sentinel-1 satellites' revisit time, the time elapsed between two observations of the same point on earth is 3 days at the equator and less than 1 day at the Arctic. Therefore, it is able to cover Europe, Canada and main routes in 1-3 days. [112].

Table 2.1: Sentinel-2 spectral bands description

Band	Description	Central Wavelength (μm)	Spatial resolution (m)
1	Coastal aerosol	0.443	60
2	Blue	0.490	10
3	Green	0.560	10
4	Red	0.665	10
5	Vegetation red edge	0.705	20
6	Vegetation red edge	0.740	20
7	Vegetation red edge	0.783	20
8	Near Infrared (NIR)	0.842	10
8A	Narrow NIR	0.865	20
9	Water vapour	0.945	60
10	Short wavelength infrared (SWIR)	1.375	60
11	Short SWIR (SSWIR)	1.610	20
12	Long SWIR (LSWIR)	2.190	20

2.2.2 Sentinel-2

Sentinel-2 satellites are equipped with high-resolution, multi-spectral imaging sensors and a revisit time (~ 2 -3 days at European latitudes), aimed at monitoring variability in land surface conditions. Each satellite carries an optical instrument payload that samples 13 spectral bands, at different spatial resolutions: four bands at 10 m, six bands at 20 m and three bands at 60 m [38, 108].

Band 1 is sensible to the concentration of aerosols in the atmosphere, which may be used to refine the atmospheric correction procedures and it can provide a closer inspection of the coastal and inland waters. Bands 2, 3, 4 are sensible to the visible light, representing the image in the classical red, green, and blue configuration. Bands 5, 6, 7 are sensible to the vegetation red edge, which is a region in the red-NIR transition zone of vegetation reflectance spectrum and marks the boundary between absorption by chlorophyll in the red visible region, and scattering due to leaf internal structure in the NIR region. Band 8 is sensible to vegetation type, density, water content, and general plants health. Band 8A is designed to avoid contamination from water vapour and be sensitive to iron oxide content for soil. Bands 9 and 10 are sensitive to water vapour absorbing the light that comes from the earth surface. Band 9 presents a weak water absorption level and is used for atmospheric correction. Band 10 presents strong water absorption and it is commonly used to detect high clouds, such as cirrus. Finally, bands 11 and 12 are usually adopted for applications such as snow, ice or cloud detection [26, 161, 92]. A summary of the spectral bands, their respective wavelengths and resolutions are shown in Table 2.1.

After the acquisition, the raw data is subjected to four preprocessing steps before being available to the users. The first available product, named Level-1C (L1C), is composed of $100 \times 100 \text{ km}^2$ tiles (ortho-images in UTM/WGS84 projection). It results from using a Digital Elevation Model (DEM) to project the image in cartographic geometry. Per-pixel radiometric measurements are provided in Top Of Atmosphere (TOA) reflectances along with the parameters to transform them into radiances. Level-1C products are resampled with a constant Ground Sampling Distance (GSD) of 10, 20 and 60 m depending on the native resolution of the different spectral bands.

Another processing level, built on top of L1C products, is Level-2A (L2A). L2A products are subjected to atmospheric correction, providing Bottom Of Atmosphere (BOA) reflectance images, which result in clearer ground applications. In L2A products, Band 10 is omitted, as it does not contain surface information.

This introduction on satellites, their bands and products is needed for deeply understanding the scientific literature on burned areas and floods delineation, which are the topics of the next sections.

2.3 Literature review for burned areas delineation and damage severity estimation

In general, Bands 8 and 12, corresponding to NIR and LSWIR wavelengths, have been found to provide stronger burned area discrimination than visible wavelengths, and most burned area mapping algorithms are based on detecting decreased reflectance at these wavelengths. However, other non-fire surface changes, such as shadows or agricultural harvesting, can generate comparable spectral shifts, which, depending on the methodology and wavelengths employed, can result in misleading burned area detection [129, 130]. Visible wavelengths are more vulnerable to atmospheric noise and smoke aerosols than longer non-visible wavelengths. For that reason, they are not generally suited for burned area identification [95]. In literature, it is common to leverage spectral indices, computed from the combination of Sentinel-2 spectral bands, to highlight burned regions and distinguish among damage severities. Ideally, if a spectral index is appropriate for detecting a physical change in the area of interest, then both the changed region and the direction of the changing displacement in spectral feature space should have a linear relationship [70].

2.3.1 Delineation of burned areas

Regarding the delineation of burned areas, the Normalized Burnt Ratio (NBR) is one of the most popular indexes adopted in the field. As specified in Equation 2.1, it involves Bands 8, which naturally reacts positively to leaf area and plant productivity, and Band 12, which positively responds to non-vegetated surface characteristics. Band 12 presents high water absorption by green vegetation and moist surfaces, including wet soil and snow, just the opposite of Band 8. Because NBR measures the difference between B8 and B12, it is positive when B8 is greater than B12. This is the case over most lush vegetated areas. When it is near zero, B8 and B12 are about equal, as occurs with clouds, and drier soils or rock. When NBR is negative, this is suggested by severe water stress in plants and the non-vegetative traits created due to burns [97].

$$\text{NBR} = \frac{\text{NIR} - \text{LSWIR}}{\text{NIR} + \text{LSWIR}} = \frac{\text{B08} - \text{B12}}{\text{B08} + \text{B12}} \quad (2.1)$$

Other indexes used for the same purpose of NBR are the Mid Infrared Burn Index (MIRBI) and the Normalized Burnt Ratio 2 (NBR2), which were designed for shrub-savannah vegetation type, where NIR wavelengths are less effective. The indexes, shown in Equations 2.2 and 2.3, were developed using Bands 10 and 12 and its performance was proven to be relatively stable over time in savannah ecosystems, with promising results in the assessment of post-fire vegetation recovery in the shrublands of California and sclerophyll forests of Australia [133, 159, 151, 68, 128].

$$\text{MIRBI} = 10 \cdot \text{LSWIR} - 9.8 \cdot \text{SSWIR} + 2 = 10 \cdot \text{B12} - 9.8 \cdot \text{B11} + 2 \quad (2.2)$$

$$\text{NBR2} = \frac{\text{SSWIR} - \text{LSWIR}}{\text{SSWIR} + \text{LSWIR}} = \frac{\text{B11} - \text{B12}}{\text{B11} + \text{B12}} \quad (2.3)$$

Also, in 2018, Filipponi et al. proposed the Burned Area Index for Sentinel-2 (BAIS2), a revisited version of the Burned Area Index (BAI) largely used in the past in with Landsat satellites. BAIS2, introduced in Equation 2.4, considers the use of a band ratio in the red-edge spectral-domain (Bands 6 and 7), which aim to describe vegetation properties, combined with a band ratio that involves Bands 8A and 12, recognized to be efficient in the determination of burned areas [51]. Compared to other satellites such as Landsat and Spot, Sentinel-2 reduced the length of bands sensitive to NIR, identified in Band 8 and Band 8A. They are designed to avoid contamination from water vapour yet still be able to represent the NIR for vegetation and be sensitive to iron oxide content for soil [25]. However, as admitted by the author of the BAIS2 index itself, the adoption of Band 8A instead of B8 in the calculation of spectral indices for burned areas is demonstrated not to bring significant advantages.

$$\text{BAIS2} = \left(1 - \sqrt{\frac{\text{B06} \cdot \text{B07} \cdot \text{B8A}}{\text{B04}}}\right) \cdot \left(\frac{\text{B12} - \text{B8A}}{\sqrt{\text{B12} + \text{B8A}}} + 1\right) \quad (2.4)$$

Although spectral indices may produce good burned area discrimination for a particular location and time they may not perform well elsewhere. In order to determine the suitable index, according to the examined area of interest, the Separability Index (SI), presented in Equation 2.5 is used to estimate the effectiveness of individual bands and spectral indices to discriminate between burned and unburned land. The separability index, also known as normalized distance [133], is defined as follows:

$$\text{SI} = \frac{|\mu_b - \mu_u|}{\sigma_b + \sigma_u} \quad (2.5)$$

where μ_b and μ_u are the mean values, and σ_b and σ_u are the standard deviations of the considered indices for burned and unburned areas. The higher the separability index SI, the better the discrimination. Values of SI higher than one indicate good separability, while values lower than one represent a large degree of histogram overlap between the burned and unburned classes. Once the best index has been assessed, the burned area is discriminated from the unburned region through the definition of a threshold value, that may significantly vary from place to place. Therefore, the definition of that value is usually subjected to manual supervision.

Classical Machine-Learning algorithms, such as Random Forest, Artificial Neural Networks (ANN) and Support Vector Machines are also employed, leveraging the spectral bands and indexes computed before and after the wildfire event. Among them, it is worth mentioning: (i) approaches for specific regions, such as forests, or deserts [64, 27, 71], (ii) generally applicable approaches [11, 124, 125, 141]. For instance, in their work Ramo et al. carefully assess the performances between Decision Trees, Random Forests, Feed-Forward Neural Networks, and Support Vector Machines, using pre and post-wildfire MODIS acquisitions with the resolution of ~500m per pixel [124].

2.3.2 Damage Severity assessment

Inferring damage severity from a burned area is an advancement of the delineation task. In their work, Key et al. proposed the composite burn index (CBI): an in-field method to evaluate burn severity [77]. The CBI provides a semi-quantitative index of severity, being computed by considering measurable aspects, evaluating in the burned area: (i) material lying on the floor, (ii) short shrubs and small trees (<1 m tall) (iii) tall shrub and sapling trees (<5 m tall), (iv) intermediate trees (5–20 m tall), and (v) large trees (>20 m tall). All those properties are combined to create the Composite Burned Index (CBI), which is the best approximation of damage severity: it is far fine-grained than the grading maps provided by EMS, but its

computation is not feasible at a large scale because too much data must be collected manually. In the same article, the authors suggest the use of the Differenced Normalized Burn Ratio (dNBR), computed as follows: $dNBR = NBR_{PRE} - NBR_{POST}$. The dNBR requires to compute the NBR over an image acquired before the wildfire event, comparing it with another acquisition taken after the wildfire exhaustion. Then, the dNBR is quantized to obtain different ranges of severity. This approach is considered very accurate and it is still largely shared by the scientific community: the Copernicus programme itself, through its EFFIS programme, provides adjusted thresholds for dNBR in order to identify four different severity levels, shown in Table 2.2 [88, 49, 33].

Table 2.2: Differenced Normalized Burn Ratio (dNBR) thresholds proposed by the European Forest Fire Information Service (EFFIS).

Fire Severity Class	Range of dNBR
Unburned/Very Low	< 0.1
Low	$0.1 - 0.255$
Moderate	$0.256 - 0.41$
High	$0.42 - 0.66$
Very High	> 0.66

Techniques based on manual or automatic thresholding are widely applied because they are computationally fast and efficient [166, 152, 48, 5, 86, 102]. However, they rely on acquisitions taken before the wildfire event, which might be trivial to obtain, considering the average revisit time. In this respect, weather conditions can widely affect the atmosphere and the morphology of the ground: clouds can cover the area of interest, vegetation can be subjected to variations, especially across seasons, etc. All these aspects can make pre- and post-wildfire acquisitions harder to compare. Moreover, most of the works in literature are validated in few places, possibly because of the need to manually assess hyperparameters and thresholds according to the morphology of the area of interest. Also, Copernicus EMS requires days to complete a mapping process, which always requires both pre- and post-wildfire acquisitions. Manual intervention is usually foreseen in the mapping process since homogeneity between pre and post-event data is not commonly guaranteed [24].

2.3.3 Research contributions

My contributions in this topic concern the assessment of the delineation capacity of neural networks, leveraging (i) only visible wavelengths, (ii) all spectral wavelengths. The assessment on visible wavelengths concern only Sentinel-2 bands

2, 3, and 4: it aims to discover whether a limited amount of data could be sufficient to accomplish the task. In the positive case, more frequent inspections with low-cost cameras mounted on aircraft could significantly improve the mapping activities with fast delivery of the results. Then, the results are compared to the ones obtained by considering all the spectrum. In this case, I evaluated the performance achievable using just the actual acquisition, avoiding pre-fire images. Finally, I worked on the severity estimation task, proposing a novel approach able to operate with just post-wildfire acquisitions. Using pre-wildfire images concern very often a manual supervision to avoid the presence of undesirable phenomena in the area of interest, such as: (i) the presence of high atmospheric noise caused by fog, pollution, or water vapour, (ii) cloud coverage over the burned area, (iii) variation of soil and vegetation, due to seasonal changes. Therefore, using just post-wildfire acquisitions for evaluating the affected areas and estimating the damage severity speeds up the mapping process by halving the required information and reducing the manual intervention.

2.4 Literature review for flooded areas delineation

During flood events, clouds represent occlusions from the satellite point of view. In those scenarios, Synthetic Aperture Radar (SAR) sensors have been extensively used in the last decade to monitor many flooding events by taking advantage of their ability to operate independently of cloudy conditions 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 concerns the use of satellites equipped with SAR instrument, like RADARSAT, TerraSAR-X, COSMO-SkyMed and also Sentinel-1. During the acquisition process, SAR data are inherently affected by speckle noise, which requires proper despeckling operations, or filtering techniques, to reduce noise preserving all the relevant scene features, such as radiometric and textural information. Generally, filtering techniques are drawn from signal processing topics, and concern Gaussian filtering [59], Frost filtering [164], Gamma-MAP filtering [99], and more recently, Nonlocal means filtering [142, 163].

Then, the literature presents several threshold-based works for water segmentation, based on backscatter histogram thresholding, region growing, change detection, and fuzzy logic approaches. Backscatter is the portion of the outgoing radar signal that the target redirects directly back towards the radar antenna. Generally, water presents low backscatter, due to its smooth surface, while ground or urban areas present higher values [85]. Thresholds are usually determined by analyzing the histogram of SAR backscatter intensity and estimating the probability distributions of water and non-water pixels. In regions with a considerable amount of surface

water, the histogram of SAR images presents two maxima, corresponding to water and non-water regions, respectively. Among the published approaches, tiling and thresholding [98, 23] or the combination of Otsu thresholding and region growing [83, 93, 122] have been recognised among the most successful methods. The delineation of flooded areas requires some knowledge about natural water sources to avoid their misclassification with the flood. On purpose, the literature proposes techniques based on change detection or fuzzy-logic approaches. Change detection approaches leverage two (or more) acquisitions of the same region during the time, to spot the areas subjected to the highest modifications [134, 58]. Instead, fuzzy logic approaches combine SAR data with different sources of information, like Digital Elevation Maps (DEM) and Water Body data [160, 121, 98], to remove potential water-lookalikes [160]. DEMs, maps representing the elevation of the ground with respect to the sea level, are used to compute the Height Above Nearest Drainage (HAND) index, a binary exclusion mask calculated to separate flood-prone from non-flood prone areas data [111]. Water Body data are water masks depicting permanent water bodies (related to normal water levels), and are used to identify inundated areas [19].

With the growing development of Artificial Intelligence, supervised machine learning classifiers have been used to delineate flood extent. In particular, recent works have proposed the use of Support Vector Machines [74, 149, 4], and Random Forest [3, 146].

2.4.1 Research contributions

Like in the burned area delineation, approaches are generally validated in few areas, due to: (i) the limited availability of large datasets, and (ii) their applicability in large areas with high resolution. In this context, my research contribution consisted in the assessment of recent techniques in several places across Europe, with the purpose to limit the amount of data needed to accomplish the delineation task. Methods that demonstrate to work with high reliability in different and heterogeneous areas can be operationalized and adopted in production environments. Moreover, avoiding pre-flood acquisitions limits manual supervision. Usually, those acquisitions are carefully chosen to be used as a reference for the current flood event: for instance, it must be verified that in the upcoming dates before the pre-flood date the weather conditions were stable and that there were not abundant rains or storms that could have altered the dimensions of the natural water bodies. The work assessed the performance achieved by machine learning and deep learning approaches, providing an ablation study on the gain of accuracy provided by means of pre- and post-processing steps, and the use of cartography as extra data. Performances are evaluated with official flood delineation maps, provided by Copernicus EMS.

2.5 Social Media data analysis during flood events

In recent years, smartphones and IoT devices have become even more important in our daily lives. Social media in general represent a new way for us to communicate: among private profiles, such as Twitter, Instagram, or Facebook, it is easy to find personal information about our activities, hobbies, thoughts, but also news about real-time events. During natural disasters, social media has become a massive source of data from which, if properly processed, valuable information for emergency management can be extracted, especially during the response phase [34, 143, 94].

In the context of flood events, approaches in the literature are generally focused on the detection of contents about the emergency, in order to build systems able to filter relevant posts [67, 53, 50, 81, 91]. Then, considering only flood-related content, it is possible to focus on information about the context of the emergency. For this purpose, this section explores the literature about computer vision approaches applied to social media posts on flood events.

2.5.1 MediaEval conference: a Benchmarking Initiative for Multimedia Evaluation

Most of the research activities that I personally carried out in the context of social media are related to the MediaEval conference. MediaEval is a benchmarking initiative dedicated to evaluating novel approaches for multimedia access and retrieval. Every year, several tasks are proposed that can be solved by combining different types of information (e.g. images, text, audio, video, metadata, geolocation, etc.) by using multimodal approaches. Among the tasks proposed in the conference, I participated in the "Multimedia and Satellite Task for Emergency Response during Flooding Events", during the editions of 2018 and 2019. The Multimedia and Satellite task focuses on flooding events and uses social media as a source of visual, text and metadata information to retrieve flood-related content. Its main focus is about floods, presenting different goals every year. In 2018 the goal was detecting roads and assessing their passability based on the severeness of the flood (if any) using tweeter posts with images. In 2019 the goal was determining if, in case of a flood, the tweet presented people potentially in danger, estimating whether the water level was below or over the knee of at least one person. Even if the objectives are different, the works present similarities in dealing with the multimodal data. Also, given the specificity of both tasks, the works presented in MediaEval represent the current state of the art.

2.5.2 Literature review for computer vision approaches applied to Twitter data about flood events

Tweets contain two types of information: metadata and pictures. Metadata is composed of textual information (e.g. the text of the post, title) and punctual information, such as GPS-coordinates, creation date, author reference, etc. (an exhaustive explanation of the parameters is given in Chapter 6), while the images are generally .png or .jpg pictures. Therefore, multimodal approaches are required to exploit properly the data and infer meaningful information.

Several approaches exploit Metadata information. In their work, Zhao et al. [46] manually defined a set of rules that leverage tweets' text, looking for n-grams of lexical items expected to occur in tweets related to road passability. Other works, like Hanif et al. [63] and Mounztidou et al. [107] started with a pre-processing of the tweet texts: first removing hyperlinks, punctuations and symbols and performing the word tokenization, then removing the stop-words and performing word stemming. The processed information was enriched by adding other metadata features, like user tags. Another work by Kirchknopf et al. [79] proposed to check the metadata language feature and it incremented the number of English tweets by translating those written in other languages. The words are then translated into a vectorial representation with word embeddings (i.e. leveraging fasttext [46, 10], Word2Vec [105] or GloVe [118]) and/or by computing the Term Frequency - Inverse Document Frequency (TF-IDF) [14]. For final classification, techniques like Spectral Regression-based Kernel Discriminant Analysis (SRKDA) [17], Support Vector Machines (SVM) or Convolutional Neural Networks (CNNs) for sentence classification [78] have been used.

In order to deal with pictures, two approaches were mainly adopted: (i) using visual descriptors and (ii) extracting features from pre-trained CNNs. In the first case, several descriptors were already available from the dataset: Color and Edge Descriptor (CEDD) [20], Color Layout (CL) [76], Fuzzy Color and Texture Histogram (FCTH) [21], Edge Histogram (EH) [117], Joint Composite Descriptor (JCD) [169] and Scalable Color Descriptor (SCD) [96]. In the latter case, state-of-the-art CNNs such as AlexNet [80], DenseNet201 [69], InceptionV3 [156], InceptionResNetV2 [155], ResNet [66], VGG [145] or YOLOv3 [126] were taken pre-trained on popular and wide datasets such as ImageNet [29], Places365 [171] or VOC [57]. Those datasets present information about single entities (ImageNet, VOC) or common contexts about places (Places365): therefore, the last hidden layer of the pre-trained models provides a vectorial representation for each picture. Most of the MediaEval proposed works exploited this approach, extracting the visual features by feeding to the network(s) the pictures from the dataset of the challenge and taking their last layers activations [46]. The features and descriptors mentioned above can be referred to as *global* features, as they are extracted using the whole picture. In addition, Bischke et al. [7] and Zhao et al. [170] combined

also information related to single entities (i.e. cars, boats, persons), named *local* features. Then, the extracted features are used for classification in several manners. One option [46, 107, 30] was to feed them as input for a neural network having few fully connected layers and using softmax for classification. Other approaches used other state-of-the-art machine learning algorithms, such as Support Vector Machine (SVM) [63, 79, 7, 170, 132], Multinomial Naive-Bayes, Random Forest and SRKDA [63]. Finally, *early* and *late fusion* strategies were utilized to combine metadata and visual information. *Early fusion* aggregates features before the classifier computes them, whereas *late fusion* averages the predictions of the techniques developed separately for the two domains.

2.5.3 Research contributions

The MediaEval conference proposed novel challenges, that helped to define benchmarks and metrics that allowed the participants to compare themselves with other approaches, ensuring the research quality of their work. Sharing the same goals with other teams helped to network with researchers interested in similar topics as well as getting to know other approaches to solve those problems. Beyond personal experience, we were acknowledged as the winners of the Multimedia and Satellite Task in 2018 with the work on roads passability. After the conference we furtherly investigated and improved the approach making it lighter, and suitable for operational purposes. Also, we participated in the conference in 2019, where we provided a solution for detecting critical flood depth and inferring people potentially in danger.

Chapter 3

Post-wildfire assessment using Satellite data

This chapter presents my works on the automatic mapping of burned areas from satellite acquisitions, once the wildfire is completely extinguished.

As introduced in Section 1.3, this activity is carried out in the Recovery Phase of the Emergency Management cycle. It consists of the realization of cartographic maps, reporting the affected areas and estimating the severity of the damage caused by a wildfire. The census of those areas is needed for estimating the economic damage and for planning a complete restoration of the environment. During my PhD, I worked on models that could improve the mapping results and limit the need for data, to speed up the mapping process.

The chapter is structured as follows. Section 3.1 introduces and analyses the data used in the assessments. Sections 3.2, 3.3, and 3.4 present approaches to solve the problems of burned areas (i) delineation and (ii) damage severity estimation. Section 3.5 describes the experiments, introducing the dataset preprocessing steps, the testing process and discussing the results.

3.1 Data acquisition and analysis

As introduced in Chapter 2, through its Rapid Mapping and Risk & Recovery Mapping services, Copernicus EMS provides (i) delineation maps, which define the perimeter of the event extent, and (ii) grading maps, which add a layer of information about the severity of the damage to delineation maps. Both refer to an Area of Interest (AoI) and to a reference date (marked as “Situation as of” in the map’s cartouche). The AoI is a rectangular region that includes the area/s hit by the wildfire. It is composed of two tuples of coordinates $\langle \textit{Longitude}, \textit{Latitude} \rangle$, which indicate the top-left and bottom-right edges of the region. The reference date is the post-wildfire date used as a reference for the analysis by the domain experts.

Sentinel-2 L2A acquisitions were downloaded from SentinelHub [140], a web service that makes Earth observation imagery easily accessible through Application Programming Interfaces (APIs). Downloaded data refers to the same AoI and reference date specified in the Copernicus EMS grading maps (delineation maps can be obtained from grading maps considering just the wildfire extent).

3.1.1 Data description and constraints

Sometimes Sentinel-2 L2A data may not be available for the specified AoI and the reference date. Commonly, the reason can be twofold: (i) AoI was not explored or partially explored by satellites in the reference date, or (ii) the AoI is mostly covered with clouds. Therefore, Sentinel-2 acquisitions were subjected to three constraints: (i) the satellite acquisition must be equal to the reference date, (ii) data must be available for at least the 90% of the AoI, and (iii) cloud coverage must not exceed the 10% of the AoI. While the data availability is given by the Sentinel-Hub service, the cloud coverage value was estimated according to the method proposed by Braaten, Cohen and Yang [13].

Overall, 21 Copernicus EMS grading maps have been collected (and associated with each suitable Sentinel-2 acquisition) from 5 European regions: Portugal, Spain, France, Italy, and Sweden. Sentinel-2 data have been downloaded at the highest resolution available. Then, all the AoIs were split into 7 folds, according to two different constraints: (i) a fold must include at least two AoIs, and (ii) areas of interest must be geographically close. A representation of the geographical distribution of the wildfire-affected regions and their categorization in folds is shown in Figure 3.1. Sentinel-2 data are images having dimensions ($W \times H \times D$). W and H are the acquisition Width and Height, and they are up to 5000×5000 pixels. D , the Depth, is the number of spectral bands, which is 12 for Sentinel-2 L2A imagery. Copernicus EMS grading maps are images of size $W \times H$, having the same dimensions as the satellite acquisitions. Each pixel value of the Copernicus annotation ranges between 0 and 255, and determines: (i) unburned locations, with the value equal to 0, or (ii) burned locations, with the value greater than 0. Damage intensities are expressed with values greater than zero, that have been rescaled to 1 for "Negligible to slight damage", 2 for "Moderately Damaged", 3 for "Highly Damaged", and 4 for "Completely Destroyed".

Details about the collection, especially the Copernicus EMS annotations, the dates in which Sentinel-2 data were acquired, and the fold they were assigned to are reported in Appendix A.1. Moreover, in order to allow the proposed methods to generalize among different kinds of vegetation, another aspect considered in the dataset is the heterogeneity of land use, which includes inland areas with dense vegetation (i.e. red fold), areas characterized by cropland and small or sparse trees (i.e. fuchsia fold), coastal areas (i.e. blue fold) and rural areas with little or no vegetation (i.e. yellow fold). A detail of the land use distribution for every AoI



Figure 3.1: Map of the areas affected by wildfires in this study, divided by fold. The position of the wildfires considered in this study is determined by circles, while the color of each circle defines a specific fold: circles of the same color identify areas of interest assigned to the same fold.

included in the dataset is reported in the Appendix [A.2](#).

3.1.2 Data analysis

Data analysis has been performed to assess the data informativeness for the tasks of delineation and damage severity estimation.

Delineation of burned areas

First, the Separability Index (SI), reported in Equation [2.5](#), has been computed for all the spectral bands and the spectral indices used in the literature: BAIS2, MIRBI, NBR, NBR2. SI evaluates the difference between the statistical distributions of pixels belonging to burned areas and the ones belonging to unburned areas, in the domain \mathbb{R}^+ . Generally, values of SI higher than one indicate good separability, while values lower than one represent a large degree of overlap between burned

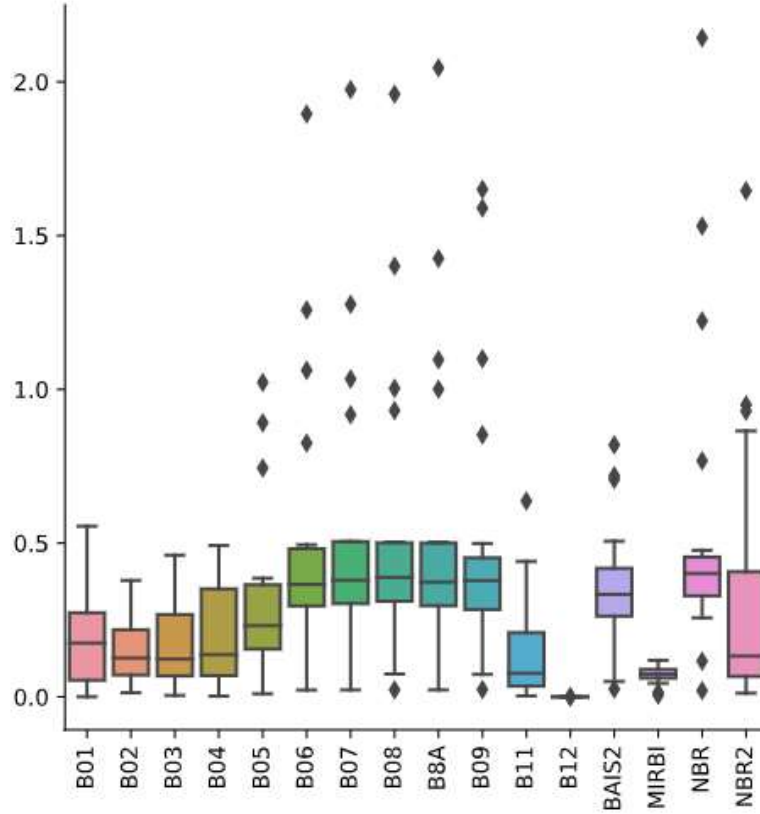


Figure 3.2: Separability Index computed on Sentinel-2 L2A spectral bands (B01-B12) and on spectral indices used for burned areas delineation.

and unburned distributions [162, 133].

As shown in Figure 3.2, both the spectral bands and indices present low SI values. This indicates that the dataset presents ambiguous regions: considering the bands and indices alone, some unburned regions are very similar to the burned ones and vice versa. The best indicators are evaluated considering both the SI in the whole distribution and the median value (expressed by the 50th percentile). Considering spectral bands, the best ones are B06, B07, B08, and B09. Except for B09, the others are used in the computation of BAIS2 (B06, B07, B8A) and NBR (B08), which means that they carry most of the discriminative information alone. The best spectral indices resulted to be BAIS2, NBR, and NBR2. NBR2 is among the worst indicators if considering the distribution below the 25th percentile, and presents the highest SI value considering its 75th percentile. This behaviour is explained by the literature: the index was created to be sensitive to vegetation typical of savannah areas (i.e. small shrubs), while it is less effective with other vegetation types. A positive fact is that NBR2 is computed using B11 and B12, that alone did not provide significant information. BAIS2 presents high SI values

over the 25th percentile of its distribution, but suffer from the same lack of informativeness of NBR2 for the values below. In the end, NBR, which is computed using Bands 8 and 12, presents the highest SI values considering its median value and the whole distribution range. The other bands present lower separability with respect to burned areas, but their informativeness has not to be underestimated, because they are employed in literature for the delineation of other land regions, such as coastal areas (B01), vegetation (B04), clouds (B03, B11, B12), and ground. Therefore, they can be employed to improve the disambiguation capability between unburned and burned regions.

Estimating damage severity

The Pearson’s correlation between the spectral bands of the gathered AoIs, the dNBR, and the Ground Truth (GT) (designated as the Copernicus EMS damage level) was explored in order to determine the spectral bands that can provide useful information for the detection of damage severity. The dNBR index was considered because it is the main reference used for producing grading maps. The dNBR requires pre-wildfire acquisitions to be computed: therefore, only for its computation, the same pre-wildfire images used in the grading maps were downloaded.

The Pearson’s correlation is a measure computed between two variables and it ranges between -1 and 1 . High correlation is determined at the extremes of its domain: a highly positive or a highly negative value means the two variables are directly or indirectly related, respectively. More in detail, no correlation is expressed by 0 , it is low for values between -0.35 and 0.35 , and medium to strong for the remaining values [84]. In order to compute the correlation coefficient, a transformation was applied to each image and annotation. Each spectral band, the dNBR, and the GT, all having dimensions $(W \times H)$, have been flattened into a vector of length $(W \times H)$, in order to resemble statistical variables. In the following, GT is used to refer to the target variable of this analysis, the damage severity level.

As shown in Figure 3.3, the spectral bands presenting noticeable correlations (medium or high) with both dNBR and the target variable GT are B06, B07, B08, B8A, and B09. They are the same bands that presented a high separability value in the delineation task, and therefore, they are the most sensitive in wildfire contexts. Moreover, it is worth considering the very high correlation between dNBR and GT, which is coherent with the literature and confirms the dNBR to be a good estimator of the GT. However, the spectral bands are characterized by high correlation values concerning the target variable and are produced by evaluating only the post-fire image, whereas the dNBR index requires two images (pre- and post-fire).

	B01	B02	B03	B04	B05	B06	B07	B08	B8A	B09	B11	B12	DNBR	GT
B01	1.000	0.876	0.833	0.786	0.777	0.576	0.498	0.456	0.447	0.512	0.688	0.723	0.001	-0.059
B02	0.876	1.000	0.972	0.936	0.886	0.679	0.598	0.591	0.546	0.538	0.779	0.795	-0.078	-0.124
B03	0.833	0.972	1.000	0.978	0.950	0.783	0.708	0.702	0.659	0.637	0.853	0.831	-0.170	-0.232
B04	0.786	0.936	0.978	1.000	0.961	0.761	0.685	0.678	0.640	0.614	0.897	0.882	-0.144	-0.211
B05	0.777	0.886	0.950	0.961	1.000	0.868	0.803	0.771	0.765	0.734	0.927	0.860	-0.263	-0.335
B06	0.576	0.679	0.783	0.761	0.868	1.000	0.988	0.962	0.975	0.926	0.772	0.589	-0.558	-0.606
B07	0.498	0.598	0.708	0.685	0.803	0.988	1.000	0.974	0.992	0.940	0.711	0.505	-0.611	-0.650
B08	0.456	0.591	0.702	0.678	0.771	0.962	0.974	1.000	0.974	0.918	0.687	0.476	-0.628	-0.651
B8A	0.447	0.546	0.659	0.640	0.765	0.975	0.992	0.974	1.000	0.946	0.688	0.463	-0.648	-0.679
B09	0.512	0.538	0.637	0.614	0.734	0.926	0.940	0.918	0.946	1.000	0.670	0.448	-0.639	-0.686
B11	0.688	0.779	0.853	0.897	0.927	0.772	0.711	0.687	0.688	0.670	1.000	0.937	-0.190	-0.286
B12	0.723	0.795	0.831	0.882	0.860	0.589	0.505	0.476	0.463	0.448	0.937	1.000	0.084	-0.027
DNBR	0.001	-0.078	-0.170	-0.144	-0.263	-0.558	-0.611	-0.628	-0.648	-0.639	-0.190	0.084	1.000	0.853
GT	-0.059	-0.124	-0.232	-0.211	-0.335	-0.606	-0.650	-0.651	-0.679	-0.686	-0.286	-0.027	0.853	1.000

Figure 3.3: Correlation Matrix computed on burned regions included in the dataset. The spectral bands (B01–B12) refer to the post-fire Sentinel-2 acquisition, dNBR is the Delta Normalized Burn Ratio, and Ground Truth (GT), is the damage severity level used as target variable (i.e., the Copernicus EMS grading map).

3.2 Unsupervised delineation on visible light with pre- and post-wildfire data

In preliminary work, we proposed Burned Area Estimator (BAE), a novel unsupervised approach to delineate burned areas using information from visible light in pre- and post-wildfire acquisitions [45].

Within BAE, we wanted to provide a location-independent technique that could be applied without any training. We chose to limit the available information to the set of the visible spectral bands to approximate the feasibility of the task with different equipment, such as normal cameras mounted on airplanes, which can improve the frequency and the coverage of the monitoring.

3.2.1 Problem statement

Consider $I_b, I_a \in \mathbb{R}^{w \times h \times n}$, two Sentinel-2 acquisitions of the same area of interest, taken before the beginning of a wildfire event (I_b) and after its extinction (I_a), in which: (i) w represents the acquisition width, (ii) h represents the acquisition height, and (iii) n is the number of considered spectral bands. In this problem, only the spectral bands related to the visible light B04, B03, and B02 are considered, therefore $n = 3$. The goal is to predict the binary mask $I_m \in \{0,1\}^{w \times h}$, in which pixel values set to 1 refer to burned regions, 0 otherwise. Therefore, the problem

is configured as a binary segmentation task, also known as delineation task in the geospatial context.

3.2.2 Methodology

The BAE algorithm, represented in Figure 3.4, works as follows. First, the two acquisitions are preprocessed separately by the Normalization & HSV Preprocessing module, which performs a Z-Score Normalization and a lossless conversion from RGB to Hue Saturation lightness Value model (HSV). Then it applies a transformation keeping the same H and setting both S and V to a constant value. We chose to select their maximum value (S_{max} , V_{max}) because this allows increasing the distance between color values in the RGB domain and hence the values can be clustered more easily in the following steps. This step is key to make the color component comparable between the two images while removing the differences that can result from images taken at different conditions (e.g. time of the day). The HSV Preprocessing module outputs the isolated H component, which is sent to the Hue Difference Segmentation Strategy module, and the (H, S_{max}, V_{max}) , which is the input of the Color-based segmentation strategy module. Both strategies are detailed in the next subsections.

Hue Difference Segmentation Strategy (HDSS)

This strategy is based on the assumption that, in an area affected by a wildfire, the greatest changes in terms of pure colors (H) between the pre- and the post-wildfire images are due to the burned areas. However, not only wildfires produce significant changes of hue during a short period, but at this stage, this strategy aims to detect every area that has been subjected to a modification during the two times in which acquisitions were taken. Hence, the *Windowed “H” difference module* computes the difference between the “H” components of the two images.

To reduce errors due to objects e.g., metallic surfaces, that change color when exposed to different kinds of sunlight, we consider a 5x5 matrix of pixels in the pre-wildfire acquisition and compute the minimum pixel-wise difference with the pixel corresponding to the center of the matrix in the after image.

Let $HA_{i,j}$ be the “H” component of a pixel in position (i, j) in the pre-processed post-wildfire acquisition. Then, consider $W_{i,j}$ as a squared Window of odd size (w, w) centered in position (i, j) in the pre-processed pre-wildfire acquisition. This module generates a Hue Difference matrix HD having the same size of the input images, in which each pixel $HD_{i,j}$ represents the minimum distance between $HA_{i,j}$ and the pixel values p in $W_{i,j}$, which is computed as follows:

$$HD_{i,j} = \min_{p \in W_{i,j}} \text{angdist}(HA_{i,j}, p), \quad (3.1)$$

where $\text{angdist}(x, y)$ is the angular distance between x and y . The angular distance is necessary because the “H” component is expressed in degrees, from 0 to 360. In HD,

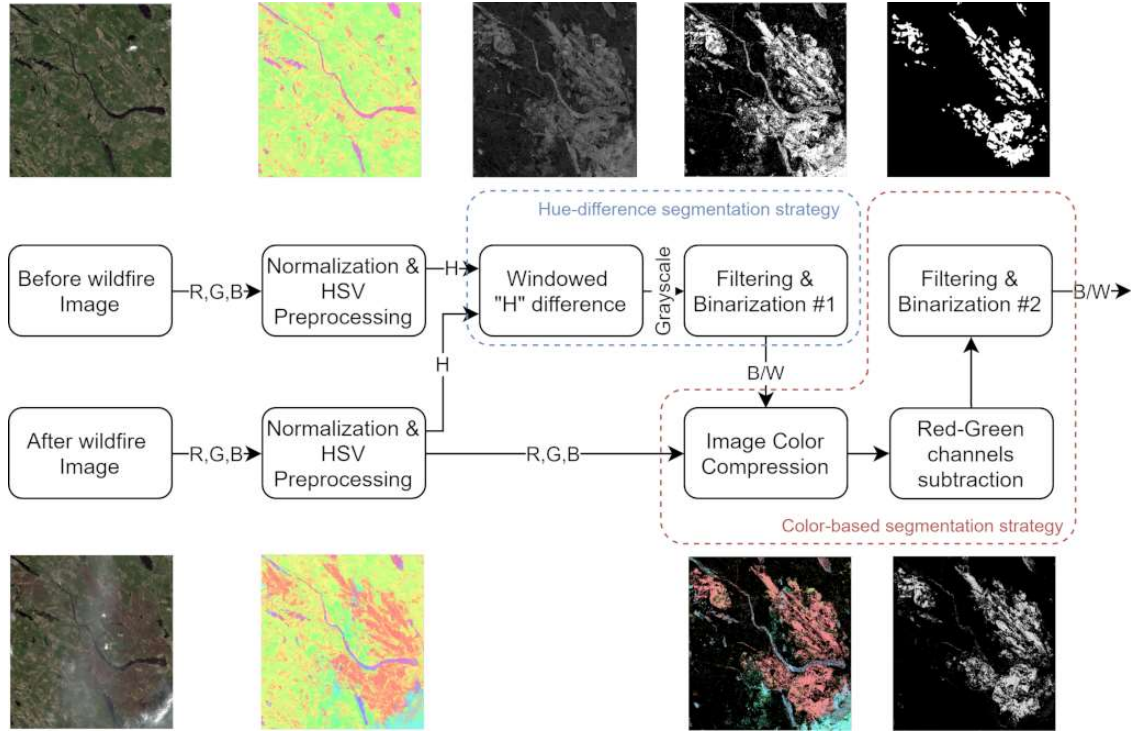


Figure 3.4: Architecture of the BAE's approach for segmenting burned regions. The rectangular boxes indicate the main steps of the algorithm, while in the arrows the input/output types are presented. Large boxes with dotted borders enclose the two different segmentation strategies.

0 values mean that there is no difference in that location with respect to the before-event situation, while positive values indicate the variation magnitude. Finally, the Filtering and Binarization #1 module applies a Gaussian Filter of dimensions (5,5), smoothing differences and facilitating the computation for automatic thresholding to binarize the image, performed by means of the standard Otsu's algorithm [115].

Color-based Segmentation Strategy (CSS)

After the application of the HDSS strategy, the CSS is performed. The CSS is based on the assumption that burned areas in the same image are characterized by similar colors. The first step of this strategy, named *Image Color Compression*, works solely on the regions identified by the white color of the binary mask generated in the previous step. Accordingly to that mask, this module selects the colors of the HSV processed after the wildfire image to reduce the color space and cluster similar areas. Then, the second step, named *Red-Green Channels subtraction*, isolates the burned regions performing a subtraction between the Red and Green channels of the image. Finally, a second filtering and binarization step fine-tune

the segmentation. The mentioned steps are detailed in the following.

After converting the HSV post-wildfire image back to RGB, which we name Pre-processed Image (*PI*), the *Image Color Compression module* selects only the regions affected to a significant change, by selecting the colors in the regions that resulted as white in the binary image returned by the HDSS strategy. Then, it reduces the number of colors to “force” similar regions to be represented by the same RGB triple, i.e., we aim to “cluster” similar regions associating them to the same RGB triple.

To accomplish this task, we adopt a Self-Organizing Map (SOM) [65], which is an unsupervised Artificial Neural Network (ANN) that maps the input image while preserving its neighbourhood relations. The SOM can be represented as a lattice of dimensions (w, h) , composed of $w \times h$ neurons. Each neuron n shares the same dimensions of the pixels in the input image: in our case, it is defined as follows: $n \in \mathbb{R}^3$. Firstly, the neurons are initialized in the multidimensional space. Then, SOM neurons are iteratively updated in order to resemble the distribution of the input data. At each step, neurons are updated according to two parameters: (i) the learning rate $\eta \in \mathbb{R}^+$, that determines the module of update and, (ii) the neighbourhood function $f(n)$, that modulates the learning rate for each neuron. One of the most popular neighbourhood functions corresponds to the probability density function of the normal distribution $\mathcal{N}(\mu, \sigma^2)$, in which (i) μ is determined according to the Best Matching Unit (BMU), the neuron that minimizes its average distance with the pixels in the input image (see Figure 3.5), and (ii) σ^2 is arbitrarily chosen. Therefore, the closer the neurons to the BMU are, the higher their weight update will be.

As explained before, the *HSV Preprocessing module* increases the distance of the pixels in the RGB space, facilitating the SOM training process in producing more representative neurons (see Figure 3.6).

The *Image Color Compression module* normalizes the input image by using the *min-max normalization*, which maps the RGB components from the range $[0, 255] \in \mathbb{N}$ to $[0, 1] \in \mathbb{R}$ and feed that to the SOM, which should be carefully sized and initialized to be effective. We empirically set the network size to $(3, 3)$, while we uniformly initialize the network weights in normalized RGB space. The network is adaptively trained for each image with an increasing number of epochs, until convergence. We show in Figure 3.6 the RGB representation of the after-wildfire image before and after the HSV Preprocessing, as well as the initialized SOM neurons. We select the Euclidean distance to evaluate the distance between pixel and neuron values, in order to determine the BMU during the SOM training. We define TN as the set of Trained Neurons of the SOM and CC as the Color Compressed image that the SOM outputs. Every CC pixel $CC_{i,j}$ is assigned to the color of the neuron $n \in TN$ that minimizes the Euclidean distance from the

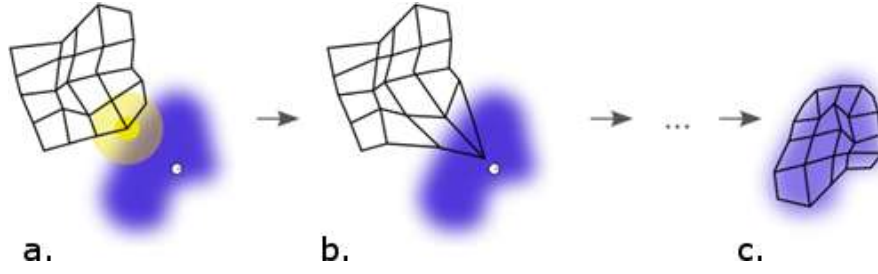


Figure 3.5: Simplified illustration of the Self-Organizing Map training phase. The dataset is represented by the violet area. The SOM size (5, 5) is represented by the black mesh, having a neuron at each intersection. The phases are described as follows: (a) The BMU is identified: it is depicted as the yellow-circled neuron, whose radius indicates the neighbourhood function that affects the weight updates of the other neurons; (b) result of the weights update; (c) SOM's neurons displacement after the end of the training. Illustration is from Wikipedia [135].

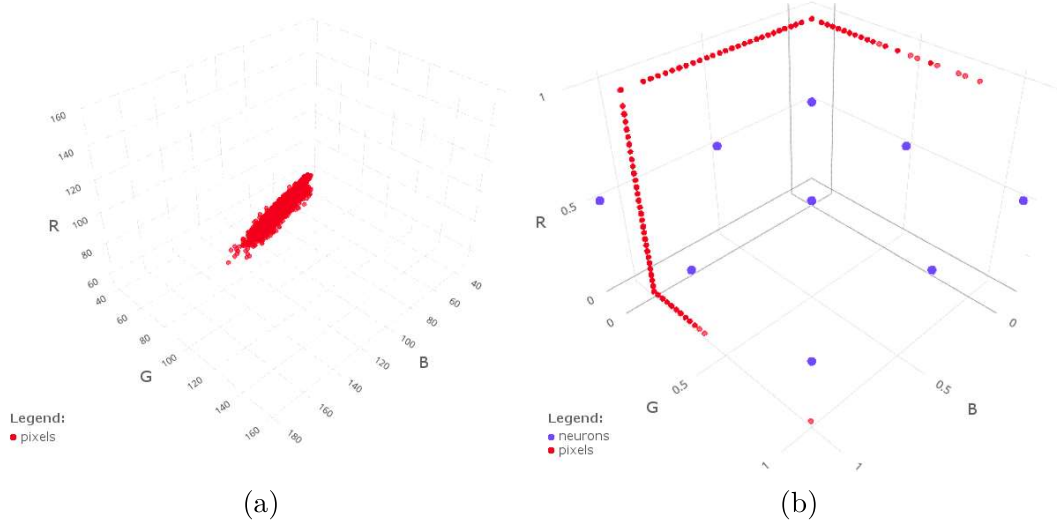


Figure 3.6: Illustration of post-wildfire acquisition in the RGB space. (a) pixel values in the raw acquisition (red dots), (b) pixel values in the preprocessed acquisition (red dots) and initialized SOM neurons (blue dots).

corresponding $PI_{i,j}$ pixel:

$$CC_{i,j} = \operatorname{argmin}_{n \in TN} \|PI_{i,j} - n\|_2 \quad (3.2)$$

Therefore, in the white-colored regions of the HDSS output, CC is an RGB image having a reduced number of distinct colors equal to the number of the network neurons. While, in the remaining regions, it is colored in black.

The step performed by the Image Color Compression module made similar colors closer to each other and, at the same time, it increased the distance concerning

different colors.

At this point, a module that allows highlighting common characteristics of burned regions is needed. In an unsupervised manner, this implies that no previous knowledge about the data can be exploited, but just a generic intuition is allowed. The idea behind the *Red-Green Channels subtraction* module is that considering the hue of burned regions, they are prominent to red/violet colors and, at the same time, they present a near-to-zero level of green. Therefore, that module subtracts the green component from the red one. We do not consider the blue channel because, even if it is highly relevant in blue or light-blue areas like rivers or lakes, it is also relevant in violet regions, which can characterize burned areas. Finally, the *Filtering & Binarization #2* module is equivalent to the one adopted in the HDSS, with the addition at the end of a median filter, which removes possible noise generated by the binarization phase.

3.3 Delineation assessment with Convolutional Neural Networks using post-wildfire data

In a second work, published at the International Conference on Information Systems for Crisis Response and Management, we assessed the performance of supervised approaches, trained on historical post-wildfire training data, using information from (i) visible light and (ii) all the available spectrum in post-wildfire acquisitions [44]. The goal was to assess their performance without the need to acquire pre-wildfire or any other extra data. The work has been acknowledged as a runner up for the Best Student Paper Nominee in 2020.

3.3.1 Problem statement

The problem involves a post-wildfire Sentinel-2 acquisition and it is split into two tasks, which consider different spectral bands. The first task considers only spectral bands related to visible light, namely B04, B03, and B02, while the other one considers the whole spectrum. The goal is the same as the previous section: producing a binary mask, whose pixels assume values equal to 1 to indicate burned regions, 0 otherwise. To accomplish the task, a set of annotated data, consisting of Sentinel-2 acquisitions and binary masks, is available for supervised approaches.

3.3.2 Methodology

From a geometrical point of view, burned areas resemble spots: circumscribed shapes presenting irregular borders, sometimes presenting branches or protruding parts. With some abstraction, this rough description can be applied to biological

cells, as shown in Figure 3.7. Those similarities drove our search for promising approaches to U-Net [127] and CuMedVision [22]: two popular Convolutional Neural Networks (CNNs) used in the medical field for cells segmentation.

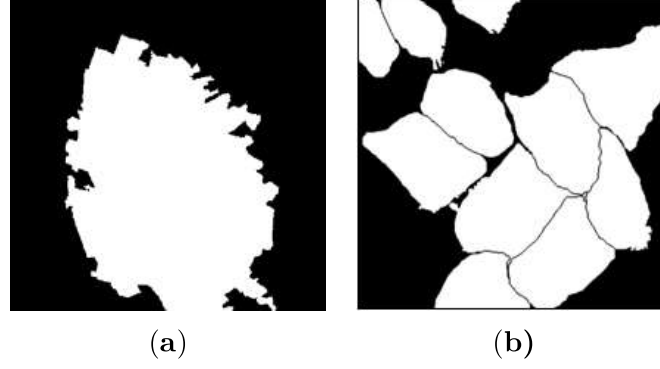


Figure 3.7: Geometrical similarities between different ground truths of (a) burned regions, (b) biological cells, the picture is from the work of O. Ronneberger et al. [127].

U-Net

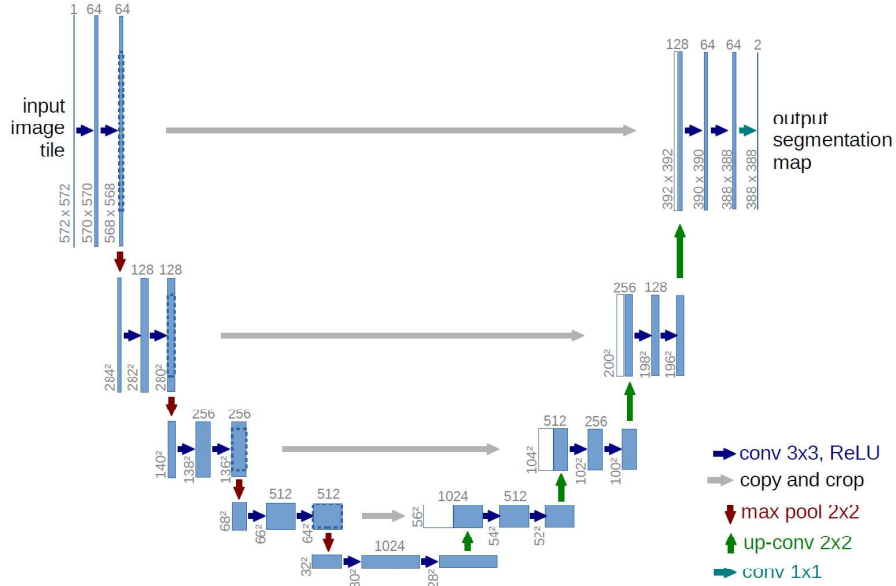


Figure 3.8: U-Net architecture. The picture is from the work of O. Ronneberger et al. [127].

The U-Net architecture, depicted in Figure 3.8, is composed of two sequential parts of convolutional and pooling layers, which gives the U-shaped form: the contracting and the expanding path. Like a generic CNN architecture, the contracting path interleaves convolutions and max-pooling layers, which gradually reduce both the width and height of the image, while increasing its depth, enabling it to focus on a larger receptive field. Convolutional operations, together with downsamplings of the max-pooling layers, let this path focus on the features that best describe the subject to be detected. However, at the end of the contracting path, the spatial information about the subject to be detected is lost: on purpose, the goal of the expanding path is to restore that information, transforming the original input into a binary segmentation. The expanding path increases the dimensions of the feature vector through up-convolutions. To enhance the precision of the spatial information reconstruction, at every step of the expansion path, the output of the upsampling operation is concatenated, through skip-connections, with the feature maps from the contracting path at the same level.

In the original architecture, U-Net splits the input image in tiles of size 388×388 pixels, but it takes in input an extra area, used to give context to the network during the inference. Therefore, its input is of size 572×572 pixels, while the segmented region is the central part, of 388×388 pixels.

In our work, we adopted the original structure in terms of the number of convolutional layers and operations, but we introduced some modifications to simplify its applicability.

Variations are related to (i) the net input and output dimensions, and (ii) the loss function. The net input depth is adapted to the number of considered spectral bands of the acquired data, which is 3 in the case of visible wavelengths and 12 in the case of the whole spectrum. Also, width and height from input and output tiles are set to 480×480 pixels, the same dimension adopted for the other methods compared in this study. In order to keep the same input and output dimensions. To keep the same net input and output dimensions, the convolutional layers were adapted as follows: (i) the input to each convolutional layer was padded by mirroring the layer input itself, avoiding the reduction of the layer output dimensions, which is naturally induced by the convolution, and (ii) each max pooling operation (of size 2×2) in the contracting path was performed with stride 2, with the effect of halving both the width and the height of the next layer. In the expanding path, the up-convolution operation doubles both the width and the height of the next layer, restoring the initial dimensions in the net output.

Concerning the loss function, the original U-Net used a Cross-Entropy loss combined with a pixel-wise weight map, which depends on the nearest cell and penalizes more errors made in contour pixels. In the biological context, cells used to be very close to each other. Instead, in our context, there is either one burned region per acquisition, or few ones presenting considerable distance. Computing weight maps in our context would lead most of the values near zero, thus making neural networks

training more error-prone.

CuMedVision

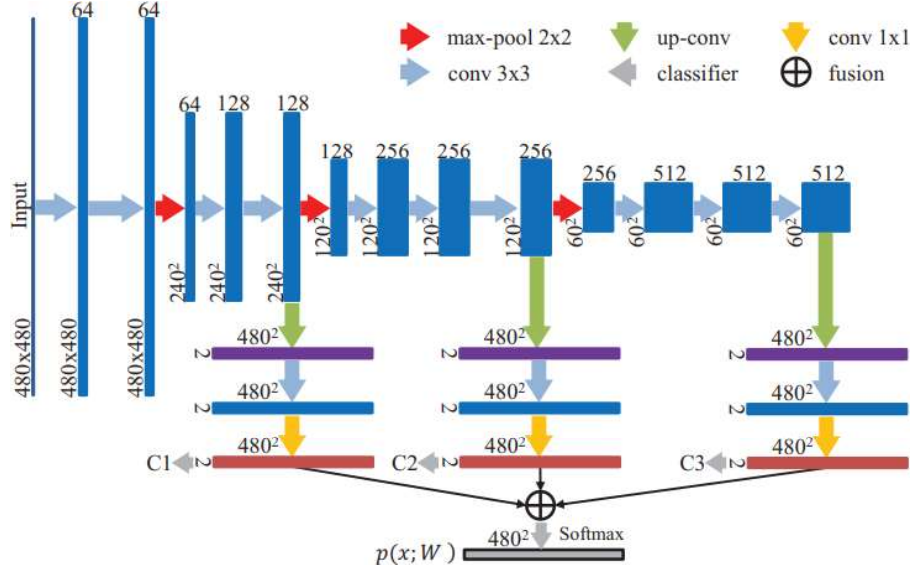


Figure 3.9: CuMedVision architecture. The picture is from the work of H. Chen et al. [22].

CuMedVision architecture, depicted in Figure 3.9, achieved the best scores in the ISBI dataset, proposed for neural cells segmentation challenge, overcoming the results obtained by the original U-Net. It is made of two main components: a contracting path and three parallel expanding paths. Like U-Net, the contracting path aims to extract features useful to identify the subjects for the segmentation, reducing the resolution of the input images. Then, three expanding paths aim to reconstruct the spatial information, leveraging on different convolutional layers of the contracting path. Finally, the three paths outputs are summed up and modulated by means of the Softmax activation function.

In our work, the net input depth is adapted to the dimension of the acquired data (as explained for the U-Net model) to (i) 3, when considering only visible light information, and (ii) 12, when considering the whole spectral data.

3.4 Damage severity estimation using post-wildfire data

The third and final work on burned areas from satellite imagery concern the damage severity estimation task. As seen in Chapter 2, this task is usually performed through in-situ inspections. However, approximations computed from pre- and post-wildfire satellite acquisitions are officially accepted. In our work, published in MDPI Applied Sciences journal, we present a novel supervised approach that aims to provide accurate estimates of damage severity, only leveraging on post-wildfire acquisitions [43]. Moreover, the approach aims to be land-type-independent and, therefore, to be applicable to any European region.

3.4.1 Problem Statement

Considering all the spectral bands of a post-wildfire Sentinel L2A acquisition, the goal is to predict a matrix, whose elements can assume continuous values in the range $[0,4] \in \mathbb{R}$. The matrix is an approximation of the Copernicus EMS grading map values, whose severity levels are natural numbers within the same range. More precisely, severity levels are the following: 0 means "Unburned area - no damage", 1 is associated with "Negligible to slight damage", 2 corresponds to "Moderately damaged", 3 is "Highly damaged", and 4 stands for "Completely Destroyed". The problem is configured as a bidimensional regression task because the target variable is a numerical feature used to represent ordered severity values. To accomplish the task, a set of annotated data, consisting of Sentinel-2 acquisitions and grayscale masks, is available for supervised approaches.

3.4.2 Methodology

The main contribution of this study is the modification of the original U-Net, empowering its ability to distinguish between ordered classes, as in the case of damage severity. The U-Net adopted in Section 3.3 was proposed for solving a segmentation task, being able to identify a specified entity in an image. Therefore, it is able to recognize relations and features (i.e. borders and gradients) among the pixel values belonging to the searched entity. Intuitively, in the context of this work, the problem can be configured to be solved at once as a regression task. Conversely, we considered the problem as composed of two sub-tasks: (i) to identify areas affected by fire, and (ii) to determine damage severity in the burned areas. In the first sub-task, the goal is to distinguish burned areas from unburned regions, like in a classical segmentation task (bidimensional classification task). The second sub-task takes into account the areas affected by the fire and discriminates between four consequent levels the severity of the damage (bidimensional regression task).

As we will see later on in this section, splitting the problem into subsequent subtasks demonstrated to be more effective.

Models' architecture

The two sub-tasks are solved with two different building blocks: the “Binary Classification U-Net” (BCU) and the “Regression U-Net” (RU). Both BCU and RU consider the adjustments made in Section 3.3 for the original U-Net, with small variations: (i) BCU uses the Dice loss instead of the Binary cross-entropy loss, and (ii) RU, being a regression task, uses the Mean Squared Error as loss function and avoids the softmax activation function in the last layer.

The output map in both building blocks is a 480×480 matrix, with each element referring to the pixel in the same position of the input tile. In the proposed solution, BCU and RU are combined together to outperform the prediction quality of the approach based solely on a single Regression U-Net block.

Firstly, the Binary Classification U-Net is trained for segmentation purposes: given a tile having dimensions $480 \times 480 \times 12$ of a post-wildfire Sentinel-2 L2A acquisition, the network assigns to each pixel the probability of belonging to a burned area, thanks to the application of the softmax activation function. Thus, the generated output is a binary mask of size 480×480 with values $\{0, 1\}$ (i.e., unburned or burned), where each pixel is assigned to the class with the highest probability.

Second, the Regression U-Net is used to provide the severity level estimation. Given the input tile, the model generates a map of the same size with values in the range $[0, 4]$.

By combining BCU and RU differently, we have considered three different approaches:

- Single U-Net, a regression-only approach in which just the Regression U-Net is used;
- Parallel U-Net, a parallel approach, in which the two building blocks are employed separately in parallel. As shown in Figure 3.10, the final output is obtained by multiplying the two outputs pixel-wise;
- Double-Step U-Net, a two-step approach in which the building blocks are concatenated. The burned regions in the input tile are first predicted using the Binary Classification U-Net. Its output, the binary mask, is used to limit the information passed to the Regression U-Net, filtering out the pixel values predicted as belonging to unburned regions. Then, the Regression U-Net provides the damage severity estimation. Figure 3.11 shows the simplified architecture of the proposed solution with sample images, showing only RGB channels for simplicity.

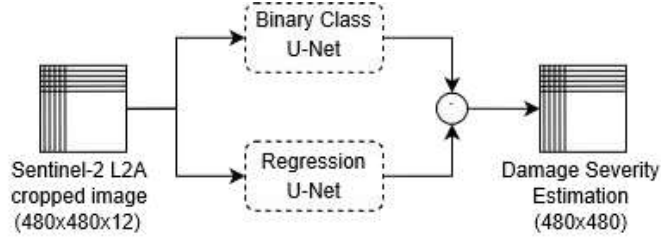


Figure 3.10: Simplified Parallel U-Net diagram. The burned/unburned binary mask, the output of the Binary Classification U-Net, is multiplied pixel-wise with the Regression U-Net output. This operation filters out the unburned regions from the grading mask produced by the Regression U-Net. The final output is the estimate of the damage severity in the area of interest.

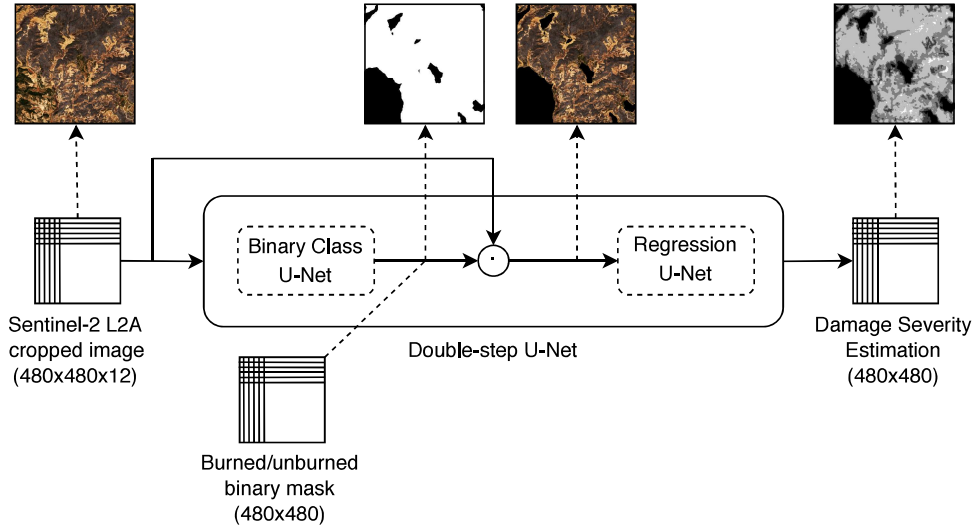


Figure 3.11: Simplified Double-Step U-Net diagram. The damage severity estimation is computed in two steps: (i) burned area delineation through the Binary Classification U-Net, and (ii) damage severity estimation by means of Regression U-Net. The Regression U-Net receives as input the Sentinel-2 L2A tile filtered with the binary segmentation mask.

The idea behind those architectures is hidden in the problem of Damage Severity estimation. Even if similar, we think that the two subtasks diverge when considering the unburned area. When trained, the BCU is able to segment the burned regions, therefore its convolutional layers will be able to recognise characteristics, such as geometries and values, of those regions. Consequently, it is able to do the same for the unburned ones. Differently, in the second subtask, the network must focus on identifying dissimilarities in the same class: the burned region. Therefore:

- with the Single U-Net approach, we want to set a baseline to the whole

problem;

- with the Parallel U-Net we want to prove that the whole problem is too difficult to be solved at once. In particular, we want to compare the performance obtained on the severity level 0 (unburned region), which is considered as a normal severity level by the Single U-Net approach, while it is considered as a different class by the BCU in the Parallel U-Net;
- with the Double-Step U-Net we want to prove that the RU benefits from both: (i) the knowledge on burned areas acquired by the BCU, and (ii) the masking operation on the unburned areas, which isolates the predicted damage severity levels and allows the RU to focus on the second subtask.

A detailed version of the Double-step U-Net architecture is shown in Appendix [A.3](#).

3.5 Experiments

This section presents the experiments for the three problems previously introduced. First, the raw satellite data illustrated in Section [3.1](#) are preprocessed to prepare the dataset. Then, the testing process and the evaluation metrics are presented. For what concerns supervised approaches, the initialization and regularization techniques, as well as the hyperparameters used during the experiments, are detailed. Finally, the results are discussed.

3.5.1 Dataset preparation

As introduced in Section [3.1](#), the satellite acquisitions were split into seven folds. The grouping criteria were determined based on the geographical distance between the acquisitions' locations, in order to include geographically adjacent regions with similar morphology and land cover characteristics, such as vegetation types, infrastructure, and agricultural areas, in the same fold.

Generally, the high-resolution images retrieved from Sentinel-2 (and consequently, the grading maps) have dimensions up to 5000×5000 pixels. Currently, due to GPU memory limitations, this size is too big to be processed by any Deep learning model at once and it needs to be re-adapted. A first solution would be to shrink the data to fit the maximum input size supported, with a consequent loss of information. This option could be useful for fast detection approaches, but it is not suitable for our purposes.

In our work, we opted for preserving all the information by tiling the original post-fire acquisitions in smaller crops of size 480×480 pixels, maintaining the original spectral information. Furthermore, the dataset only includes crops with at

least one pixel classified as burned (a damage severity level between 1 and 4). In the end, the dataset contains a total of 135 crops, distributed in folds as follows; blue fold: 8, brown fold: 9, fuchsia fold: 30, green fold: 16, orange fold: 18, red fold: 12, yellow fold: 42. The dataset’s folds, as illustrated in Figure 3.12, have imbalanced damage severity levels, as easily predictable.

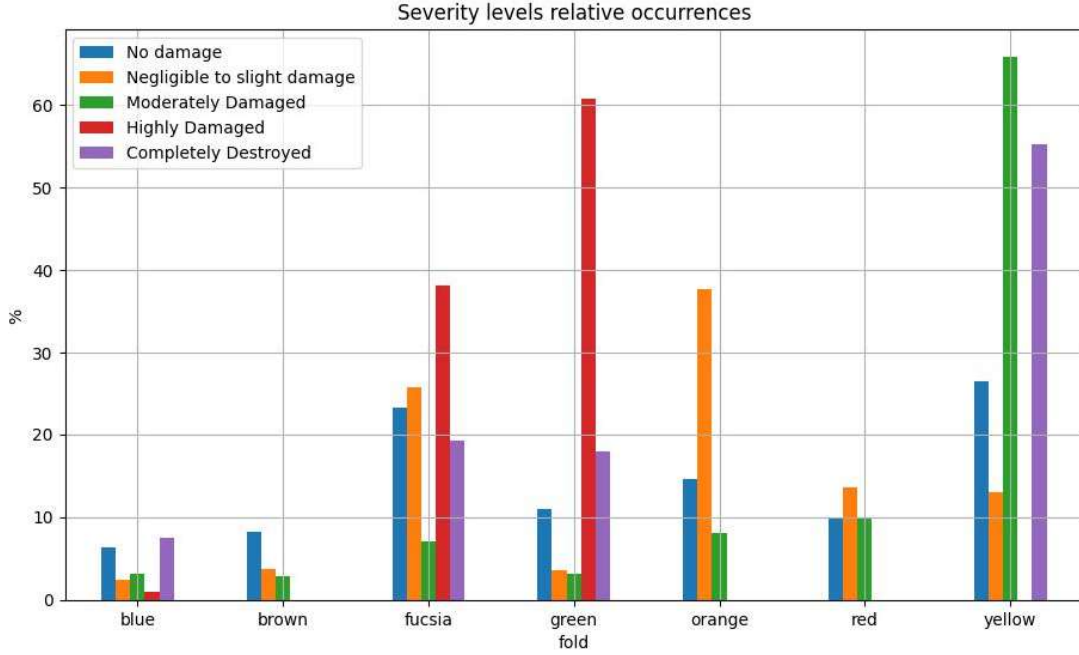


Figure 3.12: Severity level distribution for each fold. The percentages shown in the histograms are computed considering the whole dataset. Therefore, the percentage associated to each severity level has to be considered with respect to all the other folds.

3.5.2 Experiments and evaluation processes

The goal of the experiments is to evaluate the approaches in different areas, proving their ability to operate independently from morphological and geological aspects. Therefore, their performances will be measured for each fold. In supervised approaches, the models’ performances are evaluated through a cross-validation approach: for each iteration, five folds are used as the training set, one as the validation set, and the remaining fold as the test set. As explained later in this section, the early stopping criteria was used as a regularization approach, leveraging on the validation set to compute the model’s performance. A common prerequisite in supervised learning algorithms is that the training, validation and test data arise from the same distribution and are independent and identically distributed [110]. Therefore, a validation set should closely reflect the data distribution of the test

set. However, as illustrated in Figure 3.1, each fold presents a unique distribution of severity levels. In a real-world situation, there is no chance to know the distribution of burned areas and severity levels a priori. Therefore, the choice fell on a fold that contains all severity levels and which could generalise the most, presenting a distribution of severity levels that tends to a uniform distribution. Considering all those aspects, we chose the “fucsia” fold as the validation set for each test set, except for itself: in that case, we chose the “green” fold.

Performances Evaluation

The problem of *burned areas delineation*, configured as a binary segmentation task, is evaluated with the Precision, Recall, and F1-Score metrics [52]. Precision considers the purity of the predictions: among the pixels predicted as belonging to a certain class, e.g., belonging to a burned region, it indicates the percentage of matches with the GT. Recall verifies the ability of the estimator to recognize all the pixels belonging to a certain class, specified in the GT. Therefore, given the whole set of pixels belonging to a certain class (referring to the GT), the recall is the percentage of correctly predicted pixels among the whole set.

The F1-Score is the harmonic mean between Precision and Recall. It is as a measure of accuracy with the property to take into account the class imbalance.

The problem of *damage severity estimation*, concerning the distinction between 5 severity levels and configured as a regression task, is evaluated with the Root Mean Squared Error (RMSE) metric. Given the ordinal relationship between damage severity levels, the RMSE gives a measure of distance between the prediction and the ground truth.

3.5.3 CNNs hyperparameters tuning, regularization techniques, and training process

To increase CNNs models generalization, data augmentation was used on each fold of the training set. For each epoch of the training phase, the tiles were subjected to the following four transformations, used to create newer augmented tiles having the same dimensions as the original ones: random rotation, random horizontal flip, random vertical flip and random shear. Each transformation had the 50% of probability to be applied for each tile. When applied, random rotation and random shear randomly selected the transformation angles within specific ranges, as reported in Table 3.1.

Table 3.1: Data augmentation parameters.

Transformation	Probability	Parameters
Random rotation	50%	Angle: -50° , $+50^\circ$
Random horizontal flip	50%	-
Random vertical flip	50%	-
Random shear	50%	Angle: -20° , $+20^\circ$

Hyperparameters tuning

To ensure test reproducibility, for each training of the cross-validation process we initialized the CNNs with the same weights, generated with the same seed number, using a normal distribution and the Glorot initialization [60]. All the training were performed using Adam optimizer with a learning rate of 1×10^{-4} , 50 epochs and a batch size of 8.

Loss functions have been chosen according to the problem to be solved. For the delineation of burned area, U-Net (or BCU) and CuMedVision used the Dice Loss [148], which is equivalent to the F1-Score and therefore it benefits from the advantage of being robust in unbalanced datasets. For the damage severity estimation problem, the Regression U-Net used the Mean Squared Error (MSE) loss.

Regularization techniques

During the training process, three techniques for regularization were adopted: early stopping, dropout, and batch normalization.

Early stopping was implemented to avoid overfitting and to stop the training process in case no further improvements were seen in the validation loss. A patience of 5 epochs was used with minimum improvements of 1×10^{-2} on the validation loss. At the end of each training process, the model’s best weights determined by the early stopping mechanism were restored.

Dropout layers were enabled during the training process before each transposed convolution with a probability of 25%. Moreover, after each convolutional layer, batch normalization was performed.

Training process

For a single CNN, the training process starts with the Glorot initialization and continues with the weights update epoch by epoch, until either the maximum epoch is reached or the early stopping criteria is matched.

For the Double-step U-Net, the process is composed of three phases. First, the Double-Step U-Net weights are set according to the Glorot initialization. Then, the network is trained on the binary segmentation problem: during the epochs, the

weights of the Binary Classification U-Net are updated, while the weights of the Regression U-Net are kept frozen. Once the BCU is trained, the Double-Step U-Net starts again the training process, this time to solve the damage severity estimation problem: therefore, the BCU weights are kept frozen, while the Regression U-Net weights are updated.

3.5.4 Results on delineation problems

In this section we discuss the results of the experiments performed for burned areas delineation problems introduced in Sections 3.2 and 3.3.

Performance evaluation. Tests using visible light compare BAE, the unsupervised approach data which leverages on pre- and post-wildfire acquisitions, with U-Net and CuMedVision, the supervised approaches used for the assessment that consider only post-wildfire acquisitions.

According to the results shown in Table 3.2, BAE tends to be conservative, presenting reliable precision (> 0.72) in the majority of the folds (4/7), but lower recall. Therefore, it is quite accurate in finding burned regions, but it is not able to identify a good portion of the affected surfaces. Best F1-Scores were achieved in blue, brown, and green folds, in regions presenting green vegetation, such as forests or grasslands. Considering the whole dataset, its average F-Score is about 0.59.

Compared to BAE, CuMedVision and U-Net achieved, on average, higher performances in all the metrics. Even if they are more accurate, they tend to overestimate the burned region, being more prone to segment the whole affected area, presenting the highest recall performances, but lower precision. Both the approaches tend to misclassify regions presenting water sources or bare soil, like bare rocks or arable lands. Overall, U-Net demonstrated to be the best approach, presenting an average F1-Score of about 0.70.

Table 3.2: Burned areas delineation results using *visible light* data. BAE is an *unsupervised* approach and leverages on in *pre- and post-wildfire acquisitions*, while CuMedVision and U-Net are supervised approaches and only leverage on post-wildfire data. (\dagger) marks best Precision values, (\star) marks best Recall values, and **bold text** marks best F1-Score values.

Fold	BAE			CuMedVision			U-Net		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
blue	0.88 \dagger	0.64	0.74	0.33	0.99	0.49	0.34	0.99 \star	0.51
brown	0.61	0.63 \star	0.62	0.98 \dagger	0.15	0.22	0.44	0.39	0.41
fuchsia	0.49	0.41	0.45	0.89	0.67 \star	0.77	0.95 \dagger	0.54	0.69
green	0.75	0.61	0.67	0.86	0.93 \star	0.95	0.98 \dagger	0.89	0.93
orange	0.48	0.42	0.45	0.86 \dagger	0.45	0.59	0.74	0.61 \star	0.66
red	0.73	0.48	0.58	0.23	0.99 \star	0.37	0.80 \dagger	0.91	0.85
yellow	0.83 \dagger	0.44	0.57	0.82	0.84	0.83	0.80	0.97 \star	0.87
Avg.	0.68	0.52	0.58	0.71	0.72	0.60	0.72 \dagger	0.76 \star	0.70

Tests using all spectral data on post-wildfire events are shown in Table 3.3. According to the analyses performed in Section 3.1.2, we concluded that the NBR index was the most suitable for delineating burned areas. Therefore, we used that index as a baseline to compare the results with the other approaches. Normally, the NBR is manually thresholded by domain experts, or default thresholds are set, according to the environmental characteristics of the examined areas. In our case, we selected the thresholding value that performed best in each fold. In this way, we ensured that NBR performances obtained by traditional thresholding approaches would be always less or equal to the ones we obtained. Overall, NBR achieved accurate results with an average F1-Score of 0.79, while the lowest and highest F1-Scores were 0.63 and 0.87, respectively.

CuMedVision and U-Net significantly improved their performances in every fold, showing more stable results if compared to the tests on visible light. However, U-Net is confirmed to be the best model under both precision (4/7 folds) and recall (4/7 folds) metrics, and achieving the highest average F1-Score, equal to 0.86. Moreover, F1-Score below 0.82 is achieved only once, in the brown fold.

Table 3.3: Burned areas delineation results using *all spectral bands* data in *post-wildfire acquisitions*. (†) marks best Precision values, (★) marks best Recall values, and **bold text** marks best F1-Score values.

Fold	NBR (Best Threshold)			CuMedVision			U-Net		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
blue	0.55	0.98★	0.63	0.42	0.96	0.58	0.91†	0.95	0.93
brown	0.80†	0.94	0.85	0.79	0.83	0.81	0.45	0.98★	0.61
fuchsia	0.90	0.75	0.82	0.85	0.97	0.90	0.93†	0.98★	0.95
green	0.92	0.83	0.87	0.98	0.95★	0.96	0.99†	0.91	0.95
orange	0.80†	0.77	0.74	0.64	0.99	0.78	0.71	0.99★	0.82
red	0.78	0.83	0.81	0.73	0.98	0.84	0.84†	0.99★	0.91
yellow	0.75	0.92	0.80	0.94†	0.87	0.91	0.78	0.99★	0.87
Avg.	0.79	0.86	0.79	0.76	0.94	0.83	0.80†	0.97★	0.86

Computation time evaluation. Performances were also evaluated according to the complexity and the inference time of the assessed approaches, as shown in Table 3.4. Times were measured from the beginning of the inference process, to the time the delineation map of an acquisition tile of 480×480 px was returned (all the dataset was considered for this study). Performances were evaluated running the approaches both on CPU (Intel Core I9 7940x with 128 GB RAM) and on GPU (NVIDIA 1080 Ti). Tests on visible light data highlight that:

- BAE is the lightest model in terms of the number of parameters, but it is the slowest one (~2 seconds per tile). Being unsupervised, BAE needs to retrain the neurons’ weights for each tile;

- CuMedVision is lighter than U-Net, but both present a high number of parameters (> 20 Mln). Computation times are linearly proportional to the number of parameters. Execution times are lower than 0.72 seconds per tile on CPU, are about 15 times faster on GPU.

Considering all spectral bands, both CuMedVision and U-Net increase their number of parameters by $\sim 10\%$, which results in computation times lower than 0.8 seconds per tile on CPU and lower than 62 ms on GPU. NBR computation times are added for comparison, considering the computation of the spectral index and its binarization using a pre-defined threshold: in real contexts, that threshold is manually assessed for each tile.

Table 3.4: Inference times of the assessed methods for the delineation task, considering input tiles of dimension 480×480 px.

Bands	Method	# params	Inference time (ms)			
			Avg (CPU)	Std (CPU)	Avg (GPU)	Std (GPU)
RGB	BAE	< 100	1980	455	-	-
	CuMed.	21 Mln	516	20	41	0.2
	U-Net	28 Mln	719	27	45	0.3
ALL	NBR	-	2	3	-	-
	CuMed.	24 Mln	624	22	47	0.3
	U-Net	31 Mln	796	30	61	0.4

Overall considerations

BAE, which was designed to work without any training set and only with visible light data, proved to be precise in most of the test sets, resulting to be suitable for a preliminary delineation of burned areas.

CuMedVision and U-Net demonstrated to be valid approaches not only in the biomedical field but also in geospatial contexts. They proved to work on visible light data, and they demonstrated to be highly accurate when considering the whole spectrum. In both cases, the greatest advantage is that the information carried out by spectral bands is sufficient to determine reliable mappings, without the need for pre-wildfire acquisitions. Furthermore, CuMedVision and U-Net are suitable to provide near-real-time mappings, especially on GPU hardware. However, U-Net demonstrated to be the best model in both the tests cases, achieving accurate and reliable results in the whole dataset.

Qualitative considerations

Considering all the tests, there are qualitative aspects that can be noticed about the presented approaches, especially when considering particular regions, like (i)

coastal areas, (ii) forests areas, and (iii) arid regions.

Figure 3.13 shows an example of mappings in a coastal area. (a1) and (a2) are the satellite acquisition, visualised using visible light bands (B04, B03, B02), and the ground truth, the official delineation map, respectively. In the GT, white pixels describe burned regions, while black pixels describe unburned regions. (a3) Shows the BAE’s prediction: it is able to find the burned regions, but it underestimates them. False positives errors are made in arable land and sandy soils. (a4, a5) show CuMedVision’s and U-Net’s predictions using visible-light data. Both the approaches correctly identify the position of the burned areas. Segmentations are appropriate, even if they tend to overestimate the burned regions and contours are smooth, lacking details present in the GT. Both the approaches misclassified the water source. However, U-Net’s prediction is more appropriate than CuMedVision’s one. (a6, a7) show CuMedVision’s and U-Net’s predictions, using the whole spectral bands. Compared to the past predictions, both approaches improved the segmentation accuracy. However, CuMedVision still misclassifies water sources, while U-Net correctly identifies the burned regions and improves contour details.

Figure 3.14 shows an example of mappings in a forest region, with the presence of settlements. (b1) and (b2) show the satellite acquisition and the ground truth, respectively. (b3) Shows the BAE’s prediction: in this area, it is able to delineate the burned region with higher precision, even if it still tends to underestimate it. In this case, the presence of false positives is negligible. (b4, b5) show CuMedVision’s and U-Net’s predictions using visible-light data. Both approaches correctly identify the position of the burned areas. Also, in this case, segmentations are appropriate and they tend to overestimate the burned regions, presenting smooth contours which lack details in the GT. Both the approaches misclassified small portions of settlements, that are confused with burned regions. (b6, b7) show CuMedVision’s and U-Net’s predictions, using the whole spectral bands. Compared to the past predictions, both the approaches improved the segmentation accuracy and corrected the settlements misclassification, improving contour details. However, U-Net’s prediction tends to be more appropriate than CuMedVision’s one.

Figure 3.15 shows an example of mappings in an arid area, made of mountains presenting bare rocks, arable lands and shrubs. (c1) and (c2) show the satellite acquisition and the ground truth, respectively. (c3) Shows the BAE’s prediction: in this area, it is able to find the burned region, but it underestimates it. However, the presence of false positives is negligible. (c4, c5) show CuMedVision’s and U-Net’s predictions using visible-light data. Both approaches roughly delineate the area, overestimating the affected regions. Also, in this case, they present smooth contours which lack most of the details present in the GT. In this case, unburned regions appear to be very similar to the burned ones, and the prediction is approximated. (c6, c7) show CuMedVision’s and U-Net’s predictions, using the whole spectral bands. Compared to the past predictions, both approaches improved the segmentation recall. However, U-Net tends to be more precise than CuMedVision,

being also in this case the best model.

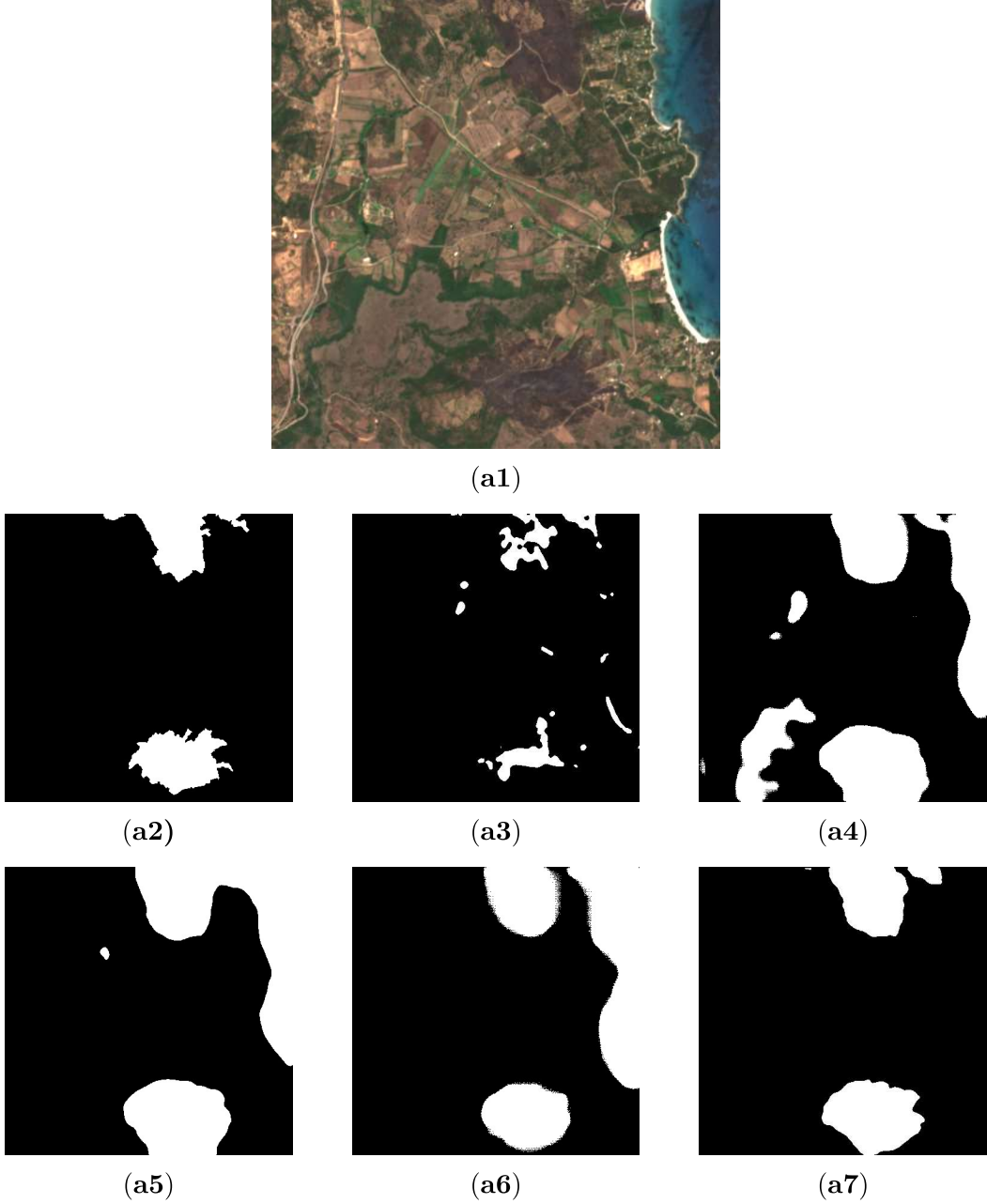


Figure 3.13: Burned area segmentation in a coastal area. (a1) Satellite acquisition of the burned region, realised using visible light spectrum; bands 4, 3 and 2 correspond to R, G, B channels, respectively. (a2) Ground Truth, derived from the Copernicus EMS delineation map. White pixels represent burned regions, while black pixels represent unburned regions. (a3) BAE's prediction, using visible light data. (a4, a5) CuMedVision's and U-Net's predictions, using visible light data. (a6, a7) CuMedVision's and U-Net's predictions, using all spectral bands.

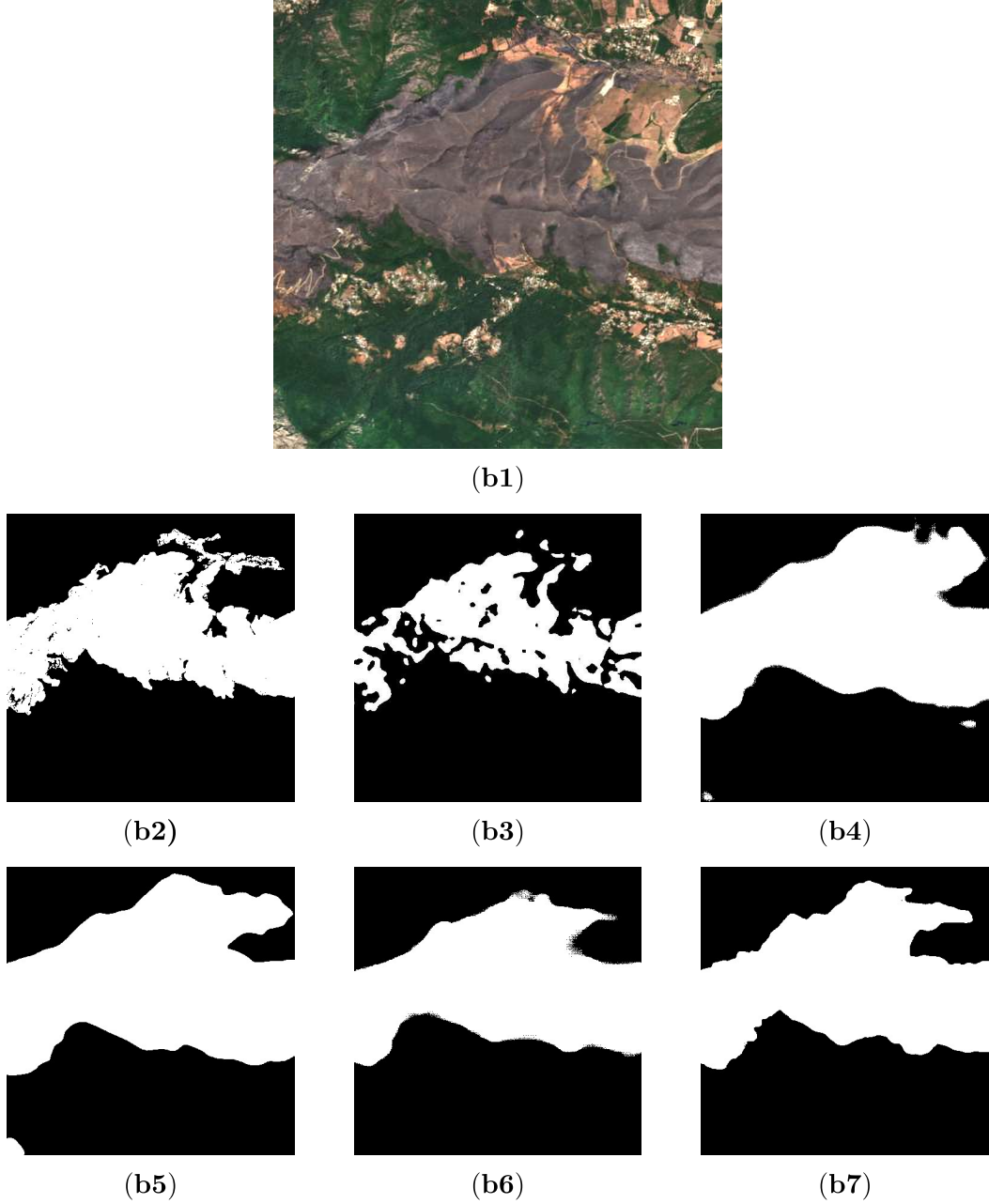


Figure 3.14: Burned area segmentation in a forest region, with the presence of settlements. (b1) Satellite acquisition of the burned region, realised using visible light spectrum; bands 4, 3 and 2 correspond to R, G, B channels, respectively. White pixels represent burned regions, while black pixels represent unburned regions. (b2) Ground Truth, derivated from the Copernicus EMS delineation map. (b3) BAE's prediction, using visible light data. (b4, b5) CuMedVision's and U-Net's predictions, using visible light data. (b6, b7) CuMedVision's and U-Net's predictions, using all spectral bands.

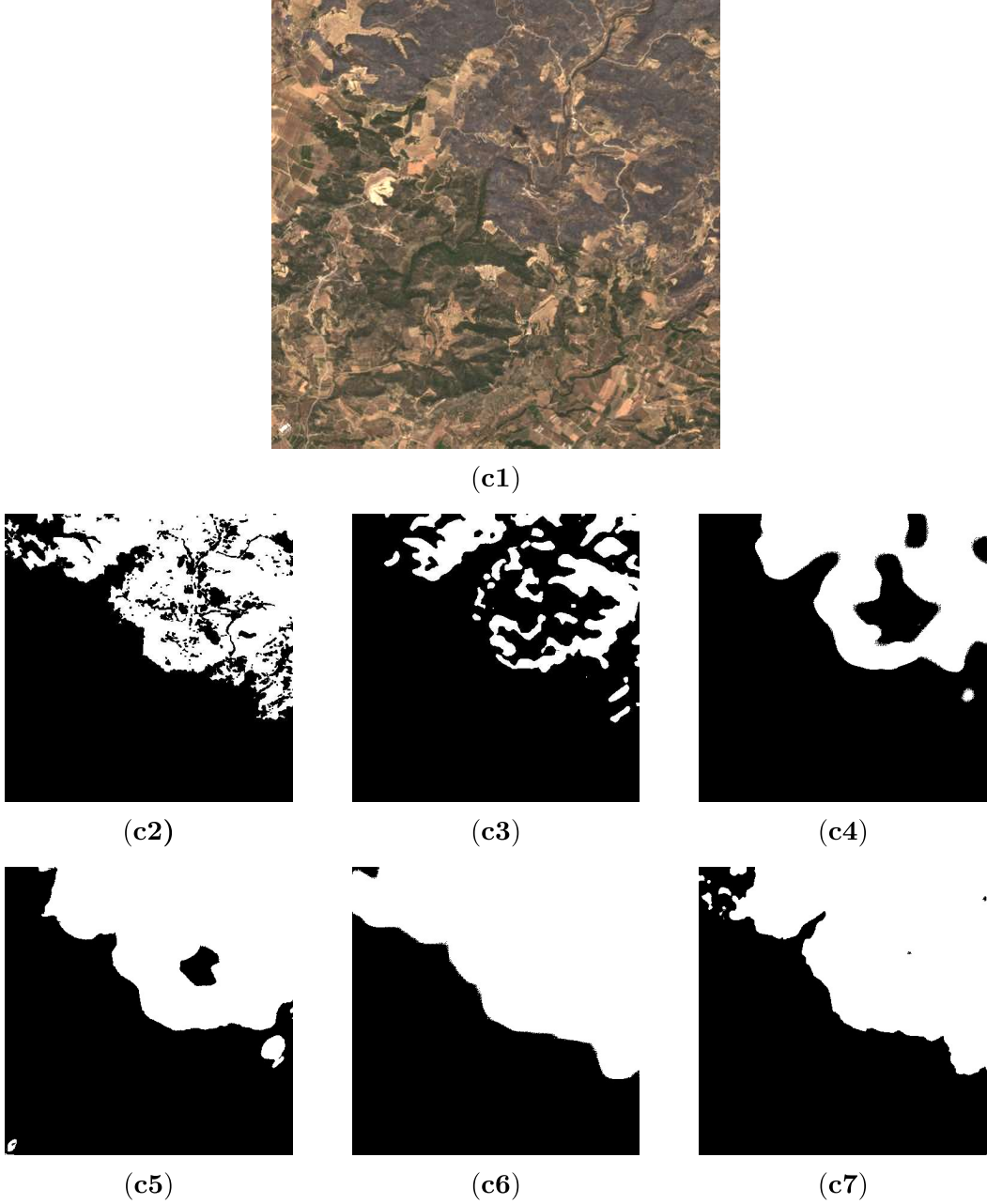


Figure 3.15: Burned area segmentation in an arid area, made of mountains presenting bare rocks, arable lands and shrubs. (c1) Satellite acquisition of the burned region, realised using visible light spectrum; bands 4, 3 and 2 correspond to R, G, B channels, respectively. (c2) Ground Truth, derived from the Copernicus EMS delineation map. White pixels represent burned regions, while black pixels represent unburned regions. (c3) BAE's prediction, using visible light data. (c4, c5) CuMed-Vision's and U-Net's predictions, using visible light data. (c6, c7) CuMedVision's and U-Net's predictions, using all spectral bands.

3.5.5 Results on Damage Severity estimation problem

As introduced in Section 3.4, Single U-Net, the Parallel U-Net, and Double-Step U-Net leveraged on all spectral data available from post-wildfire satellite acquisitions for making their predictions. A detailed performance report of the results achieved in every fold for the estimation of damage severity is shown in Table 3.5. The table presents the RMSE evaluated on every fold, for every severity level, reported as an ordinal number for the sake of space. Severity levels are mapped as follows; 0 stands for *No damage*, 1 stands for *Negligible to slight damage*, 2 stands for *Moderately Damaged*, 3 stands for *Highly damaged*, and 4 stands for *Completely destroyed*. To avoid ambiguities, we named Single U-Net the approach based on the Regression U-Net only.

Furthermore, the results were subjected to statistical tests, in order to determine whether the models performances were significantly different. The tests were performed using the RMSE scores of each acquisition in the dataset. Firstly, the Friedman test [52] was performed to determine whether, in at least one approach, the distribution of the scores was different from the other approaches. Within the Friedman test, we formulated the following hypotheses:

- Null hypothesis (H_0): the RMSE achieved by the assessed approaches are not statistically different;
- Alternative hypothesis (H_1): at least one approach achieved statistically different scores compared to the others.

If the Null hypothesis is rejected, the Nemenyi test [52] is performed to compare the approaches in pairs and to determine statistically different scores. Given A and B two different approaches, we formulated the following hypotheses:

- Null hypothesis (H_0): approaches A and B are equal;
- Alternative hypothesis (H_1): either A or B is better than the other.

Results comparison between U-Net-based approaches

Performance evaluation. In a first analysis, we do not consider the dNBR column, but we focus only on the three networks performances. The best score for each row, considering only the U-Net-based approaches, is marked with the star symbol (\star). Compared to the Single U-Net, the approaches in which the outputs of the Binary U-Net and Regression U-Net are combined showed better overall performances. This is due to the improvement of the segmentation performances for the unburned regions (severity level 0), brought by the output of the Binary U-Net in the Parallel U-Net. However, misclassified unburned regions by the Binary U-Net slightly worsened the RMSE in the remaining severity levels (e.g. in the yellow

fold, severity levels 1 and 4). The Double-Step U-Net is the most accurate in the discrimination between severity levels (1 to 4), achieving best results in 5 folds out of 7 (blue, fuchsia, green, orange, and yellow). The only exception is for the brown fold, which, as seen in the previous section, is the one where the Binary U-Net achieved a lower F1-Score.

In Double-Step U-Net, Regression U-Net is strongly dependent on the Binary U-Net’s performance: false positives not recognized in the delineation task are considered as belonging to a damage level by the Regression U-Net, with the result to increase the overall error. However, the RMSE values of the brown fold for the Double-Step U-Net result to be comparable to the RMSE values of other folds (i.e., orange and yellow). Also, Double-Step U-Net results to be robust in regions presenting strong differences from the ones that are used to train the model.

Statistical evaluation. The presented results were subjected to statistical tests to assess the real significance of the achieved performance, as shown in Table 3.6. The Nemenyi test, with $\alpha = 0.05$, was performed comparing: (i) Single and Parallel U-Net, (ii) Single and Double-Step U-Net, and (iii) Parallel U-Net and Double-Step U-Net. The test was conducted considering each fold and each severity level in the dataset. Single and Parallel U-Net showed statistical significance on the majority of the folds (5/7) only for the severity level 0. This behaviour confirms the improvement brought by the Binary U-Net for the identification of unburned regions in the Parallel U-Net approach.

Instead, the Double-Step U-Net shows statistical significance for every severity level and fold in the comparisons with Single and Parallel U-Nets.

Computation time evaluation. Performances were also evaluated according to the complexity and the inference time of the assessed approaches, as shown in Table 3.7. Times were measured from the beginning of the inference process, to the time the delineation map of an acquisition tile of 480×480 px was returned (all the dataset was considered for this study). Performances were evaluated running the approaches both on CPU (Intel Core I9 7940x with 128 GB RAM) and on GPU (NVIDIA 1080 Ti).

The Single U-Net, which is based on a Regression U-Net, show a computation time similar to the one achieved by the Binary U-Net, already assessed in Section 3.5.4 (U-Net, all bands). Being based on both the aforementioned U-Nets, the Parallel U-Net doubles the number of employed parameters and execution times. Similarly, the Double-Step U-Net show comparable parameters and computation times of the Parallel U-Net. All the severity estimation approaches are able to provide their estimate in either about 1.5 seconds on CPU or about 100 ms on GPU hardware.

Table 3.5: Cross-validation performance per fold. (*) indicates the best RMSE per severity category among the three U-Net versions. (†) indicates the best RMSE per severity category, dNBR included.

Fold	Severity	Performance (RMSE)			
		dNBR	Single U-Net	Parallel U-Net	Double-Step U-Net
Blue	0	0.78	1.06	0.23 *†	0.27
	1	1.07	0.89	0.89	0.73 *†
	2	1.23	0.71	0.80	0.62 *†
	3	0.82	0.63	0.65	0.52 *†
	4	0.62 †	0.93 *	0.96	1.44
Brown	0	0.65	0.22	0.20 *†	0.47
	1	0.97	0.94	0.94	0.92 *†
	2	1.01	0.65 *†	0.65 *†	0.86
	3	0.70	0.35 *†	0.35 *†	0.39
	4	0.48 †	1.26 *	1.28	1.49
Fucsia	0	0.82	0.39	0.16 *†	0.24
	1	1.37	1.40	1.41	1.02 *†
	2	1.12	1.35	1.35	1.00 *†
	3	1.10	0.97	0.97	0.75 *†
	4	1.67	1.26 *†	1.28	1.49
Green	0	0.20	0.28	0.04 *†	0.18
	1	0.64 †	1.03	1.03	0.80 *
	2	1.18 †	1.78	1.78	1.40 *
	3	1.46	1.87	1.90	1.38 *†
	4	1.09	1.57	1.58	1.00 *†
Orange	0	0.42 †	0.40	0.39 *	0.43
	1	1.10 †	1.68	1.68	1.47 *
	2	1.04	1.14	1.14	1.02 *†
	3	-	-	-	-
	4	-	-	-	-
Red	0	0.20	0.21	0.15 *†	0.33
	1	0.66 †	0.71 *	0.71 *	1.21
	2	0.80	0.56 *†	0.56 *†	0.97
	3	-	-	-	-
	4	0.58 †	1.96	1.96	1.21 *
Yellow	0	1.31	0.37	0.25 *†	0.54
	1	0.83 †	0.83*	0.84	1.04
	2	1.24	0.89	0.89	0.71 *†
	3	-	-	-	-
	4	0.99 †	1.70	1.71	1.18 *

Table 3.6: Statistical significance between grading maps produced by the approaches shown in Table 3.5, considering different folds (shortened to the second letter) and severity levels. The Nemenyi test was performed with $\alpha = 0.05$. Check marks (✓) highlight statistical relevance (null hypothesis is rejected). Dashes (-) mark unavailable severity for the corresponding fold.

Test	Severity	Fold						
		Bl	Br	Fu	Gr	Or	Rr	Ye
Single U-Net - Parallel U-Net	0	✓		✓	✓		✓	✓
	1							
	2							
	3					-	-	-
	4						-	-
Single U-Net - Double-Step U-Net	0	✓	✓	✓	✓	✓	✓	✓
	1	✓	✓	✓	✓	✓	✓	✓
	2	✓	✓	✓	✓	✓	✓	✓
	3	✓	✓	✓	✓	-	-	-
	4	✓	✓	✓	✓	✓	-	-
Parallel U-Net - Double-Step U-Net	0	✓	✓	✓	✓	✓	✓	✓
	1	✓	✓	✓	✓	✓	✓	✓
	2	✓	✓	✓	✓	✓	✓	✓
	3	✓		✓	✓	-	-	-
	4	✓	✓	✓	✓	✓	-	-
dNBR - Double-Step U-Net	0	✓	✓	✓			✓	✓
	1	✓	✓	✓	✓	✓	✓	✓
	2	✓	✓	✓	✓		✓	✓
	3	✓		✓	✓	-	-	-
	4	✓	✓	✓		✓	-	-

Table 3.7: Inference times of the assessed methods for the damage severity estimation task, considering input tiles of dimension 480×480 px and 12 bands.

Method	# params	Inference time (ms)			
		Avg (CPU)	Std (CPU)	Avg (GPU)	Std (GPU)
dNBR	-	3	2	-	-
Single UN	31 Mln	788	31	62	0.3
Parallel UN	62 Mln	1582	43	104	0.5
Double-Step UN	62 Mln	1511	53	103	2

Overall results discussion

Considering the dNBR in the evaluation, the best performances per row in Table 3.5 are marked with the dagger symbol ([†]). In order to compare the dNBR with the GT, its values were thresholded according to the default values [77]. In this case, best results vary from fold to fold, but generally, the are matched by U-Net based approaches. It must be considered that the dNBR is computed using both pre- and post-fire acquisitions, whereas U-Net’s approaches consider only post-fire acquisitions. In order to summarize the performance, the average RMSE value for each severity level is shown in Table 3.8. Compared to the Single U-Net, the Double-Step U-Net results to be a better approach, achieving the best RMSE on each severity level. Moreover, with reference to the dNBR, Double-Step U-Net achieves comparable performance, with a noticeable improvement for the detection of the unburned area (severity 0), using only half of the information. Note that, on average, the Double-Step U-Net is the only approach to achieve better results (lower RMSE) than the dNBR.

In Table 3.6, the statistical test between dNBR and Double-Step U-Net highlight that, even if they provide comparable performances in terms of RMSE, their results are significantly different for every severity level.

The reason behind the success of Double-Step U-Net is hidden in the problem split. First, the neurons of the Binary U-Net are employed to identify burned regions. Its prediction will mask the spectral values of unburned regions, leaving only the information related to burned areas to the Regression U-Net. Therefore, the latter network will employ its neurons in finding differences between correlated values (severity levels 1 to 4).

It is worth mentioning that the masking operation performed by the Binary U-Net prediction introduces a new and uncommon value in the spectral information fed as input to the Regression U-Net: the 0 value. Areas identified as unburned will be “cancelled” by replacing their original value with 0, which is not present in nature. Therefore, a bad classification from the Binary U-Net can lead the Regression U-Net to make more mistakes because it will consider 0-valued-regions as unburned and every other unburned region not detected by the Binary U-Net will be considered as burned.

In Figure 3.16, a comparison between predictions of dNBR, Single U-Net, and Double-Step U-Net is shown in two areas of the green fold. At a first glance, delineating the wildfire contours just looking at the RGB acquisition (pictures a1 and b1) seems feasible, but assigning different severity levels appears to be more challenging. In both the acquisitions, the Binary U-Net predictions resulted to be highly accurate (pictures a3 and b3), compared to the Copernicus EMS annotation (GT, pictures a2 and b2). The dNBR (pictures a4 and b4) show a good match

Table 3.8: Average performance for severity level. (*) indicates the best RMSE per severity category among the three U-Net versions. (†) indicates the best RMSE per severity category, dNBR included.

Severity	Overall Per-Class Performance (RMSE)			
	dNBR	Single U-Net	Parallel U-Net	Double-Step U-Net
0	0.62	0.42	0.20 *†	0.35
1	0.95 †	1.07	1.08	1.03 *
2	1.09	1.01	1.02	0.94 *†
3	1.02	0.95	0.97	0.76 *†
4	0.91 †	1.45	1.46	1.30 *
Avg.	0.92	0.98	0.94	0.88*†

with the GT, except for some noise in the vast unburned regions. The Single U-Net (pictures a5 and b5) correctly identifies the burned region and the contours of different burned areas, but it tends to underestimate the severity. In the end, the Double-Step U-Net (pictures a6 and b6) improves the prediction of the Single U-Net, resulting to be more similar to the GT.

Considering the computation time evaluation, all the approaches are suitable to provide near-real-time mappings, especially on GPU hardware.

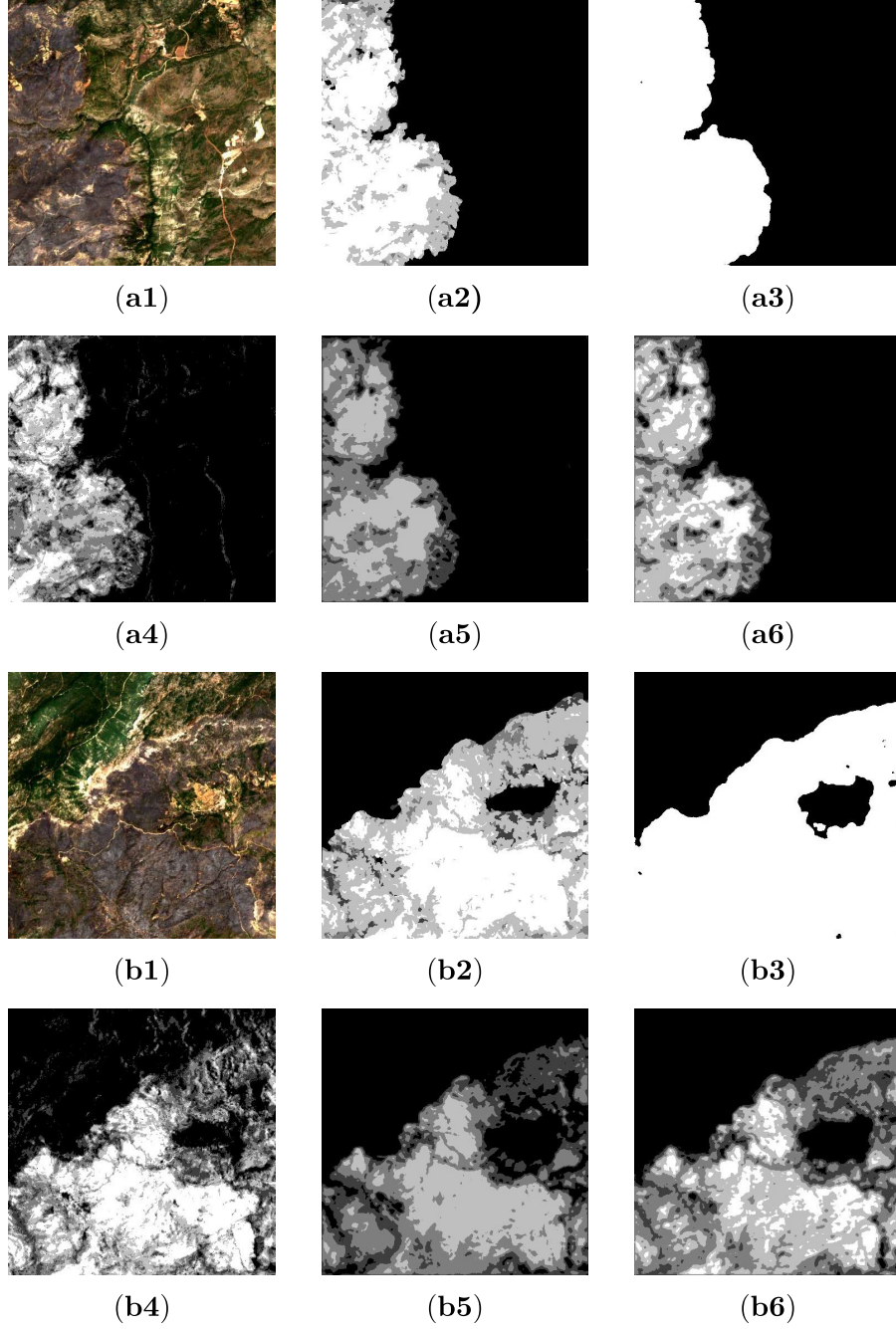


Figure 3.16: Grading maps for the the estimation of the damage severity level. Severity levels are presented though five shades of grey, ranging from black (severity = 0) to white (severity = 4). **(a1, b1)** Sentinel-2 L2A acquisition; **(a2, b2)** Copernicus EMS grading map (GT); **(a3, b3)** Binary mask generated by the Binary U-Net: black and white colors indicate unburned and burned regions, respectively; **(a4, b4)** Thresholded dNBR, obtained from pre and post-fire acquisitions; **(a5, b5)** Single U-Net prediction; **(a6, b6)** Double-Step U-Net prediction.

3.6 Summary

This chapter presented the problem of burned area mapping from Sentinel-2 acquisitions, for which we addressed both delineation and grading tasks. Compared to the approaches presented in the literature, which require pre- and post-wildfire imagery to deliver accurate mappings, the goal of this chapter was to present reliable solutions to the problem, that could only leverage post-wildfire imagery. Using only one acquisition avoids human intervention, usually needed either to validate proper pre-fire data or to tune spectral indexes on post-fire data, other than halving the needed information to accomplish the same task.

Firstly, we assessed the delineation task considering either a small portion of the spectrum related to the visible light or the whole spectral data. For that task, we proposed BAE, an unsupervised approach, which was compared with CuMedVision and U-Net, two supervised approaches. On average, BAE shows similar precision (~ 0.68) with respect to other approaches, but it achieves lower recall. The best model, in terms of both Precision (~ 0.72), Recall (~ 0.76) and F1-score (~ 0.70) was U-Net. When considering the whole spectrum, we selected the NBR as the spectral index having the maximum separability, and we used it as the baseline. To simulate the human manual intervention, we chose the best threshold values for the NBR in the dataset samples to maximize the results. Then, the baseline was compared with CuMedVision and U-Net, in which the latter was confirmed to be the best approach in all the three metrics, achieving, on average, Precision of 0.80, Recall of 0.97 and F1-Score of 0.86.

Given the reliability of the U-Net approach for the delineation task, we assessed that model also for the grading mapping task. In this scenario, we proposed a new model named Double-step U-Net, based on the intuition that the problem to be solved could be split into two sub-problems: to distinguish between burned and unburned areas and to discriminate the right severity in the burned regions. Within Double-step U-Net, we proved that the knowledge acquired for solving the first subtask positively influence the solution of the second subtask. We compared U-Net and Double-step U-Net, trained on post-wildfire data, with dNBR, the spectral index computed using both pre- and post-wildfire data, that is the official approach used by Copernicus. As a result, Double-step U-Net achieved the best RMSE on all the severity levels, if compared with U-Net, and it gives comparable results with respect to the dNBR but using only half of the information.

3.7 Relevant Publications

Farasin, A., Nini, G., Garza, P., & Rossi, C. (2019). Unsupervised Burned Area Estimation through Satellite Tiles: A multimodal approach by means of image segmentation over remote sensing imagery. CEUR Workshop Proceedings, 2019, 2466

[45]

Farasin, A., Colomba, L., Palomba, G., Nini, G., & Rossi, C. (2020, May). Supervised Burned Areas delineation by means of Sentinel-2 imagery and Convolutional Neural Networks. In Proceedings of the 17th International Conference on Information Systems for Crisis Response and Management (ISCRAM 2020), Virginia Tech, Blacksburg, VA, USA (pp. 1060-1071). [44]

Farasin, A., Colomba, L., & Garza, P. (2020). Double-Step U-Net: A Deep Learning-Based Approach for the Estimation of Wildfire Damage Severity through Sentinel-2 Satellite Data. *Applied Sciences*, 10(12), 4332. [43]

Chapter 4

Flood delineation assessment using Satellite data

This chapter presents two works that bring relevant information in the context of the emergency management of flood events. The first work uses Sentinel-1 acquisitions to provide automatic mappings of flooded areas. The main objective is to identify a reliable approach, able to produce flood delineation maps leveraging only on a single SAR acquisition of the area of interest and, optionally, on cartographic information about natural water sources. The second work proposes an expert system able to evaluate persisting flooded areas in cities. Given an area of interest, the approach exploits a time series of Sentinel-2 acquisitions and determines the presence of non-natural water bodies that persist all along with the considered time range. Those works aim to increase the understanding of the affected regions, supporting activities of prompt intervention (response phase) and providing information useful to update risk maps and to plan a proper restoration of the environment (recovery phase).

The chapter is structured in two sections, presenting the two works. Section [4.1](#) is about the automatic delineation of flooded areas, while Section [4.2](#) presents the expert system for the identification of persisting flooded areas. Both sections share a similar structure: after introducing the main objective and the context of the application, a subsection about Data sources describes the data employed in the work. Then, the problem to be solved is formalized in the Problem Statement subsection. After, the Methodology is explained, presenting the proposed approaches. Furthermore, Experiments describe the training/testing process and the metrics used to evaluate the approaches. Finally, results are presented and discussed.

4.1 Flood delineation using Sentinel-1 data

In this section, we present an assessment for the delineation of floods from Sentinel-1 acquisitions and cartographic information, exploiting supervised approaches: (i) supervised machine learning algorithms used in literature, and (ii) U-Net, which has been largely explored in the previous chapter. We evaluate their performances on several Copernicus EMS flood delineation maps distributed in different geographical regions [116]. The objective of the assessment is to identify the best model, able to create fast mappings of the flood extension automatically. The model will contribute to saving time and workforce for creating delineation maps, which will be particularly important during the Response and Recovery phases. The capacity to map the extent of flooded areas in a timely and accurate manner is critical for two reasons: (i) 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, and (ii) for the creation of a fast mapping service, that can be used to provide extra information to first responders during the emergency response phase.

4.1.1 Data sources

In this work, we considered Copernicus EMS delineation maps of flood produced between 2014, the year in which Sentinel-1 satellites were launched, and 2018, when we performed this study.

To use a supervised approach, we require ground truth masks, which we created from the vector data provided in the EMS delineation maps. For each map, we created a binary mask identifying flooded areas, whereas pixels belonging to flooded regions are set to 1, 0 otherwise. Also, we considered cartographic information to create hydrography maps of natural water sources. In this case, hydrography was obtained by OpenStreetMap [114], a service that creates and distributes free geographic data for the world. Cartographies were obtained using the same AoI specified in the Copernicus delineation maps and were transformed into binary masks, where pixels equal to 1 indicate a water source, 0 otherwise.

Concerning satellite data, Sentinel-1 has the advantage of operating at wavelengths not impeded by cloud cover or a lack of illumination and can acquire data over a site during day or night time under all weather conditions. However, because Sentinel-1 satellites gather data in stripes while following an orbit, it is possible that certain acquisitions that meet the aforementioned conditions are incomplete, covering only a fraction of the targeted region; in such circumstances, the data were discarded. Sentinel-1 data is downloaded from the Sentinel-Hub Service as images with a spatial resolution of 10x10m using IW mode and the RGB_RATIO configuration, which maps the input bands given by the different polarizations of the SAR instrument into a false RGB image. It uses the VV channel for red, 2 times the

Table 4.1: Copernicus EMS delineation maps considered in the study.

Country	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

value of the 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 geo-referenced and orthorectified. Depending on the requested AOIs, the downloaded GeoTIFF has a size ranging between 1000-2000 x 2000-3000 pixels.

Our dataset is 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 4.1 the composition of the dataset, displaying the country to which the maps belong, the map code, and the activation date.

4.1.2 Problem Statement

The problem involves a Sentinel-1 acquisition, taken during a flood event. Considering the available data obtained using the SAR IW mode, namely VV and VH, and a binary mask of the natural water sources, the goal is to predict a binary mask of the flooded regions, where pixels equal to 1 mean flood, and pixels equal to 0 mean non-flood. Note that only the flooded area must be identified, not the natural water sources, i.e. rivers, lakes. Therefore, the problem is configured as a binary segmentation task, also known as delineation task in the geospatial context. To accomplish the task, a set of annotated data, consisting of Sentinel-1 acquisitions and binary masks, is available for supervised approaches.

4.1.3 Methodology

The steps involved in flood delineation process are depicted in Figure 4.1. Firstly, the raw Sentinel-1 satellite acquisition is taken from the Sentinel-Hub service. Raw SAR acquisitions are affected by speckle noise that is generated during the acquisition process, due to back-scattered waves from multiple distributed targets. To reduce the noise, the second step foresees a despeckling operation. In this study, we used an approach largely adopted in literature, the Non-Local Means (NL-means) filter [16]. For each pixel p , it recomputes its value as a weighted average of the square neighbourhood of fixed size $k = 5$ centered at p , where the weights depend on the distance between the pixel p and its neighbourhood. Then, cartography about natural water sources is employed: pixel values are initially inverted (I), setting non-flooded areas to 1 and flooded ones to 0; then it is multiplied pixel-wise with the despeckled image. Therefore, all the natural water regions are set to 0 and diversified from other water regions. Finally, the input data is processed by one of the examined approaches: Support Vector Machines (SVM), Random-Forest, and U-Net, which provides the delineation map. According to the model that is applied, the input data is properly pre-processed, as will be explained later in this section.

Baseline: Support Vector Machine

In 2019, a solution proposed by Benoudjit and Guida [4], performs the mapping of the flood extent on SAR images using the Stochastic Gradient Descent (SGD) algorithm to optimize the loss function of a supervised classification algorithm. SGD is an iterative method for optimizing an objective function, employed in several machine learning algorithms, like Support Vector Machines (SVM), K-Means, and Feed-Forward Neural Networks (FFNN) [12]. Also, it is recently known as one of the weights optimization algorithms for Deep Neural Networks [61]. The authors used that approach to optimize the cost function of a SVM classifier, applied pixel-wise on input data.

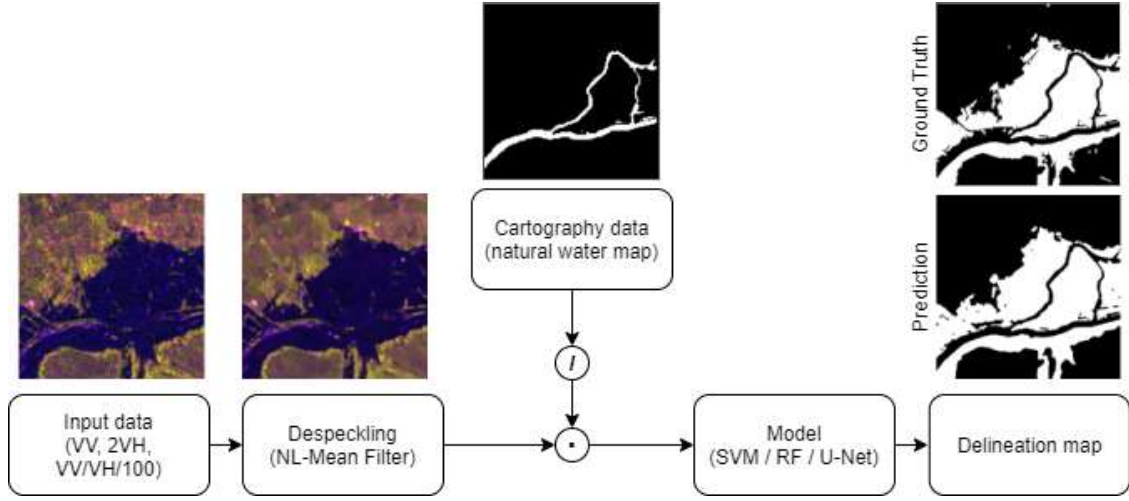


Figure 4.1: Flow-diagram of the flood delineation process. Input data is subjected to despeckling operation, which removes noise. Cartography mask is inverted (I) and multiplied pixel-wise (\cdot) with the despeckled data, highlighting natural water sources. Finally, the model segments the flooded regions.

Before training the model, both the input data and the ground truth, having dimensions ($width, height, 3$), are reshaped into two vectors of lengths $width \times height \times 3$. Then, the SVM is trained using the SGD to optimize the cost function, defined as follows:

$$L(x_j, y_j) = H(x_j, y_j) + L2 = \max(0, 1 - y_j \cdot (\omega x_j + b)) + \alpha \|\omega\|_2 \quad (4.1)$$

H is the Hinge loss function, where ω and b are the optimized model parameters, while x_j and y_j are the j -th element of the input and the ground truth vectors, respectively. $L2$ is a regularization terms, also known as Ridge regularization term, used to help the model to generalize to unlabeled data, preventing the overfitting [52]. The incidence of the regularization term, is tuned by the parameter $\alpha > 0$.

In our work, we followed the same procedure, but we extended the set of features to be evaluated by the SVM, adding some 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 the human eye), would be better recognized. Instead of considering one pixel at a time, SVM will also consider a region around the pixel to be evaluated. Therefore, SVM will consider a squared region having dimensions $(w \times w)$ pixels, where w is an odd number. The pixel at the center of each squared region, p_c is the one subjected to classification. For p_c in the border of the original input data, the missing part of surrounding pixels is replaced by mirroring the portion of the considered data.

The ground truth remains a vector that specifies the class of each pixel: flooded or

non-flooded.

Random Forest

The first model tested is a Random Forest Classifier (RF), well known to be one of the most versatile Machine Learning algorithms suitable for classification. Random Forest is an ensemble model, based on the training of a pre-defined number of Decision Trees on different subsets of features over the same dataset, where each Decision Tree learns to classify new samples using a subset of the feature set. Also, in this case, the input data is not processed at once, due to large dimensions. Instead, we adopted the same windowing approach applied with SVM.

U-Net

The third and last approach compared in this work is U-Net, the deep convolutional neural network largely discussed in the previous chapter. Indeed, the same architecture described in Chapter 3 for binary classification is used in this work. Similarly to delineation maps of burned areas, flooded areas present common aspects: they appear as shapes of different sizes having irregular borders, which sometimes present protrusions. Flood delineation maps may identify either one or many regions in the inspected area of interest, which determines the unusability of the loss function presented in the original paper. Given the promising results obtained in the previous chapter, we decided to maintain the same loss function: the Dice loss.

As in the cases of SVM and RF, input data is too large to be computed at once, also by U-Net. Therefore, both input data and ground truth are tiled, generating images of dimensions $(480 \times 480 \times 3)$ and $(480 \times 480 \times 1)$, respectively.

4.1.4 Experiments

This section presents the experiments for the three approaches previously introduced. First, the raw satellite data illustrated in Section 4.1.1 is preprocessed, preparing the dataset. Then, the testing process and the evaluation metrics are presented. Finally, the results are discussed.

Dataset preparation

As introduced in Section 4.1.1, the satellite acquisitions are related to five different countries where flood events occurred: Australia, Greece, Ireland, Italy, United Kingdom. However, for computational limitations, none of the approaches previously presented is able to process a whole acquisition at once. Therefore, we opted to tile each acquisition in smaller crops of size 480×480 pixels, maintaining the original information. In order to avoid the presence of many tiles without flood,

we considered in the dataset only tiles containing at least one pixel classified as flooded. In the end, the dataset contains a total of 64 tiles, distributed in folds as follows; AU fold: 8, GR fold: 8, IR fold: 21, IT fold: 11, UK fold: 16.

Each tile in the dataset has the same dimension of the U-Net’s input, while it has to be adjusted for SVM and RF. In particular, for the two approaches we considered scrolling windows of dimensions (7×7) pixels, composed of: a central pixel (p_c), and 48 neighbour pixels. Therefore, each tile having dimensions (480×480) is transformed into a matrix of 480×480 rows, by $7 \times 7 \times 3$ columns (consider that each pixel is composed of 3 values).

Testing process

The goal of the experiments is to evaluate the approaches in different areas, proving their ability to operate independently from morphological and geological aspects. Therefore, the dataset is split into five folds, grouping each acquisition to the respective country. In order to evaluate the capability of each model to obtain good results on different geographical areas, we compute the model performances using the Cross-Validation approach [150]. A total of 5 tests is foreseen: every test is performed on a different fold, while the remaining four are used to train the models. In the case of U-Net, one of the four remaining sets is used for validation purposes.

Moreover, we want to assess the contribution brought by both despeckling operation and cartography. Therefore, we performed an ablation study, identifying three test cases: (i) using raw data, (ii) using despeckled data, and (iii) using both despeckled data and cartography about natural water sources. To provide a fair comparison, we run cross-validation for each of the three test cases.

Evaluation metrics

After analyzing the dataset, it emerged that the number of non-flooded pixels is considerably higher than the number of flooded ones: precisely, the ratio of the non-flooded pixels with respect to the total pixel count is 80.7%. This means, in the Machine Learning domain, that the classes are imbalanced. This situation can lead models to underfit the training data, with the consequence to be more error-prone during the test phase.

For the same reason using *accuracy* to assess the performances is not reliable at all: in this situation of class imbalance, any approach could achieve about the 80% of accuracy by just classifying all the pixels as non-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, we chose an evaluation metric that is reliable with imbalanced classes, the F1-Score. Also, we report Precision and Recall metrics to better evaluate the three approaches.

Hyperparameters settings

SVM is trained according to the parameters provided in the original paper: the model is trained for 1000 iterations, while the regularization coefficient α is set equal to 0.0001.

RF is trained using 19 decision tree models. That number is empirically chosen as follows. Considering the third test case, which involves both despeckling and cartography, we performed the cross-validation using: 3 sets as training, 1 as test and 1 is excluded. In this phase, the test set is used to assess the best number of trees: in the actual experiments, the real test set will be the set we excluded in this phase. We tested a number of trees which ranged from 2 to 50 and we determined that the best performances were obtained with 19 trees. Each decision tree uses the Gini index as a measure of the split quality. Finally, *RF* uses the majority vote to determine the prediction for each pixel [52].

U-Net is trained according to the specifications provided in Section 3.5.3 for the binary classification problem, with the only exception of not using augmentation, in order to avoid that U-Net was trained on more data than the one available for *SVM* and *RF*.

4.1.5 Results

In this section, the models' performances are evaluated using the same dataset and performing the same cross-validation process, then the results are compared. Furthermore, the models are trained and evaluated on the dataset in three distinct configurations to investigate the influence of pre-processing activities on the results: (i) using only raw data, as acquired from the Sentinel-1, (ii) using raw data, after being processed with a despeckling operation to reduce the noise caused by the acquisition process, and (iii) using despeckled data with the addition of the hydrography mask. Then, statistical tests assess similarities among the delineation maps generated by the studied approaches. Finally, computation times are compared and discussed.

Performance evaluation

Table 4.2 reports the experimental results obtained for the first test case, which considers raw data only. The *SVM* achieves high Recall (0.79), but just sufficient Precision (0.61). The speckle noise, inherently contained in the raw data, induce the linear model to detect more false positives, and therefore to overestimate the

predicted flooded area, lowering the Precision score. Instead, RF shows an opposite behaviour compared to SVM. It is more accurate in avoiding false positives, achieving the best Precision score (0.78), but it tends to underestimate the flooded area. However, it shows a higher F1-Score, of about 0.75. The best model in this test case is undoubtedly the U-Net, which presents both high and balanced Precision and Recall, which result in the highest F1-Score, on average, of about 0.80.

Table 4.2: Cross-validation results on Test case #1 (Raw data only). (†) marks best Precision values, (★) marks best Recall values, and **bold text** marks best F1-Score values.

Fold	SVM			RF			U-Net		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
AU	0.62	0.76	0.68	0.77	0.62	0.69	0.83 [†]	0.94 [★]	0.88
GR	0.54	0.90	0.68	0.87	0.89	0.88	0.88 [†]	0.90 [★]	0.89
IR	0.58	0.71	0.64	0.70 [†]	0.77	0.74	0.69	0.80 [★]	0.74
IT	0.65	0.68 [★]	0.66	0.73 [†]	0.51	0.60	0.68	0.61	0.65
UK	0.69	0.88 [★]	0.77	0.80 [†]	0.76	0.78	0.79	0.86	0.82
Avg.	0.61	0.79	0.69	0.78 [†]	0.71	0.74	0.77	0.82 [★]	0.80

Table 4.3 reports the experimental results obtained for the second test case, which considers raw data preprocessed with the despeckling operation. The reduced noise largely improves the performances of SVM, especially for the Precision score, which raises by +18% from the previous test case. However, the artefacts introduced by the preprocessing operation made the Recall score lower of -6%. Note that the average scores for the SVM in this test case are comparable to the ones achieved by RF in Test case #1.

The despeckling operation barely affected RF and U-Net performances, which slightly improved their scores by decimals. However, U-Net was confirmed to be the best model, achieving best Precision and Recall scores in the majority of the folds (3/5), and best F1-Scores in almost every fold (4/5).

Table 4.3: Cross-validation results on Test case #2 (Despeckled data). (†) marks best Precision values, (★) marks best Recall values, and **bold text** marks best F1-Score values.

Fold	SVM			RF			U-Net		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
AU	0.66	0.96 [★]	0.78	0.80	0.71	0.75	0.92 [†]	0.86	0.89
GR	0.79	0.74	0.76	0.87	0.84	0.85	0.89 [†]	0.88 [★]	0.89
IR	0.74	0.70	0.72	0.69	0.76 [★]	0.73	0.79 [†]	0.72	0.75
IT	0.83 [†]	0.55	0.66	0.73	0.54	0.62	0.64	0.66 [★]	0.65
UK	0.94 [†]	0.71	0.81	0.79	0.77	0.78	0.84	0.82 [★]	0.83
Avg.	0.79	0.73	0.75	0.78	0.72	0.75	0.82 [†]	0.79 [★]	0.80

Finally, Table 4.4 reports the experimental results obtained for the third test case, which considers despeckled data, with the addition of hydrography layer. In this case, the SVM presented slight improvements compared to the previous test case. Instead, both RF and U-Net widely improved their performances. The greatest improvements are achieved by RF, presenting an average of +11% on both Precision and Recall. The large increment in Precision is directly justified by the presence of pixels into the hydrography regions that allows the model to reduce false positives. Also, flooded regions close to natural water sources are better detected, resulting in an improvement in the Recall score.

Even if their average F1-Score is similar, 0.85 for RF and 0.86 for U-Net, we acknowledge U-Net as the best model, because it presents the highest Recalls and F1-Scores in the majority of the folds.

Table 4.4: Cross-validation results on Test case #3 (Despeckled data + Hydrography). (\dagger) marks best Precision values, (\star) marks best Recall values, and **bold text** marks best F1-Score values.

Fold	SVM			RF			U-Net		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
AU	0.56	0.97 \star	0.71	0.83	0.96	0.89	0.91 \dagger	0.84	0.87
GR	0.78	0.73	0.76	0.92	0.92 \star	0.92	0.94 \dagger	0.91	0.93
IR	0.78	0.74	0.76	0.88 \dagger	0.83	0.85	0.85	0.86 \star	0.86
IT	0.93	0.62	0.74	0.94 \dagger	0.59	0.72	0.69	0.82 \star	0.75
UK	0.96 \dagger	0.72	0.82	0.90	0.87	0.88	0.88	0.90 \star	0.89
Avg.	0.80	0.76	0.76	0.89 \dagger	0.83	0.85	0.85	0.87 \star	0.86

Statistical evaluation

The delineation maps generated in the three test cases were subjected to a statistical test, the McNemar’s test, in order to check whether the outputs in assessed approaches could be interpreted as equal. The McNemar’s test is used to determine if there are differences on a dichotomous dependent variable between two related groups [82]. In this context, the dichotomous dependent variable is represented by the value of each pixel in the delineation map, while the two related groups are the models’ predictions. For each fold, we considered the predictions of all the delineation maps generated, comparing the approaches in pairs. Given A and B two different approaches, we formulated the following hypotheses:

- Null hypothesis (H_0): delineation maps generated by A and B are equal;
- Alternative hypothesis (H_1): delineation maps generated by A and B are significantly different.

Therefore, we considered the following groups: (i) SVM - RF, (ii) SVM - U-Net, (iii) RF - U-Net. The McNemar’s tests were performed for each fold in the dataset and repeated for each Test case analyzed above (Raw data, Despeckled data, Despeckled data + Hydrography maps).

As a result, with a significance level $\alpha = 0.05$, all the tests rejected the null hypothesis: all the algorithms provide significantly different delineation maps.

Computation time evaluation. Performances were also evaluated according to the complexity and the inference time of the assessed approaches, as shown in Table 4.5. In this evaluation, we do not provide any distinction between test cases, as the approaches performed similarly on each test case. Times were measured from the beginning of the inference process, to the time the delineation map of an acquisition tile of 480×480 px was returned (all the dataset was considered for this study). Performances were evaluated running the approaches both on CPU (Intel Core I9 7940x with 128 GB RAM) and on GPU (NVIDIA 1080 Ti). SVM is the lightest model in terms of the number of parameters and also the fastest one, considering the test on CPU (~ 217 ms per tile). Random Forest shows a high number of parameters (> 20 Mln), considering all the Decision Trees in the ensemble. However, it is slightly faster than U-Net on CPU, generating a delineation map in less than 600 ms, on average. U-Net computation time is in line with the results achieved using visible light bands for burned area delineation, as explained in Section 3.5.4.

Table 4.5: Inference times of the assessed methods for the delineation task, considering input tiles of dimension 480×480 px and 3 channels.

Method	# params	Inference time (ms)			
		Avg (CPU)	Std (CPU)	Avg (GPU)	Std (GPU)
SVM	< 100	217	15	-	-
RF	21 Mln	593	33	-	-
U-Net	28 Mln	716	24	47	0.4

Overall considerations

Overall, U-Net demonstrated to be, also in this context, the most reliable approach, achieving the highest results in all test cases. The despeckling operation resulted to be useful when using linear approaches, like the SVM. In the end, the hydrography map demonstrated that none of the examined approaches is able to distinguish natural water sources from the flooded areas properly. Therefore, it represents essential information to enable the model to provide highly accurate delineation maps. Figure 4.2 shows an example of the performance gains that have

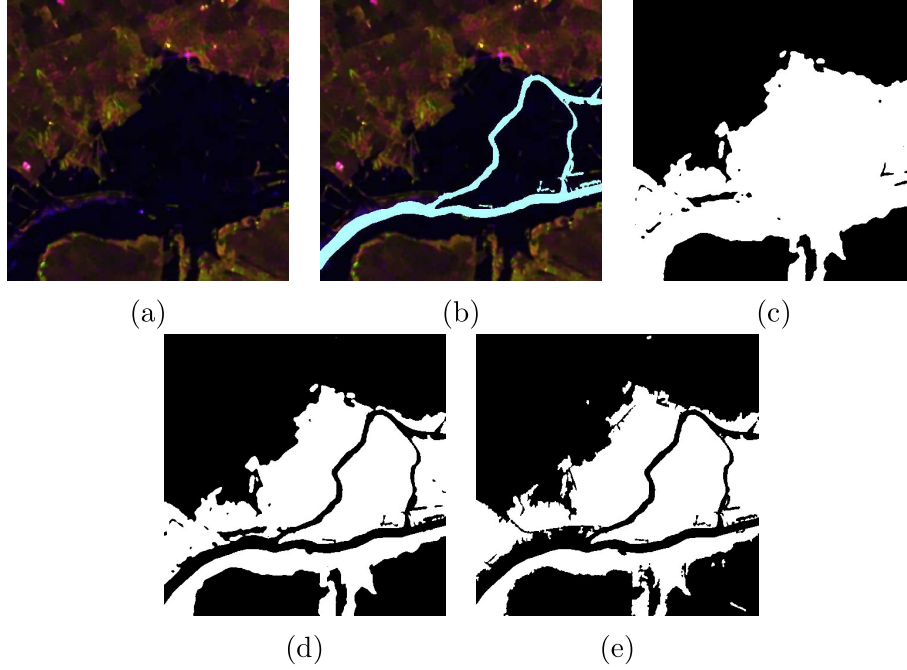


Figure 4.2: Example of how the addition of a hydrography layer improved the U-Net’s performance (EMSR149 - 13PORTUMNA): (a) Despeckled data (no hydrography), (b) Despeckled data with hydrography (colored in light blue), (c) Delineation obtained from despeckled data without hydrography (F1-Score = 0.89), (d) Delineation obtained from despeckled data with hydrography (F1-Score = 0.97), (e) Ground truth.

resulted from the usage of hydrography. In Figures 4.3 and 4.4 we report the predictions of the examined models in two areas of the Test case #3 where U-Net performed best and worst, respectively. In both areas, all three techniques are able to provide qualitative results. However, the main differences are noticeable between SVM and RF / U-Net predictions. SVM tends to be more prone to errors, presenting false positives in Figure 4.3 and with false negatives in Figure 4.4. RF and U-Net present very similar and qualitative results, errors are generally related to false negatives.

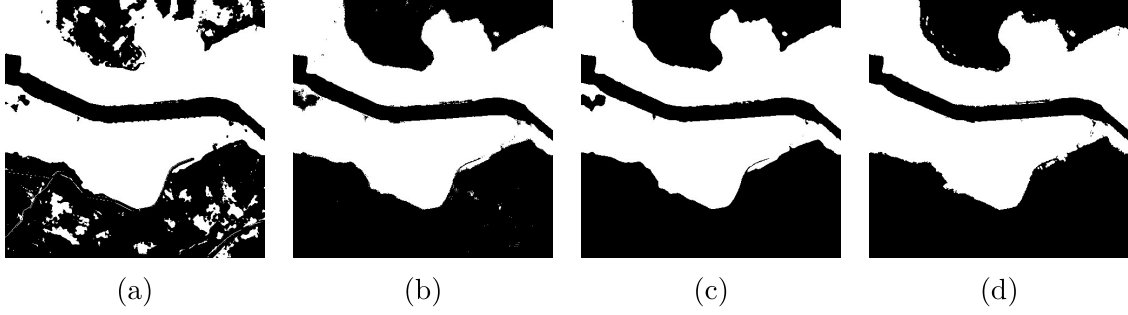


Figure 4.3: Performances comparison on U-Net model best result (EMSR122 - STRYMONAS): (a) SVM (F1-Score = 0.91), (b) RF (F1-Score = 0.98), (c) U-Net (F1-Score = 0.99), (d) Ground truth

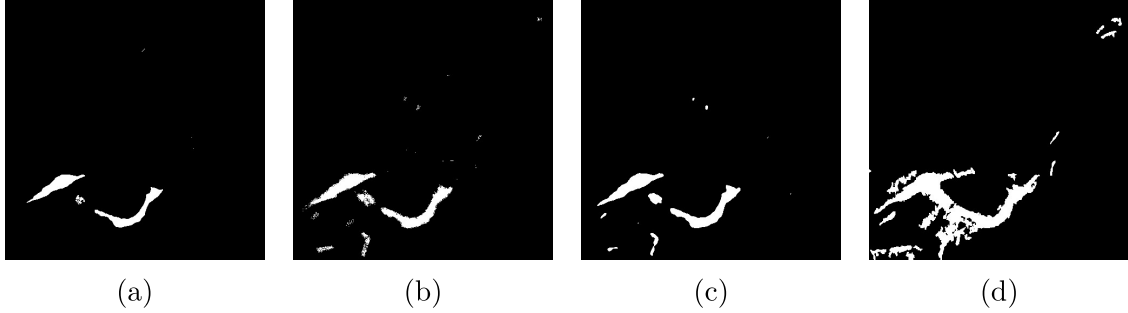


Figure 4.4: Performances comparison on U-Net model worst result (EMSR192 - 13SALE): (a) SVM (F1-Score = 0.40), (b) RF (F1-Score = 0.56), (c) U-Net (F1-Score = 0.55), (d) Ground truth

4.2 Long-lasting flood event detection in cities

In this section, we present the work submitted to the MediaEval 2019 challenge. It competed for the subtask "City-centered satellite sequences", which was part of a broader task, named "Multimedia Satellite: Emergency Response for Flooding Events" [168]. The task involved a set of sequences of Sentinel-2 images that depicted a certain city over a certain length of time. The goal was to determine whether there were flooding events ongoing in that city, whereas at least one region was flooded during all the flooding events in the considered time range.

4.2.1 Dataset

The dataset used in this work contains 335 image sequences, where each image corresponds to a Sentinel-2 L2A acquisition, which includes all the 12 spectral bands. Each sequence contains the acquisitions captured 45 days before and 45 days

after a flooding event officially stated by the Copernicus Emergency Management Service. Also, for each image there is extra metadata information that provides: (i) DATE: the satellite acquisition date, (ii) FULL-DATA-COVERAGE: whether the acquisition presented satellite measurements on all the areas of interest, and (iii) FLOODING: if in that date, there was a flooded area somewhere in the area of interest. The Ground Truth (GT) is given as a label in binary form. It is created considering the intersection of the mapped flood extend (that was not part of the available data). Therefore, if there was a flooded region that lasted during all the examined period, the GT was equal to 1. Otherwise, the GT was set to 0. The dataset was balanced: an image sequence had 50% of probabilities to present a flood.

4.2.2 Problem Statement

This problem concerns a dataset of time sequences about daily Sentinel-2 L2A acquisitions and metadata associated with each acquisition, as described in the previous section. The main objective is to determine, for each sequence, whether there is at least one flooded area that lasted during all the flooding events in the examined time range. In the positive case, the time sequence is labelled as 1, 0 otherwise. Therefore, the problem is configured as a binary classification task.

4.2.3 Methodology

We built an *expert system* which leverages on both spectral and metadata information. Water regions are segmented in both sets using the a spectral index specifically designed to be sensitive to water segmented by means of in the ac with no water regions belonging to flood events, but it and the first one contains acquisitions related to flooded areas, while the other contains the remaining acquisitions. Note that in the latter set, containing no flooded areas, and the other leverage on the notation In Figure 4.5, a diagram show the principal steps of the algorithm. Firstly, the algorithm computes a binary mask for each image in the sequence, in which white pixels represent areas with the presence of water, while black pixels represent the other regions. The binary masks are obtained: (i) by computing, for each pixel, the Modified Normalized Difference Water Index (MNDWI) [32] adapted for Sentinel-2 bands (S2), according to Equation (4.2); (ii) by setting to white the pixels having $MNDWI_{S2} \geq 0$, while others are set to black.

$$MNDWI = \frac{\rho_{green} - \rho_{swir1}}{\rho_{green} + \rho_{swir1}}, MNDWI_{S2} = \frac{B03 - B11}{B03 + B11} \quad (4.2)$$

Assuming that the dataset does not have missing values lasting for the whole time series, we set the pixels related to uncovered areas to white. Then, we performed the pixel-wise intersection among two sets of images: (i) the binary masks computed

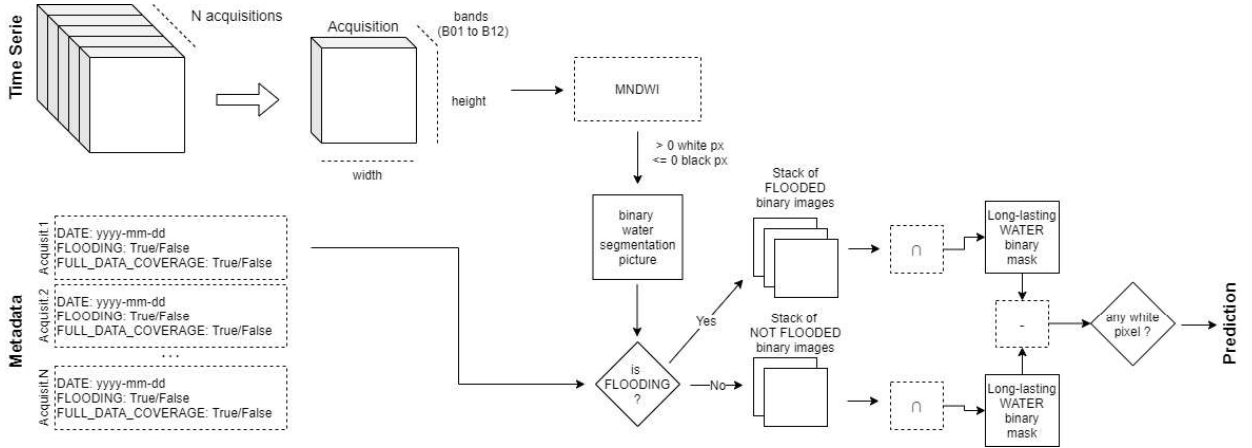


Figure 4.5: Flow-diagram of the expert system presented to the challenge.

for images marked as presenting a flooded area in the metadata file (FLOODING = True), and (ii) the ones which do not present any flooded area (FLOODING = False) in the metadata file.

The two intersections depict the water persistence in case of flood and non-flood: water bodies in the first set include both flooded areas and natural water sources, while the second set includes only natural water sources. Finally, to discriminate flooded regions from normal water sources (like rivers or lakes) a pixel-wise difference among the two sets is computed. Even if a binary mask representing the residual flood extent is available, to be compliant with the subtask, the approach returns 1 if there is still any white region in the resulting binary mask, 0 otherwise.

4.2.4 Experiments

The task organizers made the image sequences in the dataset available to the participants. The dataset was split into two sets: 80% of the image sequence belonged to the development set, while the remaining 20% belonged to the test set. The GT was made available only for the development set. Official runs for the challenge considered the performance obtained on the test set. To know the performances achieved by its approach, a participant had to send the predictions on the test set to one of the task organizers, for a maximum of 5 attempts. Being an expert system, our approach does not need to be trained on any dataset. Therefore, we assessed the performances using all the development set, before sending the official run on the test set.

Evaluation metric

In order to evaluate the approaches, task organizers used the F1-Score metric for this subtask. The metric computes the harmonic mean between Precision and

Recall for the corresponding class.

4.2.5 Results

In Table 4.6, the results for both the development and the test set are shown. Generally, our approach presented high results in the development set, but even higher (+3% F1-Score) in the official run. This score proves that our model, being an expert system, does not present typical problems encountered with the supervised approach, such as overfitting. Given the high score, the tests confirm both MNDWI as an accurate index and B03 and B11 as highly informative bands for water segmentation on optical satellite sensors. Also, our approach is able to provide a reliable delineation map of the persisting flood, localizing the permanently inundated regions for prompt intervention.

Table 4.6: Results for the subtask of "City-centered satellite sequences" of MediaEval 2019. Results refer to F1-Score metric. * Subset of the Development set, which ranges from 10% to 30% of its size.

Approach	Development set	Test set (official run)
Y. Feng et al. [47]	0.978*	0.971
P. Jain et al. [75]	1.00*	0.970
B. Bischke et al. [6]	0.926*	0.963
Our (Expert System)	0.885	0.912
S. Andreadis et al. [2]	0.835	0.866
H. Ganapathy et al. [56]	-	0.720
K. Ahmad et al. [1]	-	0.588

In the challenge, our approach achieved the highest scores among the solutions that did not require any training on data. Limitations on our algorithm concern the adopted threshold in the NDWI index, which may slightly vary from region to region. Wrong thresholds may both underestimate and overestimate water sources, and therefore bring to either false positives or false negatives.

Better scores are achieved by supervised methods, that leverage convolutional neural networks and recurrent neural networks for their predictions. For instance, the winner approach uses a DenseNet121 pre-trained on ImageNet without its last layer to extract deep features from each image in the time sequence. Then, deep features for each image are fed into an LSTM, that predicts the final outcome.

4.3 Summary

This chapter presented two works that leverage satellite data during flood events. In the first work, we assessed the performances of three machine learning models on SAR data, acquired from Sentinel-1. As for the previous chapter, the goal was to predict reliable delineation maps based only on a single acquisition, in order to speed up the whole mapping process. In order to distinguish natural water sources from flooded areas, we integrated our data with hydrography maps: in literature, other approaches make use of pre-event acquisitions, that must be manually assessed. Then, we compared the performances of Support Vector Machines (used as a baseline), Random Forest, and U-Net in three test cases: (i) using raw SAR data, (ii) using raw SAR data, preprocessed with a despeckling operation, and (iii) using despeckled data, with the addition of hydrography. SVM is the method mostly affected by the despeckling operation, which added about 9% to the average F1-Score, while hydrography slightly improved the performances. The best-averaged F1-Score achieved is 0.76. RF demonstrated to work better than the baseline using Raw data (+6% of F1-Score), but it is not subjected to significant improvements when using despeckling operation. Instead, the information added by the hydrography map is crucial for improving RF performances by 11%, achieving an F1-Score of 0.85, on average. U-Net is able to provide highly reliable predictions using just raw data from Sentinel-1, achieving an F1-Score of 0.8. That result is not particularly improved by despeckling operation, but hydrography map gives it a boost of +6% of F1-Score, on average. Moreover, it achieved the best scores in the majority of the folds (4/5). Finally, whereas CNNs often require a large number of training samples to function effectively, the U-Net model achieves good results with a relatively small number of samples.

The second work proposed an expert system able to evaluate persisting flooded areas in cities, leveraging on a time series of Sentinel-2 acquisitions. During floods, it is common to have the area of interest covered by clouds occluding the region to be assessed, therefore the time series is needed both to have more chances to spot unoccluded areas, and to monitor the evolution of flooded regions. Our proposed system, which is able to work without training, demonstrated reliable performances, achieving an F1-Score of 0.88 and 0.91 in the development and in the test sets, respectively. Compared to other solutions to this problem, our work resulted to be the best among the approaches that do not need any training process.

4.4 Relevant publications

Palomba, G., Farasin, A., Rossi, C. (2020). Sentinel-1 Flood Delineation with Supervised Machine Learning. In Proceedings of the 17th International Conference on Information Systems for Crisis Response and Management (ISCRAM 2020),

Virginia Tech, Blacksburg, VA, USA (pp. 1072-1083). [[116](#)]

Zaffaroni, M., Lopez-Fuentes, L., Farasin, A., Garza, P., Skinnemoen, H. (2019). AI-based flood event understanding and quantification using online media and satellite data. CEUR-WS: Aachen, Germany, 2670. [[168](#)]

Chapter 5

Rapid Mapping and Damage Assessment platform

This chapter introduces the Rapid Mapping and Damage Assessment (RMDA) platform, designed and developed to bring the research of chapters 3 and 4 to an operational level. In practice, the best models previously discussed were prepared to deliver accurate delineation and grading maps to any granted third party, such as civil protection, public bodies, or private companies. The platform was developed through a collaboration with LINKS Foundation - a private research centre located in Turin - within the context of two European H2020 projects: I-REACT (G.A. 700256) and SHELTER (G.A. 821282). The two projects address relevant topics of the European Commission, related to the protection of the society and its historical values against the effects of natural disasters.

The chapter is structured as follows: Section 5.1 shows the whole architecture, describing its functionalities and presenting the main modules; Section 5.2 provides proper details about the logic of each module and its inputs and outputs. Section 5.3 presents the technological stack used to develop the solution. Finally, Section 5.4 presents a preliminary analysis of the platform performances, assessing the computational time taken to accomplish the mapping activity.

5.1 The architecture: a big picture

The architecture was designed to enable any authorized client to perform a mapping request: given few details, such as the region of interest and a short time range, the platform must provide automatically high-quality delineation/grading map in a brief period of time. After the request, the platform must be able to: (i) detect the request, (ii) handle the mapping flow, (iii) look for the best available satellite data (from external services) that matches the provided constraints, (iv) provide an accurate delineation/grading map, operationalizing the models discussed

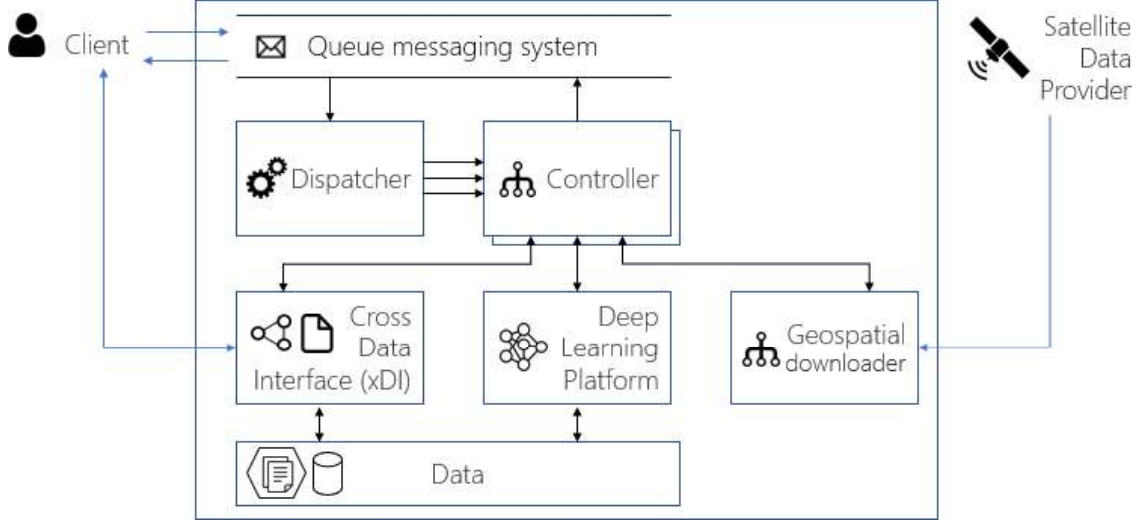


Figure 5.1: Architecture of the Rapid Mapping and Damage Assessment Platform architecture.

in the previous chapters, and (v) return the mapping output to the client.

In order to accomplish those tasks, six modules were identified: the Queue messaging system, the Dispatcher, the Controller, the Geospatial Downloader, the Deep Learning Platform and the Cross-Data Interface, as shown in Figure 5.1. Their responsibilities can be summarized as follows:

- **Queue messaging system:** handles the communications between external clients and RMDA;
- **Dispatcher:** selects from the message queue the requests to be processed and regulates their concurrence, according to the available hardware resources;
- **Controller:** handles the whole mapping task from the request to the final output. It regulates the flow of data through the RMDA modules;
- **Geospatial Downloader:** is responsible for the acquisition of satellite raw data that matches minimum quality requirements;
- **Deep Learning Platform:** is responsible for the operationalization of custom deep learning models;
- **Cross Data Interface (xDI):** is responsible for the management (load, store, and update) of the RMDA's geospatial data, which includes raw satellite acquisitions, maps, and metadata.

Every module was thought to take a specific responsibility and (except for the Controller) is designed to be independent of the other modules. Therefore, the

whole architecture can be easily modified or integrated with other platforms, resulting to be versatile for more specific applications.

In the next section, the RMDA modules will be examined deeper, focusing on their inputs/outputs and main logic.

5.2 Modules

The Rapid Damage Assessment platform allows authorized clients to submit a new mapping request to the *Queue messaging system* module. Supported requests concern either the delineation and grading of regions affected by wildfire, or the delineation of flooded areas. Then, the platform will process the request asynchronously and will inform the client through the same queue once the mapping task is finished. In the end, the client can explore the result through the *Cross Data Interface* that will be presented later on in this section.

5.2.1 Queue messaging system

The Queue messaging system is the main access for the mapping request. It collects multiple requests in real-time, allowing to process of them asynchronously, according to the system's computational capacity. It implements the Publish-Subscribe pattern [35], a common messaging architectural pattern used for delivering the same message to multiple recipients.

The advantage of using the publish-subscribe pattern is that the message sender only needs to know where to send (or publish) the message, without worrying about the recipients. The message destination is called topic: it is pre-defined in the Queue messaging system and it represents a sink that collects messages related to the same subject. All the actors interested in receiving messages from a specific topic must be registered to it. Once a message is sent to a specific topic, the Queue messaging system forwards the message to a dedicated queue for every registered recipient.

In the Queue messaging system of the RMDA platform, the following topics are set:

- map-request: topic for submitting a new mapping request. Any authorized client can send a new request that will be received and handled by the RMDA Dispatcher;
- <stakeholdername>: topic for receiving the notification of a completed mapping task. <stakeholdername> is a placeholder for a generic group or company authorized in sending map requests. For instance, researchers of Politecnico di Torino can subscribe to the topic "polito" to be notified of the completion of a mapping task.

Input

Messages sent to the "map-request" topic must contain the following information:

- Event delineation area: details the region to be analyzed. This is composite information, which includes the following geographical properties:
 - Area of Interest (AoI): rectangular-shaped region to be mapped. It is specified as a tuple of coordinates <latitude, longitude> of the upper-left and lower-right corner of the region;
 - Resolution: spatial resolution of the satellite acquisition;
 - Coordinate reference system: reference system used to map the geospatial coordinates.
- Event delineation date range: date range in which focus the analysis. Generally, a mapping request refers to a single date, for which satellite data may be unavailable for some reason, like the presence of many clouds over the requested area. Therefore, a date range is suggested to let the RMDA platform assess the best available acquisition to make the evaluation of the hazard.
- Hazard: type of hazard to be evaluated (Wildfire/Flood).
- Map type: type of map to be generated (delineation or grading for wildfires, delineation for floods).

Furthermore, other metadata concerning the sender information are forwarded with the message. This information will be used afterwards to prepare and deliver the reply to the right topic.

Output

After that the mapping task is completed, the Controller module, will prepare and send the reply to the topic retrieved from the metadata information of the request. The reply is structured as follows:

- Result: a brief report of the result of the mapping task, it describes whether it is succeeded or failed. In the latter case, it reports the error;
- Resource: link to the map and the satellite acquisition used for the mapping process.

5.2.2 Dispatcher

The Dispatcher is the module that enables the execution of multiple mapping tasks concurrently. A maximum number of allowed parallel tasks is determined by the hardware capabilities. If the number of tasks is below the limit, the Dispatcher fetches a new request from the dedicated queue of the ‘map request topic and starts a new instance of the Controller module that will process the request. The Dispatcher regulates the number of active tasks by monitoring the activity of the Controllers instantiated until their completion.

5.2.3 Controller

The Controller is the module that handles the execution of the mapping task. It is the only specific module in the whole RMDA platform architecture, being the one that coordinates the data processing and data flow through the other modules. Its main steps can be summarized as follows:

1. request raw satellite data from the Geospatial Downloader according to the parameters specified in the client’s request;
2. preprocess the satellite data, splitting the original acquisition into tiles;
3. iteratively send the tiles to the Deep Learning Platform and retrieve the mapping result;
4. rearrange the mapped tiles and build the mapping result;
5. register and store the retrieved raw satellite data and the mapping result, sending them to the Cross-Data Interface;
6. sends a message to the Queue messaging system, acknowledging the mapping result.

The Controller is started by the Dispatcher, which forwards the parameters of the client’s request. Using this information, the Controller requests the best available satellite acquisition in the specified time range to the Geospatial Downloader module (1). According to the Mapping type, the Controller requests either Sentinel-2 acquisitions for post-wildfire mappings, or Sentinel-1 acquisitions for flood mappings. In the latter case, the Controller also requests cartographic information about natural water sources in the same AoI. The water sources data represents a new layer of information that is added to the Sentinel-1 acquisition. Then, the acquisition must be preprocessed to fit the input criteria of the models operationalised by the Deep Learning platform, which in both cases have fixed resolution width = height = 480px. In case the acquisition had a lower resolution, it will be rescaled proportionally in order to fit the smallest edge to 480px. Then,

the acquisition is tiled, leaving a minimum overlap between tiles in the edges. The overlap is needed to give a smoother transition between the tiles returned as a result of the mapping task (2).

The Deep Learning Platform allows to process a set of tiles concurrently: therefore, the Controller iteratively sends a subset of the tiles to be mapped and collects the results, until the whole satellite acquisition is fully mapped (3).

Then, the mapped tiles are rearranged and merged to build the mapping result. During the merging phase, the value of each overlapping pixel is determined by weighting its original value in the corresponding tiles, according to the relative position of the pixel to the edge of the two tiles. The closer the pixel is to the edge of the tile, the lower its weight will be (4).

At this point, the Controller sends the satellite acquisition and the merged map to the Cross-Data Interface, which will store the data and will return an URL that allows access to the resources (5).

Finally, the Controller prepares the message containing the result of the mapping process and the link to the resources, then it sends the reply to the client through the Queue messaging system (6).

5.2.4 Geospatial Downloader

The Geospatial Downloader is responsible for downloading satellite data that match minimum quality criteria. It provides acquisitions from Sentinel-1 or Sentinel-2 (both L1C and L2A), Digital Elevation Maps and Water Source maps. According to the spatial constraints defined in the client's request, raw satellite data and DEM maps are acquired from the Sentinel-Hub portal [154], while water source maps are acquired from OpenStreetMap [114].

Data downloaded from Sentinel-Hub is subjected to a quality check, according to two parameters: cloud coverage and data coverage. Cloud coverage is the percentage of the acquisition that is covered by clouds (this applies only for Sentinel-2 data), while data coverage is the percentage of data captured by the satellite sensors in the acquisition.

Regions of the acquisition covered by clouds are estimated by means of the cloud test, proposed by J. Braaten et al. [13] and shown in Equation 5.1 (spectral bands reported in the equation refers to Sentinel-2 data).

$$Cloud \quad Test = \left[(B03 > 0.175) \wedge \left(\frac{B03 - B04}{B03 + B04} > 0 \right) \right] \vee (B03 > 0.39) \quad (5.1)$$

Pixels meeting the cloud test criteria are classified as clouds: their percentage pixels with respect to the total number of pixels considered in the test represents the cloud coverage.

The data coverage is dependent on the satellite orbits: it may happen that for a given place and date, the satellites did not get any data, resulting in a portion of

the acquisition without any information. Therefore, the data coverage is computed as the percentage of informative pixels in the acquisition.

Setting very strict thresholds on the two parameters may introduce the risk of not finding any suitable acquisition, especially if the time range specified in the client's request is short. On contrary, soft thresholds may accept scarce quality acquisitions, which can compromise a proper mapping. Empirically, the Geospatial Downloader sets the maximum cloud coverage to 10% and the minimum data coverage to 90%.

Data downloaded from OpenStreetMap - maps about natural water sources - is converted into a binary image (in which white pixels represent natural water) and returned as a GeoTIFF file.

Input

Parameters available in the request are the following ones:

- Area of Interest (AoI): rectangular-shaped region to be mapped. It is specified as a tuple of coordinates <latitude, longitude> of the upper-left and lower-right corner of the region;
- Resolution: spatial resolution of the satellite acquisition;
- Coordinate reference system: reference system used to map the geospatial coordinates;
- Date Range: date range used to evaluate the best acquisition;
- Data Source: data source used to download data, it can be Sentinel-2, Sentinel-1 or Digital Elevation Map;
- Product type (only for Sentinel-1 and Sentinel-2 data): related to SAR data, it determines the polarization level (IW or EW); related to the optical data, it determines the preprocessing level (L1C or L2A);
- Minimum Data Coverage (default 90%): quality parameter, used to assess the data availability in the acquisition;
- Maximum Cloud Coverage (default: 10%, only for Sentinel-2 data): quality parameter used to assess the percentage of acquisition covered by clouds;

Output

The module returns a GeoTIFF file, containing:

- Metadata: information about the area of interest, the resolution, and the time of acquisition;

- Product: a $W \times H \times D$ tensor of the acquisition, where W is the width, H is the height, and D is the depth. The Depth is related to the number of channels returned from the data source and the product. It is equal to: 13 for Sentinel-2 L1C data, 12 for Sentinel-2 L2A data, 2 for Sentinel-1 data, 1 for DEM products.

5.2.5 Deep Learning Platform

The Deep Learning Platform handles deep learning models and enables them to be triggered as a service, process the data and return the result. Moreover, the platform acts as a model repository, providing the functionality to upload any model exported in the Onnx format [113]. Its main responsibilities can be summarized as follows:

1. handle the upload of a new deep learning model;
2. operationalize a model, loading it in memory and enabling it to receive inputs and return inference predictions;
3. handle a model inference request;
4. store and return the inference results.

Those responsibilities are matched by means of internal modules, namely the Web service, the Worker module and an internal Queue messaging system, represented in Figure 5.2. The Web service handles requests from external modules, like the Controller, fulfilling responsibilities 1, 2, and 3, mentioned above. The Worker module operationalizes the deep learning models (one for each operational deep learning model), enabling them to match responsibilities 3 and 4. Finally, the internal Queue messaging system enables asynchronous delivery of the requests about mappings and it makes available the results from the Controller.

As introduced before, the upload of a new deep learning model (1) is handled by the Web service, which stores the model in file storage in the cloud and registers its existence in a structured database.

When it receives the request of making a model operational (2), the Web service creates a new queue in the Queue messaging system and triggers a new instance of the Worker. The Worker will load in memory the deep learning model and will wait for a mapping request in the queue just created.

When the Web service receives a new mapping request from the Controller (3), it stores the tiles of the geospatial data in the file storage. Then, if it is the first time that receives data from the Controller, the Web service creates a new queue, where the final result will be delivered and communicates the queue name to the Controller. Then, according to the type of mapping requested, the Web service forwards the request to the right Worker through its dedicated queue. Asynchronously,

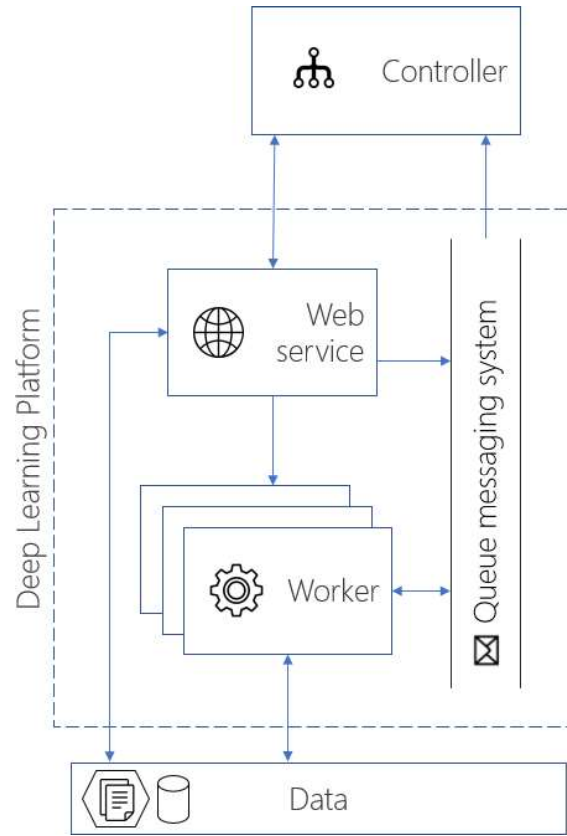


Figure 5.2: Overview of the Deep Learning Platform architecture

the Worker will fetch the request, start the model's inference, and store the final mapping in the file storage. Finally, it prepares the message with the link to the file storage and sends the reply to the queue created for the Controller (4).

5.2.6 Cross Data Interface (xDI)

The Cross-Data Interface handles heterogeneous data in order to make them available and easily accessible to the client by means of a web portal.

In the Rapid Damage Assessment platform, the xDI receives from the Controller all the information retrieved during the mapping process, such as the best available satellite acquisition, the mapping result and all the metadata related to the request. All the data are stored both in file storage and in a structured database. Then, the xDI makes available all the information through its portal. An authorized client can access the portal, read the information, and possibly add further information.

5.3 Technological stack

In this section, we briefly introduce the technologies used to implement the RMDA platform. We do not enter into configuration or implementation details, because it is not the purpose of this thesis, but we want to give a brief context on the frameworks and services used for the realization of the platform.

As shown in Figure 5.3, the global architecture has been enriched with the logos of the adopted technologies. All the components, except for the Data module, run into Docker containers (marked with the surrounding dashed blue box) [31]. Containers are isolated environments that allow to package and run applications independently from the host operative system.

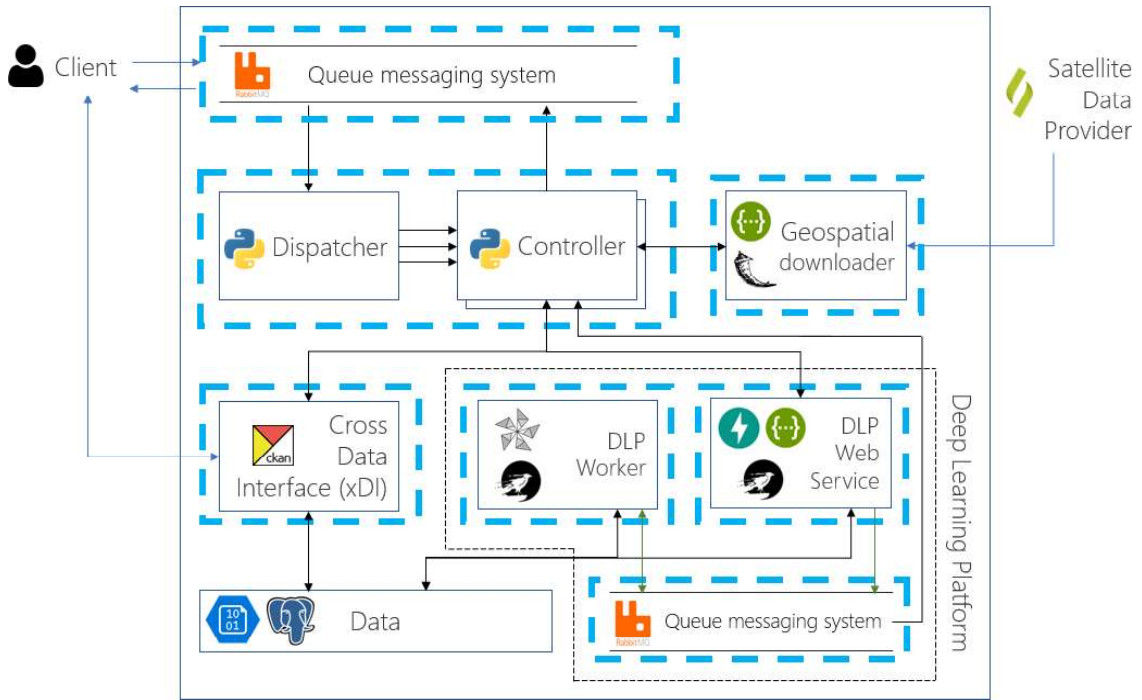





Figure 5.3: Architecture of the Rapid Mapping and Damage Assessment Platform architecture, enriched with the technologies used for the implementation. Blue dashed boxes identify modules running on Docker containers.



The Data module consists in a DBMS for storing structured data, and a File Storage for saving files, such as satellite acquisitions. In the first case we used PostgreSQL (🐉) [62], in the latter we used Azure Blob Storage (📁) [103].

The Queue Messaging System(s) and the Cross-Data Interface are based on existing services, that accomplish their functionality. In the first case, we used RabbitMQ (🐇) [120], a message-oriented middleware that supports several messaging protocols, such as AMQP, STOMP, and MQTT. In the latter, we used Comprehensive Knowledge Archive Network (CKAN) (📊) [54], an open-source open data

portal for the storage and distribution of open data.

All the other modules in the RMDA platform were implemented using Python language. The Dispatcher and the Controller modules are fully implemented in Python () and do not leverage on any particular framework.

The Geospatial Downloader and the DLP Web Service leverage on two web frameworks to handle web requests and replies. They use Flask () [55] and FastAPI () [157], respectively. Moreover, they leverage on Swagger () [147] to easily show and document, through a web page, the APIs that trigger their functionalities.

The DLP Worker uses Dramatiq () [131] to schedule and monitor background jobs used during the computation of delineation and grading maps, and Onnx runtime () [104] to manage Deep Learning models instances and inference processes on GPU. Note that Dramatiq is also used in the DLP Web Service, to properly address, through the Queue messaging system, the type of background operation to be performed by the DLP Worker.

5.4 Performance Evaluation

In this section we assess the mapping performances of the core module of the RMDA platform, the Deep Learning Platform, using the most complex of the models illustrated in the previous chapter, the Double-Step U-Net. The DLP is tested on a server with the following capabilities:

- CPU: Intel Core I9 7940x;
- RAM: 128 GB, DDR4;
- GPU: 1 x NVIDIA 1080 Ti;

The Operating System is Ubuntu 18.04, with Cuda version 10.2.

The test is performed using all the post-wildfire acquisitions in the dataset illustrated in Section 3.1.

Each acquisition was tiled in images of dimensions 480×480 pixels, simulating the behaviour of the Controller. Then, through a Python script, we provided the tiles to the DLP, through the internal Web service. In this experiment, we evaluated the execution times using batches of either 1, 2, or 4 tiles for each request sent to the DLP. Moreover, timings are recorded after three steps: (i) the request, the time needed by the DLP to receive the batch of tiles, (ii) the inference, the time spent by the model to produce the prediction for the received batch, and (iii) the response, the time spent to send back the model's predictions. In Table 5.1, the average and the standard deviations of the timings needed for each step are given.

The first aspect to be noticed is that the measured times show linear dependence with the batch size: sending a set of tiles, instead of one at a time, has the only

Step	Time (ms)					
	Batch size: 1		Batch size: 2		Batch size: 4	
	avg	std	avg	std	avg	std
Request	139	10	268	16	505	23
Inference	110	7	215	14	443	31
Response	66	3	140	3	305	10
Total	313	20	624	33	1253	64

Table 5.1: Computation times of the main steps of the Deep Learning Platform, during the operationalization of the Double-Step U-Net. Also, times are compared by varying the batch size of the analyzed satellite acquisition tiles.

effect of reducing the number of requests to be performed. For a single tile, the execution times are about 139 ms for the request, 110 ms for the inference, and 65 ms for delivering the response. In general, the mapping time for a single tile is about 313 ms, with a standard deviation of 20 ms. Considering that Sentinel-2 products have a maximum resolution of 10m per pixel (measured in the diagonal), a single tile covers a region of about 1629 km². That surface is sufficient to cover the city of Rome (1285 km²) or more than 12 times the city of Turin (130 km²).

5.5 Summary

In this chapter, we presented the Rapid Mapping and Damage Assessment platform, a tool designed and developed during the I-REACT and SHELTER European projects, able to handle the whole mapping process automatically.

It is composed of six modules that handle different aspects of the mapping process: (i) the Queue messaging system regulates the communications between clients and the platform, (ii) the Dispatcher regulates the workload to be carried by the platform, (iii) the Controller manages the mapping process by requesting satellite data, tiling the acquisition and regulating the data flow for the inference of the mapping models, (iv) the Geospatial Downloader downloads the best available satellite data for the mapping process, (v) the Deep Learning platform operationalizes pre-trained models, producing the actual delineation or grading maps, and (vi) the Cross-Data Interface, which manages the geospatial data and metadata associated to the mapping request. Furthermore, we presented the frameworks and services used to implement the modules.

Currently, the RMDA platform is able to provide delineation and grading maps for wildfire events and delineation maps for flood events. It operationalizes the best models presented in Chapters 3 and 4, namely the U-Net and the Double-Step U-Net for delineation and grading tasks, respectively.

Finally, we assessed the performances of the Deep Learning Platform on the

wildfire damage severity estimation task for a portion of a Sentinel-2 acquisition, having dimensions 480×480 pixels ($\sim 1629 \text{ km}^2$). The test achieved an average computation time of 313 ms.

Chapter 6

Knowledge extraction from Social Media during flood events

During natural disasters, situational awareness is crucial to understand the environmental characteristics, comprehend their meaning, and respond accordingly. To this end, the large volume of data provided by social media can contribute to increasing the general knowledge about the context to operate. Compared to satellite imagery, which allows having a wide and detailed perception of the extension and the severity of a hazard, social media contents, especially if geo-referenced, can provide a punctual and on-site view of the situation, from which it is possible to infer further details. Certainly, social media is not widely recognized or used as an emergency reporting tool, but there is evidence [167] of a great number of posts providing direct proof of natural hazards, which, if properly handled, might be a major assistance in dealing with the emergencies. The research community's sensitivity to natural disasters and the variety of data to deal with makes the community itself an active player on such themes, facilitating the organization of several conferences. On purpose, in this chapter, we present approaches that, leveraging on crowdsourced data from Twitter, focus on flood-related posts to assess viable roads for transporting emergency support to victims. A second work focuses on the detection of people in potential danger, through the evaluation of flood depth. Due to the data availability, this problem is tested on news articles presenting similar information content like social media posts. However, the provided solutions are compatible to be adopted with social media information.

The chapter is organized into two sections. Section 6.1 focuses on the problem of detection of passable roads during flood events. The problem, originally proposed in the "Multimedia task on Emergency Response for Flooding Events", is presented detailing the main objectives, the dataset, and the approaches proposed in the competition. Moreover, an extension of this work introduces newer approaches that aim to simplify the solution, keeping similar performances, discussed at the end of the section. Finally, Section 6.2 presents the problem of flood

depth estimation, proposed in the flood-related track of MediaEval2019, discussing the proposed solutions.

6.1 Detection of roads and estimation of their viability in flooded areas

In this section, we present several approaches for the problem proposed by the flood-related track in MediaEval2018 [89]. The objective is, given a collection of social media posts that include images related to floods, determine whether: (i) there is *Evidence of Roads (ER)* and, in positive case, (ii) there is *Evidence of Roads Passability (ERP)*. In the first task, we are concerned with determining whether a road is present (or mentioned) in the post content: this means that the road can be directly visible or that enough elements justify its presence, such as the presence of traffic lights or vertical signs, or that it might be mentioned in the message. In the second task, we determine whether the identified road is in good enough shape to be traveled. In the context of flooding, the evidence of road passability means that the road can be completely clear or partially or entirely covered by water, but cars or persons must be able to cross it.

At the competition, we provided the best approach according to the evaluation metric chosen by the organizers. However, as it will be explained, ours presents a high computational cost to be implemented in operational contexts. Therefore, after the competition we worked to simplify the approach and to develop lighter methods able to achieve similar performances [90].

6.1.1 Problem Statement

The problem is related to the analysis of social media posts related to flood events. For each post, it is required to estimate: (i) the evidence of road, and, in positive case, (ii) the evidence of road passability. Therefore, the problem presents two binary classification tasks, which can be considered either independently or somehow related. Social media posts contain heterogeneous information, such as text, discrete values, and images, which can be properly exploited for the classifications. In both tasks, posts presenting actual evidence are marked as 1, 0 otherwise. Moreover, posts not presenting evidence of roads are not considered for the evaluation of the road passability task.

6.1.2 Dataset

The dataset was distributed by the organizers of the Multimedia Satellite task [8]. It consisted of a list of 11070 tweet ids, split into two sets: a development set, containing 7387 tweet ids, and a test set, containing 3683 tweet ids. The tweets were

collected during three big Hurricane events of 2017: Harvey, Maria and Irma. They were collected on Twitter by filtering the texts of the tweets with the keywords “flooding” and “flood” during the time-frames of the three events. Moreover, the dataset was already prefiltered from duplicates, as reported in [100].

Using the tweet ids, participants were able to download the data directly from Twitter. A considerable number of tweets were no longer available by the time they were collected, resulting in a development set of 5818 tweets and a test set of 3017 tweets.

The ground truth (GT) provided for the tweets was generated manually through a crowdsourcing task. For the evidence of road subtask, the GT was set to 1 if the road presence can be deduced from the tweet information, 0 otherwise. Only for tweets presenting evidence of road, a second binary class label was attributed for the actual road passability. A road is passable when, according to the water level and the surrounding context, it is practicable by conventional means (no boats, off-road vehicles, or agricultural equipment). Therefore, in the evidence of road passability subtask, the GT is set to 1 when it is considered as passable, 0 otherwise.

In general, the dataset is significantly imbalanced towards the non-evidence of road, having only $\sim 36\%$ of the tweet content containing roads. Among the tweets labelled as containing roads, $\sim 45\%$ have evidence of positive road passability. In Table 6.1, we provide the number of tweets per class for both the development set and the test set. Moreover, for the development set, we provide the cardinality of each class: this information is not available for the test set, because its GT was not made public. Note that the development set is biased to no evidence of road (63%) and no road passability (55%).

Dataset	Total	# Evid. of Roads		# Passable Roads	
		YES	NO	YES	NO
development set	5818	2130	3688	951	1179
test set	3017	-	-	-	-

Table 6.1: Dataset composition: for each set and class, the number of tweets is shown according to the class label. Note that only tweets presenting evidence of road are labelled for the road passability task.

Tweet’s information

Tweets in the dataset contain two types of information: metadata and image. Metadata contains a set of 29 punctual data, which can be strings, dates, and discrete numbers related to the tweet, such as the text, the username of the tweet sender, and the date on which it was posted. Table 6.2 briefly lists the most relevant fields contained in tweets. Since many of them are empty or semi-empty ($\sim 90\%$), we only report the fields (16 out of 29) without missing values in the MediaEval

Field	Description	Type
Created at	UTC time when this tweet was created	datetime
Entities	Dictionary of the entities which have been parsed out of the text, such as the hashtags	object
Extended entities	Dictionary of entities extracted from the media, such as the image size	object
Favorite count	Indicates how many times the tweet has been liked	int64
Favorited	Indicates whether the tweet has been liked	bool
Id	Unique identifier of the tweet	int64
Id str	String version of the unique identifier	string
Is quote status	Indicates whether this is a quoted tweet	bool
Lang	Indicates the language of the text (machine generated)	string
Possibly sensitive	When the tweet contains a link it indicates if the content of the URL is identified as containing sensitive content	bool
Retweet count	Indicates how many times has the tweet been retweeted	int64
Retweeted	Indicates whether the tweet has been retweeted	bool
Source	Utility used to post the tweet	object
text	Text written by the user	string
Truncated	Whether the value of the text parameter was truncated	bool
User	Dictionary of information about the user who posted the tweet	object

Table 6.2: Description of the metadata information contained in Tweets.

2018 tweets.

Most of the images contained in the dataset are related to floods since the tweets have been retrieved using flood-related tags. Among the images that have been classified as without evidence of road, some of them contain charts or weather maps, some others contain information about floods that is not related to roads, whereas some other images do not contain any flood information.

Images presenting evidence of passable roads usually depict cars crossing the road or present enough surrounding contextual information that allows inferring that the water level is not very high. Instead, images classified as presenting evidence of roads, but not passable sometimes show cars stuck in roads and people crossing the street with boats. Some examples of the images contained in the dataset are given in Figure 6.1. Sometimes the differences between positive and negative road passability are very subtle and not very objective (e.g., see Figure 6.1i and Figure 6.1j), while we believe others could be wrongly classified (e.g., see Figure 6.1k and Figure 6.1l).

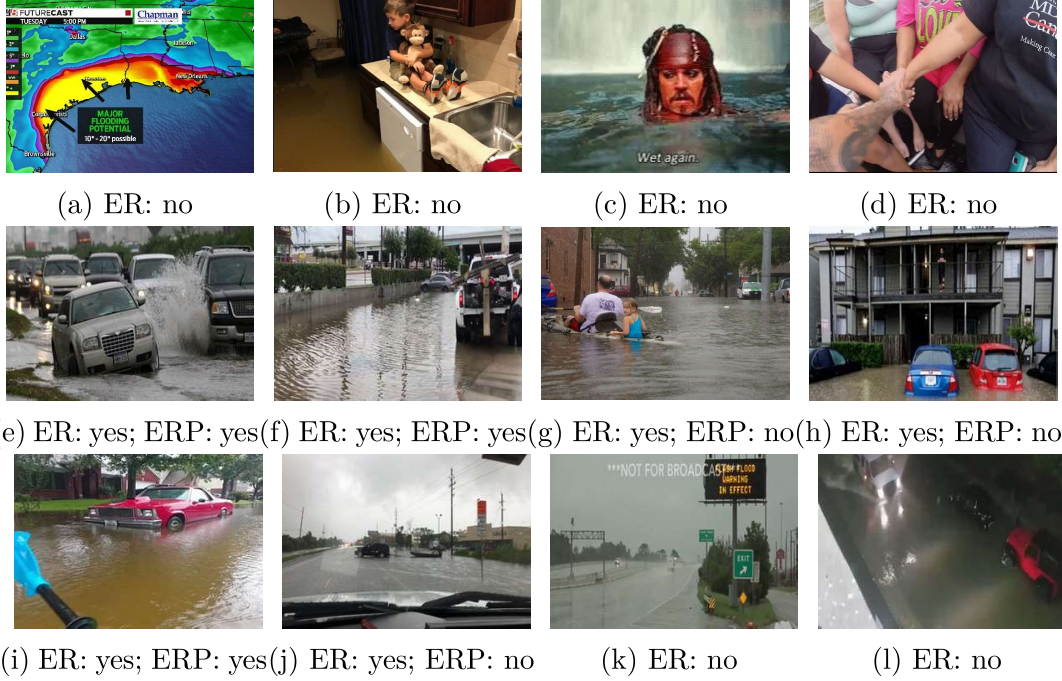


Figure 6.1: Examples of images from the dataset. The first row (a–d) contains images classified as not containing Evidence of Roads (ER), while the second row (e–h) contains images classified as presenting evidence of roads and their corresponding Evidence of Road Passability (ERP). The third row (i–l) corresponds to images that were difficult to classify or wrongly classified.

6.1.3 Methodology

In this section, we describe the approaches that we proposed to accomplish to solve the problem. First, we considered visual information and metadata separately, then we combined both the features.

Approaches based on visual information only

Considering only the visual information, therefore the tweet images, the two tasks have been reformulated as follows: (i) detecting if the image presents evidence of roads and, in positive case (ii) determining whether the road is passable or not. Another aspect to be considered is whether the two tasks are related enough to be solved jointly by the same network, or they must be treated separately. In fact, there is evidence in the literature about networks trained to solve two related tasks simultaneously, that achieve better performance on both tasks than if they were trained separately [87].

In the following, we describe the approaches that we proposed as solutions to the problem. The first one, named “Networks Ensemble” (NE), was the solution

presented to MediaEval2018 and considered the two tasks as different binary classification problems. Then, we present the “Double-ended Network”, a solution that we developed after the challenge, that considers the two tasks as related and as one-class classification problems.

Networks Ensemble. This solution aims to solve the two tasks separately, involving the following 9 state-of-the-art CNNs: DenseNet121, DenseNet201, InceptionResNetV2, InceptionV3, MobileNet, NaSNetLarge, VGG16, VGG19 and Xception. Those models were employed to build two ensemble models to solve the two tasks separately. Each network used the binary cross-entropy as a loss function, considering each task as a binary classification problem.

In order to prevent overfitting while exploiting the whole dataset, we performed cross-validation on the development set using 5 different train-validation folds. Each fold was generated using a random split of the development set into 75% train and 25% validation. Each CNN was trained in each fold and for each task separately, resulting in a total of 90 networks, 45 networks for each task (or 5 networks per network architecture and task).

To enhance their generalization capability, each network was pre-trained on ImageNet [80]: during the training, we kept the first half of each network frozen and we fine-tuned the parameters from the second half, as depicted in Figure 6.2.

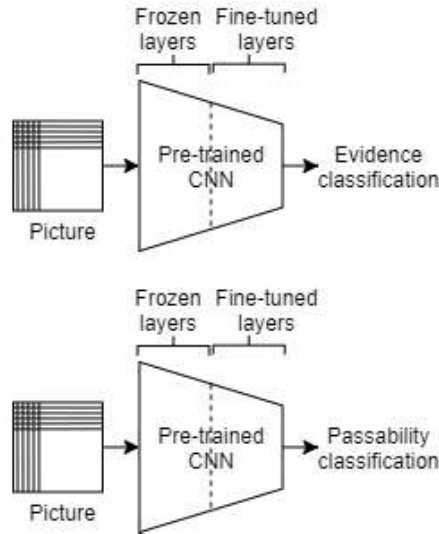


Figure 6.2: Simplified schema of a CNN trained for the ER and for the ERP tasks. In both tasks, the network is trained by keeping the first half of its layers frozen and fine-tuning the second half.

Moreover, the networks trained to solve the evidence of road passability task were trained using only the images with evidence of road, according to the ground truth. Finally, early stopping criteria were adopted to chose and store the best

model for each fold.

The output of each network is a real number between 0 and 1, which represents the probability of the picture containing: (i) evidence of road, for the first ensemble and (ii) evidence of road passability, for the second ensemble. In order to make the final prediction for each ensemble, the output of the networks is combined according to an aggregation strategy. Within this work, we proposed an aggregation strategy, which is a combination of an average aggregation on predictions with a majority vote aggregation, defined as:

$$\text{pred}(p_1, \dots, p_n, x, y) = \begin{cases} 1 & \text{if } (\bar{p} > 0.5 - x \text{ and } \text{voting}(p_1, \dots, p_n) \geq \frac{n}{2}) \text{ or} \\ & (\bar{p} > 0.5 \text{ and } \text{voting}(p_1, \dots, p_n) > \frac{n}{2} - y), \\ 0 & \text{otherwise.} \end{cases} \quad (6.1)$$

where:

- 1 and 0 are the possible outcomes for the task to be solved, that is either ER or ERP. 1 corresponds to a positive outcome (actual evidence), while 0 corresponds to the negative outcome (no evidence);
- n is the number of networks;
- p_i is the probability of the i^{th} -picture of belonging to Class 1, which corresponds to positive ER or positive ERP respectively for each task;
- \bar{p} is the average of p_i for all $1 \leq i \leq n$ and voting is given by $\text{voting}(p_1, \dots, p_n) = |\{i \mid p_i > 0.5, 1 \leq i \leq n\}|$, where $|\cdot|$ is the set cardinality;
- $x, y \in \mathbb{R}$ are tunable parameters, added to weaken the constraints.

Thresholding over the average of the predictions \bar{p} or making the majority voting are two largely adopted approaches to deal with ensemble models predictions. However, their combination through a logical “and” tends to benefit the prediction of negative outcomes (no ER, nor ERP) with the result of lowering the number of matches with the ground truth. Therefore, we added two parameters x and y to weaken the constraints, that must be assessed during the training phase.

Despite being a simple and effective model, in fact, the winning solution of the challenge, this solution requires a long training process as well as high computation cost and time during testing. Moreover, the solution is fairly heavy in terms of storing space, since we are saving the parameters trained on 90 different networks.

Double-ended network. The Networks Ensemble relies on networks trained and tested separately to solve each task individually. However, since we are using a pre-trained network and freezing half of the model, both tasks share the first

parameters of the model. Thus, we reorganized the solution as a single model, where the first part shares the parameters for both tasks and then diverges into two branches, each one with the specific parameters learned for each task, as represented in Figure 6.3.

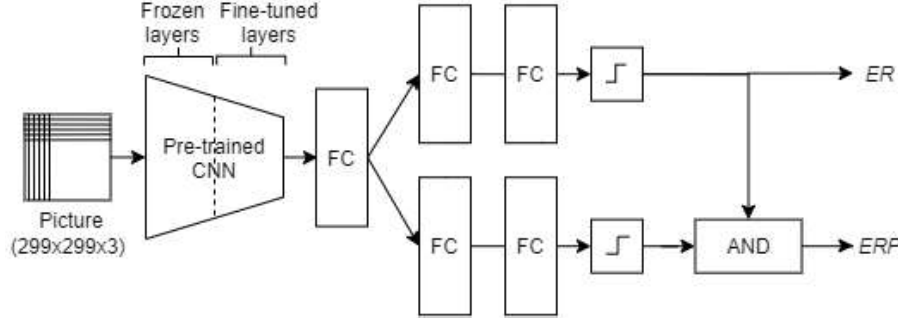


Figure 6.3: Double-ended network architecture. The first part, a pre-trained CNN, is shared between the two tasks. The first half parameters are kept frozen, while the second half is fine-tuned during training. Its last layer is replaced with a Fully Connected (FC) layer, which extracts the image features. From that, two branches start one for each task. The first one solves the Evidence of Roads task, while the other one solves the Evidence of Road Passability task. At the end of each branch, a square function thresholds the prediction. To avoid the inconsistent prediction of not having evidence of any road, but having at the same time evidence of roads passability ($ER = 0$ and $ERP = 1$).

The shared CNN was implemented with InceptionV3, pre-trained on ImageNet. According to the network definition, the input picture is resized to have dimensions of 299x299 pixels, each one consisting of 3 scalars (representing the R, G, and B components) which have been normalized, ranging from 0 to 1. Its last layer is replaced with a fully connected layer of 1024 neurons, to extract the image features. From that layer, two branches start one for each task. Each branch is composed of two fully connected layers, which output two real values in the range (0,1). They represent the percentage of belief to be in the condition of evidence of roads and evidence of passable roads, respectively. The two outputs are then rounded according to a threshold of 0.5. Therefore, the first output is the classifier for the ER class. On the other hand, to avoid inconsistent classifications (i.e. $ER = 0$ and $ERP = 1$), the second output is multiplied by the first one, determining the classifier for the ERP class. The binary cross-entropy is chosen as a loss function. Compared to the architecture presented in Figure 6.2, this solution is lighter, end-to-end and computationally less expensive since we do not run the image twice through the same layers.

Furthermore, we tried to improve the network by rethinking the problem we were solving and modifying the loss function. If the two tasks are considered as binary classification problems, the architecture proposed in Figure 6.3 can simply use the binary cross-entropy as a loss function. However, both tasks could be interpreted as one-class classification problems, where the target class is either the “evidence of road” for the ER task ($ER = 1$) or the “evidence of passable road” for the ERP task ($ERP = 1$). The advantage of considering the tasks as a one-class classification problems rather than binary classification problems consists in the following: one-class classification algorithms only consider the target class (in our case either $ER=1$ or $ERP=1$, depending on the task to be solved), without considering the other samples in the dataset. Therefore, those algorithms aim to learn the features distribution of samples in the target class. Any sample not matching the distribution of the target class is considered as “anything else”. Instead, in binary classification problems algorithms are trained to identify the features distribution of both the considered classes. With this intuition we aimed to bring the advantages of one-class classifications solutions to our binary classification problems. Therefore, taking inspiration from [119], in each branch of the Double-ended architecture we combined two different loss functions, named *Descriptiveness loss* and *Compactness loss*, described as follows:

$$\hat{g} = \max_g \mathcal{D}(g(t)) + \lambda \mathcal{C}(g(t)) \quad (6.2)$$

where:

- g is the deep feature representation for the training data t ;
- λ is a positive constant;
- \mathcal{D} is the *Descriptive loss function*, that aims to maximize the feature distance between different classes;
- \mathcal{C} is the *Compactness loss function*, it aims to maximize the “compactness among features of the target class”, providing a similar feature representation for different images of the target class (either $ER = 1$ or $ERP = 1$, depending on the task).

Both loss functions were implemented according to the ones proposed in [119]: (i) the binary cross-entropy as a Descriptive loss, and (ii) the variance of the feature representation of the samples in the target class as the Compactness loss. Given the chosen loss functions, maximizing the distance between the feature representation of different classes equals to minimizing the binary cross entropy. Similarly, maximizing “the compactness among features of the target class” equals to minimizing their variance. Therefore, the objective function \hat{g} implemented in this approach is:

$$\hat{g} = \min_g \mathcal{D}(g(t)) + \lambda \mathcal{C}(g(t)) \quad (6.3)$$

The mathematical formulation of the Compactness loss described as follows:

$$l_C = \frac{1}{nk} \sum_{i=1}^n \mathbf{z}_i^T \mathbf{z}_i \quad (6.4)$$

where $\mathbf{z}_i = \mathbf{x}_i - \mathbf{m}_i$, being $\mathbf{x}_i \in \mathbb{R}^k$ the samples of the batch of size n for all $1 \leq i \leq n$ and $\mathbf{m}_i = \frac{1}{n-1} \sum_{\substack{j=1 \\ j \neq i}}^n \mathbf{x}_j$, the mean of the remaining samples. As it is proved in [119], this compactness loss is in fact a scaled version of the sample variance given by

$$l_C = \frac{1}{nk} \sum_{i=1}^n \frac{n^2 \sigma_i^2}{(n-1)^2}, \quad (6.5)$$

where σ_i^2 is the sample variance for all $1 \leq i \leq n$.

In order to implement the backpropagation, we need to compute the gradient of l_C with respect to the input x_{ij} . In [119], the derivation of the backpropagation formula obtained from the gradient of l_C with respect to x_{ij} contained a mistake. We fixed the derivation, which is reported in the Appendix in Section A.4, while the final gradient is shown in the following equation:

$$\frac{\partial l_C}{\partial x_{ij}} = \frac{2}{(n-1)nk} \left[n \cdot (x_{ij} - m_{ij}) - \sum_{l=1}^n (x_{lj} - m_{lj}) \right]. \quad (6.6)$$

Approach based on metadata information only

As explained in Section 6.1.2, each tweet contains 29 different fields, but only 16 of them had non-empty values in at least 90% of the tweets. Therefore, the other 13 features were discarded since they do not contain enough information to give any statistical significant information. Moreover, we discarded the following features: (i) “*Created at*”, which contained the date in which the tweet was posted. Since the tweets were collected during specific hurricane events (namely, Harvey, Irma and Maria), we considered this field to have a very limited time coverage with the risk of being biased and therefore not useful. Specifically, the development set contains tweets from 38 different days. (ii) “*Extended entities*”, which contains structural information about the tweet, such as the icon and images sizes, their URLs and ids and therefore it does not provide any relevant information; (iii) “*Id*” and “*Id str*” fields are automatically generated to guarantee uniqueness to the tweet thus, not containing any meaningful information; (iv) “*Truncated*” contains a constant value, which is equal for each tweet in the development set; (v) “*Source*” and “*User*”: contained features pertinent to Twitter and the user profile, such as “id”, “profile image URL”, “friends count”, which is information not relevant to our purposes. Additionally, we verified that the development set rarely contained multiple posts

from the same user: this lack of information prevented the extraction of data for determining a possible positive (or negative) influence on our goals.

As for the “*Lang*” feature, since most of the tweets were in English and all the other languages were very minority, we transformed it into a binary value “*originally_en*” to state whether the language of the tweet in English. To make sure that all features would contribute equally to the loss, we normalized the features “*Favorite count*” and “*Retweet count*” between 0 and 1, which we named “*favorited_norm*” and “*retweeted_norm*” respectively. Finally, we also discarded the features corresponding to “*Favorited*” and “*Retweeted*” since they are subsumed by the former ones.

To determine a correlation between the normalized fields: “*favorited_norm*”, “*is_quote_status*”, “*originally_en*”, “*possibly_sensitive*” and “*retweeted_norm*” and the task at hand, we built a point-biserial correlation matrix between each feature and the “ER” and “ERP” ground truth using the Pearson correlation coefficient. As seen on the point-biserial correlation matrix from Figure 6.4, none of the features has a very strong correlation with the ground truth, however we decided to keep the fields “*favorited_norm*”, “*originally_en*” and “*retweeted_norm*” since they are the highest correlated features.

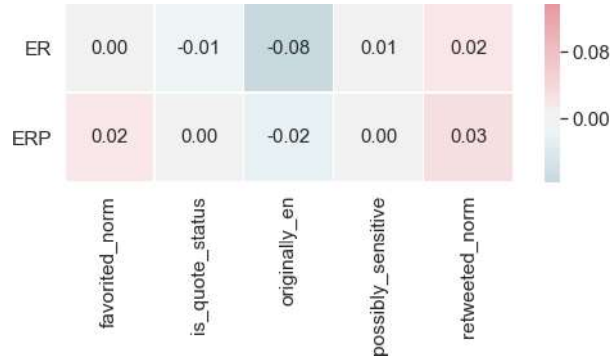


Figure 6.4: Correlations of selected metadata features with respect to class labels.

We expect the text written by the user (“*Text*”) and the hashtags of the tweet (“*Entities*”) to be the most informative features, which we concatenated, obtaining a single sentence. To help the training, we translated all the texts into English, tokenized the words, filtered stopwords (i.e. emojis, URLs, special characters, articles, conjunctions) and lemmatized the sentence. Finally, the sentences were transformed into a matrix using a word embedding initialized with GloVe [118] weights, transforming each word into a vector of 200 dimensions. To be processed by a neural network, the matrices generated from *Text* and *Entities* were standardized to have the same number of word vectors: sentences shorter than 30 words (the maximum length of a processed sentence in the dataset) were filled with zero paddings.

A representation of the architecture is shown in Figure 6.5. As other state-of-the-art works [165], the 30x200 matrices have been fed in a Bidirectional Long Short-Term Memory (BiLSTM) network. Then, the output has been concatenated with the *extra fields* and fed into two parallel fully-connected (FC) layers with a softmax classifier, one per task. In each FC layer, we used the binary cross-entropy $H(y, \bar{y})$ as a loss function, where y is the class annotation and \bar{y} is the model prediction. Denoting by $H_{ER}(y, \bar{y})$ the loss function for the ER task and $H_{ERP}(y, \bar{y})$ the loss function for the ERP task, the overall loss $H_{TOT}(y, \bar{y})$ is set to be the sum of the preceding two. Finally, the outputs from the two FC networks have been thresholded (with the threshold set to 0.5). The first FC layer output is the prediction for the *ER* task, while the second FC layer output, which represents the prediction for the *ERP* task, is combined with the first output through a logical AND operation. This operation avoids the network to predict inconsistent situations, such as having evidence of roads passability while there is no evidence of roads.

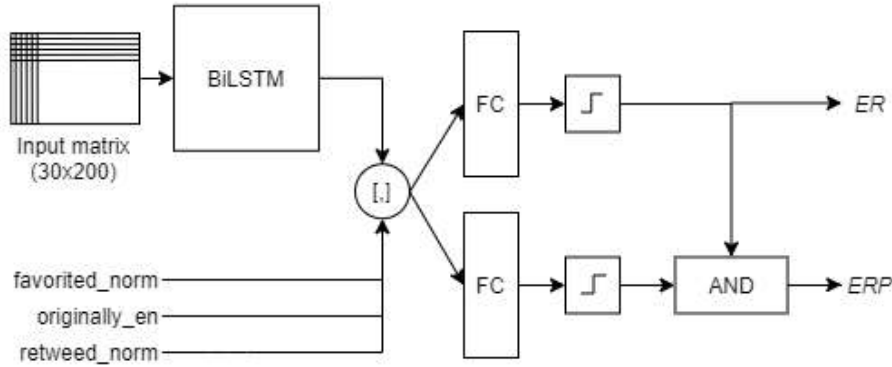


Figure 6.5: Diagram of the metadata-only approach.

6.1.4 Combining metadata and visual information

Any of the previously described solutions for the image-only architecture can be merged with the metadata-only architecture by concatenating the features collected from the bi-directional LSTM with the features extracted from the convolutional network, as shown in Figure 6.6. This approach was assessed using both loss functions proposed in the visual information the only section.

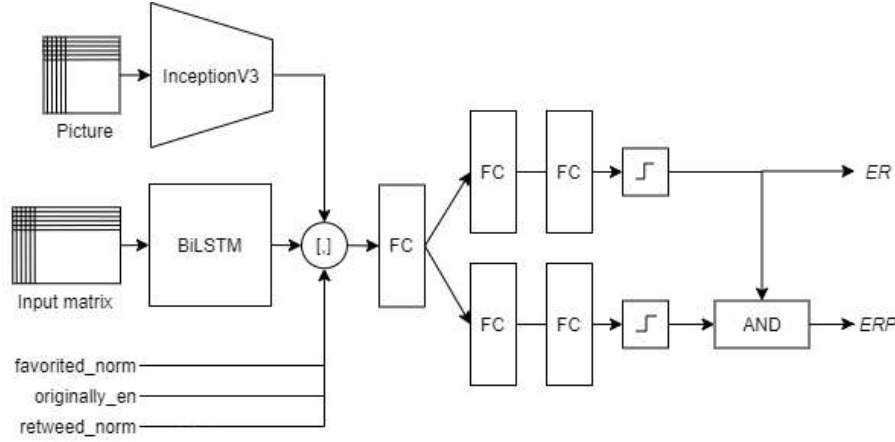


Figure 6.6: Double-ended architecture, modified to process both visual and meta-data information.

6.1.5 Evaluation and Results

In this section, we compare the results of the proposed approaches with the results of the other participants of the workshop.

However, since we did not have access to the ground truth for the test set, we used the development set to assess the performance achieved by the work done after the challenge. Therefore, we split the development set (5818 images) into training (4074 images), validation (872 images) and test (872 images) sets. The images from the new test set are solely utilized to present the final results in order to keep the setup as similar to the original challenge as possible. We defined a validation set from our training set to validate the models and tune the hyperparameters, in order to make the comparison as fair as feasible. In addition, we invited four people to solve the tasks on a subset of 50 photographs so that we could compare the outcomes. These persons were not involved in the research but were familiar with artificial intelligence and computer vision topics. They were given a verbal description of the tasks that were similar to the challenge organizers' explanation, but no examples were provided before they started annotating.

The official evaluation metric used in the challenge is the F1-Score, the harmonic mean of precision and recall. For the human annotators, we will give the results as the average of their F1-Score. It is worth noting that the second objective, classifying whether or not a road is passable, is dependent on the first. If an image is classified as not presenting evidence of road passability, it will not be evaluated for the second task. Therefore, a false negative detection in the first task (an image wrongly classified as not containing evidence of road passability) will also count as a false negative in the second task, regardless of its ground truth. At the same time, a false positive in the first task (an image wrongly classified as containing evidence of road passability) will also count as a false positive for the second task. Due to

this error propagation, the performance of the second task can not be higher than the performance of the first task.

The results of the proposed models are presented in the same order as they were in the previous section.

Results using visual information only

Firstly, we introduced the Networks Ensemble, which was presented in the MediaEval 2018. The algorithm is primarily focused on performing iterative cross-validations to train and ensemble the models, which output is used by the proposed aggregation strategy. Due to the unavailability of the test set used in the challenge, we determined a new training and test split. To ensure a fair comparison, we re-trained the Networks ensemble on the new training set, and we tested it on the new test set. In Figure 6.7 we show how the F1-Score evolves for both tasks as we ensemble more models and the difference between the different ensembling techniques. As the charts show, both tasks benefit from the ensembles, especially for lower cardinalities, while for higher cardinalities the performances stabilize. However, the Networks ensemble is more effective in the ER task than the ERP task, which presents less stable performances for higher cardinalities of the ensemble. We think this is due to different difficulties between the two tasks.

When the cardinality of the ensembles is greater than about 30 networks, the performances of both tasks start worsening slowly. That is because (i) we are adding different architectures and some of them yield better results on average than others, (ii) we have stacked the networks in order of the architectures' average performance and thus it gets a point in which adding more architectures starts degrading the results. Given the information from both graphs, the ensemble of more than 30 models (up to 90, in our test case) does not significantly improve the performance. The Networks Ensemble of 90 networks (45 per task) resulted to be the winning architecture presented in the MediaEval competition, and its performances are shown in Table 6.3.

This is the only architecture for which we have results on both the challenge and our own test sets. The results on the MediaEval test set are very close to the results obtained on our own test set, which indicates that the difficulty of both sets should be quite similar. Also, some differences might be because we had to retrain all the networks to fit them to the new training, validation and test set.

The usage of the ensemble models is acceptable for a competition, however, it might not be suitable for a real-life application, since it tends to be computationally expensive and time-consuming. Therefore, we focused our analysis to compare the best available model, obtained without taking into account computational limitations, with a lightweight version, proposed in this work. We started reducing the ensemble to the minimum set, therefore using a single CNN per task, that we named "Single CNN", and then we compared it with the Double-ended network, presented

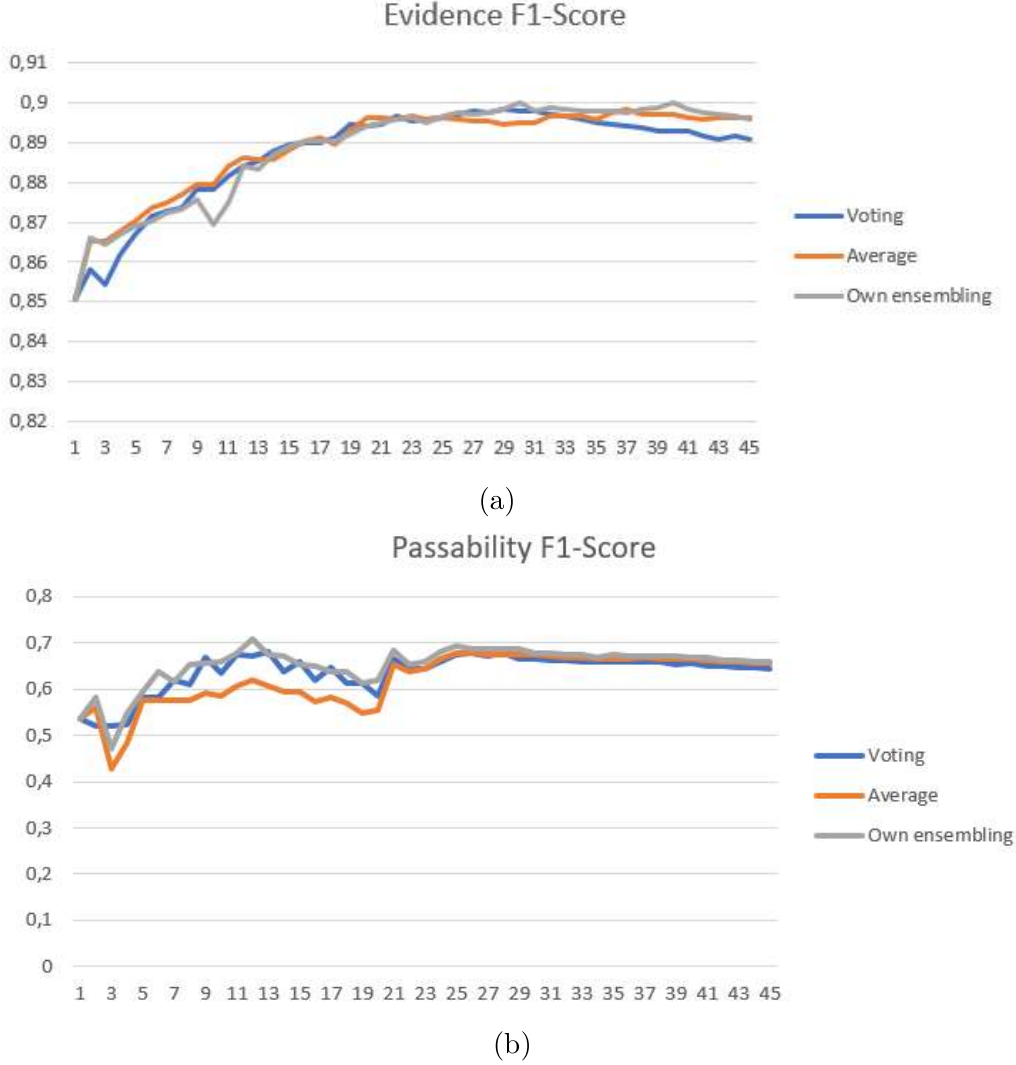


Figure 6.7: F1-Scores achieved through the variation of the number of ensemble models in the (a) evidence of road, and (b) evidence of passable road tasks. Three ensembling approaches are compared: (i) the majority voting strategy, (ii) the average voting strategy, and (iii) the aggregation strategy proposed at the beginning of Section 6.1.3.

in the previous section. The performances of the Double-ended network are firstly assessed using only the binary cross-entropy as a loss function, therefore configuring the tasks as binary classifications problems, like the Single CNN. Then, the Double-ended network is tested using compactness loss to assess the configuration of the two tasks as one-class classification problems.

Approach	EVIDENCE OF ROAD [%]			EV. OF ROAD PASSABILITY [%]		
	Validation set	Test set (MediaEval)	Test set (Own)	Validation set	Test set (MediaEval)	Test set (Own)
Human annotation	87.32*	-	-	47.71*	-	-
Networks Ensemble (90)	90.14	87.79	90.17	64.33	68.38	65.91
Networks Ensemble (30)	88.91	-	89.45	70.18	-	65.28
Single CNN	86.48	-	84.88	62.84	-	59.99
Double-ended network	88.73	-	85.00	67.51	-	67.91
Double-ended with comp. loss	87.78	-	86.42	67.49	-	68.53
Y. Feng et al. [46]	-	-	-	-	64.35	-
M. Hanif et al. [63]	-	74.58	-	-	45.04	-
Z. Zhao et al. [170]	-	87.58	-	-	63.13	-
A. Moutzidou et al. [107]	-	-	-	-	66.65	-
A. Kirchknopf et al. [79]	-	-	-	-	24	-
N. Said et al. [132]	-	-	-	-	65.03	-
D. Dias [30]	-	-	-	-	64.81	-
B. Bischke [7]	-	87.70	-	-	66.48	-

Table 6.3: F1-Scores achieved using only tweet images. Results are compared both on the official test set used in the MediaEval challenge, and on our own set. In the latter, the result of the Network ensemble of 90 models is used as reference. *Results on a subset of 50 images.

To reduce the randomness associated with the training process, the three architectures have been initialized with the same weights and used the same hyperparameters and stopping criterion. The improvement brought by the Double-ended network is significant especially in the passability task, as shown in Table 6.3. We think that, when the passability and evidence tasks are trained separately, the passability task has significantly fewer images to train, making it more difficult for the model to generalize to new data, whereas when they are trained together, the passability task can benefit from what the evidence task has learned. The Double-ended network has fewer parameters than two Single CNNs (required to accomplish both tasks), making it lighter and less computationally expensive, still being an end-to-end architecture.

Although there is no direct evidence that the compact loss improves the results, the outcomes of the Double-ended network with compactness loss appear to generalize to the test set, since the results from validation and test sets are more similar than the ones without compactness loss.

Overall, the Double-ended network achieves about the same performance as the ensemble of 30 models, implying that the technique is nearly 30 times faster and lighter while maintaining similar performance.

Results using metadata only

In Table 6.4 we collect the results of the model using metadata information only. As the table shows, even in the case of human annotators, performance is generally poor. We think that not much metadata contain relevant information about the

two tasks.

Approach	EVIDENCE OF ROAD [%]		EV. OF ROAD PASSABILITY [%]	
	Validation set	Test set	Validation set	Test set
Human annotation	51.48*	-	18.18*	-
Metadata only	59.93	62.56**	56.82	57.05**
Y. Feng et al. [46]	-	-	-	32.8
M. Hanif et al. [63]	-	58.30	-	31.15
Z. Zhao et al. [170]	-	32.60	-	12.86
A. Moutzidou et al. [107]	-	-	-	30.17
A. Kirchknopf et al. [79]	-	-	-	20

Table 6.4: F1-Scores achieved using only metadata. *Results on a subset of 50 images. **Results given on our own test set.

Remarkably, the results using images are considerably better than the ones using metadata not only in our case but also for humans or other participants. In the ER task, the Double-ended network with compactness loss on visual information gains about 24% of F1-Score if compared to the metadata only approach and more than 35% of F1-Score if compared to the human test. It must be considered that the dataset was created and annotated only based on visual data: therefore, we are confident that images provide sufficient information to solve the problem, but we are not certain that metadata could provide such distinctive information.

Results using both visual information and metadata

As a final step, we combined the best Double-ended model on visual information with the metadata approach. Since from previous results was not clear the performance improvement brought by the compactness loss, we firstly tested the hybrid model performance using only the binary cross-entropy as loss function. Then, in a second test we used the compactness loss, as shown in Table 6.5.

Approach	EVIDENCE OF ROAD [%]		EV. OF ROAD PASSAB. [%]	
	Validation set	Test set	Validation set	Test set
Double-ended architecture	78.96	86.99*	61.06	62.96*
Double-ended with comp. loss	77.85	84.56*	73.61	75.93*
Y. Feng et al. [46]	-	-	-	59.49
M. Hanif et al. [63]	-	76.61	-	45.56
Z. Zhao et al. [170]	-	87.58	-	63.88
A. Moutzidou et al. [107]	-	-	-	66.43
A. Kirchknopf et al. [79]	-	-	-	35

Table 6.5: F1-Scores achieved using both image and metadata. *Results given on our own test set.

By using both visual and metadata information we can notice a considerable improvement in the model with compactness loss relative to the one without it. In

fact, even if the model achieves 3% below the best score in the evidence of road task, it obtains almost a 10% improvement in the evidence of road passability task compared to the second-best participant. Finally, it seems like adding the metadata information improves the road passability task. For the sake of readability, in Table 6.6 we report the results of the Double-ended classifiers and the Metadata approach in the three case studies: visual information (V), metadata information (M), visual and metadata information (VM).

	Approach	EVIDENCE OF ROAD [%]		EV. OF ROAD PASSAB. [%]	
		Validation set	Test set	Validation set	Test set
V	Double-ended network	88.73	85.00	67.51	67.91
	Double-ended (comp. loss)	87.78	86.42	67.49	68.53
M	Metadata approach	59.93	65.56	56.82	57.05
VM	Double-ended network	78.96	86.99	61.06	62.96
	Double-ended (comp. loss)	77.85	84.56	73.61	75.93

Table 6.6: Summary of the results achieved by the proposed Double-ended network approach for the three test cases: visual information only (V), metadata information only (M), visual and metadata information (VM). The result is reported with the F1-Score metric.

To understand how the metadata information can help to improve the results, in Figure 6.8 are gathered some tweet examples which were incorrectly classified by the image only model but correctly classified by the model which combined visual and metadata information. These tweets contain some very informative keywords such as: "flooded street", "stalled cars" and "drive-through".

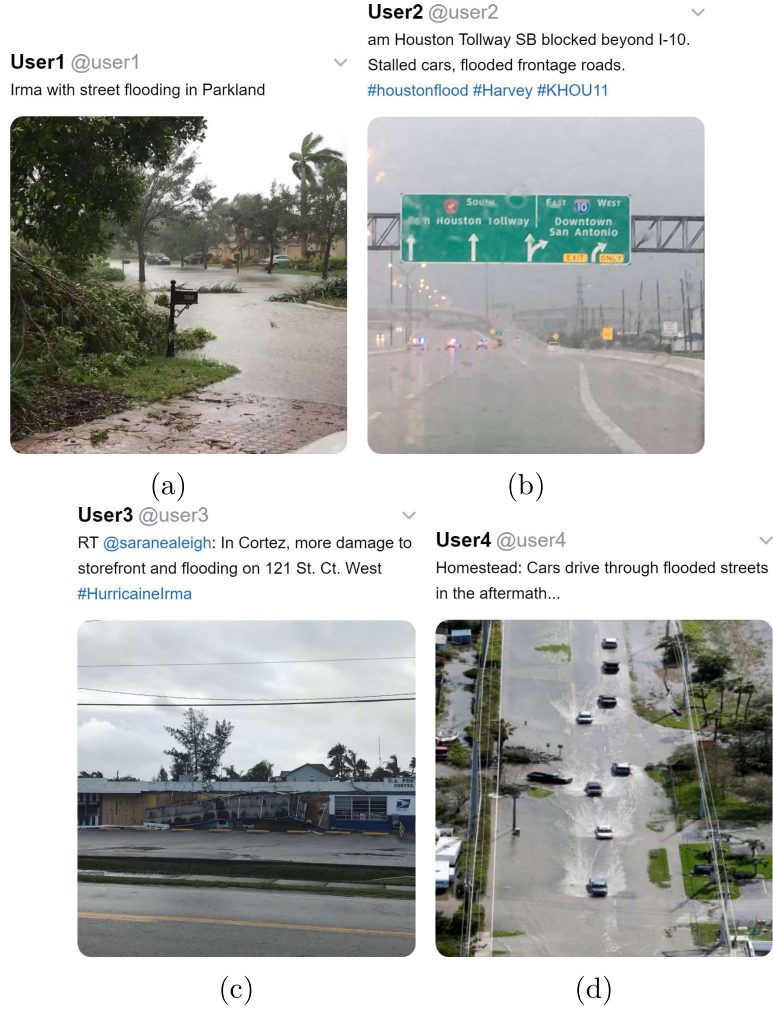


Figure 6.8: Tweets containing informative text that helped the classifier to disambiguate the visual content for the correct prediction of the evidence of road passability task.

6.2 Flood depth estimation

In this section we present the work submitted to the MediaEval 2019 challenge, competing for the track "Multimodal Flood Level Estimation from News" [168]. The track involved news articles about flood events, which included pictures depicting at least one person. The goal is to predict whether in the article there is evidence of people standing in water above knee level. The difficulty of this task concern the evaluation of a measurable quantity, i.e. the flood depth, with relative references, the knee height. That reference is highly variable because it depends on many factors, like age and gender. However, the challenge did not provide any detail about those aspects, even it is supposed that for a rigorous evaluation, the typical

knee height might be referred to an adult of an average height. In the scope of the challenge, critical water levels occur when the flood depth is greater than the knee height: the detection of those situations is essential to support emergency response in identifying people at risk and intensive floods, that could harm buildings and streets.

6.2.1 Problem Statement

The problem is related to the analysis of news articles related to flood events. Each news article contains both text and image, the latter depicting at least one person. The goal is to estimate whether the water level reached dangerous depth, stated as the knee height of a standing person. Therefore, the problem is configured as a binary classification task: news articles presenting dangerous water level are labelled with 1, 0 otherwise.

6.2.2 Dataset

The dataset consists of 6166 articles from local newspapers from African countries, composed of a textual part and an image. All the articles present words like "flood", "floods" or "flooding" in the text, and contain an image depicting at least one person. The subsets have been identified for the challenge: the development and the test sets, containing 4932 and 1234 articles, respectively.

The ground truth was manually annotated that considered only the image content. They considered people in general, without any distinction between adults or children. Each image is labelled as 1 if there is at least one person in the image standing in water, and the water level is above the knee height, 0 otherwise.

Knees dataset generation

Given the specificity of the problem, we decided to extend the actual dataset with extra information generated from the images. In particular, for each image, we wanted to extract a sub-image, or crop, of each visible knee, in order to establish if it was either above or under the water level. We automated the crop-generation process (i) by using a multi-person pose estimation algorithm [18] to detect knees, and (ii) by extracting a tile of dimensions 48×48 pixels around each identified knee, as depicted in Figure 6.9.

The ground truth was manually annotated by humans for the knee-crops generated by full-sized images labelled as 1 (water level above the knee height). In fact, in this case, there might be humans standing in water above their knees. Therefore, crops depicting a knee under the water were labelled as 1, 0 otherwise. In the other case, when the full-sized images were labelled as 0 (water level under the knee height), all the knee-crops generated were automatically labelled as 0.

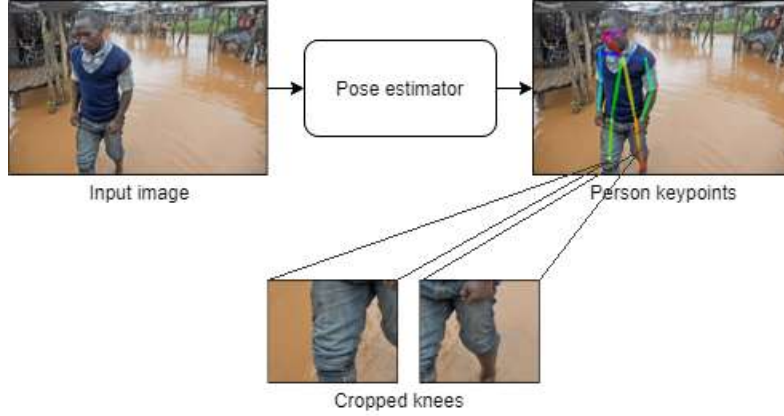


Figure 6.9: Diagram on the generation process of image crops depicting knees.

It is important to note that even if a person’s knee is not visible in the image because it is below the water, most times the pose estimator will estimate where the legs of the person should be based on the estimation of the body located above the water, see Figure 6.10. However, if there is an upper body detected with no lower body pose estimation, that is also relevant information for the algorithm since chances are, the lower part of the body is not being detected because it is below the water.

6.2.3 Methodology

In the challenge, we developed three models, based on: visual-only, textual-only, both textual and visual information.

Approach based on visual information only

Our major contribution is related to the first model, represented in Figure 6.11, which combines semantic information of both local and global aspects of the image. Local aspects are evaluated by the upper branch, which takes image crops of people’s knees as input. Global aspects are evaluated by the lower branch, which takes as input the full image of the scene, and predicts if the image presents water level above the knee. Before being processed, knee crops and the full-size image are resized to 224×224 pixels, and each channel is linearly rescaled in the range $[-1, 1]$. Both inputs are fed into a VGG19 [144] pre-trained on ImageNet [28], that extracts deep features of the images, followed by a fully-connected layer (FC). Then, the output of the two branches is concatenated to combine the semantic features of the knee with the context information provided by the full resolution image. Then, the model ends with two branches, each one composed of an FC layer, to deliver the output.



Figure 6.10: Example of the output of the pose estimator algorithm. In this image, even though the legs of the person are not visible because they are below the water, the pose estimator algorithm makes an estimation of where they should be.

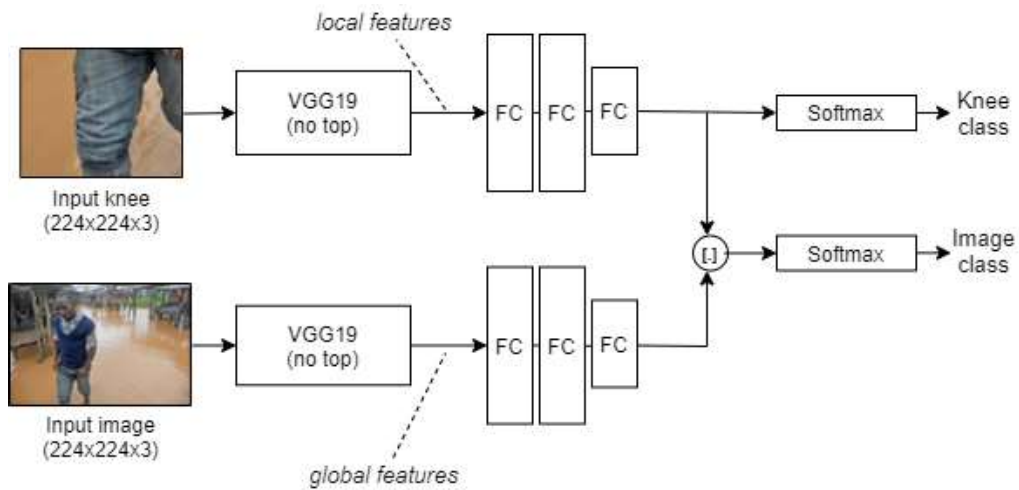


Figure 6.11: Double branched model to estimate the depth of the water by determining if the water is above or below the knee. The upper branch of the model gets as input knee crops while the lower part gets the full image.

The upper branch learns to improve its prediction thanks to the focus on the knee, while the lower branch learns how to make a better prediction on the knee

thanks to the context. Training the CNN without any distinction between global and local data would lead the network to predict flooded images as a positive class because it lacks specific data about the knees in the scene and so associates the features of a flooded area as a positive class because it solely composed by these examples. The final classification for an image in the dataset concern the evaluation of that image with each crop of knee present in the scene. If the classifications of both branches are positive in at least one couple $\langle \text{crop of knee, image} \rangle$, the prediction for that image is that the flood depth is greater than the knee height, 0 otherwise.

Single and mixed approaches on textual information

Textual data were processed similarly to the metadata approach presented in Section 6.1.3. Compared to tweet messages, newspapers articles present more articulated contents, with a well-defined structure, organized in paragraphs. Usually, titles and subtitles summarize the content of the articles, focusing on the main concepts in short sentences. In our approach, we considered articles title and subtitles as a source of information for textual data. In case the subtitle was absent, we considered the first paragraph of the article. Both texts are merged in a unique string, then they are processed as follows: (i) the punctuation was removed and each word was tokenized, (ii) stopwords were filtered, i.e. conjunctions, articles, and (iii) the remaining words were lemmatized, i.e. bringing nouns to the normal form, verbs in their infinitive form. The preprocessed text was transformed into a matrix, using a word embedding initialized with GloVe. Then, each word in the sentence was transformed into a vector of 200 dimensions, while the length of each sentence was standardized to 100 words, filling with zero paddings shorter sentences. The matrix represents the input data for a bidirectional Long Short Term Memory (BiLSTM) network. In the textual-only approach, the BiLSTM output is processed by an FC layer, which uses the Softmax activation function to return the prediction. In the mixed approach, the BiLSTM output is concatenated to the last FC layer of the image classifier.

6.2.4 Evaluation and Results

In the challenge, official performances were evaluated using the F1-Score on the test set predictions. We divided the development set into training and validation sets with an 80-20 split. The training set was used to actually train the models, while the validation set was used to assess the performances of the model on unseen data. During the training process, we used early stopping as a regularization approach to avoid model overfitting on training data. Also, dropout regularization with a probability of 50% was applied on every FC layer in the proposed architectures. Binary cross-entropy was used as a loss function in all the approaches.

The results on both validation set and test set are shown in Table 6.7. Considering the approach on visual information, we performed an ablation study to assess the improvements brought by using both local and global information together, instead of using just one of them. In the validation set, combining both information makes the approach achieving 79% of F1-Score, while the performances degrade to 3% when using only local information, and by 8% when using only global information.

Table 6.7: Results of the challenge, evaluated on validation and test sets. For the visual approach, performances are also evaluated for the upper and lower branches, separately. Results refer to F1-Score metric.

Approach	Validation set	Test set (official run)
Visual (upper branch)	0.76	-
Visual (lower branch)	0.71	-
Visual	0.79	0.54
Textual	0.52	0.50
Visual + Textual	0.54	0.53

Generally, the visual approach resulted to perform better than the textual-only and the mixed approaches in both validation and test sets. However, performances of the visual approach in the test set were quite low, if compared to ones on the validation set. A reason might be related to the frequent presence in the test set of images depicting people near water sources: their reflection on the water source is identified by the pose estimator as a separate person, which makes it generate false knee crops. Depending on the angle of reflection, the mirrored image could omit body parts, leading the visual approach to make prediction errors.

The winner approach in the challenge was presented by C.Q. Kahan-An et al. [123] and was related to visual information only. The approach shares the same idea of using both global and local information, but it is more complex. Besides the information about knees, it analyzes a variety of conditions related to the human body, such as the hip position related to the water, the ratio between thighs and the upper body. Moreover, it detects other aspects, like the presence of "swimming" people in the scene and it segments water bodies. Leveraging on the combination of Faster R-CNN, a custom CNN called "WaterClassifier", ResNet-50, OpenPose, and Mask R-CNN, the approach achieved an F1-Score of 0.88 in the test set.

6.3 Summary

This chapter presented approaches able to extract valuable information for the emergency response phase during flood events. In particular, we assessed solutions for the problems of (i) evidence of roads detection and evidence of roads passability detection from social media posts, and (ii) the detection of people in potential danger, through the estimation of flood depth from news articles.

The first problem was initially solved for the MediaEval2018 Flood classification challenge, which considered only images and dealt with the ER and ERP task separately, through a network ensemble of 90 CNNs (45 per task). This approach was the best solution proposed to the challenge, achieving F1-Scores of 87.79 % and 68.38 % for the ER and ERP tasks, respectively. The solution achieved the best score, but it is computationally expensive to be used in real applications. Therefore, we extended that work in order to simplify the solution and make it usable during emergency operations. Due to the unavailability of the test set used in the challenge, we created a new one, to be used as a reference for further experiments. In this dataset, the Networks Ensemble of 90 CNNs achieved similar F1-Scores: 90.17 % and 65.91 % for the EP and ERP tasks, respectively.

The first simplification consisted of the assessment of the adequate number of networks in the ensemble able to achieve comparable results to the winning approach. As a result, 30 CNNs (15 per task) were sufficient to achieve F1-Scores of 89.45 % and 65.28 % in the two tasks.

Then, we developed a new architecture, the Double-ended network, based on a single CNN with two endings, one per task. The peculiarity of this approach consists in considering the two subtasks as related. Within this assumption, we believed that training part of the network on both the problems simultaneously will benefit both the predictions, other than reducing the size of the network. This architecture was adapted and tested on three contexts: (i) using visual information only, (ii) using metadata only, and (iii) using both visual and metadata information. Considering the network complexity and the results, the best model was obtained by using visual information only, which achieved F1-Scores of 85.00 % and 67.91 % for the ER and ERP tasks, respectively.

Finally, we tested the double-ended architecture by using a different loss function, the compactness loss, which allows to solve each task as a one-class classification task, instead of as a binary classification task. When using this loss, we spot an error of its derivation in the paper in which this loss was introduced [119] and we corrected it. We tested the new loss in the contexts in which the double-ended architecture worked best: using visual information only and using both visual and metadata information. While for the ER task there was not a significant improvement, for the ERP task the combination of both visual information and metadata brought to a significant improvement of F1-Score, where the approach achieved 75.93 %: far better than the winner approach of the MediaEval2018 challenge: the

ensemble of 90 CNNs.

Finally, leveraging on double-ended architecture with the compactness loss we have a single network that achieves slightly lower results in the evidence of road task, but higher results for the evidence road passability task than the Networks ensemble. Moreover, it is almost 90 times faster, lighter and end-to-end, making it a viable solution for real-world applications.

The second problem was proposed for the MediaEval2019 Flood classification challenge, which considered text and image data from news articles about flood events. The aim was to detect articles presenting evidence of people standing in water above knee level. We proposed three approaches, which leveraged on (i) visual information only, (ii) textual information only, and (iii) both visual and textual information. The approach that achieved the highest results were based on visual information only. It consisted of the combination of both local information about knees present in the image and global information about the whole picture. We demonstrated that using both local and global information together improves the results instead of using them separately: in the validation set the combined approach achieved an F1-Score of 0.79, compared to 0.76 and 0.71 for the approaches exploiting local information only and global information only, respectively. In the test set, the visual approach lowered the performance, obtained in the validation set, achieving an F1-Score of 0.54.

6.4 Relevant publications

Lopez-Fuentes, L., Farasin, A., Skinnemoen, H., Garza, P. (2018, October). Deep Learning Models for Passability Detection of Flooded Roads. CEUR-WS: Aachen, Germany, 2283. [89]

Zaffaroni, M., Lopez-Fuentes, L., Farasin, A., Garza, P., Skinnemoen, H. (2019). AI-based flood event understanding and quantification using online media and satellite data. CEUR-WS: Aachen, Germany, 2670. [168]

Lopez-Fuentes, L., Farasin, A., Zaffaroni, M., Skinnemoen, H., Garza, P. (2020). Deep Learning Models for Road Passability Detection during Flood Events Using Social Media Data. Applied Sciences, 10(24), 8783. [90]

Chapter 7

Conclusions

The fight against natural hazards still represents a big challenge that involves every country in the world. In Europe, the European Commission and its Member States are actively supporting and coordinating national actions and promoting cross-border cooperation. Through funding programs like H2020, the European Commission promotes research and innovation activities among universities, research centres and industries, promoting the development of projects that can improve and support emergency management operations.

Part of the works presented in this thesis represent my contributions to I-REACT and SHELTER projects, developed in collaboration with LINKS Foundation, a private research centre based in Turin. Research activities were focused on supporting the response and the recovery phases of the emergency management cycle for wildfire and flood events, leveraging satellite data and social media data.

Within satellite data, research activities were addressed to propose methods that could leverage limited data, in order to boost the mapping process and limit the human intervention. In this regard, we adapted machine learning approaches and we assessed their performances in the delineation and damage severity estimation tasks. We provided novel approaches, able to solve the same tasks with higher performances and using less information than the ones currently adopted in the literature.

In Chapter 3, we focused on burned areas caused by wildfire events, assessing the feasibility of delineation tasks using (i) visible wavelengths, and (ii) the whole spectrum. In the first case, we assessed supervised approaches and we proposed BAE, a novel unsupervised approach. Then, we considered the whole spectral data, assessing the best approach to accomplish the task. The U-Net, which provided the highest results in the delineation task, was used as a baseline to estimate the damage severity on affected areas. Major improvements were brought by the novel approach Double-step U-Net, which resulted in more accuracy than our baseline and provided better results than the standard approach, but just half of the information.

In Chapter 4 we dealt with flood events. Firstly, we leveraged SAR data, assessing

the best approach among machine learning models for the delineation task. We studied their performances variation in three test cases which concerned raw data, preprocessing, and hydrography maps used as extra data. In a second work, we leveraged on time series of Sentinel-2 acquisitions, in order to spot persisting flooded areas. We provided a novel approach, able to determine the existence of those areas and to segment them, through delineation maps.

In Chapter 5, we developed a platform able to operationalize the approaches for both delineation and damage severity estimation tasks, providing an end-to-end mapping service that can be actively used by external stakeholders.

Within social media data, research activities aimed to provide solutions that could extract valuable information for the coordination of emergency operations during ongoing flood events. In Chapter 6, we dealt with heterogeneous information which included textual and visual data, providing novel approaches for (i) the detection of people potentially in danger, through the evaluation of flooded sources depth, and (ii) the detection of flooded roads that could be still viable, useful for transporting emergency support to victims. In particular, in the latter problem, we took part to a challenge providing the best solution. Then, we worked to simplify the winning architecture, providing a lighter approach that achieved similar performances and could be operationalized with a limited amount of resources.

General reproducibility

The source code of the algorithms and the datasets mentioned in this thesis are not publicly available, as subjected to the regulation of the I-REACT and Shelter EU H2020 research projects.

Future works

Future works will certainly expand the applications of the proposed approaches to other natural hazards, such as earthquakes, where satellite data can help to delineate and evaluating the severity of damaged areas, while social media or in-situ sensors can provide crucial information in detecting damaged buildings or structures like hospitals, schools, bridges and roads in time.

Through the combination of Sentinel-1 and Sentinel-2 data, newer approaches will be studied to reconstruct missing optical information due to occlusions like clouds, with the aim to increase the amount of affordable data for mappings.

Finally, through the analysis of time series of satellite acquisitions combined with extra data, such as meteorological measurements and Digital Elevation maps, newer approaches will be studied for predicting the evolution of wildfires and floods in order to produce risk maps and organize a prompt intervention.

Appendix A

A.1 Dataset

Legend

- *ISO* stands for ISO-3166 Country Code (<https://www.iso.org/obp/ui/#search&3166>);
- *EMSR* stands for Copernicus Emergency Management Service (EMS) - Rapid Mapping (R) Activation Code (<https://emergency.copernicus.eu/mapping>);
- *BB_TL_LON* and *BB_TL_LAT* is the couple of coordinates (LONgitude, LATitude) of the Bounding Box (BB) for the Top Left (TL) corner;
- *BB_BR_LON* and *BB_BR_LAT* is the couple of coordinates (LONgitude, LATitude) of the Bounding Box (BB) for the Bottom Right (BR) corner;
- *PRE Date* and *POST Date* stand for Pre Fire acquisition Date, and Post fire acquisition Date, respectively. They describe the date on which the Sentinel-2 satellite registered data for the specified bounding box, which complies with the availability and cloud coverage criteria outlined in Section 3.1.
- *FOLD* indicates in which fold the product belongs, according to the colors represented in Figure 3.1.

ISO	EMSR	BB_TL_LON	BB_TL_LAT	BB_BR_LON	BB_BR_LAT	PRE Date	POST Date	FOLD
FR	221_01	9.300647504	42.886608180	9.505183140	42.763508590	06-07-2017	15-08-2017	blue
	371_01	9.644467300	39.921109700	9.687531600	39.856982400	18-07-2018	18-07-2019	
SE	290_03	16.260418860	59.853342900	16.321359390	59.828218650	20-05-2018	09-06-2018	brown
	298_02	16.361547000	63.140440500	16.447381400	63.099673880	26-06-2018	24-07-2018	
	298_06	15.357103890	62.915099290	15.574015210	62.833943810	27-06-2018	24-07-2018	
ES	248_01	-6.196452795	41.659029510	-6.009019170	41.544015950	14-07-2017	08-08-2017	fucsia
	248_03	-5.635075322	40.498431780	-5.506317882	40.418039390	02-04-2017	04-09-2017	
	248_04	-5.095913397	40.401377180	-5.017491612	40.352353570	01-07-2017	20-08-2017	
	248_05	-4.999988815	40.416147790	-4.903491087	40.355830440	15-08-2017	04-09-2017	
	368_01	-5.078459714	40.337887420	-4.823074137	40.208122740	18-05-2018	01-07-2019	
ES	216_01	-2.372395304	38.474711270	-2.314204884	38.425632380	28-06-2017	04-08-2017	green
	216_02	-2.314223218	38.474528270	-2.256145364	38.425540850	03-07-2017	04-08-2017	
	216_04	-2.314236645	38.425632380	-2.256299384	38.376741980	03-07-2017	04-08-2017	
	216_05	-2.430549330	38.455364920	-2.372358909	38.406286030	03-07-2017	04-08-2017	
ES	365_01	0.412424949	41.454152670	0.770854806	41.165836340	31-05-2019	30-06-2019	orange
	373_01	-0.630882068	41.824559980	-0.508149384	41.754863330	18-07-2018	25-07-2019	
ES	302_01	-6.673627731	37.809332970	-6.460693904	37.671156100	16-07-2018	05-08-2018	red
	302_06	-6.603647894	37.733337620	-6.530111345	37.685598470	16-07-2018	05-08-2018	
	302_07	-6.530284692	37.733470790	-6.456643427	37.685736730	16-07-2018	05-08-2018	
PT	250_01	-9.163435178	40.036735710	-8.705635612	39.636755460	27-09-2017	17-10-2017	yellow
	372_04	-8.192177749	39.796433080	-7.948217694	39.609595860	13-08-2018	24-07-2019	

Table A.1: Areas of Interest (AoIs) considered in this work. Each AoI reports information about the Country (ISO code), the grading map identifier for Copernicus EMS (EMSR), the coordinates of the AoI's top-left and bottom-right corners, the Pre-fire (PRE Date) and Post-fire (POST Date) Sentinel-2 acquisition dates, and the related fold.

A.2 Dataset Land Use

Dataset Land use - Legend: Land use details for the areas of interest considered in the dataset Tables A.2 and A.3. They are specified in the grading maps cartouche, using the hectare (ha) as the unit of measurement. The land use types are specified as follows (Land use types used in this work refer to the official Copernicus EMS notation, available at <https://emergency.copernicus.eu/mapping/ems/domains>):

- Residential/Industrial: urban areas involving residential or industrial buildings;
- Arable land: also specified as cropland, non-irrigated arable land areas, permanently irrigated land, and rice fields;
- Grassland: natural grassland;
- Forests: broad-leaved forest, coniferous forest, mixed forest;
- Heterogeneous agricultural areas: annual crops associated with permanent crops, complex cultivation, land principally occupied by agriculture, agro-forestry areas;
- Open spaces with little or no vegetation: beaches, dunes, sand plains, bare rock, sparsely vegetated areas, and glaciers;
- Pastures: ground covered with grass or herbage, used or suitable for the grazing of livestock;
- Permanent Crops: vineyards, fruit trees and berry plantations, olive groves;
- Shrub and/or herbaceous vegetation association: natural grassland, moors, and heathland, Sclerophyllous vegetation, transitional woodland shrub;
- Inland wetlands: inland marshes, peat bogs;
- Woodland shrub: transitional woodland shrub.

For sake of space the table is split into two parts: the first reports land-use attributes from “Arable land” to “Open spaces with little or no vegetation”, the other from “Pastures” to “Woodland”.

EMSR	Res. / Burnt	Ind. (ha) AoI	Arable Land (ha) Burnt AoI	Grassland (ha) Burnt AoI	Forests (ha) Burnt AoI	Het. agric. (ha) Burnt AoI	Open sp. (ha) Burnt AoI
221_01	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
371_01	3.5	29	123.3	646.8	36.5	98.9	503.8
290_03			115.6		17.8	86.7	
298_02			262.2		197.4	56.4	
298_06					350.5	7661	
248_01		112.1	1048.1	7562.7	84.1		
248_03			286.2	2617.4	247.8		
248_04			85.2	573	36.4		
248_05		58.8	247.6	1330.4	5.6		
368_01		12.1		32.2	58.2	10.8	1010.8
216_01	43	185	169.7	244.3	46.5		
216_02			37	265.6			
216_04		466		646.5	185.3		
216_05		57	3.3	129.4			26.2 1872.3
365_01		365.1		1323.3	1538	2287	35,995.5
373_01			208.3	2899.4	173	122.6	810
302_01		761	630.2		708	3148.8	
302_06	24	4650		29.3	377.8	27.2	
302_07	3	54			54.9	50.3	
250_01				120.7	1129.3	50.8	636.2 2356.4
372_04	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.

Table A.2: Details on land use for the areas of interest considered in this study. This table is partial and continues in Table A.3. It reports, in hectares: Residential/Industrial areas, Arable lands, Grasslands, Forests, Heterogeneous agricultural areas, and Open spaces with little or no vegetation. For each land use type, burnt regions are reported (Burnt).

EMSR	Pastures (ha)		Perm. Crops (ha)		Shrubs / herb. (ha)		In. Wetlands (ha)		Woodland (ha)	
	<i>Burnt</i>	<i>AoI</i>	<i>Burnt</i>	<i>AoI</i>	<i>Burnt</i>	<i>AoI</i>	<i>Burnt</i>	<i>AoI</i>	<i>Burnt</i>	<i>AoI</i>
221_01	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
371_01			25	267.9	381.8	936.3				
290_03		13.9			18.3	120.1		14		
298_02					7.9	112				
298_06					52.2	1626.2				
248_01					1642.4	8521.42			56.1	1989.9
248_03					674.7	2980.1			836	1565.6
248_04					437.4	2353.5			47.3	220.6
248_05					588.9	1513			166.7	2457.5
368_01		1708.6		2896.1	1471.3	15,130.6				
216_01					796.1	924			860.9	1482.4
216_02					658.2	1834.5			172.8	643
216_04					99.7	821.6				980.1
216_05						236.1			148.8	483.82
365_01			316.8	6695.8	880.2	8877.5		38.3		
373_01					366.3	918.3				
302_01	199.4		499.8	958	13,469.9					
302_06					391.9	960.4				
302_07					83.6	2578.9				
250_01		300		642	1172.2	6596.2				
372_04	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.

Table A.3: Land use details for the AoIs considered in this work. For the sake of space, this table is partial, and continues from Table A.2. It reports, in hectares: Pastures areas, Permanent croplands, Shrubs or herbaceous vegetation areas, In-land wetlands areas, and Woodlands. For each land use type, the areas affected by wildfire are reported (Burned).

A.3 Double-Step U-Net architecture

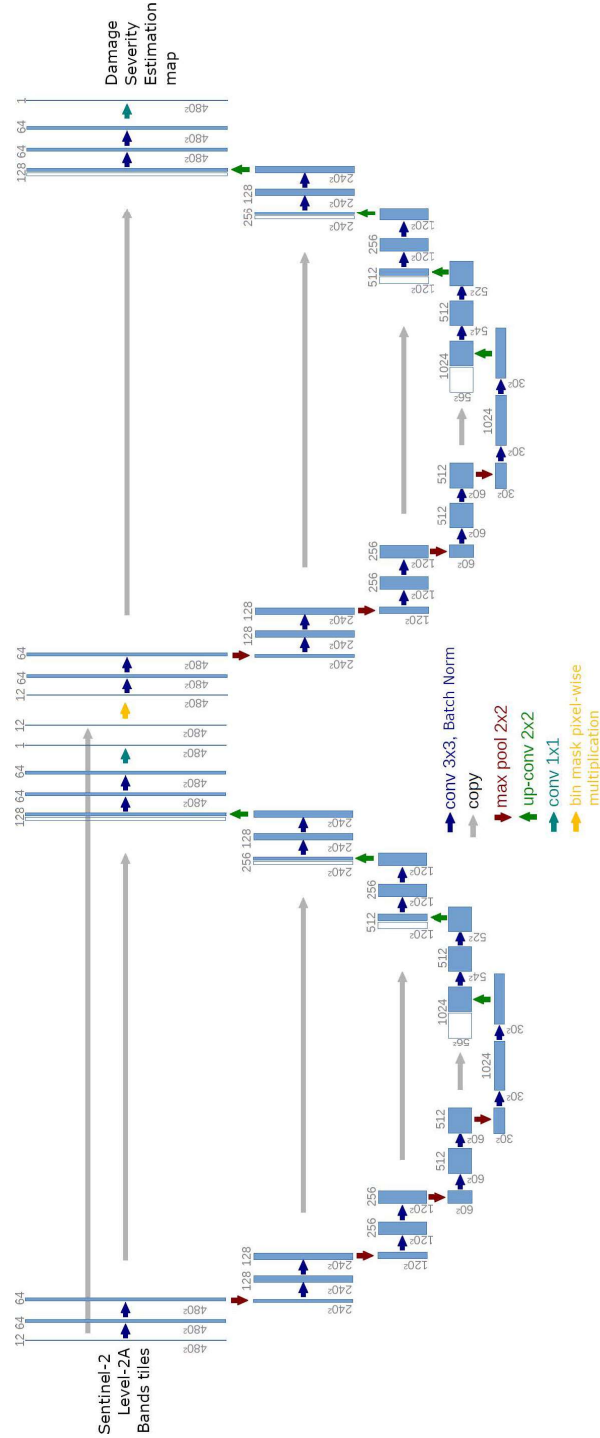


Figure A.1: Double-Step U-Net architecture.

A.4 Compactness loss gradient derivation

In [119], the derivation of the backpropagation formula obtained from the gradient of l_C with respect to x_{ij} contains a mistake. Indeed, in Appendix A in [119], it is stated that the gradient is given by the following equation

$$\frac{\partial l_C}{\partial x_{ij}} = \frac{2}{(n-1)nk} \left[n \times (x_{ij} - m_{ij}) - \sum_{l=1}^n (x_{il} - m_{il}) \right]. \quad (\text{A.1})$$

The first mistake is within the summation since the samples \mathbf{x}_i have k components, not n . However, as we will prove, this is not the unique mistake.

Let us compute the gradient of l_C with respect to x_{ij} . Using the definition of the inner product, we have that $\mathbf{z}_i^T \mathbf{z}_i = \sum_{t=1}^k z_{it}^2$. Thus, l_C can be written as

$$l_C = \frac{1}{nk} \sum_{l=1}^n \sum_{t=1}^k (x_{lt} - m_{lt})^2.$$

Now, taking partial derivatives of l_C with respect to x_{ij} for all $1 \leq i \leq n$ and $1 \leq j \leq k$, we obtain

$$\frac{\partial l_C}{\partial x_{ij}} = \frac{2}{nk} \sum_{l=1}^n (x_{lj} - m_{lj}) \cdot \left(\frac{\partial (x_{lj} - m_{lj})}{\partial x_{ij}} \right).$$

This first step is already wrong in [119]. The rest of the proof follows similarly. Let us check it. Note that

$$\frac{\partial (x_{lj} - m_{lj})}{\partial x_{ij}} = \begin{cases} 1 & \text{if } l = i, \\ -\frac{1}{n-1} & \text{otherwise.} \end{cases}$$

Thus, we obtain that

$$\begin{aligned} \frac{\partial l_C}{\partial x_{ij}} &= \frac{2}{nk} \left[x_{ij} - m_{ij} - \frac{1}{n-1} \sum_{\substack{l=1 \\ l \neq i}}^n (x_{lj} - m_{lj}) \right] \\ &= \frac{2}{nk} \left[\frac{n}{n-1} \cdot (x_{ij} - m_{ij}) - \frac{1}{n-1} \sum_{l=1}^n (x_{lj} - m_{lj}) \right] \\ &= \frac{2}{(n-1)nk} \left[n \cdot (x_{ij} - m_{ij}) - \sum_{l=1}^n (x_{lj} - m_{lj}) \right], \end{aligned}$$

retrieving finally

$$\frac{\partial l_C}{\partial x_{ij}} = \frac{2}{(n-1)nk} \left[n \cdot (x_{ij} - m_{ij}) - \sum_{l=1}^n (x_{lj} - m_{lj}) \right].$$

Bibliography

- [1] Kashif Ahmad et al. “Multi-Modal Machine Learning for Flood Detection in News, Social Media and Satellite Sequences”. In: *arXiv preprint arXiv:1910.02932* (2019).
- [2] Stelios Andreadis et al. “Multimedia Analysis Techniques for Flood Detection Using Images, Articles and Satellite Imagery”. In: *Working Notes Proceedings of the MediaEval 2019 Workshop, Sophia Antipolis, France*. 2019, pp. 27–30.
- [3] C Bayik et al. “Exploiting multi-temporal Sentinel-1 SAR data for flood extend mapping”. In: *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci* 42.3 (2018), W4.
- [4] Abdelhakim Benoudjit and Raffaella Guida. “A Novel Fully Automated Mapping of the Flood Extent on SAR Images Using a Supervised Classifier”. In: *Remote Sensing* 11.7 (2019), p. 779.
- [5] Wu Bin et al. “A Method of Automatically Extracting Forest Fire Burned Areas Using Gf-1 Remote Sensing Images”. In: *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*. IEEE. 2019, pp. 9953–9955.
- [6] Benjamin Bischke, Simon Brugman, and Patrick Helber. “Flood Severity Estimation from Online News Images and Multi-Temporal Satellite Images using Deep Neural Networks”. In: (2019).
- [7] Benjamin Bischke, Patrick Helber, and Andreas Dengel. “Global-Local Feature Fusion for Image Classification of Flood Affected Roads from Social Multimedia”. In: *Proc. of the MediaEval 2018 Workshop (Oct. 29-31, 2018)*. Sophia-Antipolis, France, 2018.
- [8] Benjamin Bischke et al. “The Multimedia Satellite Task at MediaEval 2018: Emergency Response for Flooding Events”. In: *Proc. of the MediaEval 2018 Workshop (Oct. 29-31, 2018)*. Sophia-Antipolis, France, 2018.
- [9] Dan Bina et al. “Flood Severity Estimation in News Articles Using Deep Learning Approaches”. In: *CEUR-WS: Aachen, Germany* (2019), p. 2670.

- [10] Piotr Bojanowski et al. “Enriching word vectors with subword information”. In: *Transactions of the Association for Computational Linguistics* 5 (2017), pp. 135–146.
- [11] Luigi Boschetti et al. “MODIS–Landsat fusion for large area 30 m burned area mapping”. In: *Remote Sensing of Environment* 161 (2015), pp. 27–42.
- [12] Léon Bottou. “Stochastic gradient descent tricks”. In: *Neural networks: Tricks of the trade*. Springer, 2012, pp. 421–436.
- [13] Justin D Braaten, Warren B Cohen, and Zhiqiang Yang. “Automated cloud and cloud shadow identification in Landsat MSS imagery for temperate ecosystems”. In: *Remote Sensing of Environment* 169 (2015), pp. 128–138.
- [14] Max Bramer. *Principles of data mining*. Vol. 180. Springer, 2007.
- [15] Pierrick Bruneau and Thomas Tamisier. “Transfer learning and mixed input deep neural networks for estimating flood severity in news content”. In: *MediaEval Multimedia Evaluation Workshop*. 2019.
- [16] Antoni Buades, Bartomeu Coll, and J-M Morel. “A non-local algorithm for image denoising”. In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)*. Vol. 2. IEEE. 2005, pp. 60–65.
- [17] Deng Cai, Xiaofei He, and Jiawei Han. “Speed up kernel discriminant analysis”. In: *The International Journal on Very Large Data Bases* 20.1 (2011), pp. 21–33.
- [18] Zhe Cao et al. “Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. July 2017.
- [19] Mark L Carroll et al. “A new global raster water mask at 250 m resolution”. In: *International Journal of Digital Earth* 2.4 (2009), pp. 291–308.
- [20] Savvas A Chatzichristofis and Yiannis S Boutalis. “CEDD: color and edge directivity descriptor: a compact descriptor for image indexing and retrieval”. In: *International Conference on Computer Vision Systems*. Springer. 2008, pp. 312–322.
- [21] Savvas A Chatzichristofis and Yiannis S Boutalis. “FCTH: Fuzzy color and texture histogram-a low level feature for accurate image retrieval”. In: *2008 Ninth International Workshop on Image Analysis for Multimedia Interactive Services*. IEEE. 2008, pp. 191–196.
- [22] Hao Chen et al. “Deep contextual networks for neuronal structure segmentation”. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 30. 1. 2016.

- [23] Marco Chini et al. “A hierarchical split-based approach for parametric thresholding of SAR images: Flood inundation as a test case”. In: *IEEE Transactions on Geoscience and Remote Sensing* 55.12 (2017), pp. 6975–6988.
- [24] European Commission. *Copernicus EMS Rapid Mapping Manual*. Accessed: 2021-05-09.
- [25] *Comparison of Spatial Resolution and Wavelength Characteristics of Sentinel-2, LANDSAT-8, and SPOT Instruments*. <https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-2/heritage>. 2020.
- [26] *Copernicus Sentinel-2 Radiometric Resolutions*. <https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi/resolutions/radiometric>. Accessed: 2019-12-19. 2021.
- [27] Fernando Moreira De Araujo and Laerte G Ferreira. “Satellite-based automated burned area detection: A performance assessment of the MODIS MCD45A1 in the Brazilian savanna”. In: *International Journal of Applied Earth Observation and Geoinformation* 36 (2015), pp. 94–102.
- [28] J. Deng et al. “ImageNet: A Large-Scale Hierarchical Image Database”. In: *CVPR09*. 2009.
- [29] Jia Deng et al. “ImageNet: A large-scale hierarchical image database”. In: *2009 IEEE conference on computer vision and pattern recognition*. IEEE. 2009, pp. 248–255.
- [30] Danielle Dias and Ulisses Dias. “Flood detection from social multimedia and satellite images using ensemble and transfer learning with CNN architectures”. In: *Proc. of the MediaEval 2018 Workshop (Oct. 29-31, 2018)*. Sophia-Antipolis, France, 2018.
- [31] Inc. Docker. *Docker*. <https://www.docker.com/>. 2020.
- [32] G. Donchyts et al. “A 30 m resolution surface water mask including estimation of positional and thematic differences using landsat 8, srtm and openstreetmap: a case study in the Murray-Darling Basin, Australia”. In: *Remote Sensing* 8.5 (2016), p. 386.
- [33] *EFFIS: Sentinel-2 satellite Spatial Resolution*. <https://effis.jrc.ec.europa.eu/about-effis/technical-background/rapid-damage-assessment>. Accessed: 2021-03-23. 2021.
- [34] Amany Elbanna et al. “Emergency management in the changing world of social media: Framing the research agenda with the stakeholders through engaged scholarship”. In: *International Journal of Information Management* 47 (2019), pp. 112–120.
- [35] *EMS Rapid Mapping Product Portfolio*. https://en.wikipedia.org/wiki/Publish-subscribe_pattern. Accessed: 2021-03-02. 2019.

- [36] *EMS Rapid Mappings Portfolio*. https://emergency.copernicus.eu/mapping/sites/default/files/files/CopernicusEMS-Service_Portfolio-Rapid_Mapping.pdf. Accessed: 2021-03-18. 2019.
- [37] *ESA Sentinel Product Overview: Polarimetry*. <https://sentinel.esa.int/web/sentinel/user-guides/sentinel-1-sar/product-overview/polarimetry>. 2019.
- [38] *ESA Sentinel-2 Satellite - Spatial Resolution*. <https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/resolutions/spatial>. Accessed: 2019-12-14. 2019.
- [39] *EU - What is Horizon 2020?* <https://ec.europa.eu/programmes/horizon2020/en/what-horizon-2020>. Accessed: 2021-03-11. 2020.
- [40] *European Civil Protection and Humanitarian Aid Operations - European Disaster Risk Management*. https://ec.europa.eu/echo/what/civil-protection/european-disaster-risk-management_en. Accessed: 2021-03-10. 2017.
- [41] *European Commission - Overview of natural and man-made disaster risks the European Union may face*. https://ec.europa.eu/echo/sites/echo-site/files/overview_of_natural_and_man-made_disaster_risks_the_european_union_may_face.pdf. Accessed: 2021-03-12. 2020.
- [42] *European Copernicus Programme*. <https://www.copernicus.eu/en/about-copernicus>. Accessed: 2021-03-18. 2021.
- [43] Alessandro Farasin, Luca Colomba, and Paolo Garza. “Double-Step U-Net: A Deep Learning-Based Approach for the Estimation of Wildfire Damage Severity through Sentinel-2 Satellite Data”. In: *Applied Sciences* 10.12 (2020), p. 4332.
- [44] Alessandro Farasin et al. “Supervised Burned Areas delineation by means of Sentinel-2 imagery and Convolutional Neural Networks”. In: *Proceedings of the 17th International Conference on Information Systems for Crisis Response and Management (ISCRAM 2020)*, Virginia Tech, Blacksburg, VA, USA. 2020, pp. 1060–1071.
- [45] Alessandro Farasin et al. *Unsupervised Burned Area Estimation through Satellite Tiles: A multimodal approach by means of image segmentation over remote sensing imagery*. 2019.
- [46] Yu Feng et al. “Ensembled Convolutional Neural Network Models for Retrieving Flood Relevant Tweets”. In: *Proc. of the MediaEval 2018 Workshop (Oct. 29-31, 2018)*. Sophia-Antipolis, France, 2018.
- [47] Yu Feng et al. “Flood level estimation from news articles and flood detection from satellite image sequences”. In: *Development* 52.75.56 (1973), pp. 73–23.

- [48] Alfonso Fernández-Manso, Oscar Fernández-Manso, and Carmen Quintano. “SENTINEL-2A red-edge spectral indices suitability for discriminating burn severity”. In: *International journal of applied earth observation and geoinformation* 50 (2016), pp. 170–175.
- [49] Alfonso Fernández-Manso, Carmen Quintano, and Dar A Roberts. “Can Landsat-Derived Variables Related to Energy Balance Improve Understanding of Burn Severity From Current Operational Techniques?” In: *Remote Sensing* 12.5 (2020), p. 890.
- [50] Cornelia Ferner et al. “Automated Seeded Latent Dirichlet Allocation for Social Media Based Event Detection and Mapping”. In: *Information* 11.8 (2020), p. 376.
- [51] Federico Filipponi. “BAIS2: Burned Area Index for Sentinel-2”. In: *Multi-disciplinary Digital Publishing Institute Proceedings*. Vol. 2. 7. 2018, p. 364.
- [52] Peter Flach. *Machine learning: the art and science of algorithms that make sense of data*. Cambridge University Press, 2012.
- [53] J Fohringer et al. “Social media as an information source for rapid flood inundation mapping”. In: *Natural Hazards and Earth System Sciences* 15.12 (2015), pp. 2725–2738.
- [54] Open Knowledge Foundation. *Comprehensive Knowledge Archive Network (CKAN)*. <https://ckan.org/>. 2020.
- [55] Open Knowledge Foundation. *Flask*. <https://flask.palletsprojects.com/en/2.0.x/>. 2020.
- [56] Hariny Ganapathy et al. “Deep learning models for estimation of flood severity using Satellite and News Article Images”. In: (2019).
- [57] Ross Girshick et al. “Rich feature hierarchies for accurate object detection and semantic segmentation”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014, pp. 580–587.
- [58] Laura Giustarini et al. “A change detection approach to flood mapping in urban areas using TerraSAR-X”. In: *IEEE transactions on Geoscience and Remote Sensing* 51.4 (2012), pp. 2417–2430.
- [59] Laura Giustarini et al. “Probabilistic flood mapping using synthetic aperture radar data”. In: *IEEE Transactions on Geoscience and Remote Sensing* 54.12 (2016), pp. 6958–6969.
- [60] Xavier Glorot and Yoshua Bengio. “Understanding the difficulty of training deep feedforward neural networks”. In: *Proceedings of the thirteenth international conference on artificial intelligence and statistics*. 2010, pp. 249–256.
- [61] Ian Goodfellow et al. *Deep learning*. Vol. 1. 2. MIT press Cambridge, 2016.

- [62] PostgreSQL Global Development Group. *PostgreSQL*. <https://www.postgresql.org/>. 2020.
- [63] Muhammad Hanif, Muhammad Atif Tahir, and Muhammad Rafi. “Detection of passable roads using Ensemble of Global and Local Features”. In: *Proc. of the MediaEval 2018 Workshop (Oct. 29-31, 2018)*. Sophia-Antipolis, France, 2018.
- [64] Leonardo A Hardtke et al. “Semi-automated mapping of burned areas in semi-arid ecosystems using MODIS time-series imagery”. In: *International Journal of Applied Earth Observation and Geoinformation* 38 (2015), pp. 25–35.
- [65] Simon S Haykin et al. *Neural networks and learning machines*. Vol. 3. Pearson education Upper Saddle River, 2009.
- [66] Kaiming He et al. “Deep residual learning for image recognition”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.
- [67] Benjamin Herfort et al. “Exploring the geographical relations between social media and flood phenomena to improve situational awareness”. In: *Connecting a digital Europe through location and place*. Springer, 2014, pp. 55–71.
- [68] Samuel Hislop et al. “Using landsat spectral indices in time-series to assess wildfire disturbance and recovery”. In: *Remote sensing* 10.3 (2018), p. 460.
- [69] Gao Huang et al. “Densely connected convolutional networks”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 4700–4708.
- [70] Haiyan Huang et al. “Separability analysis of Sentinel-2A multi-spectral instrument (MSI) data for burned area discrimination”. In: *Remote Sensing* 8.10 (2016), p. 873.
- [71] M Hughes, S Kaylor, and Daniel Hayes. “Patch-based forest change detection from Landsat time series”. In: *Forests* 8.5 (2017), p. 166.
- [72] IPCC - *Climate Change 2013: The Physical Science Basics - Technical Summary*. https://www.ipcc.ch/site/assets/uploads/sites/4/2020/07/03_Technical-Summary-TS_V2.pdf. Accessed: 2021-03-11. 2013.
- [73] IPCC - *Global Warming of 1.5 °C*. <https://www.ipcc.ch/sr15/>. Accessed: 2021-03-11. 2018.
- [74] Gareth Ireland, Michele Volpi, and George Petropoulos. “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 (2015), pp. 3372–3399.

- [75] Pallavi Jain, Bianca Schoen-Phelan, and Robert Ross. “MediaEval2019: Flood Detection in Time Sequence Satellite Images”. In: (2019).
- [76] Hamid A Jalab. “Image retrieval system based on color layout descriptor and Gabor filters”. In: *2011 IEEE Conference on Open Systems*. IEEE. 2011, pp. 32–36.
- [77] Carl H Key and Nate C Benson. “Landscape Assessment (LA). FIREMON: Fire effects monitoring and inventory system”. In: *Gen. Tech. Rep. RMRS-GTR-164-CD, Fort Collins, CO: US Department of Agriculture, Forest Service, Rocky Mountain Research Station* (2006).
- [78] Yoon Kim. “Convolutional Neural Networks for Sentence Classification”. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics, Oct. 2014, pp. 1746–1751. DOI: [10.3115/v1/D14-1181](https://doi.org/10.3115/v1/D14-1181). URL: <https://www.aclweb.org/anthology/D14-1181>.
- [79] Armin Kirchknopf et al. “Detection of Road Passability from Social Media and Satellite Images”. In: *Proc. of the MediaEval 2018 Workshop (Oct. 29-31, 2018)*. Sophia-Antipolis, France, 2018.
- [80] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. “Imagenet classification with deep convolutional neural networks”. In: *Advances in neural information processing systems*. 2012, pp. 1097–1105.
- [81] Anna Kruspe, Jens Kersten, and Friederike Klan. “Detection of informative tweets in crisis events”. In: *Natural Hazards and Earth System Sciences Discussions* (2020), pp. 1–18.
- [82] Peter A Lachenbruch. “McNemar test”. In: *Wiley StatsRef: Statistics Reference Online* (2014).
- [83] Ning Li et al. “Robust river boundaries extraction of dammed lakes in mountain areas after Wenchuan Earthquake from high resolution SAR images combining local connectivity and ACM”. In: *ISPRS journal of photogrammetry and remote sensing* 94 (2014), pp. 91–101.
- [84] Nicola Linty et al. “Detection of GNSS Ionospheric Scintillations Based on Machine Learning Decision Tree”. In: *IEEE Transactions on Aerospace and Electronic Systems* 55.1 (2018), pp. 303–317.
- [85] CHANG Liu. “Analysis of Sentinel-1 SAR data for mapping standing water in the Twente region”. MA thesis. University of Twente, 2016.
- [86] Meng Liu, Sorin Popescu, and Lonesome Malambo. “Feasibility of burned area mapping based on ICESAT- 2 photon counting data”. In: *Remote Sensing* 12.1 (2020), p. 24.

- [87] Xuebo Liu et al. “Fots: Fast oriented text spotting with a unified network”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 5676–5685.
- [88] Rafael Llorens et al. “A methodology to estimate forest fires burned areas and burn severity degrees using Sentinel-2 data. Application to the October 2017 fires in the Iberian Peninsula”. In: *International Journal of Applied Earth Observation and Geoinformation* 95 (2021), p. 102243.
- [89] Laura Lopez-Fuentes et al. “Deep Learning Models for Passability Detection of Flooded Roads.” In: *MediaEval*. 2018.
- [90] Laura Lopez-Fuentes et al. “Deep Learning Models for Road Passability Detection during Flood Events Using Social Media Data”. In: *Applied Sciences* 10.24 (2020), p. 8783.
- [91] Laura Lopez-Fuentes et al. “Multi-modal Deep Learning Approach for Flood Detection.” In: *MediaEval* 17 (2017), pp. 13–15.
- [92] Jérôme Louis et al. “Sen2Cor Atmospheric Correction with Meteorological Aerosol Optical Thickness”. In: (2018).
- [93] Jun Lu et al. “Automated flood detection with improved robustness and efficiency using multi-temporal SAR data”. In: *Remote sensing letters* 5.3 (2014), pp. 240–248.
- [94] Sergio Luna and Michael J Pennock. “Social media applications and emergency management: A literature review and research agenda”. In: *International journal of disaster risk reduction* 28 (2018), pp. 565–577.
- [95] A Lyapustin et al. “Multiangle implementation of atmospheric correction (MAIAC): 2. Aerosol algorithm”. In: *Journal of Geophysical Research: Atmospheres* 116.D3 (2011).
- [96] Bangalore S Manjunath et al. “Color and texture descriptors”. In: *IEEE Transactions on circuits and systems for video technology* 11.6 (2001), pp. 703–715.
- [97] Agnese Marcelli et al. “Large-scale two-phase estimation of wood production by poplar plantations exploiting Sentinel-2 data as auxiliary information”. In: (2020).
- [98] Sandro Martinis, Jens Kersten, and André Twele. “A fully automated TerraSAR-X based flood service”. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 104 (2015), pp. 203–212.
- [99] Sandro Martinis, André Twele, and Stefan Voigt. “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 (2009), pp. 303–314.

- [100] *MediaEval 2018 Multimedia Satellite Task*. <http://www.multimediaeval.org/mediaeval2018/multimediasatellite/>. Data released: 31 May 2018. 2018.
- [101] *MediaEval2019: Multimedia Satellite task*. <http://www.multimediaeval.org/mediaeval2019/multimediasatellite/>. Accessed: 2021-04-25. 2021.
- [102] Andrea Melchiorre and Luigi Boschetti. “Global analysis of burned area persistence time with MODIS data”. In: *Remote Sensing* 10.5 (2018), p. 750.
- [103] Microsoft. *Azure Blob Storage*. <https://azure.microsoft.com/en-us/services/storage/blobs/>. 2020.
- [104] Microsoft. *Onnx Runtime*. <https://www.onnxruntime.ai/>. 2020.
- [105] Tomas Mikolov et al. “Efficient estimation of word representations in vector space”. In: *Proceedings of the International Conference on Learning Representations (ICLR 2013)*. Workshop poster. Available at <https://arxiv.org/pdf/1301.3781.pdf>. 2013.
- [106] Muhammad Hanif Mir Murtaza, Muhammad Atif Tahir, and Muhammad Rafi. “Ensemble and Inference based Methods for Flood Severity Estimation Using Visual Data”. In: (2019).
- [107] Anastasia Moutzidou et al. “A multimodal approach in estimating road passability through a flooded area using social media and satellite images”. In: *Proc. of the MediaEval 2018 Workshop (Oct. 29-31, 2018)*. Sophia-Antipolis, France, 2018.
- [108] *MultiSpectral Instrument (MSI) overview*. <https://earth.esa.int/web/sentinel/technical-guides/sentinel-2-msi/msi-instrument>. Accessed: 2019-12-03. 2019.
- [109] *Natural Hazards and Climate Change in European Regions, United Nations International Strategy for Disaster Reduction (UNISDR)*. https://www.espon.eu/sites/default/files/attachments/20130704_ESPON_TERRITORAL_07_CS6_CM_Final.pdf. Accessed: 2021-03-10. 2013.
- [110] AY Ng. “Proceedings of the twenty-first international conference on Machine learning”. In: (2004).
- [111] Antonio Donato Nobre et al. “Height Above the Nearest Drainage—a hydrologically relevant new terrain model”. In: *Journal of Hydrology* 404.1-2 (2011), pp. 13–29.
- [112] Sentinel-1 ObservationScenario. *Sentinel-1 Observation Scenario*. 2019. URL: <https://sentinel.esa.int/web/sentinel/missions/sentinel-1/observation-scenario>.
- [113] *Onnx*. <https://onnx.ai/>. 2020.

- [114] *OpenStreetMap*. <https://www.openstreetmap.org/>. Accessed: 2021-04-21. 2021.
- [115] Nobuyuki Otsu. “A threshold selection method from gray-level histograms”. In: *IEEE transactions on systems, man, and cybernetics* 9.1 (1979), pp. 62–66.
- [116] Giulio Palomba, Alessandro Farasin, and Claudio Rossi. “Sentinel-1 Flood Delineation with Supervised Machine Learning”. In: *Proceedings of the 17th International Conference on Information Systems for Crisis Response and Management (ISCRAM 2020)*, Virginia Tech, Blacksburg, VA, USA. 2020, pp. 1072–1083.
- [117] Dong Kwon Park, Yoon Seok Jeon, and Chee Sun Won. “Efficient use of local edge histogram descriptor”. In: *Proceedings of the 2000 ACM workshops on Multimedia*. ACM. 2000, pp. 51–54.
- [118] Jeffrey Pennington, Richard Socher, and Christopher Manning. “Glove: Global vectors for word representation”. In: *Proceedings of the 2014 conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2014, pp. 1532–1543.
- [119] Pramuditha Perera and Vishal M Patel. “Learning deep features for one-class classification”. In: *IEEE Transactions on Image Processing* 28.11 (2019), pp. 5450–5463.
- [120] Pivotal. *RabbitMQ*. <https://www.rabbitmq.com/>. 2020.
- [121] L Pulvirenti et al. “An algorithm for operational flood mapping from synthetic aperture radar (SAR) data based on the fuzzy logic”. In: *Natural Hazard and Earth System Sciences* (2011).
- [122] Luca Pulvirenti et al. “Discrimination of water surfaces, heavy rainfall, and wet snow using COSMO-SkyMed observations of severe weather events”. In: *IEEE transactions on geoscience and remote sensing* 52.2 (2013), pp. 858–869.
- [123] Khanh-An C Quan et al. *Flood event analysis base on pose estimation and water-related scene recognition*. 2019.
- [124] Ruben Ramo and Emilio Chuvieco. “Developing a random forest algorithm for MODIS global burned area classification”. In: *Remote Sensing* 9.11 (2017), p. 1193.
- [125] Ruben Ramo et al. “A data mining approach for global burned area mapping”. In: *International journal of applied earth observation and geoinformation* 73 (2018), pp. 39–51.
- [126] Joseph Redmon and Ali Farhadi. *YOLOv3: An Incremental Improvement*. 2018. arXiv: [1804.02767](https://arxiv.org/abs/1804.02767) [cs.CV].

- [127] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. “U-net: Convolutional networks for biomedical image segmentation”. In: *International Conference on Medical image computing and computer-assisted intervention*. Springer. 2015, pp. 234–241.
- [128] E Roteta et al. “Development of a Sentinel-2 burned area algorithm: Generation of a small fire database for sub-Saharan Africa”. In: *Remote sensing of environment* 222 (2019), pp. 1–17.
- [129] David P Roy et al. “Prototyping a global algorithm for systematic fire-affected area mapping using MODIS time series data”. In: *Remote sensing of environment* 97.2 (2005), pp. 137–162.
- [130] David P Roy et al. “The collection 5 MODIS burned area product—Global evaluation by comparison with the MODIS active fire product”. In: *Remote sensing of Environment* 112.9 (2008), pp. 3690–3707.
- [131] ClearType S.r.l. *Dramatiq*. <https://dramatiq.io/>. 2020.
- [132] Naina Said et al. “Deep learning approaches for flood classification and flood aftermath detection”. In: *Proc. of the MediaEval 2018 Workshop (Oct. 29-31, 2018)*. Sophia-Antipolis, France, 2018.
- [133] Lennert Schepers et al. “Burned area detection and burn severity assessment of a heathland fire in Belgium using airborne imaging spectroscopy (APEX)”. In: *Remote Sensing* 6.3 (2014), pp. 1803–1826.
- [134] Stefan Schlaffer et al. “Flood detection from multi-temporal SAR data using harmonic analysis and change detection”. In: *International Journal of Applied Earth Observation and Geoinformation* 38 (2015), pp. 15–24.
- [135] *Self Organizing Maps, illustration reference*. https://en.wikipedia.org/wiki/Self-organizing_map. Accessed: 2021-03-15. 2019.
- [136] *Sentinel-1 Extra Wide Swath*. <https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-1-sar/acquisition-modes/extra-wide-swath>. Accessed: 2021-03-19. 2021.
- [137] *Sentinel-1 Interferometric Wide Swath*. <https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-1-sar/acquisition-modes/interferometric-wide-swath>. Accessed: 2021-03-19. 2021.
- [138] *Sentinel-1 StripMap*. <https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-1-sar/acquisition-modes/stripmap>. Accessed: 2021-03-19. 2021.
- [139] *Sentinel-1 Wave*. <https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-1-sar/acquisition-modes/wave>. Accessed: 2021-03-19. 2021.

- [140] *Sentinel-Hub*. <https://www.sentinel-hub.com/>. Accessed: 2021-04-01. 2019.
- [141] Tianchan Shan et al. “A Burned Area Mapping Algorithm for Chinese FengYun-3 MERSI Satellite Data”. In: *Remote Sensing* 9.7 (2017), p. 736.
- [142] Francescopaolo Sica et al. “The offset-compensated nonlocal filtering of interferometric phase”. In: *Remote Sensing* 10.9 (2018), p. 1359.
- [143] Tomer Simon, Avishay Goldberg, and Bruria Adini. “Socializing in emergencies—A review of the use of social media in emergency situations”. In: *International Journal of Information Management* 35.5 (2015), pp. 609–619.
- [144] K. Simonyan and A. Zisserman. “Very Deep Convolutional Networks for Large-Scale Image Recognition”. In: *ICLR* (2015).
- [145] Karen Simonyan and Andrew Zisserman. “Very deep convolutional networks for large-scale image recognition”. In: *arXiv preprint arXiv:1409.1556* (2014).
- [146] Mrinal Singha et al. “Identifying floods and flood-affected paddy rice fields in Bangladesh based on Sentinel-1 imagery and Google Earth Engine”. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 166 (2020), pp. 278–293.
- [147] SmartBEAR. *Swagger*. <https://swagger.io/>. 2020.
- [148] Toufique A Soomro et al. “Strided U-Net model: Retinal vessels segmentation using dice loss”. In: *2018 Digital Image Computing: Techniques and Applications (DICTA)*. IEEE. 2018, pp. 1–8.
- [149] S Sreechanth and Kiran Yarrakula. “Multi-Temporal Analysis of Sentinel-1 SAR data for Urban Flood Inundation Mapping-Case study of Chennai Metropolitan City”. In: *jiP* 2 (2017), p. 1.
- [150] Mervyn Stone. “Cross-validatory choice and assessment of statistical predictions”. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 36.2 (1974), pp. 111–133.
- [151] Emanuel A Storey, Douglas A Stow, and John F O’Leary. “Assessing post-fire recovery of chamise chaparral using multi-temporal spectral vegetation index trajectories derived from Landsat imagery”. In: *Remote Sensing of Environment* 183 (2016), pp. 53–64.
- [152] Emanuel Arnal Storey, Krista R Lee West, and Douglas A Stow. “Utility and optimization of LANDSAT-derived burned area maps for southern California”. In: *International Journal of Remote Sensing* 42.2 (2021), pp. 486–505.
- [153] Julia Strebl et al. “Flood Level Estimation from Social Media Images”. In: (2019).

- [154] *Synergise Sentinel Hub Overview*. 2019. URL: <https://www.sentinel-hub.com/about>.
- [155] Christian Szegedy et al. “Inception-V4, Inception-ResNet and the impact of residual connections on learning”. In: *Thirty-First AAAI Conference on Artificial Intelligence*. 2017.
- [156] Christian Szegedy et al. “Rethinking the inception architecture for computer vision”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 2818–2826.
- [157] Tiangolo. *FastAPI*. <https://fastapi.tiangolo.com/>. 2020.
- [158] Kevin E Trenberth. “Changes in precipitation with climate change”. In: *Climate Research* 47.1-2 (2011), pp. 123–138.
- [159] S Trigg and S Flasse. “An evaluation of different bi-spectral spaces for discriminating burned shrub-savannah”. In: *International Journal of Remote Sensing* 22.13 (2001), pp. 2641–2647.
- [160] André Twele et al. “Sentinel-1-based flood mapping: a fully automated processing chain”. In: *International Journal of Remote Sensing* 37.13 (2016), pp. 2990–3004.
- [161] FD Van der Meer, HMA Van der Werff, and FJA Van Ruitenbeek. “Potential of ESA’s Sentinel-2 for geological applications”. In: *Remote sensing of environment* 148 (2014), pp. 124–133.
- [162] S Veraverbeke, Sarah Harris, and Simon Hook. “Evaluating spectral indices for burned area discrimination using MODIS/ASTER (MASTER) airborne simulator data”. In: *Remote Sensing of Environment* 115.10 (2011), pp. 2702–2709.
- [163] Sergio Vitale et al. “Guided patchwise nonlocal SAR despeckling”. In: *IEEE Transactions on Geoscience and Remote Sensing* 57.9 (2019), pp. 6484–6498.
- [164] Kaupo Voormansik et al. “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 (2013), pp. 562–577.
- [165] Peilu Wang et al. “Learning distributed word representations for Bidirectional LSTM recurrent neural network”. In: *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 2016, pp. 527–533.
- [166] Mira Weirather, Gunter Zeug, and Thomas Schneider. “Automated Delineation Of Wildfire Areas Using Sentinel-2 Satellite Imagery”. In: *GI_Forum 2018*, 6 (), pp. 251–262.
- [167] Clayton Wukich et al. “Social media use in emergency management”. In: *Journal of Emergency Management* 13.4 (2015), pp. 281–294.

- [168] Mirko Zaffaroni et al. “AI-based flood event understanding and quantification using online media and satellite data”. In: *CEUR-WS: Aachen, Germany* (2019), p. 2670.
- [169] Konstantinos Zagoris et al. “Automatic image annotation and retrieval using the joint composite descriptor”. In: *2010 14th Panhellenic Conference on Informatics*. IEEE. 2010, pp. 143–147.
- [170] Zhengyu Zhao, Martha Larson, and Nelleke Oostdijk. “Exploiting Local Semantic Concepts for Flooding-related Social Image Classification”. In: *Proc. of the MediaEval 2018 Workshop (Oct. 29-31, 2018)*. Sophia-Antipolis, France, 2018.
- [171] Bolei Zhou et al. “Places: A 10 million image database for scene recognition”. In: *IEEE transactions on pattern analysis and machine intelligence* 40.6 (2018), pp. 1452–1464.

This Ph.D. thesis has been typeset by means of the T_EX-system facilities. The typesetting engine was pdfL^AT_EX. The document class was `toptesi`, by Claudio Beccari, with option `tipotesi=scudo`. This class is available in every up-to-date and complete T_EX-system installation.