MASTER'S DEGREE THESIS

Unsupervised classification of flowing condition in non-perennial rivers

Supervisor:

Prof. Paolo Vezza

Co-supervisor:

Prof. Guillermo Palau Salvador

Eng. Giovanni Negro

Candidate:

Isabelle Brichetto

Department of Environment, Land and Infrastructure Engineering

A.Y. 2022 / 2023

Che cosa vuoi fare da grande? Del Bene.

A tutte le persone che sognano

ed hanno sognato con me.

Abstract

Non-perennial rivers are the most common type of rivers on Earth today. Anthropogenic pressures, such as land use change and water withdrawals, as well as climate change, are leading to significant shifts, modifying rivers flowing conditions from perennial to non-perennial. Change in surface water volume usually causes serious consequences on river ecosystems and aspects of human life, therefore it is crucial to efficiently detect the existence of surface water, extract its extent, quantify its volume and monitor its dynamics. Knowledge about the frequency and duration of flowing condition of non-perennial rivers is severely limited by the small number of streamflow gauges and reliable prediction of surface water presence by hydrological models. In this contest, the satellite images can provide a useful medium for studying ecological and hydrological processes in such systems, offering effective ways to observe and monitor water surface dynamics with spatially explicit and temporally frequent data. The combination of Remote Sensing (RS) and machine learning algorithms, in particular with the use of Convolutional Neural Networks (CNN) that are designed for processing and analyzing visual data such as images, have recently begun to be used also in environmental studies concerning water bodies. In this study, multispectral Sentinel-2 images were used to detect and monitor changes in surface water presence along two non-perennial Mediterranean rivers located in the province of Valencia (Eastern Spain). Thanks to the match between satellite images and ground truth data, consisting in field survey and high-resolution images, the spectral signatures of sediments, vegetation, and water in the river corridors were evaluated, founding that the Sentinel-2's bands combination of SWIR, NIR, and RED allows an optimum discrimination of water on the riverbed. False-Color composite Images (FCI) were then used for the identification of three distinct flowing conditions of non-perennial rivers in the considered reaches: "flowing" (F), "ponding" (P) and "dry" (D). In particular, six years of Sentinel-2 images (period 2017-2023) were classified in terms of identified flowing and cloud-cover conditions, adding the "cloudy" (C) class. The obtained dataset allowed to train a CNN for the unsupervised classification of rivers' flowing status, basing on FCI visualization only. 6-layer CNNs were used, adopting ResNet50 as base model for adding the train dataset at classification head. Three models were used, for 2 (water-nonwater), 3 (F-P-D) and 4 (F-P-D-C) classes problems. Their performances in terms of

accuracy have been 0.7-0.88, always lower than 0.9, indicating a not optimal learning of the models. Despite dataset balancing in pre-processing, the "ponding" condition is the most critical to identify, causing the majority of misclassifications. Subsequent improvements such the extension and the diversification of the dataset are required, but this innovative technique has large application potential for the study of non-perennial rivers.

Table of contents

List of Figures

List of Tables

1 Introduction

Surface water bodies, such as rivers, lakes and wetlands, are critical freshwater resources, for both human and ecological systems. Water helps preserve the biodiversity in riparian and wetland ecosystems by providing habitats to a bunch of flora and fauna species, but also touches multiple aspect of human lives, from drinking water supply to industrial processes, passing from agriculture, electricity production and transportation.

Surface water bodies are dynamic in nature as they shrink, expand, or change their appearance or course of flow with time, owing to different natural and human-induced factors (Huang, 2018). Climate change is within these factors, since in recent decades all the components of the global water cycle have been modified from it (high confidence). The global increases in temperature and evapotranspiration, with changes in rainfall patterns, are significantly modifying the hydrological cycle of rivers. The Assessment Report 6 of IPCC concluded with medium evidence and high agreement that trends in annual streamflow have generally shown changes in regional precipitation and temperature since the 1950s. Observations over recent decades, together with current global-scale climate change models, indicate that the change in precipitation and temperature patterns is increasing the temporal variability, with higher frequency of extreme events such as floods and supra-seasonal droughts (Magand, 2020). Stream flows showed decreasing trends in parts of Southern Europe, Western and Central Africa, Eastern Asia, Western North America and Eastern Australia, and increasing trends in Northern Asia, Northern Europe, and Northern East America. In particular the Southwestern Europe is one of the semi-arid regions in which water bodies, including Mediterranean rivers, are most vulnerable, with high probability to face acute water scarcity problems, also due to land cover changes and increased water demands by the irrigation sector (IPCC, 2023). Change in surface water volume usually causes serious consequences, therefore it is crucial to efficiently detect the existence of surface water, extract its extent, quantify its volume and monitor its dynamics (Huang, 2018). The extraction of water bodies from satellite images is crucial for studying ecological and hydrological processes, urban planning, disaster management, updating geospatial datasets, detecting droughts, monitoring floods, navigation and other applications (Parajuli, 2021). Remote sensing technologies offer effective ways to observe and monitor surface water dynamics at various spatial scales, with spatially explicit and temporally frequent observational data of the Earth's surface. In last years, remote sensing

has exponentially evolved and so have other fields such as computer vision. In addition, a greater number of satellite imagery has become publicly available. Notable examples are Landsat and Sentinel-2 constellations.

1.1 Non-Perennial Rivers

Non-Perennial Rivers (NPRs) are ubiquitous watercourses characterized by the occurrence of non-flowing periods, strongly depending on the season and on water recharge, in form of rainfall or snowmelt, over their catchments. This definition encompasses all watercourses in which some data testify the absence of water along the channel, including a wide range of intermittencies, from streams with episodical water presence to quasi-perennial rivers. The separation between perennial and non-perennial river is not fixed in time. Anthropogenic pressures, such as climate change, land use change and water withdrawals could lead to significant shifts from perennial to non-perennial condition. It's observed that many perennial rivers are gradually becoming intermittent (Magand, 2020). In a recent research of Messager et al. estimated the percentage of NPRs around the world, obtaining that the condition of intermittency is the most common. It has been predicted that water ceases to flow for at least one day per year, on interannual average, along 41% of the river with mean annual flow > 0.1 m^3/s and along 51% (with conservative approach) to 60% of the watercourse with mean annual flow ranging from 0.1 m^3/s to 0.01 m^3/s (Messager, 2021). From these results, the consideration of the non-perennial as the common condition. Under every climate change scenario this river type is predicted to expand in the future and not only in arid and semi-arid climatic regions. Due to the combination and succession of both terrestrial and aquatic ecosystems, the time-averaged biodiversity of NPRs is extremely high. These watercourses provide a variety of benefits, such as the provision of materials (water and timber), freshwater and riparian biodiversity, regulation of biogeochemical cycles and functioning as ecological corridor for wild and herded animals. Drying and rewetting processes, timing and duration of the different aquatic phases influence the biodiversity, the ecosystem functioning and so the provision of ecosystem services (Magand, 2020). However, suffering from negative perceptions and being historically overlooked by researchers compared to perennial rivers, NPRs are deteriorating at alarming rates, with difficulties in being recognized in the legislations as a group of rivers with specific characteristics and needs (Fritz, 2017).

The Water Framework Directive - WFD 2000/60/EC (EU-Parlament, 2000) has been adopted in year 2000 to establish a new framework for the protection and sustainable management of water resources. From the adoption, its purpose has been the classification of the ecological quality of surface water bodies in order to achieve a "good ecological status" in all inland and coastal waters in EU. For doing that, it's necessary to define the reference condition for each surface water body, retrieving the undisturbed case to be compared with subsequent analysis and monitoring. The definition of the reference condition in NPRs is not a simple task, since it requires the knowledge of the degree of intermittence and of the occurrence of the different flowing conditions (Cavallo, 2022). Only some countries in the Mediterranean region integrated the WFD with national implementations, introducing specific classifications for NPRs. In 2008, Spain and Italy, respectively with "ORDEN ARM/2656/2008" and "Decreto Ministeriale 16 giugno 2008, n. 131" (D.M. 131/2008), defined a method for the distinction of this type of river (Manfreda, 2023).

Perennial	Water courses have natural flow regime conditions flow during the whole year.
Temporary or seasonal	Water courses where natural flow regime conditions present a marked seasonality, showing reduced flow or dry riverbed in summer, and flow is present during an average period of 300 days in a year.
Intermittent or strongly seasonal	Water courses where natural flow regime conditions present a high temporality, and flow is present during an average period between 100 and 300 days in a year.
Ephemeral	Water courses that in natural flow regime conditions only flow sporadically, mainly in storm episodes, during an average period less than 100 days in a year.

Figure 1: Classification of non-perennial rivers in the ORDEN ARM/2656/2008, Spain, for the implementation of WFD 2000/60/EC (IPH, 2007)

Stream type	Description
Temporary	Watercourse that can dry out completely and/or at some stretches
• Intermittent	Water is present more than 8 months a year. It may dry out in some river stretches and/or several times a year
• Ephemeral	Water is present <8 months a year. Disconnected pools may remain
• Episodic	Water only present after heavy rains, once every 5 years

Figure 2: Classification of non-perennial rivers in the D.M. 131/2008, Italy, for the implementation of WFD 2000/60/EC (MATTM, 2000)

Main obstacles for the implementation of correct management policies are the lack of information on the ecological functioning of NPRs and the high spatial and temporal variability of this river type. To have clear patterns of river intermittency, is fundamental to know the hydrologic conditions that rivers can experience together with their frequency, duration and seasonality. Hence, the comparison of real-time data with a long-term dataset, functioning as reference condition, is mandatory for understanding the alteration of the natural flow regime (Magand, 2020). In this context, there are no adequate instruments to detect and monitor duration and frequency of non-flowing periods (Cavallo, 2022). The common approaches for hydrologic regime characterization are: field surveys, gauging stations, logger sensors and hydrological models.

Field surveys are one of the best solutions for acquiring real-time data on the metrics of the flow regime and can assess with high accuracy which hydrologic condition is present on time but are difficult to replicate. Gauging stations, instead, within the best solutions to obtain long-term data, allowing to obtain the evolution of flow discharge over years, but they are rarely present in intermittent stretches and not effective in measuring small flows and neither the presence/absence of water during ponding phases. Logger sensors may detect the movement of wetting and drying fronts in real-time and medium-term, but the main drawbacks are the difficulties in distinguishing flowing and standing water and the possibility of the instruments to be swept away or buried during floods. The hydrological models still biased in predicting the variability of flow discharge in NPRs and they require the knowledge of many information, mainly related to the interaction surface-groundwater and the riverbed soil condition (Manfreda, 2023).

Recently, remote sensing has defined significant opportunities for monitoring NPRs. Within the used technique, the airborne surveys allow the execution of rapid and extended surveys on intermittent reaches, even if the riverbed is complicated to reach by an operator. Then, satellite platform can provide periodical images of the entire river network, making possible the constant monitoring of the NPRs flowing conditions. The

main limitation of the satellite image is their spatial resolution, that impedes the recognition of small objects and thus the application on narrow rivers (Manfreda, 2023).

1.2 Remote Sensing

Remote sensing is the science and the technology that allows to acquire qualitative and quantitative data on remote objects and the surrounding environment by measuring the electromagnetic energy that is emitted, reflected or transmitted by the area of interest. The data acquisition takes place thanks to remote sensors mounted on platforms, such as drones, airplanes and satellites, that allow to detect the electromagnetic energy coming from a scene and to convert it into information. Each picture elements, the pixel, has a discrete value in units of Digital Number – DN – (NASA, 2013). If the primary energy source is the Sun, instruments are passive sensors, while if they illuminate themselves the surfaces (e.g. radar), they are passive sensors. Passive sensors measure energy from the optical regions of the electromagnetic spectrum: visible, near infrared, shortwave infrared and thermal infrared. The resulting images have different characteristics and resolutions depending on the sensor used for the detection.

A significant advantage of satellite remote sensing compared to other remote platforms, is the possibility of monitoring wide areas with various temporal (the so called revisit time) and spatial resolutions, providing a global geographic coverage. Moreover, some satellite archives (e.g. Landsat) provide time series longer than 40 years. The wide use of satellite data is also encouraged by the free distribution policy adopted by some space agencies, like the Copernicus program of the European Space Agency (ESA), and by research and education programmers set up by private companies, like Planet and Esri. In most of the cases, data acquired by passive sensors do not require long and complicated pre-processing steps. They are frequently corrected radiometrically and geometrically by various providers and can be directly exploited by users. Furthermore, by associating the blue, green and red bands to the corresponding channels, a true color image can be obtained, through an additive color-synthesis operation, which can be easily interpreted even by non-expert users. The main drawback of passive sensors like satellite, is the inability of observing the Earth's surface in the presence of clouds; long periods without observation may occur in areas with frequent precipitation, ending with non-useful observations.

A variety of satellite missions have followed over the years. Among them Landsat, MODIS (Moderate Resolution Imaging Spectroradiometer) and, more recently, the Sentinel-2 constellation. The choice of the satellite to be used in a specific monitoring application depends on many factors such as the object size (length and width of the target), spatial resolution required, physical properties of the target, duration of the observation and frequency with which changes need to be tracked. In general, spatial, temporal and spectral resolutions are in relative opposition: data with coarse spatial resolution are available with high revisit time and vice versa. Hence, the choice of the most suitable satellite datasets to be used is a challenging task, due to inevitable trade-offs between spatial and temporal resolution (Manfreda, 2023).

The use of satellite data for monitoring NPRs has so far been limited by two main factors: the spatial resolution of the satellite images and the availability of images at affordable costs, since high spatial and temporal resolutions are required for monitoring NPRs. Very high-resolution images (space resolution of the order of 0.5 m) are available for commercial use, but the long-time monitoring is limited by the high products cost. Among the multispectral images with global coverage freely distributed, the Sentinel-2 mission provides the highest spatial resolution and revisiting frequency, so that its images have been widely exploited to map water surfaces along perennial rivers. Satellite imagery allows the computation of the spectral signature of the different land cover classes, meaning the reflectance values as a function of wavelength. Since each material has a unique signature, it is possible to distinguish objects easily creating land cover maps, extraction of land cover classes created by evaluating the spectral content of each pixel along the bands. Some authors have developed supervised and unsupervised classification methods to generate such maps for perennial rivers. Carbonneau et al. used a supervised classification method to extract land cover classes (water, vegetation and sediment) and delimit active channels from Sentinel-2 images on four Italian rivers: Po, Sesia, Paglia and Bonamico River (Carbonneau, 2020). There are various approaches now available for extracting water surfaces from multispectral sensors, but none of them is equally applicable to all case studies, and the challenge of extracting land cover from multispectral satellite frames is constantly evolving.

Only recently some authors have begun to explore the potential of satellite data in monitoring the presence of water along NPRs. Seaton et al. examined the utility of various multispectral indices derived from Sentinel-2 and Landsat-8 imagery for the identification and mapping of water surface areas along three NPRs located in South Africa, Western Cape (Seaton, 2020). Maswanganye et al., instead, investigated the use of multi-source remotely sensed data to monitor water pools distribution and dynamics along three NPRs,

retrieving that the visualizations with indices NDWI, MNDWI, and NDVI are the best for extracting water and for their consequent use into a supervised Random Forest algorithm (Maswanganye, 2022). Furthermore, Wang and Vivoni developed a new approach to establish the presence of surface flow in a NPR of Arizona, USA, by exploiting the commercial CubeSat imagery of Planet and using an index based on NIR band to determine the flow condition in different reaches of the case study; the temporal evolution of this index showed a high degree of convergence with the observed flow data recorded by a gauging station located in the surveyed river reach (Wang Z. &., 2022). Finally, Cavallo et al. used multispectral Sentinel-2 images to detect changes in water surface presence and flowing condition along three Mediterranean NPRs located in southern Italy. By evaluating the spectral signature of the riverbed-cover classes of water, sediment and vegetation, thanks to the contemporaneity between ground truth and Sentinel-2 data, they retrieved that false color image with band SWIR, NIR, and RED is the one that better allows the distinction between the components of the river corridor; the obtained images were then used to train a locally calibrated Random Forest model to fulfill temporal gaps between satellite images and to predict the occurrence of the NPRs flowing condition (F/P/D) by using local meteorological and hydrological data (Cavallo, 2022).

1.3 Convolutional Neural Network

Machine learning is an evolving subset of artificial intelligence (AI) that focuses on the development of algorithms and models that enable computers to learn and to make predictions or decisions based on data, without being explicitly programmed for every possible scenario. Deep learning is a subset of machine learning that focuses on using neural networks with multiple layers (Deep Neural Networks) to learn patterns and representations from data. These neural networks, also known as Artificial Neural Networks, are inspired by the structure and functioning of the human brain's interconnected neurons. Deep learning has been highly successful in various tasks, particularly those involving complex data like images, text, and audio.
Convolutional Neural Networks (CNNs) are a specific type of Deep Neural Network designed for processing and analyzing visual data, such as images and videos. CNNs have revolutionized computer vision tasks by automatically learning hierarchical features directly from raw pixel data, through a self-optimization of the constituent neurons

(O'Shea, 2015). CNNs are composed by three types of layers: convolutional, pooling and fully-connected layers. These layer types constitute the CNN architecture, being organized in an input layer that takes and adapts data to the CNN requirement, a hidden layer where calculations are performed (with convolutional and pooling layers) and an output layer that deliver the outcomes of calculations and extractions (typically a fully-connected layer). A CNN is a feedforward network where the neurons that compose one layer are interconnected with the ones of next layer, basing on the quantitative similarity results (scores) produced at each step by activation functions. Activation functions introduce the non-linearity into the neural network, allowing the capturing of complex relationships within data. Common activation functions include sigmoid, softmax, ReLU (Rectified Linear) and its variants.
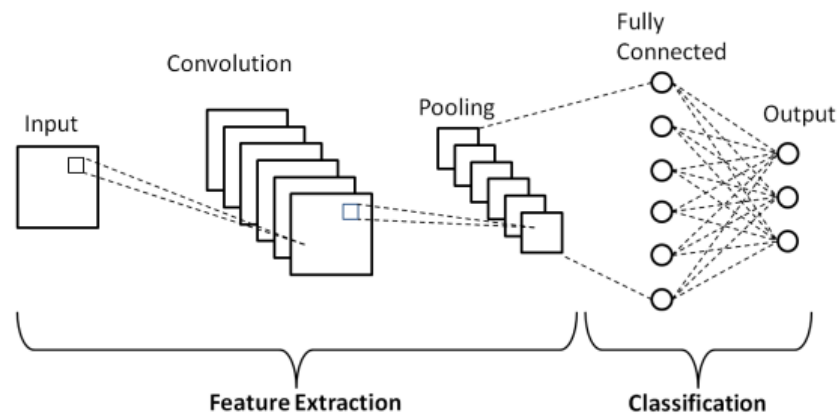
Figure 3: CNNs architecture (Balaji, 2020)

Convolutional layers apply filters (also called kernels) to the input image in a sliding window manner. Each filter extracts specific features from the image, such as edges, textures, and patterns and the results of these convolutions are passed to subsequent layers. The feature map produced by these filters is location-dependent, recording the precise positions of features in the input. What pooling layers provide is the "translational invariance" which makes the CNN invariant to translations, so that, even if the input of the model is translated, the CNN will still be able to recognize the features in the input. Pooling layers achieve this by simplifying the input image with an action of downsampling (Figure 4). With the gradual shrink of the image dimensions, the number of parameters and computations in the network is minimized. Max pooling and average pooling are common techniques that reduce the sample size of the data by considering, respectively, the maximum and average value within small regions. After passing through multiple convolutional and pooling layers, the extracted features are flattened and passed into the fully-connected layer, named in this way because neurons have full connectivity

within input and outputs. Here neurons apply a linear transformation to the input vector, calculating the dot product between the input vector (e.g. in Figure 5 of dimension 1x9) and a matrix of weights (e.g. 9x4); then a non-linear transformation is applied through an activation function, that returns an output (e.g. of dimension 1x4) of regression or classification type, basing on the purpose of the faced task. This procedure can be expressed by the following formula:

$$y_j(x) = f\left(\sum_{i=1}^{N} w_{ij} x_i + w_{j0}\right)$$

where x is the input vector (with elements x_i), y_j are the elements of the output vector y, N is the length of the input, w_0 are the bias terms and f is the applied activation function. As a result, all possible connections layer-to-layer are present, meaning every input of the input vector influences every output of the output vector (Unzueta, 2022).
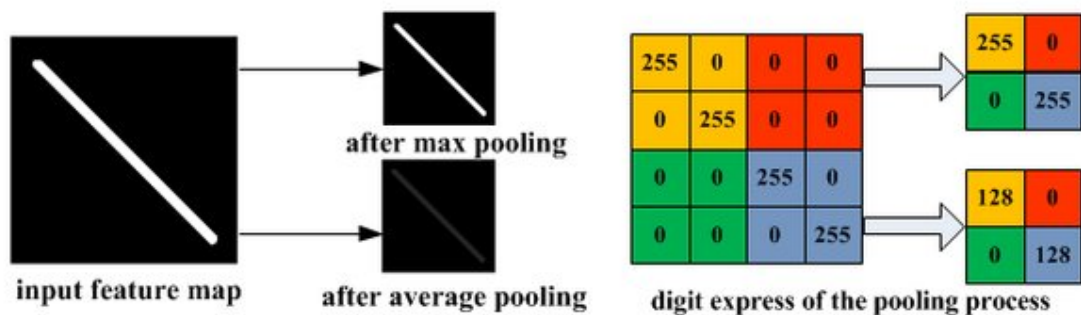
Figure 4: Illustration of max and average pooling (Yu, 2014)
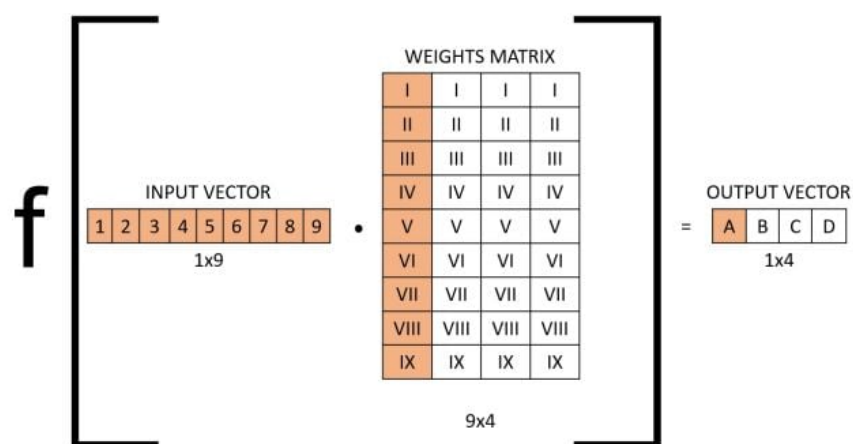
Figure 5: Illustration representing the fully connected layer, in which the inputs is multiplied by the weights matrix to receive the output vector (Unzueta, 2022)

The combined use of CNNs and Remote Sensing have recently begun to be used also in environmental studies concerning water bodies. It was proposed by Pu et al. to classify water quality of Chinese inland lakes from Landsat-8 images. A 4-layered CNN was developed by attributing to the images a water quality level basing on the data collected from official website and using a transfer learning strategy for dealing with the lack of measurement information. The results of CNN outperformed the ones of Random Forest and Support Vector Machine in the classification task (Pu, 2019). Wang et al. extracted urban water from Landsat imageries by combining a Multi-Scale Convolutional Neural Network (MSCNN) with Google Earth Engine (GEE), using satellite frames over China as case studies. The methodology proposed to train a complete MSCNN model offline to then extract water in urban areas online on GEE, by using the model previously trained. The Offline Training Online Prediction (OTOP) method was concluded to be accurate for thematic map extraction on different temporal epoch and spatial location (Wang Y. D., 2019). Also Chen et al. proposed a method for urban water extraction using Self Adaptive pooling Convolutional Neural Network on a dataset of Chinese high resolution multispectral images of ZY-3 and Gaofeng-2 satellites. The images were firstly segmented into superpixels using an iterative clustering method, for then functioning as input of the CNN, aimed to extract water body in an innovative way and to distinguish shadow pixels from and water ones (Chen, 2018).

Similarly did Li et al. using very high resolution Gaofeng-2 satellite images (0.8 m spatial resolution) to extract water body in the south of the Beijing metropolitan area using Fully Convolutional Network; even if trained on a small amount of labelled data, the model significantly outperformed the Normalized Difference Water Index, the Support Vector Machine and the Sparsity Model based methods in the water body extraction procedure (Li, 2019). In the research of Parajuli, surface water features from Sentinel-2 images were extracted using state-of-art approaches of deep learning, for finally integrate them with a new approach having better performance. In fact, the CNN created by integrating DenseNet and AttResNet resulted to be more effective than index-based neural network in extracting water from riverbanks and small surfaces (Parajuli, 2021).

2 Methodology

2.1 Study area

The study area comprehends reaches of two rivers in the Valencian Community, Spain: Rio Palancia and Barranc del Carraixet. River's hydrographic basins are included in the Demarcación Hidrográfica del Júcar (DHJ), geographically located in the extreme Central East of the Iberian Peninsula, with an area of 42'851 km². This territory is hydrologically managed by the Confederación Hidrográfica del Júcar (CHJ) and spans five autonomous communities (Figure 6): the Valencian Community, with a territorial share of almost 50% of the total area, the Community of Castilla-La Mancha with 37.6% share, Aragon with 12, 6%, Catalonia with 0.20% and Murcia with 0.15% (CHJ, s.d.). Geomorphologically it is described as an inland mountainous area, with points and chains of high altitude that act as barrier forcing the moisture-laden clouds mass coming by the sea to rise to higher atmospheric layers for then favouring precipitation, and a coastal area, made up of plains commonly known as 'planas', comprehending formation as lagoons (like Albufera), marshes and wetlands. Over CHJ there is a typical Mediterranean climate, with warm summers and mild winters. For the Köppen-Geiger classification the climate over Valencian Community is of type "BSk", meaning cold semi-arid (steppe) climate (Mindat, s.d.). Average annual temperatures range from 14 to 16º C and the thermal maximum is recorded in July and August during the dry season. The average annual rainfall is 500 mm, but there is large spatial variability, with values of 300 mm in the southern regions and other areas where annual means greater than 750 mm are reached. The rainfall events occurring with great intensity and short duration are more frequent in autumn and a secondary in spring. Most of the surface of the CHJ territory is covered by permeable materials which promote infiltration of surface water into underground strata (Hispagua, 2021).

In the DHJ, 304 water masses have been defined in the river category, of which 257 correspond to natural rivers, 43 to highly modified water masses (16 assimilable to rivers and 27 to reservoirs) and 4 to artificial water masses. The total longitudinal extension of this net is 5467 km and the two main rivers are Turia and Júcar, respectively of 512 and 280 km length. These 300 water masses (the 4 artificial water masses were not

considered) were classified following the system defined by the order ARM/2656/2008 of the Spanish Government. The classification results counted 216 water masses as perennial, 10 seasonal (or temporary rivers), 4 intermittent (or strongly seasonal) and 70 ephemerals. High number of NPRs stands out in the water courses that are located to the north of the town of Castellón de la Plana and on the left bank of the middle section of the Júcar (CHJ, s.d.). It's evident that NPRs have a great importance in CHJ, representing more than the 20% of the rivers type, so that it's fundamental to study their behaviour for setting compliant legal, environmental and hydrological objectives (CHJ, s.d.).

The selection criterion for rivers and underlying reaches to consider in this study was based on both quality factors, such as frequent and significant phenomena of ceasing of flowing along the year, goodness of Sentinel-2 visualization over the area (e.g. images without shadows in river corridor), satellite detection of water in past events, channel width, distance of the study area from big anthropic intervention, and on practical factors, such as the channel reachability for ground truth data collection and the proximity to meteorological stations for having representative associated ground-recorder-data. After a preliminary field surveys stage, two reaches of Rio Palancia and one reach of Carraixet were chosen.

Figure 6: Territorial shares of Demarcación Hidrográfica del Júcar (CHJ, s.d.)

Figure 7: Classification of surface water bodies over Demarcación Hidrográfica del Júcar (CHJ, s.d.)

Rio Palancia is within the natural rivers, characterized as "mineralized river of low Mediterranean mountain" from the system established by the WFD in its Annex II and developed in the Hydrological Planning Instruction - IPH, and "perennial" according to the order ARM/2656/2008 of IPH. It has a basin of 976 km^2, in which the river develops for a length of 90.7 km, for ending in the sea with a typical alluvial fan (CHJ, s.d.). The river valley is located between the Calderona and Espadà mountains, with a typical Iberian orientation (NW-SE). At the headwaters the river is semi-perennial, with flow provided by local perched aquifers. The Regajo dam (6.6 Mm^3) has regulated the flows in the upper basin since the year 1959 (mean river flow is 1.3 m^3/s) and supplies the irrigation canals of the coastal plain. The Algar reservoir (6 Mm^3) was built in year 2000, 15 km downstream the first dam, in order to control floods and recharge the coastal plain aquifer; however, it is always empty due to impounded area permeability. Downstream the dams, the river becomes ephemeral in its last 25 km, only carrying continuous flow after heavy rains (Sanchis-Ibor, 2017).

For studying purposes in single channel river, it is considered representative a sub-portion of the river, called reach, included between 10 and 20 times the average width or the active channel (Rinaldi, 2014).

The first considered reach was the segment BC ("Palancia 1"), long 2.7 km, being more than 10 times the width of the channel that is of 230 m, with a slope of 0.81% and human pressure of medium level. The second reach is the segment AB ("Palancia 2"), with an approximative length of 2.6 km for an average river width of 120 m, a slope of 0.59% and low human pressure. Both reaches are classified as sinuous and confined, with bedload sediment of medium size made up by cobbles and gravels (40–80 mm of diameter). Three meteorological stations have been identified in the surrounding of the selected reaches, reporting a mean annual precipitation of 500 mm. Gauging station data from Sistema Automático de Información Hidrológica (SAIH, s.d.) are only available in correspondence of Regajo dam, detecting a mean annual release of 0.4 m^3/s (CHJ and TYPSA, 2020).
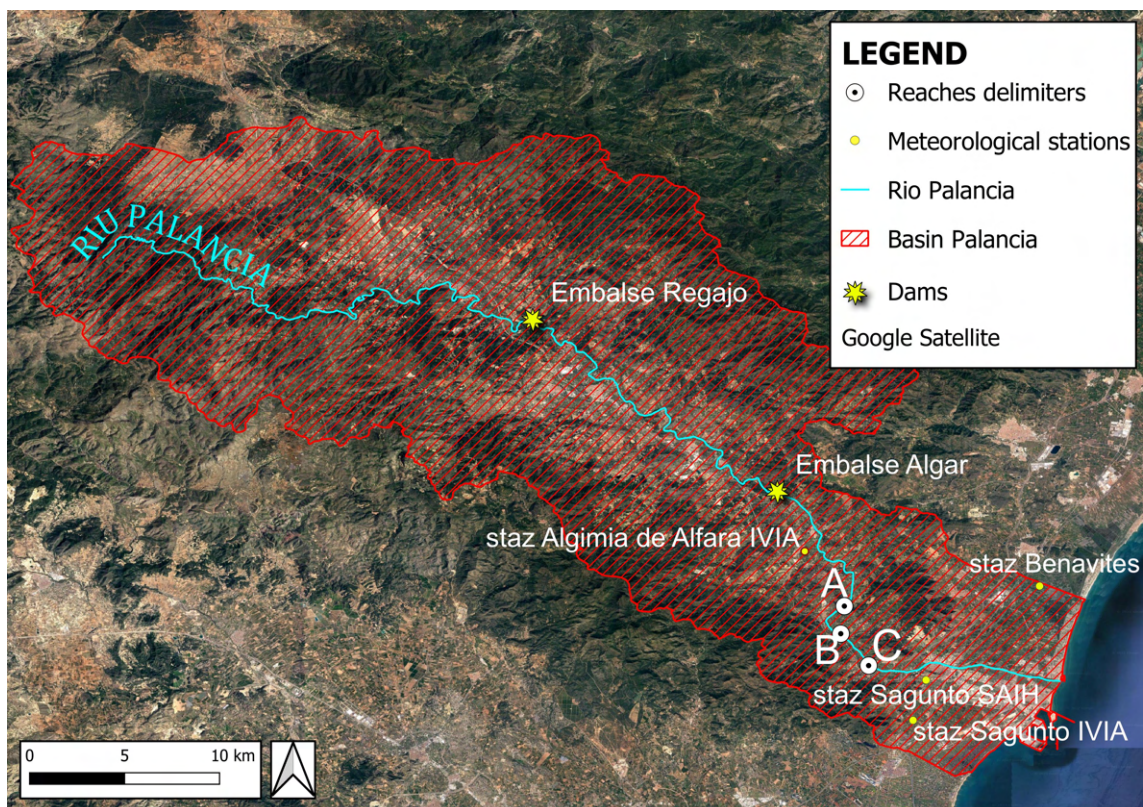
Figure 8: Rio Palancia case study

Figure 9: Location of reaches Palancia 1 - BC and Palancia 2 - AB

The Barranc del Carraixet is a natural river, characterized as "mineralized river of low Mediterranean mountain" and "ephemeral". It is 33.3 km long and drains a basin of 311 km^2 (CHJ, s.d.). The stream flows into the Mediterranean Sea passing in the north of Valencia, so that the catchment includes multiple areas of intensive land use. The upper river basin is dominated by coniferous forests and transitional woodland-shrub areas, the middle one by natural grasslands and sclerophyllous vegetation, while urban and agricultural uses dominate the lower basin with high presence of fruit trees plantations and complex cultivation patterns. Frequently absent base flow characterizes the hydrology, as common in Mediterranean ephemeral streams (García, 1999).

The segment AB represents the reach under study ("Carraixet"). It is located along the intermediate area of transition of the river, that connects the headwaters to coastal plains. The reach has a length of 1.6 km for an average width of 80 m, a slope of 0.34% and medium human pression, due to water abstraction for irrigation. Here the channel is confined and sinuous, geomorphologically with a prevalence of limestone and sandstone relief (Camarasa-Belmonte, 2016). Three meteorological stations have been identified in the river basin, detecting an annual rainfall that ranges approximately between 270 and 500 mm. In the hydrological year 2022-23 a mean flow of 0.01 m^3/s has been registered

at the SAIH gauging station of Carraixet-Bétera, located 8 km upstream the study reach (SAIH, s.d.).
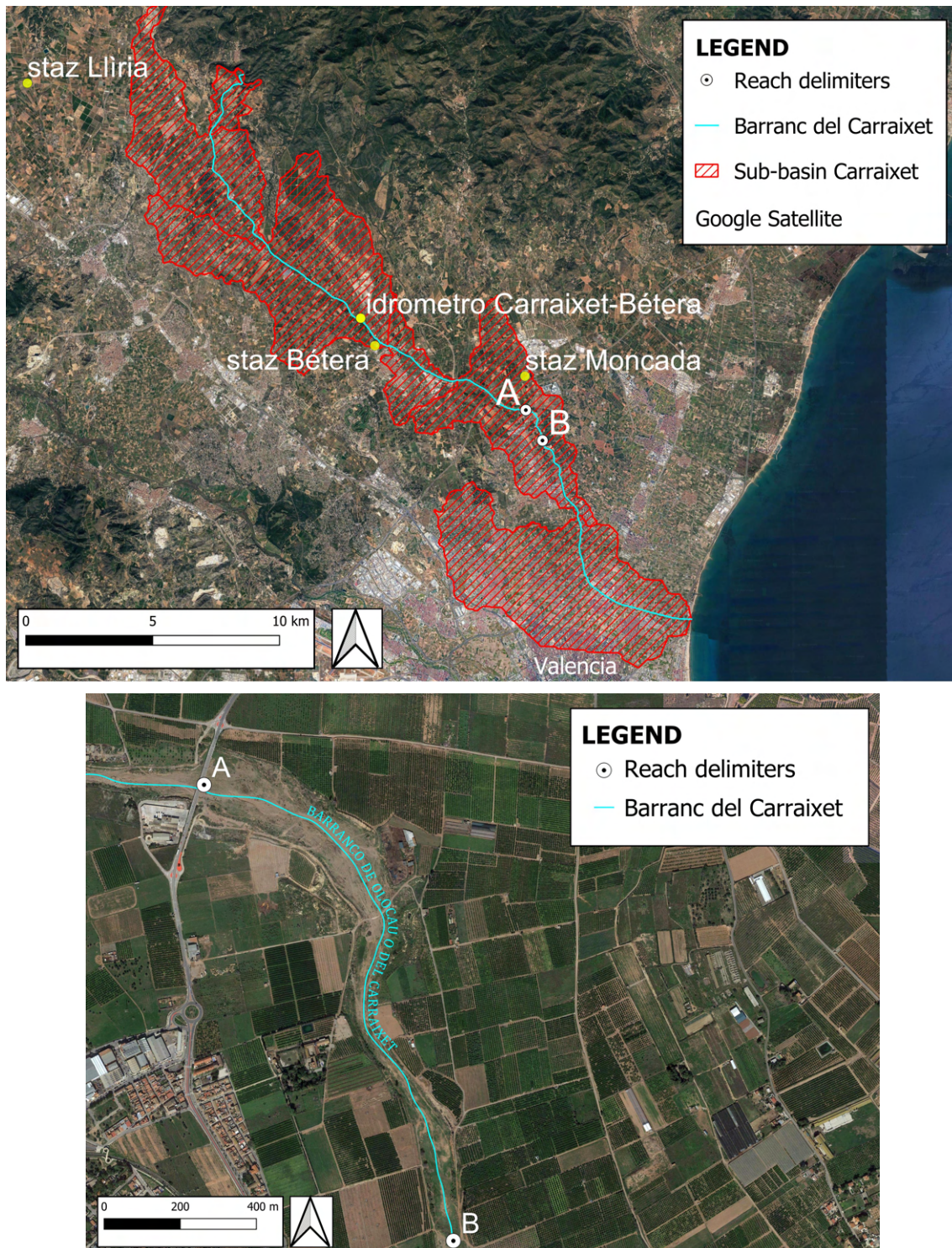
Figure 10: Barranc del Carraixet case study (above) and location of reach Carraixet – AB (below)
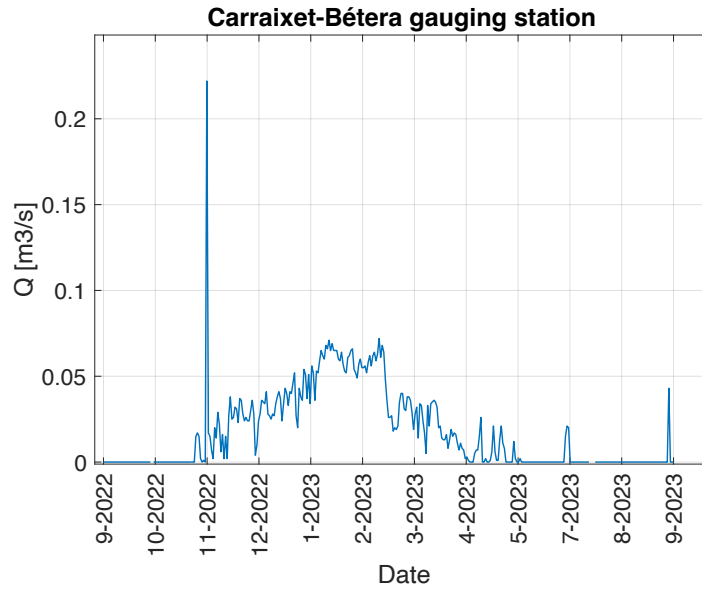
Figure 11: Measured flow rate at the Carraixet-Betera gauging station, h.y. 2023
(SAIH, s.d.)

2.2 Ground truth

The field surveys were executed preliminary for choosing the study area and consequently for collecting data and having references for assessing flowing conditions. In particular, photographic documentation and riverbed-cover polygons mapping, were gathered in order to have ground truths of the studied areas. The identified riverbed-cover classes were water, sediments, vegetation-bushes and vegetation-grass. The spectral signatures (SS) of these classes have been extracted thanks to the overlapping of on-site acquired Regions Of Interest (ROIs) and Sentinel-2 satellite images. For this reason, when possible, the field surveys have been executed in cloud-free days, when the passage of one of the Sentinel-2 satellites was scheduled. For georeferenced polygon acquisition the following instrumentation was used:

- Laser rangefinder
- Rugged computer with GIS software and Map Stream plug-in integration
- Tripod
- Device with GPS (e.g. smartphone)

Figure 12: Field survey instrumentations (left) and example of rugged computer (Nomad, center) and laser rangefinder (Laset Tech, right)

The ISPRA lineguides for hydromorphological units acquisition were followed for polygons detection (Vezza, 2017). Very high-resolution images, provided free of charge by Google Earth Pro, were used as additional ground truth for SS and features extraction, in order to have a riverbed-cover description for every season.

2.3   Sentinel-2 imagery

Sentinel-2 multispectral images have been used for this study, thanks to their optimal match between temporal and spatial resolution and the open-access data availability.

On the studied areas the revisit time is of 5 days, with a spatial resolution from 10 to 60 meters on the ground and a wide -swath width of 290 km. The Sentinel-2 mission is part of the Copernicus Earth Observation program led by the European Commission and operated by the European Space Agency (ESA). This mission comprises a constellation of two polar-orbiting satellites. The first one, Sentinel-2A, was launched on 23rd June 2015 and provides images with a revisit time of approximately 10 days at the equator; since the launch of the second satellite, Sentinel-2B, on 7th March 2017, the overall revisit time has become of 5 days under the same viewing condition (European Space Agency, s.d.). The mean orbital altitude of the satellites is 786 km and the mean local solar time at the descending node is 10:30 AM, this value chosen as a compromise between a suitable level of solar illumination and the minimization of potential cloud cover (European Space Agency, s.d.). Both satellites are equipped with an opto-electronic Multispectral

Instrument (MSI), able to acquire a total of 13 spectral bands in different spectral regions: visible (bands 1-2-3-4), red-edge (bands 5-6-7), near-infrared (NIR, bands 8-8a), shortwave infrared (SWIR, bands 9-10-11-12). The bands have different spatial resolutions and their values are reported in the following table.

Wavelenghth range [nm]	Spatial resolution [m]			Spectral region
	10	20	60	
423–463			B1	Coastal aerosol
458–523	B2			Blue
543–578	B3			Green
650–680	B4			Red
698–713		B5		Red Edge
733–748		B6		Red Edge
773–793		B7		Red Edge
785–899	B8			NIR
855–875		B8a		NIR narrow
925–965			B9	Water-Vapour
1350–1410			B10	SWIR-Cirrus
1565–1655		B11		SWIR
2100–2280		B12		SWIR

Figure 13: Characteristics of Sentinel-2 MSI, (Cavallo, 2022)
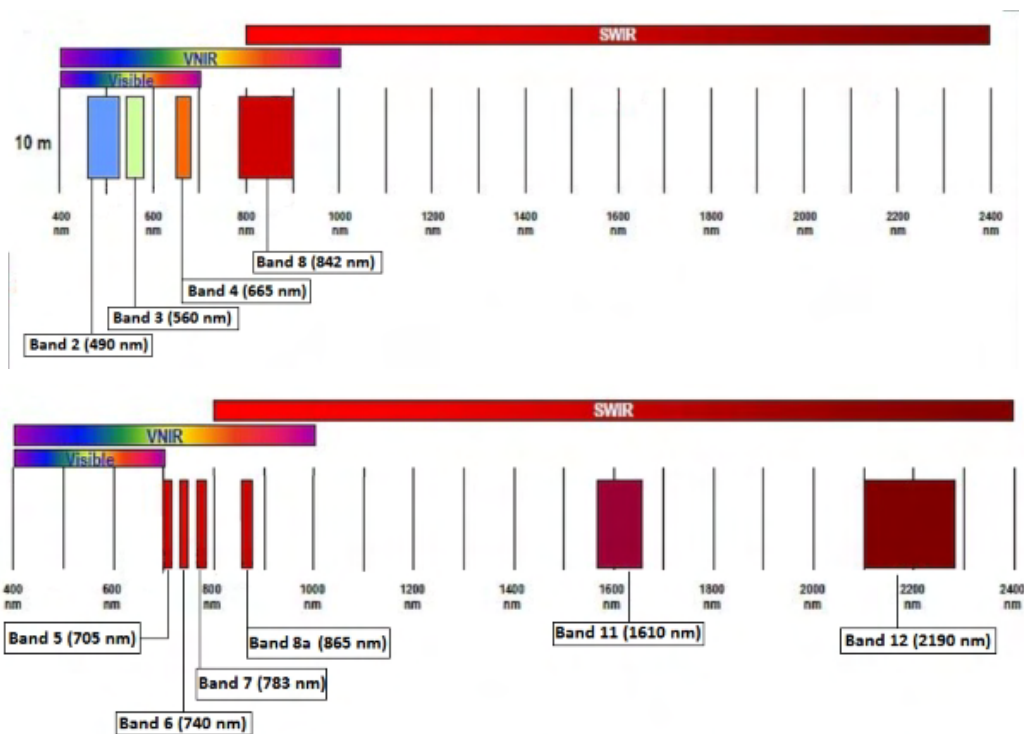
Figure 14: Spectral distribution of 10 m (above) and 20 m (below) spatial resolution bands, (European Space Agency, s.d.)

The Copernicus Open Access Hub from ESA provides free and open access and downloading of Sentinel data (ESA - Copernicus HUB, s.d.). Sentinel Hub EO Broswer, instead, provides easy browsing, visualization and consultation of Sentinel-2 imagery (Sentinel HUB, s.d.). The images of "Sentinel-2 MSI: MultiSpectral Instrument, Level-2A" dataset have been accessed directly from Google Earth Engine (GEE, s.d.), a computing platform for Earth science that allows users to access data and run geospatial analysis at planetary-scale. The Level-2A dataset contains orthoimages already atmospherically corrected, being Surface Reflectance (SR) or Bottom of Atmosphere (BOA) product. From the code editor, available in Python and JavaScript, images of the dataset on required dates have been download and used for the analysis. The shapefiles representing the classes polygons collected in field survey or drawn on high resolution images directly on Google Earth Pro, have been imported in GEE environment and overlapped to the Sentinel-2 image of correspondent (or closest) date. Spectral signatures represent reflectance in percentage plotted versus the bands or related central wavelengths. The reflectance response of water, sediment and vegetation (split into bushes and grass) classes have been extracted for all the available season, in order to catch the seasonal reflectance variation of the river corridor elements. Once obtained, the combination of bands that best allowed the discrimination of the identified classes was chosen as RGB triplet for False Color Images (FCIs) visualization. The FCIs obtained in this way were then used into the following steps for reducing the chance of possible images misclassifications.

Using the Google Colaboratory, a cloud-based Jupyter notebook environment that execute phyton code, a six-year dataset of the available images over the study reaches, for the period April 2017- May 2023, has been created cutting the images with masks of reaches' active channel and manually determining (in a supervised mode) the occurrent flowing and cloud cover conditions. Also levels of confidence, distinguishing between high and low, have been attributed during the classification process. The FCIs were stored in organized nested-folders as GeoTiff images, named with the collection date in the format "yyyymmdd.tif" and with related metadata, distinguishing the Sentinel-2 FCIs in 4 classes: "cloudy", "dry", "flowing" and "ponding". For the supervised classification, the criterion for flowing condition definition has been:

- Dry (D) - absence of water along the whole segment as visible from FCI
- Flowing (F) - continuity of water flows along the whole segment
- Ponding (P) - at least a trait in which the continuity of flow is interrupted as visible from FCI

Cloudy images have been stored as side class (C) and used only in one of the classification tasks, since the lower interest in this image type under the hydrological point of view. To standardize the image classification process, the same interpreter classified the whole archive at the fixed scale of first visualization of 1:30'000.

2.4 CNNs

2.4.1 Pre-processing

Google Colaboratory (Colab) was used for code and model development, since it is free to use, requires no setup and has direct access to images and data stored in cloud (e.g. Google Drive), with files that can be simply shared and edited. It also contains as pre-installed TensorFlow, an end-to-end open-source platform for machine learning that has now incorporated Keras, the most popular high level Application Programming Interface (API) for building and training neural network. A bench of related libraries can then easily be imported. By default, Colab notebooks run on CPU, but it can be switched to GPU or TPU for significant speed up of the processes. Datasets have been handled for matching the CNN input data requirement. The most common image formats used for classification with CNNs are JPEG (or JPG) and PNG. The choice of the one to be used for a classification task depends on the size and complexity of the dataset, the memory and storage constraints of the system and on the type of application. PNG is the format that better preserves all the details of the original image, generally being of larger size compared to JPEG files. PNG is also a good option when the dataset is small and the size of the image size is not a major concern. Due to these reasons, GeoTiff have been converted into PNG format and subdivided into folders according to the classification task to feed. In this phase only the images with "high confidence" of classification have been used. Dataset creation and organization is a step of paramount importance in CNN applications. For each river reach, the following subfolders were prepared for their subsequent use into the models:

- 2 classes: nonwater, water (NW, W)
- 3 classes: dry, flowing, ponding (D, F, P)
- 4 classes: cloudy, dry, flowing, ponding (C, D, F, P)

Dataset were loaded directly from cloud, in a way that class names and successively labels are automatically inferred to the images, basing on subfolder name. Class names are

reported in alphabetic order, since is in this way that the model attributes the label indexes (e.g. C [0], D [1], F [2], P [3] for 4-classes reading).
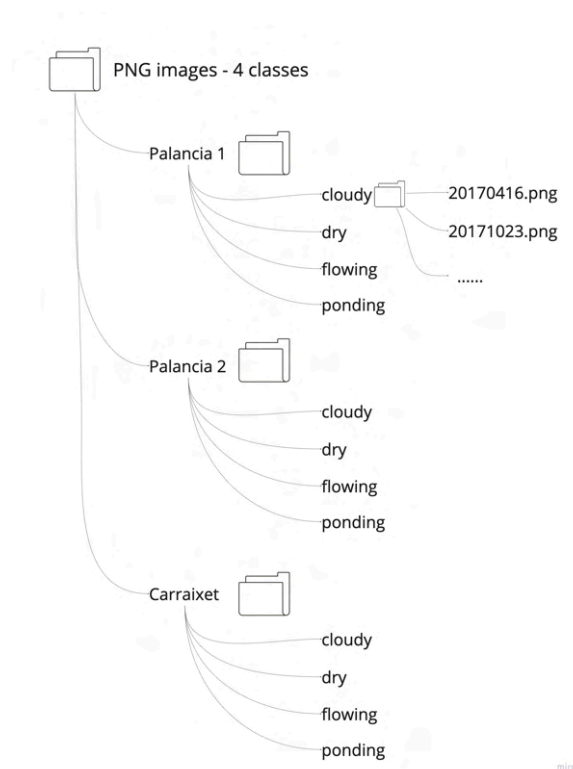
Figure 15: Tree-folder organization of the 4-classes dataset

The images were further handled and pre-processed to meet the CNNs input requirements. The pre-processing steps have been (1) dataset exploring, (2) image batch grouping, (3) image resizing, (4) oversampling, (5) dataset split into train, validation and test subset, (6) data augmentation and (7) prefetching.

While assessing pre-processing the common practice is to start from data visualization; the dataset can be explored by plotting some of the contained images for image size check, object data type and pixel content in terms of Digital Number (DN).

Consequently, since CNNs are typically used to process large images and processing one image at a time can be computationally inefficient, the images are grouped in batches: by processing multiple images in parallel as a batch, it's possible to optimize hardware utilization. The number of the image contained in each batch, and so processed at once at every CNN epoch before updating the weights and biases during the training process, is defined by the batch size. Large batch sizes can lead to faster training times because the CNN updates its parameters less frequently, but they also require more memory and computational resources; smaller batch sizes may be more computationally efficient but requiring more training time to converge. The batch size is a power of two and commonly

it is chosen in the range 16 – 512, with size of 32 as rule of thumb for good initial choice. Once grouped in batch, the images can be more easily handled in the other pro-processing steps.

A CNN algorithm requires that the input objects all have the same size, so that image resizing is a compulsory step. Once extracted the major image height and width, the resizing has been conducted by adding a black pad to the objects of smaller dimension. In this way, the original images resolutions were maintained, without applying zooming or compression to the pixel, and the images homogenized according to the purpose.

Once obtained in homogeneous size, the images were split into the train, validation and test subsets (Figure 16). The training dataset (Train DS) contains sample of data used to fit the model, the validation one (Val DS) the ones used to provide an unbiased evaluation of a model fit at the end of every training epoch, tuning model hyperparameters, and the test one (Test DS) the samples used to provide an unbiased evaluation of a final model (Brownlee, 2020). Test DS is only used once the model is completely trained on both Train and Val DS and represent a small number (with respect to the other subset) of unseen images that the model must classify basing on previous learning. The dataset split was achieved by setting a percentage for each subset and randomly assign dataset objects to them; percentages of 0.7, 0.2 and 0.1, respectively for training, validation and test subset, were chosen.

Another step of paramount importance is the training dataset balance, since a model trained on an unbalanced set can led to false high accuracy performance, only predicting all the images as the majority class. Only the training subset was balanced in order not to have images in the validation and test sets that the model had already and previously seen. The number of images per each class in the training dataset was counted for then performing the oversampling procedure, that involves a random resampling of the objects of the minority class (or classes) to match the same initial condition into the learning process. Generally, for small dataset, the oversampling approach is preferred comparing to other balancing method like undersampling, that is instead useful for large amount of data since it reduces the number of objects contained in the majority class (or classes).

In order to reduce the risk of overfitting, data augmentation was applied on the training subset. This technique artificially introduces sample diversity by applying random but realistic transformations to the training images, such as rotation and horizontal flipping. This practice is strongly suggested for small dataset since it exposes in a variated way the images to the model, so that it can learn and recognize features even if they have different spatial location.

Finally, prefetching during image loading was set. This technique consists in overlapping the pre-processing and model execution during a training step and it allows to reduce the computing time and to increase data performance. If tuned on, while the model is executing a training step s, the input pipeline is reading the data for step s+1, speeding up the loading process and the overall execution (TensorFlow, s.d.).
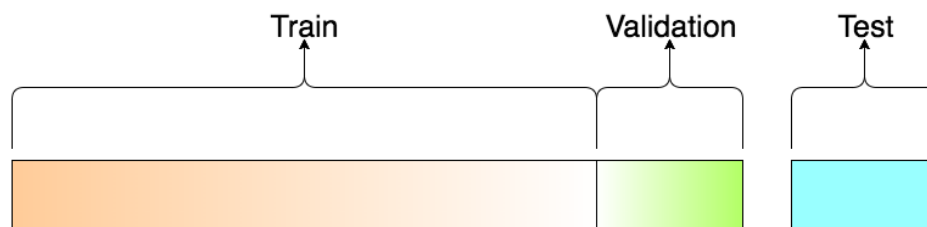
Figure 16: Generic split into Train, Val and Test subsets (Shah, 2017)

2.4.2 Model construction

Dealing with image classification tasks, a reasonable number of samples for training an algorithm is 10'000. When having too little data to train a full-scale model from scratch, it's common to construct own model starting from a pre-trained network, that is a model constructed on large and general enough dataset. This technique, known as transfer learning, consists of taking features of the visual world learned on one problem, and leveraging them on a new, similar problem. It directly addresses the smart parameter initialization for training neural networks, allowing to speed up the training and to overcome small dataset size (Prakash Maheswari, 2019).

CNN models have been constructed by following the workflow of transfer learning application.

At first was selected the pre-trained model to be used as base model. For image classification task, the Residual Network (ResNet) models are within the best in terms of performance (Huilgol, 2020). In particular the ResNet50 (Keras, s.d.) was used as base model for the CNN construction and it has been pretrained and weighted on ImageNet, a large research-training dataset consisting of 1.4M images belonging to 1000 different classes. During this pre-training, only the bottleneck layer was used for feature extraction. In fact, a common practice is to focus on the very last layer before the flatten operation, the bottleneck layer, that is ideal for feature extraction since it retains more generality as compared to the final/top layer, that instead is not very useful (TensorFlow, s.d.). This is achieved by specifying in the base model definition 'include_top=False' as a function argument.

After pre-training the convolutional base was freeze, in order to maintain the learned patterns, preventing the weights of the layers from being updated during the consecutive training on the target database.

At this point, the proper model was built by chaining together pre-processing steps (constituted by the input and the sequential layers), the base model (constituted by a functional layer), and the classification head (that includes pooling layer, dropout layer and dense prediction layer). In fact, the classification head is formed by some new, trainable layers, that will learn to turn the old features into predictions on a new dataset. It is called head since it is directly incorporated on top of the frozen layers. During this built passage the model architecture was defined, inserting six layers.

Finally, the target dataset was used to train the new layers, by compiling and fitting the model with the Train plus Val DS, tuning learning rate and number of epochs in order to reach the best model performance. During model compiling and fitting important parameters, evaluation functions and metrics are set. Model compiling configures the network for training by defining the optimizer, the loss function and the metrics. The optimizer controls the learning rate, meaning how quickly the model's optimal weights are calculated. A smaller learning rate may result in more accurate weights (up to a point), but it will take longer to compute the weights. Adam optimizer was used, since it adjusts the learning rate in training throughout stochastic optimization and it is generally a good optimizer for many cases. A loss function is a function that compares the target and predicted output values, measuring how well the neural network is modelling training data. When training, the aim is to minimize the loss between the predicted and target outputs by adjusting the hyperparameters. Classification loss functions are the one used in classification problems and they comprehend Binary Cross-Entropy, for binary classification, Categorical Cross-Entropy and Sparse Categorical Cross-Entropy functions, this last two for multi-class predictions. The metrics parameter from which it's possible to choose during training and testing compiling are accuracy, precision, recall and F1-score. In model fitting, the train and validation dataset are assigned to the model and the number of epochs is specified. The number of epochs is a hyperparameter that defines the number times that the learning algorithm will work through the entire training dataset (Train DS plus Val DS). With 100 epochs, the model will be exposed to or pass through the whole dataset 100 times. It's a kind of for-loop over the training dataset. Within this for-loop is another nested for-loop that iterates over each batch of samples, where one batch has the specified "batch size" number of samples. It is common to create line plots, called learning curves, that show epochs along the x-axis as time and the

performance of the model on the y-axis. These plots can help to diagnose whether the model has over learned, under learned, or is suitably fit to the training dataset.

After the fitting step, the model has been initially evaluated on the Test DS in order to obtain the first performance on this subset.

Consequently, the fine-tuning was added. This procedure is optional, but it has been implemented due to its potentially to achieve meaningful improvements, by incrementally adapting the pretrained features to the new data. In fact, the fine-tuning consists in unfreezing a-part-of or the entire model obtained above for re-training it on target data with a very low learning rate. While in the early steps of transfer learning it's necessary to freeze the entire convolutional base, rendering it entirely untrainable and only using weight initialization, after having done the initial training it's possible to fine-tune the weights of the latter layers of the overall models, comprehending in the trainable layers also some of the base model. It's important to note that without training a classifier on the frozen base first, initial epochs would overwrite any useful representations encoded in the pre-trained model, resulting in an abundance of noise.

After this change the model was compiled again, with a learning rate of 1/10 of the one used in the initial stage, and re-fitted, so that it went on in the learning process. As last step, the obtained CNN was evaluated on the Test DS in order to extract the final performance scores on totally unseen data (Keras, 2020).

2.4.3 Evaluation metrics

Model evaluation is an essential part of every CNN. There are different types of evaluation metrics available for measuring the performance of a model and the most representative one depends on the type of problem faced and on the input features. As seen before, the loss is a base parameter during model compiling and fit. Accuracy, precision, recall and F1-score are then the mostly used evaluation metrics to assess the performance of classification models (Saturn Cloud, 2023).

Loss

The loss is calculated by loss function and represents the error or discrepancy between the predicted output of the model and the true label, quantifying how well the model is performing in terms of minimizing such difference. As already sad, the goal of the training process is to minimize the loss.

Binary Cross-Entropy or Log Loss (BCE) is the loss function used in binary classification models, where the model takes an input and has to classify it into one of two pre-set categories. It's calculated using the formula:

$$BCE\ Loss = \frac{1}{N}\sum_{i=1}^{N} -\left(y_i \cdot \log(p_i) + (1 - y_i) \cdot \log(1 - p_i)\right)$$

where N is the number of observations, i is the given observation, y the actual true value, p the prediction probability (p_i is the probability of the positive class, class 1, and ($1-p_i$) the one of the negative class, class 0) and log refers to the natural logarithm.

In binary problems, the positive class is the one with a higher level of significance for the faced problem, with the particular features to identify, while the negative class is its complementary (Yathish, 2022). The prediction probability is a value between 0 or 1 obtained by applying a sigmoid function as activation to the prediction logits, that instead are not in a fixed range, being unnormalized raw predictions. Positive logits typically represent the scores for the positive class (index [1]), while negative logits the scores for the negative one (index [0]). The sigmoid activation function is represented by:

$$\Phi(\vec{z}) = \frac{1}{1 + e^{-z}}$$

where z is the vector with logits score.
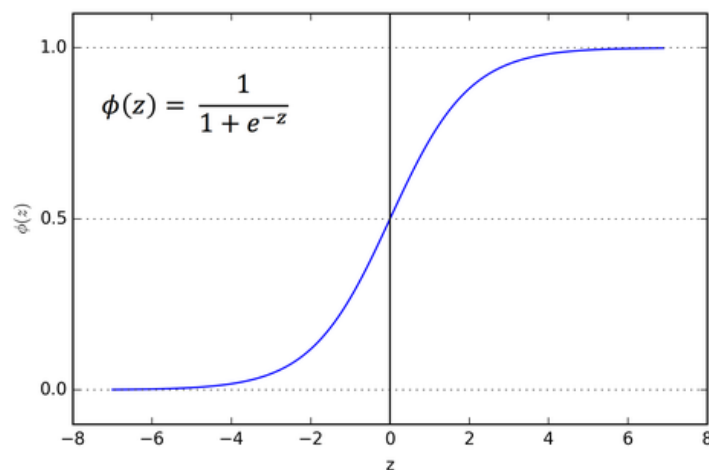
Figure 17: Sigmoid activation function (Sharma, 2017)

Categorical Cross-Entropy Loss (CCE) is the function to utilize in case the number of classes is greater than two, with an approach very similar to binary cross-entropy, but using and producing one-hot encoded labels. For example, the one-hot encoded label of an image belonging to the 3rd class out of 4, is [0 0 1 0]; the prediction vector for the

image may be [.3 0 .5 .2], where the inside values represent the belonging probability, summing up to 1, then converted to the prediction output [0 0 1 0].

$$CCE\ Loss = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} y_{ij} \cdot \log(p_{ij})$$

where N is the number of observations and M the number of classes. The softmax activation function is used for multiple classes tasks for the conversion of the logits vector into the prediction vector and it's expressed by:

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{M} e^{z_j}}$$

where $\vec{z}_i$ is i-th vector of logits in input to the softmax function $\sigma(x)$, M in the number of classes, e^{z_i} is the standard exponential function for input element and e^{z_j} the one for the output element, functioning as normalization term. The function returns a prediction vector with values in [0, 1] probability range.

Sparse Categorical Cross-Entropy Loss (SCCE) is the equivalent of the Categorical Cross-Entropy, but, instead of one-hot encoded vectors, uses integer labels, starting from class [0]. For example, the integer label of an image belonging to the 3rd class out of 4, is [2]; the prediction vector for the image may be [.3 0 .5 .2], where the values represent the index of belonging probability, from which is then extracted the position of the maximum argument and predicted [2]. It uses the same loss and activation function of CCE.

Confusion matrix

The confusion matrix gives a lot of information about the model's performance. It is the visual match between prediction and true sample, being a 2-D array of shape [n, n], where n is the number of valid classes for a given classification task. The matrix columns represent the prediction labels and the rows represent the real labels. In the simplest case of binary classification, the confusion matrix has 4 essential components (Figure 18):

- True Negatives (TN) - number of samples correctly predicted as "negative"
- True Positives (TP) - number of samples correctly predicted as "positive"
- False Positives (FP) - number of samples wrongly predicted as "positive"
- False Negatives (FN) - number of samples wrongly predicted as "negative"

Confusion matrix for multiple classes is similar, having on the diagonal the numbers of correctly predicted samples and in other positions the number of wrongly predicted ones. From the confusion matrix it's possible to understand which classes are wrongly predicted and in which sense, for further model improvement. Its components allow then the computation of precision, recall and F1-score metrics.
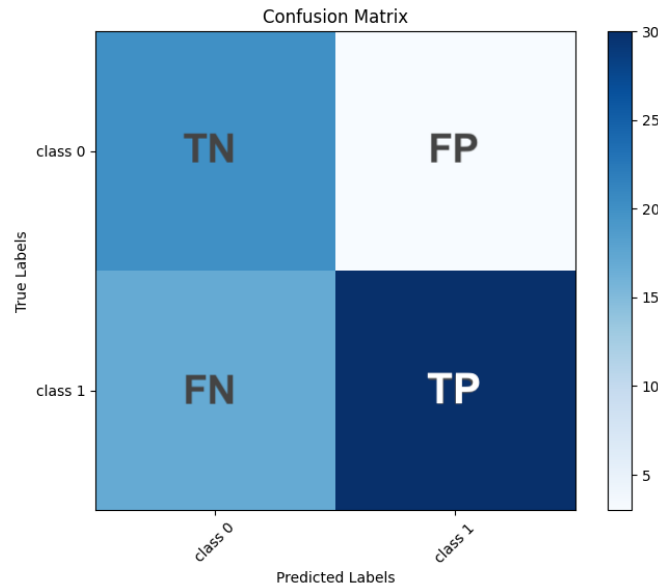
Figure 18: Confusion matrix of binary classification model with 'class 0' and 'class 1'; TrueNegatives (TN), FalsePositives(FP), FalseNegatives (FN), TruePositives (TP)

Accuracy

Classification accuracy, or simply accuracy, is the ratio of number of correct predictions to the total number of input samples. Considering the terms of confusion matrix, the accuracy is defined as the true terms over the total number of predictions, meaning the diagonal terms over the sum of all the terms. In training phase, as the model improves its performance, the loss should decrease, and the accuracy increase. Accuracy is a realistic metric only if there are equal number of samples belonging to each class. For example, considering an unbalance dataset with 98% samples of class A and 2% samples of class B on the training set, the model can reach 98% of training accuracy by simply predicting all the sample to class A. If the same model is evaluated on a test set with 60% samples of class A and 40% of class, the test accuracy falls to 60%. Hence, classification accuracy is a great metric, but returns the false sense of achieving high accuracy in unbalance problems.

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ prediction\ made}$$

Precision

Precision is a class specific metric. By considering the positive class, class 1, as the one under interest, the related precision is the number of correct positive results divided by the number of total positive results (made up of true and false positives) predicted by the classifier. This metric focus on the prediction column of the confusion matrix, only considering the class (and so the column) under interest. Also the precision of the negative class, class 0, can be retrieved considering in the same way the two negative terms. This metric stays in the range [0,1] and the expression for the positive class precision is given by:

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives}$$

Recall

Also recall is a class specific metric. Taking the expression for the positive class, it is defined by the number of correct positive results divided by the number of all relevant samples (all samples that should have been identified as positive, made up of true positives and false negatives). In the confusion matrix, the recall focus on the true label row of the class under interest. Hence, it can also be obtained for the negative class, class 0, by considering the true and false terms in the first row as substitutes of the ones in the second row of the confusion matrix. Recall stays in the [0, 1] range. The related formula for the positive class is:

$$Recall = \frac{TruePositives}{TruePositives + FalseNegatives}$$

F1-score

Precision and recall are not particularly useful metrics when used in isolation. F1-score is the harmonic mean between precision and recall of the same class and it takes into account both false positives and false negatives. It tells how precise the classifier is (how many instances it classifies correctly), as well as how robust the model is (not missing a significant number of instances). Maximizing the F1-score for having better performance, implies simultaneously maximizing precision and recall, two metrics that for their construction are in trade-off. Once again, the metric is class specific and range in the [0, 1] interval. It is mathematically defined as:

$$F1-score = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} = 2\frac{Precision \cdot Recall}{Precision + Recall}$$

3 Results and Discussion

3.1 Ground truth

Three field surveys have been performed on each studied river, considering the reach "Palancia 1" and "Carraixet", due to the easier reachability of their riverbed. High resolution images (orthophoto) of Google Earth Pro have been used as additional ground truth for SS and features extraction. The acquisition dates are reported in the table below, with specification of the ground truth type, belonging season and date of the closest available Sentinel-2 observation. When the ground truth acquisition and the satellite passage were not contemporary (the maximum time lapse was of 3 days), it has been checked that no precipitation event interposed, so that the only possible changes could be due to evaporation and consequent reduction in water surface area. For each river, rows of same color have been considered as gathered for the SS extraction, identifying, according to the available data, three main period: late winter, late autumn + winter and spring for Palancia; late winter, autumn and spring for Carraixet.

Table 1:Palancia - resume of ground truth source, date and related Sentinel-2 image (W:winter, SP:spring, A:autumn)

River	Acquisition	Ground truth source	Ground truth date	Season	Sentinel-2 date
Palancia	1	Google Earth Pro	11/03/21	Late W	11/03/21
	2	Google Earth Pro	27/05/22	SP	30/05/22
	3	Field survey	17/12/22	Late A	16/12/22
	4	Field survey	20/01/23	W	20/01/23
	5	Field survey	16/04/23	SP	20/04/23
	6	Google Earth Pro	03/05/23	SP	05/05/23

Table 2: Carraixet - resume of ground truth source, date and related Sentinel-2 image

River	Acquisition	Ground truth source	Ground truth date	Season	Sentinel-2 date
Carraixet	1	Google Earth Pro	11/03/21	Late W	11/03/21
	2	Google Earth Pro	07/11/22	A	06/11/22
	3	Field survey	26/11/22	A	cloudy
	4	Field survey	16/12/22	A	16/12/22
	5	Google Earth Pro	15/04/23	SP	15/04/23
	6	Field survey	24/04/23	SP	25/04/23

Summer period is out of the descriptor since no observational data were available in recent years for this season. In order to catch all the river's flowing conditions, the field surveys were conducted both in days after rain event, both in no precipitation period. An important precipitation event occurred over Valencian Community on November 10th-12th (P_{cum}=70 mm at Algimia de Alfara – Palancia reaches, P_{cum}=40 mm at Moncada – Carraixet reach) as shown in Figure 19. A trial of field investigation has been done in date 26/11 and no water was found in the rivers' channels. The matching satellite image, together with all the ones subsequent the event, was cloudy and no remote observations after this precipitation occurrence have been available. This fact highlights the unpredictability of working with satellite data, that strongly depend on sky-cover condition. A precipitation event of lower intensity interested the zone on December 11th-13th (P_{cum}=20 mm at Algimia de Alfara, P_{cum}=15 mm at Moncada). The first open-sky available image has been on December 16th and two field surveys have been conducted on dates 16-17/12, detecting flowing conditions of "ponding" for Palancia and "dry" for Carraixet. During the field survey period, no water was detected in Carraixet river channel.
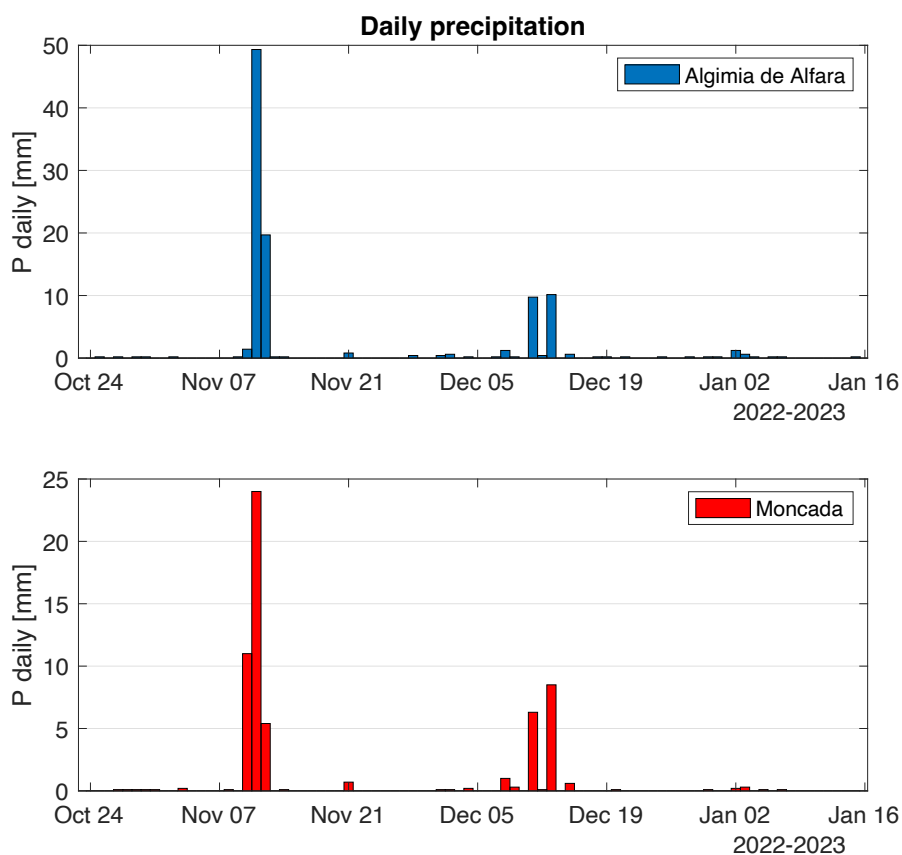
Figure 19: Recorder daily precipitation at Algimia de Alfara (reference for Palancia reaches) and Moncada station (reference for Carraixet reach) in period Oct 22 – Jan 23, (IVIA, 2023)

For Palancia reaches, the field survey on 26/11 showed that 15 days after the precipitation event of P_{cum}= 70 mm, the water was already disappeared, with river found in "dry" condition. As a result of the precipitation event of December, instead, water remained longer time, allowing "ponding" status, with ponds decreasing in size, until 15/01/23 (detected thanks to satellite images). Considering the absence of further precipitation intake, this phenomenon can be related to soil saturation and surface-groundwater interaction processes. Georeferenced polygons of various shape and linear dimension from 2 to 50m, representing the perimeter of found water bodies, sediment and vegetation riverbed-cover classes, were collected with the laser rangefinder.

The photographic documentation of the December 16th -17th field surveys is reported below. In this survey, a number of 10 georeferenced water ponds was acquired for Palancia (Table 3).
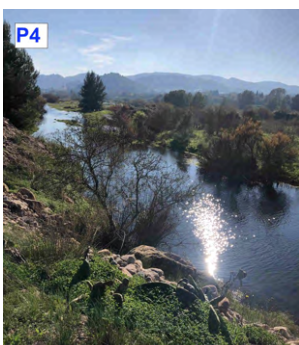
Figure 20: Field survey on Palancia, date 17-12-22 and related Sentinel-2 FCI of 16-12-22
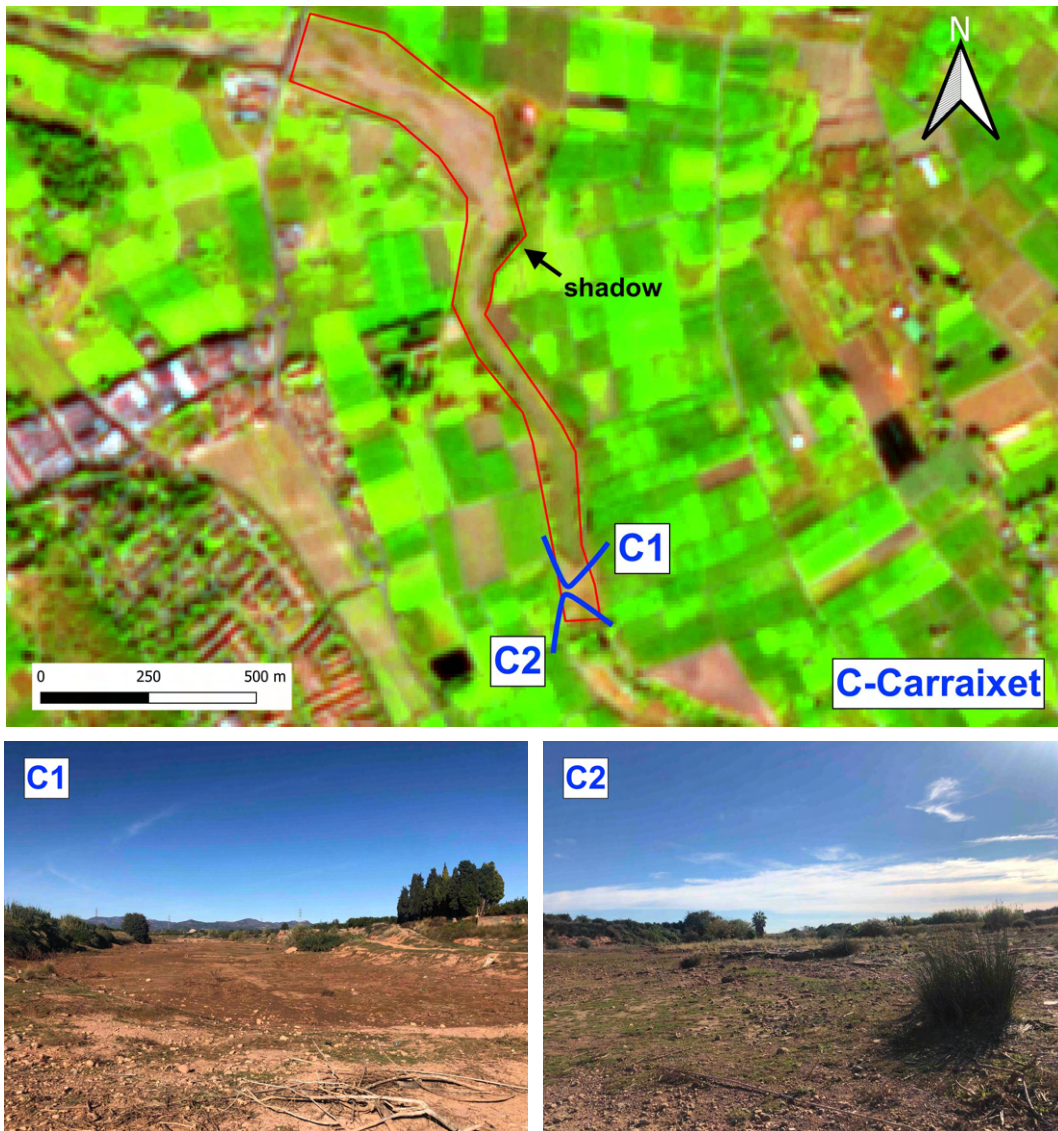
Figure 21: Field survey on Carraixet, date 17-12-22 and related contemporary Sentinel-2 FCI

In the table below are reported the dimensions of the acquired water pools of Palancia in the survey of December 17th. Average length and width are considered with respect to flow directions. The related positioning is shown in Figure 22.

Table 3: Pool acquisition for Palancia, date 17/12/22

Date	Pool n.	Average length [m]	Average width [m]
17/12/22	1	55	30
	2	60	11
	3	78	13
	4	112	34
	5	88	40
	6	2	7
	7	82	6
	8	34	10
	9	65	44
	10	93	15
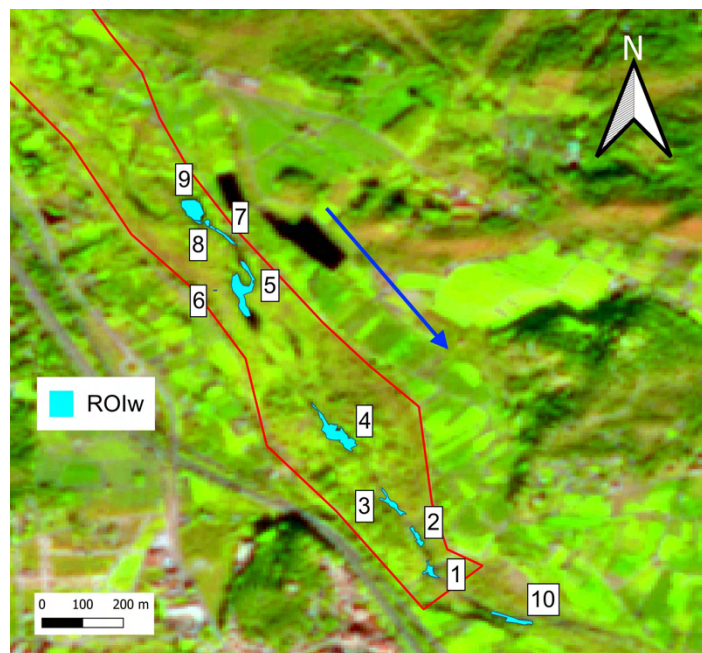
Figure 22: Distribution of pool acquired on Palancia in date 17/12/22

Recognizing NPRs, catching photos and collecting water coverage in different flowing conditions, helps to improve the environmental and social perceptions of this type of river, that has been historically overlooked and abandoned to degradation. The understanding of water heterogeneity in distribution along the channel and intermittency is necessary for a correct management of NPRs.

Figure 23: Palancia "ponding"-date 16/12/22 over "dry"- date 16/04/23 condition

Figure 24: Carraixet degradation, date 24/04/23

3.2 Sentinel-2 imagery

3.2.1 Spectral Signature extraction

Ground true georeferenced polygons of the defined classes were combined with the temporally closest Sentinel-2 image (Tables 1,2) in order to extract the spectral behaviour of the riverbed-cover classes. Bands are representing the layers of a multispectral image. Each layer is made of pixels, that contain, according to the used source database, values of surface reflectance. By considering the pixel under a defined ROI, the SS of the object

over the available bands can be extracted (Figure 22). Since every pixel represents an area on the ground of 10 m x 10 m and it is not possible to distinguish sub-pixel objects, the SS extraction is of a generalized frame. The analysis showed that the minimum width of ponds identifiable by FCIs is highly variable. Water objects with a linear dimension lower than 7 m were not visible from the FCIs. In some cases, wet channels and ponds with widths between 7 and 10 m were identified, while in other, wet channels of width between 10 and 15 m were not detected. This strongly depends on the relative positioning of the object with respect to the pixels: when a water surface is centered and spread over a single pixel, the identification is well performed, while if it covers only partially two or more pixels, the identification is more complex.

Palancia's seasonal spectral signatures of available reference data are reported below. Late winter extraction is matching the Google Earth Pro ground truth with the contemporary satellite passage on 11/03/21. Late autumn and winter refer to the dates 16/12/22 and 20/01/23. Spring 2023 (date 03/05/23) was compared with spring 2022 (27/05/22) due to the strong difference in land cover among these two years. In fact, low precipitation on the hydrologic year 2022-23 caused the dry up of soil and vegetation.
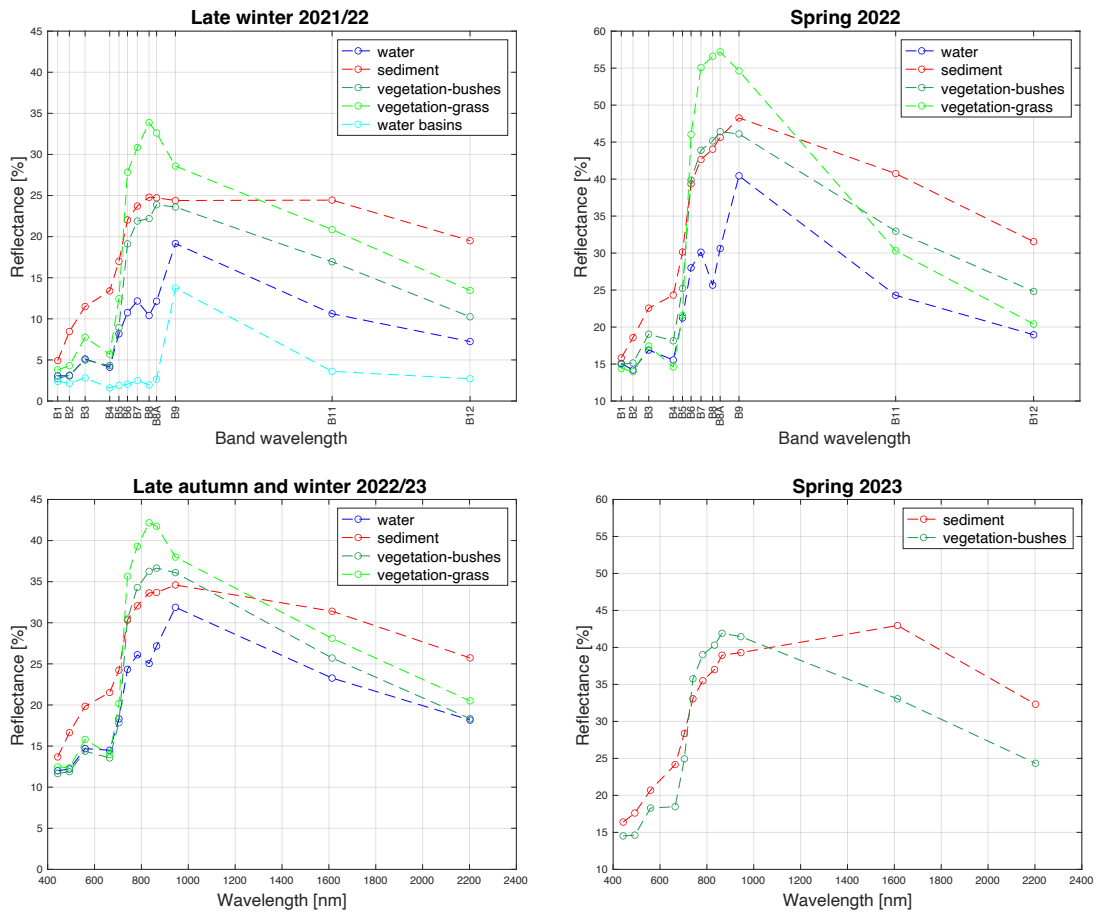
Figure 25: Seasonal Spectral Signatures, Palancia

As shown in Figure 25, on late winter the reaches were in "flowing" condition and SS of water in wet channel was extracted. In the same plot the SS of water basins used for irrigation scope, that surround the river channel, was retrieved. Size, uniformity and linear geometry of water basins, together with the deep of such water bodies, allow better visualization into satellite image and consequent lower values of surface reflectance, that in the visible (400-700 nm) and red edge (700-800 nm) are close to the one of "clear water" (Figure 26). In fact, the SS of clear water (suspended solids <10 mg/l) peaks in the green wavelength band and decreases with increasing wavelength, reaching near-zero reflectance in NIR (750-1350 nm). Water of wet channel and pools, when detected, has instead higher reflectance values among all the spectra. This is related to the fact that shallow water contains solid particles representing turbidity and its reflectance is affected by the response of the underlying sediment bed. Among all the seasons water SS shows a peak in B9, not an intrinsic property of water, that is low reflective in visible and completely transmissive for greater wavelengths, but it is instead a remark of soil and sediments. Generally, soil reflectance varies according to its chemical and physical composition and the main drivers are the moisture and organic substance content, texture

and structure. The reflectance of the soil increases with the wavelength and decreases proportionally to the moisture content in correspondence to the water absorption peaks (1400, 1900, 2700 nm). As for water, the spectral response of vegetation varies with the wavelength, and depends on multiple factors, such as the type of vegetation, density, state of growth and moisture content. In the visible, surface reflectance is related to the presence of pigments, such as chlorophyll, that determines the absorption of the blue and red radiation and reflection of the green one. Along NIR lengths, the SS is influenced by the structure of the leaf, while in the short-wave infrared (1350- 2700 nm) by the water content. Healthy vegetation tends to show greater reflectance on NIR wavelengths.
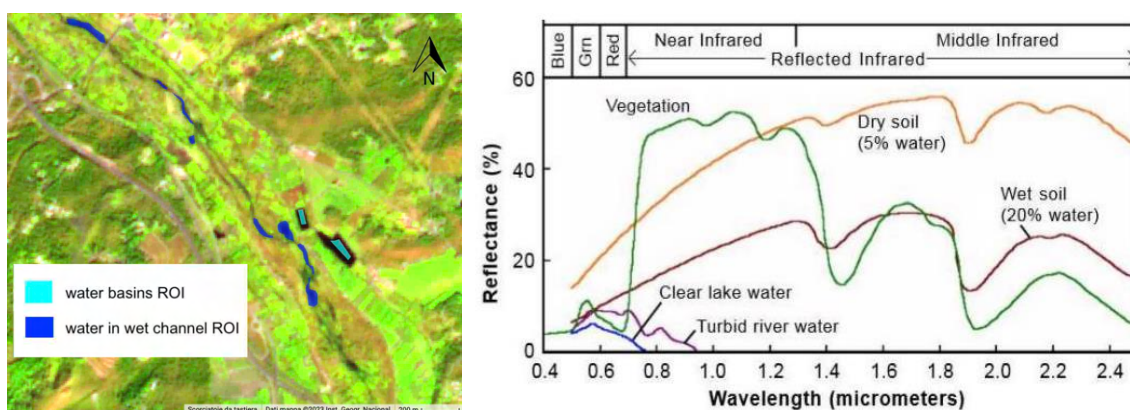
Figure 26: Extraction of water Regions of Interest (ROIs) belonging to basins and wet channel in Late Winter 2021/22 (left) and (right) common Spectral Signature plot (Progressive Gardening, 2022)

In spring 2022 the SS are shifted upward with respect to the previous season and this can be related to the difference in the solar radiation incident angle and to different image acquisition conditions. Here, after a local peak in B3 - green, both bushes and grass vegetation peak in NIR, demonstrating a higher vegetative activity. In late autumn 2022 the classes' reflectance decreased and approached, being not well distinguishable. Here water class represent pond ROIs, since the river was found in "ponding" condition. It is worth to note that these spectral behaviours have been extracted only starting from field survey, passing polygons acquired with a different point of view with respect to remote sensed image, generating more inaccuracies. At last, spring 2023 shown strong differences with respect to spring 2022, both in SS and on visual detection. Grass sub-class was no more present, bushes reached lower reflectance values in NIR, symptom of decreased activity and water along the channel was absent, determining "dry" condition over Palancia reaches. These modifications are related to the significantly lower precipitation intake over the hydrologic year 2023, phenomenon that caused the land cover change over the entire studied area. During this water year less than 300 mm of rain

were detected by the Algimia de Alfara station, 200 mm lower than the mean annual sensed precipitation.

Figure 27: Google Earth Pro images of Spring 2022 (left) and 2023 (right)
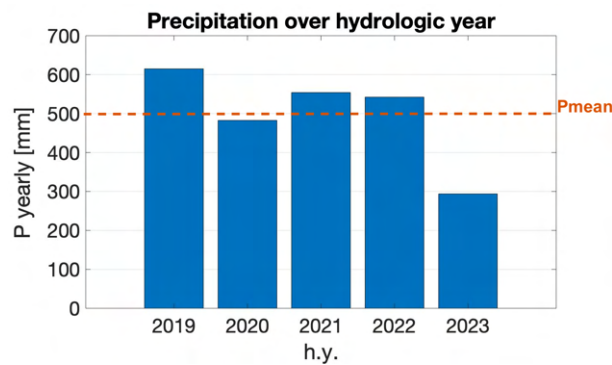
Figure 28: Precipitation over hydrologic year (h.y.) in the period 2019-2023 and relative P_{mean} recorded at Algimia de Alfara (IVIA, 2023)

For the studied Carraixet reach, no water was found in the ground truth references. As sad, it is classified as ephemeral stream, as the term "barranc" indicates, with only sporadic flow after storm events. An adjacent lower reach, with different hydromorphologic characteristics and permanent water presence, was used for the field survey acquisition and the SS extraction. Here, water from canals of the irrigation system is released in the Carraixet channel, so that downstream the conjunction point the river is in artificial "flowing" condition. This lower reach was not taken as additional case study because of the high human influence and the absence of "dry" condition over the 2017-2023 period. Late winter refers to the date 11/03/21, autumn to dates 06/11/22 and 16/12/22 and spring to the acquisitions 15-25/04/23.
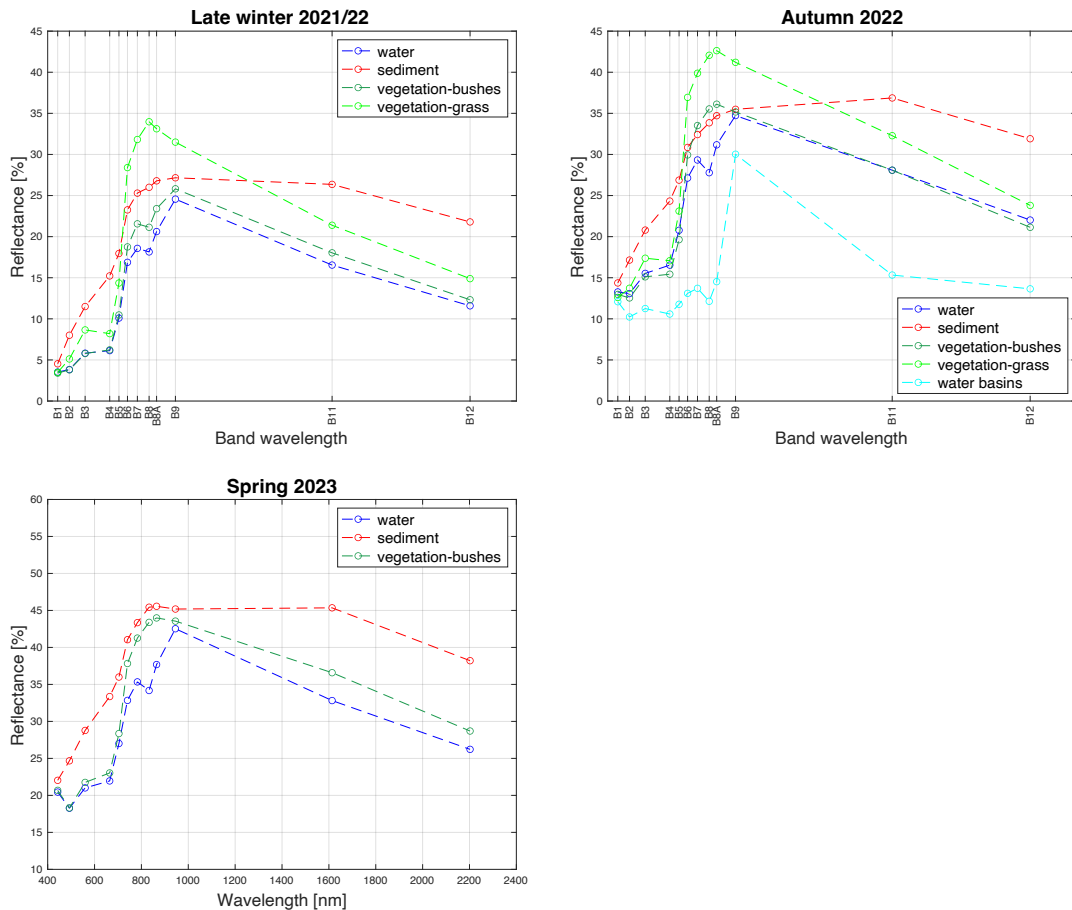
Figure 29: Seasonal Spectral Signatures and extraction of different water Regions Of Interest (ROIs) in Autumn 2022, Carraixet

Similar discussions can be conducted for Carraixet. In late winter the reflectance responses are particularly low, due to the differences in acquisition conditions. For autumn the SS of water basins was added, retrieving a behaviour more similar to "clear water", with an anomalous peak in B9, related to the sediment and vegetation dirtying effect on the pixel while acquiring with a spatial resolution of 10 m. This effect is amplified in the water ponds SS, since ROIs have more complex geometries that cannot be well described and extrapolated by FCIs. Due to the presence of marsh reed that in autumn and winter have brownish stems, the reflectance of bushes in these seasons reach lower values. In spring, instead, stems are greenish and the SS reaches an higher reflectance value. Once again, despite the artificial presence of water in the reach, in this season no grass was detected and the surrounding of the channel was particularly dry.

Descriptions of the spectral responses of the classes in summer period is missing, not having a complete knowledge of the reflectance variations along the year. Despite this, it is not possible to catch a unique seasonal response due to the year-by-year alteration, as demonstrates the comparison in Palancia between spring 2022 and 2023. In the extraction where a unique ground truth was used, there is a strong dependence on the satellite acquisition conditions, for instance for cloudy dimming and shadows covering. About the significancy of the results, it's necessary to note that despite a precise and accurate caption of classes' ROIs, it's impossible to have SS that perfectly describe the spectral characteristics of the pure classes. Despite it's expected that water inside river channel undergoes the influence of the underlying soil and sediments layer due to its shallowness, the obtained water reflectance of wet channel and ponds are compromised by the overlap within classes. Also water basins undergo the sub-pixel overlapping effect during satellite acquisition. Here stays the limit of using Sentinel-2 images, with finer spatial resolution of 10 m, while dealing with objects and covers-type that start from the meter scale (e.g. for pool with 2 m width), having a resolution mismatch that causes inaccuracies on the extrapolated data.

3.2.2 RGB-triplet selection

By viewing the previous charts it's possible to state that for visible bands (B2, B3, B4) the water and vegetation have similar spectral signature, sometimes overlapped, while sediment is well distinguishable due to the higher reflectance values on these bands, especially in B4. In red edge spectral region (B5, B6, B7) there is less separability between all signatures due to the rise of vegetations and water reflectance until the approach of sediment's one. Along NIR wavelengths (B8, B8A) the different classes' SS are sufficiently separable. In autumn, vegetation and sediment SS approach, due to the lower reflectance values that bushes and grass show among this season, while water always has significantly lower values; due to this fact, B8 and B8A are the best bands for water discrimination. The SWIR spectral region (B11, B12) is good for the different classes' separation, since sediment, vegetation and water occupy distinct reflectance levels. There are only in some cases in which bushes and water show similar reflectance among B12. Since the goal is the water class discrimination, taking into account the obtained results and the different bands' spatial resolution, selecting the finer ones, the RGB triplet R - B11, G - B8 and B - B4 was chosen for the FCIs visualization. Respectively, the bands resolutions are 20 m, 10 m, 10 m. The B11 provides a good

overall separability, B8 allows the distinction of the water and has higher spatial signature than B8A, and B4 permit to recognize sediments from water and vegetation, composing the optimum combination for water detection. This choice is coherent with the study of Cavallo et al. (Cavallo, 2022). The selected RGB composite returns FCIs where water is represented by dark color (black-dark blue), vegetation by green (grass by acid green and bushes by darker tones) and soil and sediments by straw yellow. By comparing the FCI with the correspondent True Color Image (TCI) and high-resolution orthophoto, it can be noticed that in the FCI the wet channel clearly stands out from the other component of the river corridor, while it is not easily distinguishable in the TCI, retrieved by the same multispectral acquisition and neither by the orthophoto, despite it has a much finer resolution. The comparison made illustrates how visualizing FCI in B11-B8-B4 allows a better differentiation of the classes, improving ease and accuracy for supervised classification and thus dataset creation. Since among all the spectra, including on the selected bands, the shadows naturally present in Sentinel-2 imagery have a SS similar to the one of water, in dataset creation images masked with the active riverbed channel were used.
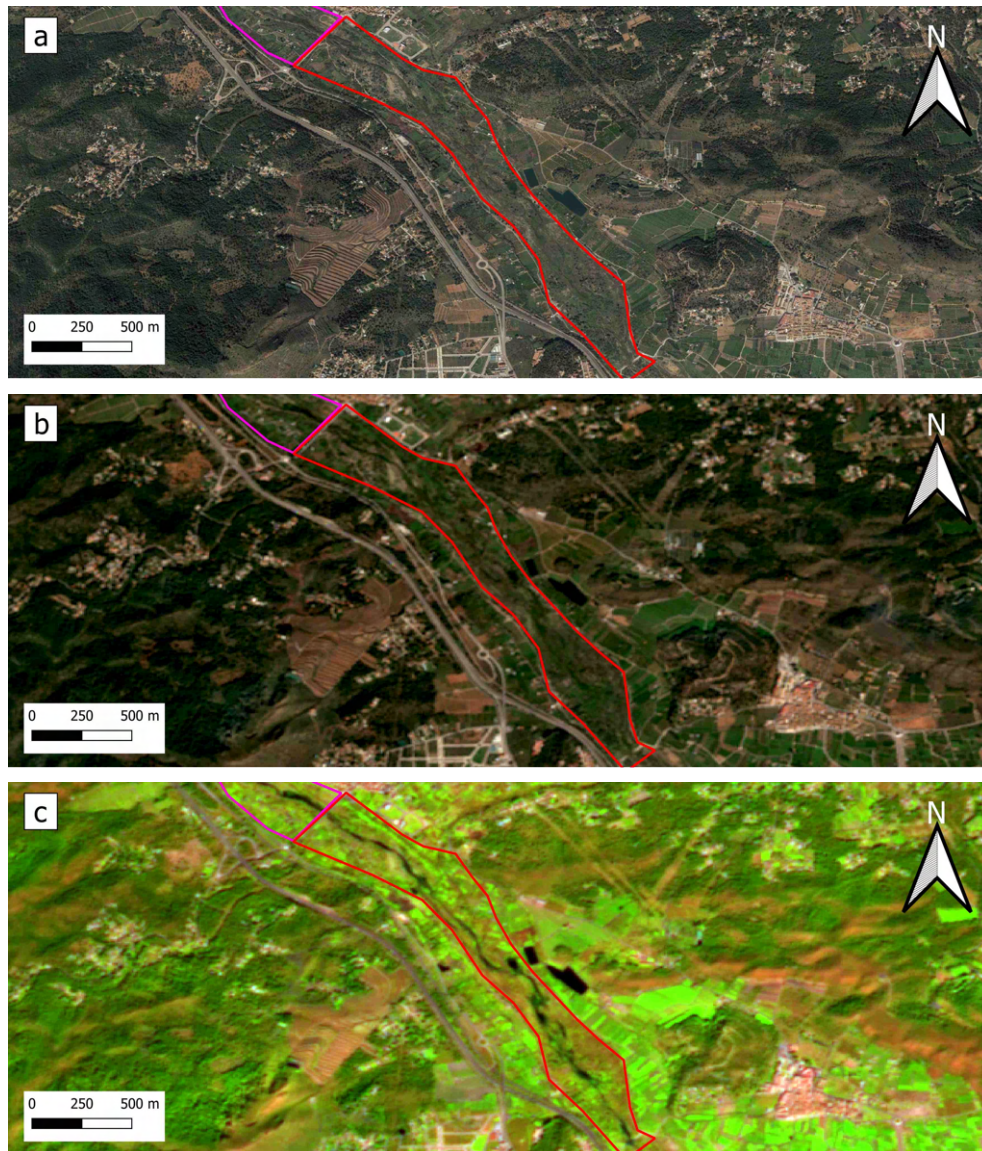
Figure 30: High resolution image (a), TCI (b) and FCI (c) visualization of date 11/03/21

3.2.3 Dataset creation

The GeoTiff images dataset has been created by clipping and classifying in a supervised mode the FCIs of the three reaches under study for the six-year period 2017-2023. A total of 1265 images was retrieved, subdividing them in the classes "cloudy" (C), "dry" (D), "flowing" (F) and "ponding" (P) and attributing the related classification confidence (high/low) since the attribution is not always trivial. Images clipped with active-channel mask were used in order to focus on the river channel only, avoiding the problem of having surrounding shadows and artificial water basins that can lead to image

misclassification when using unsupervised algorithms. The start of the period under study was determined by the availability of the Surface Reflectance products of Sentinel-2, April 2017, and the analysis ended on June 2023. The specification of dataset composition is given below (Figure 31). Among reaches, similar numbers of images were classified, almost all being of high confidence. In reach "Palancia 1", a majority of cloudy images were found (contribute of 37% on the total classified images), followed by the dry (30%), ponding (23%) and flowing one (10%). Within the flowing status the "dry" one is the most common, but considering the water status as joint, there is a balanced distribution of the "nonwater" (NW) and "water" (W) classes, sign that the reach experiences several situations with water. With quantitative data, it's possible to state that clouds are the major concern of working with satellite observation, highly reducing the quantity of useful images. In cloudy class have been considered both the images completely and partially covered by cloud that didn't make possible the attribution of a flowing condition. Reach "Palancia 2" is upstream with respect to the previous one, but it has different hydromorphological characteristics. In particular, the channel is thinner and generally has been detected lower water presence. In fact, even considering a unique water class (made up by flowing -12% and ponding – 9%), there is a strong dataset unbalance, since the majority class are cloudy and dry, both with a contribution of 40%. From the related bar chart is possible to view that there was a significant lower number of ponding images, while flowing status was always seen in almost 50 images. Despite the two reaches are adjacent, the cloudy classifications are different since cloud coverage can also interest little sub-portion of the image. The last reach, being Carraixet an ephemeral stream, showed a prevalence of dry phase (42% of the classified images) and a high cloudy percentage (39%). The flowing condition, with a continuity of water flows along the whole segment, is extremely rare (2%). More common is the ponding phase (17%) with water present in a heterogeneous and disconnected way, principally concentrating in a pool of large width naturally formed the lower part of the reach. Once again, the dataset is unbalanced, even considering the water-nonwater distinction.

Almost the 40% of the images detected by the satellite was cloudy, over each studied reach. This strongly limits the data availability on flowing condition. In several cases cloud covers exactly the "flowing" phase, being causing precipitation. This is also the phase more ephemeral, lasting in general shorter time. Ponding phase, instead, is much frequent and last longer since it comprehends a variety of situation, from the almost-flowing to the almost-dry one. The analysis of phases' frequency was necessary for the characterization of the rivers and for consequent dataset balancing operations.

Meteorological data has to be considered as coupled with the frequency analysis for assessing the duration of each phase and retrieving precipitation threshold for channel activation, operation yet disturbed by the indeterminacy of cloudy observations. Cavallo et al. proposed the use of Random Forest (RF) models able to fill temporal gaps between satellite images and to predict the occurrence of the flowing status on a daily scale, using cumulative rainfall and air temperature data (prior to the date of satellite acquisition) as predictors. The performances of RF models were very high, with total accuracy of 0.82–0.97 and true skill statistic of 0.64–0.95 (Cavallo, 2022). This suggest the goodness of the model, that can be applied also to different case studies, including Palancia and Carraixet, for retrieving the duration of each flowing phase.
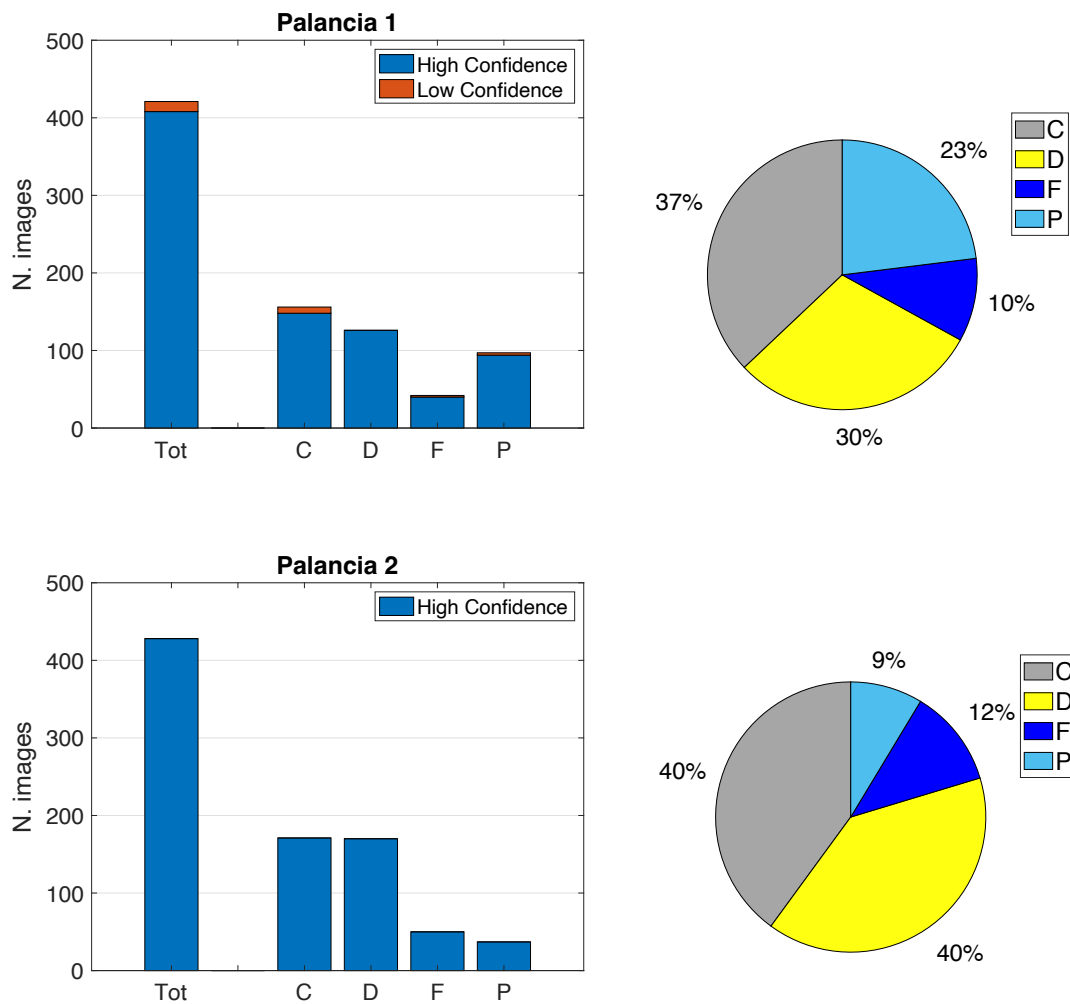
Figure 31:Distribution of the images classified in a supervised mode over Palancia reaches. Class C:Cloudy, D:Dry, F:Flowing, P:Ponding.
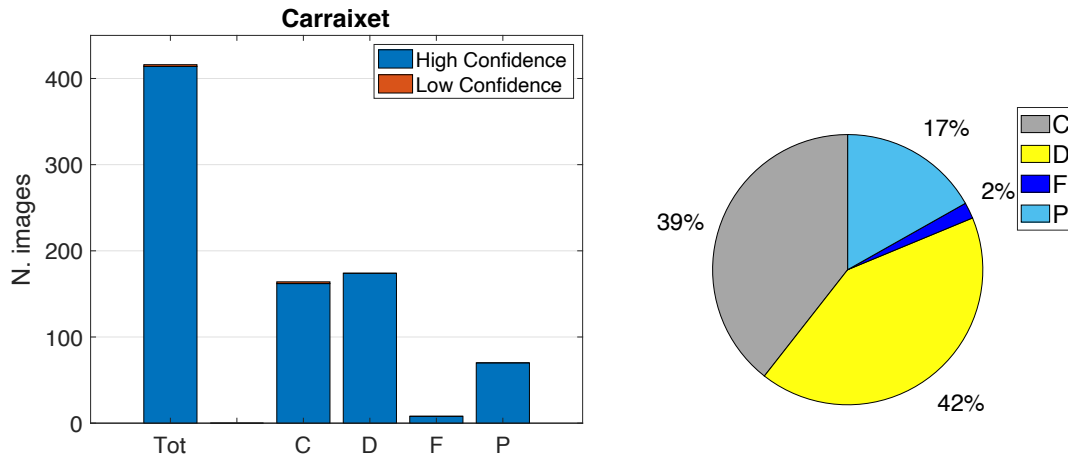
Figure 32: Distribution of the images classified in a supervised mode over Carraixet reach. Class C:Cloudy, D:Dry, F:Flowing, P:Ponding.

3.3 CNNs development

Images of created dataset have then adapted to feed the CNNs, following all the necessary steps to deal with a small dataset, having 1250 high-confidence samples, distributed as in Figure 33. Then, for each classification task only the images of the classes of interest were used: 769 samples for the 2 and 3-classes tasks, the total of 1250 for the 4-classes task. The pre-processing workflow was applied for obtaining a uniform dataset. After a trial-and-error procedure, a batch size equal to 32 was attributed as best image grouping size for extracting details from a small dataset. Features of Carraixet channel are less distinguishable compared to the ones of the other two case studies, since this reach has lower width, representing then a smaller object in which features do not appear with clarity. In order not to apply zooming operation, smaller images were framed with black pixels for obtaining a consistent size, a CNN input requirement. The overall unbalance of the class was solved case by case, by implementing oversampling procedure. The oversampled dataset was then split into Train, Val and Test DS, representing a percentage of 0.7, 0.2 and 0.1 respectively. Data augmentation processing was constructed to be applied consequently in the model construction.

Overall High Confidence image distribution
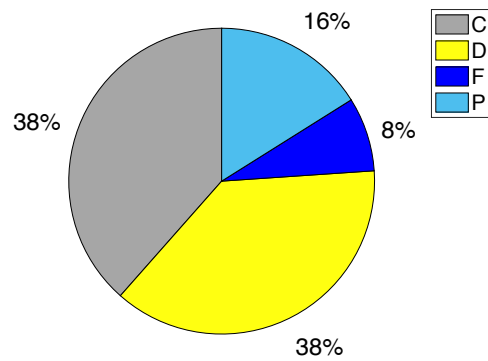
Figure 33: Distribution of the overall high confidence dataset. Class C:Cloudy, D:Dry, F:Flowing, P:Ponding.
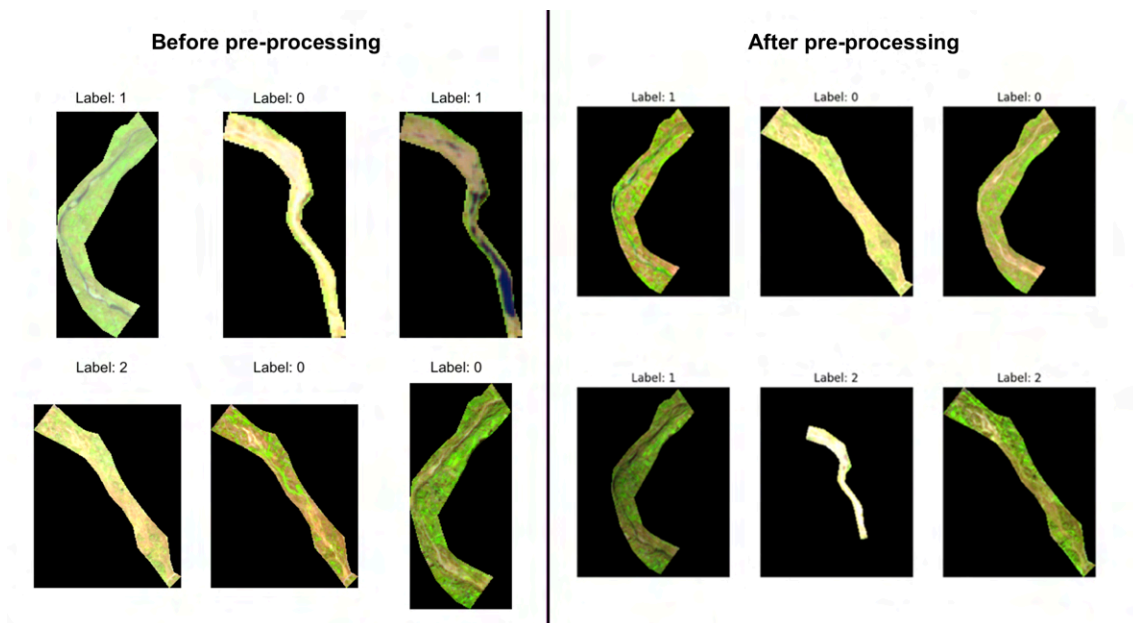
Figure 34: Images of the 3-classes database before (left) and post (right) pre-processing. Label [0]:Dry, [1]:Flowing, [2]:Ponding.

Specifications of models' construction are reported below. Basically, the same model was used, adapted then to the specific requirement of every task (e.g. binary or 3/4 multi-classes). Figure 35 shows the six layers that constructed the model, being in logical order: input layer, sequential layer, ResNet50 functional layer, pooling layer, dropout layer and dense layer. The input layer serves as the initial entry point for data. Its primary function is to accept the raw inputs in the form of images and prepare them for the further processing of the subsequent network layers. For every layer, the first dimension of the input and output object refers to the batch size. Since the datasets have numbers of samples not multiple of the batch size (equal to 32) and so the number of image per batch

is not constant, the first dimension is replaced with "None". In the first three layers the second dimension, "234", refers to the images height in pixels, while the third one, equal to "188" in these layers, to the images width, both values standardized during the resizing pre-processing step in which objects with smaller dimension were framed with black pixels. The last dimension, instead, refers to the number of color channel and it is equal to "3" since the three RGB channel characterized the images. By proceeding in analyzing the models' architecture, the second layer is the sequential one. It performs data augmentation, a technique that artificially introduces sample diversity by applying random but realistic transformations to the training images only, such as rotation and horizontal flipping, in a way that images do not change in size (height and width).

The functional layer with the ResNet50 base model aims to learn and extract meaningful features from the pre-trained model. ResNet50 involves a high number of parameters (23'587'712) for basic image recognition and the functional layer generates predictions from the block of features. The dimension "(None, 8, 6, 2048)" provides information about the shape and structure of the feature maps. Feature maps are the outputs of a convolutional layer representing specific features in the input image; in this case they are generated by the final convolutional layer of ResNet50 model and they are grouped in batch, in which each feature map has a height of 8 pixels, a width of 6 pixels and a depth of 2048 channels. Going forward into the model's architecture, the pooling layer (GlobalAveragePooling2D) simplifies the image and perform downsampling along the spatial dimensionality, averaging the information contained in each feature map. In the model developed, it converts object of size "(None, 8, 6, 2048)" into the simplified one of "(None, 2048)", resulting in a feature vector of length 2048 for each sample in the batch. The dropout layer serves as regularization technique to help preventing overfitting, since it randomly deactivates a fraction of the neurons of the layer, avoiding the model from fitting too closely the training data. In the constructed CNN, this layer inserts between the pooling and the dense layer. The dense layer is the responsible for producing the network's predictions. The number of neurons (units) in this layer corresponds to the number of classes or categories of the classification problem. For the binary classification model, the dense layer returns an object of size "(None, 1)", where the second dimension indicates that the raw prediction (logits vector) has a single prediction per image: positive numbers predict class [1], negative numbers predict class [0]. For the 3-classes task the model layering was the same of the previous case, with the unique difference of having a second dimension of the dense layer equal to 3, since every line of the logits vector should contain as aligned the classification score for classes [0], [1], [2]. The same occurs in the 4-classes model, where the dense layer has a dimension of 4, scoring over [0], [1], [2],

[3]. This predictive layer used activation function for converting the raw logits, that represent the classification scores for the model, into probability values between 0 and 1. The sigmoid activation function was used for the binary classification, while softmax activation function for multi-classes tasks. These probabilities are then the input of the loss function, that returns a metrics for evaluating the model, the loss. The Binary Cross-Entropy or Log Loss function was used for the 2-classes problem and the Sparse Categorical Cross-Entropy Loss function for multi-classes ones.


```
Layer (type)                    Output Shape              Param #
=================================================================
input_2 (InputLayer)            [(None, 234, 188, 3)]     0

sequential (Sequential)         (None, 234, 188, 3)       0

resnet50 (Functional)           (None, 8, 6, 2048)        23587712

global_average_pooling2d (      (None, 2048)              0
GlobalAveragePooling2D)

dropout (Dropout)               (None, 2048)              0

dense (Dense)                   (None, 1)                 2049

=================================================================
Total params: 23589761 (89.99 MB)
Trainable params: 2049 (8.00 KB)
Non-trainable params: 23587712 (89.98 MB)
```

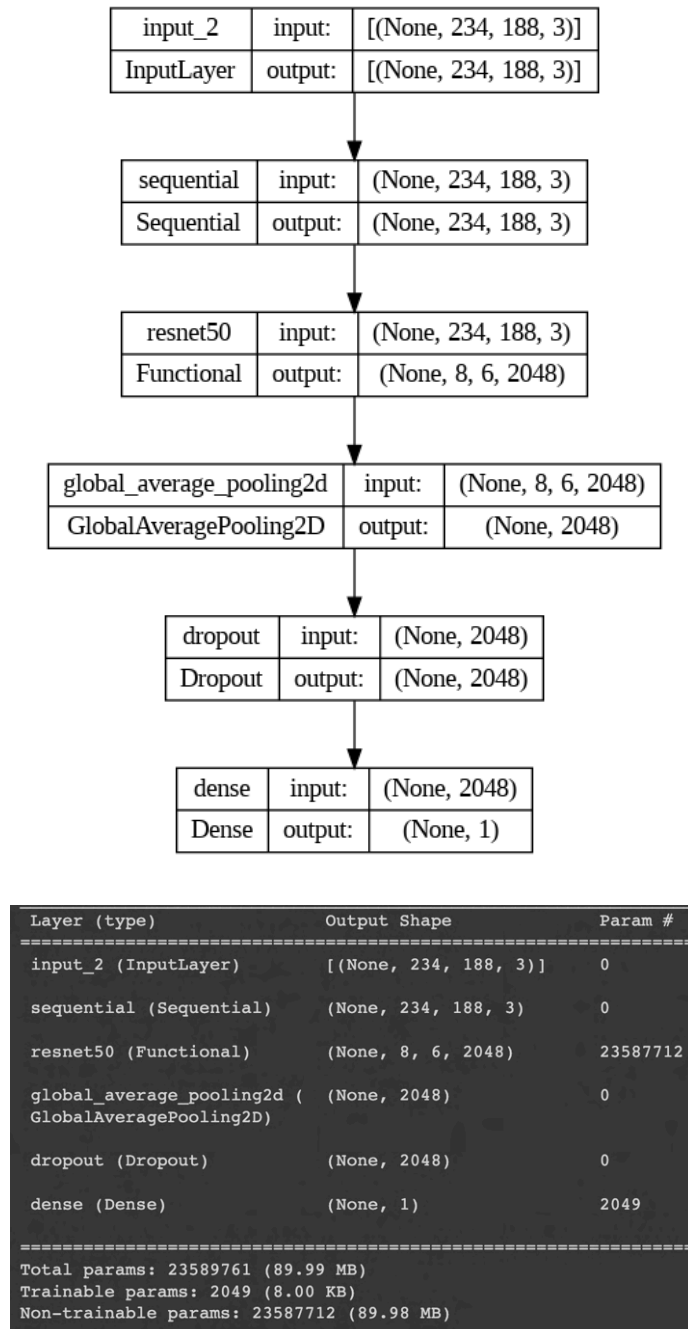*Figure 35: Architecture (above) and initial parameters (below) of used binary classification model*

```
Layer (type)                 Output Shape            Param #
=================================================================
input_2 (InputLayer)         [(None, 234, 188, 3)]   0

sequential (Sequential)      (None, 234, 188, 3)     0

resnet50 (Functional)        (None, 8, 6, 2048)      23587712

global_average_pooling2d (   (None, 2048)            0
GlobalAveragePooling2D)

dropout (Dropout)            (None, 2048)            0

dense (Dense)                (None, 3)               6147

=================================================================
Total params: 23593859 (90.00 MB)
Trainable params: 6147 (24.01 KB)
Non-trainable params: 23587712 (89.98 MB)
```

```
Layer (type)                 Output Shape            Param #
=================================================================
input_2 (InputLayer)         [(None, 234, 188, 3)]   0

sequential (Sequential)      (None, 234, 188, 3)     0

resnet50 (Functional)        (None, 8, 6, 2048)      23587712

global_average_pooling2d (G  (None, 2048)            0
lobalAveragePooling2D)

dropout (Dropout)            (None, 2048)            0

dense (Dense)                (None, 4)               8196

=================================================================
Total params: 23,595,908
Trainable params: 8,196
Non-trainable params: 23,587,712
```

Figure 36: Layets and parameters used for the 3-classes (above) and 4-classes (below) models

For the selection of the parameter to use during compiling and fitting procedures, a trial-and-error procedure was used. For the compiling, 0.001 was chosen as start value for the learning rate; the model response at tuning this parameter to multiple and submultiple of the start value was observed, until the identification of the optimal learning rate, equal to 0.0001 for all the models. The optimum number of epochs to use during fitting instead changed with tasks. To retrieve this number, an initial long run was conducted, setting a high number of epochs and activating the early stop tuning to shut the fitting when the performance of the model, and in particular the validation loss, keeps decreasing. For each task, multiple runs of the model were made in order to identify the optimum number of epochs for both the initial fitting and the fine tuning fitting. The use of the early stop tuning allows to avoid overfitting while maximising the learning of the CNN.

3.4 CNNs performance

3.4.1 Model 2-classes

For the binary classification model, the sigmoid activation function and the Binary Cross-Entropy Loss function were used. In this model, class labels attribution was [0] – nonwater, NW, and [1] – water, W, since water is the feature under interest to identify within the binary classification. The performance of the model was evaluated, as practice, on the Test DS. In fact, this subset is populated by images completely new to the model, not viewed and used for learning, that must be classified in an unsupervised manner. The model was run 5 times (a suggested number between 3 and 10) on the same dataset, in order to catch its different responses: at every run the weight initialization was different, as well as the content of the Train, Val and Test DS in terms of images. Following this procedure, the performance results reported in Table 4 were obtained. The accuracy for the binary classification on the test set ranged in the 0.70-0.78 interval, with an average value within the five runs of 0.77. While the accuracy should reach high values, the loss should minimize, so that the related test loss, from the worst to the best score, ranged in the 0.62-0.42 interval, with an average value of 0.52. The learning curves are the evolution of the model performance in terms of accuracy and loss over epochs. The learning curves for run n.1 and n.5, respectively the one with lowest and highest performance representing the range's extremes of the analysis, are reported in Figure 37. A total number of 21 epochs was used, subdividing them into 7 for the initial fitting (epoch from 0 to 6) and 14 for fine tuning (from 7 to 20). In the initial stage, since the totality of the layer of the base model al freeze, the increasing of the performance was limited and sometimes characterized by a plateau, like in the very first epochs of run n.5; by analyzing the predictions at the end of the initial fitting, it's possible to retrieve that in the plateau epochs the model was predicting an unique class, and in particular class water, for all the images of the validation dataset, having low performance since only the 35% of the images in the Val DS belonged to this class. Once in the fine tuning fitting, the learning performance improved, thanks to the contribution of an higher number of involved layers and parameters, obtained by unfreezing a part of the base model. In this stage the model learned more features from the Train DS, improving its guess on the validation one. In the learning curve n.1 both training and the validation accuracy reached a value of 0.70 and this performance was maintained also in predicting the images of the test set. In learning curve n.5 the validation accuracy reached the value of 0.75 and loss values lower than 0.5, for then having a slightly better performance while predicting the

Test DS, with a test accuracy of 0.78. The shape of the learning curve, with validation and training curves that gradually approach at latter epochs, demonstrated that a good fit was performed, not obtaining overfitting neither unrepresentative validation. Although this, the model had limited performance in both terms of accuracy and loss, demonstrating its not optimal learning. By analyzing the confusion matrixes and F1-score metrics, directly computed from the first cited, it's possible to retrieve the major cause of prediction failure. The F1-score metric is class specific, so that this metrics is present for both NW and W class. The average F1-score for NW class is 0.77 (range 0.67-0.85), higher than the one of W of 0.64 (range 0.59-0.75). This indicates that a higher number of correct predictions occurred for the NW class. The visualization of the confusion matrixes (Figure 38) confirmed that, despite in run n.2, the test samples correctly predicted as NW class are higher in number. The model mainly failed by attributing the W label to images that in reality are NW, having a lot of FalsePositives. Also the number of FalseNegatives was relevant and the overall False attribution limited the F1-score of both NW and W.

Despite it's not explicit in the binary classification model, this misclassification is probably caused by the presence of the ponding images in the water class. This in-between status, in fact, covers a bench of situation from the quasi-dry to the quasi-flowing condition, so that it can led to classification errors, considering dry images as ponding and thus W and viceversa. The inability of the model to reach higher performance is also probably related to the limited dataset size. Despite different expedients were used to minimize the effect of the dataset size, the model was not able to learn in an optimal way for the dedicated classification task.

Table 4: Performance results of 2-classes model

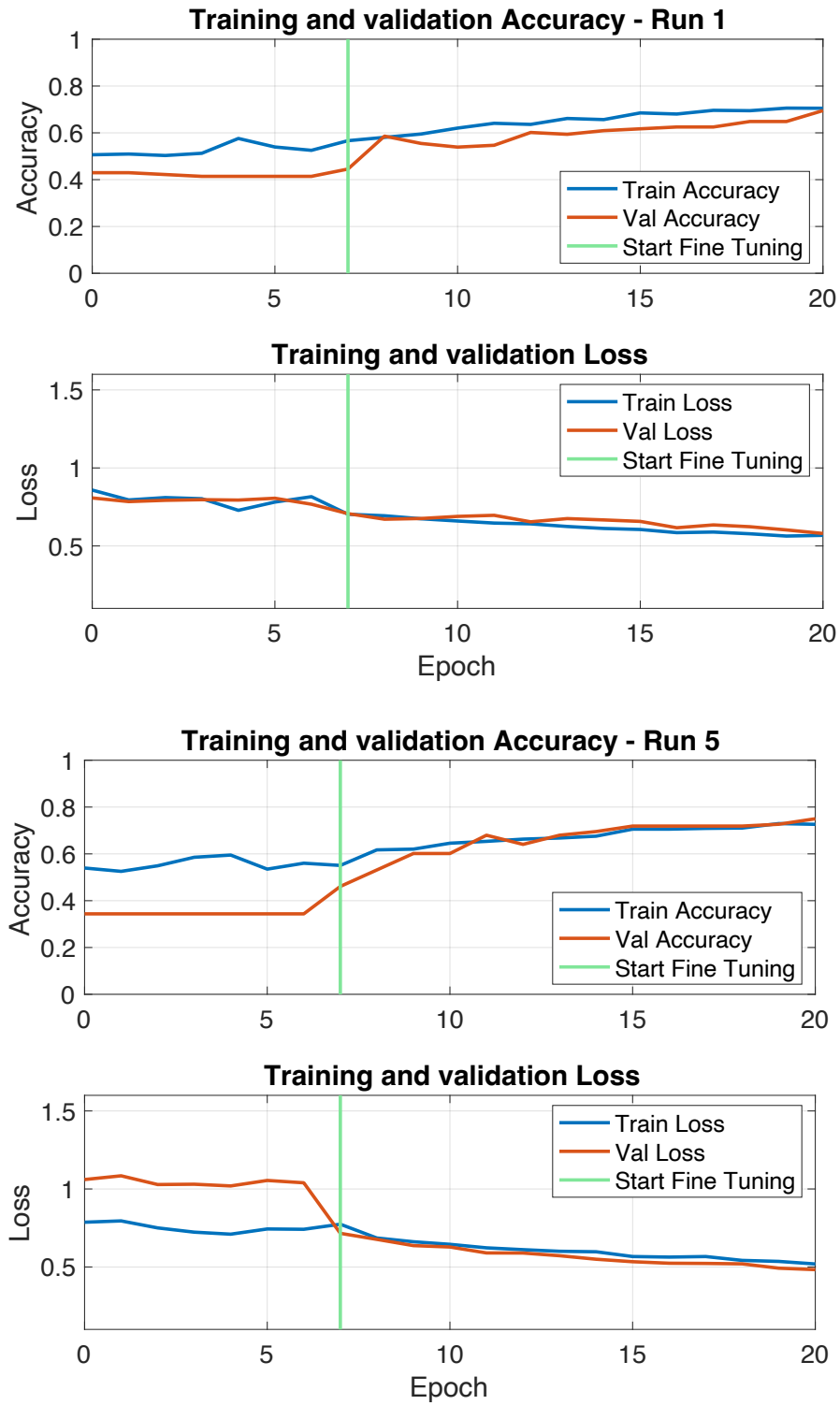| Run n.	Test Loss	Accuracy	F1-score NW	F1-score W
1	0.62	0.70	0.75	0.60
2	0.48	0.72	0.67	0.75
3	0.58	0.72	0.78	0.62
4	0.50	0.74	0.80	0.63
5	0.42	0.78	0.85	0.59
AVG	**0.52**	**0.73**	**0.77**	**0.64**

Figure 37: Learning Curve of run n.1 and n.5. Total of 21 epochs, fine tuning at 7th epoch
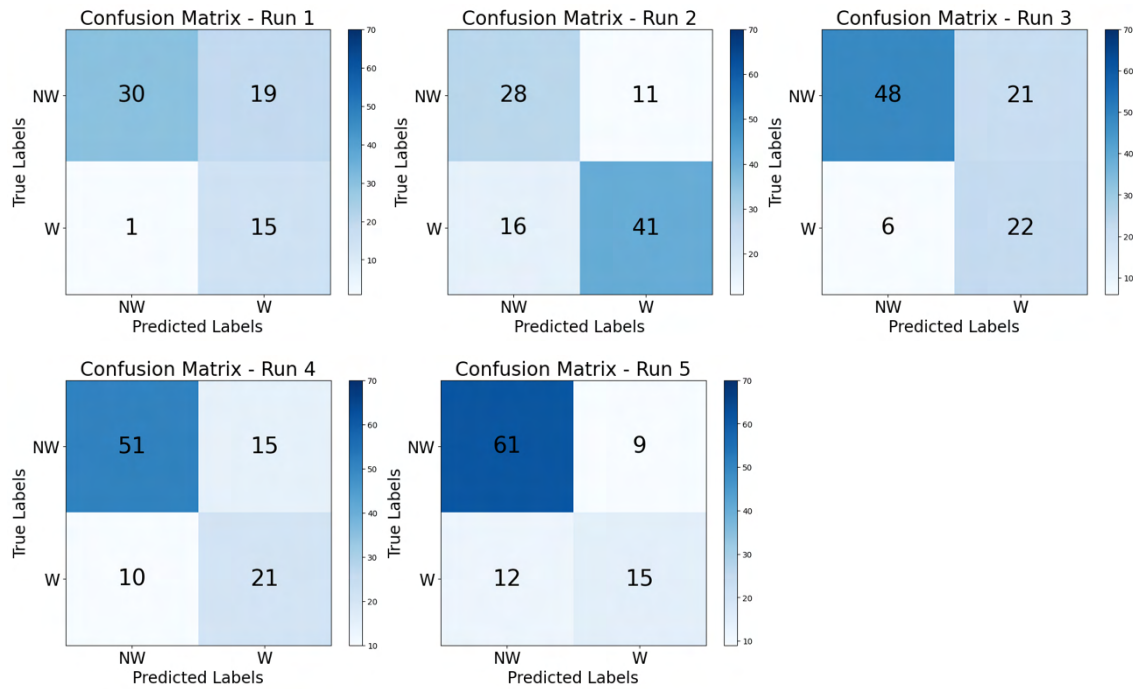
Figure 38: Confusion Matrix relative to the 2-classes model runs

3.4.2 Model 3-classes

In the 3-classes model, the considered categories were: [0] – dry, [1] – flowing, [2] – ponding. Once again, the model was evaluated on the test set along 5 runs. Since it was a multi-classes classification model, the softmax activation function and the Sparse Categorical Cross-Entropy Loss function were used. For model fitting, a total number of 21 epochs was adopted, using 5 of them for the initial fitting and the remaining 16 for fine tuning. The registered performances are reported in Table 5. Along the different runs, the average accuracy on test data was of 0.78, reaching a minimum of 0.72 in run n.5 and a maximum of 0.85 in run n.4. The correspondent losses were 0.5 on average, with the worst value of 0.61 and best of 0.38. These scores, in terms of accuracy in the range 0.72-0.85, were higher than the ones of the 2-classes model, instead with accuracy in 0.70-0.78 range, demonstrating that, overall, the model was better in distinguishing the water classes (flowing and ponding) as separated instead of incorporated. The F1-scores clarified that the dry and flowing class were sufficiently well classified, while the majority of ponding samples were incorrectly predicted. The dry class was the best identified, with F1-scores in the range 0.82-0.94, followed by the flowing one, that ranged in 0.69-0.87 and lastly by the ponding one, that ranged in 0-0.31. The extremely low results on the ponding class were driven by the inability of the classifier to identify ponds of water. The

confusion matrixes in Figure 39 shows that in run n.1 and n.4, none of the sample in the Test DS was predicted as ponding and that, in run n.5, a single image was attributed to this class, but erroneously, since its true label stated the belonging to the dry class. Run n.2 and n.3 were the unique ones in which the F1-score was higher than 0; the confusion matrix of the first one shows a single sample correctly predicted as ponding, over the 12 truly present in the Test DS, and 4 samples as FalsePositives, meaning wrongly predicted as ponding; the confusion matrix of the second one, instead, shows that the model correctly identified 2 images of the ponding class, having no FalsePositives, but not recognizing all the true ponding images in the Test DS, missing 9 of them. Despite the images in the Train DS were balanced in order to have the same initial condition along the classes during the learning phase and despite in the Val and Test DS the number of ponding images is slightly higher than the one of flowing samples (Figure 33), for the classification of ponding class the model had the lowest performance, revealing difficulties in recognizing ponding features. From the practical point of view, the obtained results showed the limitation of the CNN identify the feature of ponds of water with limited size in the given FCIs. During the supervised classification, the attribution to ponding class of some of the images were made knowing that even small pixels in a precise zone of the channel were representing water. Since the CNN model was trained solely on images, without other site's information, the distinction of this status is more muddled. The RGB triplet selected for FCIs visualization is the one that highlights water presence and for the CNN this was effective for the recognition of water in wet channels characterizing the flowing condition, since in these images a greater number of dark tones pixels, with spatial continuity all along the river channel, was present. The choice made was not effective in the same way for the opportune discrimination of isolated dark pixels, that resulted not sufficiently highlighted for the CNN. The evolution of accuracy and loss along epochs for run n.3, under interest since the one that bring to the best identification of the ponding class, and n.4, the one that reached the highest test accuracy score, can be viewed in the learning curves in Figure 40. In the first 2 epochs of the learning curve relative to the run n.3, the accuracy on the Val DS is higher than the one on the Train DS, but at that point the curves still had to reach their stability, so that it's not a significant behaviour and may be related to the guess of the majority of the images as dry. Over the rest of the epochs the two accuracies increased, both reaching a value close to 0.8 (validation accuracy of 0.78 and training accuracy of 0.8) in the latter one. The score on the Test DS for run n.3, in fact, was of 0.77, similar to the one on Val DS. In the learning curve of run n.5 along the epochs the two learning curves are closer, not having sharp rise neither drop. In particular the losses reach values lower than 0.5 and the accuracy values

higher than 0.8 (0.83 for both training and validation accuracy), then predicting correctly the 85% of the samples in the Test DS.

Table 5: Performance results of 3-classes model

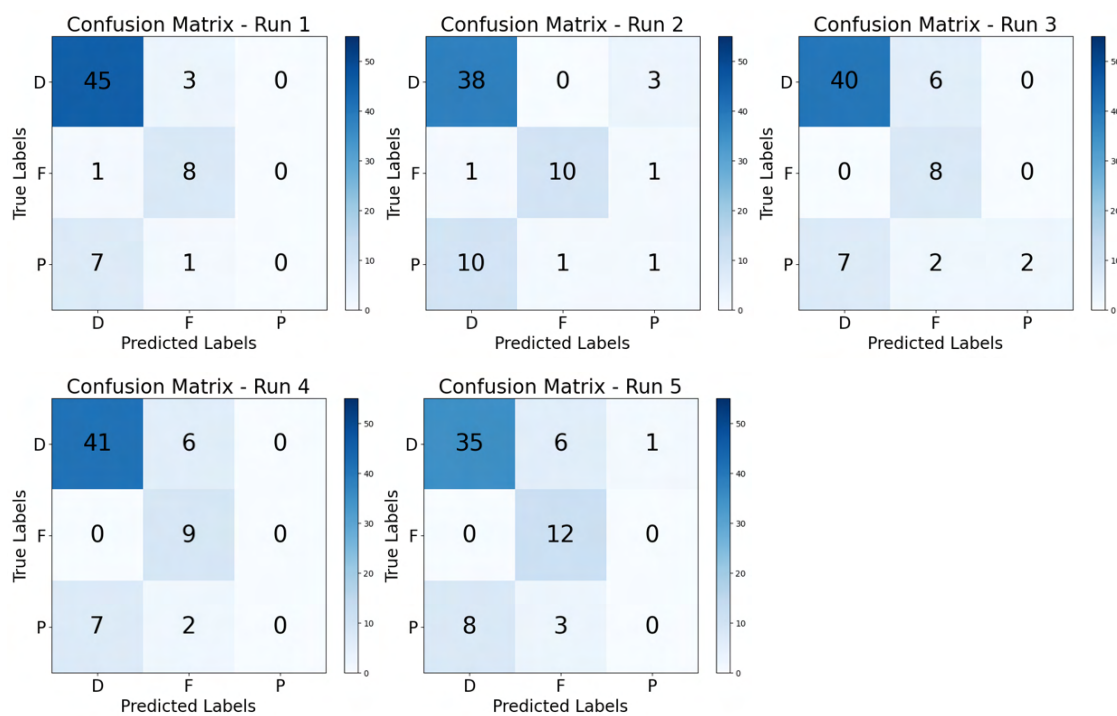| Run n.	Test loss	Accuracy	F1-score D	F1-score F	F1-score P
1	0.47	0.82	0.89	0.76	0.00
2	0.49	0.75	0.84	0.87	0.12
3	0.56	0.77	0.86	0.69	0.31
4	0.38	0.85	0.94	0.81	0.00
5	0.61	0.72	0.82	0.73	0.00
AVG	**0.50**	**0.78**	**0.87**	**0.77**	**0.09**

Figure 39: Confusion Matrix relative to the 3-classes model runs
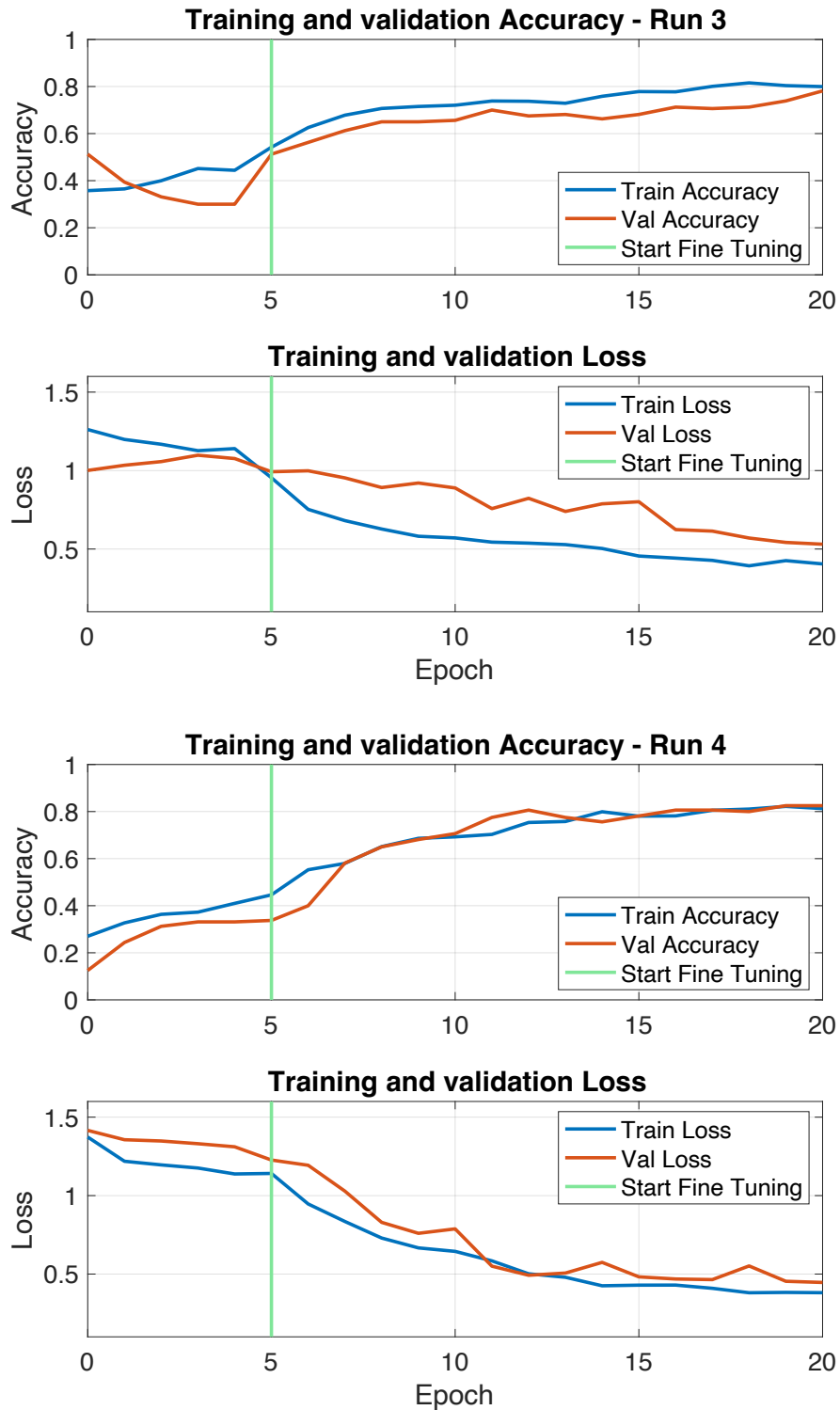
Figure 40: Learning curves of run n.3 and n.4. Total of 21 epochs, fine tuning at 5th epoch

3.4.3 Model 4-classes

For the 4-classes model the considered classes, with related labels, were: [0] – cloudy, [1] – dry, [2] – flowing, [3] – ponding. This model was evaluated on the totality of the created dataset because also the cloudy images is considered and stored in a separated class. Since it was a multi-classes classification model, the softmax activation function and the Sparse Categorical Cross-Entropy Loss function were used. The performance of the model was recorded along 5 runs, using for the fitting a total number of 40 epochs, using the first 20 (0-19) for the initial fitting and the other 20 (20-39) for fine tuning. The results, reported in Table 6 show the highest performance within the developed model in terms of accuracy and loss. The accuracy ranged in 0.79-0.88, with a medium value of 0.85, while the loss ranged in 0.42-0.33 (from the worst to the best score), with a mean of 0.36. The model optimally recognized the cloudy samples, with F1-scores always higher than 0.9 and that reach a value of 0.99 in run n.5. Also the F1-scores related to the flowing class were high, ranging in 0.87-0.91, followed then by the dry class, with F1-scores in the range 0.80-0.88 and finally by the dry class, that had the lowest scores, in the range 0.17-0.52. In run n.1 the prediction of cloudy and flowing was very precise, while a higher number of misclassification were made on dry and on ponding class; in particular the 9 samples were predicted as ponding, but only 3 correctly. In run n.2 a slightly worst classification on the flowing samples was made, for a slight improving in the ponding recognition; in fact, on the total of 13 ponding image present, only 4 were correctly classified, with other 4 predictions to this class made erroneously. The run n.3 showed within the best scores for cloudy, dry and flowing images, but had a relatively low F1-score for ponding class. In run n.4 the overall accuracy is quite low, reflecting the lowest F1-scores of the classes. Finally, in run n.5 the accuracy increased, having the highest F1-scores among all the four classes. Here, the ponding score is higher than 0.5, since on the 16 images with ponding label, 6 were correctly classified, having only one case of erroneous prediction to this class. After this discussion, it's possible to state that the cloudy class didn't mine the model performance, but instead allowed to reach higher stability, going ahead in epochs for the improvements of the overall performance. In fact, this dataset is more populated, since cloudy images represented almost the 40% of the total frame classified in a supervised mode. In Figure 43 are reported the learning curves of run n.3 and n.5. Both the runs reached an accuracy on the test set of 0.88 but following different metrics evolution along epochs. At the 40th epoch the training and validation accuracies reached a value of 0.87 in the first graph and of 0.88 in the second one. In the 4-classes model, the higher number of epochs allowed the model to reach a highest performance while

remaining stable, without overfitting. In fact, when the number of epochs used to train a neural network model is more than necessary, the training model learns patterns that are specific to training sample data to a great extent, bringing overfitting and reducing the reliability of the model.

Table 6: Performance results of 4-classes model

Run n.	Test Loss	Accuracy	F1-score C	F1-score D	F1-score F	F1-score P
1	0.34	0.85	0.96	0.85	0.91	0.32
2	0.36	0.84	0.97	0.84	0.87	0.38
3	0.33	0.88	0.97	0.88	0.89	0.29
4	0.42	0.79	0.92	0.80	0.88	0.17
5	0.37	0.88	0.99	0.88	0.90	0.52
AVG	**0.36**	**0.85**	**0.96**	**0.85**	**0.89**	**0.33**

Figure 41: Learning curvers of run n.4 and n.3

Figure 42: Confusion Matrix relative to the 4-classes model runs
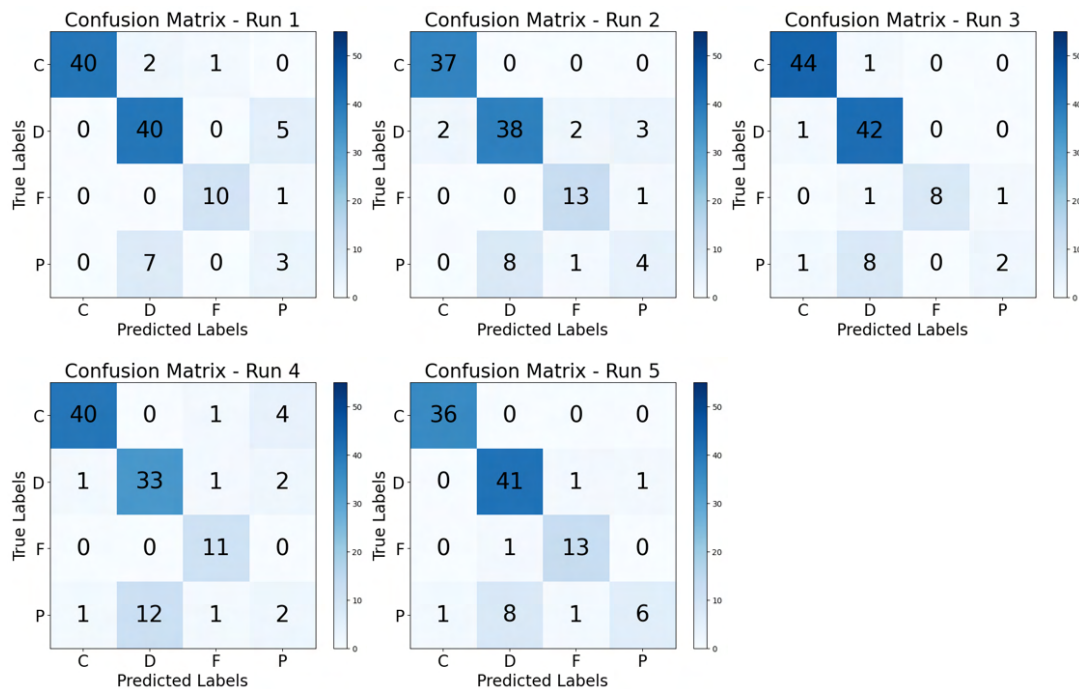
Figure 43: Learning curves of run n.3 and n.5. Total of 40 epochs, fine tuning at 20th epoch

In the end, the obtained accuracy ranged in 0.70 - 0.78 for the 2-classes model, in 0.72 – 0.85 in the 3-classes model and in 0.79 - 0.88 in the 4-classes one. Hence, the last is the one reporting the better performances, while the first one is the roughest. When present,

the class C is the best identified, with a F1-score ranging in 0.92 - 0.99, due to the strong difference of cloudy images, that most of the times appears as white objects with no features on the ground. In both the 3 and 4-classes CNNs, the flowing condition is well identified, with a F1-score of 0.60-0.98, with a higher frequency of high values in the 4-classes model. Along all the developed CNN, the dry class is sufficiently well classified, even if a high number of sample belonging to other class were attributed to D and this is the main reason why the F1-score has on average values lower than 0.9. When explicitly separated, the ponding class was the worst identified. In the 3-classes CNN in three runs over five, none of the sample was correctly attributed to this class. The situation improved a bit in the 4-classes model, but still this is the main cause of misclassification. The low performances reached in the binary classification is probably related to the difficulty in attributing the ponding samples to the water class. The cloudy presence in the last considered case is not introducing big concern, but instead allows the model to be stable over a higher number of learning epochs.

The obtained results demonstrated that this approach is a good starting point for assessing the flowing condition of NPRs in unsupervised mode. The not optimal learning of the CNNs has to be related to both the data quality and data quantity. In fact, despite the Sentinel-2 resolution mission was selected as the best compromise between spatial, spectral and temporal resolution, the effect of pixel shelling summed to the fact of having a limited dataset, made the small water objects, such as the ponds, difficult to discriminate for the neural networks. The recognition of images with clear features, over all the cloudy images, with the quasi-totality of white pixels and few information of the ground, is almost perfect. The CNN classification potential and validity were also demonstrated by the performance in identifying the flowing samples; also the images in the flowing conditions have clear features, specifically enhanced by selecting a FCI visualization: water in wet channel appears as a bench of pixels with dark tones and spatial continuity. Also dry images alone should have clear features, being without water and without white-cloudy pixels and so well distinguishable, but the introduction of the ponding class, that the majority of the times in the used dataset was more similar to the dry condition rather than to the flowing one, led to feature mistake and so misclassification, lowering the overall performances. Despite their cryptical functioning, thanks to the bench of evaluation metrics that can be computed and the visualization of the model learning curves, it's relatively easy to comprehend why the network is not performing well, for then bringing the necessary modifications. In fact, the three developed model clearly showed that the problematic class is the ponding one. Hence, for better classification performances it's necessary to work on this prediction, increasing the number of ponding

images from which the modal can learn and trying different tuning approaches. Generally, the overall dataset extension and diversification are requires. The images used to feed the models were to scarce, so that the ponding features were not sufficiently distinguished from the dry ones. Further improvements are related to the possibility to modify the base model for fine tuning, testing some with variated features and possibly with higher number of satellite images. With a more populated dataset, other type of balancing operations can be considered. By changing the above, also model tuning will require important modifications related to number of epochs, learning rate and freezing and unfreezing procedures. Also, the k-folder cross validation can be added to directly retrieve metrics representative of multiple runs.

Conclusion

The study proposed an innovative methodology for the unsupervised recognition of the flowing conditions in non-perennial rivers, by combining Remote Sensing and Convolutional Neural Network disciplines and using them for environmental and hydrologic application. For this implementation, the satellites of Sentinel-2 mission were found to be the best providers of observational data, due to the high spatial resolution and revisit time. The methodology was applied using the images detected in the period 2017-2023 over three reaches taken as case studies. In particular, reaches of two different Mediterranean NPRs located in Eastern Spain were considered. The overlapped ROIs collected during field survey or drawn on high resolution orthophoto permitted the extraction of the spectral signatures of the classes water, sediment and vegetation, the last distinguished into bushes and grass. The obtained spectral responses were compromised by the coarser resolution of satellite image with respect to ground truth, but still allowed the identification of the bands triplet in which the spectral differences between classes were the greatest. This combination was obtained by attributing B11-B8-B4 to the RGB channels for obtaining FCIs. Starting from the visible, in B4 water and vegetation have similar reflectance but can be distinguished from sediment. In NIR region, B8 allows a clear identification of water, while among B11 in the SWIR all the classes can be differentiated fairly well. FCIs were then used for the identification of the flowing condition among reaches and so classified in a supervised mode, subdividing the images into "cloudy", "dry", "flowing" and "ponding" classes. Almost the 40% of the total detected images were cloudy, thus an important fraction of observational data was not useful for retrieving the flowing condition. This is the main limitation while dealing with satellite acquisitions. Three CNNs were developed, one for binary classification (distinguishing between nonwater – water), one for 3-classes task (dry – flowing – ponding) and one for 4 multi-classes (cloudy – dry – flowing – ponding), and the FCIs database was used to feed them. Firstly, the images have been pre-processed in order to meet the input requirements of the models. The networks started from the same 6-layer model, composed by input, sequential, functional, pooling, dropout and dense layer, the latter adapted to the number of classes that the predictor had to consider. Since most of the time the population of the classes in the dataset was unbalanced, balancing operations were performed on the training subset only, by applying the oversampling of minority classes. Methods to avoid overfitting were used and while tuning the models' parameters

a low level of complexity was preferred, selecting a number of epochs between 21 and 40 for the different models and a learning rate of 0.0001. The obtained accuracy ranged in 0.70 - 0.78 for the 2-classes model, in 0.72 – 0.85 in the 3-classes model and in 0.79 - 0.88 in the 4-classes one. Hence, the last was the one reporting the better performances, while the first one was the most rough. The cloudy presence is not introducing big concern. In fact, when present, the cloudy class is the best identified, with a F1-score ranging in 0.92 - 0.99, due to the strong difference of cloudy images, that most of the times appears as white objects with no features on the ground. In both the 3 and 4-classes CNNs, the flowing condition was well identified, with a F1-score of 0.60-0.98, with higher frequency of high values in the 4-classes model. Along all the developed CNN, the dry class is sufficiently well classified, even if a high number of sample belonging to other class were attributed to the dry one and this limited its F1-score to values lower than 0.9. Among all the models, the ponding condition was the most critical to detect, negatively affecting the performance. These results confirm that the proposed methodology is a good starting point for mapping water presence and assessing flowing conditions of NPRs in an unsupervised mode. The not optimal learning of the CNNs has to be related to both the data quality and quantity. In fact, despite the Sentinel-2 resolution mission was selected as the best compromise between spatial, spectral and temporal resolution, the effect of pixel shelling summed to the fact of having a limited dataset, made the small water objects, such as the ponds, difficult to discriminate from the neural networks. The correct recognition of images with clear features, over all the cloudy and flowing ones, was demonstrated by the high performance in attributing these classes. Alone, dry images should have clear features, but the introduction of the ponding class, that the majority of the times was more similar to the dry condition rather than to the flowing one, led to features mistake and so to misclassification, lowering the overall performances. For better classification performances, a focus on the ponding class is needed, increasing the number of samples from which the model can learn ponding feature and trying different model tuning after the modifications in other. Hence, model improvements are required, generally starting from the dataset extension and diversification, adding reaches with different hydromorphological characteristics and flowing status distribution, in order to reduce the specificity of the models for better performance while generalize. In fact, the images used to feed the models were to scarce and specific of the Valencian Community. With a more populated dataset, other type of balancing operations may have to be considered. Further improvements are related to the possibility to variate the base model for fine tuning, testing some with different features and possibly with higher number of satellite images. By changing the above, also model

tuning will require important modifications related to number of epochs, learning rate and freezing and unfreezing procedures. Also, k-folder cross-validation can be added to directly retrieve metrics representative of multiple runs. In the end, the free and global coverage of the source data, Sentinel-2 satellite mission imagery, and the base of the methodology developed, may allow a cost-effective application for the identification of the flowing condition in NPRs in different river reaches worldwide. A collaboration between scientists and computer science researchers is necessary for further improvements and for placing NPRs under a magnifying glass. The output of a similar procedure can then feed model that also consider meteorological data for assessing the frequency and duration of the NPRs flowing conditions in a substantial number of reaches.

Bibliography

(s.d.). Tratto da Mindat: https://www.mindat.org/climate-BSk.html

(2021). Tratto da Hispagua:
https://hispagua.cedex.es/en/instituciones/confederaciones/jucar

Balaji, S. (2020). Tratto da Medium: https://medium.com/techiepedia/binary-image-classifier-cnn-using-tensorflow-a3f5d6746697

Brownlee, J. (2020). Tratto da Machine Learning Mastery:
https://machinelearningmastery.com/difference-test-validation-datasets/

Camarasa-Belmonte, A. M. (2016). Flash floods in Mediterranean ephemeral streams in Valencia Region (Spain). *Journal of hydrology, 541*, 99-115.

Carbonneau, P. E. (2020). UAV-based training for fully fuzzy classification of Sentinel-2 fluvial scenes. *Earth surface processes and landforms, 45(13)*, 3120-3140.

Cavallo, C. P. (2022). Exploiting Sentinel-2 dataset to assess flow intermittency in non-perennial rivers. *Sci Rep 12, 21756.*

Chen, Y. F. (2018). Extraction of urban water bodies from high-resolution remote-sensing imagery using deep learning. *Water, 10(5)*, 585.

CHJ. (s.d.). Tratto da Confederación Hidrográfica del Júcar: https://www.chj.es/es-es/medioambiente/cuencahidrografica/Paginas/Presentaci%C3%B3ndelacuenca.aspx

CHJ and TYPSA. (2020). *Estudio y caracterización de varios tramos de ríos de la demarcación hidrográfica del Júcar para la adaptación del régimen de caudales ecológicos.*

ESA - Copernicus HUB. (s.d.). Tratto da Copernicus Open Access HUB:
https://scihub.copernicus.eu/

EU-Parlament. (2000). WFD 2000/60/EC.

European Space Agency. (s.d.). *Sentinel Online*. Tratto da Sentinel Copernicus:
https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi/revisit-coverage

Fritz, K. C. (2017). Governance, Legislation, and Protection of Intermittent Rivers and Ephemeral Streams. In *Intermittent Rivers and Ephemeral Streams* (p. 477-507).

García, M. J. (1999). Use of geomorphological units to improve drainage network extraction from a dem: comparison between automated extraction and photointerpretation methods in the Carraixet catchment (Valencia, Spain). *International Journal of Applied Earth Observation and Geoinformation, 1(3-4)*, 187-195.

GEE. (s.d.). Tratto da Google Earth Engine: https://earthengine.google.com/

Huang, C. Z. (2018). Detecting, Extracting, and Monitoring Surface Water From Space Using Optical Sensors: A Review. *Reviews of Geophysics*.

Huilgol, P. (2020). *Top 4 Pre-Trained Models for Image Classification with Python Code*. Tratto da Analytics Vidhya: https://www.analyticsvidhya.com/blog/2020/08/top-4-pre-trained-models-for-image-classification-with-python-code/

IPCC. (2023). *AR6.*

IPH. (2007). *Instrucción Planificación Hidrológica.* Tratto da Gobierno de Espana: https://www.chminosil.es/es/2013-05-08-16-21-21/80-chms/807-

IVIA. (2023). Tratto da Instituto Valenciano de Investigaciones Agrarias: http://riegos.ivia.es/meteorologia

Keras. (2020). Tratto da Transfer learning & fine-tuning: https://keras.io/guides/transfer_learning/

Keras. (s.d.). *ResNet and ResNetV2*. Tratto da https://keras.io/api/applications/resnet/

Li, L. Y. (2019). Water body extraction from very high spatial resolution remote sensing data based on fully convolutional networks. *Remote Sensing, 11(10)*, 1162.

Magand, C. M. (2020). *Intermittent Rivers and Ephemeral streams: What water managers need to know.*

Manfreda, G. B. (2023). *Best practices in identification of Temporary Rivers, Rivertemp.*

Maswanganye, S. E. (2022). Use of multi-source remotely sensed data in monitoring the spatial distribution of pools and pool dynamics along non-perennial rivers in semi-arid environments, South Africa. *Geocarto International*, 1-20.

MATTM. (2000). Tratto da Gazzetta Ufficiale della Repubblica Italiana: https://www.gazzettaufficiale.it/eli/id/2008/08/11/008G0147/sg

Messager, M. L. (2021). Global prevalence of non-perennial rivers and streams. *Nature 594*, 391–397.

NASA. (2013). Tratto da USGS: https://www.usgs.gov/landsat-missions/landsat-7-data-users-handbook

O'Shea, K. &. (2015). An introduction to convolutional neural networks.

Parajuli, J. (2021). *Extracting surface water bodies from Sentinel-2 imagery using Convolutional Neural Networks.*

Prakash Maheswari, J. (2019). Tratto da Towards Data Science: https://towardsdatascience.com/breaking-the-curse-of-small-data-sets-in-machine-learning-part-2-894aa45277f4

Progressive Gardening. (2022). Tratto da https://www.progressivegardening.com/agricultural-meteorology/pft-[erm-e-x-100.html

Pu, F. D. (2019). Water-quality classification of inland lakes using Landsat8 images by convolutional neural networks. *Remote Sensing, 11(14)*, 1674.

Rinaldi, M. S. (2014). *IDRAIM-Sistema di valutazione idromorfologica, analisi e monitoraggio dei corsi d'acqua.* ISPRA-Manuali e Linee Guida, 113.

SAIH. (s.d.). Tratto da CHJ: http://saih.chj.es/chj/saih/glayer?t=p

Sanchis-Ibor, C. S.-B.-C. (2017). Channel forms recovery in an ephemeral river after gravel mining (Palancia River, Eastern Spain). *Catena, 158*, 357-370.

Saturn Cloud. (2023). *A Comprehensive Guide to Metrics in tf.keras.model.compile.* Tratto da https://saturncloud.io/blog/a-comprehensive-guide-to-metrics-in-tfkerasmodelcompile/

Seaton, D. D. (2020). Use of multi-temporal satellite data for monitoring pool surface areas occurring in non-perennial rivers in semi-arid environments of the Western

Cape, South Africa. *ISPRS Journal of Photogrammetry and Remote Sensing, 167*, 375-384.

Sentinel HUB. (s.d.). Tratto da Sentinel HUB: https://www.sentinel-hub.com/

Shah, T. (2017). Tratto da Toward Data Science: https://towardsdatascience.com/train-validation-and-test-sets-72cb40cba9e7

Sharma, S. (2017). *Activation Functions in Neural Networks*. Tratto da Toeards Data Science: https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6

TensorFlow. (s.d.). Tratto da https://www.tensorflow.org/

Unzueta, D. (2022). Tratto da BuiltIn: https://builtin.com/machine-learning/fully-connected-layer

Vezza, P. Z. (2017). *Manuale tecnico-operativo per la modellazione e la valutazione dell'integrità dell'habitat fluviale.* ISPRA–Manuali e Linee Guida, 154.

Wang, Y. D. (2019). Extracting urban water by combining deep learning and Google Earth Engine. *arXiv preprint arXiv:1912.10726.*

Wang, Z. &. (2022). Detecting Streamflow in Dryland Rivers Using CubeSats. *Geophysical Research Letters, 49(15).*

Yathish, V. (2022). *Loss Functions and Their Use In Neural Networks*. Tratto da Towards Data Science: https://towardsdatascience.com/loss-functions-and-their-use-in-neural-networks-a470e703f1e9#:~:text=A%20loss%20function%20is%20a,network%20models%20the%20training%20data

Yu, D. W. (2014). Mixed Pooling for Convolutional Neural Networks. *In Rough Sets and Knowledge Technology: 9th International Conference, RSKT 2014, Shanghai, China, October 24-26, 2014, Proceedings 9 ,* 364-375.