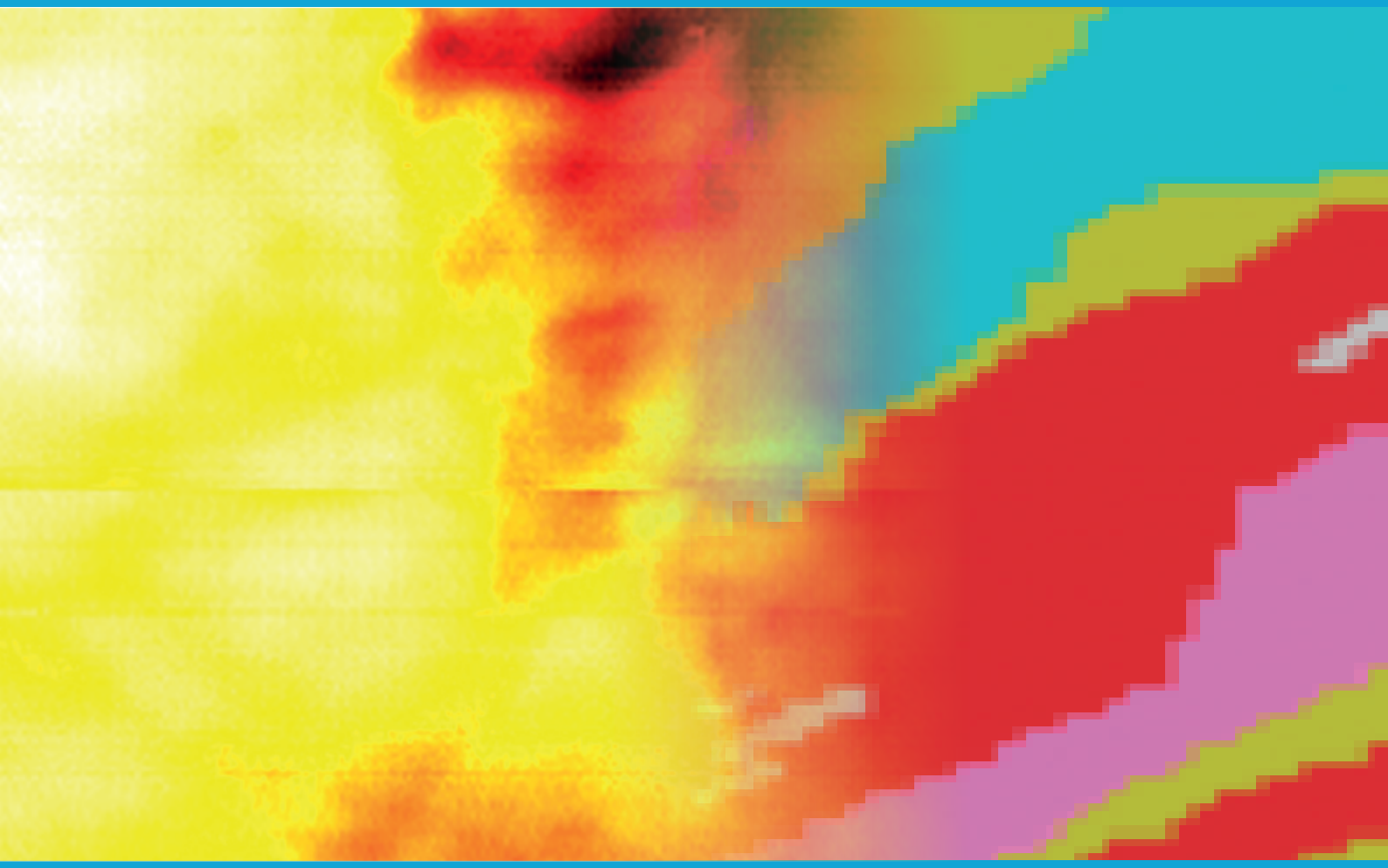
MSc thesis in Geomatics

Rock and soil classification using thermal and SAR data in deep learning models

Javier Martínez

2025



TU Delft

MSc thesis in Geomatics

Rock and soil classification using thermal and SAR data in deep learning models

Javier Martínez

October 2025

A thesis submitted to the Delft University of Technology in partial fulfillment of the requirements for the degree of Master of Science in Geomatics

The work in this thesis was carried out in the:

Geo-Database Management Center
Delft University of Technology

Supervisors:	Dr. A. (Azarakhsh) Rafiee
	Dr. RJG (Rémi) Charton
Co-reader:	Dr. RC (Roderik) Lindenbergh

Abstract

Fields like geology, mining, and construction have long relied on manual rock and soil classification methods, which remain time-consuming and labour-intensive. Accurate rock and soil classification is necessary for hazard assessment, urban management or resource exploration, which the rapid advances in artificial intelligence and remote sensing might facilitate. Though optical imagery has dominated remote sensing in the last decades, thermal data offers unique advantages such as sensitivity to material properties, sunlight independence or seasonal and diurnal changes responsiveness. This study explores the use of thermal imagery, along with SAR and NDVI data as complementary data, to leverage and automate the classification process with deep learning. Therefore, the aim is to develop two deep learning models capable of rock and soil segmentation with multi-modal datasets, and evaluate the performance of these models, Convolutional Neural Networks (CNNs) and Convolutional Long Short-Term Memory networks (ConvLSTMs) across diurnal, seasonal and multi-source scenarios as well as the effect of vegetation on the prediction results.

Results show CNN tends to have a strong overall performance, especially with thermal and SAR data. Instead, ConvLSTM excels at capturing temporal dependencies, improving the classification of most datasets and variations. The findings demonstrate the potential of thermal imagery to be used as a powerful classification tool when combined with deep learning methods for rocks and soils. They also demonstrate the slightly negative influence of vegetation on the predictability of the outputs. By combining spatial and temporal dynamics, the models offer an automated, scalable, fast and cost-effective approach to more traditional workflows. This work intends to contribute towards modernizing geological practices for more informed decision-making in geology, mining, urban planning, hazard and resource management.

Acknowledgements

First, and foremost, I would like to start thanking both my supervisors, Azarakhsh Rafiee and Rémi Charton, for the continuous guidance, patience and help throughout this project. Their involvement and professionalism made this work, not only possible, but enjoyable.

Of course, it goes without saying, that I would also like to thank my parents and uncles for providing me with constant support, my uncles and cousins for providing me with a caring home while carrying out my studies, and everyone who was around me during this time. They each formed a piece I needed to complete this puzzle. Thank you…

Contents

List of Figures

List of Tables

Acronyms

1. Introduction

For centuries, fields such as geology, mining, and construction have relied on traditional, manual methods—visiting outcrops, taking in-situ measurements and interpolating inaccessible areas—which, although still reliable, have seen relatively little change in recent decades (Marjoribanks, 2010). This is particularly surprising given the rapid advancement of technology in recent years and the availability of powerful technological tools like Geo Artificial Intelligence (AI) techniques to which other similar disciplines like agriculture, geomatics, environmental monitoring or urban planning have adapted and included to actively leverage these technologies in order to enhance their efficiency, scalability, performance and decision-making tasks (Sharifi & Mahdipour, 2024). Sharifi and Mahdipour (2024) explain how the recent trends in Geo AI for remote sensing applications include the integration of deep learning models which process and analyse satellite imagery, multispectral data to improve land cover classification and the use of SAR data for terrain mapping and disaster response.

All of the fields mentioned—including geology, mining, construction, geomatics, and urban planning—rely, to some extent, on accurate rock and soil classification (Marjoribanks, 2010). This is essential for risk assessment, such as identifying vulnerable areas to hazards like landslides (due to unstable materials) or flooding (due to heavy rains or nearby river floodplains) or even target prospection (Marjoribanks, 2010). At the same time, it supports prediction and safety measures, optimized resource allocation to reduce time and costs, smarter planning for suitable development of land or even high resolution mapping of geological structures and urban areas. With the recent advancements in remote sensing and computational methods (Sharifi & Mahdipour, 2024), new tools and techniques can now be adapted to enhance these tasks. Even though field work plays a crucial role in some of these areas, integrating in an efficient way these technologies into the aforementioned traditional disciplines holds great potential to reduce costs and time-consuming tasks by optimizing and complimenting the workflow, resulting in more informed decision-making.

Traditionally, optical imagery has been the dominant approach for rock and soil classification when remote sensing is employed. This has overshadowed other modalities which offer unique advantages. Thermal imagery, in particular, provides several technical strengths which can help perform a better classification over other types of imagery such as Radio Detection and Ranging (RADAR), yet it is often not associated to any classification procedure. It captures temperature-related physical properties of materials allowing for distinction through varied mineral compositions. Additionally, it is independent to sunlight, responsive to diurnal and seasonal changes and can perfectly complement other types of data like SAR or optical (Red, Green, Blue light model (RGB)), making it a valuable application for rock and soil identification. Thus, as thermal imagery becomes increasingly accessible and reliable, its value for these tasks grows substantially. In the second place, to explore these opportunities, the thesis will employ the use of Deep Learning (DL), a newer type of data-driven Machine Learning (ML) which employs multiple neural network layers to obtain high-level features from inputted datasets (Dehghani et al., 2023a). This will be done through two architectures: CNN and ConvLSTM. The research builds on the work of van Capel (2024) which focused on Local Climate Zone (LCZ) classification using CNN based solely on thermal imagery. In contrast, the present study expands the approach by integrating complementary imagery and a different DL architecture which focuses on temporal variation, ConvLSTM.

Nevertheless, even though the advancements prove promising, there are some persistent challenges faced such as limited availability of high-quality training data, robust and interpretable models, and the difficulty behind the integration of multi-source datasets (Sharifi & Mahdipour, 2024). Therefore, the thesis aims to exploit both the temporal and spatial relationships in thermal data—and complementary SAR data—, through LST and emissivity, to accurately classify rocks and soils, as well as face and attempt to solve some of the challenges previously described. By feeding thermal images (for CNN) and time-stacked sequences (for ConvLSTM) as inputs into the models, the goal is to classify a wide variety of rocks and soils based on thermal patterns, textures, properties, and contextual relationships to finally form fully predicted geological maps. Further on, the integration of auxiliary data—SAR

as input and Normalized Difference Vegetation Index (NDVI) for analytical purposes—will be tested and evaluated on its impact to the performance of these models. As a result, an optimized, robust model capable of performing reliable classification tasks under varying environmental and temporal conditions can be obtained.

1.1. Thesis outline

The project is divided into 7 chapters, each dealing with a different part of the workflow. In the first place, Chapter 2 consists on an extended review of several theoretical aspects in which the thesis is based. At the same time, it explains the background behind the topic of this thesis and concepts that are important to understand why the project works the way it does. On the contrary, Chapters 3 and 4 take a more practical approach, explaining the different sources, tools and methods (Chapter 3) as well as the parameter-tuning decisions (Chapter 4), carried out during the project to obtain the final results. Moreover, Chapter 5 displays the results obtained after the training and testing period of the deep learning models in addition to the experiments performed. Chapter 6 includes a general discussion on the impact that these decisions had on the results, as well as insights on the experiments and the limitations involving certain parts of the project. Chapter 7 summarises the findings and conclusions made from this study while providing some pointers into any possible future work. The appendices of the paper include a self-reflection, the geology maps used as ground truth and the tables reflecting the dates and times used for training and testing the neural networks. All the source code developed for pre-processing, training, and evaluation, along with the results obtained are openly available in the following repository: GitHub.

1.2. Motivation

The motivation for this thesis stems from several factors, both personal and purposeful, aiming not only to serve as a first step to solving a problem but to also contribute valuable insights to the fields of geology and geomatics. As mentioned previously, other fields should be able to also benefit from the results of this project. Accurate soil and rock classification is necessary in a wide range of disciplines including geology, geomatics, agriculture, hydrology, construction, and environmental management. The distribution and properties of surface materials is essential for hazard mitigation, resource management or land suitability. Instead, inadequate soil classifications can lead to construction failures or inefficient land use. Traditional geological mapping techniques often involve field sampling and manual interpretation, which still remain very valuable, but are also time consuming, labour-intensive, and spatially limited in certain cases. Therefore, consistent mapping can become challenging in many regions, to which advances in machine learning can offer opportunities to modernize them, offering supporting tools and data.

In studies like geology, field trips display some of the limitations behind traditional and manual methods employed in rock mapping and classification. In many cases, programming tools are barely employed during mapping, and only a few remote sensing techniques are performed for geological mapping. Even though these processes are still effective, to some extent, they can be costly, prone to human error, and labour-intensive. With the world of deep learning and advanced remote sensing in constant growth, integration of some of the methods used in these modern areas into these traditional fields could be combined to streamline geological workflows by producing tools that can support their current approaches. The thesis aims to bridge the gap between these two domains: applying modern geomatic approaches to enhance and support traditional geological methodologies. Therefore, exploring the potential of such techniques in geological applications. By introducing automated, data-driven approaches, geologists, urban planners and related professionals could all benefit from improved decision-making, reduced costs and safety planning-particularly in areas where access is restricted due to remoteness or budgetary constraints. These real-world challenges serve as motive to investigate alternative approaches such as thermal imagery and deep learning. The project represents an effort to modernize geology and combine it with the world of geomatics, making classification tasks accurate, efficient and cheaper as well as scalable and applicable to many complex environments through a supporting tool.

1.3. Objectives and Scope

The aim, or objective, of this work is divided into two parts:

1. Firstly, building deep learning models that can perform proper rock and soil classification with the use of thermal and complementary data.
2. Secondly, finding and refining the best model to execute such a task, which involves comparing two different U-Net architecture neural networks (CNN and ConvLSTM)—under varying conditions—for multi-class segmentation purposes in order to evaluate which one can yield better classification results.

In doing so, the scope of the thesis includes several parts such as:

1. Obtaining and pre-processing thermal and SAR data, as well as geological ground-truth labels extracted from digitized versions of geology maps.
2. Building, training and testing two different DL models for multi-class segmentation.
3. Discuss the influence of temporal dynamics like diurnal and seasonal data, and vegetation coverage in the performance of the models and determining the fittest model for the task.

Thus, the goal consists of producing a relevant classification process which different scientific sectors and companies can benefit from. Such stakeholders could leverage informed decisions in fields like urban management, hazard assessment, resource exploration, construction or terrain analysis using these models as supporting tools. The purpose of the research is to combine the fields of geology, geomatics and programming to provide a method that aids in automating tedious processes recently mentioned by reducing both the time and costs of these tasks.

1.4. Hypotheses and Research questions

To follow a guideline for the experiments conducted during the thesis, both hypotheses and research questions were developed based on the literature review made prior to the initiation of the models' development. The hypotheses tackle a more theoretical approach, to test how two different soils react to the temperature changes throughout the day, based on the investigations made by Abu-Hamdeh (2003) and their statement on thermal diffusivity of both materials. Since it works as a baseline to understanding the behaviour of certain soils, it can help determine how different thermal behaviours might be favoured or under-represented by the models in the final outputs.

- Clay soils gain heat slowly throughout the day but retain heat longer due to its higher moisture content and fine-grained texture.
- Sandy soils gain heat quickly throughout the day but loose it rapidly given its lower moisture content and coarse-grained texture.

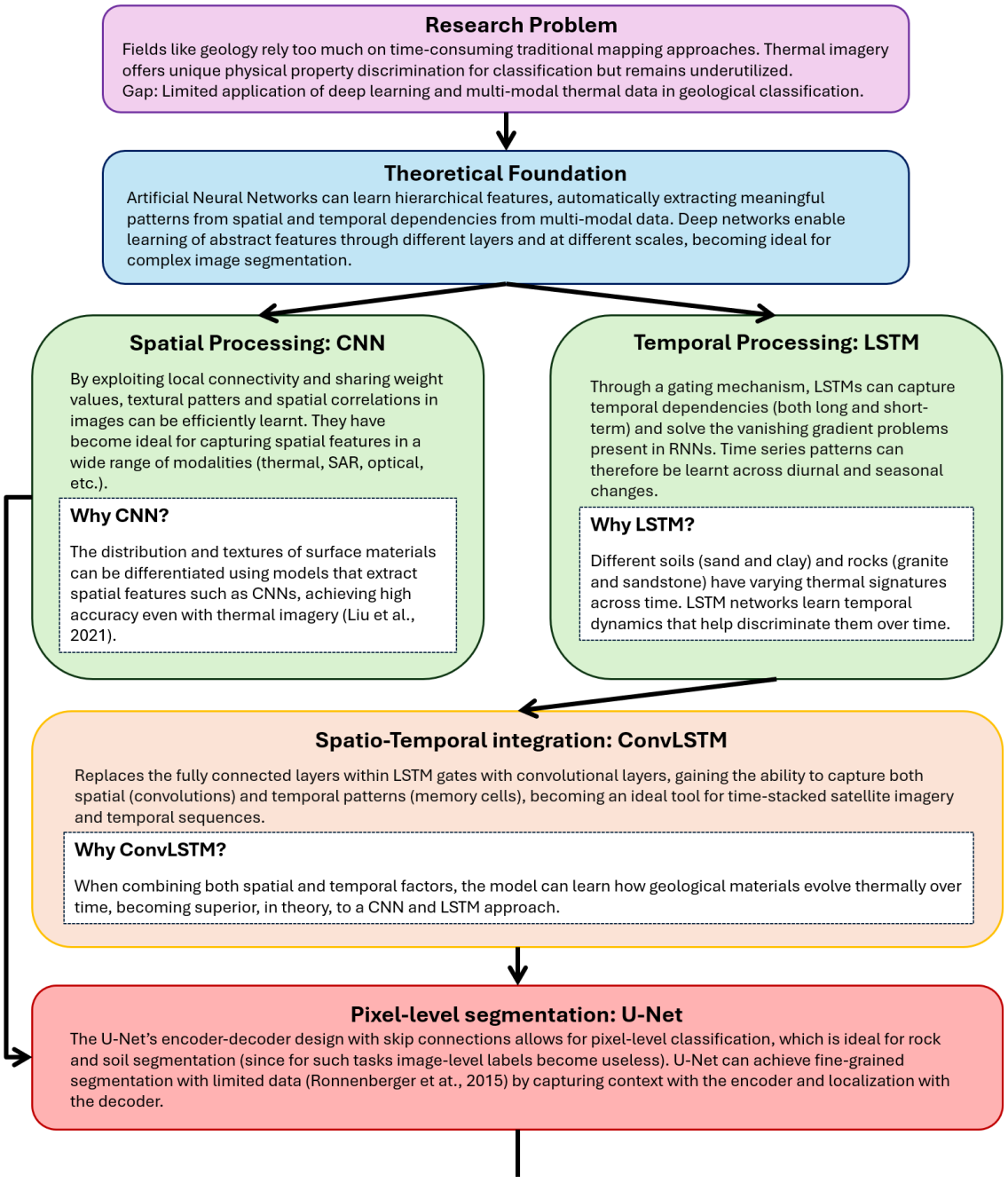
On the other hand, the research questions focus on a more practical side of the thesis, including the comparison of models, their performance, the effect of diurnal and seasonal factors or the effect of complementary data in the process. Several studies make emphasis on the challenges the use of thermal data presents (Rubio et al., 1997) for classification purposes, especially with emissivity ranges overlapping for similar and different rock and soil types. Instead, other authors confirm the potential of overcoming these issues through the use of several thermal channels and complementary data like SAR, (Zhu et al., 2022), (Ye et al., 2022), (Ndikumana et al., 2018). As stated by S. Liu et al. (2019), CNNs have become increasingly useful for remote sensing classification tasks, proving to be valuable for this study's purpose. Furthermore, since thermal values vary through time (during the day and overall throughout the year), the results could be compared to a spatio-temporally dependent model, to evaluate the importance of time in the performance of the models.

- To what extent can CNN and Recurrent Neural Network (RNN) deep learning methods help identify/detect different soil/rock types with thermal imagery?
- How does the model deal with very similar rock/soil attributes?

- How do temporal factors (diurnal and seasonal changes) affect the classification performance?
- If other types of data are included (SAR), does the performance/outcomes improve?

1.5. Theoretical framework

The theoretical framework of this study establishes the scientific foundation supporting the integration of both remote sensing and deep learning techniques to achieve automated rock and soil classification. Grounded in geology and geomatics, it explains how several data sources can be combined to capture spatio-temporal properties of Earth's surface materials. The framework provides a conceptual basis for understanding the decision made before and throughout the project, better reflected in Figure 1.1:



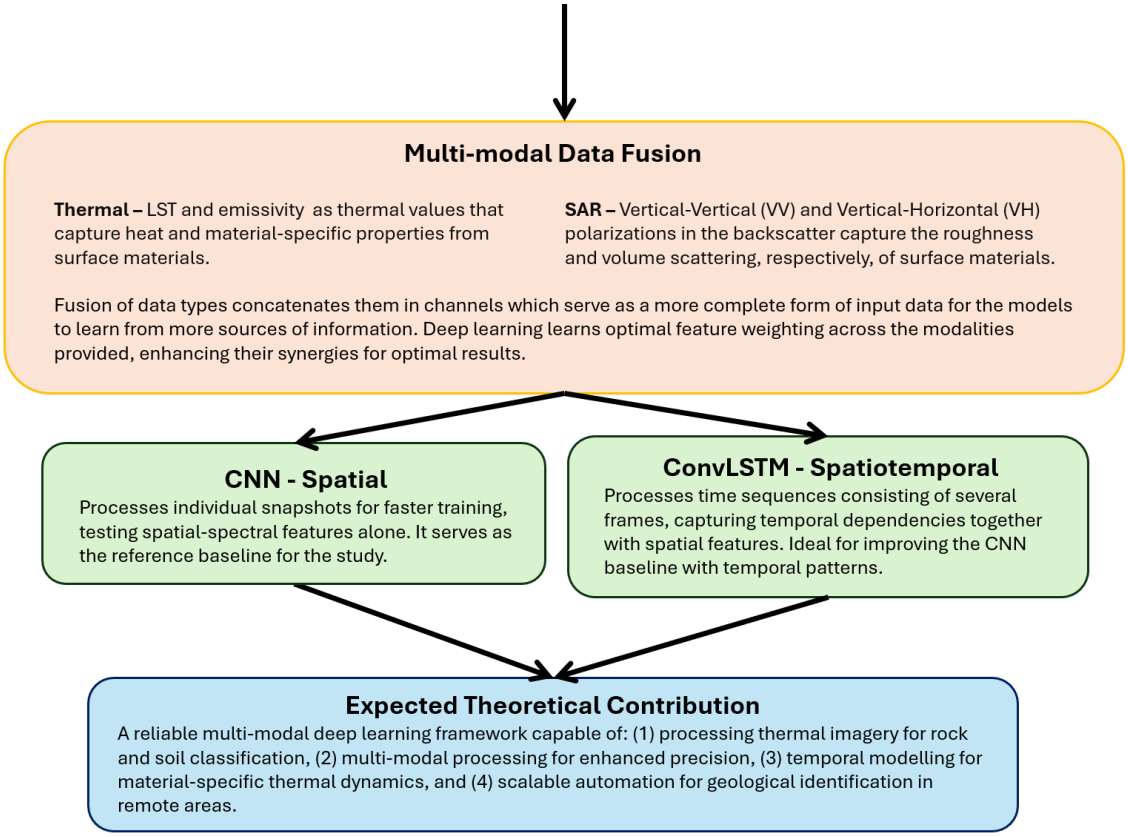


Figure 1.1. Methodological Framework for thesis workflow

2. Literature review and Conceptual Framework

2.1. Rock and soil classification systems

Rocks and soils both follow multiple and varied classification systems such as sedimentary, igneous and metamorphic for rocks—in which systems like the QAPF (Streckeisen, 1974), depositional factors (Hallsworth & Knox, 1999) or textures (Nelson, 2000) are common— and the wide range of soil systems present like United States Department of Agriculture (USDA)'s Soil Taxonomy (Staff, 1999) or the USCS (Harlianto et al., 2017a). They aim to standardize their description and interpretation across regions and disciplines. Classifying the surface is a necessary task given the variability of resources. Without classification systems, it is hardly possible to identify different materials from different locations. In turn, it is impossible to determine the suitability of a rock and, therefore, carry out any management practices. Classification systems offer a common structure and framework for understanding and managing rocks and soils and exploit their resources with an appropriate approach. In addition, they offer a practical way to perform geological maps that can be understood worldwide. Since the first geological map developed in 1815, the evident importance of geological maps led most of the advanced countries to rapidly explore and document this new way of representing geological land distribution (Berra et al., 2024). The same authors, Berra et al. (2024), reveal the fundamental importance of a unified system for the production of homogeneous coverage of entire nations and continents. And although this process has evolved with every generation, especially in the second half of the XIX and XX centuries, and most of the bulk work has been performed, all these techniques would still benefit from the workflow presented in this project, to test their reliability and aid in local areas of hard access.

Both rocks and soils present diverse physical, chemical and biological properties which include, for instance, colour, texture, structure, mineral composition or the presence of organic matter. All of these vary based on climate, topography and living organisms for soils and on time, tectonic, temperature, pressure, and transportation for rocks. These conditions work in complex and distinctive ways, resulting in a extensive list of surface materials across different parts of the world. With all this spatial variability, robust classification systems are necessary for researchers to communicate results and make informed decisions.

2.1.1. Rocks

Rocks have many different classification systems which help describe and understand Earth's materials. For simplicity, rocks are traditionally categorized based on their origin: igneous, metamorphic and sedimentary. Nevertheless, each category contains more detailed metrics and systems which allow to distinguish the characteristics of rock diversity. Some of these are mentioned below:

- **Sedimentary rocks -** these rocks are formed from the accumulation in deposits of pre-existing rocks or skeletons of past living organisms. They are mainly categorized into three subdivisions: detritic, carbonatic and organic. Then, for further precision, they are distinguished based on their depositional texture, transport, composition and grain size. The combination of these factors reflect their formation history. Each category is detailed in Hallsworth and Knox (1999), as each subcategory also contains their own classification system.
- **Igneous rocks -** formed from hot, molten magma that crystallizes and solidifies. Unlike sedimentary rocks, igneous rocks are distinguished based on their mineral or chemical composition. Based on their cooling process, igneous rocks can be divided into two main categories:

volcanic and plutonic (further explained in section 3.1). Classification of rocks inside these two classes are performed with two main systems that fulfil this purpose: the Quartz-Alkali feldspar-Plagioclase-Feldspathoid (QAPF) developed by Streckeisen (1974) focusing on the mineral composition and the Total Alkali-Silica (TAS), designed by Maitre (2002) and distinguishing most volcanic rocks based on their proportions of silica and alkalis.

- **Metamorphic rocks -** starting out as some other type of rocks, they are changed from their original igneous, metamorphic or sedimentary form due to high pressure and heat, or even mineral-rich fluids. These, on the other hand, are primarily classified based on their texture—notably, whether they are foliated or non-foliated, indicating if their minerals are aligned or not, respectively—and the grade, a term used to describe the temperature and pressure they underwent (Nelson, 2000). Other metrics used include petrogeny and facies.

2.1.2. Soils

Soil bodies are usually considered as Quaternary sediments (Han et al., 2023) and are fundamental to the Earth's surface. They were originally rocks that through erosion, weathering, and stripping turned into looser materials. Therefore, their properties differ depending on their geological conditions, formation, time, and environment (Han et al., 2023). Just like rocks, soils also present several classification procedures, given that soil is not uniform but it is made up of a blend of minerals, water, air and organic matter (Enviroliteracy Team, 2024). It is a crucial framework necessary to make precise decisions for fields like agriculture (fertility assessment or water retention), construction (feasibility and hazard mitigation), and geology (interpretation of surface processes and depositional environments). The purpose of a soil classification system, as explained by several studies Brevik et al. (2016), FAO and IUSS (2022), and Hempel et al. (2013), is to provide a common scientific language that facilitates the comparison, exchange and extrapolation of soil information, results, and experience on issues among scientists, unifying everything into a soil nomenclature. Scientists defend two different points of view regarding a single classification system (Nikiforova, 2019). The first one involves the use of different systems for different purposes (Ibáñez & Boixadera, 2002) and the second one a unified single system (Fridland, 1986). Although this ambiguous and split state for soil classification has remained the same throughout the years, the aim is still to generate a universal system, which is still considered challenging due to the complexity, nature and spatial variability of soils (Fridland, 1986; Ibáñez & Boixadera, 2002). Some of the current most common systems include:

- **USDA's Soil Taxonomy -** a highly detailed system that focuses on a hierarchical and complex structure of six levels, inluding: Order, Subsorder, Great Group, Subgroup, Family and Series (Staff, 1999). All of them based on horizons, climate, soil behaviour and properties. It offers a framework for understanding soil genesis and potential management influencing other systems worldwide as a result. For Staff (1999), soils is a natural body comprised of solids, liquids and gases occurring in the land surface characterized by horizons and layers distinguishable from each other from additions, losses, transfers, and transformations of energy and matter.
- **World Reference Base (WRB) for Soil Resources -** on a more international level, this system also contains a hierarchical structure (Reference Soil Groups (RSG)) focusing, in this case, on material composition and properties (FAO & IUSS, 2022) rather than formation processes as in Soil Taxonomy. This system defines soil as any material within two meters of the Earth's surface that is in contact with the atmosphere.
- **Unified Soil Classification System USCS -** a simpler approach for geotechnical mapping rather than petrological, Harlianto et al. (2017a) provide a guide into a classification system based on two letters. The first one indicates the dominant type of soil while the second one describes the plasticity or gradation as seen in Figure 2.1. It is widely employed in civil engineering and infrastructure development due to its insights into soil mechanical behaviour, capacity and permeability.

Unified Soil Classification System (USCS)

First and/or second letters		Second Letter	
Letter	Definition	Letter	Definition
G	gravel	P	Poorly graded (uniform particle sizes
S	sand	W	Well graded (diverse particle sizes)
M	silt	H	High plasticity
C	clay	L	Low plasticity
O	organic		

Source: Harlianto et al. (2017b)

Figure 2.1. Soil classification system presented by the USCS

2.2. Thermal imagery

Prakash (2000) defines thermal remote sensing as a branch of remote sensing that focuses on the acquisition, processing and interpretation of data that are acquired mainly in the Thermal Infrared (TIR) region of the Electromagnetic (EM) spectrum. In the study, the process is described as measuring the radiation 'emitted' from the target's surface. Totally contrary to what is done in the more commonly used optical imagery, where it is the 'reflected' radiation that is measured. Thus, in principle, thermal remote sensing is very different to the optical region. And, though it is still not explored to full extent, it contains potential for a wide range of applications.

TIR is described in Mineo and Pappalardo (2021) as the spectral range used to collect thermal data. The same study explains the use of Infrared Thermography (IRT) is the application of TIR sensors, used to visualise or analyse thermal patterns, becoming exceptionally useful since it can penetrate smokes, aerosols, dust and mists. It is Ye et al. (2022) that performed a study stating the use of remote sensing techniques, which involve TIR data, contains useful information that helps distinguish features in the land surface, hence, deeming it as reliable for land classification by identifying differential patterns in thermal imagery. When analysing thermal data, especially emissivity and LST, large differences in the shape of emissivity curves through time imply that hyperspectral TIR remote sensing techniques are the ones that help distinguish subtle features of the land surface. Due to the richness in information of these spectral channels (Ye et al., 2022), land classification is possible.

Thermal imagery is obtained with the help of several instruments. Sensors like ECOSTRESS, ASTER or Landsat TIRS are equipped with special TIR detectors that employ physical laws such as Stefan-Boltzmann's, Planck's or Wien's laws to record the radiance, which is later on converted into actual temperature values after a processing stage. Despite this, thermal remote sensing relies on two major laws as described in Payra et al. (2023):

- **Stefan-Boltzmann law -** the surface temperature of any object displays its specific properties, calculated by the Boltmann formula:

$$M(T) = \sigma T^4 \tag{2.1}$$

  where:

  - $M(T)$ is the total radiant existence ($\mathrm{W{\cdot}m^{-2}}$),
  - σ is the Stefan–Boltzmann constant ($5.6697 \times 10^{-8}\ \mathrm{W{\cdot}m^{-2}{\cdot}K^{-4}}$),
  - T is the thermodynamic temperature of the blackbody (in Kelvin, K).

  According to this law, change in temperature results in a large change in radiation.

- **Wien's displacement law -** describes the relation between temperature recorded and the wavelength of radiation. The wavelength of radiation that a blackbody emits is inversely proportional to the temperature of it.

$$\lambda_{\max} = \frac{A}{T} \tag{2.2}$$

  where:

  - $\lambda_{\max}$ is the wavelength of maximum radiation intensity (μm),
  - A is Wien's constant (2898 $\mu\text{m} \cdot \text{K}$),
  - T is the thermodynamic temperature (K).

Once thermal imagery is acquired, there are different ways in which data can be interpreted and analysed. Typically in the form of radiance, further processing allows the extraction of several physical properties of the Earth's surface materials. Amongst the many approaches and data that can be extracted from thermal imagery, the following tend to be used commonly:

- **LST -** also referred to as the kinetic temperature, it consists of the temperature at the surface of the ground or target body (Figure 2.2). It is a measure of the amount of heat energy the target contains (Prakash, 2000), measured in either Kelvin (K), degrees Celsius (ºC) or Fahrenheit (ºF). On its own, LST might not be the best approach for rock and soil classification, since it highly depends on the climate and atmospheric temperature, however it can provide valuable thermal information when temporal series and changes are taken into account. This is due to the fact that it can help differentiate classes based on how rapid these materials heat up or cool down throughout diurnal and seasonal conditions.
- **Emissivity -** corresponds to the ability to emit of a real material in comparison to a black body (Prakash, 2000). This property varies with the composition and geometric configuration of the targets. It is a geophysical parameter, that ranges from 0-1, derived using a ratio between the actual emitted radiance of the target compared to a perfect blackbody at the same thermodynamic temperature (Norman & Becker, 1995; Salisbury & Walter, 1989). Some of the studies mentioned indicate it can be used to determine composite variations as well as the structure of the terrain elements analysed. As mentioned in the following section, although complicated at times, it can be used as a suitable thermal attribute for differentiation purposes.

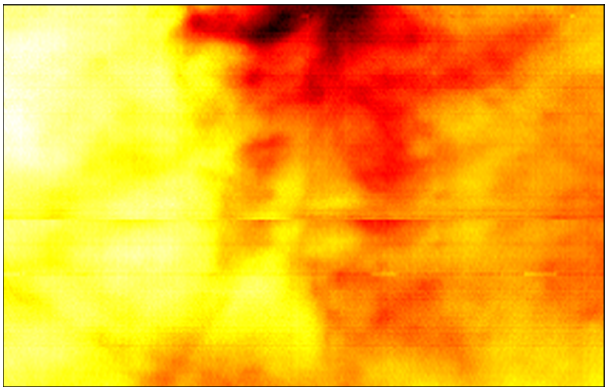
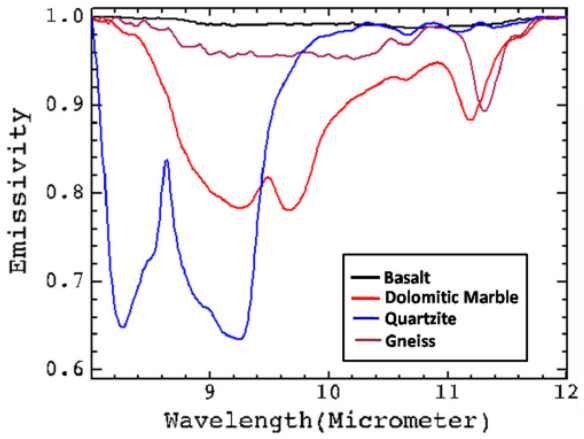


Figure 2.2. Example of an ECOSTRESS thermal LST image obtained from one of the areas of interest, in this case, Santa Olalla del Cala

In the case of rocks and soils, their thermal parameters are not only governed by their chemical composition but also impacted by surrounding vegetation cover growing above them (Guha & Kumar, 2014). Based on Rani et al. (2018), temperature variations are sensitive to other factors like land cover and moisture, hence TIR is deemed extremely useful to distinguish and differentiate minerals contained in materials. Nevertheless, many studies suggest that the values of some minerals can overlap the values of others, especially in the case of emissivity. Rubio et al. (1997), Mineo and Pappalardo (2021), Rani et al. (2018), and Guha and Kumar (2014), amongst other literatures, demonstrate a wide spectrum of emissivity values recorded for different rock and soil categories, demonstrated in Figure 2.3. They

also state that, given the varied mineralogy some of these rock categories contain, emissivity tends to be a weighted average of their mineral composition. Just like NASA's ECOsystem Spaceborne Thermal Radiometer Experiment on Space Station (ECO) Spectral Library, the studies display the results of emissivity for several examples and serve as reference in terms of reflectance spectra for rocks and soils. Using a wide range of sedimentary, igneous and metamorphic rocks, Mineo and Pappalardo (2021)'s results indicate that is it quite complicated to obtain univocal values, and instead, the majority of emissivity values are described as ranges, which fall between 0.83-0.99: sedimentary rocks standing between 0.91-0.99, metamorphic at 0.89-0.99, and igneous having the largest range with 0.83-0.99. Regarding soils, Rubio et al. (1997) provides the values of emissivity for soils, demonstrating their results to be between 0.900-0.999. This work also states that bare rock lies in a similar range. It was Guha and Kumar (2014) that suggested the use of 0.96 as the acceptable approximation of emissivity values for geological materials.



Source: Rani et al. (2018)

Figure 2.3. Laboratory derived emissivity spectra of several rock examples collected using FTIR spectrometer

Even though the values of emissivity for rocks and soils tend to be similar and, in occasions, overlap, Zhu et al. (2022) demonstrated that they present differences in thermal properties. The literature relates how soils and Quaternary loose sediments contain better heat storage capability while bedrocks present higher thermal conductivity. Even inside soil categories, there are differential thermal variances, showing a decrease in thermal conductivity based on increasing grain size. In addition, rocks' heat absorption effect is much better than that of looser sediments, however, soils retain heat better due to larger water content. The study gave valuable insight on ground surface material behaviour under thermal imagery, possibility of temporal sequences having a great impact on the results and, in turn, in developing the research questions for this thesis.

As mentioned before, taking into consideration all these studies, with the emissivity results they present, it is hard to find a clear way to classify rocks and soils based on this concept. Most of the literatures state the difficulty of using emissivity as the only classification factor, since, in some cases, value ranges can overlap (Rubio et al., 1997). Thus, the same study suggests that combinations of LST and emissivity can enhance this material discrimination.

2.3. SAR as complementary data

As a secondary part of the project, SAR data will be implemented as a complementary data type to enhance, as far as possible, the classification process. There are several reasons as to why SAR has been chosen as additional information to feed the models.

Sentinel-1 is able to acquire images regardless of the weather conditions present at the time of taking the measurements (Ndikumana et al., 2018). Therefore, cloud-covered skies are not a problem. SAR characterizes structural and dielectric properties of the ground surface, which when combined with

optical data, yields a comprehensive observation of ground targets (S. Liu et al., 2019). Thus, it possess great potential to enhance the accuracy of many classification processes. Studies like S. Liu et al. (2019), explore the idea of employing LSTM—explained in the following section—for agricultural classification purposes. It indicates that temporal dependencies in NN, with SAR data, improve the models' performance. Instead, Ndikumana et al. (2018) integrates CNN to map land use and cover with the use of SAR data. They explain how the implementation of both optical and SAR improves effectively the classification process, indicating the potential of SAR when used in combination with other data types.

Therefore, both studies suggest and demonstrate that SAR data as complementary data are reliable enough to improve the accuracy of both types of DL methods—hence the reason why it will be used on par with thermal imagery at the end of the project.

2.4. RGB as complementary data

Optical imagery has long been the most frequently used form of remotely sensed data for land use and land cover mapping (S. Liu et al., 2019). For instance, the Sentinel-2 optical sensor is well suited for monitoring agricultural environments, and when combined with other modalities such as SAR, it can provide a valuable complement to other observations (Ndikumana et al., 2018). Although cloud cover can hinder the availability of usable optical images, multi-source integration with optical data has been, in the literature, identified as an important avenue for future research.

RGB optical imagery in particular offers extremely high spatial and temporal resolutions at comparatively low acquisition costs (Mei et al., 2022), since most of them are freely available open source portals. Due to their many applications because of their rich structural and textural information, they have become important for object identification (Mei et al., 2022). When incorporated alongside thermal imagery, these additional details can potentially help CNNs and ConvLSTMs better distinguish between different soil and rock units. This has been proven effective when integrated with other sensing modalities in other studies. For example, Shebl et al. (2021, as cited in (Han et al., 2023)) demonstrated that combining optical Sentinel-2 data with gamma-ray imagery enhanced the diagnostic characterization of lithological units in Egypt. Therefore, using optical data together with thermal information has the potential to improve feature discrimination in classification tasks, which was tested as a potential complementary data during this project but was ruled out of use due to the limited amount of available imagery during temporal alignment with thermal data.

2.5. NDVI as complementary data

The NDVI was introduced by Kriegler et al. (1969) (as cited in Huang et al. (2021)) to simplify vegetation detection in multispectral imagery. It consists of a simple band transformation where red radiation (Band 4 - 665nm) is subtracted from Near-Infrared (NIR) radiation (Band 8 - 842nm) and then divided by NIR radiation plus the red radiation. The result is a newly simplified image called the NDVI. It has become a useful tool to quickly delineate the spatial and temporal distribution of vegetation in many areas like agriculture, land-use or environmental management (Huang et al., 2021; Lasaponara et al., 2022).

The reason to add NDVI arises from the fact that vegetation has a great impact on thermal values. Therefore, utilizing a vegetation index can help determine the amount of vegetation in certain areas which might be affecting the model's predictions. Afrasiabi Gorgani et al. (2013) state in their study the negative correlation between NDVI and LST. In their work, they indicate how the lowest LSTs are usually found in areas of high NDVI and, therefore, areas with less vegetation tend to experience higher land surface temperatures. Such an impact on thermal values plays a crucial role in the classification performance of the models, as the thermal value being recorded can sometimes simply belong to that of a tall tree instead of a surface material. Since vegetation can have such an effect on a pixel's thermal value, it is important to account for it when carrying out classification processes based on land surface. Thus, NDVI has the potential to become a very complete auxiliary data type that helps determine the

most vegetated areas and their impact on the classification results, indicating the degree of accuracy the models can reach in more or less vegetated areas.

2.6. Deep learning

In the geological environment, which involves both soils and rocks, reducing human intervention is crucial for achieving accuracy and scalability, particularly as resources are over-consumed due to rapid population growth. As noted by P. Srivastava et al. (2021), failing to adopt modern automated methods could hinder the efficiency needed to meet these demands. To maintain the sustainability of human society, advances in computational capacity have accelerated. Thus, AI has become a fast-growing trend in remote sensing methodologies for analysis, especially when large and complex datasets are involved (Amato et al., 2023).

AI refers to machines, with intelligent software, designed to carry out tasks that usually require human cognition—such as image recognition, speech processing, and decision-making (Goodfellow et al., 2016). However, traditional AI rely on hard-coded knowledge, making them rigid and limited at times, especially in unfamiliar scenarios (Goodfellow et al., 2016). Hence, ML was developed as the capability of AI to acquire its own knowledge by learning and extracting patterns from raw data without explicit programming. As datasets grow in complexity into unstructured and high-dimensional data (e.g., satellite imagery), a software closer to human-level understanding is required. Therefore, DL emerged as a special branch of ML capable of learning hierarchical features (Goodfellow et al., 2016), enabling the software to build complex features out of simpler ones. Contrary to ML methods which require more manual feature engineering, DL instead—especially Artificial Neural Network (ANN)-based ones—automatically learn to discriminate features, resulting in more scalable and accurate classification systems (P. Srivastava et al., 2021).

According to Amato et al. (2023), DL models excel in analysing unstructured geospatial data due to their ability to capture spatial and temporal dependencies, making them particularly effective for remote sensing, where time-series and pixel variability are essential and its applications. The study also defends the idea that when spatial and temporal information is involved, DL approaches outperform classical ML in accuracy for tasks like land cover classification, thermal anomaly detection, and lithological mapping. Deep learning for image classification in remote sensing is normally used in one category - semantic segmentation. Semantic segmentation assigns a class label to each pixel in an image (Shibuya & Hotta, 2022). These techniques always involve training certain ANNs on fully annotated datasets, often collected from the same sensor, in order to classify new, unlabelled images with high accuracy and precision. Thus, becoming usable in practice and in many different contexts.

2.6.1. Artificial Neural Networks

The human brain is a layered structure, where each layer consists of a collection of nerve cells which receive the name neurons. As explained by Shibuya and Hotta (2022), information in the brain is fed forwards and backwards between lower and upper layers and vice-versa, enabling dynamic learning and feedback. The lower levels handle simpler signals and the upper ones more abstract features. Neurons play an important role in processing and propagating the information received. Inspired on this concept, ANNs are computational models, a complex function that connects many layers with artificial neurons. Neurons receive the inputs, apply a non-linear mathematical function, and pass the output to the following layer. The higher the number of layers, the deeper the network is, enhancing the capacity to learn complex features. They are designed to implement deep learning models (P. Srivastava et al., 2021). Just as the brain, their core elements are:

- **Neurons -** the processing units, where each performs a weighted sum of the input received followed by an activation function (e.g., ReLU, sigmoid, etc.) (Goodfellow et al., 2016).
- **Layers -** groups of neurons that transform the data. There are three types of layers: input, hidden and output.
- **Connections (weights) -** numerical parameters that adjust themselves throughout training in order to optimize the performance of the model.

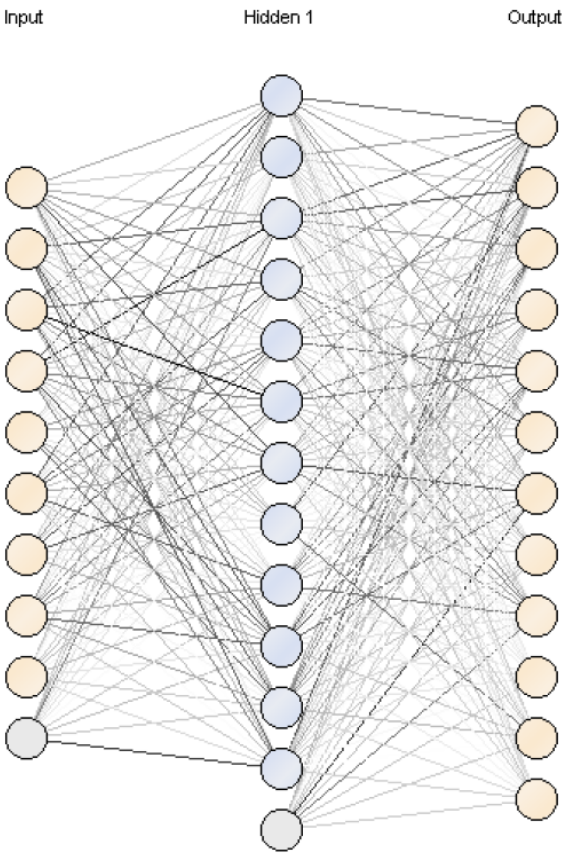


Fig. 2. Neural Network Configuration

Source: Harlianto et al. (2017b)

Figure 2.4. Demonstration of the configuration of a NN as explained in section 2.6.1

ANNs are structured, normally, intro three types of layers, defined by Harlianto et al. (2017b) and displayed in Figure 2.4:

- **Input -** receives the raw input data such as pixel values.
- **Hidden -** carries out the internal transformations with non-linear functions. The number and size of these layers determine the network's learning capacity.
- **Output -** generates the final prediction for classification results, often with 'softmax' or 'sigmoid' activations (explained in section 2.6.2).

A decision tree (a variety of binary tree) is used to make the decision, at each node, of a single input variable that will be passed on. Leaf nodes—the last node of the tree—contain the output variable, which is the one used to make the final prediction. These predictions are reached after making split decisions while walking through the nodes of the trees until reaching a leaf node and yielding its class. These type of trees are fast to learn and predicting, having a high variance and the ability to output accurate predictions (Harlianto et al., 2017b).

Among the various types of ANN architectures, the two most widely used models for remote sensing techniques with classification purposes are the CNN and RNN. The main difference between these two types of architectures is the flow of information. CNNs are built to extract spatial features, making them efficient for image classification, where data flow in one direction—from input to output (Yin et al., 2019). Instead, RNNs are designed for sequential time-series data, performed with recurrent connections where data follow a bidirectional flow through feedback loops (Yin et al., 2019). This allows the model to preserve temporal dependencies across time steps, becoming suitable, for instance, for satellite time-series analysis (Hochreiter & Schmidhuber, 1997).

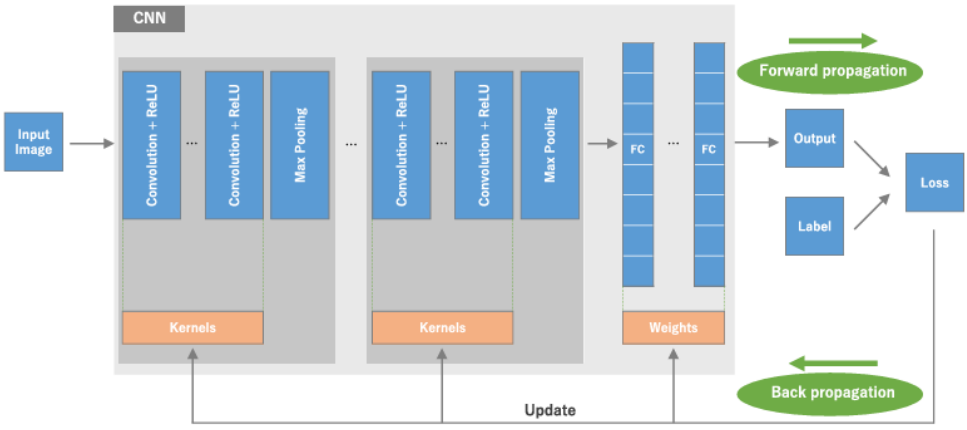
Since there is no human influence, the models have a strong reliance on the quality of the input data for the training stage. The model must learn from raw data by obtaining relevant patterns after performing crucial pre-processing steps that involve normalization and structuring (P. Srivastava et al., 2021).

2.6.2. Convolutional Neural Network

CNN are a powerful type of feed-forward NN that specialises in pattern recognition. It was developed with the purpose of managing image recognition tasks and, nowadays, perform remote sensing classification (S. Liu et al., 2019). In geological contexts, for example, CNNs can be used for classification, and in some cases, achieve high levels of accuracy when applied to emissivity images (H. Liu et al., 2021).

These type of networks exploit spatial hierarchies in grid-shaped data (like images) through local connectivity, where each neuron is only connected to a small patch of the previous layer (Dehghani et al., 2023b). It enables the network to learn, in an efficient way, about textural features and spatial relationships including edges, patterns and so on (S. Liu et al., 2019). Typically, a CNN contains three core components denominated building blocks described in Dehghani et al. (2023b) and Yamashita et al. (2018) and visualised in Figure 2.5:

- **Convolutional layer -** uses learnable kernels (e.g., 3×3 filters) that slide across the input image, computing dot-products to produce feature maps based on textures, patterns or edges. These filters enable the network to automatically extract features from low- to high-level abstractions. Figure 2.6 demonstrates the idea clearly.
- **Pooling layer -** reduces the dimensionality of feature maps via operations like max or average pooling, lowering computational load yet retaining the essential information for the next layer.
- **Fully connected or Dense layer -** feature maps are flattened and learn non-linear combinations of the extracted features after several convolutional and pooling layers. At this stage, each input node's value is multiplied by an associated weight—representing the importance of that input—and its corresponding bias is added. The sum of these weighted values is passed into an activation function that helps determine the final output. It is often passed through a softmax activation function, which generates the final prediction.



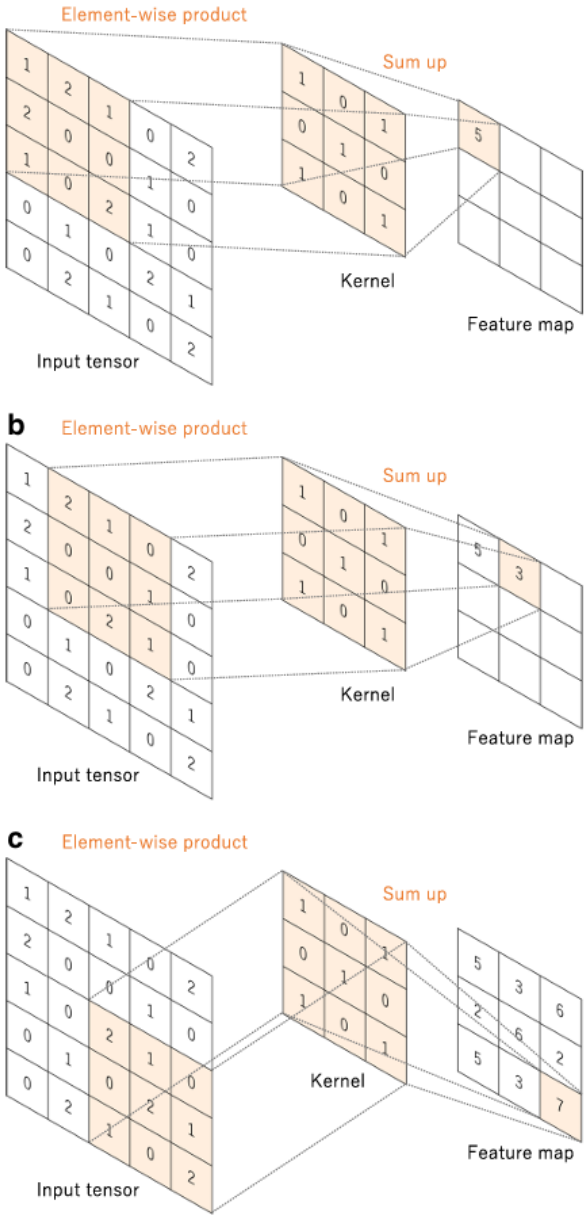
Source: Yamashita et al. (2018)

Figure 2.5. Representation of the architecture and training process a CNN goes through. Consists of building blocks with convolutional and pooling layers and final fully connected layers

The use of so many activations is due to the fact that CNNs benefit from non-linear activation functions such as ReLU or sigmoid, and even the use of batch normalization (detailed in section 3.5) to improve training stability and performance (S. Liu et al., 2019). The final feature map is run through 'ReLU'. It sets all negative values to 0 and all the positive values yielded to 1. Thus, it creates nonlinearity and a new feature map while speeding up training. This new feature map suffers one last convolution,

in a process called max pooling, which obtains patches from the input feature map, outputting the maximum values in each patch and discarding the rest (Yamashita et al., 2018). At the same time, the average value can be calculated with mean pooling. In addition, batch normalization standardizes activations to reduce internal covariance shifts (S. Liu et al., 2019).

The hierarchical design of CNNs lets early layers capture simple features, whilst deeper layers manage to handle and learn from more abstract representations of objects or regions (Yamashita et al., 2018). Further on, pooling layers will extract the most significant activations, preventing common problems such as overfitting by reducing the complexity (Yamashita et al., 2018). The final dense layer combines the newly learned features through spatial dimensions and performs the classification. Thus, these models become suitable for remote sensing applications, especially for classification purposes, due to the fact that they make use of the correlation that exists in certain images, tolerating small shifts.



Source: Yamashita et al. (2018)

Figure 2.6. Convolution operation with a standard kernel (3x3). It is applied across the input, producing element-wise products between each element of the kernel. At each location, the input is calculated and summed to obtain an output value in that position, called feature map

2.6.3. Convolutional Long-Short Term Memory

Time series data consist of observations collected over successive periods of time (Xavier, 2019). They generate temporal datasets where the order and timing of events is crucial. Therefore, to capture the dependencies between measurements across time steps, special models are required. In the context of NNs, they become reliable for classification or anomaly detection. To handle these dependencies, RNNs were built. They consist of a family of NNs designed to handle sequential data by keeping a hidden state capturing those temporal features (Wenjie, 2023). Unlike CNNs, RNNs feed the output of currently computed neurons using previous data from previous time steps. The output at time t is influenced not only by the input X_t but also by the hidden state h_{t-1} from the previous step (Ndikumana et al., 2018). The recurrent behaviour permits them to extract temporal patterns through time. The same study details the variants inside RNNs, including the standard RNN, the Gated Recurrent Unit (GRU), and LSTM. These last two improve the traditional RNN by perfecting the training process particularly dealing with vanishing and exploding gradient problems.

One critical limitation to standard RNNs is capturing long-term temporal dependencies. As just recently mentioned, it is caused by vanishing and exploding gradient problems (Yan, 2016). Gradients consist of values which indicate the amount by which certain parameters need to be changed to mitigate the loss function error in a NN (Engati, 2023). This is performed through a process known as Backpropagation Through Time (BPTT), described in the same paper as the process in which the chain rule of calculus is applied to propagate the error from an output layer to an input layer. During this process, the weights and biases of the network are updated, in practice, improving the performance of the network. Unfortunately, when the derivatives of the activation functions are too small (less than 1 or 0), repeated multiplication across several time steps causes the gradients to shrink exponentially, yielding the vanishing gradient. The model now fails to learn properly these long-term patterns. Opposite to the vanishing gradient problem is the exploding gradient problem. It occurs when derivatives are large, causing gradients to grow exponentially (larger than 1). In turn, the training destabilizes and yields erratic behaviours. These issues motivated the development of more robust models, whose architectures could solve these important gradient problems, and the solution was the LSTM.

Hochreiter and Schmidhuber (1997) introduced LSTM networks to address the limitations of traditional RNNs. They are used for prediction of future things from past events over a relatively short period of time. A generative model used to predict next data based on previous data (Xavier, 2019). The LSTM unit maintains two distinct cell states:

- **Cell state (C_t)-** stores the long-term dependencies and is considered the long-term memory as it runs through the whole sequence. Thanks to the forget, input and output gates, it can add, remove or retain information.
- **Hidden state (h_t)-** captures the short-term memory output dynamics, getting updated at every time step, holding only the information to the immediately relevant output. It then passes it to the next step.

The flow of information in ConvLSTM is controlled by three different gates defined in the paper as:

- **Forget gate (f_t)-** states what portion of the previous cell state should be discarded.
- **Input gate (i_t)-** regulates how much new data are added to the cell state.
- **Output gate (o_t)-** controls how much of the cell state is exposed as the hidden state.

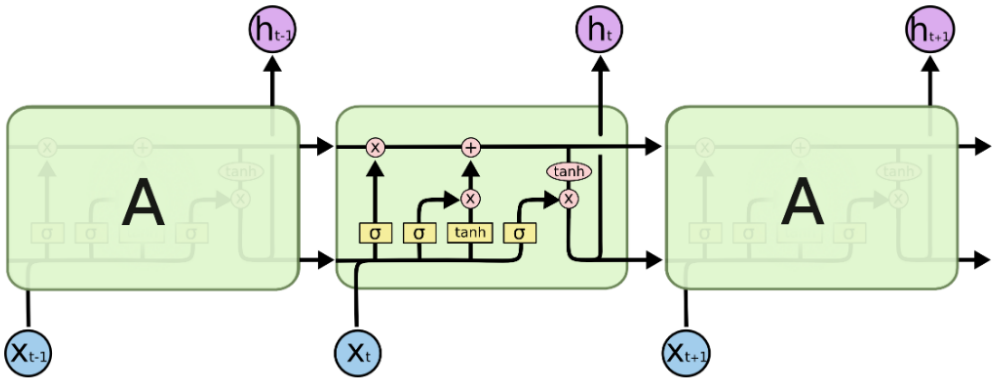
The gates are computed through sigmoid activations, which produce values between 0 and 1 acting as a filter for information flow; 1 meaning passing and 0 no passage.

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

where x is the input to the function and $e \approx 2.718$ is Euler's number, the base of the natural logarithm. The temporary cell state candidate $\tilde{C}_t$ is built using 'tahn' activation, producing values between -1 and 1, enabling both positive and negative updates (activations) to memory.

$$\tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$$

All these components work per element in the data to update the internal memory and hidden state at the same time. Thus, in comparison to traditional RNNs, LSTMs are complex due to the additional gates and states, yet are much more efficient when learning long-term dependencies (Dehghani et al., 2023b; Ndikumana et al., 2018; Olah, 2015). Figure 2.7 shows the structure of this system:



Source: Olah (2015)

Figure 2.7. Representation of the structure of the LSTM

ConvLSTM networks extend the traditional approach of LSTMs by integrating convolutional operations to their architecture, resulting in very effective models for spatio-temporal data, for instance, satellite imagery. Contrary to LSTM models which rely on fully connected layers, this improved version replaces matrix multiplications with convolutions in both the input-to-state and state-to-state transitions (Xavier, 2019). This new adaptation makes the model preserve, in addition to temporal dependencies and dynamics, spatial correlations like CNNs. Nevertheless, they still preserve a similar LSTM structure: input, forget and output gates, as well as the memory cell, however it uses convolutional layers instead of dense layers to carry out gate computations. Thus, this type of network becomes really powerful when dealing with grid-like data structures over time (Yin et al., 2019). In summary, it manages to analyse spatial features in satellite data while, at the same time, capture the evolution throughout the time steps of those features. Furthermore, if multiple ConvLSTM layers are stacked, deep architectures can be built, acting as encoders and decoders for spatio-temporal sequences (Xavier, 2019). They can learn complex temporal dependencies and, simultaneously, leverage local spatial patterns, becoming increasingly useful in tasks like remote sensing (Dehghani et al., 2023b; Xavier, 2019).

For classification purposes that involve the use of temporal data, LSTM units can be combined with fully connected output layers such as Softmax (Ndikumana et al., 2018). It maps the hidden states to class probabilities. In addition, according to the study, it is usually preferred over sigmoid functions in multi-class classification since it normalizes the outputs into probability distributions. The constant updating of LSTM models allow to make temporally informed decisions based on these probabilities and retain information along long sequences.

In MathWorks (2023), a hybrid CNN-LSTM approach to ConvLSTM is proposed. In this case, CNN layers are employed to extract the spatial features while the LSTM layers are used to handle the temporal modelling. It becomes different to ConvLSTM since convolutions are not applied within the recurrent structure itself. These hybrid models are described in the paper as particularly useful when spatial features can be obtained before applying any type of temporal modelling (MathWorks, 2023).

2.6.4. U-Net

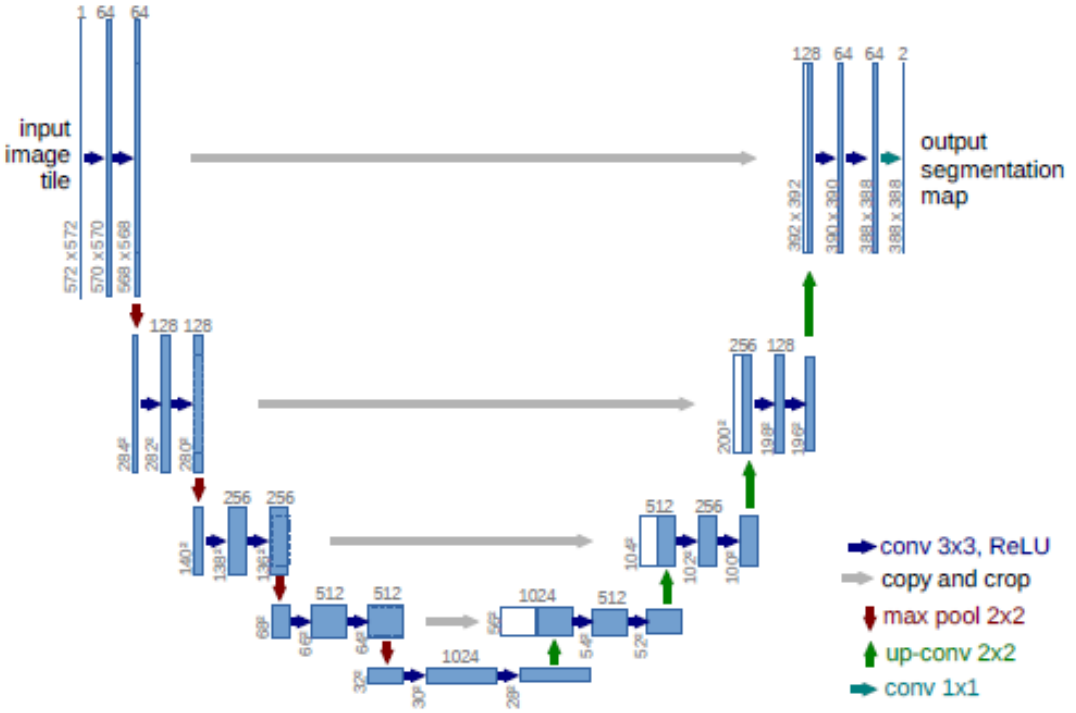
U-Net is a CNN architecture with the objective of producing image segmentation, assigning a class label to each pixel rather than classifying the entire image as a whole (Ronneberger et al., 2015). It is this fine-grained segmentation that makes U-Nets unique among other traditional neural networks. In earlier stages of development, pixel-level predictions employed sliding-windows, classifying the pixel by analysing its surrounding patches. However, it increased the amount of training data and

suffered from two major drawbacks: computational inefficiency from redundant overlapping patches and trade-offs between patch size and spatial precision (Ronneberger et al., 2015).

To overcome such issues, the U-Net introduced a more elegant and efficient design, effective when limited labelled data are available. It performs good segmentation with modest training times as stated in Ronneberger et al. (2015). The paper describes the core idea, in which a contracting path—that captures context through patches—and an expansive path—that enables precise localization— are combined. Hence, they become symmetric and produce the characteristic U-shaped architecture. It is the expansive path that allows the high-resolution features to be reused from the contracting path, allowing for both contextual awareness and spatial precision (Ronneberger et al., 2015).

In Ronneberger et al. (2015) the architecture's structure is described in detail. This structure is better visualised in Figure 2.8. It consists of two main parts:

- **Encoder -** a contracting path. It follows the traditional CNN structure of applying kernels during convolutions (3x3 with ReLU activation) as well as max pooling operations (2x2) for downsampling. At each step, the number of feature channels (filters) is doubled, increasing the network's capacity to detect high-level features (Ronneberger et al., 2015). The last layer is known as the bottleneck, and marks the end process of the encoder path.
- **Decoder -** an expansive path. It begins at the bottleneck layer, with upsampling of the feature maps. Then, 2x2 "up-convolutions" are carried out, halving the number of filters and concatenating them with the corresponding feature map from the encoder. This strategy is enabled due to skip connections, which link previous layers (before downsampling) with later layers (after upsampling). Afterwards, two 3x3 convolutions with ReLU activations are performed and, at the final layer, a 1x1 convolution maps the feature vectors to the required number of output classes (Ronneberger et al., 2015).



Source: Ronneberger et al. (2015)

Figure 2.8. Architecture of the U-Net model, where each blue box is a multi-channel feature map (the number of channels is denoted on top of the box and the x-y-size at the lower left edge) and the white boxes are copied feature maps. The arrows denote the different operations

To manage the border effects from unpadded convolutions, the segmentation map produced is based only on pixels which have a full context within the input image. Any missing context near those borders is extrapolated by mirroring the image (Ronneberger et al., 2015).

The work of Shibuya and Hotta (2022) has further developed and extended the architecture of this model. To begin with, it now replaces standard convolutions with ConvLSTM layers, capturing spatio-temporal dependencies, important for time sequences in inputs when remote sensing is involved. Furthermore, another key innovation produced is the addition of a feedback mechanism, where the output probability maps from the initial segmentation pass are passed again to the network together with the original image for refinement purposes a second time. During this second round, the ConvLSTM units use the previously stored features (cell state) to guide the process, in order to extract more accurate features. By involving the gating mechanisms that characterize this type of RNN—input, output and forget gates—the model manages to select and update only the most relevant and informative features at each round, yielding segmentation results with further refinement (Shibuya & Hotta, 2022). These improvements enhance the continuity of features whilst spatial patterns from earlier passes are effectively learnt.

3. Methodology

3.1. Rock and soil classification

With the abundance and complexity of classification systems described previously in section 2.1, a decision had to be made on the classification method that would be employed during the thesis. The final decision was to use twelve different classes, six for soils and six for rocks.

The decision to adopt six distinct types of rocks is based on the general geological classification system. As mentioned in section 2.1, they are typically categorized three main types: sedimentary, igneous, and metamorphic. Relying solely on them can be overly simplistic if the main purpose is accurate classification. Each main category includes a wide variety of rocks which can be further subdivided. This is normally done into two sub-classes, achieving a more balanced and meaningful scheme which preserves both the geological relevance of the classification process and the computational efficiency of the models. Nevertheless, inside these sub-classes there is still a considerable degree of variability, which is one of the reasons why so many classification methods exist within each category. For instance, distinguishing between rocks of the same class like marble and quartzite can be challenging for some models based solely on satellite imagery: hence the decision to limit the number of rocks to six classes, where a balance between providing meaningful classification and keeping a level of abstraction that the models can handle. Below is a list of the chosen rock types and their descriptions (provided by U.S. Geological Survey (2025a, 2025b, 2025c)):

- **Detritic (Sedimentary) -** made up of pieces of pre-existing rocks that have been weathered, loosened, transported and sedimented at a basin or depression. When buried deeply they become compacted and cemented. Examples include sandstones and shales.
- **Carbonate (Sedimentary) -** formed from the precipitation of carbonate minerals—mainly calcium carbonate and dolomite—due to water or skeletal remains. Limestone is the most common one.
- **Volcanic (Igneous) -** also known as extrusive rocks, they form when magma is able to exit—due to fissures or volcanoes—to the surface and cool down rapidly. Fast cooling means the minerals do not have much time to grow, producing fine-grained or glassy textures. Common examples include basalt or obsidian.
- **Plutonic (Igneous) -** also called intrusive, they are formed when magma is trapped beneath the surface, cooling down slowly until they solidify, allowing for minerals to grow to larger sizes, resulting in a coarse-grained texture. A common example is granite.
- **Foliated (Metamorphic) -** some kinds of rocks become banded or foliated, meaning its minerals are aligned, arranged in parallel. This happens due to the pressure and temperature rocks go through, which flattens and elongates their minerals, forming a sheet-like structure. Examples include slate or gneiss.
- **Non-Foliated (Metamorphic) -** as their name suggests, they do not contain bands and their formation comes from a wide range of sources—such as intrusion of a rock inside another—changing the composition of the mineral due to the high temperatures. Some of the most known examples are marble and quartzite.

In order to keep a reasonable amount of soil types included in the project, the decision was to use a simpler version of the USCS classification method, mentioned in section 2.1.2. Therefore, another six different types of soils were chosen, matching the number of rock classes and keeping a balance between both categories. The soil types chosen for this project are the following and their descriptions are acquired from U.S. Department of Agriculture, Natural Resources Conservation Service, National Soil Survey Center (2012):

- **Sand -** a coarse-grained soil of varying sizes of particles that drains water quickly and barely retains nutrients.
- **Silt -** more fine-grained with particles of smaller size than sand, making it better at retaining water and draining slower.
- **Chalk -** alkaline soil made of calcium carbonate making it poor in nutrients.
- **Clay -** very fine particles which makes it great at retaining water and nutrients, which also results in a compact and poorly drained soil.
- **Peat -** organic soil obtained from waterlogged conditions, becoming acidic and having less nutrients.
- **Loam -** a mixture between sand, silt and clay, making it fertile, well-draining and ideal for vegetation.

3.2. Tools and datasets used

The following section outlines the tools and datasets that have been employed to fulfil all the objectives set for this project.

The programming language selected for the implementation of the thesis is Python, developed in the *PyCharm Community* environment, due to the extensive community of libraries that can support data analysis and efficient computation. Key libraries used include *Rasterio* and *Scikit-Learn* to handle geospatial data and machine-learning tasks, *Numpy* for basic numerical operations, *MatPlotlib* for visualisation purposes, and both *TensorFlow* and *Keras* to develop the DL models.

Thermal data utilized originates from NASA's ECO mission, accessed through the open-access portal Application for Extracting and Exploring Analysis Ready Samples (AppEARS). Although other sources, for instance Landsat 8/9, also provide thermal imagery via the Earthdata Search portal, ECO was selected for its high revisit frequency (1-4 days versus 8-16 for Landsat), greater thermal resolution and its capability to provide diurnal coverage. ECO consists of a dedicated NASA mission designed to measure thermal properties from surface materials. Thus, it provides both LST and emissivity information at a spatial resolution of 70 metres (Laboratory, 2020). In addition, its freely available for use and one of the most recent National Aeronautics and Space Administration (NASA) missions, making it a newly form of thermal imagery to explore with. For these reasons, ECO was chosen as the thermal data type. LST tends to be commonly used in thermal imagery, however, it underperforms in some cases as it might not be a valuable asset for characterizing rocks and soils, since it is based only on the atmospheric temperature. Emissivity, being more specific to a material's physical characteristics, provides greater potential for discriminative tasks. Furthermore, complementary datasets were included, to assess if the general performance of the DL models could be enhanced by integrating a combination of multiple data types.

SAR data were obtained from the European Space Agency (ESA)'s Sentinel-1 by using a Google Engine API in Python. It operates in the C-band with microwave radar signals, making it unaffected to cloud coverage or lighting conditions, providing consistent diurnal imaging. Due to its sensitivity to surface's structural characteristics and moisture, it can become a strong complementary modality with the potential to improve the models in classifying soils and rocks.

Integrating another data type aims to increase the model's accuracy, robustness and general performance. While the primary focus remains on thermal imagery, the potential benefits of multi-sensor patches are explored during the study.

3.3. Ground-truth labels

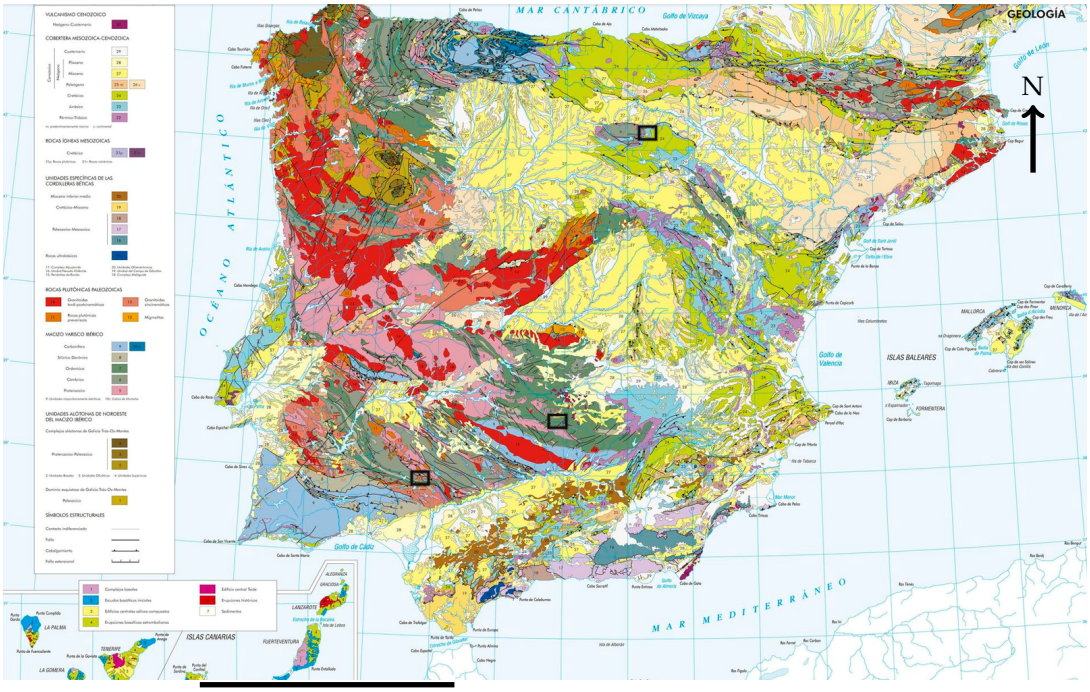
Labels play a key part in achieving an accurate classification model. Also referred to as ground truths, labels are used by the DL models as reference or correct outputs for each training sample during supervised learning (as defined by IBM Cloud Education (n.d.)). The model attempts to learn to predict new inputs using them as reference by iteratively improving its ability to predict unknown inputs. They are essential for evaluating the performance of the model, helping in computing the loss function that guides the optimization process. In this case, if a pixel contains sand as a class, the labels might be "Sand", and during the training phase, it will try to minimize the difference between the predictions it makes and the known label. Notwithstanding, the reliability of a model's predictions is dependent on the quality of the ground truth data. Inaccurate or noisy labels introduce artifacts and erroneous data from which the model learns from, resulting in distorted predictions. During this thesis, geological maps will be the ground truth that will be employed as labels for the models, with the intention of allowing models to associate thermal and auxiliary data patterns with specific geological classes.

3.3.1. Geological maps

The use of geological maps as labels stems from two different factors. Firstly, geology takes thousands or millions of years to change. Therefore, a 20 year difference in geological time is insignificant, ensuring that any geology map used within the last decades will not have suffered any major changes since it was generated. Secondly, they contain the area distribution of all the rock and soil types selected, much like land cover in urbanisation, meaning the models can perform classification using a similar procedure. However, it should be considered that geological maps are not perfect ground truths, and are instead a representation of ideas and interpolation established by geologists backed by detailed, careful observations and rationalized theory (Marjoribanks, 2010).

The geology maps used for this project are taken from the Instituto Geológico y Minero de España (IGME), the Spanish Geological Survey, whose main focus is on research on Earth Sciences. Their website provides access to geology maps for the whole of Spain, covering all the territory with maps at a scale of 1:50.000 or 1:200.000. To obtain better ground truths, 1:50.000 was selected for the labels. Each geology map defines the areas of interest used and mentioned in the previous section, however, the files provided by the Institute are not formatted labels for DL use: hence labels had to be manually generated for this project. This process hinders part of the scalability, reproducibility and speed of the workflow involved in this project. Generating the labels takes some time, however, if this workflow is enhanced in the future, the process might eventually not need the use of ground truth labels in other regions or countries once the models have learned enough geological context or continuous learning is possible. Thus, resulting in a much faster approach and a support tool that can be employed before any necessary field work.

Carrying out the project's experiments in Spain was based on the range of advantages this country offers in many areas for remote sensing applications involving thermal data. Firstly, the climatic advantages of a country with a high number of clear-sky days, distinct seasonal variations and diurnal temporal changes. Furthermore, the presence of bare soil and vegetation which provides potential case studies for this project. In addition, Spain presents a rich geological distribution of most rock and soil types present throughout the country (as seen in Figure 3.1), as well as enough data available of the geological maps due to the IGME. All the scale bars and Northern arrows used in figures containing two images refer to both of them, both in size and orientation, respectively.



Source: Instituto Geográfico Nacional (IGN) (2020)

Figure 3.1. Map of Spain showing the whole geology of the country and the selected regions of interest, marked with black boxes. Villoslada is located in the north, Santa Olalla in the southwest, and Puertollano in the east.

3.3.2. Digitalization

Due to their format in PDF, the maps had to be digitized and rasterized. Nevertheless, they provide users with georeferenced images. Thus, in QGIS, the original images where digitized by producing polygons. Each polygon corresponds to a lithology present in the original map. To each polygon, a class number was assigned, corresponding to the category they belong to. This resulted in 14 classes as follows and the map in Figure 3.2:

- 0 - Not a number (NaN)s
- 1 - Sand
- 2 - Clay
- 3 - Chalk
- 4 - Silt
- 5 - Peat
- 6 - Loam
- 7 - Detritic
- 8 - Carbonate
- 9 - Volcanic
- 10 - Plutonic
- 11 - Foliated
- 12 - Non-Foliated
- 13 - Water

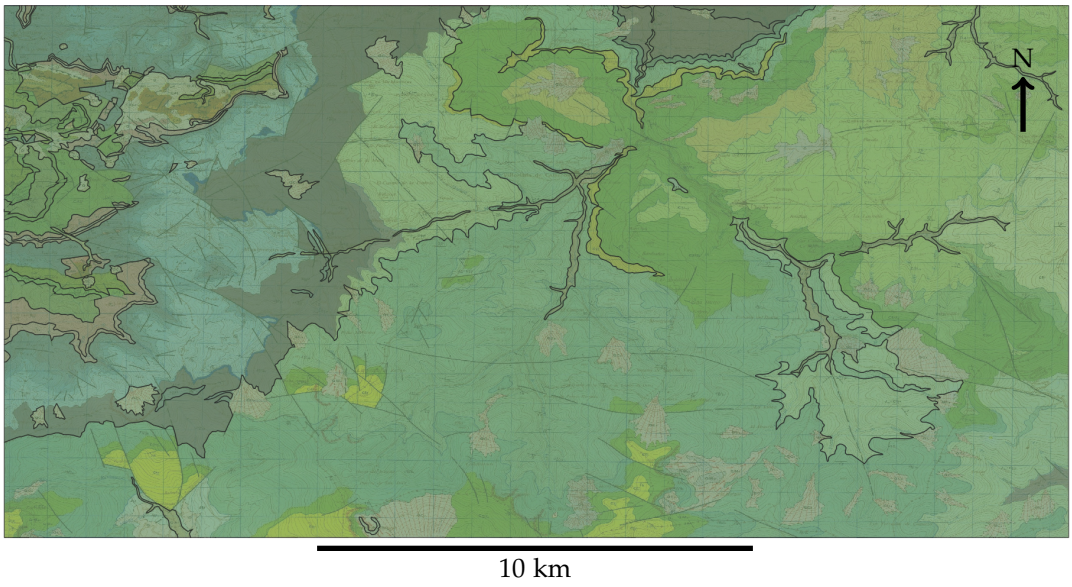


Figure 3.2. Digital lithology map example, in this case, of Villoslada de Cameros. Each polygon corresponds to a class, with some polygons containing multiple sub-classes present in the original geology map, since they belong to the same class as set up in this study.

The decision to include two more classes—water and NaN values—was based on wanting the model to perform better. If the model is able to recognize water areas and NaN values present in the labels, due to mismatches or missing data, the model could become more robust to outlier cases. In addition, water was also included for a second reason: even though the presence of water in all the areas is minimal, geologists cannot assume what lies under the water, since it cannot be seen. Therefore, since there is no intention to discover what lies beneath it in this project, it is better to leave it unknown and classify it as a class of its own.

3.3.3. Rasterization

The resulting shape file containing all the polygons had to be rasterized to '.tif' format, matching the thermal images. Unfortunately, just like with SAR, the geological maps used from IGME in Spain have a higher resolution (25m) than the thermal images from ECO (70m). The resulting raster had to be resized to a smaller shape, to match that of the thermal images. This way, during the patching process pixels would always have a corresponding label assigned, avoiding potential issues with missing data. This resizing was done in QGIS with the 'Warp(Reproject)' option. The result is Figure 3.3

Later on, the labels were included into the pre-processing stage and patched alongside the thermal images, producing the corresponding training, test and validating label sets for both models.

3.4. Pre-processing stage

Raw satellite data contain valuable information that have a wide range of uses, however feeding raw data into DL models for them to learn from will not yield positive results. It contains many different unwanted pixel values—like noise, distortions, artifacts or irrelevant background information—that hinder the performance of these models. Given CNN and ConvLSTM models employ pattern recognition during training to learn form input data, feeding them with invalid information like NaN values results in chaotic outputs. To avoid this, it is necessary to carry out a pre-processing phase, were data are filtered and prepared for the models to be able to learn from good quality pixels. Furthermore, pre-processing is not only needed to enhance signal quality or reduce redundancy, but is also essential to handle, in the case of this thesis, images from a variety of sources with varying Coordinate Reference

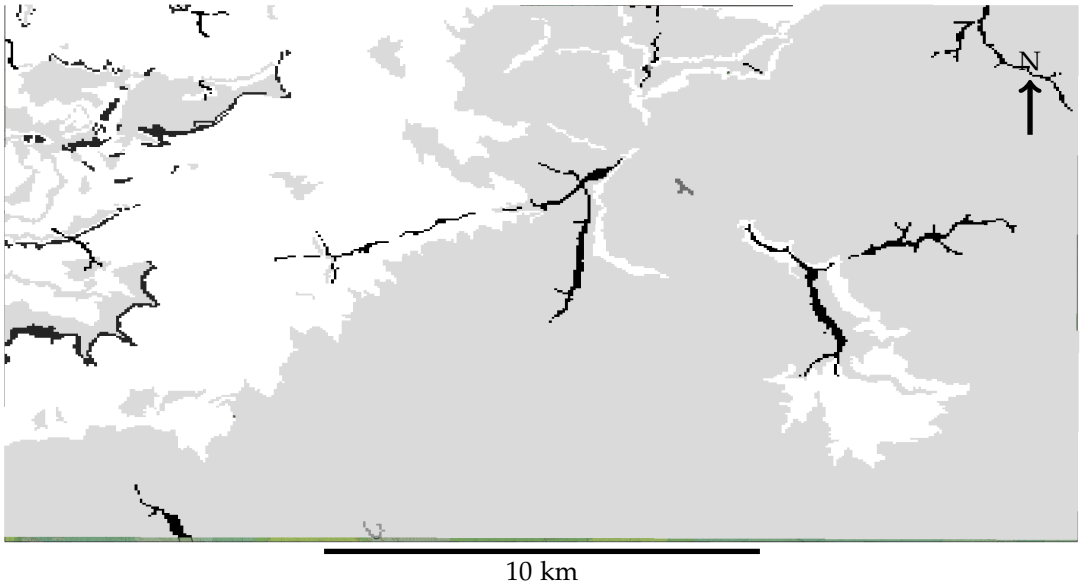


Figure 3.3. Rasterized example version of the lithologies map, in this case, of Villoslada de Cameros. The quality is lower since it is already matching the thermal image shape resolution.

Systems and sizes. The whole pre-processing workflow for the previous and following sections can be easily understood and visualised in Figure 3.4:

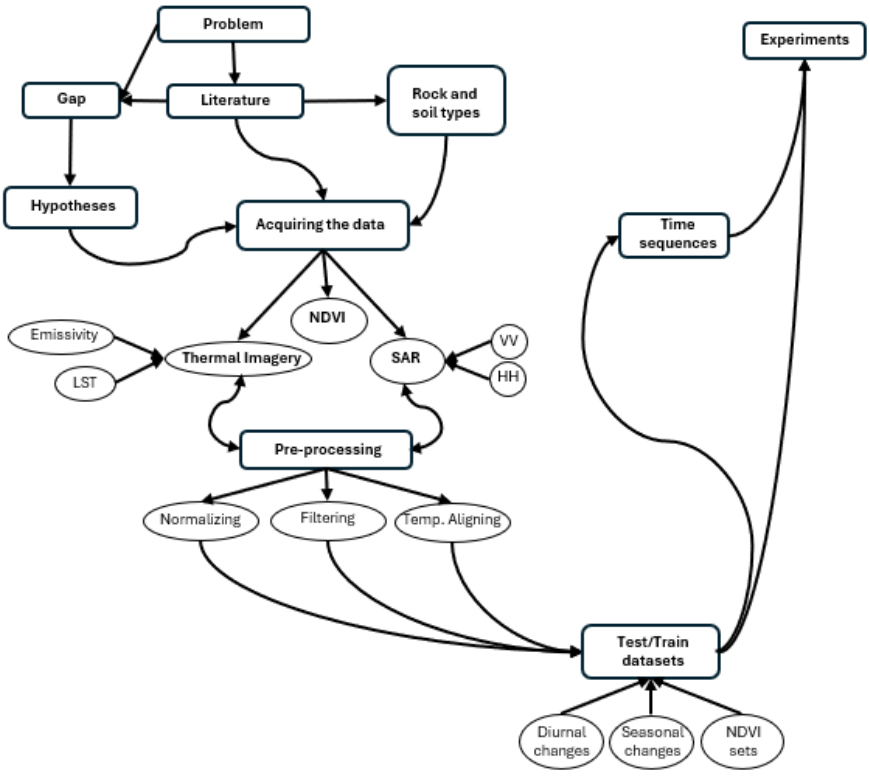


Figure 3.4. Pre-processing workflow

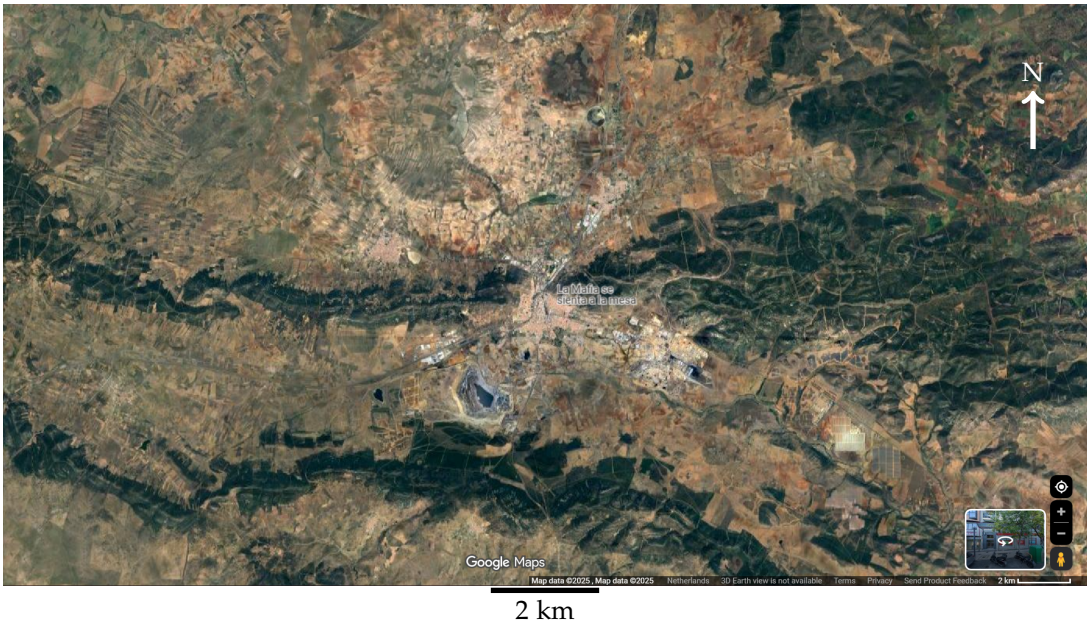
3.4.1. AOIs

The first consideration is the Area of Interest (AOI), which all the inputs and all the labels must match. For this study, up to three different areas have been selected, all of them belonging to the region of Spain. The decision behind carrying out the experimentations on Spain is because of the low general presence of cloud cover throughout the year, being lower than in many countries in Central and Northern Europe. The other reason for the AOIs to be in Spain is the rich geology, since the Iberian peninsula contains the three main rock categories. In addition, there is an abundance of geology maps for each area selected. Thus, the three regions that have been selected are:

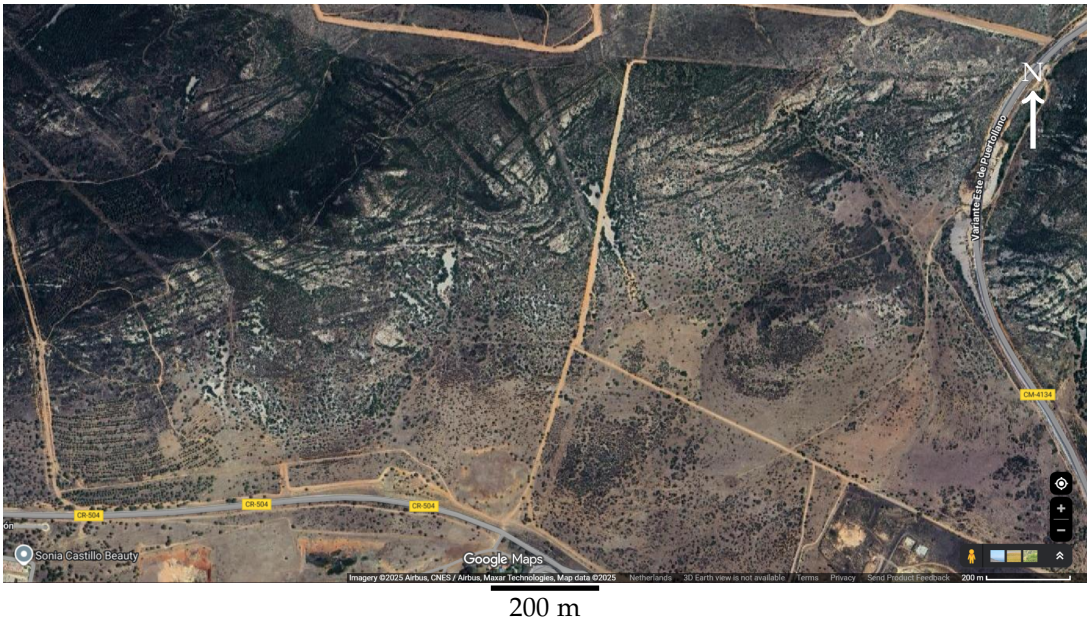
- Puertollano - NE: 38.83605,-3.85415 // SW: 38.66787,-4.18496
- Santa Olalla del Cala - NE: 37.999978411, -6.188309907 // SW: 37.832876641, -6.520675667
- Villoslada de Cameros - NE: 42.168736, -2.518803 // SW: 42.002377, -2.852681

All three areas contain a variety of rocks and soils, so all 14 classes are included with only these three regions, distributed along the centre, Southeast and Northern parts of the country. Their location in the region of Spain can be observed in Figure 3.1, representing the whole geological distribution of the Iberian peninsula. Furthermore, their respective geology maps can be observed in Figures B.1, B.2, and B.3. Puertollano, located in the central part of the peninsula, corresponds to a dry area with sparse vegetation due to its proximity to Calatrava, historically a volcanic field (Figure 3.5). This part of Spain presents a moderate amount of rains, as well as moderate to very hot temperatures throughout the year (Climate, 2025). Santa Olalla is closer to the coast in the South-West of Spain, containing abundance of spaced trees, with moderate visibility of bare soil (Figure 3.6). It also corresponds to an area or moderate to high temperatures with slightly more rainfall than Puertollano in occasions (WeatherSpark, 2025). The last area, Villoslada de Cameros is located in the North of Spain, an area known for milder weather conditions and a higher abundance of rain (Data, 2025), which results in a very vegetated area as observed in Figure 3.7.

The coordinates of the AOIs are obtained from their respective geology maps, provided by IGME in their website. Nevertheless, the final shape used for all input images and labels corresponds to the resolution of the ECO thermal images, since they had lower resolution than the geological images. Therefore, geological maps were downsampled—performed in QGIS—so that the coordinates remain unchanged but the resolution matches that of the thermal data. This is important, as projection plays a crucial role in matching an input pixel to its corresponding label. If there is no match, the program cannot perform the correct learning process. SAR data were also adapted to the thermal shape.

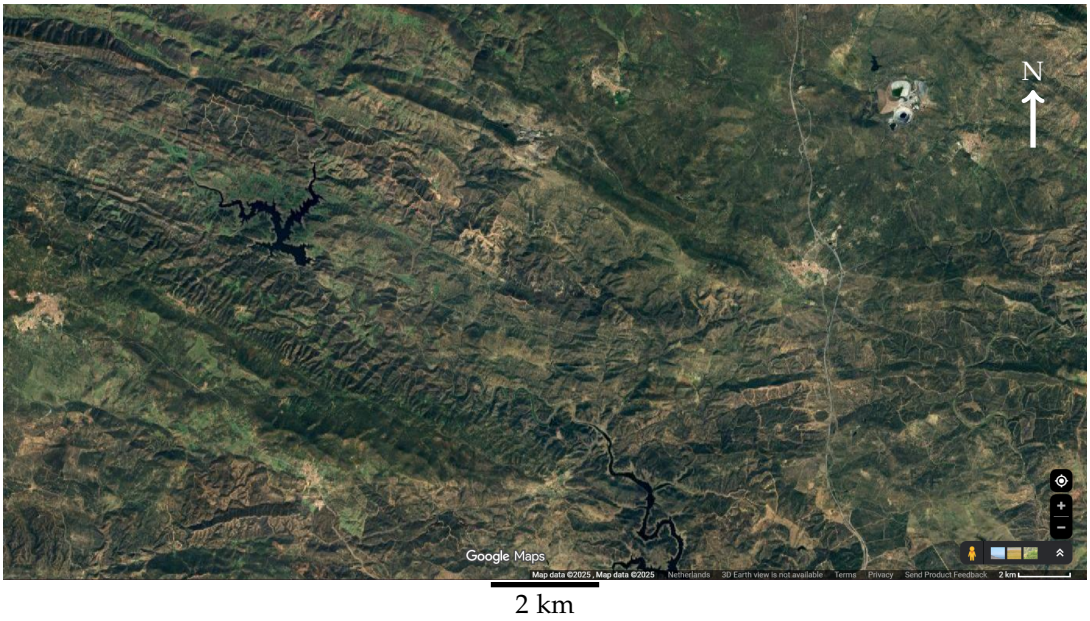


(a) Image displaying the area of Puertollano with a similar area coverage to that of the geology map used as ground truth.

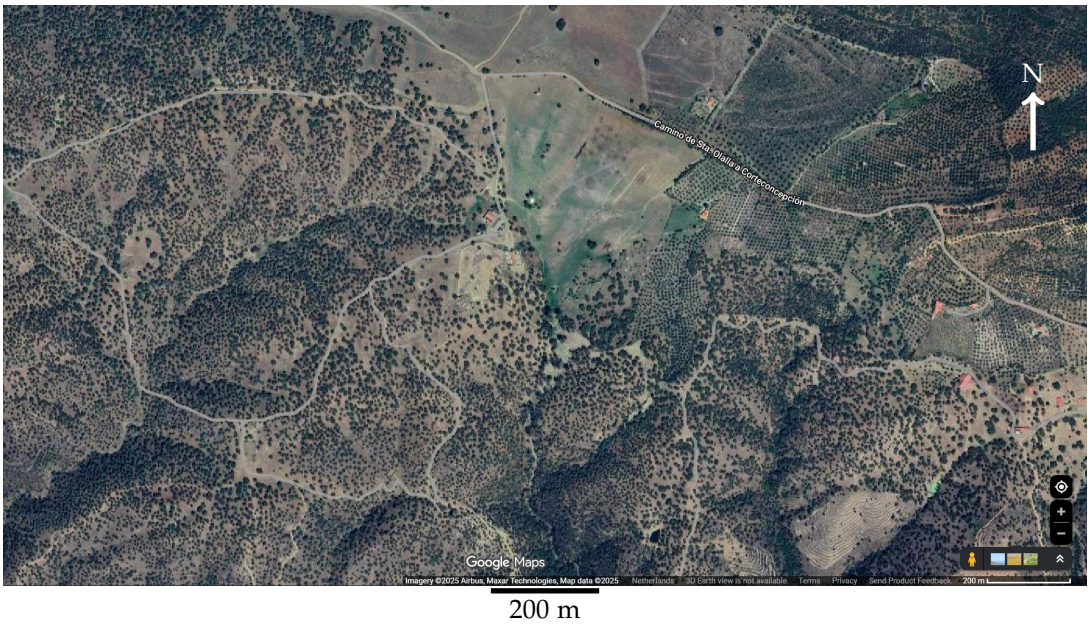


(b) Zoomed-in view of the Puertollano area for vegetation visualisation.

Figure 3.5. Optical images at different views of the area of interest of Puertollano.

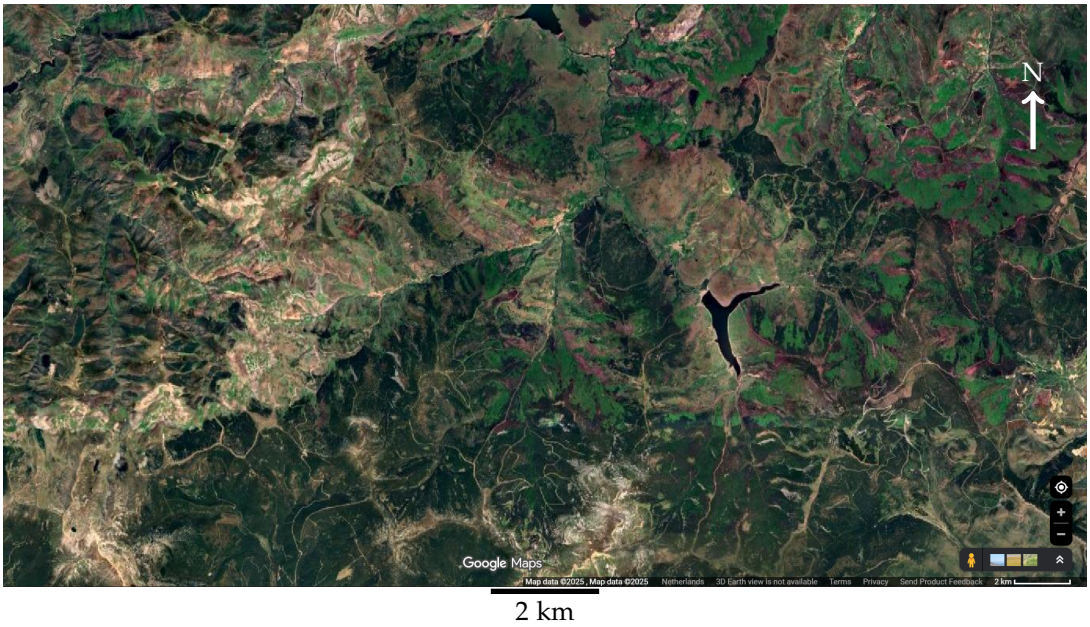


(a) Image displaying the area of Santa Olalla del Cala with a similar area coverage to that of the geology map used as ground truth.

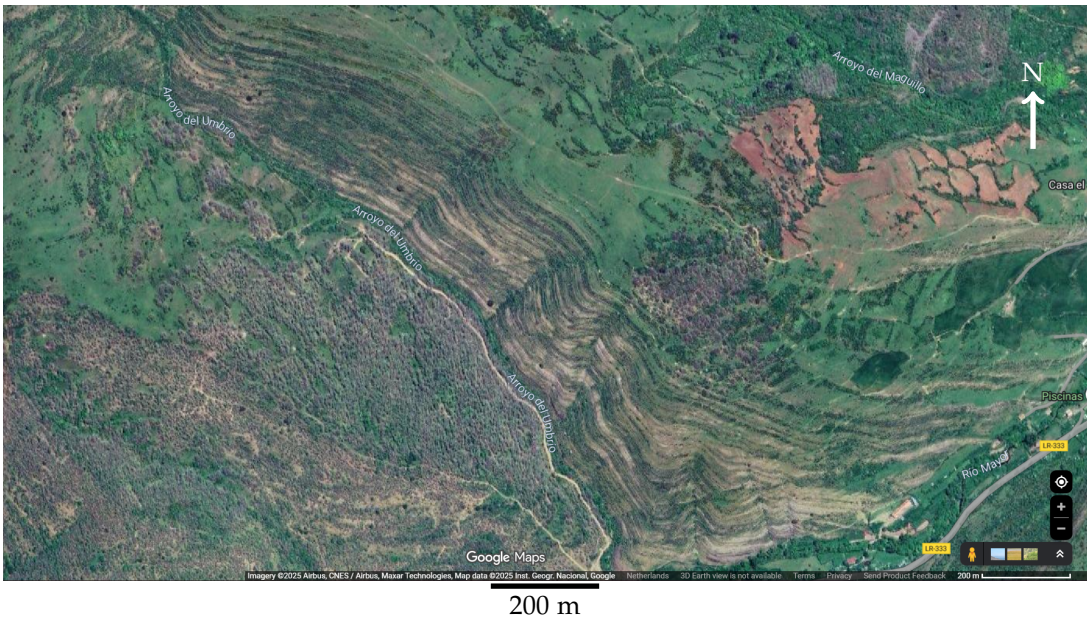


(b) Zoomed-in view of the Santa Olalla del Cala area for vegetation visualisation.

Figure 3.6. Optical images at different views of the area of interest of Santa Olalla del Cala.



(a) Image displaying the area of Villoslada de Cameros with a similar area coverage to that of the geology map used as ground truth.



(b) Zoomed-in view of the Villoslada de Cameros area for vegetation visualisation.

Figure 3.7. Optical images at different views of the area of interest of Villoslada de Cameros.

3.4.2. Pixel quality check

The presence of artifacts—such as geometric distortions due to satellite motion or radiometric errors from transmission problems—generate unwanted and invalid data. With all the images and patches at the same shape and projection, the next step involves quality assessment.

For this project, the decision was to filter out images containing either more than 10-15% of bad quality pixel data or presenting more than 10% of cloudy pixels. A threshold of around 10% was allowed since part of the aim of the models is to make them robust to varying data, and perfect quality images might not always be available. However, images presenting values higher than the threshold for cloud coverage or artifacts should not be processed or included, since the higher the amount of invalid pixels in the inputs, the worse the performance of the model. Although this is not the case for SAR data or the labels, since they are unaffected by cloud coverage, thermal data struggles with cloud coverage, altering its values and yielding problematic classifications. Hence, the 'QC' file from ECOSTRESS—a quality assessment file—was used to filter invalid pixels. By using the bit number to determine which pixels contain valid data, these files can help filter out the necessary pixels and enhance the quality of what is used as input data.

3.4.3. Normalization and NaNs

Once the quality of the images has been determined, the images containing less than 10% of invalid pixel information still have to be further processed. The presence of invalid pixel information is normally given in NaN values. They correspond to unrepresented or undefined numerical values. NaNs can also alter the learning process of DL models, since they can learn from unwanted patterns. Therefore, they also need to be substituted by other values. As observed in section 3.3, NaNs were replaced for 0, which models interpret as values not to learn from, since they have no value.

With no remaining invalid pixels or values, the rest of the pixels are normalized for all input types. Normalization in this project was necessary for several reasons. In the first place, given the use of several data types—ECOSTRESS and SAR—, it is important to keep all values in the same range, since Kelvin, reflectance and decibels all produce vastly different numerical ranges. This can produce poor generalization and the models can tend to favour larger values while learning. In the second place, the idea is to avoid possible numerical instabilities that can affect the learning process, providing the models with high-quality datasets, crucial for accurate classification (Bai et al., 2021). Lastly, NNs learn better when input ranges are lower and centred around 0.

3.4.4. Patch Positions

As an additional approach to the single patch visualisation process, ordered patch allocation was maintained to facilitate spatial and temporal reconstruction of geological maps for each date. To do so, during the pre-processing stage, while performing a sliding window approach to save the thermal images, the position of the patches is stored as metadata. This way, once the models fulfil their predictions, these initially stored positions could be employed to fully reconstruct geological maps to their original size, matching the new predictions with the original position of their respective ground-truth label. The result is a full geological map prediction for each date that covers the whole area of interest from which the model trains and tests, which can be observed in Section 5.7.1. This approach intends to simplify the comparison between the ground truth and predicted outputs, allowing for clearer visualisation and interpretation of the spatial patterns observed. Notwithstanding, this part of the methodology was only tested in the thermal-only data experiments, and not included into the diurnal, seasonal or SAR experimentations.

3.4.5. Splitting

When the final images are generated and all values have been normalized, it is important to prepare the datasets for DL. As usually performed in different studies, data are then split into training, validating and testing sets. Each set performs a key role in producing accurate predictions. The data percentages

assigned to training and testing are 85% and 15%, respectively. Inside the training set, 15% of the dataset was assigned to a validating set, used to test the model on unseen images, to analyse if the model is memorizing the input data or actually learning properly.

When performing this operation, it is common practice to randomize or shuffle the distribution of patches, to ensure statistical fairness and avoid bias. For this project, both random and ordered distributions were tested. The conclusion drawn in this case was that, for the sake of consistency and simplicity for map reconstruction, the data would not be shuffled, and training and testing datasets would be made of chronologically ordered patches. This decision was primarily motivated by the nature of the datasets and the models employed. Maintaining chronological order ensures that the temporal integrity of datasets is kept for the ConvLSTM model, which learns from those temporal dependencies from consecutive observations. However, it was still tested and shuffling the data displayed no advantages when compared to non-randomized datasets which led to the decision to keep chronologically ordered patches. Furthermore, for the remaining model, the CNN, the order of patches has no impact on it, since it processes images independently, which has no effect on its performance.

3.4.6. Temporal sequences

The input data for CNN architecture and ConvLSTM architecture are not the same. Even though they both share the U-Net architecture in this study, they cannot share the same input. Since ConvLSTM handles temporal dependencies, it requires an additional dimension called 'temporal sequence'. These temporal sequences are made of temporally continuous frames—which are the images containing all the input data channels—that may or may not overlap. It is this dimension that allows the model to learn the temporal dependencies, since, for the same area, it can detect how the thermal values change inside the sequence. To keep this temporal learning valuable, the dates should be consecutive and there should not be a great temporal variation between the first and last frame of the sequence. The temporal sequence was set to include 5 frames per sequence and have an overlap of 2. That means each new sequence starts 3 frames later than the previous one, ensuring temporal continuity between sequences and enough temporal variation inside the sequence. The process was carried after extracting the patches from thermal, SAR and NDVI images across multiple dates, resulting in chronologically ordered and stored patches. Once they are split into training and testing without shuffling, to maintain temporal consistency, the temporal sequences are created by sliding a "window" where each sequence contains 5 consecutive patches with an overlap of 2 frames (stride 3) between them—patches at positions [0,1,2,3,4] become sequence 0, patches at positions [3,4,5,6,7] become sequence 1, and so on. For the sake of simplicity, the target mask for each sequence is taken from the last frame, since the geological map does not change significantly in that time frame. For thermal and SAR data, the independent values of all 5 dates for the same pixel are stored, and instead, for NDVI the aggregated value of the five dates is stored, since vegetation does not suffer significant changes in such a short time span.

3.4.7. SAR integration

Similarly to thermal data, SAR data had to go through a pre-processing stage. Although there was no need to account for cloudy pixels, the presence of NaN values is abundant in this kind of imagery. Thus, NaN removal was necessary and, later on, normalization of values. SAR is added as two additional channels that form part of the input data, corresponding to the two most commonly used polarizations: VV (Vertical Transmit-Vertical Receive) and VH (Vertical Transmit-Horizontal Receive). VV consists of microwaves oriented in the vertical place for both the transmission and reception signal (European Space Agency, 2013). Instead, VH has microwaves oriented in the vertical for the transmitting signal but oriented horizontally for the backscattered energy the antenna receives back. Using both can improve the detection of surface features through soil moisture or surface structure.

3.4.8. NDVI

The integration of NDVI is intended for evaluating the performance of the models under different vegetated areas. Since the three regions contain varied distribution of vegetation and it has a noticeable

impact on thermal values, it is important to analyse to what extent its presence can alter the predicted results. NDVI imagery was derived from COPERNICUS' Sentinel-2 satellite mission—due to its 10 metre per pixel resolution—, specifically from the Surface Reflectance product from Google Earth Engine (GEE) with the appropriate GEE function "normalizedDifference(['B8', 'B4'])"—which represents NDVI = NDVI = $\frac{NIR-RED}{NIR+RED}$—, following a similar structure to Lasaponara et al. (2022). The same study period as the thermal images was used (2019-2024) for which each image in that time range was already atmospherically corrected and some subsequent cloud filtering—no more than 10% of cloud coverage—and masking, using the QA60 band, was applied. The NDVI values were ensured to fall between [-1, 1], with invalid pixels (NaNs, infinite values, and out-of-range values) masked to -1. NDVI images have not been normalized since they are not part of the input data for the models and are merely used as auxiliary data for comparison purposes. The images were resampled and temporally aligned to match ECOSTRESS thermal data, with a 16-day matching window. A two-week time range for temporal alignment allows for a higher rate of matching thermal and NDVI images. It is also a suitable temporal window since vegetation does not suffer drastic changes in such wide areas within shorter periods of time and the model does not need to learn from the data. Instead, it serves an analytical purpose: to understand how the models deal with vegetated areas. An example of an NDVI image for the area of Villoslada de Cameros is included in Figure 3.8:

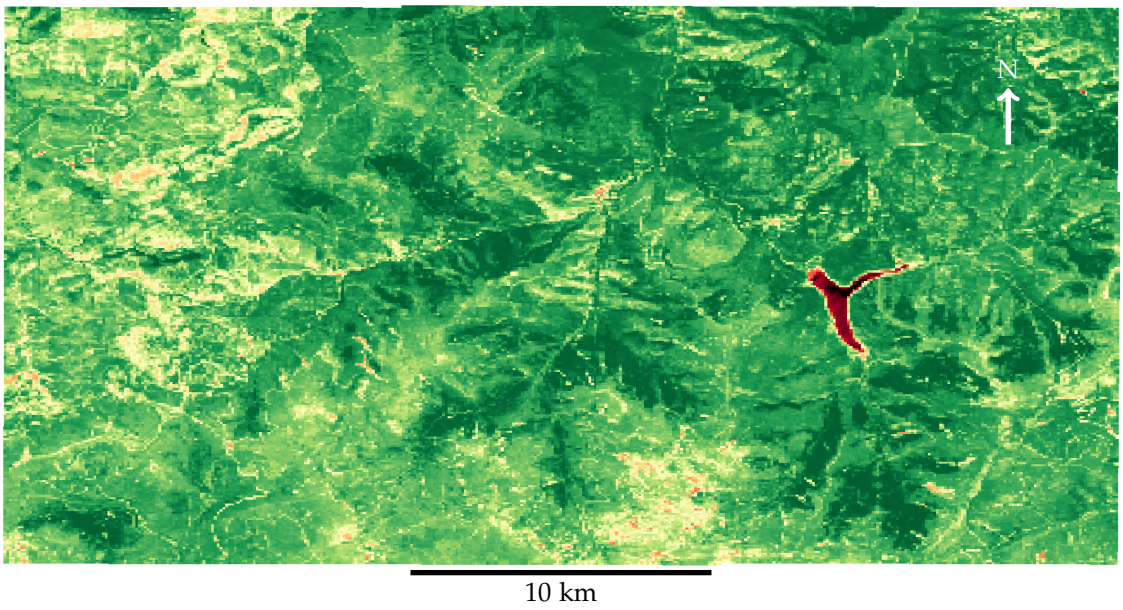


Figure 3.8. NDVI image of the Villoslada area with the bright green colours representing the most vegetated areas, the darker green pixels representing bare soil, and the black pixels the water.

During the pre-processing stage, both thermal and NDVI patches were generated using identical spatial coordinates and window sizes for proper correspondence for pixel analysis. Later on, for temporal sequence integration, the NDVI values within each sequence were aggregated, producing a single value representative of the whole valid frames. This ensures temporal information is preserved while reducing the complexity for the correlation analysis. This analysis is carried out by comparing the softmax probability map produced by the output of the model against the corresponding NDVI image, employing the Pearson correlation coefficients in the process for an enhanced analytical process. The relationship of confidence maps and NDVI were visualized through density plots and trend lines, examining performance across vegetation labels.

3.5. NN architecture

Even though CNN and ConvLSTM are not exactly the same, or work in exactly the same way, their architecture is still very similar. The main difference between the models is the use of time sequences as an extra dimension in ConvLSTM, however, this is handled easily with *Keras* and *TensorFlow* methods in Python. Thus, the rest of the architecture is fundamentally the same. Since this is the case, this

section will describe the architecture that has been developed to generate functional, efficient and effective DL models.

The first part regarding the development of the architectures is to manage the U-Net architecture. As explained in section 2.6.4, the architecture will include an encoder and a decoder, with up to 1,024 filters, to obtain as many high-level features as possible with the kernels. During the encoder phase, the model will perform the convolutions, as well as activation functions, batch normalization and dropout. The ReLU activation function is employed after each convolution, introducing non-linearity and helping mitigate vanishing gradient issues. Batch normalization is defined in GeeksforGeeks (2025) as a method used to enhance the efficiency and reliability of DL NN models. Its aim is to mitigate internal covariate shift—referring to the distribution of each layer's input changing as they are propagated through the network (GeeksforGeeks, 2025)—through normalization of the inputs at each layer. It results in faster convergence of the loss function and higher learning rates since the risk of gradient problems is reduced. Dropout is utilized to diminish possible overfitting problems (N. Srivastava et al., 2014), where validating and training loss do not converge properly, becoming unable to perform well with unseen data. Therefore, the model would perform well on training data while under-performing for testing data since these might be different or the model might simply be memorizing all the training data (Ying, 2019). To avoid this, dropout it deactivates neurons during training and, in turn, the program is forced to learn without relying too much on a single path, becoming more robust (N. Srivastava et al., 2014). Unlike at the encoder, during the upsamples at the decoder, batch normalization and dropout will not be that aggressive, since excessive uses deleted fine details—tested during parameter tunning—when reconstructing the images back in the decoder. At the end of the architecture, since none of the models perform predictions by themselves, a SoftMax layer will be used for predictions of the final multi-class layer. The layer will contain the same neurons as classes are to be predicted, considered a probability distribution performed on all classes totalling up to 1 (Ndikumana et al., 2018).

$$\sigma(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \tag{3.1}$$

Once the base architecture is developed, the loss function (including the weights) and the optimizers must be properly selected. A good initialization of weights is extremely important, since parts of the network might produce excessive activations, and others barely contribute (Ronneberger et al., 2015). Predetermined weights offered by *Keras*' SparseCategoricalCrossentropy() will be the ones employed for the thesis: however customized weights will also be explored if class imbalances are very acute. Optimizers are the other important factor when compiling the models. They minimize or maximize the loss function, aiming to find the optimal parameters which reduce its value through gradient descents (Desai, 2020). Just as with weights, initialization of optimizers strongly affects the algorithms. As suggested by Desai (2020) in their study, Adam tends to be the default choice in many cases. Although other optimizer options such as SGD, SGD with momentumm, Adagrad or RMSprop are valuable choices, the reason to employ Adam as an optimizer in this project is due to its adaptative learning rate. It can adapt to feaure-scale variations and complex landscape, common in geological maps, and deals with noise and sparse gradients efficiently due to its per-weight learning rate adaptations (Desai, 2020).

3.6. Evaluation

Evaluation of the model's results will be performed by providing the testing sets—at the end of the training phase—mentioned in section 3.4.5 to the model as input. Its predictions are then compared with the test label sets. This comparison is done in several ways, utilizing several metrics and parameters to determine how well the model deals with new unlabelled data.

All of the following metrics described in this section involve operations with true and false positives and negatives. These are described in Pedrayes et al. (2021) as follows:

- **True Positive -** number of correctly identified and classified pixels.
- **False Positive -** number of pixels classified as target class but belong to another class.

- **True Negative -** number of pixels belonging to other classes that have been deemed as correctly classified.
- **False Negative -** number of pixels which have been deemed wrongly classified as other class but are from the target class.

The combination of these parameters helps in determining a wide variety of metrics like accuracy, Intersection over Union (IoU), recall or F1 score.

The first approach is a direct simple visual comparison per patch, determining how well the model represented the new input data. This approach is called accuracy—often known as Overall Accuracy—consisting of counting the percentage of pixels whose class was correctly predicted from the whole pixel count. Lower values of accuracy indicate a high rate of incorrect wrong calls by the model. It describes the general performance of the model (Pedrayes et al., 2021), defined by:

$$\text{OA} = \frac{\text{TP}}{\text{Total number of pixels}}$$

In the case of categorical accuracy—a more class-dependent accuracy metric—, when values are lower, incorrect classification among multiple categories occurs. This often happens due to class imbalances or poor capacity.

On a similar basis is IoU. Especially used in semantic segmentation (as it involves multi-class classification), is employed as an evaluation measure consisting of the overlap ratio between prediction and ground-truth labels (Shibuya & Hotta, 2022). It can be used as an average of all classes or even for individual class analysis, helping in determining which classes perform better or the model struggles with the most. If its values decrease, the model is misclassifying or is poor at localizing features. Its equation is defined by Pedrayes et al. (2021) as:

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}} = \frac{\text{TP}}{\text{TP} + \text{FN} + \text{FP}}$$

Precision is defined as a percentage that determines the predictions that are correct from the total number of predictions made for a class (Pedrayes et al., 2021). Low precision values indicate over-predictions of certain classes:

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

Instead, the Recall metric is a percentage that describes the correctly predicted pixels from a given class (Pedrayes et al., 2021):

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

Finally, the dice and F1 score are similar metrics used to make comparisons easier. Pedrayes et al. (2021) describes it as a metric combining both Precision and Recall. It represents this combination as a single value:

$$F_1 = \frac{2 \times \text{UA} \times \text{PA}}{\text{UA} + \text{PA}}$$

Lower values of the dice coefficient describe poor similarity between predicted and actual ground-truth segments. Thus, the presence of false positives and/or false negatives is common, indicating great class imbalance.

The last factor used to determine the model's performance is the convergence. It consists of the moment where the model's loss becomes minimal and the metrics reach a plateau without any overfitting or underfitting. The loss function has then stopped decreasing as it has reached a minimum level of acceptable error (Hankare, 2023). It indicates that the model is progressively learning adequately and that it reacts well to new inputs of unlabelled data. Hence, both the training and validating curve graphs follow a similar path that decreases as the number of epochs increases. On the other hand, the rest of the metrics described in this section should increase along with the number of epochs.

4. Training and tuning

4.1. Data Preparation

For the training process, data were prepared as established throughout Section 3.4. In that process, the images were not only split into training, validating and testing sets, but also their original images size was reduced to patches of 64x64 (a common practice in DL). The reason behind this is to reduce computational efficiency by making batch processing less demanding as well as reducing training time. After the initial experiments were carried out (Section 5.2), further refinement and experimentation took place (Section 5.7). In such practice, the size of the patches was increased to 128x128 and overlap was integrated—up to 50% in this case—. The idea behind this was to increase the spatial context for the models to learn from, since at times, the predictions would struggle with borders between consecutive or neighbouring patches.

The original image size of the thermal images varied depending on the area. Due to the different locations of the three areas in Spain, there is a differing viewing angle effect from the ECOSTRESS satellite over them. This results in thermal images having different shapes. Therefore, the geology maps and SAR images had to be resized to match their corresponding thermal image's shape:

1. Puertollano shape - 415x273
2. Santa Olalla del Cala shape - 426x278
3. Villoslada shape - 397x267

All of the shapes of the geology maps correspond to their original sizes reduced to thermal image size. The geological maps made by IGME at the scale of 1:50,000 have approximately an area of 28 kilometres long and 18 kilometres wide. Since the patches correspond to areas of 64x64 pixels, the approximate area of each patch is around 19 km^2 (with each side of the patch representing roughly 4.3 kilometres). With the correct shapes and all the data needed, the patches could be constructed to fulfil all the experiments established for this study. These patches were obtained from the available number of images obtained (as explained in Section 3.4) for each area over the span of five years, from the 1st of January 2019 to the 31st December 2024, as there were no available images prior to mid 2018. As observed below, some areas contain more images than others, producing a higher number of patches:

Puertollano patches

- **thermal_only**: 100 images with 2 channels
- **thermal_sar**: 48 images with 4 channels
- **thermal_day**: 56 images
- **thermal_night**: 44 images
- **thermal_summer**: 52 images
- **thermal_winter**: 48 images

Santa patches

- **thermal_only**: 106 images with 2 channels
- **thermal_sar**: 61 images with 4 channels
- **thermal_day**: 32 images
- **thermal_night**: 74 images

- **thermal_summer**: 60 images
- **thermal_winter**: 46 images

Villoslada patches

- **thermal_only**: 309 images with 2 channels
- **thermal_sar**: 73 images with 4 channels
- **thermal_day**: 240 images
- **thermal_night**: 69 images
- **thermal_summer**: 170 images
- **thermal_winter**: 139 images

When combining patches containing different data types, temporal alignment had to be accounted for. In order for the pixels to always contain useful information and for the patches to have the same number of channels, it was necessary for the date of the image to align. Thus, only the dates from which both or all modalities are available—depending on the data types to be used—were utilized. This decision prevents any missing data or missing channels from deteriorating the training process. In addition, it ensures that the patches are describing the same physical conditions during that date. Unfortunately, the number of usable images is dependent on the revisit characteristics each satellite possess. ECOSTRESS has the shortest revisit period in Spain, in the case of the datasets obtained, this number was normally 3 days. Instead, Sentinel-1 had higher revisit times for the datasets obtained, with an average of around 6 days. Thus, SAR images are less frequent, complicating temporal alignment of dates displayed by a noticeable reduction of images available.

4.2. Training process

The patches established in the previous section were employed during the training phase. Each area contains the following files used for training, validating and testing of thermal and SAR imagery. Each file is made of 64x64 patches containing thermal and/or SAR data:

Puertollano

- CNN thermal only
- CNN thermal + SAR
- CNN thermal + SAR_thermal_only
- CNN thermal day
- CNN thermal night
- CNN thermal winter
- CNN thermal summer
- ConvLSTM thermal only
- ConvLSTM thermal + SAR
- ConvLSTM thermal + SAR_thermal_only
- ConvLSTM thermal day
- ConvLSTM thermal night
- ConvLSTM thermal winter
- ConvLSTM thermal summer

Santa Olalla

- CNN thermal only

- CNN thermal + SAR
- CNN thermal + SAR_thermal_only
- CNN thermal day
- CNN thermal night
- CNN thermal winter
- CNN thermal summer
- ConvLSTM thermal only
- ConvLSTM thermal + SAR
- ConvLSTM thermal + SAR_thermal_only
- ConvLSTM thermal day
- ConvLSTM thermal night
- ConvLSTM thermal winter
- ConvLSTM thermal summer

Villoslada

- CNN thermal only
- CNN thermal + SAR
- CNN thermal + SAR_thermal_only
- CNN thermal day
- CNN thermal night
- CNN thermal winter
- CNN thermal summer
- ConvLSTM thermal only
- ConvLSTM thermal + SAR
- ConvLSTM thermal + SAR_thermal_only
- ConvLSTM thermal day
- ConvLSTM thermal night
- ConvLSTM thermal winter
- ConvLSTM thermal summer

As observed, for each of the three AOIs, there are 22 files used for experimentation purposes. Eleven correspond to the CNN model, consisting of simple patches of multi-modal data, and the remaining eleven to the ConvLSTM model, consisting of multi-modal time sequences. The content of each patch is as follows:

- **Thermal only -** it contains only two channels composed of LST and Emissivity from ECOSTRESS imagery.
- **Thermal SAR -** it contains 4 channels composed of the two thermal LST and Emissivity as well as the two additional VV and VH channels from SAR data. Again, only the sar images that match any thermal image in date are included into the patches.
- **Thermal day -** it contains only the thermal images whose time is contained between 6:00 am and 17:59 pm.
- **Thermal night -** it contains only the thermal images whose time is contained between 18:00 pm and 5:59 am.

- **Thermal winter -** patches consisting of thermal images whose date lies between the 22nd September and 21st of March.
- **Thermal summer -** patches consisting of thermal images whose date lies between the 22nd of March and the 21st of September.

To evaluate the inclusion of SAR data improving the model's performance, on the same patches training was performed using (i) thermal and SAR channels, and (ii) thermal channels only for those same patches. This way, an actual comparison between only thermal data and thermal with SAR data can be performed adequately. Thus, ensuring any performance differences stems from additional information rather than from the numerical advantage thermal has over the other two modalities. In the case of seasonal and diurnal patches, only thermal data were employed, since the availability of SAR images was already too limited to make even further divisions. All of the experiments carried out in the following Sections 4 and 5.2 are displayed clearly in the following Figure 4.1, representing the Methodological Framework of this study:

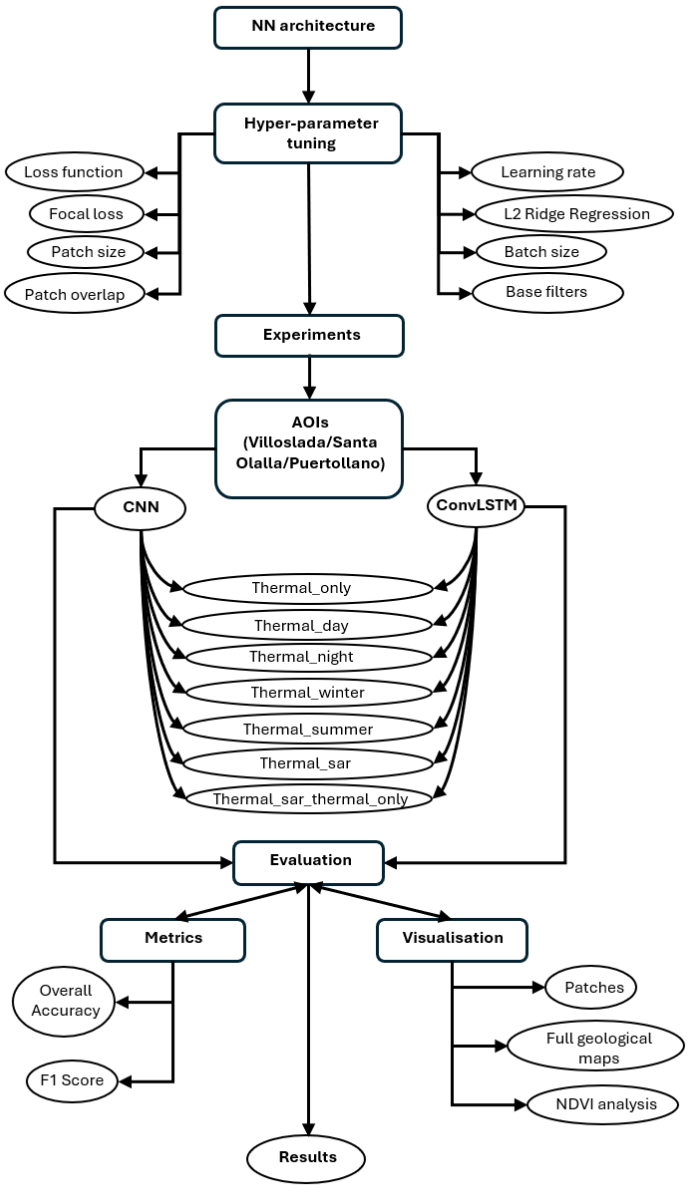


Figure 4.1. Methodological Framework for thesis workflow

4.3. Hyperparameter tuning

An essential part of the training process of a CNN and ConvLSTM involves parameter tuning. Adjustments to different parameters are necessary in order to observe any valuable changes that can enhance the performance of the models. There are numerous parameters that affect the performance, however, this study focuses on four that have shown noticeable impact during preliminary experiments: learning rate, L2 regularization, batch size and filter number. These were then selected for a more detailed evaluation.

In order to test out the optimal parameter configuration to achieve better results, two complementary methodologies were employed. The first one involved a manual search, varying a single parameter at a time while the rest remained constant. This way, the influence of each parameter could be easily assessed through its effects on the loss and accuracy of the model. The second approach to find optimal parametrisation involved the use of *Optuna*. This Python-based framework is compatible with *Keras* and *TensorFlow*, performing automated runs and exploring the predefined parameter pool through model training across different trials involving different combinations. It identifies the combination that yielded the best results.

This way, by integrating insights from both a manual exploration and an automated optimization approach, the best parameters can be determined and adopted for the remaining stages of training and testing. For every run during the hyperparameter tuning stage, the same loss function was employed: SparseCategoricalCrossentropy(). The base metrics used when testing variations are as follow: l2=0.1, lr=0.00001, batch_size=4, filters_base=16.

4.3.1. Learning rate

Throughout the experimentation phase, learning rate had a noticeable impact on the loss function decrease. It consists of a hyperparameter that determines how much the model adjusts its parameters at each step, determining whether a model managed optimal performance by minimizing the gap between predictions and real-world data (Belcic & Stryker, 2024).

A higher learning rate produces very oscillating and divergent loss functions, where convergence struggles to take place in some cases. Although in Figure 4.2 the convergence takes place, it does so too soon, finishing the program after only 18 epochs and producing a very fluctuating accuracy. The model has barely any time to learn patterns and produce actual valuable results.

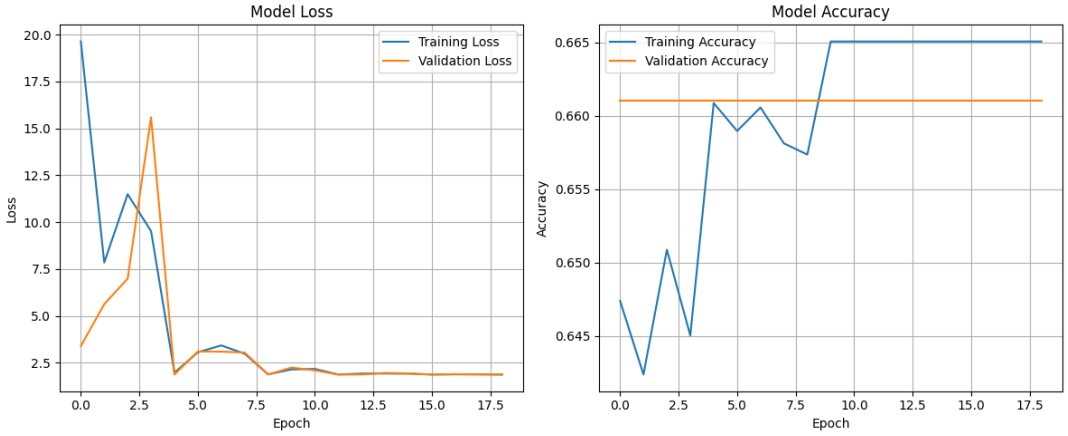
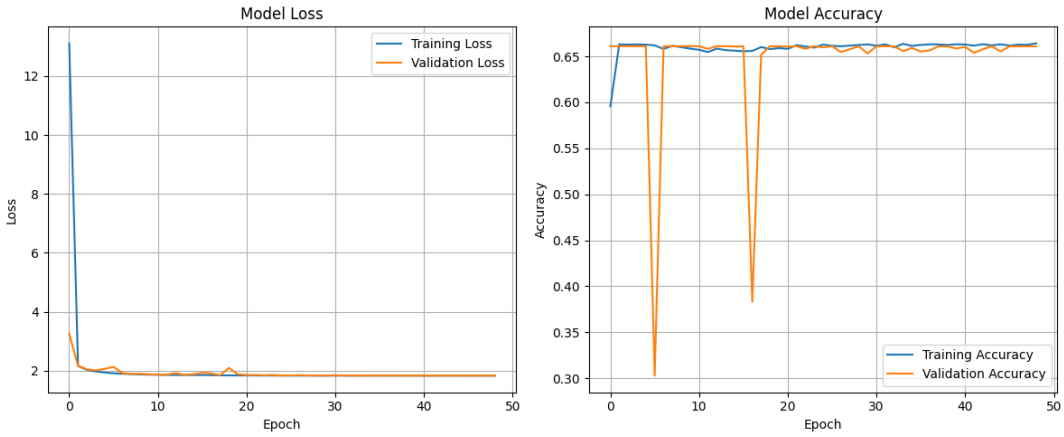
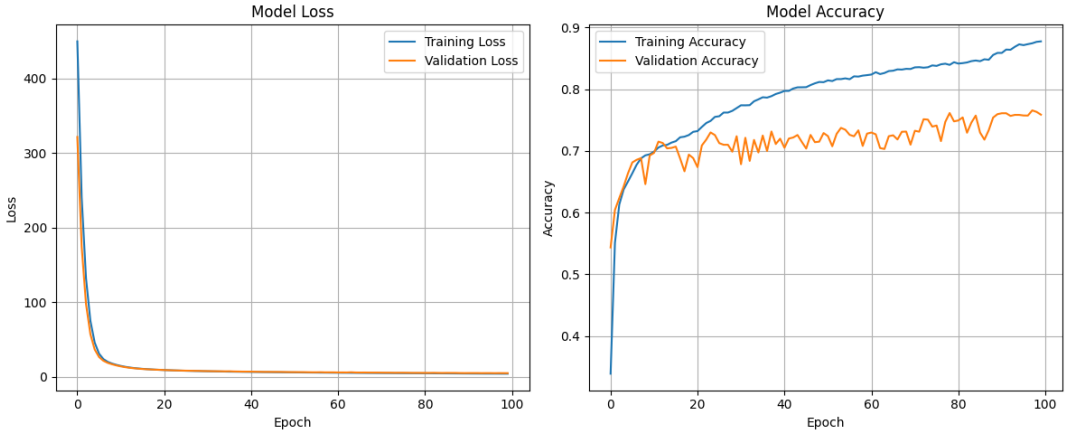


Figure 4.2. Learning rate = 0.1

Instead, as learning rates become lower, they tend to produce more stable and reliable loss functions. However, if the learning rate becomes too low, in some cases, the models can become stuck and not reach its optimal performance within the reasonable amount of epochs established—in this case it was always 100. Lower learning rates produced the results in Figure 4.3:



(a) Learning rate = 0.001



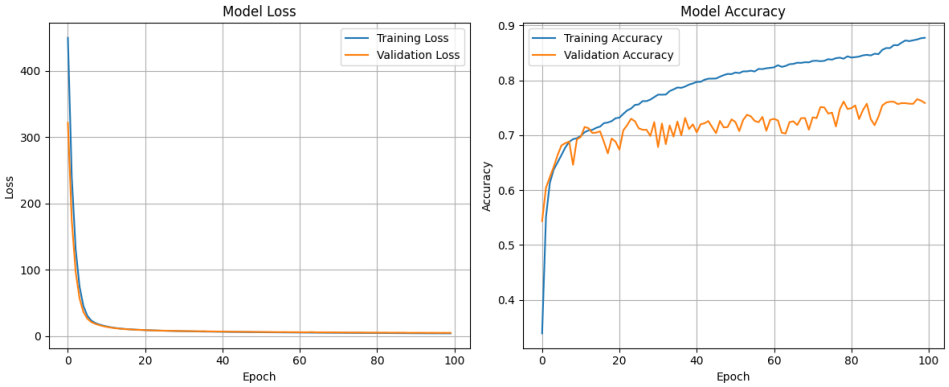
(b) Learning rate = 0.00001

Figure 4.3. Display of the plots resulting from even lower learning rates

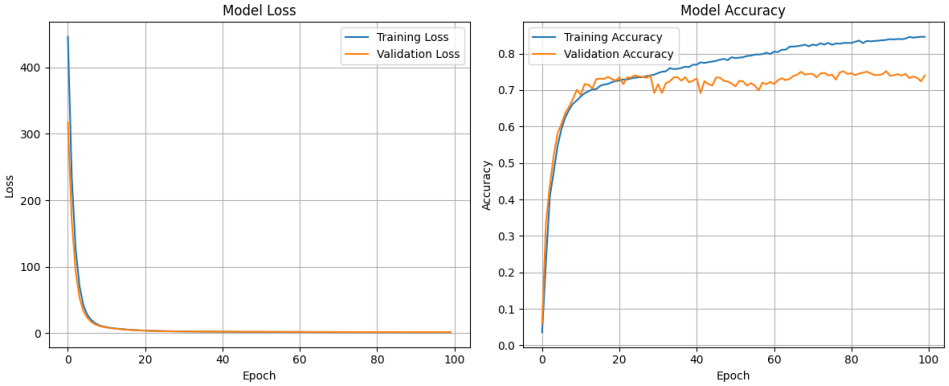
As shown in the figures, the lower learning rates generally produced better loss and accuracy outcomes. For the loss function, a learning rate of 0.00001 yielded the most stable and smooth curve, managing to reach the epoch limit without suffering some abrupt fluctuations. Instead, higher learning rates displayed irregular loss curves and never reached the epoch limit, indicating an unstable training process. Differences were more pronounced regarding accuracy. The 0.00001 learning rate achieved higher final accuracy values. Higher rates produced more erratic behaviours with fluctuations and, in some cases, little effective learning. Even lower learning rates also provided valuable results. A learning rate of 0.0000001 also facilitated learning, though the process became excessively slow, and upon reaching the epoch limit the model had not achieved satisfactory performance. Therefore, 0.00001 was deemed as the best performing learning rate, striking a balance between stable convergence and time-cost efficiency.

4.3.2. L2 factor

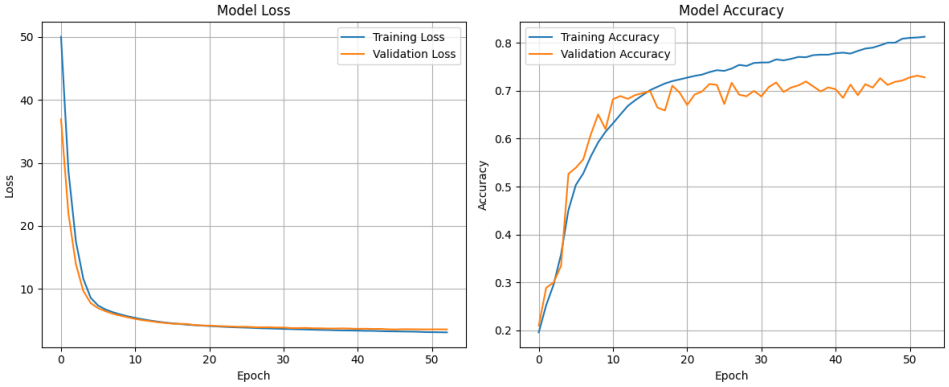
Also known as L2 Ridge Regression, the L2 regularization aims to reduce overfitting by adding penalties as the complexity of the model increases (Tewari, 2021). This penalty is proportional to the sum of squared weights to the loss. Thus, it reduces overfitting problems by preventing the weights from growing too large through forcing them to be small (Tewari, 2021). In turn, it results in the loss being slightly higher due to this penalty term, as observed in Figure 4.4.



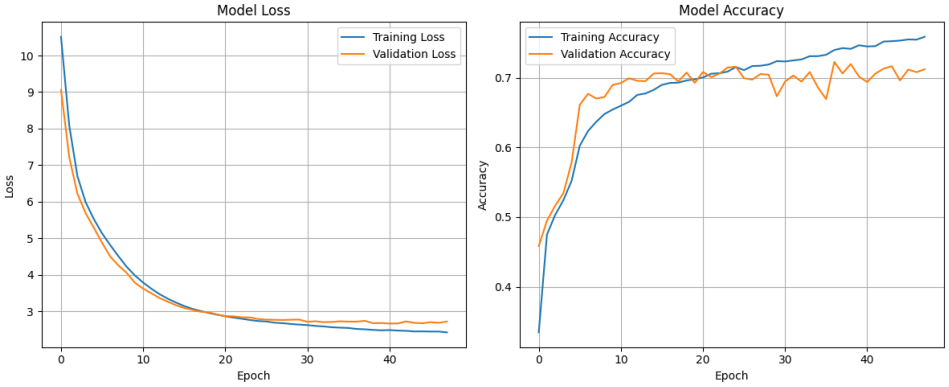
(a) L2 factor: 0.1



(b) L2 factor: 0.05



(c) L2 factor: 0.01



(d) L2 factor: 0.001

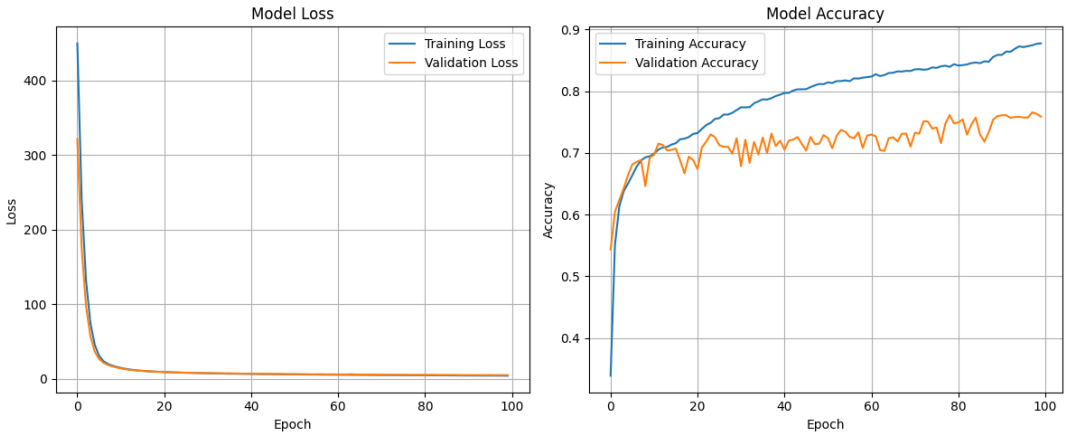
Figure 4.4. L2 regularization analysis and its effects on the loss and accuracy plots of the models

The images display the changes occurring in both the loss and accuracy metrics in the model as the L2 factor is reduced. The lower the L2 regularization factor, the lower the initial value of the loss function. Although this should be potentially positive, the results show that the accuracy achieved is not satisfactory enough. Then, a higher L2 regularization value resulted in a better training process, where the final accuracy reached is higher than with lower values. Nevertheless, overfitting becomes a problem at this point. The lower the value of the L2 factor, the lower the difference between the training and the validating accuracy. However, overfitting can also be tackled through other parameters such as dropout, hence the final decision was to employ an L2 factor of 0.05, the intermediate value. When the L2 factor is high, the initial loss value is higher, indicating how far the model is from correctly predicting a target randomly. Therefore, through a balanced learning rate, the training progress can be potentially better. In addition, the overfitting issue displayed in the highest L2 factor seemed to be slightly corrected when the L2 Ridge Regression was halved. In conclusion, an L2 regularization of 0.05 strikes a balance between reducing the overfitting issue as much as possible and maintaining a higher initial loss value from which the model can learn from.

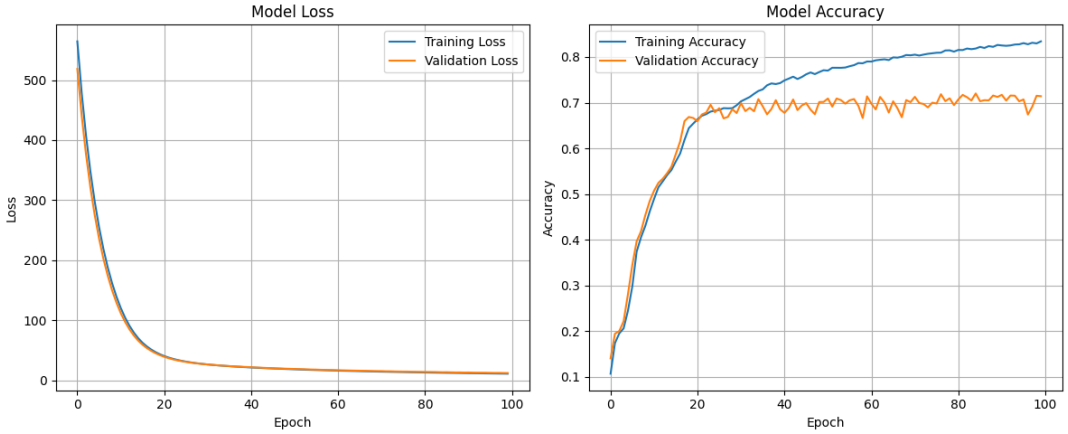
4.3.3. Batch size

The batch size is the amount of batches (subset of training dataset) the model uses one time to learn before updating the weights for the next iteration. Larger batch sizes tend to produce smoother loss function decreases, which might not end up reaching low values by the end of the epoch count. However, a lower batch size (4) produces a steeper loss function graph. Notwithstanding, fluctuations can be common. These differences can be observed in Figure 4.5. Instead, regarding the accuracy improvements, it is smaller batches that can achieve higher final accuracies, since training takes longer and the model can produce better results.

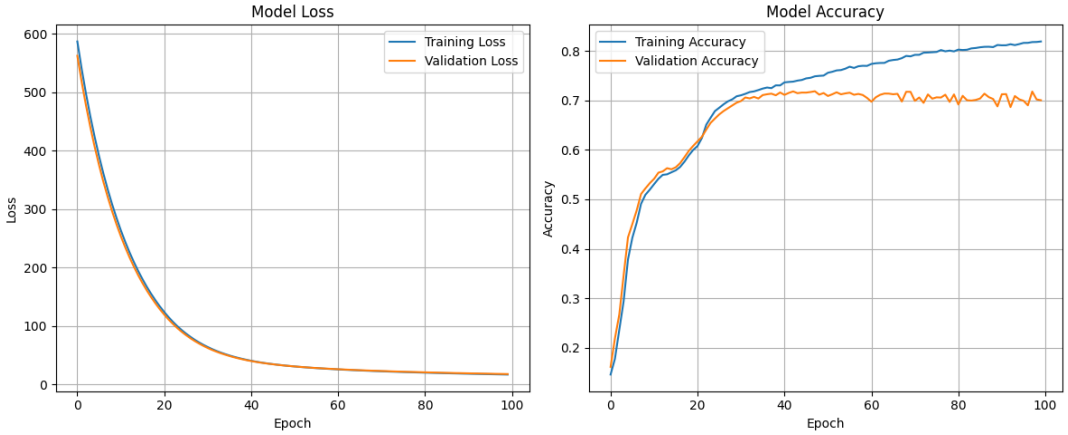
In the results displayed in Figure 4.5, the larger batch sizes struggle to obtain a final higher accuracy. When the model deals with less batches at a time, it can learn finer details more slowly compared to when it is provided with many batches at the same time. The loss function also manages to reach even lower values, which in addition to the higher accuracy, indicates the model manages to learn effectively. In conclusion, a batch size of 4 resulted in better performance and higher final accuracy across several runs.



(a) Batch size: 4



(b) Batch size: 16



(c) Batch size: 32

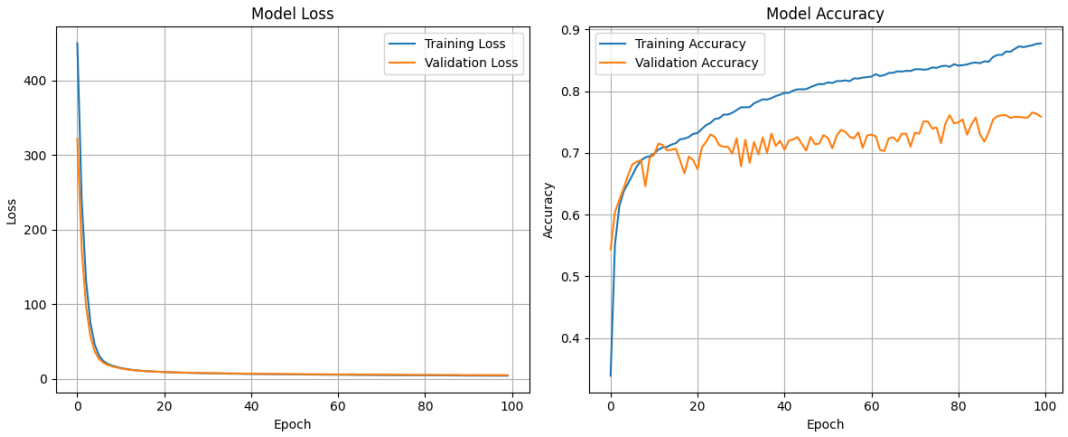
Figure 4.5. Batch size analysis and its effects on the loss and accuracy plots of the models

4.3.4. Filters

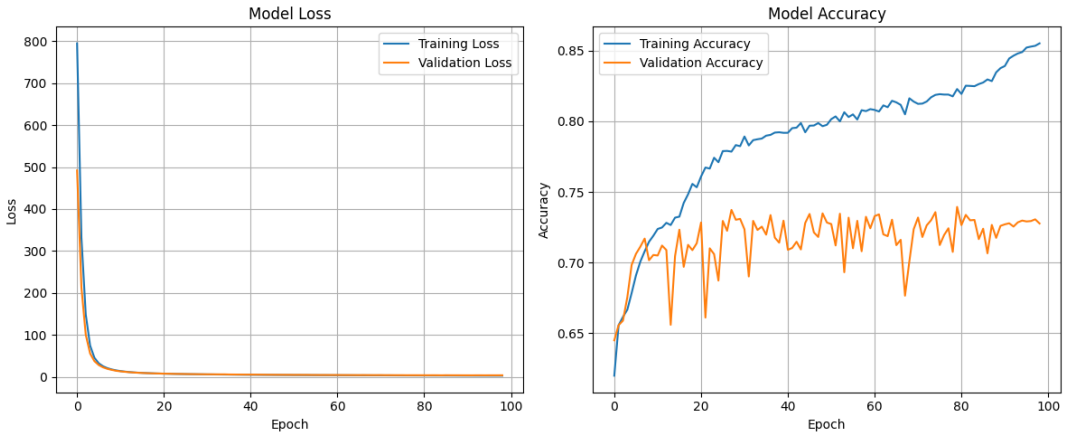
The last parameter to account for was decided to be the initial filters used, also called base filters. A higher number of filters can provide higher accuracy when complexity is high enough, though it can also lead to overfitting. Due to the structure of both the CNN and ConvLSTM and the presence of a U-Net architecture, the number of filters plays an important role in the running time, since the base

filters used will be sequentially doubled for each layer the U-Net has. Therefore, the comparison was performed on the two most common base filter numbers: 16 and 32, observed in Figure 4.6.

In the literature, both 16 and 32 tend be the most used filters. During the training process, using initially 32 filters increased exponentially the time it took for the program to complete 100 epochs. While employing 16 filters takes around 7 hours to complete, using 32 increased the time taken up to 20 hours. At the same time, as observed in the results, the higher filter number produced a noticeable overfitting problem, where the validating accuracy did not manage to adapt and learn properly. Given the clear disadvantages this presented and the little influence it had on the loss function, the decision was to employ only 16 filters as base for the rest of the experimentation process.



(a) Filters base: 16



(b) Filters base: 32

Figure 4.6. Base filter analysis and its effects on the loss and accuracy plots of the models

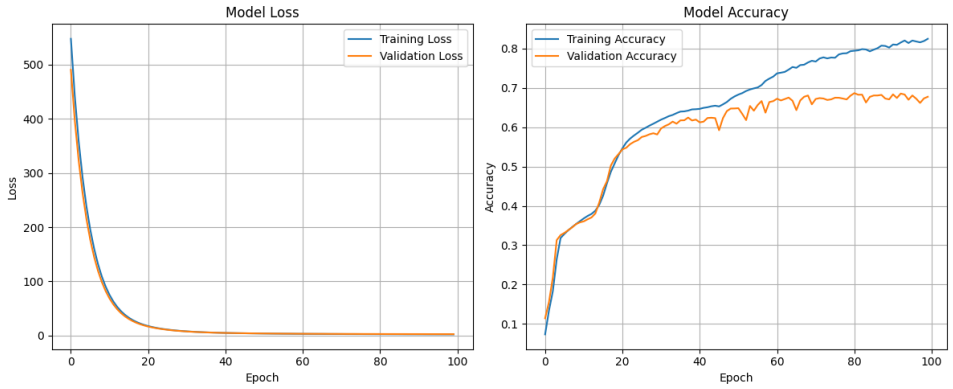
4.3.5. Loss function

The loss function in the experiments was affected by many factors. Most of them have been covered in the previous sections including the L2 regularization, the learning rate or the dropout rate. The choice of loss function had a decisive role in the performance of the models throughout this study. For the sake of fairness and consistency across all initial experiments the same loss function was used—SparseCategoricalCrossentropy() from the *Keras* library. However, later on, a second loss function design was tested to mitigate class imbalance, a recurrent challenge in the results of the experiments.

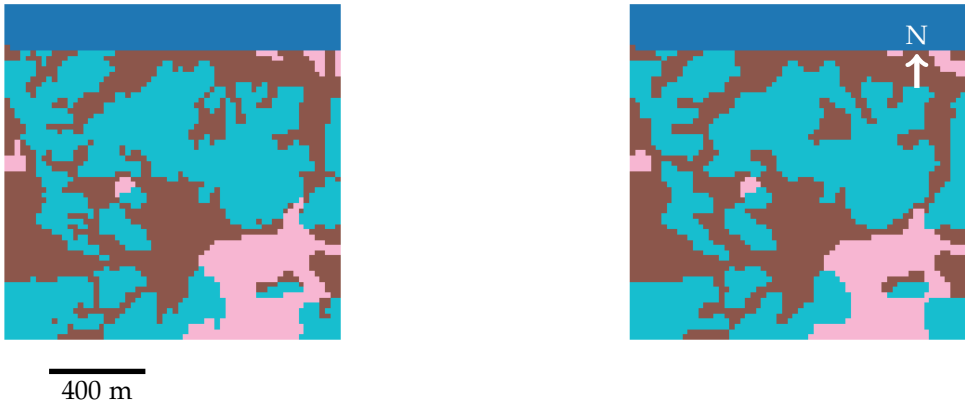
As described by Buda et al. (2018), it is a common problem when employing deep learning classification-based systems, where a higher number of examples exist in the training set for some classes. This phenomenon occurred frequently in this project, where certain rocks and soils are more frequent than

others in the selected regions, making class weighting a promising approach. Through the implementation of a customized loss function, different class weights can be adjusted, giving greater importance to those materials that are rarely predicted or entirely overlooked by the models. This new function called 'simple_weighted_loss' included variable weights for each class. During the training process, soils like Peat, Loam as well as rock classes like Volcanic and Plutonic were often under-represented in the results. In some cases, the model would completely ignore these classes. Instead, more frequent classes dominated the results—Carbonate and Detritic rocks were the most common rock in Villoslada de Cameros and Puertollano, and, Foliated and Non-Foliated rocks in Santa Olalla. To mitigate this problem, the weights assigned to less common classes were adjusted, with different configurations, searching for the most optimal combination—hence encouraging the model to better capture their presence and improve its performance during training and testing. The initial results revealed only marginal improvements for minority classes and for all core experiments, the final choice remained the Keras built-in function, ensuring stability and overall performance.

To test this, one area was chosen (in this case Puertollano), using only the CNN model for simplicity. To maintain fairness, only the loss functions were interchanged, without any modifications to other parameters. Through trial and error, different class weights were tested in search for the best classification results. Below, some examples of different combinations made with their respective results as plots and predictions are shown:



(a) Loss function and accuracy



(b) Ground-truth label (left) and model prediction (right)

Figure 4.7. Plots and predictions for SparseCategoricalCrossentropy() loss function

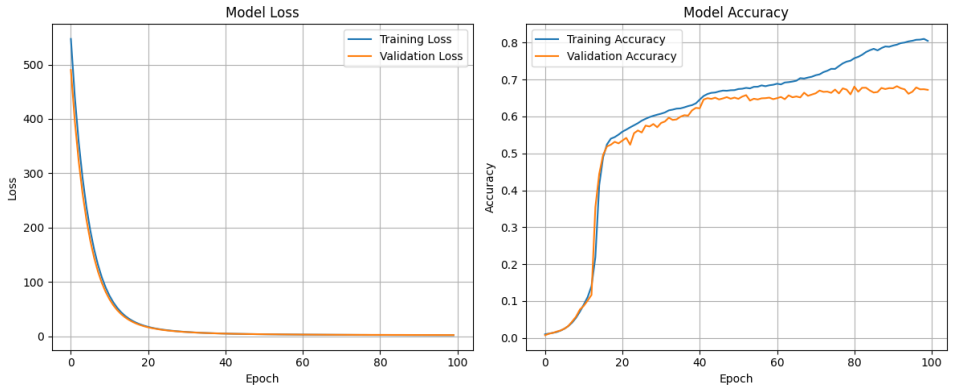
Observing Figures 4.7, 4.8, 4.9, 4.10 as well as Table 1, some distinctions can be observed. The best overall balance and performance belongs to both SparseCategoricalCrossentropy() and the equal value customized loss function. Results show this balance when class weights are kept untouched or equal performance reaches higher accuracy and a lower final loss function value. Since they, in theory, perform a similar methodology, the results are very even—though the differences one might observe could be on the way geologists interpret hidden geological strata. When uncommon classes are slightly boosted, segmentation is improved, especially for those minority classes that received a weight boost, however its overall accuracy and loss values are still worse when compared to equal weighting. Lastly,

when the weight of minority classes is further increased, the accuracy drops significantly and the model struggles more globally, introducing too many minority class pixels. Dominant classes become under-represented, sacrificing their performance without improving rare ones too much.

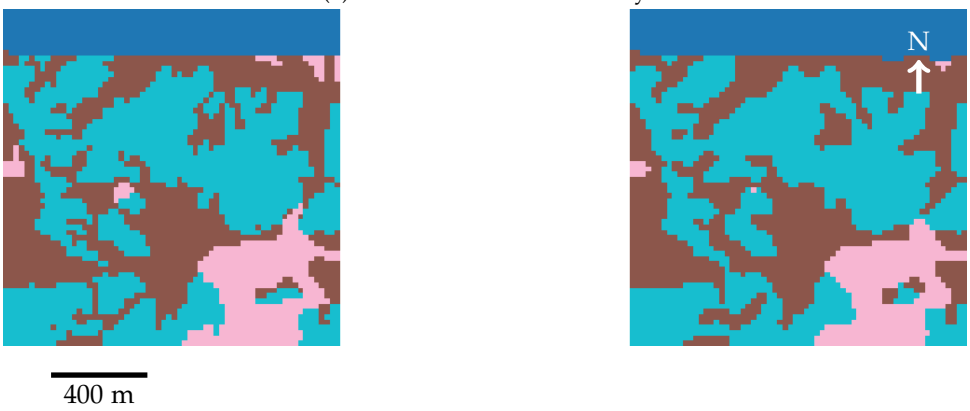
In most cases, classes like Sand and Clay are unrecognized, becoming a problem for proper segmentation results. Instead, Detritic, Non-Foliated and NaNs classes tend to perform and have consistently good F1 scores across all combinations. In conclusion, attempting to increase the performance of minority classes resulted in the loss of accuracy and prediction stability. Results becoming too inaccurate indicated that only small boosts in class weights are necessary to boost the representation of these rare classes. Since the aim of parameter tuning is to obtain the best overall stability and performance, the final decision was to stick with SparseCategoricalCrossentropy(), since it is a *Keras* built-in library that could yield similarly reliable results without risking any coding issues.

Table 1. **Comparison of model performance across different weighting strategies**

Model Setting	**Accuracy**	**Loss**	**F1**
SparseCategoricalCrossentropy	0.698	2.60	0.654
Equal values	0.692	2.52	0.651
Slightly higher (rare classes)	0.619	5.38	0.594
Higher (rare classes)	0.487	8.83	0.482

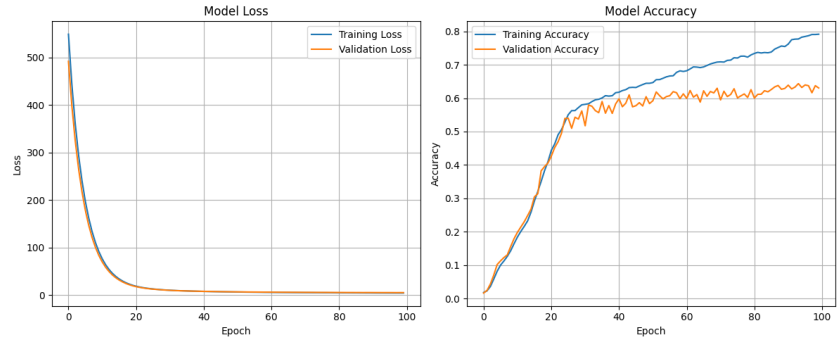


(a) Loss function and accuracy

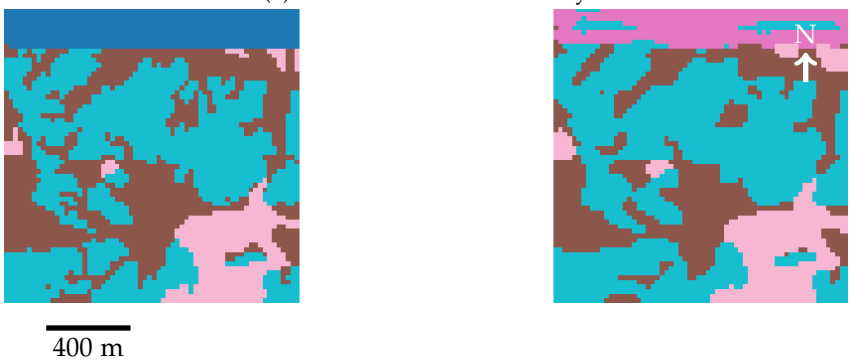


(b) Ground-truth label (left) and model prediction (right)

Figure 4.8. Plots and predictions for 'simple_weighted_loss' loss function with equal class weights

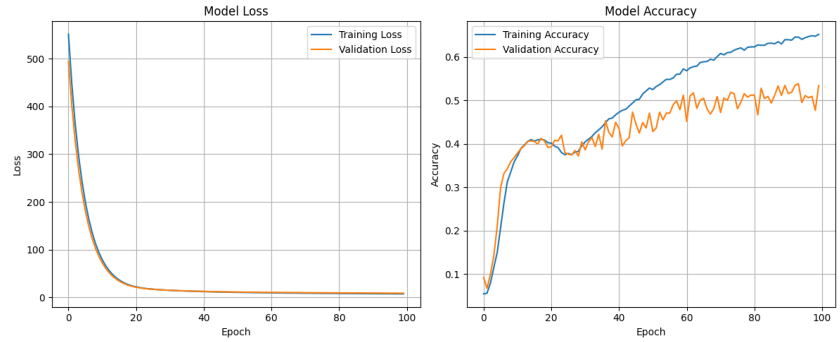


(a) Loss function and accuracy

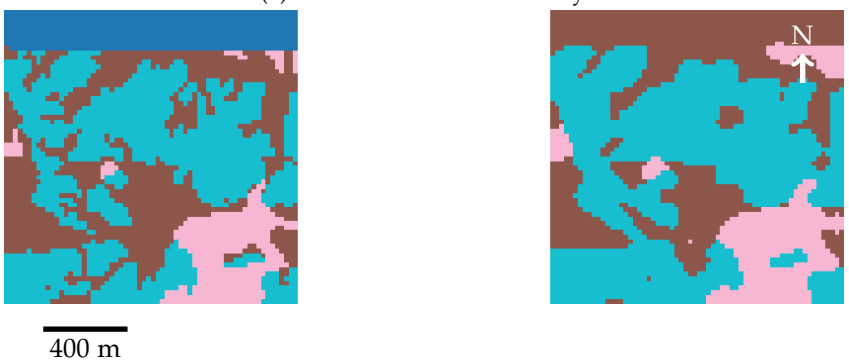


(b) Ground-truth label (left) and model prediction (right)

Figure 4.9. Plots and predictions for 'simple_weighted_loss' with uncommon classes slightly higher weights



(a) Loss function and accuracy



(b) Ground-truth label (left) and model prediction (right)

Figure 4.10. Plots and predictions for 'simple_weighted_loss' with uncommon classes much higher weights

In the later stages of this work, a new and improved customized loss function was developed. This newer version combined focal loss and the previous class weights to handle class imbalance whilst maintaining the overall prediction stability already achieved in the previous stages. Focal loss down-weighs easy examples and focuses on the hard-to-classify ones by facilitating dynamically scaled-gradient updates (Dina et al., 2023). It proved to be particularly useful to capture and improve both the minority class representation and prediction details. This testing only took place for the thermal-only datasets—therefore excluding any diurnal or seasonal experiments as well as SAR—, significantly improving the accuracy and balance of classes in comparison to previous approaches. These results can be observed more clearly in Section 5.7 where the accuracy tables and prediction outputs show major improvements with respect to initial experiments.

4.3.6. Optuna best parameters

As mentioned previously, Optuna was used to determine the best parameter configuration for both models. After testing out different combinations from a wide pool of parameters, Optuna determined the best parameters to be the following:

=== BEST PARAMETERS ===

- l2_reg: 0.001
- learning_rate: 0.0009816442226180768
- batch_size: 16
- filters_base: 16

The results for both the accuracy and the loss function of Optuna with the best parameter selection are displayed below in Figure 4.11. As observed, the accuracy would not reach higher than 0.73 and the process would stop earlier than expected, at only 60 epochs. In addition, both the validating loss and accuracy have fluctuating values, with no convergence seemingly happening towards the epoch limit.

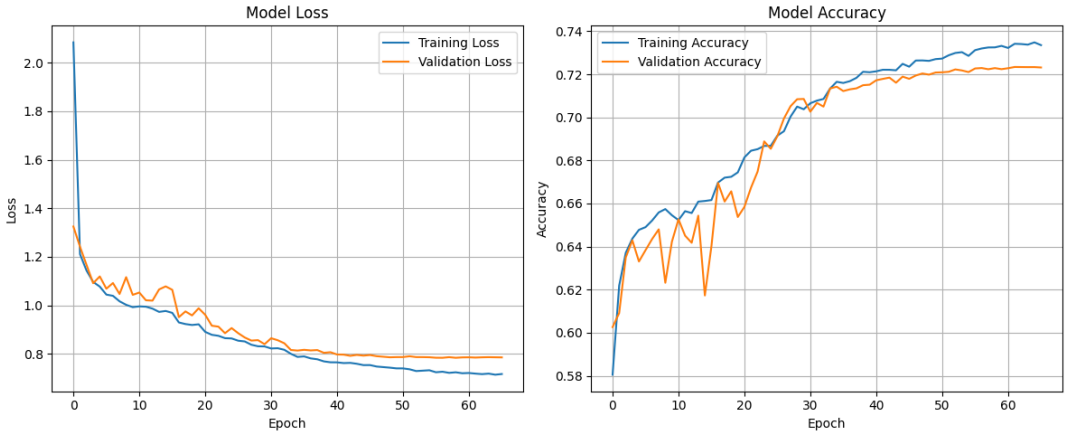


Figure 4.11. Loss and accuracy plots obtained from Optuna's best parameters combination

Although the results of Optuna do not seem to display the best results, they can still provide useful information to determine the best parameter combination when contrasting the results with the manual approach. The final results for both methods are displayed below in Table 2:

With all these results available, only one decision coincides with that of Optuna, the number of initial filters to be used (16). However, the other parameters suggested by Optuna gave worse results in practice, and those options were discarded, thus the final parameter selection considered as the optimal combination for this study is as follows:

- **Learning rate -** 0.00001
- **L2 regularization -** 0.1

- **Batch size -** 4
- **Filters base -** 16

Table 2. **Comparison of accuracies for different hyperparameter configurations**

Parameter	Setting	Accuracy
Optuna (final model)	—	0.7289
L2 factor		
	0.1	0.8045
	0.05	0.7328
	0.01	0.7457
	0.001	0.7289
Learning rate		
	0.1	0.6690
	0.001	0.6708
	0.00001	0.8045
Batch size		
	4	0.8045
	16	0.7324
	32	0.7220
Filters base		
	16	0.8045
	32	0.7492

5. Results

5.1. Thermal data analysis

The first way to analyse how well the models could perform and what classes they could struggle with is through studying their thermal signatures in LST. The thermal signatures of the rocks and soils were studied for all three areas and the results are as follows in Figures 5.1, 5.2 and 5.3:

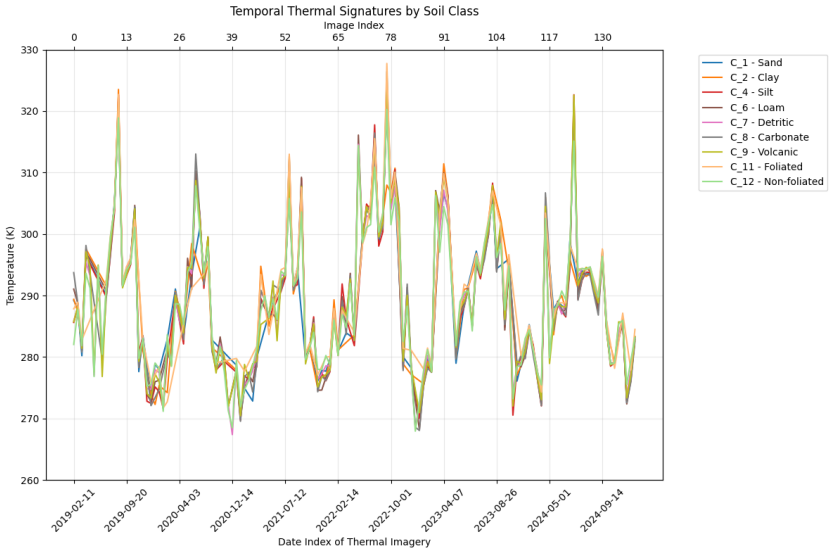


Figure 5.1. Puertollano thermal signatures

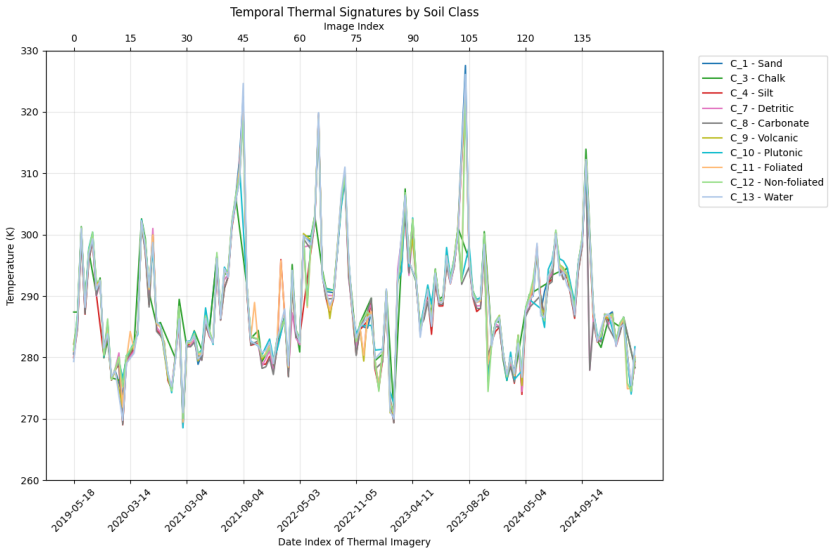


Figure 5.2. Santa Olalla del Cala thermal signatures

As observed in the graphs, all three areas present strong seasonal fluctuations. High peaks in summer and dips in winter with an amplitude of 50 to 60 Kelvin between warmer and cooler periods. In

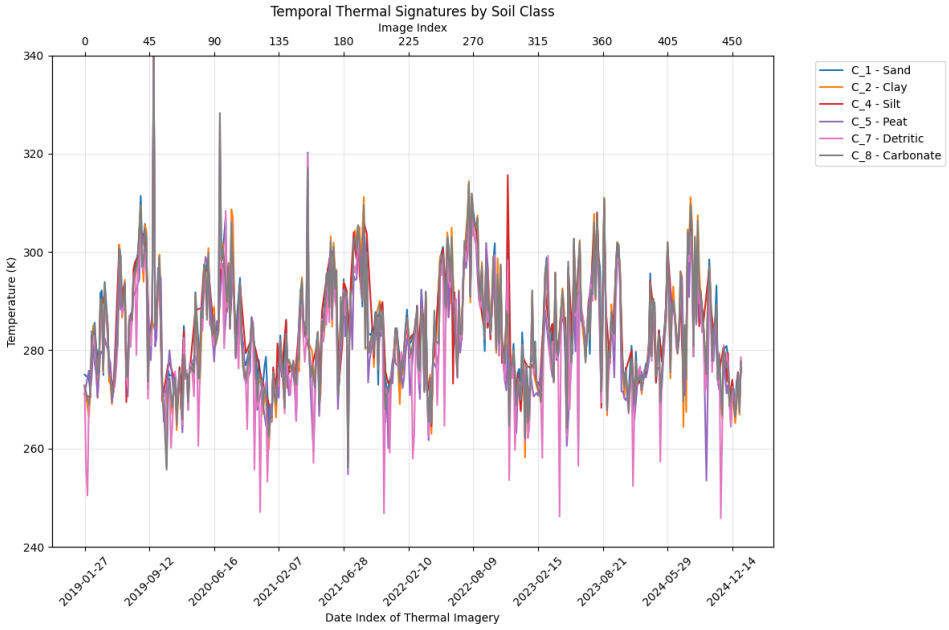


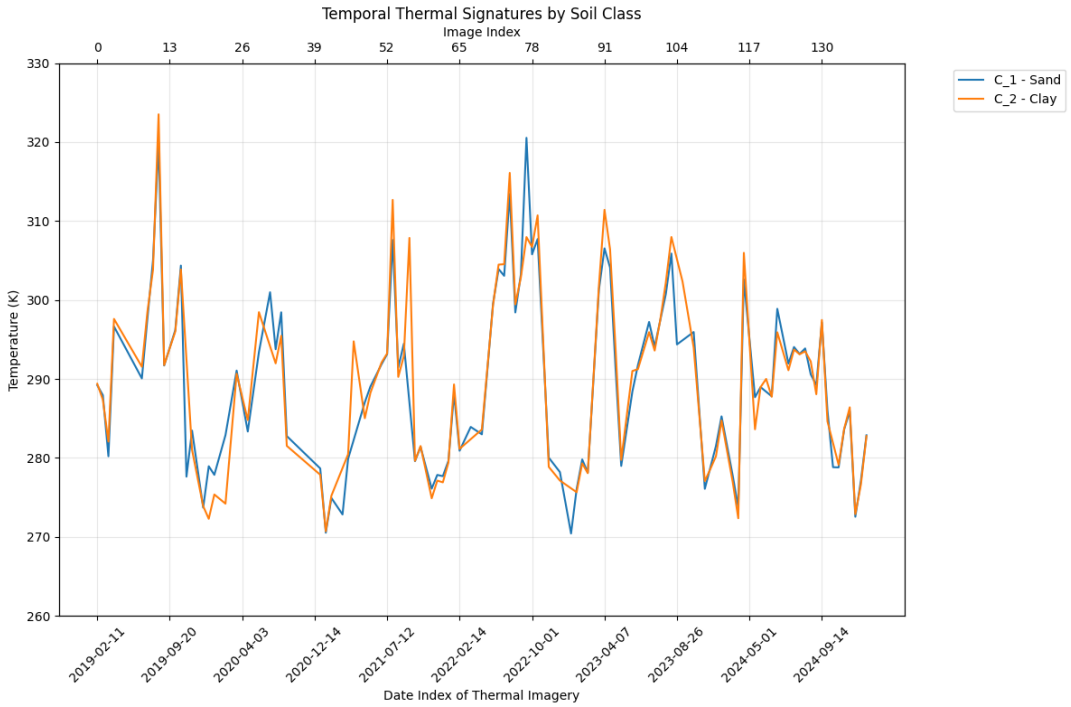
Figure 5.3. Villoslada de Cameros thermal signatures

general, different soil classes track each other closely, however sands manage to match the thermal values of some rocks. Nevertheless, rocks and soils maintain similar patterns, resulting in problems for the models to differentiate between classes. Therefore, in the majority of the initial results, class imbalances were common. Since some classes had many more pixels than others, the program would assume there was a greater chance of success if it chose the most dominant class. Both sand and peat suffer from this imbalance quite clearly, barely being represented in the predictions made by the models.

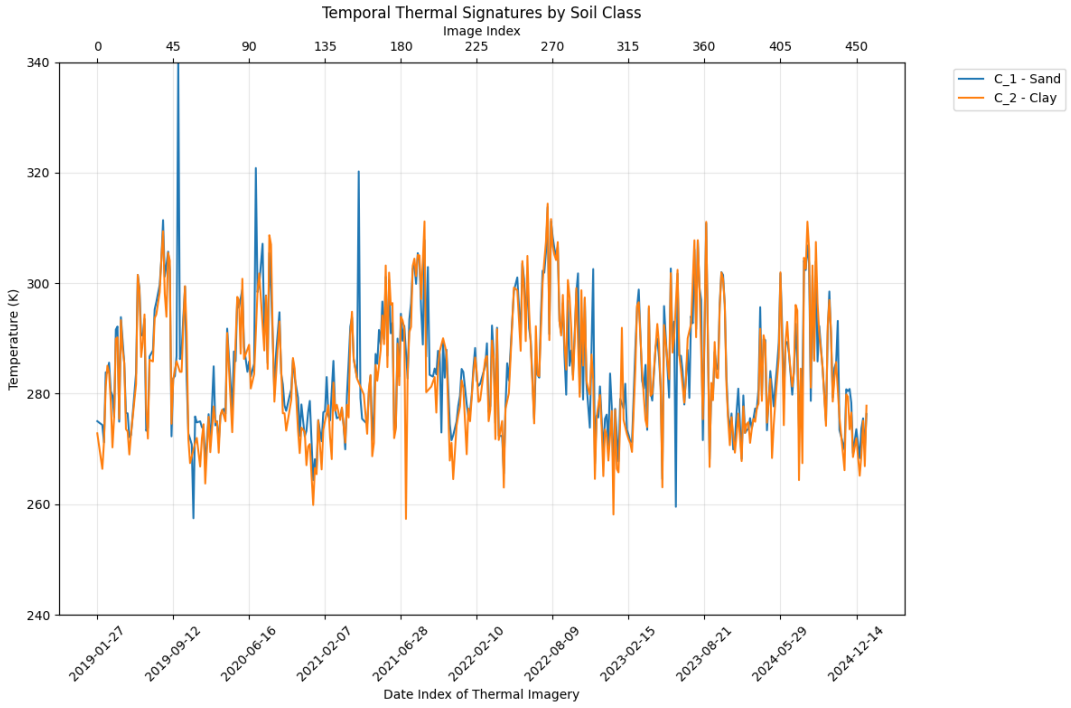
Throughout the three regions, especially for those present in all three locations, there exist some trends. For instance, Sand follows a mid-to-high range trend where it is consistently not the highest or lowest. Clay, instead, presents slightly higher peaks and dips in occasions, with Silt sitting in the mid-to-upper range with slightly cooler values than clay. Sedimentary and Metamorphic rocks stand out for consistently having higher peaks than the rest of the classes in most of the regions. This was perfectly reflected in the model's predictions, where the classes mentioned—Detritic (7), Carbonate (8), Foliated (11), and Non-Foliated (12)—were typically the most dominant classes in the results. These overshadowed other classes with less thermal influence (as observed in the graphs) like igneous rocks or peats and loams for soils.

In order to answer both hypotheses suggested for this study in Section 1.4, separate graphs were plotted to determine the differences in thermal signatures for both sand and clay more clearly. In Figure 5.4, it can be observed that clay generally shows higher average temperatures than sand across both regions. However, sand displays more extreme temperature spikes, indicating faster heating and cooling cycles with shaper and more dramatic changes in temperature. Instead, clay appears to be thermally stable, with similar values all year round, especially in Villoslada de Cameros (North of Spain). Since temperature changes are gradual, it means the material resists rapid temperature changes.

The behaviours seem to slightly align with the physical properties suggested in the hypotheses. On the one hand, due to sand's larger particle size and better drainage capabilities, thermal response is faster. On the other hand, clay, with smaller particles, has a higher thermal retention capability, however the thermal values can be affected by other factors not accounted for in the thermal images such as vegetation, which might be the cause of very similar thermal values for both soils in different regions and seasonal moments.



(a) S-C thermal signatures for Puertollano



(b) S-C thermal signatures for Villoslada de Cameros

Figure 5.4. Sand and Clay thermal signatures for both areas in which they are present: Puertollano (top) and Villoslada de Cameros (bottom)

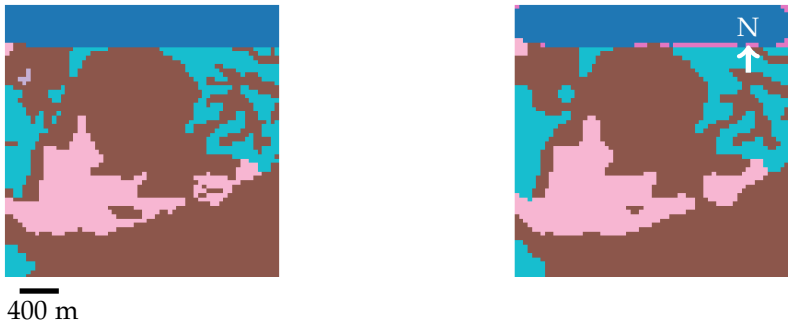
5.2. Thermal dataset

The following sections presents the results obtained using the different datasets employed for both CNN and ConvLSTM models across the three study areas: Puertollano, Santa Olalla, and Villoslada. Different patches will be displayed, to present variety and demonstrate the performance of the model in different areas and against different class distributions. The legend for visualising the classes of the patches can be observed in Figure 5.8. It is important to note than the appearance of the blue bands at the top of the images correspond to NaN values that generated when resampling and converting the original geology maps into the same CRS and resolution as the thermal images, since that extra space is out of bounds of the thermal imagery but part of the original size of the geology map, it was automatically filled with NaN values.

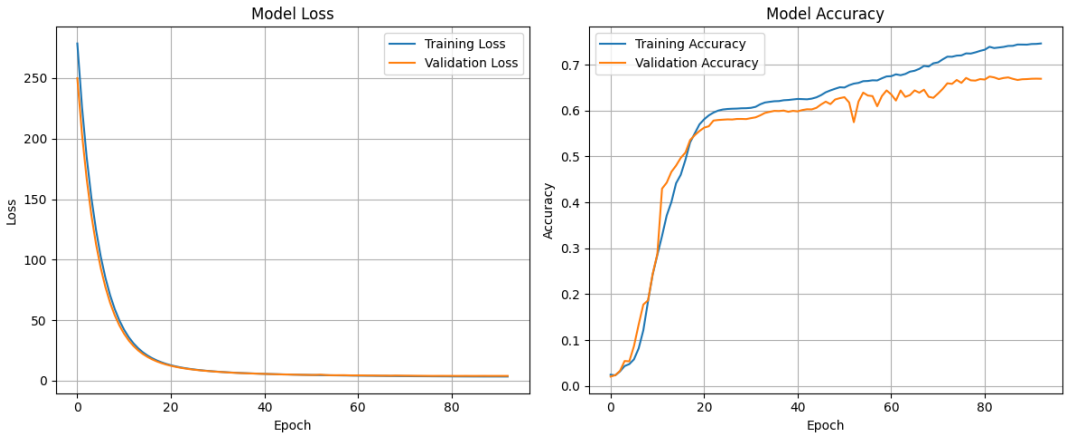
This first section shows the results obtained with only the thermal dataset. The purposes of this experiment are to establish if such approach can be used for rock and soil classification, and as a baseline of model performance with the maximum amount of thermal data available, against which the subsequent diurnal, seasonal and SAR experiments can be compared to. Therefore, determining if their integration into the models potentially enhances their performance.

Since there is a higher number of thermal patches than in other modalities, it is expected for both the CNN and ConvLSTM models to have a more complete training process which, in turn, will potentially yield better results than the rest of the data combinations. Thus, the subsequent sections of this chapter will mainly focus on answering the research questions throughout the different experiments proposed.

5.2.1. CNN

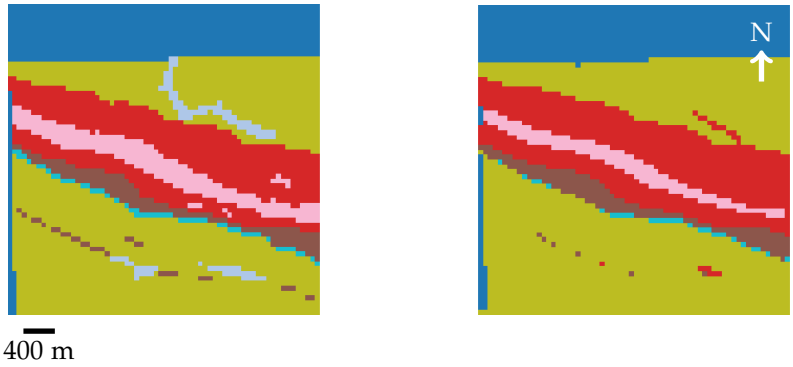


(a) Ground truth (left) and prediction (right)

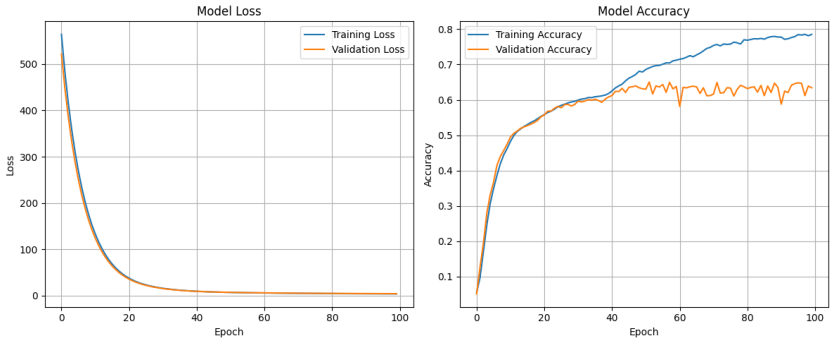


(b) Loss (left) and accuracy (right) plots

Figure 5.5. Image displaying the ground truth, predictions and plots obtained with the CNN thermal model in Puertollano



(a) Ground truth (left) and prediction (right)

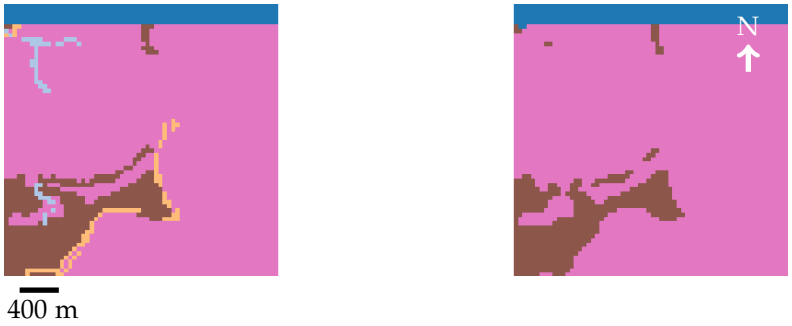


(b) Loss (left) and accuracy (right) plots

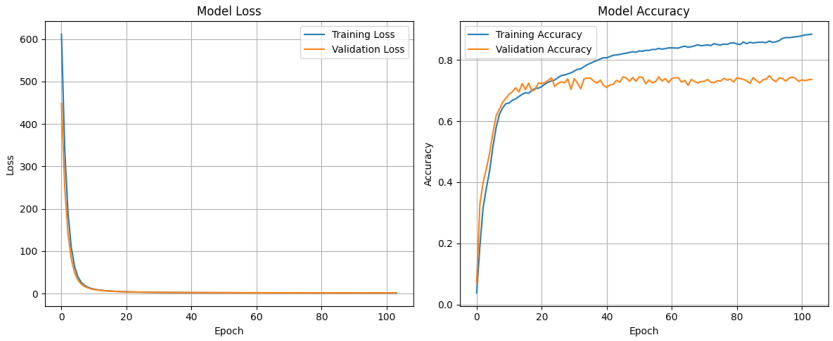
Figure 5.6. Image displaying the ground truth, predictions and plots obtained with the CNN thermal model in Santa Olalla del Cala

As the first experiment of the project, the CNN model demonstrates a solid ability to generate reasonably accurate predictions. Figures 5.5, 5.6 and 5.7 demonstrate that across all three areas, the model displays a characteristic predominance in capturing the dominant geological classes, especially classes 7 (detritic) and 8 (carbonate). Both in Puertollano and Villoslada, the predominance of these classes is quite clear, resulting in the model tending to prioritize them since they are represented frequently in the training patches.

Nevertheless, challenges arise regarding this class preference. Minority classes are rarely represented in the final predictions, excluding them constantly due to the low pixel count of these classes. The absence of less frequent classes like sand (1)—visible as grey-blue in the images—reduces the classification accuracy and IoU, resulting in broader predictions containing only the prevalent lithological units. In the results presented, minority classes where either under-represented or often misclassified as a dominant class, receiving IoU values lower than 0.3. Instead, detritic and carbonate classes managed to obtain IoUs revolving around the 0.7-0.8 mark. Table 3 shows that this model presents one of the highest accuracy levels achieved across all different modalities. As a result, the CNN approach yields satisfactory outputs, indicating that performing rock and soil classification with such a model employing only thermal imagery is suitable and can become a very reliable process if further enhanced, since as of now, it produces too much generalisation.



(a) Ground truth (left) and prediction (right)



(b) Loss (left) and accuracy (right) plots

Figure 5.7. Image displaying the ground truth, predictions and plots obtained with the CNN thermal model in Villoslada de Cameros

5.2.2. ConvLSTM

The ConvLSTM model demonstrates slightly superior performance compared to the CNN, particularly in its ability to capture less dominant classes such as sand or clay. While their representation remains imperfect, the level of detail and the inclusion of less frequent classes do improve when compared to the CNN model. This suggests that the incorporation of temporal data provide the model with additional context, which the model employs to better represent the lithological units (Figures 5.9, 5.10 and 5.11). In the case of Santa Olalla, it achieved predictions that were nearly indistinguishable from the ground truth, proving its potential for accurate geological mapping under favourable conditions. To understand the images better, the colour assigned to each class is displayed in Figure 5.8.

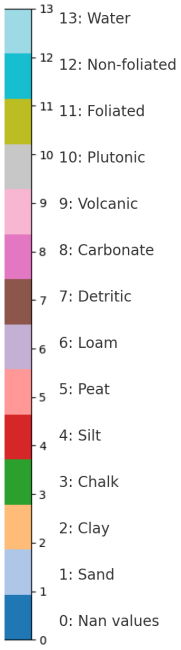
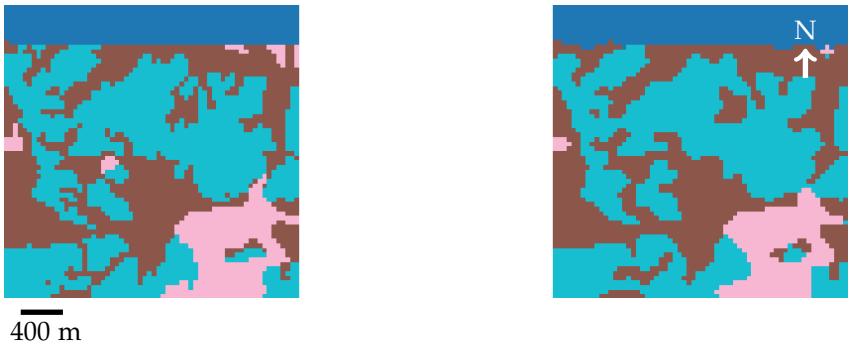


Figure 5.8. Soil and rock class legend

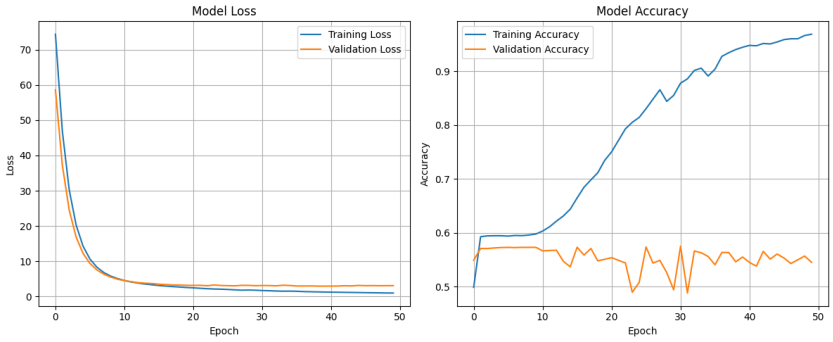
This improvement arises from the ConvLSTM architecture's capacity to retain both the short-term and long-term temporal dynamics, which enable the model to understand both the current state of the terrain and the evolution it underwent in time. As explained in Section 2.6.3, the short-term memory h_t responds to new changes between consecutive time steps, while the long-term memory C_t retains the contextual information accumulated throughout the sequence. It is this duality that allows the model to compare the new thermal data against its own previous context, resulting in stable and consistent predictions, something that a CNN model cannot reproduce due to its purely spatial nature. The ConvLSTM uses the temporal dependencies to its advantage through its sequential architecture, preserving the temporal continuity offered by time sequences to disambiguate classes that could potentially appear similar in the thermal spectrum at individual time steps but differentiate each other in temporal behaviours.

Despite the results, the model is also present with some limitations. Similar to the

CNN, some patches still deviate from expected outcomes, with some results being minor misclassifications while others are considerably deviated patches. These discrepancies contribute to the observed gap between training and validation accuracy, resulting from overfitting. Interestingly, in this study, overfitting coincided, in some cases, with improved results. This would suggest that the model may have benefited from memorizing finer spatial details. This paradox would require further investigation. It is in fact the ConvLSTM that generally presents more acute overfitting than its CNN counterpart. This only occurred with accuracy plots, since the loss function would generally converge smoothly, indicating in the end that the learning and training process performed a gradual optimization of the model's parameters. Yet, the validating accuracy would always halt its progress and produce noticeable overfitting issues. This pattern would repeat itself for most upcoming experiments, where the CNN model would normally have better converging accuracy lines between the training and validating accuracy. Strangely, ConvLSTM would still manage to produce more accurate predictions closer to the ground truth, even though the learning process seemed to not be as effective. ConvLSTM would also be the model to take always the longest to compute everything. Leveraging temporal context as an additional dimension of information comes at the cost of increased computational time and memory usage when compared to the CNN. In the end, both CNN and ConvLSTM demonstrated impressive potential in rock and soil classification (Table 3), especially when temporal parameters are involved, resulting in almost perfectly represented geological terrain.



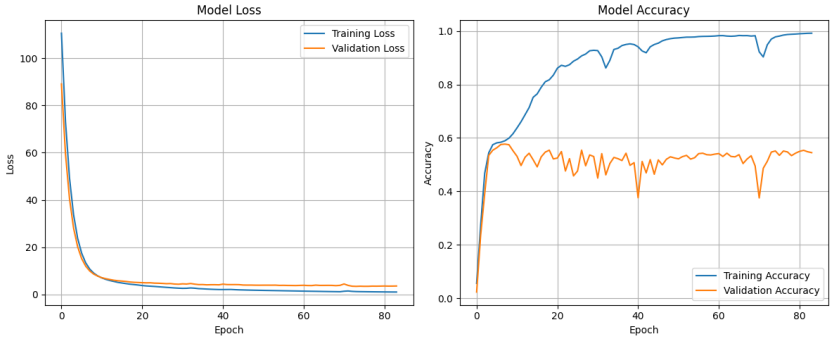
(a) Ground truth (left) and prediction (right)



(b) Loss (left) and accuracy (right) plots

Figure 5.9. Image displaying the ground truth, predictions and plots obtained with the ConvLSTM thermal model in Puertollano

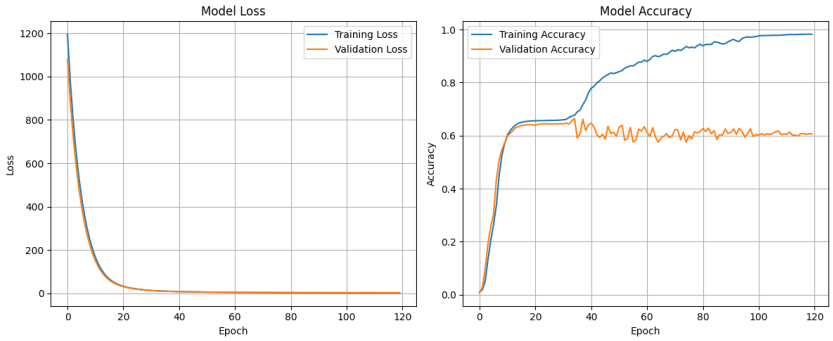
(a) Ground truth (left) and prediction (right)



(b) Loss (left) and accuracy (right) plots

Figure 5.10. Image displaying the ground truth, predictions and plots obtained with the ConvLSTM thermal model in Santa Olalla del Cala

(a) Ground truth (left) and prediction (right)



(b) Loss (left) and accuracy (right) plots

Figure 5.11. Image displaying the ground truth, predictions and plots obtained with the ConvLSTM thermal model in Villoslada de Cameros

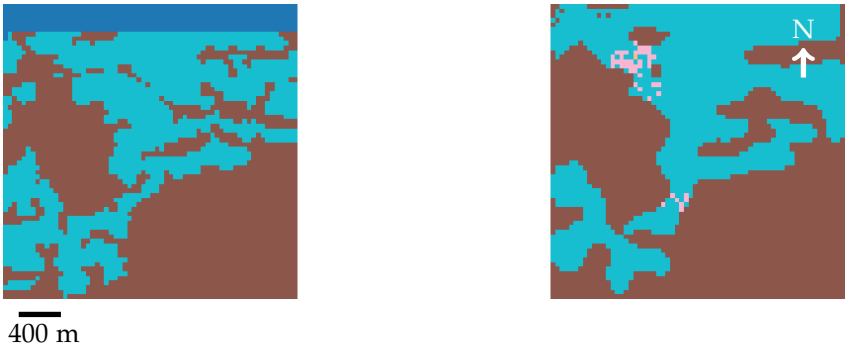
5.3. Diurnal experimentation

To investigate the effect of diurnal cycles on model performance and the thermal changes sand and clay go through in a single day, the dataset was divided into daytime and nighttime subsets. This way, illumination and rapid temperature dynamics can be explored, determining their influence on the model's performance. The following subsections display the results obtained in each model when both day and night patches are employed.

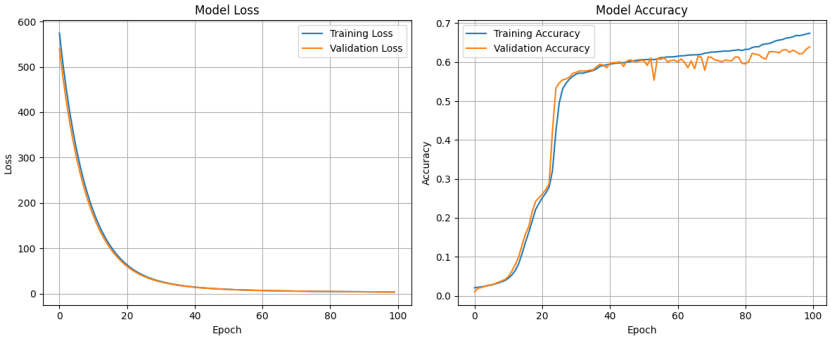
5.3.1. Day

CNN

For the daytime experiments, the CNN generally produced satisfactory results across most runs (Figures 5.12, 5.13 and 5.14 and Table 4), confirming its ability to extract meaningful spatial features. Particularly, in the case of Puertollano and Villoslada, where it managed to produce some very accurate patch representations. When compared to the model trained on the complete thermal dataset, its performance was lower—especially the CNN—in most cases, with lower accuracy results and struggling with producing much finer details (better viewed in Table 4 in the F1 scores). Once again, the CNN model performed well with broader classes but suffered to produce consistent distinctions with less frequent classes. This is likely due to the increased changes and influence of surface temperature variabilities and illumination during the day hours.

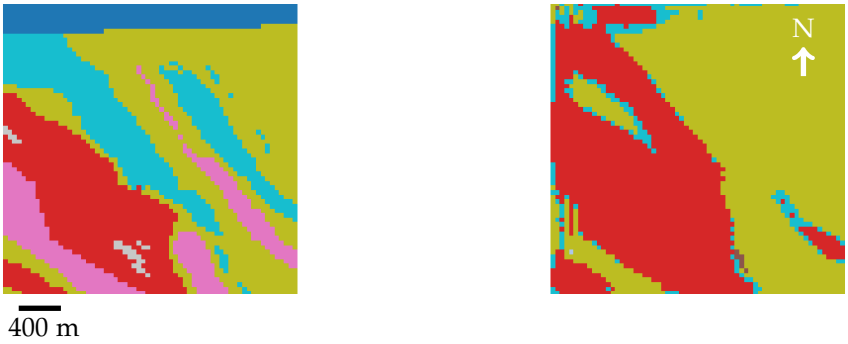


(a) Ground truth (left) and prediction (right)

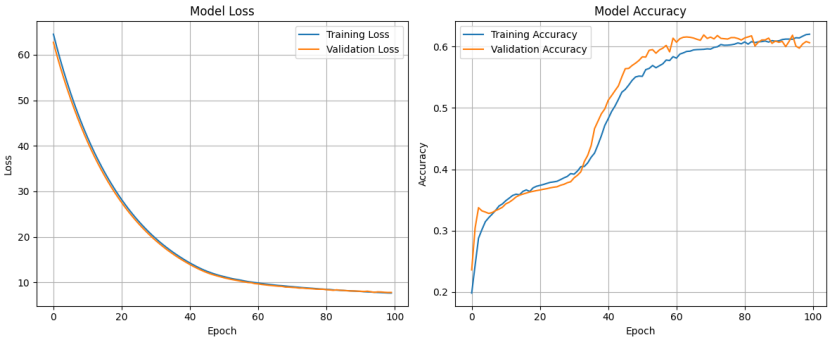


(b) Loss (left) and accuracy (right) plots

Figure 5.12. Image displaying the ground truth, predictions and plots obtained with the CNN day model in Puertollano

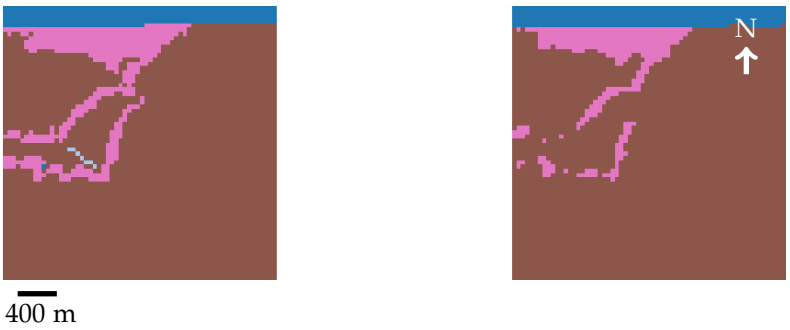


(a) Ground truth (left) and prediction (right)

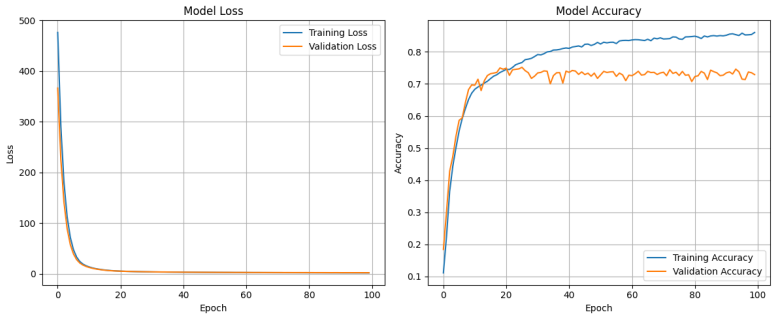


(b) Loss (left) and accuracy (right) plots

Figure 5.13. Image displaying the ground truth, predictions and plots obtained with the CNN day model in Santa Olalla del Cala



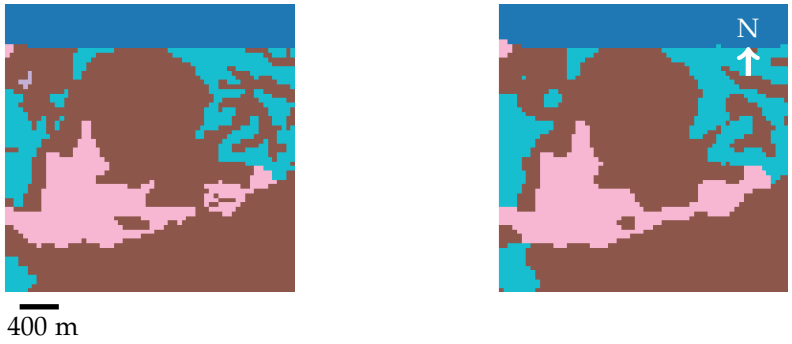
(a) Ground truth (left) and prediction (right)



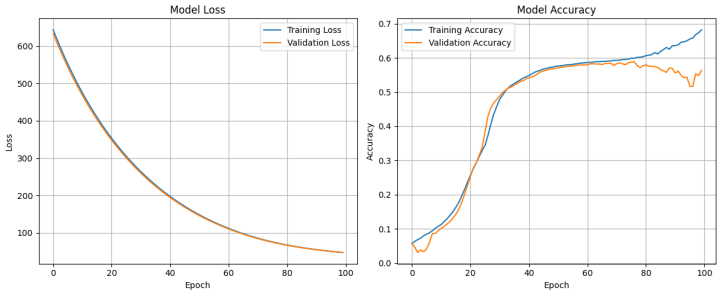
(b) Loss (left) and accuracy (right) plots

Figure 5.14. Image displaying the ground truth, predictions and plots obtained with the CNN day model in Villoslada de Cameros

ConvLSTM



(a) Ground truth (left) and prediction (right)

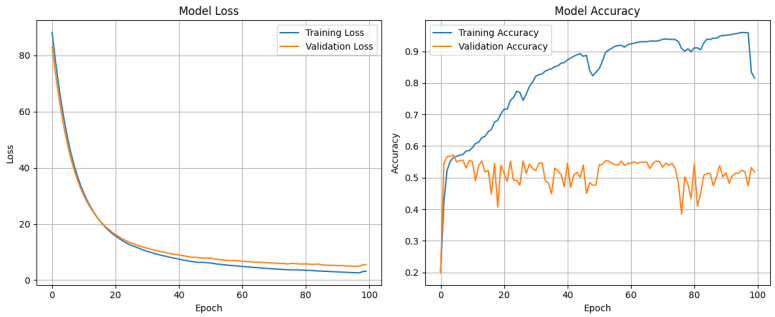


(b) Loss (left) and accuracy (right) plots

Figure 5.15. Image displaying the ground truth, predictions and plots obtained with the ConvLSTM day model in Puertollano

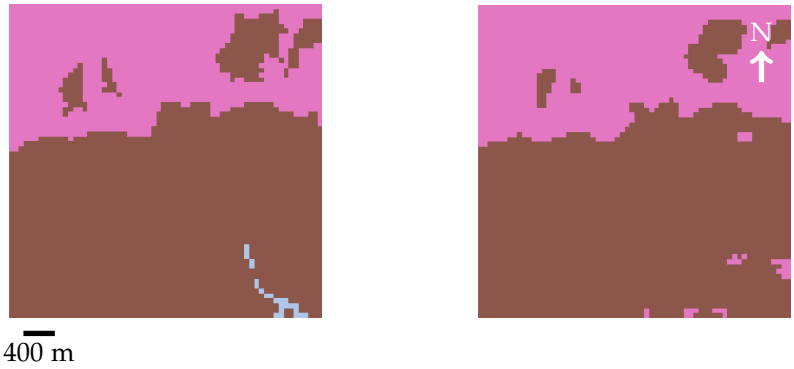
Once again, the ConvLSTM performed very well, especially in including more minority class pixels and outputting more detailed results in the final outputs. Contrary to the CNN, this model performed exceptionally well in Santa Olalla, albeit the overfitting issue, whose predictions show great resemblance to the ground truths as observed in Figure 5.16. In this case, there is a chance the model actually memorized some of the labels, reproducing them during the outputs since the overfitting issue became so accused for this specific case. Puertollano and Villoslada lack some detail in their outputs with minority classes still being a problem in the representations, but still remain as valuable predictions with noticeable improvements over the CNN model and with reduced overfitting in comparison with Santa Olalla. The ConvLSTM model proved once again its capacity to outperform the CNN, which is also reflected in Table 4. As it takes into account temporal activity, the variation in temperatures through the daytime images across several classes provides more daily context to the model, proving to have positive impact on the results.

(a) Ground truth (left) and prediction (right)

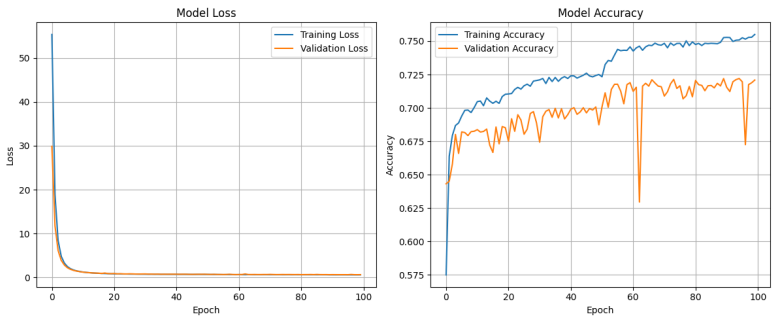


(b) Loss (left) and accuracy (right) plots

Figure 5.16. Image displaying the ground truth, predictions and plots obtained with the ConvLSTM day model in Santa Olalla del Cala



(a) Ground truth (left) and prediction (right)



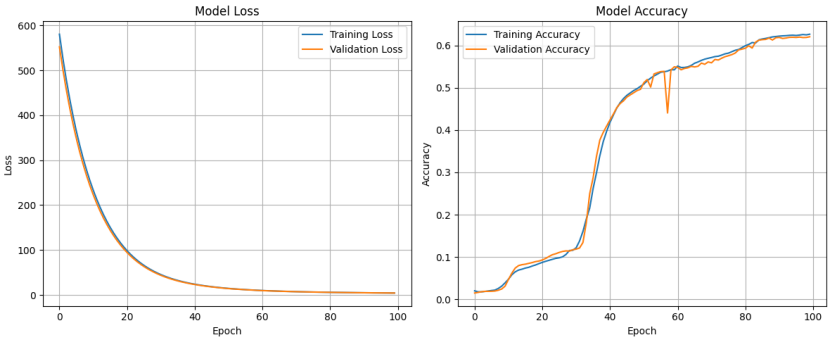
(b) Loss (left) and accuracy (right) plots

Figure 5.17. Image displaying the ground truth, predictions and plots obtained with the ConvLSTM day model in Villoslada de Cameros

5.3.2. Night

CNN

(a) Ground truth (left) and prediction (right)

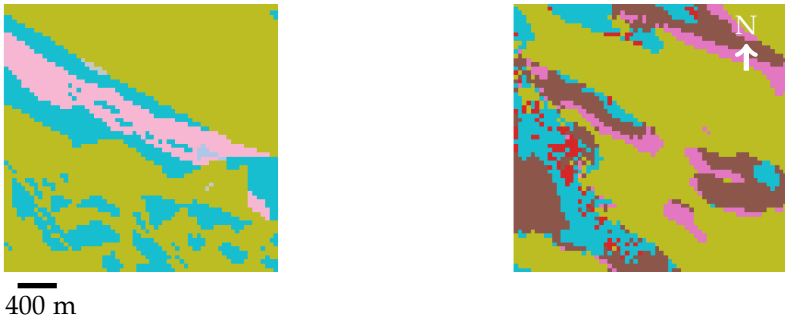


(b) Loss (left) and accuracy (right) plots

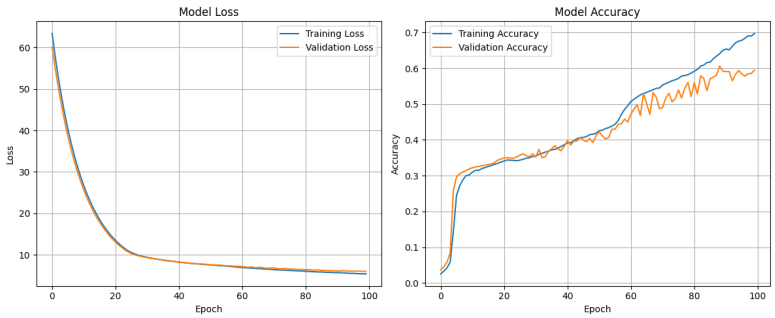
Figure 5.18. Image displaying the ground truth, predictions and plots obtained with the CNN night model in Puertollano

The results of nighttime experiments were mixed 5. Interestingly, in the case of Villoslada de Cameros (Figures 5.20 and 5.23), both in the CNN and ConvLSTM models offered promising results. The model manages to produce a good approximation of the ground-truth data, yielding a geological map that resembles the original one. Nevertheless, the presence of minority classes still remained underwhelming. For the remaining regions (Figures 5.18, 5.19, 5.21 and 5.22), the model presents some poor results, with a higher chance of misclassification of classes or even completely wrong distributions as is the case with Santa Olalla. The model can produce some valuable patches, but in the CNN the results show that the outputs lack much of the detail obtained in the daytime counterpart. Therefore, in the case of the CNN model, nighttime values seem to not be sufficient to produce a reliable product, and should instead be considered to employ the full thermal dataset including daytime data as well.

The ConvLSTM approach also responds in a similar way, producing very mixed results with frequent misclassifications and erroneous data predictions for Puertollano and Santa Olalla. Yet again, Villoslada presents a much clearer prediction. Villoslada de Cameros, located in the North of Spain, is a very vegetated area, due to frequent rainfalls throughout the year and more stable temperatures. This could partly explain the reason behind the model performing more consistently exclusively for this region. Since vegetation can hold heat for longer periods of time than bare soils and temperature extremes are less frequent than in the other two regions, the thermal values of ECO images tend to remain more stable through the night as well, aiding the model in learning the patterns better. In general, and once again, the results show a better performance than the CNN but are still underwhelming in many cases, leading to the use of daytime images being deemed as necessary if a valuable classification can be made.

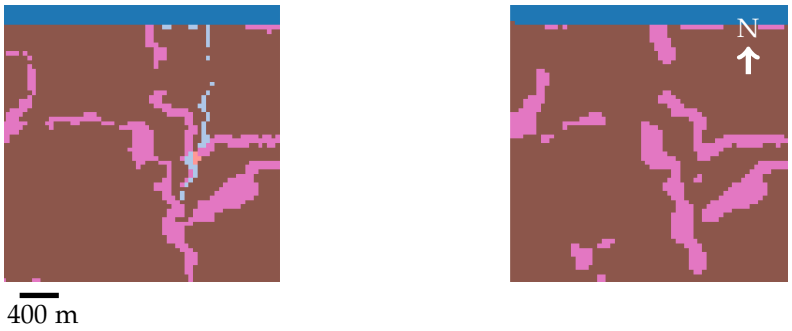


(a) Ground truth (left) and prediction (right)

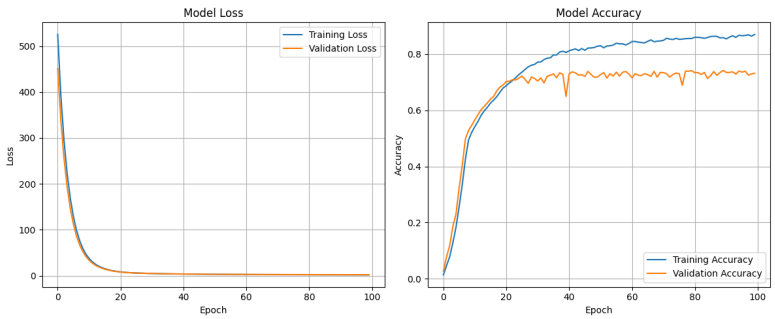


(b) Loss (left) and accuracy (right) plots

Figure 5.19. Image displaying the ground truth, predictions and plots obtained with the CNN night model in Santa Olalla del Cala



(a) Ground truth (left) and prediction (right)

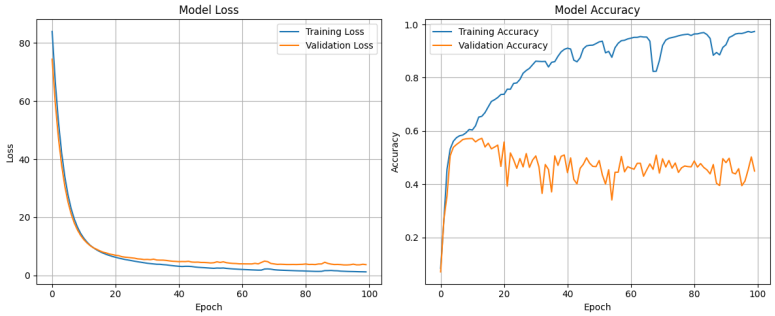


(b) Loss (left) and accuracy (right) plots

Figure 5.20. Image displaying the ground truth, predictions and plots obtained with the CNN night model in Villoslada de Cameros

ConvLSTM

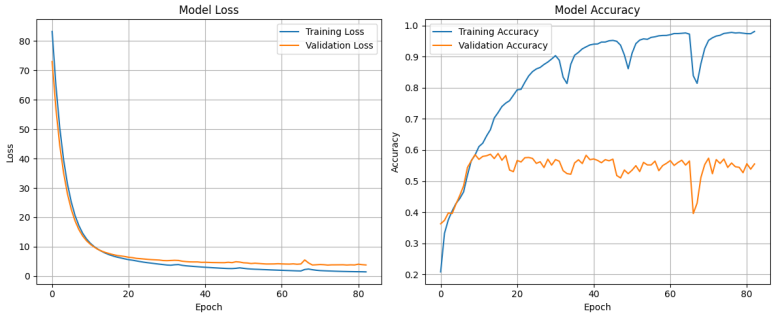
(a) Ground truth (left) and prediction (right)



(b) Loss (left) and accuracy (right) plots

Figure 5.21. Image displaying the ground truth, predictions and plots obtained with the ConvLSTM night model in Puertollano

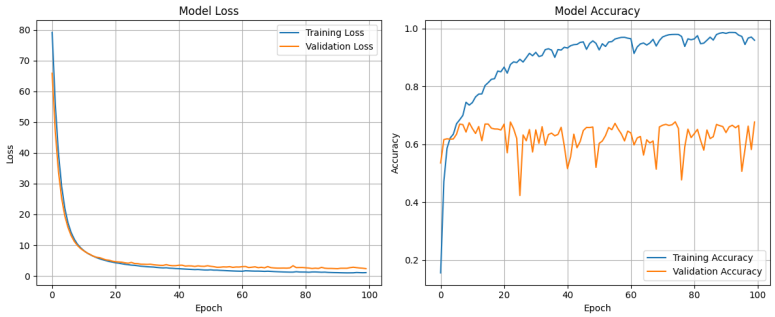
(a) Ground truth (left) and prediction (right)



(b) Loss (left) and accuracy (right) plots

Figure 5.22. Image displaying the ground truth, predictions and plots obtained with the ConvLSTM night model in Santa Olalla del Cala

(a) Ground truth (left) and prediction (right)



(b) Loss (left) and accuracy (right) plots

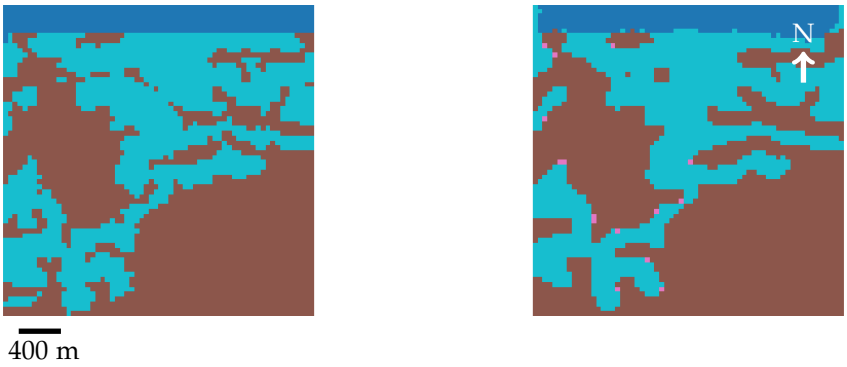
Figure 5.23. Image displaying the ground truth, predictions and plots obtained with the ConvLSTM night model in Villoslada de Cameros

5.4. Seasonal experimentation

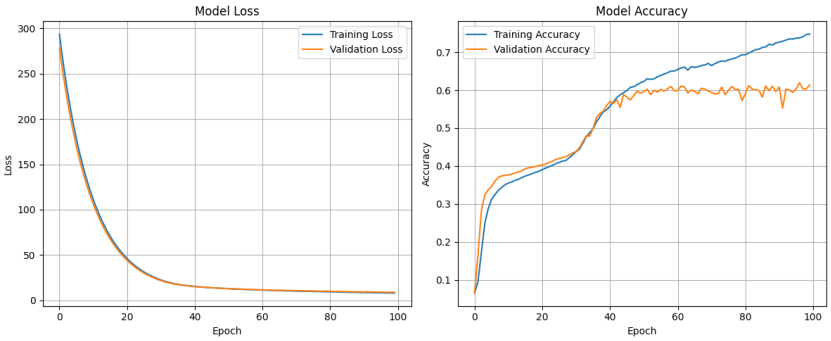
After evaluating the impact of diurnal changes in the performance of the models, the next step involved evaluating the influence of seasonal variations. To do so, input data were divided into winter and summer patches, where seasonal conditions like vegetation cover, soil moisture or thermal properties may have an impact in how the model handles these changes. The results of the classification performance of the models under such conditions are presented in the following section.

5.4.1. Winter

CNN



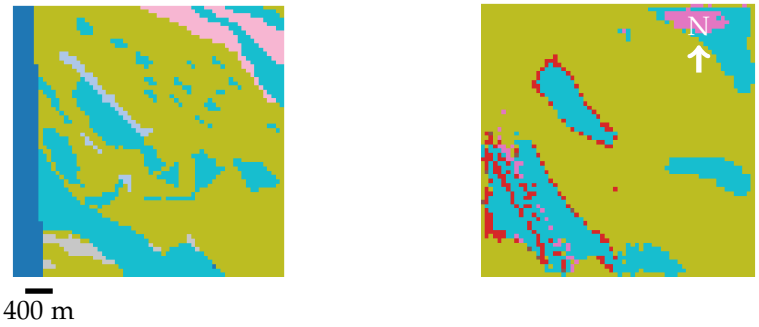
(a) Ground truth (left) and prediction (right)



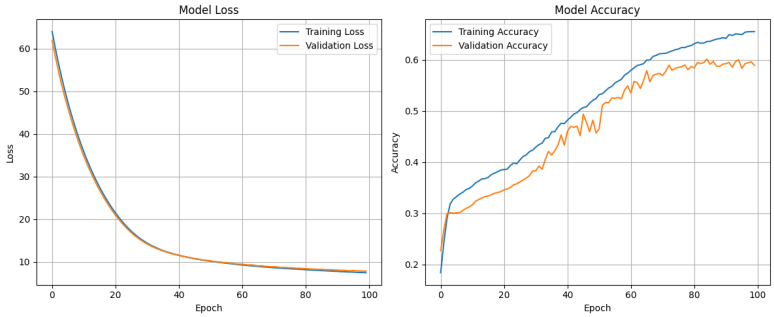
(b) Loss (left) and accuracy (right) plots

Figure 5.24. Image displaying the ground truth, predictions and plots obtained with the CNN winter model in Puertollano

The winter models managed to produce satisfactory results, at least is the case of both Puertollano (Figures 5.24 and 5.27) and Villoslada (Figures 5.26 and 5.29). Although some finer details are lost in the predictions, and once again, minority classes are not included, the terrain is represented with reasonable accuracy. In this case, both models seemed to struggle with the area of Santa Olalla (Figures 5.25 and 5.28, producing unwanted results in both cases. This discrepancy might be due to noisier thermal signals in winter in this area possibly due to its mountainous topography or vegetation cover. Instead, the remaining regions achieved better performance, possibly due to environmental conditions apparently more favourable for winter model predictions. Although the results are satisfactory in some cases (Table 6), generally, winter predictions were not very reliable, and produced a wide range of results, however the IoU of even the dominant classes would not be higher than 0.6. This inconsistent model performance indicates that the winter dataset is not very reliable due to its lack of details for rock and soil segmentation for most cases and should rather be used in combination with the summer dataset, as it will be explained in the following subsection.



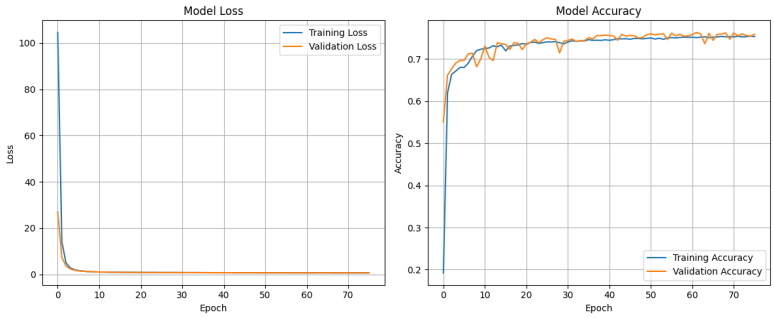
(a) Ground truth (left) and prediction (right)



(b) Loss (left) and accuracy (right) plots

Figure 5.25. Image displaying the ground truth, predictions and plots obtained with the CNN winter model in Santa Olalla del Cala

(a) Ground truth (left) and prediction (right)

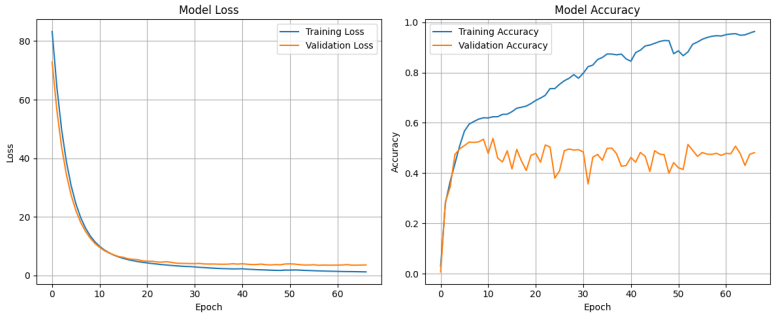


(b) Loss (left) and accuracy (right) plots

Figure 5.26. Image displaying the ground truth, predictions and plots obtained with the CNN winter model in Villoslada de Cameros

ConvLSTM

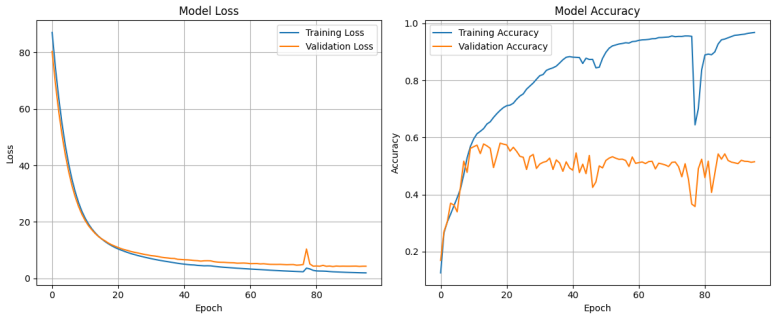
(a) Ground truth (left) and prediction (right)



(b) Loss (left) and accuracy (right) plots

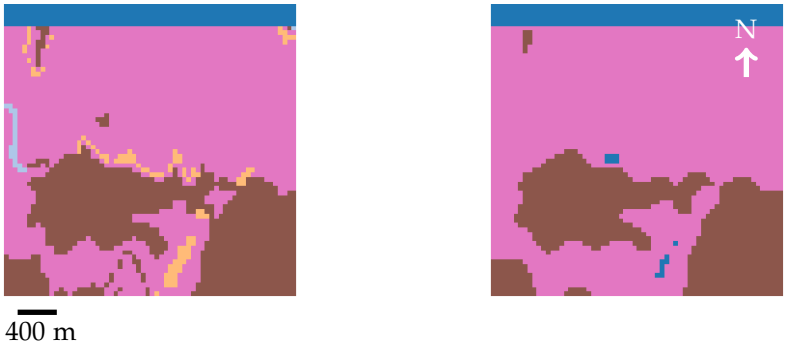
Figure 5.27. Image displaying the ground truth, predictions and plots obtained with the ConvLSTM winter model in Puertollano

(a) Ground truth (left) and prediction (right)

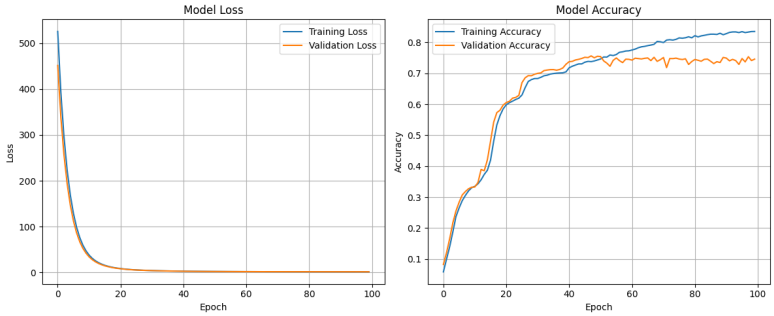


(b) Loss (left) and accuracy (right) plots

Figure 5.28. Image displaying the ground truth, predictions and plots obtained with the ConvLSTM winter model in Santa Olalla del Cala



(a) Ground truth (left) and prediction (right)



(b) Loss (left) and accuracy (right) plots

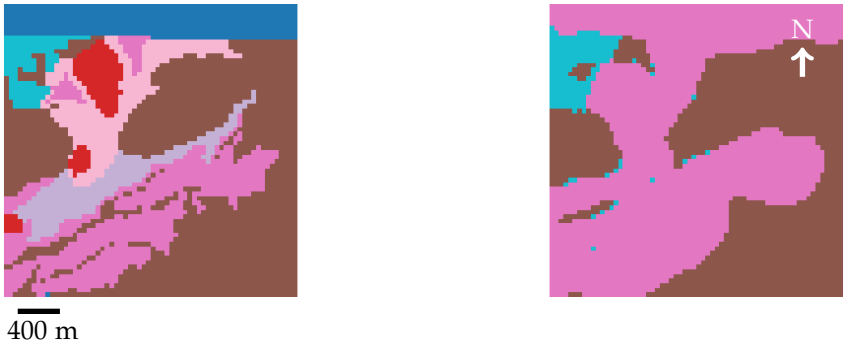
Figure 5.29. Image displaying the ground truth, predictions and plots obtained with the ConvLSTM winter model in Villoslada de Cameros

5.4.2. Summer

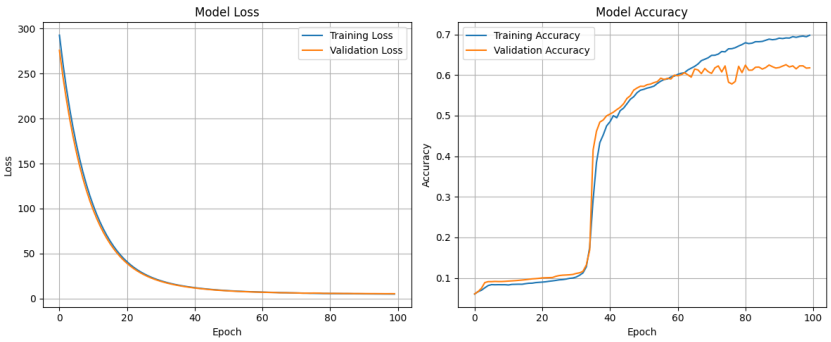
CNN

The results for the summer dataset are quite positive as reflected in Table 7. While the winter model produced some unwanted results, the summer dataset, in comparison, manages to generally produce substantially better results (Figures 5.30, 5.31, and 5.32). Both models manage to produce valid predictions, that are valuable—however once again, the CNN model struggles with class details. Minority classes again like silt or sand are barely represented if at all in the outputs. Therefore, indicating that no matter the temporal time or season, the CNN has a hard time outputting very detailed results.

Instead, the ConvLSTM model yet again demonstrated that it can consistently generate accurate predictions of the geological terrain (Figures 5.33, 5.34, and 5.35). Occasionally, it was able to detect and include minority classes like sand and clay. Nevertheless, with still some degree of imperfection and some clear overfitting. The noticeable improvement that summer patches present can be attributed to potentially clearer thermal data due to fewer atmospheric interferences taking place during the warmer months. Consequently, allowing for a higher thermal contrast and more stable thermal patterns. The lack of vegetation and the appearance of more bare soil likely enhances the model's performance, thus producing more reliable predictions than the winter dataset. Overall, the results of this dataset highlight once more the slight superiority of the ConvLSTM against the CNN, and how temporal dependencies manage to complement the spatial information available. Furthermore, the summer dataset demonstrated that thermal data were likely the most impactful one, as they enabled the model to match the results and training accuracy levels obtained with the full thermal dataset in both models. Thus, the model can still perform accordingly with only the summer months, which seem to provide more impactful data on their own than expected.

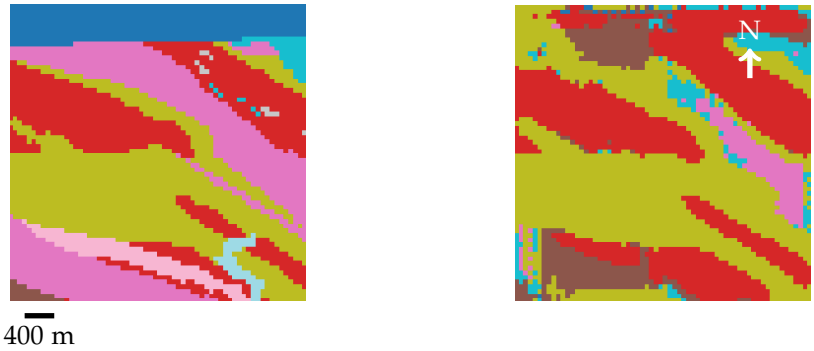


(a) Ground truth (left) and prediction (right)

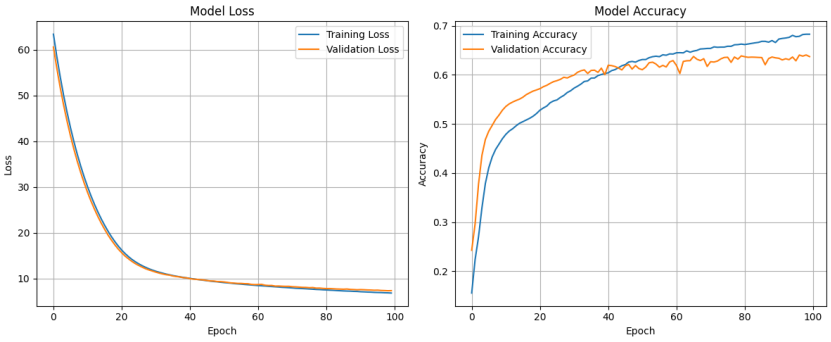


(b) Loss (left) and accuracy (right) plots

Figure 5.30. Image displaying the ground truth, predictions and plots obtained with the CNN summer model in Puertollano

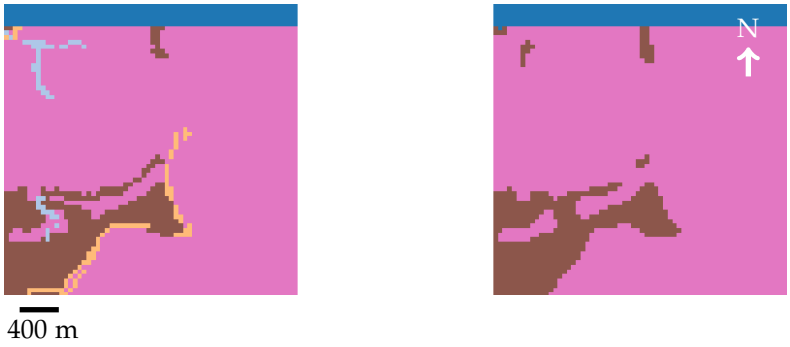


(a) Ground truth (left) and prediction (right)

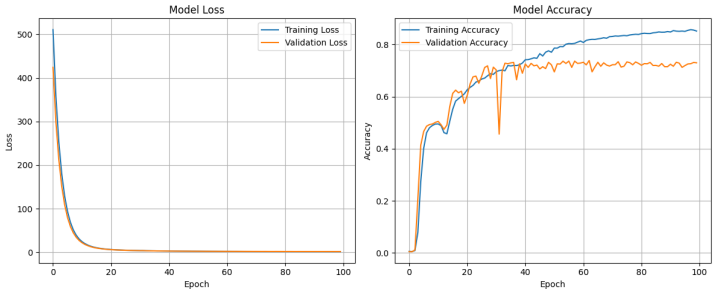


(b) Loss (left) and accuracy (right) plots

Figure 5.31. Image displaying the ground truth, predictions and plots obtained with the CNN summer model in Santa Olalla del Cala



(a) Ground truth (left) and prediction (right)

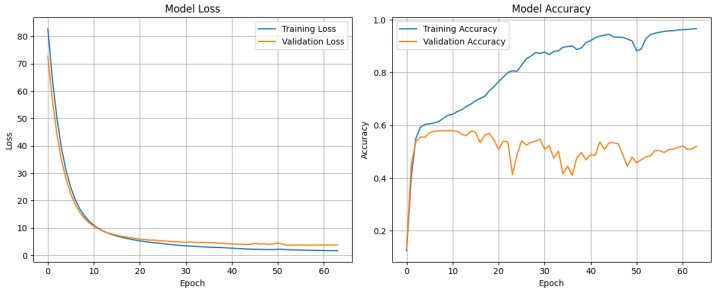


(b) Loss (left) and accuracy (right) plots

Figure 5.32. Image displaying the ground truth, predictions and plots obtained with the CNN summer model in Villoslada de Cameros

ConvLSTM

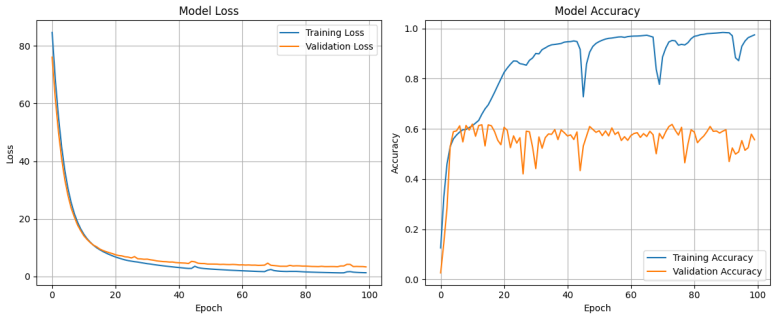
(a) Ground truth (left) and prediction (right)



(b) Loss (left) and accuracy (right) plots

Figure 5.33. Image displaying the ground truth, predictions and plots obtained with the ConvLSTM summer model in Puertollano

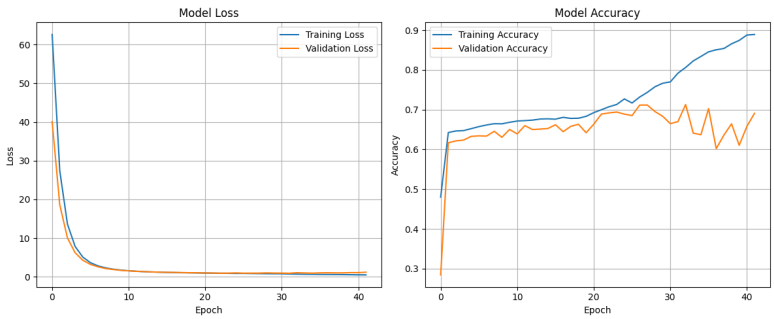
(a) Ground truth (left) and prediction (right)



(b) Loss (left) and accuracy (right) plots

Figure 5.34. Image displaying the ground truth, predictions and plots obtained with the ConvLSTM summer model in Santa Olalla del Cala

(a) Ground truth (left) and prediction (right)



(b) Loss (left) and accuracy (right) plots

Figure 5.35. Image displaying the ground truth, predictions and plots obtained with the ConvLSTM summer model in Villoslada de Cameros

5.5. SAR implementation

One of the main experiments of this study is the integration of auxiliary data to the thermal imagery. In this section, the integration of SAR data is evaluated. With the addition of the VV and VH channels, the model is expected to improve its performance or at least maintain its level of accuracy. Unfortunately, due to the limited number of matching SAR and thermal dates, the patch count is lower and had an impact on the results. Just as with the diurnal and seasonal experiments, the CNN and ConvLSTM are reported for the three sites, where, in order to compare any possible improvements, the models will be tested with SAR and thermal data in combination as well as only with the thermal imagery used for those patches.

5.5.1. SAR

Despite the reduced availability of SAR observations, the outputs produced by the model were encouraging. Although limitations exist—most notably a lack of finer details in some cases—the integration of this type of data revealed significant potential in enhancing the models' performance. The most promising effect of this implementation is the increased representation of minority classes, usually under-represented with thermal imagery.

The CNN model still showed some difficulties with certain categories such as silt in Puertollano (Figure 5.36) or sand in the case of Santa Olalla and Villoslada (Figures 5.37 and 5.38, respectively). However, it succeeded in representing other uncommon lithologies present such as volcanic rocks in Puertollano, a region where they are present but sparse. Interestingly, this same area exhibited a loss of foliated rocks, struggling to accurately represent a common class, foliated rocks, which disappears in the predictions. The reason behind may be in the texture of these rocks, since volcanic rocks are often more porous or contain cracks were water can be retained, it can provide higher VV and VH backscatter values. Instead, foliated rocks present, generally, a lower water content which might be leading to weaker radar responses.

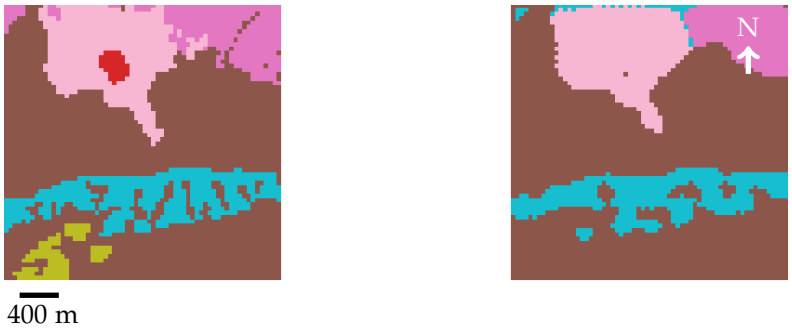
In Villoslada, the implementation of SAR influenced the representation of clay. This minority class became much more prominent and accurately represented, even to the point where, at times, it could become over-represented.This behaviour would align with the radar sensitivity to soil moisture, since clay retains water for longer periods of time, following the answers to the hypotheses, than other minority classes present in the area like sand or silt. As a result, it managed to enhance the model's ability to deal with classes associated to stable moisture content. SAR ended up reducing overfitting in the training process, especially for accuracy values, although it did introduce some more fluctuation in the validating accuracy towards the end of the process, probably due to the increased complexity of the data.

To determine the impact of SAR data, an additional expeiment was carried out employing only the thermal channels from the same set of patches. This way, a fair comparison could be made under equal data conditions (dates, thermal values, cloud coverage, etc.). By removing the VV and VH channels from the input data, the models can train based solely on thermal data, though this time with a lower patch count. Nevertheless, the results were of comparable quality. They confirmed that models combining thermal and SAR inputs generally achieved slightly higher accuracy values, nevertheless the discrepancies are not that significant. In the case of Puertollano and Santa Olalla (Figures 5.42 and 5.43, respectively), the predictions do underperform when compared to the SAR experiments. But for Villoslada (Figure 5.44), this is not the case, with some visible improvements present in the appearance of sand in the prediction (grey-blue) and a slightly better representation of detritic rocks (brown). This suggested that the added temporal coherence of thermal data partially compensated the lack of radar information. Nevertheless, this was usually not the case for most predictions involving only the thermal data.

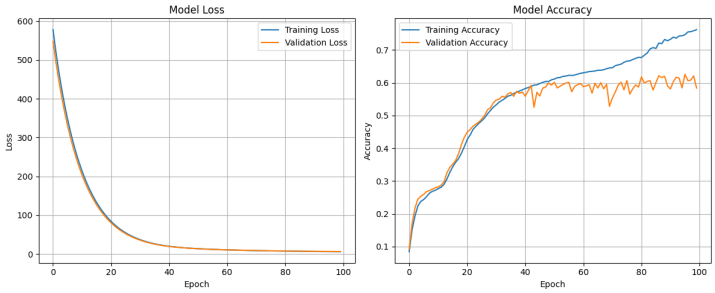
Regarding the ConvLSTM model, it also benefited from the inclusion of SAR channels. Even though spatial representation of the patches became, in occasions, less precise, it actually managed to generally capture minority classes better. As observed in Figures 5.39 and 5.45, in Puertollano, the model managed to represent silt (red) and loam (grey-purple) in the second image, in this case possibly due to their strong moisture signals. Unfortunately, it did include volcanic and non-foliated rocks into the prediction, which are out of context in that patch. Therefore, with a limited number of thermal

images available, the model struggles to find a balance with minority classes. The introduction of miss-classifcations like these indicate the sensitivity of the model to data imbalance and limited thermal availability. A similar case occurred in Santa Olalla del Cala (Figures 5.40 and 5.46), were the dominant classes got slightly under-represented. In the case of Villoslada (Figures 5.41 and 5.47), the model did struggle to produce a valuable outcome for both cases.

CNN

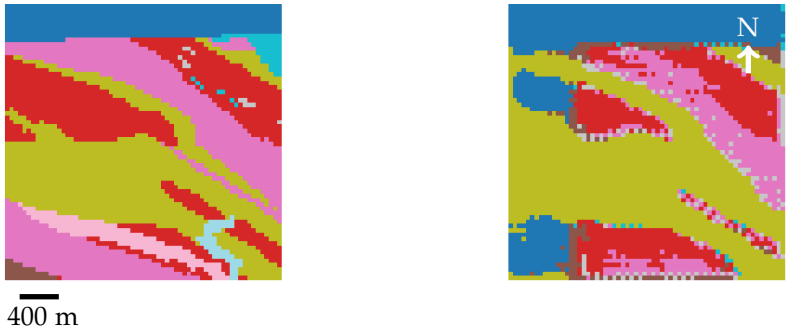


(a) Ground truth (left) and prediction (right)

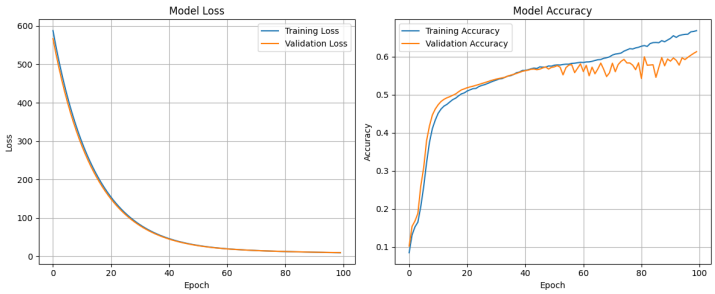


(b) Loss (left) and accuracy (right) plots

Figure 5.36. Image displaying the ground truth, predictions and plots obtained with the CNN SAR model in Puertollano



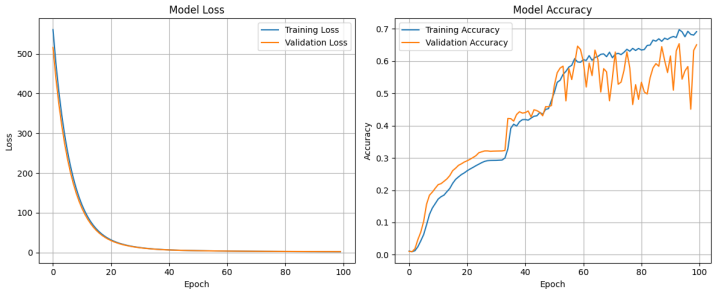
(a) Ground truth (left) and prediction (right)



(b) Loss (left) and accuracy (right) plots

Figure 5.37. Image displaying the ground truth, predictions and plots obtained with the CNN SAR model in Santa Olalla del Cala

(a) Ground truth (left) and prediction (right)

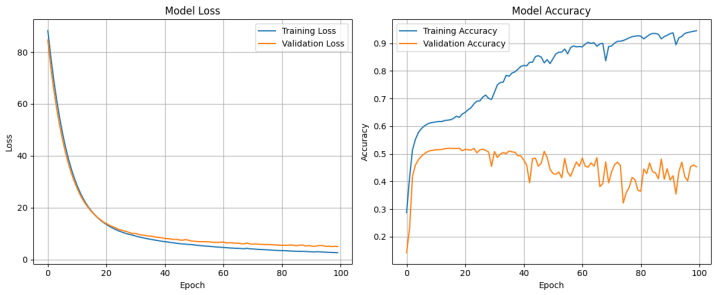


(b) Loss (left) and accuracy (right) plots

Figure 5.38. Image displaying the ground truth, predictions and plots obtained with the CNN SAR model in Villoslada de Cameros

ConvLSTM

(a) Ground truth (left) and prediction (right)

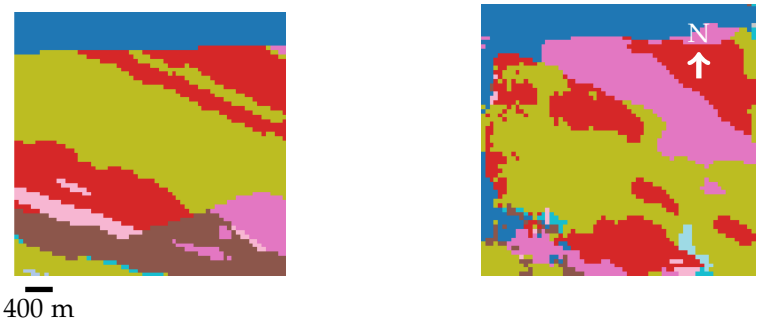


(b) Loss (left) and accuracy (right) plots

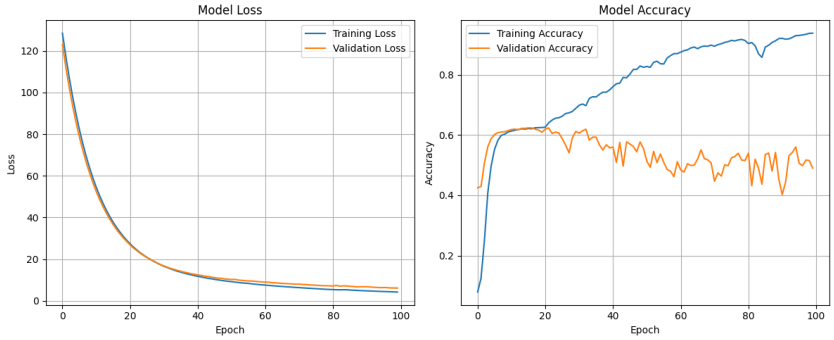
Figure 5.39. Image displaying the ground truth, predictions and plots obtained with the ConvLSTM SAR model in Puertollano

Overall, considering the constraint of a limited dataset, the performance of the models remains promising, indicating that SAR data, initially, do actually contribute positively to an enhanced predictive process. SAR appears to enhance robustness, reduce overfitting tendencies and provide valuable physical complementary data to the thermal information. Whereas ECOSTRESS thermal imagery presents sensitivity towards variations in surface temperature, heat capacity and emissivity, SAR responds to surface roughness and moisture content through its VV and VH channels. This combination improves the predictive process by integrating thermal dynamics, dielectric and structural properties of soils and rocks. If a larger number of images were to be available, the potential impact of the radar data could become even more substantial. With no sensitivity to cloud coverage or illumination, it can offer temporal continuity and density, becoming a great complementary tool.

Overall, the inclusion of SAR leads to slight improvements over thermal-only inputs in several predictions, showing great potential to support thermal-based predictions with meaningful information, especially for minority classes. Although in the case of this study it is constrained by data availability, the findings suggest that improved temporal alignment and sufficient data volume could enhance the differentiation of minority classes, reduce class imbalance as a result, and improve generalisation of deep learning models across complex terrains.



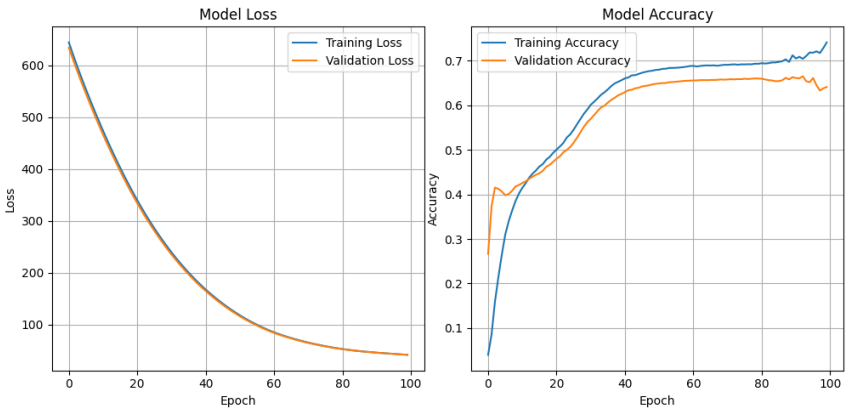
(a) Ground truth (left) and prediction (right)



(b) Loss (left) and accuracy (right) plots

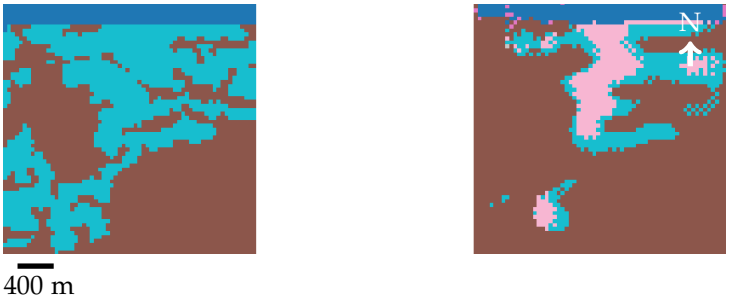
Figure 5.40. Image displaying the ground truth, predictions and plots obtained with the ConvLSTM SAR model in Santa Olalla del Cala

(a) Ground truth (left) and prediction (right)

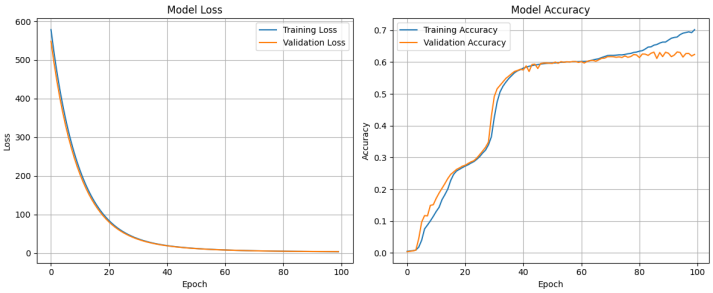


(b) Loss (left) and accuracy (right) plots

Figure 5.41. Image displaying the ground truth, predictions and plots obtained with the ConvLSTM SAR model in Villoslada de Cameros

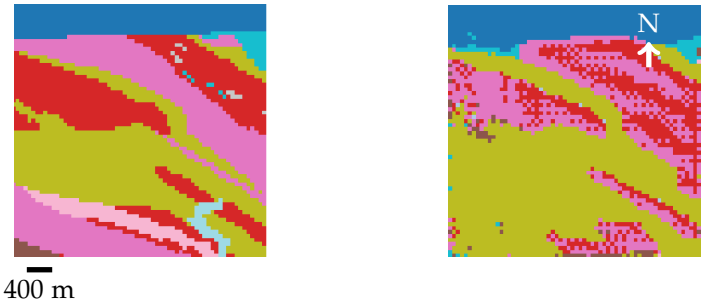


(a) Ground truth (left) and prediction (right)

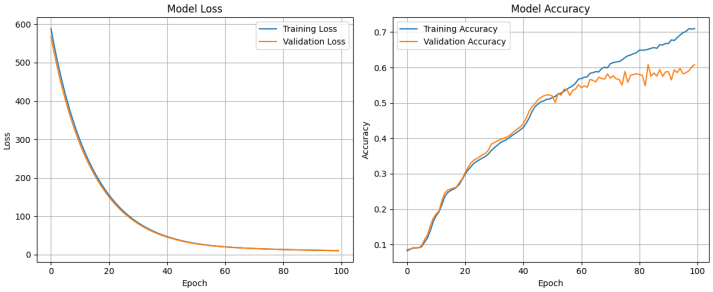


(b) Loss (left) and accuracy (right) plots

Figure 5.42. Image displaying the ground truth, predictions and plots obtained with the CNN SAR thermal-only model in Puertollano



(a) Ground truth (left) and prediction (right)

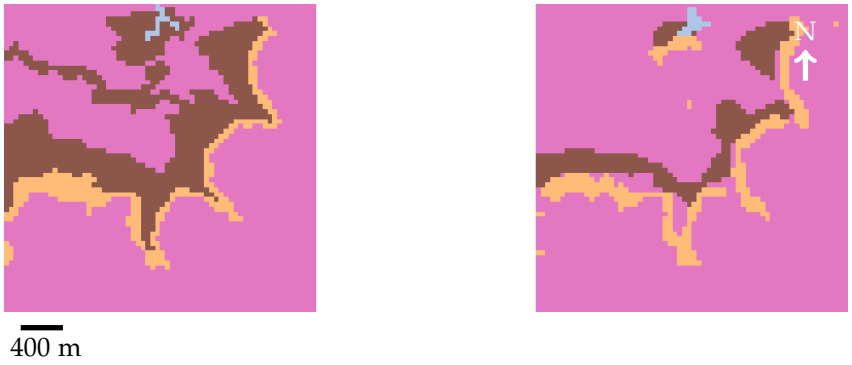


(b) Loss (left) and accuracy (right) plots

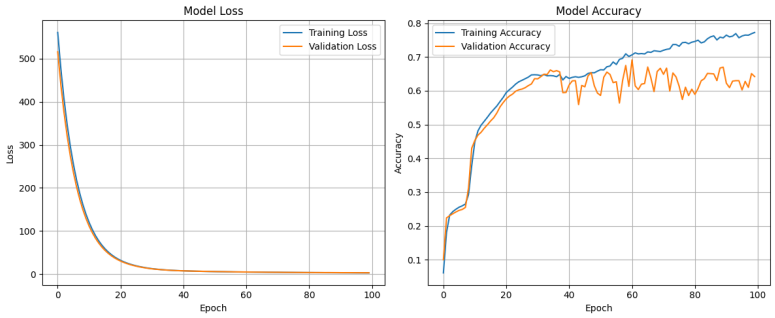
Figure 5.43. Image displaying the ground truth, predictions and plots obtained with the CNN SAR thermal-only model in Santa Olalla del Cala

5.5.2. Only thermal SAR dates

CNN



(a) Ground truth (left) and prediction (right)

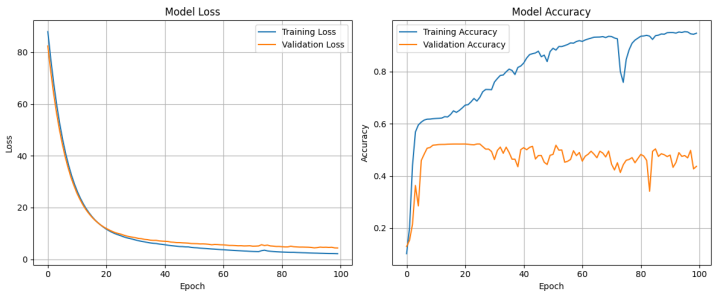


(b) Loss (left) and accuracy (right) plots

Figure 5.44. Image displaying the ground truth, predictions and plots obtained with the CNN SAR thermal-only model in Villoslada de Cameros

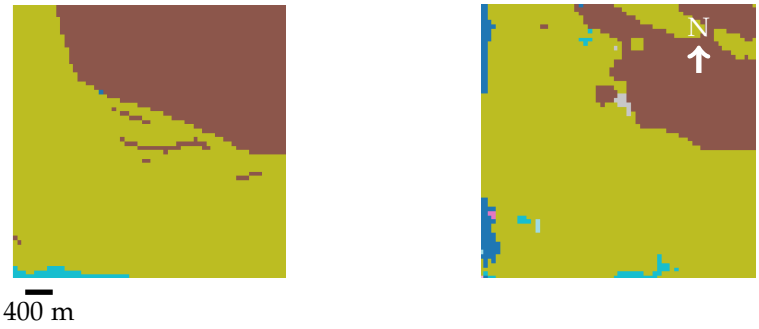
ConvLSTM

(a) Ground truth (left) and prediction (right)

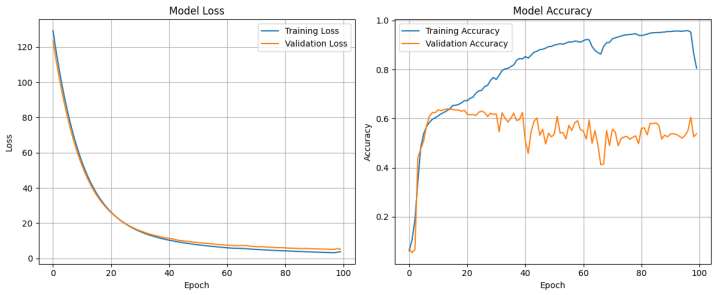


(b) Loss (left) and accuracy (right) plots

Figure 5.45. Image displaying the ground truth, predictions and plots obtained with the ConvLSTM SAR thermal-only model in Puertollano



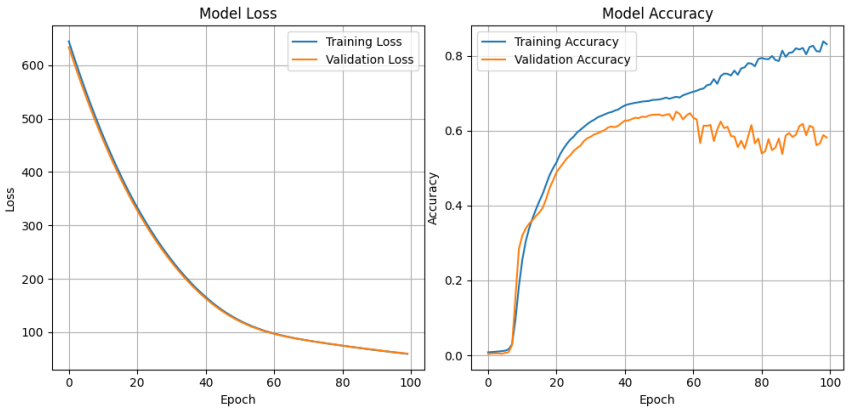
(a) Ground truth (left) and prediction (right)



(b) Loss (left) and accuracy (right) plots

Figure 5.46. Image displaying the ground truth, predictions and plots obtained with the ConvLSTM SAR thermal-only model in Santa Olalla del Cala

(a) Ground truth (left) and prediction (right)



(b) Loss (left) and accuracy (right) plots

Figure 5.47. Image displaying the ground truth, predictions and plots obtained with the ConvLSTM SAR thermal-only model in Villoslada de Cameros

5.6. Results tables

This section displays the results obtained for the accuracy and F1 score metrics across all tested models, regions, and data modalities. To ease interpretation and analysis, the results are presented with colour-coded matrices where performance levels are visually distinguished: green cells indicate high performance (>0.75), yellow cells represent moderate performance (0.60-0.75), and red cells highlight underwhelming performance (<0.60). ConvLSTM models generally outperform CNNs, while performance varies considerably across regions and when temporal sequences are constrained to daytime or nighttime or when using alternative data sources SAR. The Santa Olalla region demonstrates the highest performance with thermal data and ConvLSTM architecture, suggesting that this combination of geographical and temporal characteristics as well as the soil type diversity is particularly effective. Although Santa Olalla presents the highest single performance, it is Villoslada which consistently displays the highest overall values across most experiments, indicating that the geographical and climatic context of the area might be more beneficial for the models overall.

Table 3. Accuracy and F1 scores for Thermal data type

Region / Model	Accuracy	F1 score
Puertollano CNN	0.7128	0.6739
Puertollano ConvLSTM	0.7090	0.7299
Santa Olalla CNN	0.7443	0.6694
Santa Olalla ConvLSTM	0.8106	0.7881
Villoslada CNN	0.7576	0.5156
Villoslada ConvLSTM	0.7759	0.6952

Table 4. Accuracy and F1 scores for Day data type

Region / Model	Accuracy	F1 score
Puertollano CNN	0.6532	0.4580
Puertollano ConvLSTM	0.5804	0.5811
Santa Olalla CNN	0.5793	0.3915
Santa Olalla ConvLSTM	0.7691	0.7607
Villoslada CNN	0.7280	0.4102
Villoslada ConvLSTM	0.7424	0.6871

Table 5. Accuracy and F1 scores for Night data type

Region / Model	Accuracy	F1 score
Puertollano CNN	0.6587	0.5510
Puertollano ConvLSTM	0.6255	0.4029
Santa Olalla CNN	0.5918	0.2396
Santa Olalla ConvLSTM	0.5548	0.3538
Villoslada CNN	0.7206	0.5943
Villoslada ConvLSTM	0.7185	0.4064

Table 6. Accuracy and F1 scores for Winter data type

Region / Model	Accuracy	F1 score
Puertollano CNN	0.6820	0.6004
Puertollano ConvLSTM	0.6786	0.3512
Santa Olalla CNN	0.6203	0.4504
Santa Olalla ConvLSTM	0.5748	0.3054
Villoslada CNN	0.7417	0.6159
Villoslada ConvLSTM	0.7347	0.6021

Table 7. Accuracy and F1 scores for Summer data type

Region / Model	Accuracy	F1 score
Puertollano CNN	0.6873	0.5709
Puertollano ConvLSTM	0.7337	0.5652
Santa Olalla CNN	0.6400	0.5945
Santa Olalla ConvLSTM	0.8016	0.7761
Villoslada CNN	0.7407	0.5975
Villoslada ConvLSTM	0.7130	0.6886

Table 8. Accuracy and F1 scores for SAR data type

Region / Model	Accuracy	F1 score
Puertollano CNN	0.6472	0.4429
Puertollano ConvLSTM	0.5814	0.4699
Santa Olalla CNN	0.5701	0.4668
Santa Olalla ConvLSTM	0.5072	0.3455
Villoslada CNN	0.7010	0.5901
Villoslada ConvLSTM	0.6319	0.4355

Table 9. Accuracy and F1 scores for SAR (only Thermal) data type

Region / Model	Accuracy	F1 score
Puertollano CNN	0.4320	0.2011
Puertollano ConvLSTM	0.5570	0.4512
Santa Olalla CNN	0.6478	0.4678
Santa Olalla ConvLSTM	0.6179	0.3442
Villoslada CNN	0.7310	0.6302
Villoslada ConvLSTM	0.5793	0.5839

5.7. Further Experiments

Following the experiments presented in Chapter 5 that shape the core of this thesis, additional informative tests were proposed and subsequently conducted to further explore the performance of the models under different conditions. These supplementary experiments were designed to assess the flexibility and adaptability of the proposed methodology, demonstrating how it can be refined and extended to achieve improved classification through modified parameters and data configurations that aimed to overcome some of the limitations observed previously. However, these experiments were carried out mainly on the whole thermal dataset, leaving tests with SAR, diurnal and seasonal patches as tasks to look into in the future. The approach ensured the use of all available thermal data, avoiding temporal alignment constraints. Overall, these extended analyses strengthen the robustness of the framework, highlighting its potential for further enhancement in future work.

During this phase, many different parameters and configurations were tested to enhance the performance of the models. The aim of these additional tests was to: (1) reduce the overfitting problems consistently presented throughout the ConvLSTM results, as well as in the CNN model at times, (2) enhance the visualisation process from single patches to completely reconstructed geological map predictions for direct comparisons, and (3) perform an NDVI analysis on the results to determine how the presence of vegetation impacts the confidence with which the models predict their final outputs. Furthermore, these objectives aim to improve the depth of the analysis in this study, providing a deeper overview of the many factors that can affect the results and in which way they do so. Therefore, the following sections delve into the reconstruction of the original labels, the modifications made after the initial experiments, the impact these had on the performance of both models and the influence of vegetation through a new type of complementary data (NDVI).

5.7.1. Reconstructed geology maps

In order to ease the comparison process of ground truths and prediction patches, a full reconstruction pipeline can be developed. The best approach consists on reconstructing the entire original geological maps that were used as labels before being split into patches. The process of reconstructing the geology maps takes place in several steps involving the process mentioned in 3.4.4 through a systemic reassembly of prediction patches. The reconstruction works on both models, being adapted to time sequences and individual patches for each case, as well as working with both overlapping and non-overlapping patches.

Initially, all patches are segmented by their acquisition date, to avoid overlapping spatial coverage and to ensure that each map is reconstructed separately for each date, since each one should represent an independent snapshot. The metadata with the positions produced previously during the preprocessing stage contains all the spatial coordinates for each patch, which are used to calculate the map size that the patches should reconstruct, depending on the area where the testing is being carried out. The maps are then reconstructed with these coordinates and full map dimensions, taking into account the patch size. To handle overlapping patches, cumulative averaging is employed, since it can preserve the predictions of all patches, reduce noise by filtering out single bad predictions, handle several overlapping degrees—in this case up to 50%—and avoid arbitrary choices on selected a "winning" patch. Instead, if patches do not overlap, placement is direct. Of course, in the case of confidence maps, discussed in Section 5.7.3, the averaging process can result in lower confidences being outputted in the final reconstructed map. Reconstructed geological maps are performed for both the ground truths and the predictions, resulting in a streamlined process that allows for a direct comparison.

For all three areas, the geological maps were reconstructed and tested for visualisation purposes, which can be observed in the following section (Section 5.7.2) when comparing the outputs to the labels. The blue areas on the outside of the maps—both in the ground truths and prediction outputs—correspond to NaN values produced while resampling the geological maps to thermal imagery, since the Coordinate Reference System (CRS) and resolution were different, which was adjusted by filling the spaces with NaN values.

5.7.2. Final results

Throughout this additional phase of the study, many attempts were made at improving the performance of the models. The current outputs after the core experimentations were successful and valuable. However, in some instances they lacked enough accuracy to become consistently reliable. With overfitting being the main problem at the moment, the approach moving forward was to adapt the models to enhance their training phase. The training accuracy of the models was often quite promising, especially in the thermal datasets with values reaching between 0.8-0.9 for thermal-only datasets, day and summer and even SAR. It demonstrated the real potential of the models if the validating accuracy could match the training. This problem indicated that, for part of the training, the model was simply memorizing the patterns—and not properly learning from them—, which made it challenging to produce actual predictions during the output phase when different thermal textures or patterns were present. To fix this, a wide variety of parameters and configurations were tuned, although not all of them had a positive impact. The results of the combination of all these parameters can be better observed in Figures 5.53, 5.55, 5.57, 5.49, 5.50 and 5.51.

Firstly, the two most impactful parameters from Section 4 were tested again. Throughout the whole testing period, the learning rate slightly changed for the ConvLSTM model. The previous learning rate—which was the same as the CNN—was sufficient but could become too slow at times in such architecture, and was slightly reduced to 0.0001 to obtain faster computing times and final overall loss function. L2 was the following parameter in which new runs still demonstrated that a Ridge Regression of 0.05 value reduced the overfitting significantly, yet lower values than this would produce underwhelming results with worse predicted results just as established in Section 4. Nevertheless, these changes became even more relevant when the loss function was enhanced. The prior loss function employed throughout the experiments reflected in Section 5.2 was *Keras*' built in loss function, since it managed to produce valuable results and the previous custom loss function was not working as expected. A new parameter was considered into the customized loss function to complement the class weights. Even though the predictions reflected a similar distribution to that of the labels, they lacked detail and minority class representation. To tackle this, focal loss was included into the weighted, customized loss function to tackle the harder cases, since majority classes already had very reliable representations. This way, the model could pay more attention to detail in many areas and better represent the most uncommon classes. Focal loss had one of the highest impacts on the improvement of the results and managed to improve the overfitting issue. Moreover, the F1 score had the highest increase in value in comparison to the initial results obtained, increasing from a range of 0.5-0.7 up to a consistent value range of 0.7-0.85. As for accuracy, something similar occurred, with most values surpassing previous results by $\sim$0.2. Although it was not systematically tested on all datasets, the success presented in Table 10—along with the rest of the changes presented in this section—highlight the potential of adaptative, weighted loss for this study. The integration of focal loss in highly imbalanced geological datasets together with customized class weights for each region presents a promising advancement for the usability of these models. These parameters had the biggest impact on the overfitting, and although overfitting is still present in some cases, these parameters and the new focal loss managed to handle this problem quite effectively.

The second attempt focused on increasing the level of detail the predictions could achieve. To do so, different options were employed. After reconstructing the geological maps the visualisation was not always as expected. Though the models managed to produce successful predictions, the connectivity between consecutive patches was abrupt and, in occasions, random due to locally miss-predicted patches as observed in Figure 5.48. To solve both these problems—spatial relationships and poor levels of detail—several approaches were taken. The first one involved increasing the patch size up to double the size of the original ones used (128x128), which resulted in similar results, since the neighbouring patches would still present some sudden changes, lacking the continuity expected in the results. Therefore, the patch size had little effect on the model's performance and another approach had to be tested. A simple and common way to improve the spatial relations of neighbouring pixels is to overlap them, hence, an overlap of up to 50% was applied to the patches. This way, the model would not learn to detect which areas correspond to boundaries and instead focus more on learning those spatial relations between different classes and patches. Now that some pixels located on prior patch boundaries became patch centres, the model managed to identify the spatial distribution of many classes better. Although it only slightly improved overfitting, the predictions became much better, with more inter-connected patches. Nevertheless, since the overfitting issue is still present,

there are still some random patches or sudden breaks in a class representation due to the original patch boundaries keep appearing in the outputs (as seen in Figure 5.56b). In addition to this increase in detail in the reconstructed geological map outputs, the epoch count was increased, to allow more time for the models to learn these newer patterns and spatial relationships they did not have to employ before. The epoch count was increased by 33% (up to 150 epochs) with early stopping becoming more flexible since the run was potentially longer. This way, the model would have more time to improve its training, which resulted in better overall values and thermal pattern recognition.

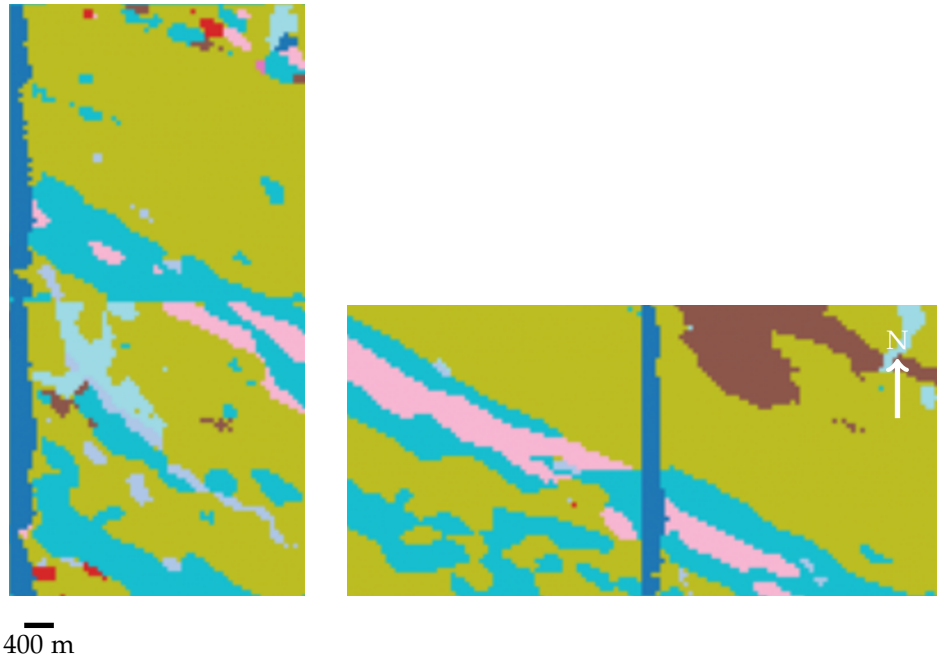


Figure 5.48. Image displaying the spatial connectivity issues between neighbouring patches during map reconstruction

Table 10. Comparison of accuracy and F1 score results for thermal-only datasets under initial and final parameter configuration

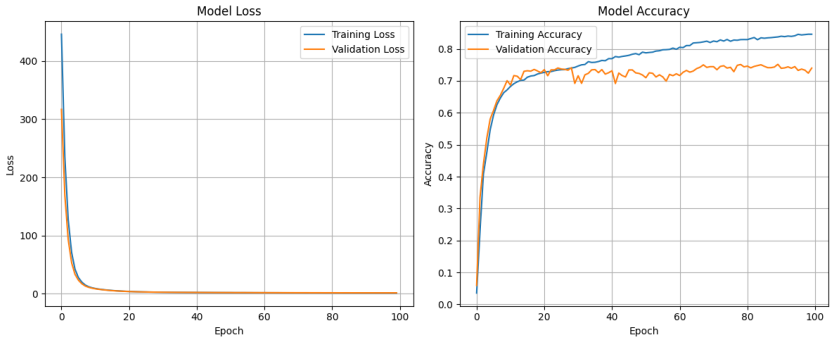
Accuracy and F1 scores of initial Thermal data tests

Region / Model	Accuracy	F1 score
Puertollano CNN	0.7128	0.6739
Puertollano ConvLSTM	0.7090	0.7299
Santa Olalla CNN	0.7443	0.6694
Santa Olalla ConvLSTM	0.8106	0.7881
Villoslada CNN	0.7576	0.5156
Villoslada ConvLSTM	0.7759	0.6952

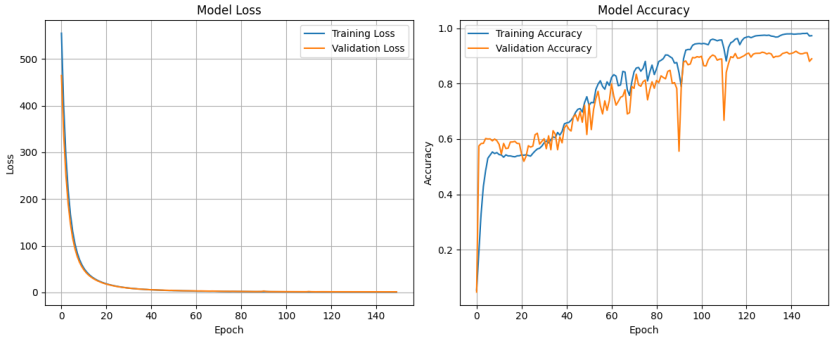
Accuracy and F1 score for final Thermal data tests

Region / Model	Accuracy	F1 score
Puertollano CNN	0.8220	0.7148
Puertollano ConvLSTM	0.9632	0.8499
Santa Olalla CNN	0.8673	0.6093
Santa Olalla ConvLSTM	0.9577	0.8018
Villoslada CNN	0.9252	0.7522
Villoslada ConvLSTM	0.9210	0.7926

The results in Table 10 display the improvements in the predictions of all areas for both models. Villoslada and Puertollano became the most profited areas from these changes, with the latter one achieving the highest accuracy mark. In the case of Puertollano, in both models, the improvements are quite important, rising by more than 1.0 and 1.5 in accuracy in its CNN and ConvLSTM models, respectively. However, the most promising part is the rise in the F1 Score for all regions—except Santa Olalla's CNN—, since it indicates that the representation of other classes has also improved and, therefore, minority classes have been predicted more often and more accurately in the new results. The rise in both accuracy and F1 score can be observed in Figures 5.53, 5.55, and 5.57, where clay, sand, and water become much more represented than before in Santa Olalla, while clay and foliated rocks are now represented in Puertollano despite their very limited presence in the area.

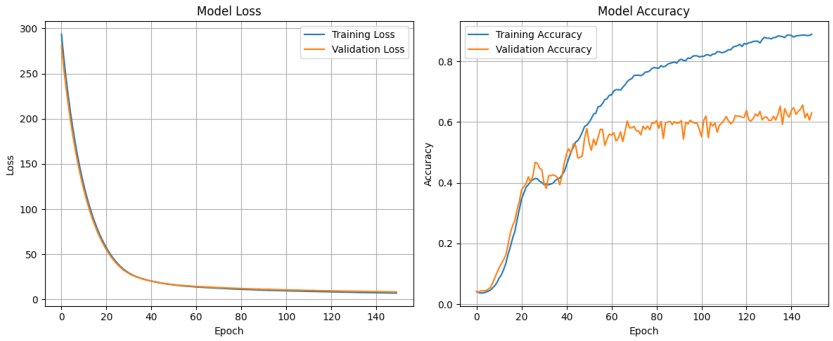


(a) Loss and accuracy plots of Puertollano's CNN

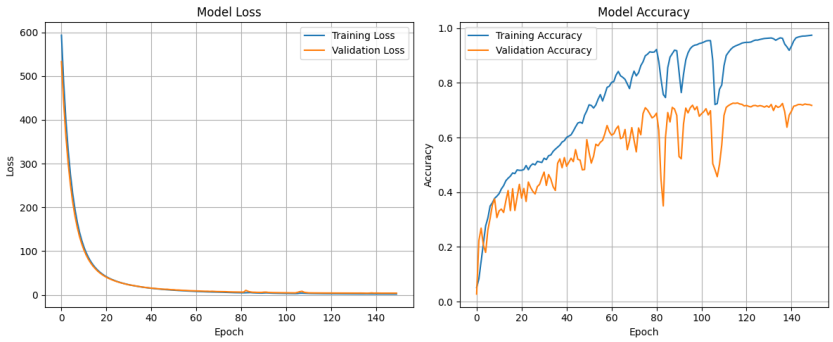


(b) Loss and accuracy plots of Puertollano's ConvLSTM

Figure 5.49. Loss (left) and accuracy (right) plots corresponding to the Puertollano area in CNN and ConvLSTM models

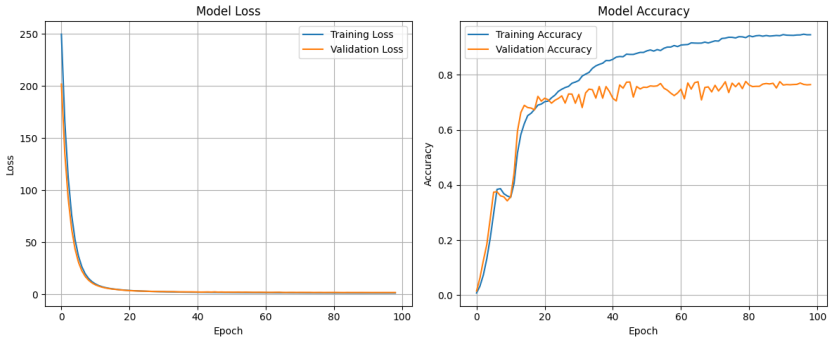


(a) Loss and accuracy plots of Santa Olalla's CNN

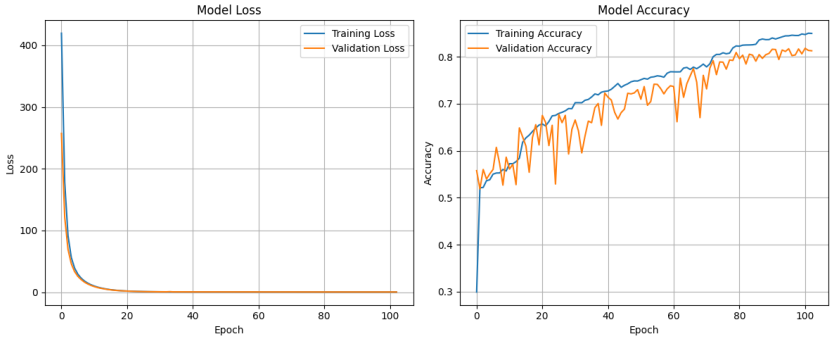


(b) Loss and accuracy plots of Santa Olalla's ConvLSTM

Figure 5.50. Loss (left) and accuracy (right) plots corresponding to the Santa Olalla area in CNN and ConvLSTM models



(a) Loss and accuracy plots of Villoslada de Cameros' CNN

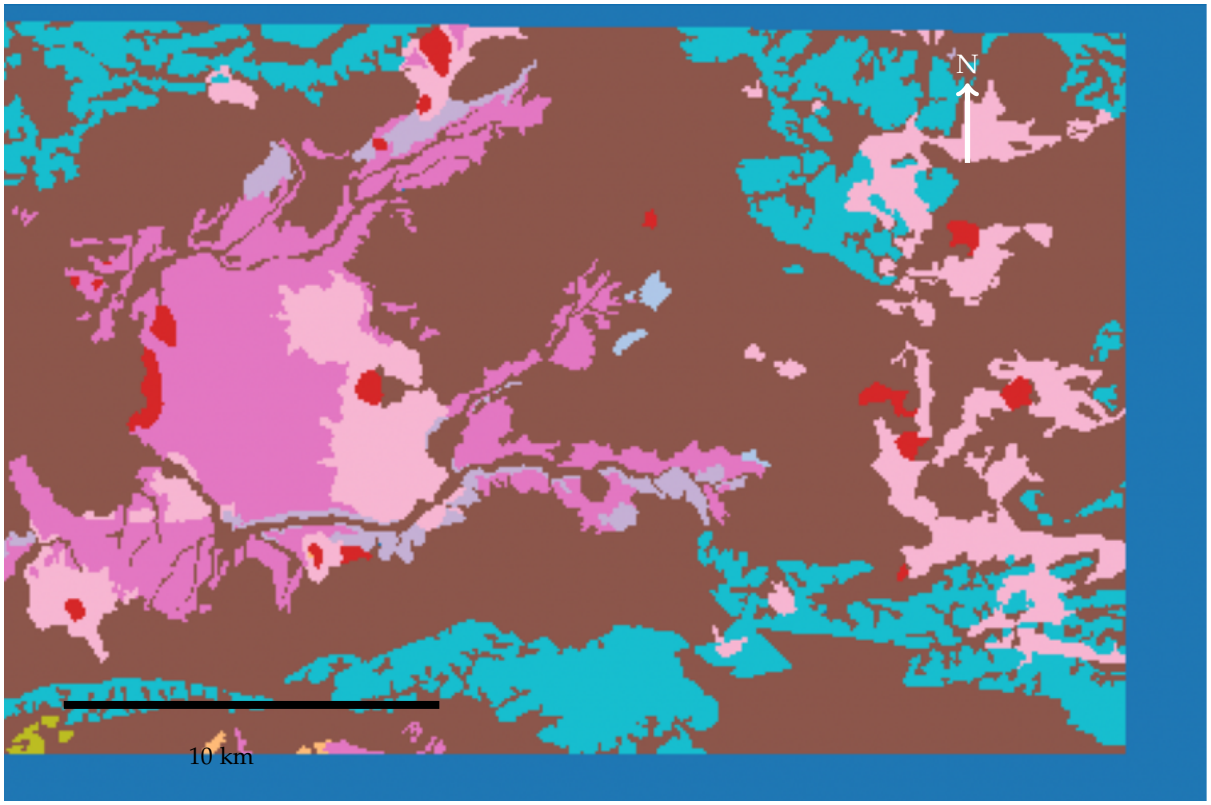


(b) Loss and accuracy plots of Villoslada de Cameros' ConvLSTM

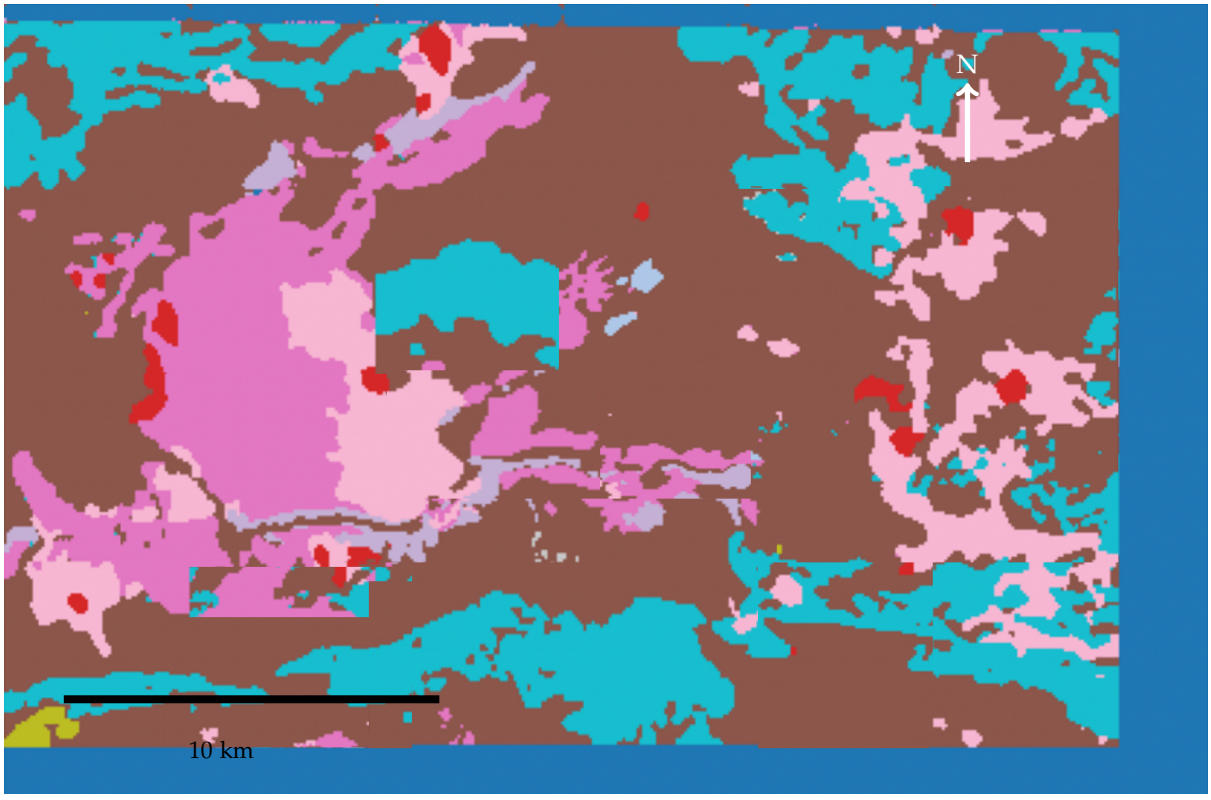
Figure 5.51. Loss (left) and accuracy (right) plots corresponding to the Villoslada de Cameros area in CNN and ConvLSTM models

The training dynamics from the plots above—visible in Figures 5.52b, 5.53a, 5.55a, 5.55b, 5.56b and 5.57a—reveal fundamental architectural differences between the two models. ConvLSTM achieved substantially lower validation losses (less than 1.0) in Puertollano and Villoslada, indicating a superior learning ability rather than memorization. In contrast, CNN exhibited more consistent overfitting across all regions. This divergence stems from the LSTM cell memory architecture in the ConvLSTM. The purely spatial CNN's architecture lacks the temporal constraint, and since it does not account for spatial relations, it still remains vulnerable to patch-level overfitting in regions with class imbalance. Although overlapping helped CNN make reliable improvements, the ConvLSTM model exploited this redundancy by penalizing temporal inconsistencies—such as abrupt class changes between patches—explaining the superior overall ability of ConvLSTM for rock and soil segmentation.

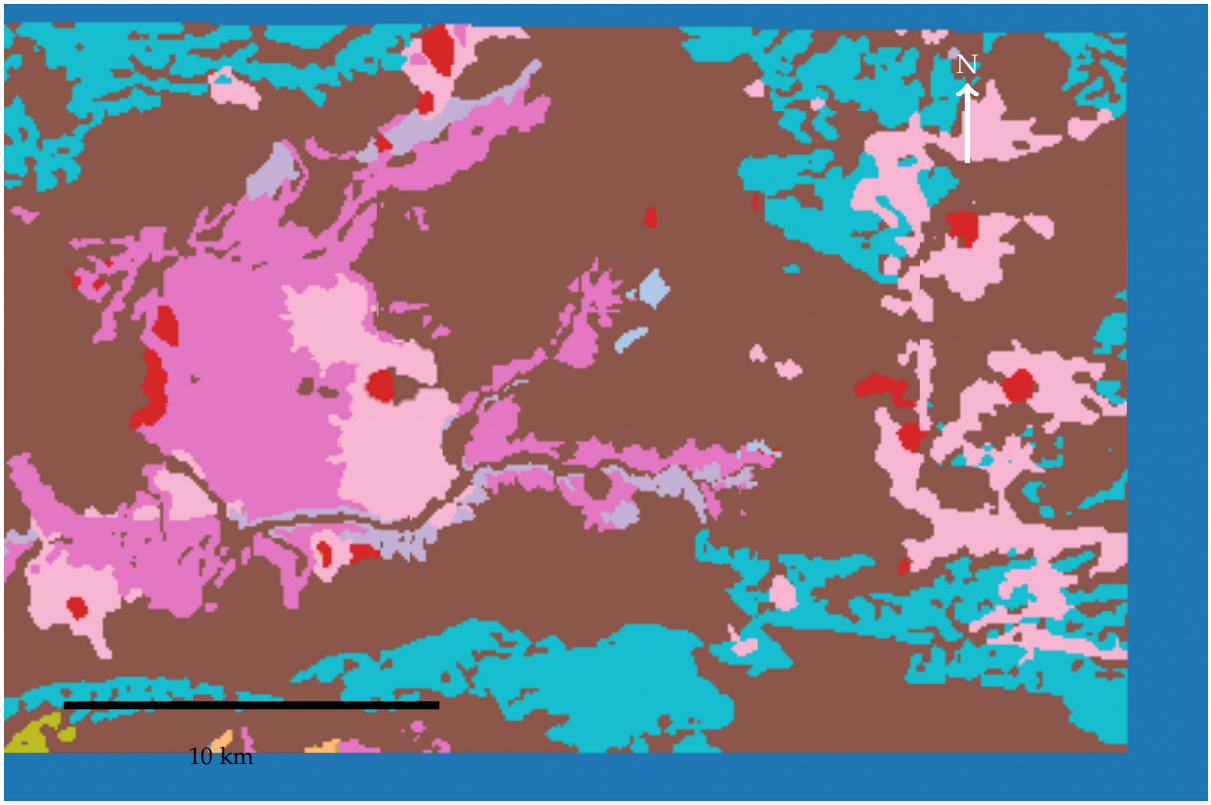
The reconstructed full-scene predictions demonstrate that, in general, ConvLSTM consistently outperforms CNN, with mean accuracy improvements of ∼10% and better minority class representation—observe silt (red) not being represented in the final output in Figure 5.56b or sand (blue/grey) in Figure 5.55a at the top left of the image. While CNN still had reasonable performance (∼0.8 accuracy), given its spatial-only design, it exhibits limitations such as incomplete class representation, patch continuation with abrupt changes at patch boundaries and random output patch positioning (visibly noticeable in Figure 5.52b). These failures are produced from non-uniform patch sampling during the pre-processing stage, since the patches of the same location might become less represented in the training data due to the quality filtering process. Furthermore, CNN tends to produce some occasional class saturation, over-representing both dominant and minority units. Although ConvLSTM also suffers similar issues, its temporal memory mechanisms manage to mitigate these effects. Despite these limitations, CNN still remains a viable baseline for single-date segmentation. However, ConvLSTM still demonstrates superior robustness, reduced boundary artifacts, and improved minority class representation—as is the case of silt when comparing Figures 5.56b and 5.57a—making it the preferred approach for rock and soil classification in this case.



(a) Ground-truth label map of Puertollano

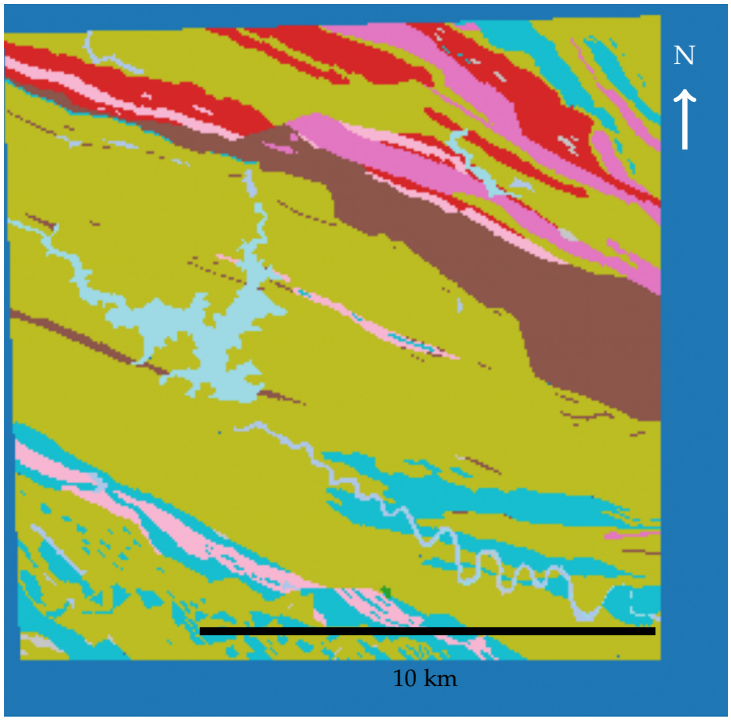


(b) CNN predicted output map of Puertollano

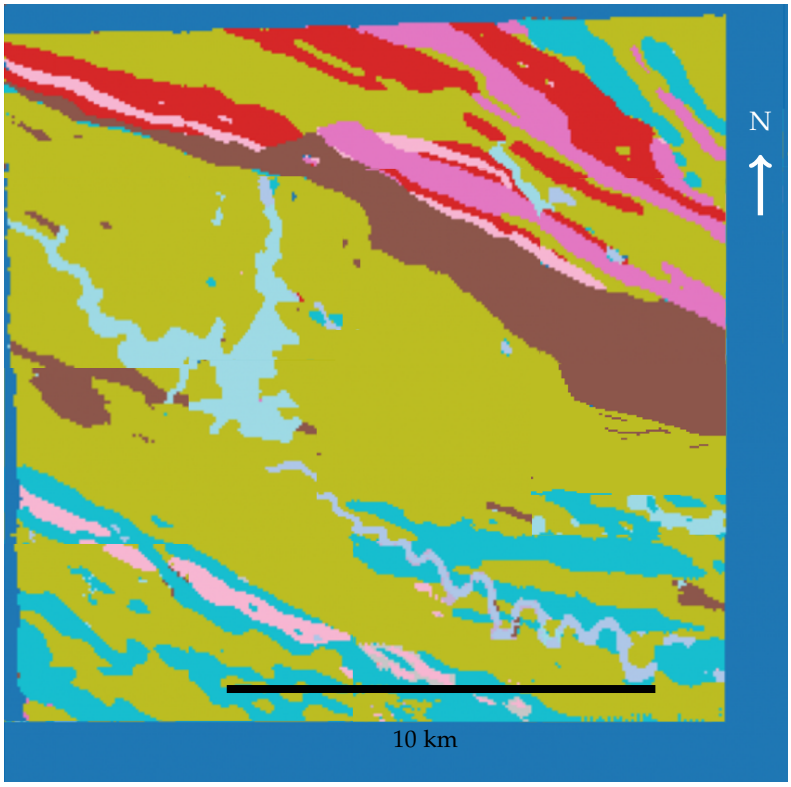


(a) ConvLSTM predicted output map of Puertollano

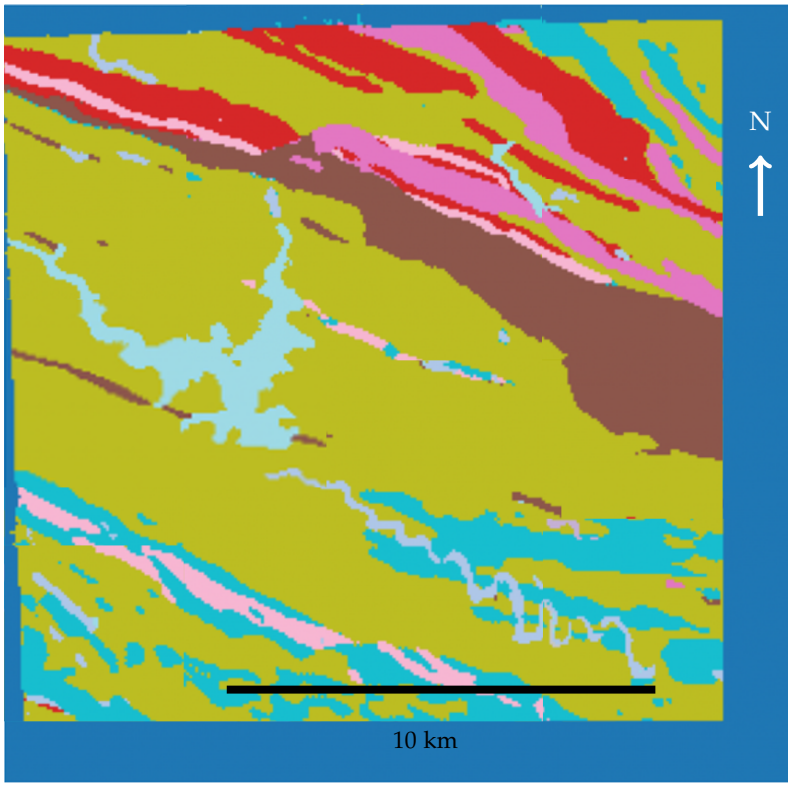
Figure 5.53. Image displaying the results of both the CNN and ConvLSTM fully reconstructed predicted geological maps in Puertollano with respect to the ground-truth labels after further experimentation



(a) Ground-truth label map of Santa Olalla del Cala

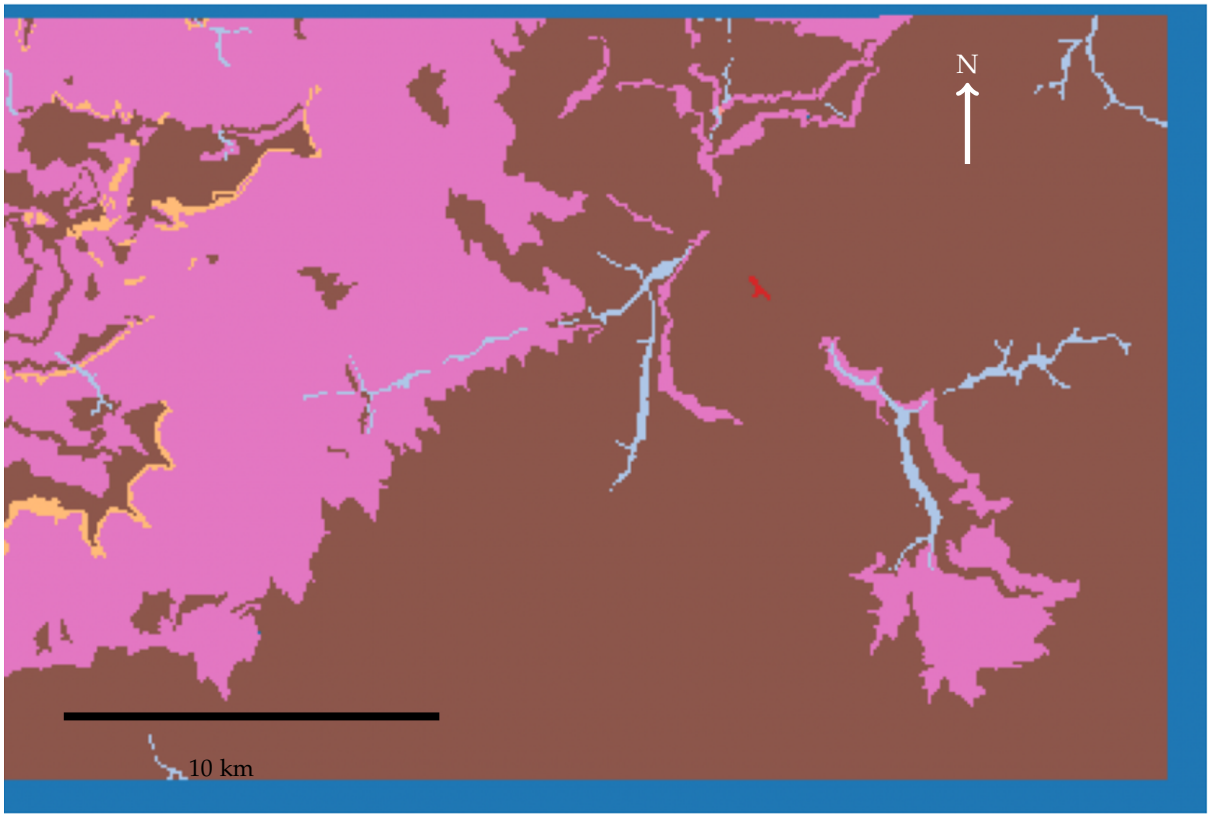


(a) CNN predicted output map of Santa Olalla del Cala



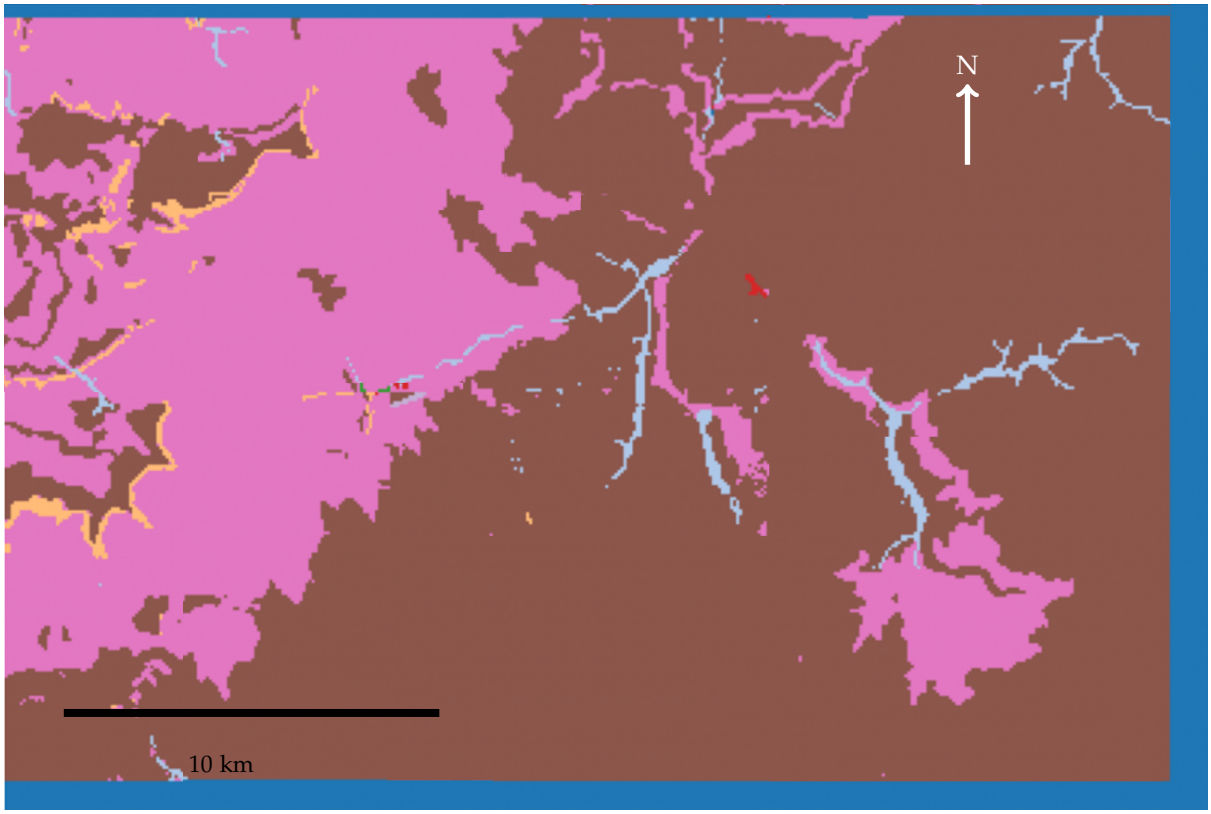
(b) ConvLSTM predicted output map of Santa Olalla del Cala

Figure 5.55. Image displaying the results of both the CNN and ConvLSTM fully reconstructed predicted geological maps in Santa Olalla del Cala with respect to the ground-truth labels after further experimentation



(a) Ground-truth label map of Villoslada de Cameros

(b) CNN predicted output map of Villoslada de Cameros



(a) ConvLSTM predicted output map of Villoslada de Cameros

Figure 5.57. Image displaying the results of both the CNN and ConvLSTM fully reconstructed predicted geological maps in Villoslada de Cameros with respect to the ground-truth labels after further experimentation

5.7.3. NDVI analysis

The section examines how vegetation cover influences model prediction confidence through a comparative analysis of softmax probability maps and NDVI data.

Since the effect of vegetation was consistent across both models, the analysis focuses on ConvLSTM for clarity. Figure 5.58 provides the classification legend providing small descriptions on the range of NDVI values.

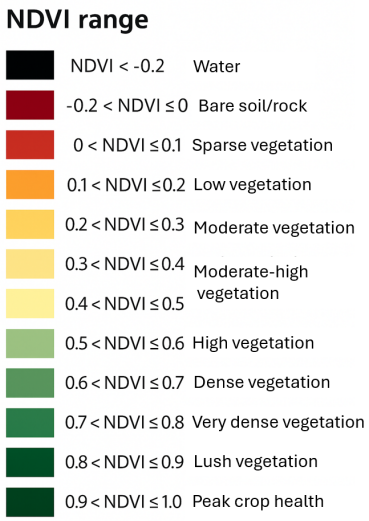
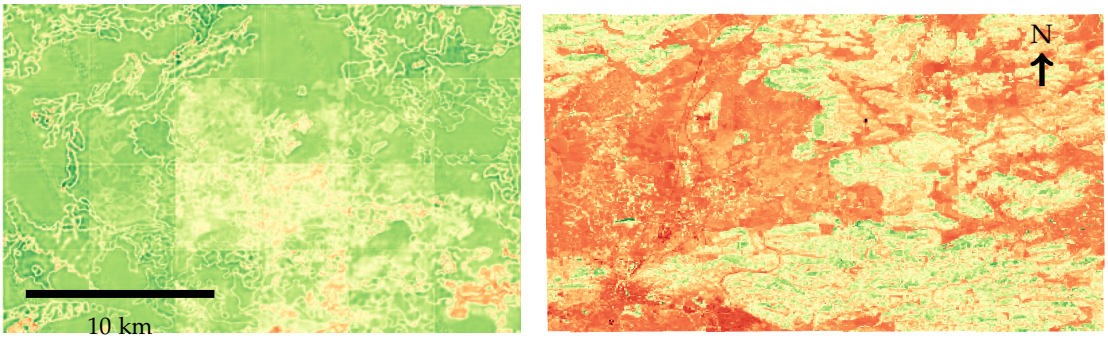


Figure 5.58. Legend with NDVI values for image interpretation.

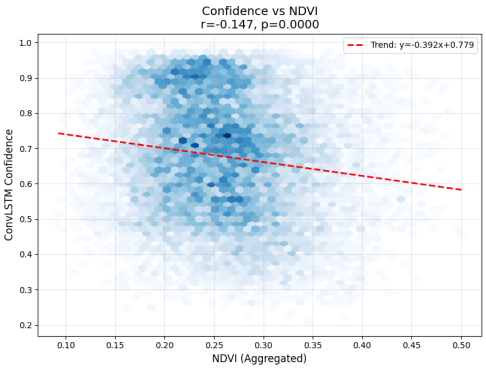
The NDVI-confidence correlation analyses (Figures 5.59, 5.60, 5.61) reveal a consistent negative correlation—indicated by Pearson's correlation coefficient 'r' being negative—across all regions. Therefore, the pattern indicates that vegetation presence (indicated by higher NDVI values) corresponds to poorer model confidence values during outputs. Furthermore, the pixel distribution—indicated by blue hexagons that get darker with higher pixel counts—reveal how each region has different regional behaviours aligned with their climate and vegetation patterns explained in Section 3.4.1. Puertollano concentrates its values between 0.1-0.5, reflecting higher presence of bare soil and sparse vegetation. In turn, it produced the highest accuracy values of all three regions, indicating that the absence of vegetation does improve the predictions, since minimal occlusion allows for thermal contrasts between geological units to be preserved. This was later corroborated with Santa

Olalla, with more intermediate climatic conditions and vegetation cover—represented by its NDVI distribution ranging from mixed cropland (0.3) and dense vegetation (0.7)—, which produced slightly poorer accuracy results. However, despite having the highest overall NDVI values (0.3-0.9), Villoslada exhibits high softmax probabilities across the landscape, contrasting with Santa Olalla's confidence distribution.

This divergent behaviour between regions reveals that vegetation cover alone does not determine the model's performance. One of the reasons behind this surprising contrast between Villoslada and Santa Olalla can potentially be explained through their climate. Since Villoslada presents much more stable annual conditions than Santa Olalla, the thermal and seasonal crop cycles become more dramatic in the southern region of Spain. These annual changes of vegetation can produce some conflicting thermal signatures, degrading the model's ability to correctly predict the same geological units. Instead, the higher vegetation density of Villoslada all year round allows for more stable and consistent thermal values from which the ConvLSTM's hidden state accumulation effectively learns the vegetation-thermal relationship, reducing misclassifications despite the higher vegetation coverage. Nevertheless, it should be considered that since vegetation can become abundant at points, the model can employ certain memorization patterns rather than actual learning, though this could not be fully determined in the current study. In addition, Santa Olalla's partial coverage might be generating heterogeneous thermal fields for the same geological units, obscuring geological boundaries and even introducing ambiguity which the models cannot resolve.

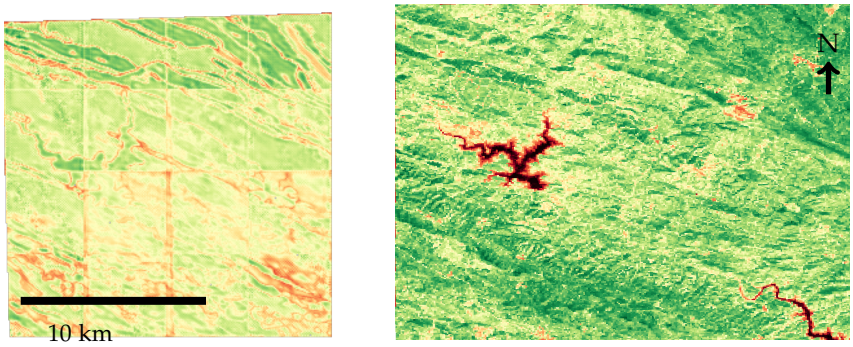


(a) Softmax confidence map (left) and NDVI image (right) for Puertollano

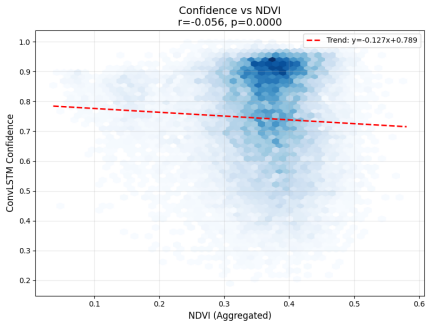


(b) NDVI-confidence correlation analysis

Figure 5.59. Comparison between softmax probability map and NDVI for Puertollano, with correlation analysis

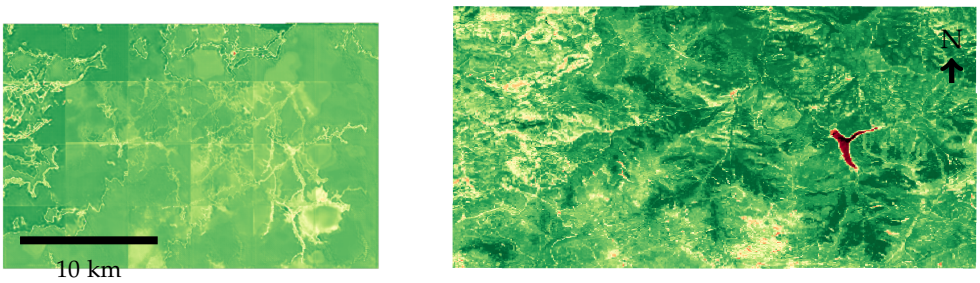


(a) Softmax confidence map (left) and NDVI image (right) for Santa Olalla del Cala

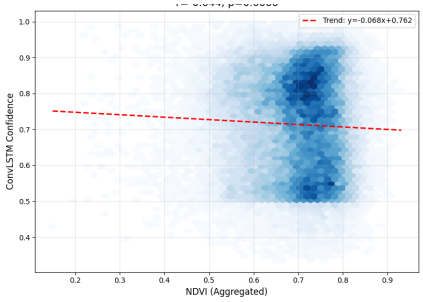


(b) NDVI-confidence correlation analysis

Figure 5.60. Comparison between softmax probability map and NDVI for Santa Olalla del Cala, with correlation analysis



(a) Softmax confidence map (left) and NDVI image (right) for Villoslada de Cameros



(b) NDVI-confidence correlation analysis

Figure 5.61. Comparison between softmax probability map and NDVI for Villoslada de Cameros, with correlation analysis

6. Discussion and limitations

6.1. Discussion

The purpose of this chapter is to interpret the findings presented in Chapters 4 and 5 following the research objectives, hypotheses, and theoretical framework introduced earlier. This section seeks to explain what the results mean, how they answer the research questions, and how they relate to existing literature on soil and rock classification with thermal and complementary data. By analysing different aspects of this study, this chapter highlights the implications of the thesis for geoscientific applications.

The analysis of thermal signatures confirmed the different behaviours consistent with their physical properties that different rocks and soils contain. It revealed insights into the different capabilities between sand and clay, addressing this study's hypotheses. Sand, generally, responded rapidly to temperature changes by displaying sharper fluctuations, due to its coarser texture and lower water retention capability. Instead, clay consistently exhibits slightly smoother transitions, resulting in higher average temperatures due to its ability to retain heat for longer periods of time given its fine-grained texture and higher moisture content (this last one facilitating its representation in SAR experiments). It is this thermal stability that might be the reason of clay's preferential representation in ConvLSTM models, since it provides more reliable training signals for the models. However, thermal patterns present a more complex result than anticipated with the hypothesis established. The similarity in thermal values of classes—such as clay and soil in this case—suggest that vegetation has a higher influence on thermal signatures than anticipated. The same applies to silt, since it was expected to present intermediate thermal values ranging in between those of sand and clay, due to its physical properties. In spite of that, it deviated from the expected behaviour and frequently showed values lower than or similar to clay. This deviation could likely be due to its favourable conditions for vegetation growth (balanced drainage and nutrient retention) which results in higher vegetation cover. These findings indicate that thermal signatures are not only dependent on the target's material but also its land cover interactions.

Since geological maps are used for reference labels, their subjectivity and cartographic generalization always introduces a certain level of uncertainty. This has probably influenced the model's learning. Boundaries between lithologies which often are gradual in nature, become sharp in the labels, reducing the accuracy of class representations already in the labels. As a result, correct predictions could be penalized during training. In addition, resampling of the original geological maps for spatial alignment with thermal images results in the loss of spatial detail and blurred class boundaries in many cases, reducing the ability to distinguish minority classes even more. Thus, a small part of the model's misclassification can be attributed to limitations in label precision. The results should be then interpreted as approximations guided by available ground truths, which remain as simplifications of a more complex geological reality.

The findings aid in answering the second research question regarding the ability of the models in dealing with rocks and soils present with similar thermal attributes. The models generally performed well in distinguishing them, with the ConvLSTM architecture showing an advantage due to its ability to capture temporal dynamics. In many cases, lithologies with similar thermal patterns were still successfully separated, indicating that the models were able to exploit subtle differences in time and space. Classes such as detritic, foliated, and non-foliated rocks displayed more distinct behaviours that were consistently captured, while clay, despite being a minority class, also achieved relatively strong representation. Nonetheless, class imbalance remained a critical factor influencing performance in some cases, especially in the CNN model, when certain classes dominated in the training data, reflected in the F1 score (0.6-0.7) when compared to the accuracy achieved (0.8-0.9), indicating slightly poor performance with minority classes despite high overall accuracy (Table 3).

The decision of reducing the patch size to 64x64 pixels proved to be computationally efficient and did manage to reduce the training time significantly. Despite the large-scale patterns being recognizable, the finer details were in occasions a problem for both models, even with this reduced patch size. Increasing the patch size to 128x128 did not present very noticeable changes, but when combined with overlapping patches, the spatial relations of neighbouring patches improved considerably. Furthermore, the temporal alignment requirement for multi-modal data integration proved to be a strong limitation, reducing the availability of training data for SAR experiments. As a result, although the study managed to obtain a model that classifies rock and soils using thermal data, this constraint limited the study's ability to fully evaluate the potential benefits of including auxiliary data in such methodology. The combined hyperparameter tuning carried out manually and by *Optuna* revealed important insights for model optimization for geological applications. More conservative parameters prioritized training stability and resulted in balanced predictions rather than rapid convergence which resulted in very mixed outputs.

To answer the first research question regarding the ability of CNN and RNNs—as well as combined—to identify the different classes in thermal imagery, the results of Section 5 display the idea very well. The comparative analysis reveals the importance of temporal information in geological classification. While CNN results demonstrated clear ability to perform proper soil and rock segmentation and potential for better predictions, ConvLSTM demonstrated superior ability in capturing minority classes in general. This can be attributed to its capacity to account for temporal dependencies, which provide more context to the model, helping it distinguish materials better and therefore, becoming valuable for identifying rocks and soils that share similar thermal values at specific time points but different thermal reactions over time. It also performs better predictions of minority classes like clay, and in some cases, sand and silt as observed in Section 5.7.1. Although the level of detail in temporally-inclusive models surpassed that of their spatial counterpart, they also exhibited more pronounced overfitting tendencies (observed in the plots in Figures 5.9, 5.10 and 5.11) in the beginning. The additional complexity of temporal dynamics leads to better spatial representation even if the metrics do not always reflect this improvement. Thus, both models prove valuable, with CNNs excelling at overall performance (see figures in Section 5.7.1), faster training times and computational efficiency, while ConvLSTM reveal to have the potential of temporal learning—if efficiency and time are not a requirement—to produce more accurate representations with the presence of minority class.

Throughout the study, the same experiments showed different results for all three regions used. The performance of the predictions varied across the three areas selected, especially with the ConvLSTM models. Villoslada de Cameros proved to consistently provide better results, achieving an average accuracy of 0.74 across all modalities compared to 0.68 for Puertollano and 0.66 for Santa Olalla. Due to its cooler and wetter conditions throughout the whole year, the thermal profiles generated were smoother, allowing for the temporal models to obtain more consistent patterns. Although Santa Olalla produced some of the best patches in certain cases (Table 3), it did generate the highest amount of noisier outcomes, particularly in winter—as the vegetation difference with respect to summer is noticeable in such a dry area. Puertollano displayed more intermediate values but struggled generally with minority classes, especially with the presence of volcanic rocks. These regional variations highlight the importance of terrain and climate characteristics and their impact on the models. While the models play a crucial role, the effectiveness of their performance is certainly tied to regional physical and ecological conditions, providing some intriguing future work for area variations.

The experiments regarding diurnal and seasonal subsets of training data provided further insights into the effect of temporal changes in the performance of the models. Daytime data generally led to more consistent results (0.69 average accuracy) due to solar illumination improvements. Instead, nighttime imagery was highly influenced by the changes in temperature throughout the night, giving slightly worse results (0.62 average accuracy) but not by a huge margin. The mixed results for day and night datasets indicate that optimal timing for thermal data collection is more location-dependent, since Villoslada has smaller variations between day and night patches than Santa Olalla (see Tables 4 and 5). For seasonal variations, summer proved to be the best subset for predicting accurate outputs out of all four temporal subsets (Table 7) due to its clearer atmospheric conditions and reduced vegetation influence. Unfortunately, though winter managed to produce some satisfactory results, it proved challenging for Santa Olalla especially (Table 6). The finding that summer datasets alone can match the complete thermal datasets suggests that targeted data collection can be considered in order to to reduce complexity and memory rather than year-round acquisition. Thus, all these findings answer the research question established, displaying the importance of temporal factors in the performance

of the model, confirming that the added value of including diurnal and seasonal cycles in the training datasets does improve the models' efficiency.

The inclusion of SAR data provided some promising results during the experiments carried out during this project. Although the implementation of multi-modal patches resulted in a limited temporal overlap which, as a consequence restricted the number of usable patches—from the initial 300 thermal patches in Villoslada, only 70 remained for SAR multi-modal patches, indicating a patch loss of around 80%—, the results consistently demonstrated that SAR meaningfully enhanced the classification process. It managed to clearly delineate some of the minority classes present in the areas—most notably clay in Figure 5.38—by leveraging on the radar backscatter's sensitivity to moisture and texture. Since clay tends to retain water for extended periods of time, it exhibits stronger VV and VH return signals in comparison to other drier surfaces like sand. It therefore provides complementary information, which reinforces thermal response for these materials, which in turn enhances the model's capacity to distinguish a variety of classes in complex terrain. Thus, the findings are in agreement with previous studies—like Ndikumana et al. (2018) and S. Liu et al. (2019)—highlighting the ability of SAR data for classification tasks. This study, proved that radar observations can become important complementary sources of information—due to intrinsic properties thermal sensors lack—hence becoming valuable additions to the perspective offered by ECOSTRESS imagery, even though in this study their full potential could not be entirely explored.

The generalizability of the proposed workflow depends on the similarity between the training and new regions testing domain. Since the models have been trained throughout three different Spanish areas with distinct geological and climate settings, a certain level of robustness to regional variability is achieved. However, this generalization is stronger for areas with comparable rock and soil behaviours, climatic conditions and contexts. In such domains, the model has a higher chance of producing similar performances, since most classes would remain within the range learned during training. Instead, due to domain mismatch, it is possible that under markedly different geological characteristics and data distribution—such as areas with a single lithology distribution or little thermal variation—the model's performance might degrade. Therefore, testing in other countries with completely new geological contexts would vary the model's performance depending on the degree of similarity to the current training domains. For instance, in a region dominated by chalk or an area covered in granite (such as the Moroccan granite fields), the model might underperform, since it has inherently learned they correspond to minority classes, instead of being capable of becoming a dominant landscape. Furthermore, if the area contains a higher distribution of previous minority classes (such as seaside areas in the Netherlands), results could be moderate, though it might still confuse these classes due to class imbalance, over-predicting previously dominant classes. Overall, while the model's design and workflow are scalable and transferable, its accuracy remains tied to the availability of representative and region-specific ground truths. Thus, the models currently would need to be trained on locally adapted labels which would then produce more favourable results for these areas of interest.

The iterative refinement of model parameters later on in the project during the further experiments phase yielded substantial improvements through partially overcoming the overfitting problem through three complementary strategies. The final configuration set-up (Table 11) incorporated an adaptative learning rate for each model, increased the patch size and an overlap of 50% for all patches with an extended training time of 150 epochs. The rest of the parameters remained the same due to lesser impact on the results. The integration of focal loss into the custom weighted loss function proved to be the most impactful parameter change, increasing F1 scores from 0.5-0.7 up to 0.7-0.85 across all regions and models, as a result, improving the accuracy by almost 0.2—a result that underscores the critical importance of the loss function for imbalanced geological datasets, as opposed to the findings in the core experiments. This mechanism forces the model to prioritize misclassified minority classes (Dina et al., 2023) rather than maximizing majority class dominance. The final class weights tailored to each region can be observed in Table 12. The second approach introduced patch overlap, up to 50%, fundamentally altering the model's learning, especially in ConvLSTM, by transforming previous patch boundaries into patch centres, which now the model could learn from to discriminate more features from and focus on the spatial relationships between adjacent and neighbouring patches as well as geological classes. Whilst this method did not eliminate overfitting, it did reduce patch-level discontinuities and enhanced the spatial coherence of the predictions. Instead, this adjustment improved the training dynamics together with the learning rate and the l2 factor, proving that overfitting in geological segmentation can stem from several sources, requiring the coordination of a wide range of parameters instead of isolated solutions.

Table 11. Final configuration of parameters for CNN and ConvLSTM models

Parameter	Value
L2 Regularization	0.05
Learning Rate (CNN)	0.00001
Learning Rate (ConvLSTM)	0.0001
Batch Size	4
Base Filters	16
Patch Size	64/128 pixels
Patch Overlap	50%
Loss Function	Custom + Focal Loss
Epoch Count	150

Table 12. Class weights assigned for each geological region to address class imbalance

Geological Class	Puertollano	Santa Olalla	Villoslada
NaNs (Background)	1.5	1.5	1.0
Sand	10.0	8.0	3.5
Clay	10.0	1.0	2.5
Chalk	1.0	8.0	1.0
Silt	7.5	4.0	4.0
Peat	1.0	1.0	5.0
Loam	4.0	1.0	1.0
Detritic	1.0	2.5	1.0
Carbonate	1.5	3.0	2.0
Volcanic	3.0	3.0	1.0
Plutonic	1.0	3.0	1.0
Foliated	15.0	1.0	1.0
Non-Foliated	2.5	4.0	1.0
Water	1.0	4.0	1.0

Notably, ConvLSTM proved to have greater sensitivity to all the changes carried out—in comparison to the CNN—, with Puertollano and Villoslada achieving validation losses under 1.0 and mean accuracy gains of $\simeq$10%. These distinctive responses support the argument that temporal memory mechanisms are better adapted to parameter optimization techniques than spatial-only architectures. The consistency of improvements suggests these modifications indicate reliable advances for thermal-based geological segmentation.

As previously observed in thermal-LST studies (Afrasiabi Gorgani et al., 2013), the correlation between thermal signatures and vegetation is inherently negative. The correlation analysis from Figures 5.59, 5.60 and 5.61 confirm this pattern, with a negative Pearson's correlation coefficient in all cases. Notwithstanding, the visual inspection of softmax probability maps reveals regional manifestations of this relationship. Firstly, Puertollano presents spatially uniform high probabilities across low NDVI values (0.1-0.5), reflecting lesser vegetation occlusion than the remaining areas and maximal thermal contrast between lithological units. Secondly, Santa Olalla exhibits patchy, variable confidence aligned with mixed vegetation coverage, with lower confidence levels being present in the highly vegetated sections. Finally, Villoslada displays consistently high probabilities across areas of very dense vegetation (0.3-0.9), contrary to the suggestion that higher vegetation coverage hinders predictions. This would suggest the model has learned to either suppress vegetation through temporal averaging, or the model has memorized certain patterns during training. The reason behind these divergent behaviours could stem from vegetation stability. Villoslada's all-year-round stable conditions generate consistent thermal signatures (as observed in Figure 5.3) throughout the temporal sequence. Consequently, ConvLSTM's hidden state LSTM accumulation learns vegetation-robust representations. In the case of Santa Olalla, the reason behind the poorer confidence maps could be the result of the landscape undergoing more dramatic seasonal changes, conflicting thermal signatures for some geological units. Thus, the negative NDVI-confidence correlation reflects thermal sensing's fundamental limitation: vegetation attenuates subsurface material signals. These findings underscore that temporal dependencies

can filter vegetation noise in certain cases, providing additional robustness for thermal-based geological segmentation in vegetated environments, although a more detailed and selective vegetation cover approach will always be considered relevant and produce further improvements to this segmentation process.

The study demonstrates that while thermal-based geological classification faces significant challenges with class imbalance and data availability, the approach displays sufficient potential for continued development and refinement. The successful predictions achieved in all regions provides a foundation for more targeted applications or methodological improvements. Overall, the findings provide strong support for the initial hypotheses, with clay and sand showing the behaviours expected initially. In addition, both CNN and ConvLSTM models prove effective with their unique strengths for rock and soil classification with thermal data. Temporal factors proved to be impactful in the results, with summer conditions being the most favourable one. Finally, the inclusion of SAR data managed to display the potential of multi-modal input data for better predictions, validating the research question on complementary data.

6.2. Limitations

Despite the contributions of this study, certain limitations must be acknowledged. Recognizing these boundaries is important for a balanced interpretation of the results obtained during the thesis. The following section describes the limitations that had an effect on the robustness and performance of the models.

- **Geology maps -** although geology maps were used as ground-truth labels for this study, they are not entirely free from uncertainty. Spanish geology maps provided by the IGME are considered reliable and widely used for scientific purposes. Yet, they are still the product of field campaigns and interpretative analysis from experts (Marjoribanks, 2010) and graduate researchers. Therefore, they always contain a certain degree of subjectivity and cartographic approximation that should always be accounted for. Certain boundaries, especially between lithological units with gradual transitions should simply serve as reference, since such sharp lithological changes do not always occur—they do not represent exact spatial positions. This inherently affects the quality of ground truth labels, since during training, predictions could be influenced by misaligned boundaries—with reality—reducing the apparent accuracy even when predictions might be geologically reasonable. Consequently, just as with geology maps, the results of this thesis should be interpreted as initial references or approximate representations of what the true geological surface might actually look like—which could possibly be somewhere between the actual geological map and the results obtained in the project—, rather than exact and definite maps. These can turn into higher-precision geological representations with further refinements or field-validated ground truths (that help with very vegetated areas).
- **User-defined labels -** as outlined in Section 2.1, numerous geological classification systems are present nowadays. Given their broadness, a simplified version of them was adopted in this study. The aim was to simplify the number of classes and the complexity behind label modelling. Yet, the use of a tailored classification system like the one used in this study makes direct comparisons with other studies more difficult. Furthermore, the chosen classes encompass many subdivisions and lithological categories, meaning the labels generated are broad and can lack some detailed rock definitions.
- **Polygon weights -** class weight assignment was introduced as part of the methodology of this project to test any possible improvements in comparison to the standard built-in *Keras* loss function: however since it was done per-class and not spatially within polygons, some details are lost. The importance of this stems from the fact that, in geology, typically the more accurate areas are found at the core of lithological units rather than at their boundaries. Not differentiating the weight of these central areas with marginal ones results in some spatial information being lost, potentially reducing the model's performance.
- **Spatial resolution -** differences in spatial resolution represented one of the main issues of this project. Due to the varied spatial resolutions of the different satellite imagery used—70m for

ECOSTRESS, 5-20m for Sentinel-1, 10m for Sentinel-2, and around 25m for geology maps—, images had to be aligned. To do so, resampling and resizing was necessary, to match the coarser resolution of thermal images (the main focus of this study), leading to loss of detail due to the resolution decrease. This became particularly problematic for the labels. Finer-scale information was smoothed out, potentially affecting the results and robustness of the model's training and testing phase. This degradation in precision for labels and input images can significantly affect deep learning outcomes, potentially confusing boundaries or small classes—in the case of labels—. The incorporation of finer-detailed geology maps and SAR imagery without resampling could have proven to be even more effective for the model's robustness and accuracy if their resolution could have been kept untouched or such improvements can be implemented in some future work.

- **Temporal resolution -** a similar issue occurred with the revisit frequency of the satellites employed in this project. As explained in Section 4.1, revisit cycles for ECOSTRESS and Sentinel-1 are very different, causing highly unbalanced data availability across sensors. In turn, in the five year study period available, thermal images were far more abundant than SAR images. Such a big mismatch of image numbers made constructing consistent multi-modal datasets challenging. Thus, performing perfect comparisons on the implementation of auxiliary data was rather difficult. In addition, it would also affect data alignment during patch creation. Since patches were ideally built from acquired images from the same date across all three sensors, the probability of having coincidental dates for thermal and SAR, data was very low, limiting the use of auxiliary images.
- **Time span limitations -** the main reason for the short supply of SAR images is in fact ECOSTRESS' launch date. The temporal coverage is restricted by ECOSTRESS' launch in 2018. Since there is no imagery available prior to that date, it leaves only five years of thermal data available at the time of analysis. Even though there are further SAR images available prior to 2018, they cannot be employed to keep experimentation fair. While this period of time is reasonable for initial thermal analysis on rocks and soils, it is relatively short when complementing it with the already limited auxiliary data.
- **Atmospheric conditions -** thermal infrared is highly sensitive to certain atmospheric conditions. Since thermal infrared is affected by varying weather, variables like cloud coverage or precipitation can distort LST values. To ensure that data quality is maintained for a proper training phase, many images had to be removed or discarded after some quality filtering. To avoid excessive removal of imagery, either up to 10% of cloud cover or missing data were allowed in each image, but not both. Even so, this process further reduced the number of usable images, limiting the volume of information available for the models to train and test on.

7. Conclusion and Future Work

7.1. Conclusion

The study set out to evaluate the potential of thermal imagery for rock and soil classification. The research successfully demonstrated the usability of employing thermal imagery for automated rock and soil classification through deep learning algorithms. It aimed to determine if CNN and ConvLSTM models could identify soil and rock types under different temporal and seasonal conditions, and whether integrating complementary data such as SAR improved the performance.

The findings demonstrate that thermal data alone proved to produce the strongest and most reliable performances, since the size of the datasets available was larger than that of the other datasets and also its ability to capture material-specific properties. The refinements of the models through coordinates optimization of loss function design, patch overlap strategies, and regularization techniques proved to obtain enhanced results. The final configuration (Table 11), with focal loss proving to have been a decisive factor, improved accuracy by approximately 1.5-0.2 across all regions. ConvLSTM proved to be substantially more sensitive to these improvements than CNN, achieving higher accuracy and F1 score values in the end, indicating an accuracy gain of around 10% with respect to the initial experiments. These findings reinforced the idea that temporal memory mechanisms provide advantages for geological segmentation tasks which spatial-only models lack. Nevertheless, both models managed to produce very consistent and reliable results, albeit the occasional patch misclassification. The seasonal analysis demonstrated the effectiveness of the summer dataset, yielding the best results probably due to the reduced cloud coverage during the summer months and the enhanced thermal contrast in ground materials. Despite the limitations present from the limited amount of available imagery, multi-modal integration proved to be promising—SAR data as complementary information helped the models account for more features present in the inputs.

The NDVI analysis revealed how vegetation does impact model confidence, where coverage stability had a higher influence than density alone. Generally, thermal imagery experiences vegetation attenuation, restricting geological values. However, vegetation cover did not pose as much impact as expected, especially for ConvLSTM, which seemed to adapt well through its temporal mechanisms learning vegetation-adjusted features. The findings established that annual landscape variations could generate fluctuating thermal values with which models struggle more to adapt to, whereas consistent climatic conditions can be handled more adequately by the models. These results indicate that in rock and soil segmentation, consideration of environmental distribution and stability is necessary to improve outputs.

While the proposed workflow does demonstrate scalability in areas with similar geological and climatic conditions, its generalizability could remain limited in areas with different lithological distributions. Future research work could improve the scalability by adjusting the models to perform continuous learning strategies (Ao & Fayek, 2023). This way, the model progressively integrates knowledge from several areas, understanding the distribution and behaviours of certain classes under different conditions—considering the limitation of catastrophic forgetting (Shin et al., 2017), ensuring new information does not cause the model to lose previously learn knowledge. Therefore, the model could become more robust and accurate towards current minority classes. Consequently, the model could also become reliable in many different contexts and regions.

Overall, the models successfully captured thermal distinctions between certain soil and rock classes—albeit with varying degrees of success— to perform proper segmentation, as well as supporting the central research hypotheses, especially with clay (due to their stable thermal behaviour). The thermal signatures revealed that the thermal behaviour of both sand and clay and its variation with time did support the hypotheses established. Moreover, the models managed to produce distinctions between different classes, especially in the case of the ConvLSTM. Most cases produced reasonable predictions

of the actual geological distribution, obtaining remarkable accuracies that closely replicated geological maps. In occasions, certain patches did exhibit deviations from the expected results through miss-classification or certain discontinuity in spatial relationships. These inconsistencies highlight the challenges present with such models and datasets, presenting class imbalances, model generalisation or even environmental variations. Vegetation coverage proved to have certain impact on the models' prediction confidence, yet both of them seemed to produce valuable results despite the challenges. These results provide insights into what can be a promising new tool for rock and soil classification.

Despite the limitations present in the project, the research aims to present an important initial step towards automating soil and rock classification with remote sensing and AI by providing a supporting reliable tool. Deep learning models can reduce reliance on manual methodologies used in geological mapping, allowing for new possibilities in mining, geology, hazard assessment or urban planning. The demonstrated capability to differentiate between rock and soil types using freely available satellite data offers significant potential for developing countries and remote regions where traditional surveying can become topographically or economically challenging. With more extensive datasets for thermal imagery, refined strategies, the approaches can turn into a more robust, cost-effective, scalable tool capable of supporting professionals across these disciplines.

7.2. Future work

The limitations identified in Section 6.2 suggest several opportunities for further research. Future studies could address the challenges faced during the thesis and expand upon the findings, thus refining the proposed methodologies. Since this research was conceived as an initial contribution to multiple fields, future work can extend and build on the approaches developed.

Firstly, the classification system implemented in this thesis can be further enhanced by increasing its complexity. The ground-truth labels produced in this project were intentionally simplified due to time constrains, prioritizing insights into model performance over generating a comprehensive rock categorization. However, incorporating more detailed geological classification systems in future work may enable richer predictions, hence supporting improved informed decision-making.

Weight class importance of rocks could also be further enhanced. As noted previously in section 6.2, rock margins are inherently less certain than central lithological areas. Introducing customized pixel weights during the rock polygon construction (e.g., in QGIS) could potentially enhance the training process of the models. By accounting for spatial uncertainty, predictive accuracy can be improved, increasing the reliability of the outputs generated for real-world practical applications.

Conducting the experiments with consistent spatial resolution and equal image counts for all data types could potentially improve the performance of the models. With such an approach, any loss of information could be prevented during the pre-processing stage. Unfortunately, achieving this may require the use of commercial satellite imagery or alternative data providers. As a result, employing imagery from other providers could benefit data availability. Increasing the availability of SAR data will produce more frequent temporal matches with thermal images, enriching the input information for the model to train from. Moreover, extending thermal data sources beyond ECOSTRESS' launch date—such as ASTER—could lengthen the temporal coverage for more comprehensive analyses. Consequently, date matches with SAR could also become more frequent too.

The models developed in this study have been tested and evaluated exclusively on regions in Spain, due to their favourable balance of atmospheric conditions and geological diversity. Future work on this project could involve the assessment of the model in different environmental conditions to further assess its robustness and improve its scalability further on. For instance, enhancing the model by training it with images from countries such as the Netherlands or Germany—-with persistent cloud coverage and challenging climatic conditions—as well as in different geological contexts and distributions could improve the performance across different scenarios. To do this, adapting the models for continuous learning can make such an approach possible while, at the same time, avoid catastrophic forgetting issues. Furthermore, although this study tackled the influence of vegetation, taking into account the impact this has on the results, as established in Section 5.7.3, a model that includes NDVI as input together with thermal channels or that removes vegetation could produce an enhanced predictive model that can perform well in different environmental contexts and countries. Moreover, the

later proposed environments could test the model performance in other urban contexts where thermal properties are highly influenced by infrastructures and human activity. Removing or accounting for urban pixels can improve the robustness of the model in large city areas. All these cases can extend the applicability of the approach suggested in this study, extending its scalability, viability and covering a wider range of fields and tasks.

In summary, future research can build upon this thesis by expanding the classification system, incorporating proximity weighting, improving data consistency and availability, or testing model robustness across diverse geographical and environmental settings. The approaches can contribute to more accurate and versatile models under varying conditions. Thus, these suggestions can strengthen the reliability of the models, resulting in a viable application for geological and environmental analysis.

A. Reproducibility self-assessment

A.1. Marks for each of the criteria

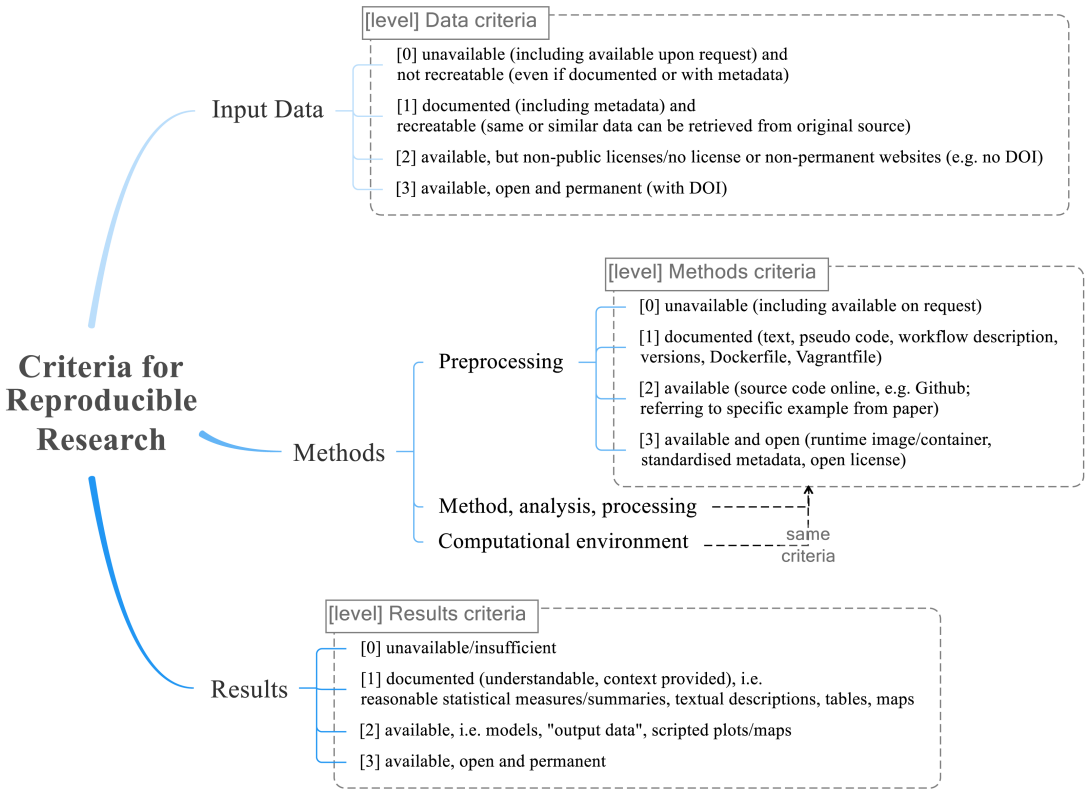


Figure A.1. Reproducibility criteria to be assessed.

To evaluate the reproducibility of this thesis, a self-assessment was carried out following the five criteria described in Figure A.1. Each criterion scored on a 0–3 scale, with higher values reflecting stronger reproducibility. The results are as follows:

- **Input data:**
  - Satellite data – 3
  - Labels – 0
- **Methods:**
  - Pre-processing – 2
  - Method, analysis, processing – 2
  - Computational environment – 1
- **Results:** 2

A.2. Self-reflection

The reproducibility of the project is mixed and dependent on the phase. Firstly, input data has a very mixed reproducibility. Thermal, SAR and NDVI imagery comes from open sources—ECOSTRESS via AppEARS and Sentinel via GEE—, where data is openly available and accessible from their original providers. Nevertheless, some sites may require simple registration. Instead, the digitized geological labels produced in this project were manually generated and are not publicly shared. Although the methodology is shared and the labelling process could be reproduced, generating the same exact labels might only be achieved by sharing data.

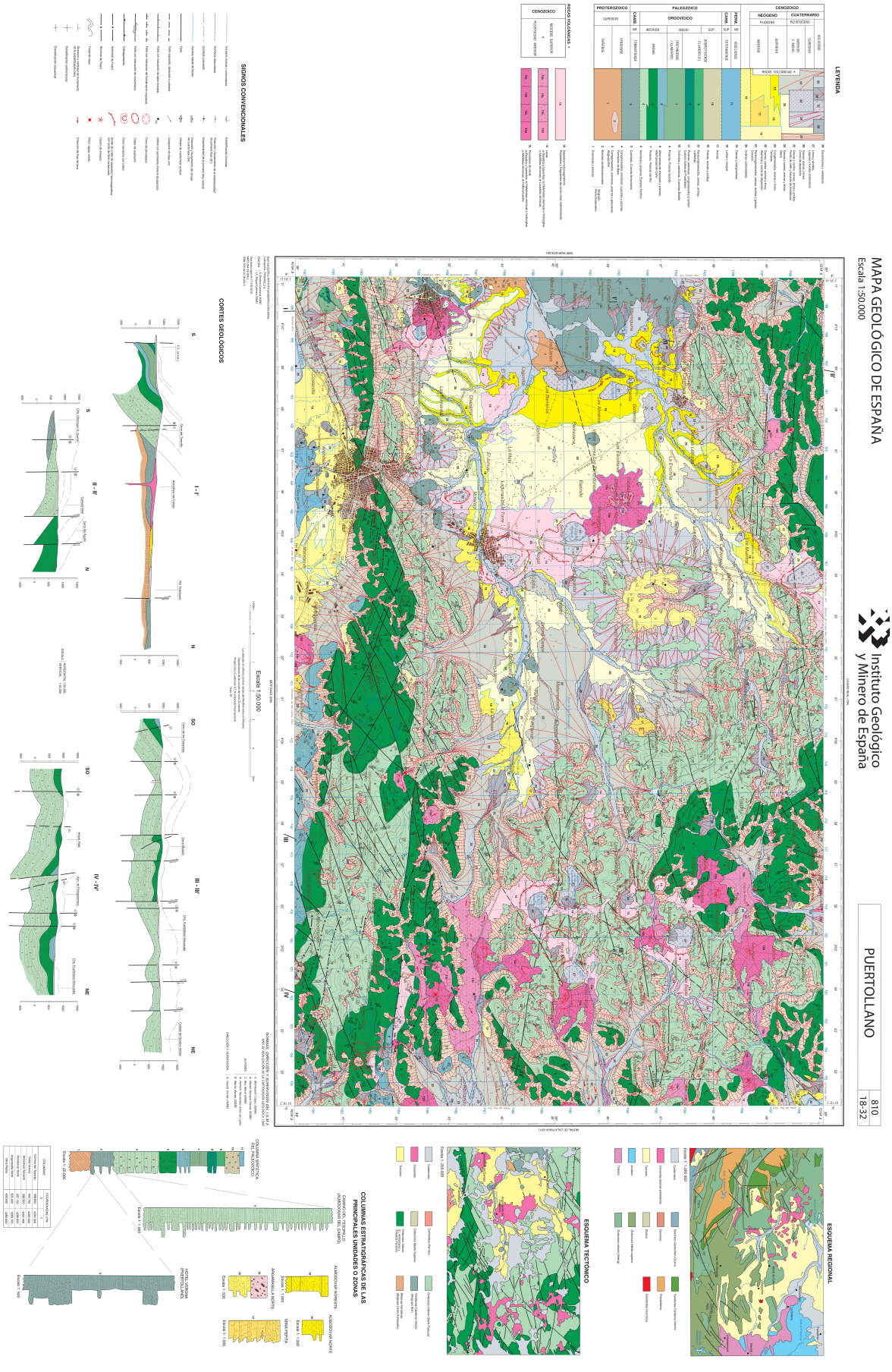
Regarding methods, the preprocessing steps and modelling workflow are described in detail in Sections 3.4 and 4. Quality controls, model architectures (CNN and ConvLSTM), hyper-parameter tuning (parameters such as learning rate, batch size, filters, etc.), and the training strategies are well reported. Sufficient information is provided to allow reproduction of the study with similar datasets by following the documentation established in those sections. However, some manual steps (e.g., digitization and QGIS processing) limit the automation of this process.

In the computational environment aspect, the programming language used (Python) and QGIS are of free use. The packages employed are mentioned but without full environment specifications (e.g., versions or OS). The environment and software used (PyCharm) do require authorisation through institutions or behind a pay-wall. Others can approximate the environment, however, exact reproducibility may be difficult.

Results are well-documented in the thesis, reported with figures, tables, and performance metrics plots. Nevertheless, without access to exact data splits, and environment, complete reproducibility (e.g., trained model weights) with exact same outputs would be challenging. Results can be partially reproducible if other users implement the methodology and available data described in this paper.

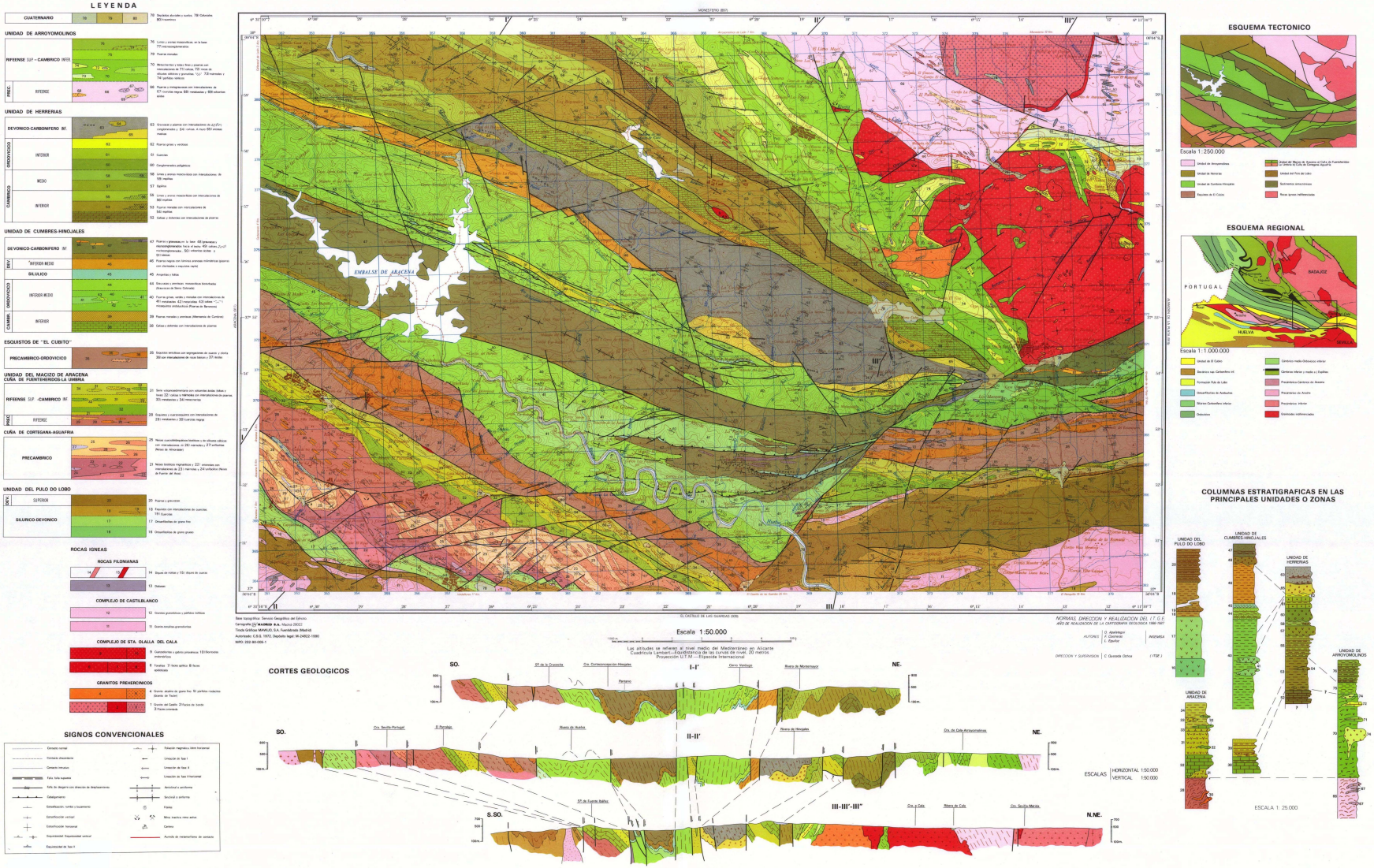
Overall, the thesis demonstrates a medium level of reproducibility: sufficient documentation and methodological transparency exist for knowledgeable researchers to replicate the study, but reproduction would require additional effort due to the lack of openly shared labels, a fully defined computational environment, and automated processing scripts.

B. Geology maps



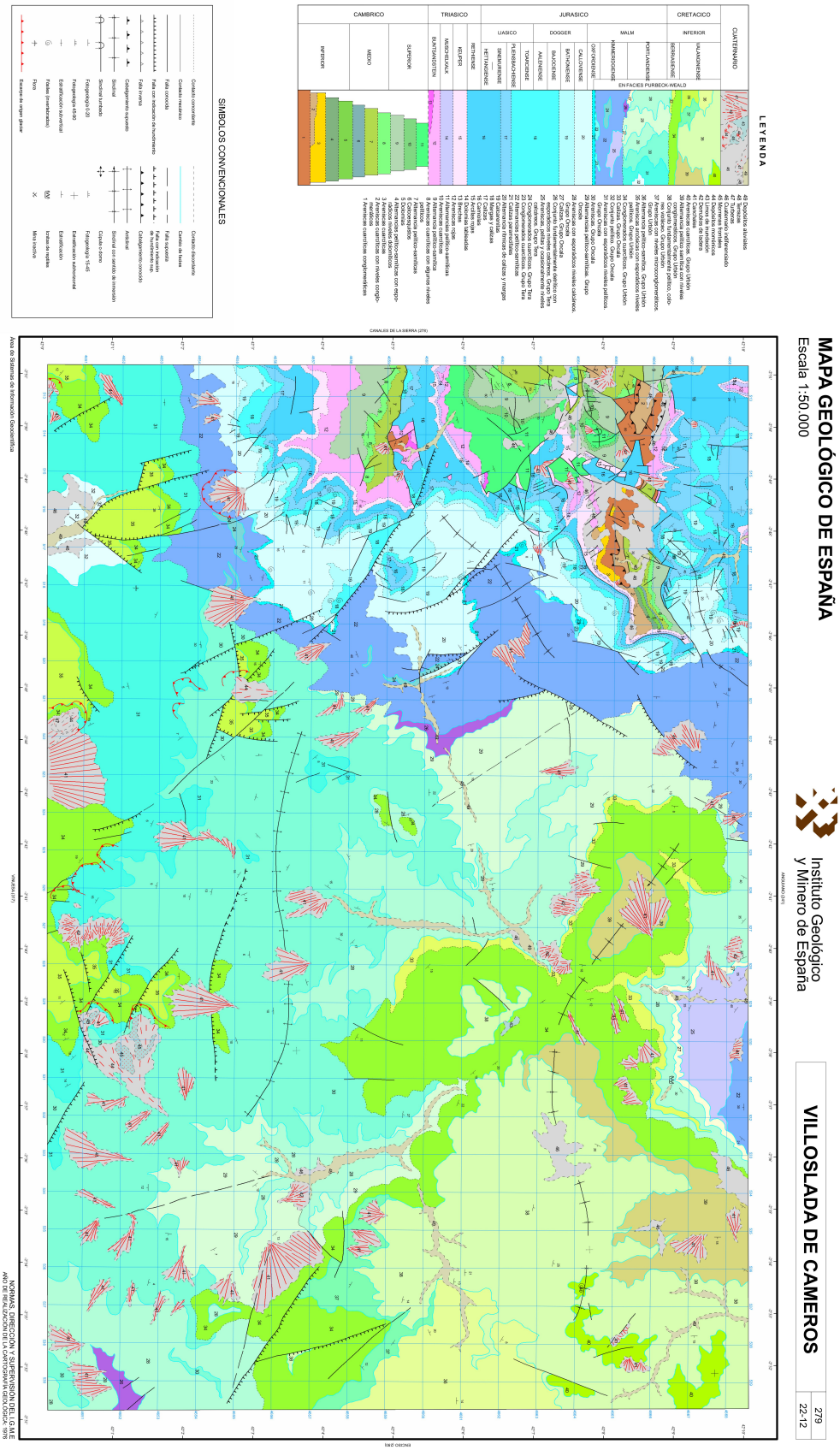
Source: Instituto Geográfico Nacional (IGN) (2020)

Figure B.1. Official geological map of Puertollano



Source: Instituto Geográfico Nacional (IGN) (2020)

Figure B.2. Official geological map of Santa Olalla



Source: Instituto Geográfico Nacional (IGN) (2020)

Figure B.3. Official geological map of Villoslada de Cameros

C. Dates and times of thermal images

Villoslada de Cameros

Date	Time	Diurnal	Season
11/02/2019	17:28:50	Day	Winter
13/02/2019	10:53:34	Day	Winter
15/02/2019	15:39:30	Day	Winter
17/02/2019	09:05:06	Day	Winter
17/02/2019	09:05:07	Day	Winter
20/02/2019	08:07:23	Day	Winter
21/02/2019	13:44:01	Day	Winter
21/02/2019	13:44:02	Day	Winter
22/02/2019	12:52:25	Day	Winter
22/02/2019	12:52:26	Day	Winter
23/02/2019	07:08:47	Day	Winter
23/02/2019	07:08:48	Day	Winter
25/02/2019	11:54:42	Day	Winter
28/02/2019	10:57:45	Day	Winter
01/03/2019	10:07:01	Day	Winter
11/03/2019	06:27:16	Day	Winter
01/06/2019	15:40:08	Day	Spring
27/06/2019	05:29:19	Night	Summer
11/07/2019	00:24:37	Night	Summer
15/07/2019	21:54:43	Night	Summer
18/07/2019	21:03:54	Night	Summer
22/07/2019	19:23:39	Night	Summer
29/07/2019	16:51:43	Day	Summer
02/08/2019	15:11:17	Day	Summer
10/08/2019	11:49:54	Day	Summer
13/08/2019	10:58:25	Day	Summer
12/09/2019	22:38:35	Night	Summer
12/09/2019	22:39:27	Night	Summer
19/09/2019	20:12:33	Night	Summer
01/10/2019	15:20:37	Day	Autumn
01/10/2019	15:21:29	Day	Autumn
05/10/2019	13:43:26	Day	Autumn
09/10/2019	12:06:22	Day	Autumn
20/10/2019	08:02:22	Day	Autumn
08/11/2019	00:42:10	Night	Autumn
11/11/2019	23:06:17	Night	Autumn
11/01/2020	22:57:33	Night	Winter
11/01/2020	22:58:25	Night	Winter
15/01/2020	21:23:00	Night	Winter
23/01/2020	18:12:55	Night	Winter
31/01/2020	15:06:13	Day	Winter
04/02/2020	13:32:24	Day	Winter
28/02/2020	04:10:44	Night	Winter
07/03/2020	01:03:29	Night	Winter

Date	Time	Diurnal	Season
14/03/2020	21:55:28	Night	Winter
22/03/2020	18:48:15	Night	Spring
03/04/2020	14:11:10	Day	Spring
07/04/2020	12:39:29	Day	Spring
20/05/2020	19:28:59	Night	Spring
20/05/2020	19:29:51	Night	Spring
28/05/2020	16:20:10	Day	Spring
05/06/2020	13:11:12	Day	Spring
13/07/2020	22:14:04	Night	Summer
14/07/2020	21:26:03	Night	Summer
14/07/2020	21:26:55	Night	Summer
17/07/2020	20:39:34	Night	Summer
25/07/2020	17:29:49	Day	Summer
06/08/2020	12:48:16	Day	Summer
14/08/2020	09:40:58	Day	Summer
22/08/2020	06:33:16	Day	Summer
24/08/2020	11:27:06	Day	Summer
28/08/2020	09:54:19	Day	Summer
31/08/2020	09:07:40	Day	Summer
08/09/2020	06:00:06	Day	Summer
30/09/2020	15:02:03	Day	Autumn
16/10/2020	08:49:06	Day	Autumn
26/10/2020	10:34:23	Day	Autumn
29/10/2020	09:47:33	Day	Autumn
30/10/2020	09:00:05	Day	Autumn
18/12/2020	07:48:31	Day	Winter
10/01/2021	22:24:51	Night	Winter
25/01/2021	16:57:29	Day	Winter
26/01/2021	16:10:07	Day	Winter
26/01/2021	16:10:59	Day	Winter
03/02/2021	13:05:29	Day	Winter
11/02/2021	09:59:14	Day	Winter
11/02/2021	10:00:06	Day	Winter
15/02/2021	08:27:01	Day	Winter
24/02/2021	11:02:56	Day	Winter
28/02/2021	09:29:55	Day	Winter
01/03/2021	08:42:09	Day	Winter
01/03/2021	08:43:01	Day	Winter
03/03/2021	02:14:46	Night	Winter
04/03/2021	07:56:48	Day	Winter
05/03/2021	07:09:09	Day	Winter
03/04/2021	13:52:39	Day	Spring
07/04/2021	12:20:15	Day	Spring
15/04/2021	09:15:04	Day	Spring
16/04/2021	08:28:04	Day	Spring
19/04/2021	07:42:19	Day	Spring
23/04/2021	06:09:20	Day	Spring
23/04/2021	06:10:12	Day	Spring
06/05/2021	00:43:39	Night	Spring
07/05/2021	06:26:13	Day	Spring
11/05/2021	04:53:15	Night	Spring
20/05/2021	19:19:06	Night	Spring
25/05/2021	16:59:56	Day	Spring
02/06/2021	13:54:21	Day	Spring
09/06/2021	11:35:41	Day	Spring

Date	Time	Diurnal	Season
13/06/2021	10:02:35	Day	Spring
21/06/2021	06:56:55	Day	Spring
22/06/2021	06:09:21	Day	Summer
25/06/2021	05:24:13	Night	Summer
05/07/2021	07:15:10	Day	Summer
09/07/2021	05:43:07	Night	Summer
19/07/2021	19:24:00	Night	Summer
19/07/2021	19:24:52	Night	Summer
08/08/2021	11:42:39	Day	Summer
09/08/2021	17:24:14	Day	Summer
12/08/2021	10:09:39	Day	Summer
16/08/2021	08:37:30	Day	Summer
20/08/2021	07:04:10	Day	Summer
20/08/2021	07:05:02	Day	Summer
22/08/2021	12:00:02	Day	Summer
24/08/2021	05:32:41	Night	Summer
26/08/2021	10:28:17	Day	Summer
26/08/2021	10:29:09	Day	Summer
08/10/2021	17:42:00	Day	Autumn
09/10/2021	16:54:22	Day	Autumn
13/10/2021	15:21:57	Day	Autumn
15/10/2021	08:55:04	Day	Autumn
17/10/2021	13:49:27	Day	Autumn
19/10/2021	07:22:22	Day	Autumn
19/10/2021	07:23:14	Day	Autumn
21/10/2021	12:17:24	Day	Autumn
02/11/2021	07:40:52	Day	Autumn
11/12/2021	16:20:00	Day	Winter
12/12/2021	15:32:25	Day	Winter
12/12/2021	15:33:17	Day	Winter
14/12/2021	09:05:41	Day	Winter
15/12/2021	14:47:09	Day	Winter
20/12/2021	12:26:40	Day	Winter
20/12/2021	12:27:32	Day	Winter
31/12/2021	08:29:48	Day	Winter
25/01/2022	16:27:07	Day	Winter
29/01/2022	14:52:05	Day	Winter
02/02/2022	13:15:57	Day	Winter
10/02/2022	10:04:57	Day	Winter
11/02/2022	15:45:33	Day	Winter
15/02/2022	14:09:47	Day	Winter
10/04/2022	16:41:01	Day	Spring
16/04/2022	08:35:01	Day	Spring
24/04/2022	05:20:44	Night	Spring
29/04/2022	09:25:56	Day	Spring
30/04/2022	08:37:34	Day	Spring
04/05/2022	07:01:26	Day	Spring
08/05/2022	05:25:19	Night	Spring
20/05/2022	18:56:18	Night	Spring
20/05/2022	18:57:10	Night	Spring
28/05/2022	15:46:08	Day	Spring
29/05/2022	14:57:24	Day	Spring
03/06/2022	19:02:55	Night	Spring
05/06/2022	12:33:53	Day	Spring
12/06/2022	10:08:50	Day	Spring

Date	Time	Diurnal	Season
13/06/2022	09:20:53	Day	Spring
20/06/2022	06:55:49	Day	Spring
22/06/2022	11:49:30	Day	Summer
24/06/2022	05:20:30	Night	Summer
04/07/2022	07:01:26	Day	Summer
07/07/2022	06:12:55	Day	Summer
08/07/2022	05:25:07	Night	Summer
20/07/2022	18:54:02	Night	Summer
20/07/2022	18:54:54	Night	Summer
24/07/2022	17:17:16	Day	Summer
25/07/2022	16:28:52	Day	Summer
28/07/2022	15:40:18	Day	Summer
01/08/2022	14:02:34	Day	Summer
03/08/2022	18:54:48	Night	Summer
05/08/2022	12:25:20	Day	Summer
09/08/2022	10:47:20	Day	Summer
12/08/2022	09:58:18	Day	Summer
13/08/2022	09:10:07	Day	Summer
14/08/2022	14:50:13	Day	Summer
16/08/2022	08:20:56	Day	Summer
17/08/2022	07:32:49	Day	Summer
20/08/2022	06:43:14	Day	Summer
21/08/2022	05:55:11	Night	Summer
22/08/2022	11:34:43	Day	Summer
26/08/2022	09:58:03	Day	Summer
02/09/2022	07:32:33	Day	Summer
06/09/2022	05:55:39	Night	Summer
01/10/2022	13:40:52	Day	Autumn
05/10/2022	12:04:59	Day	Autumn
05/10/2022	12:05:51	Day	Autumn
09/10/2022	10:29:02	Day	Autumn
12/10/2022	09:41:02	Day	Autumn
14/10/2022	14:33:50	Day	Autumn
16/10/2022	08:04:36	Day	Autumn
20/10/2022	06:28:22	Day	Autumn
26/10/2022	09:46:18	Day	Autumn
29/10/2022	08:59:11	Day	Autumn
02/11/2022	07:23:56	Day	Autumn
01/12/2022	13:37:36	Day	Winter
10/12/2022	16:06:26	Day	Winter
14/12/2022	14:30:03	Day	Winter
16/12/2022	08:00:42	Day	Winter
16/12/2022	08:01:34	Day	Winter
18/12/2022	12:53:06	Day	Winter
18/12/2022	12:53:58	Day	Winter
28/12/2022	03:12:52	Night	Winter
01/01/2023	01:37:26	Night	Winter
31/01/2023	13:29:15	Day	Winter
09/02/2023	15:57:21	Day	Winter
10/02/2023	15:09:29	Day	Winter
12/02/2023	08:40:15	Day	Winter
15/02/2023	07:52:20	Day	Winter
16/02/2023	07:04:12	Day	Winter
26/03/2023	16:05:35	Day	Spring
03/04/2023	12:55:46	Day	Spring

Date	Time	Diurnal	Season
06/04/2023	12:08:22	Day	Spring
07/04/2023	11:20:14	Day	Spring
07/04/2023	11:21:06	Day	Spring
11/04/2023	09:45:16	Day	Spring
14/04/2023	08:57:28	Day	Spring
15/04/2023	08:09:47	Day	Spring
16/04/2023	13:50:03	Day	Spring
18/04/2023	07:21:24	Day	Spring
19/04/2023	06:32:51	Day	Spring
24/04/2023	10:36:52	Day	Spring
28/04/2023	08:58:45	Day	Spring
02/05/2023	07:21:26	Day	Spring
04/05/2023	00:52:45	Night	Spring
14/06/2023	08:34:09	Day	Spring
15/06/2023	14:13:31	Day	Spring
18/06/2023	06:55:42	Day	Spring
23/06/2023	10:58:05	Day	Summer
25/06/2023	04:28:20	Night	Summer
26/06/2023	03:40:19	Night	Summer
01/07/2023	07:41:46	Day	Summer
04/07/2023	06:50:57	Day	Summer
17/07/2023	01:16:17	Night	Summer
29/07/2023	14:45:55	Day	Summer
31/07/2023	19:38:56	Night	Summer
10/08/2023	09:55:16	Day	Summer
14/08/2023	08:20:25	Day	Summer
21/08/2023	05:58:28	Night	Summer
24/08/2023	10:03:53	Day	Summer
30/08/2023	01:58:33	Night	Summer
31/08/2023	07:38:53	Day	Summer
04/09/2023	06:02:07	Day	Summer
25/09/2023	15:38:16	Day	Autumn
29/09/2023	14:02:30	Day	Autumn
10/10/2023	10:02:32	Day	Autumn
18/10/2023	06:47:38	Day	Autumn
28/10/2023	08:29:59	Day	Autumn
03/12/2023	12:26:32	Day	Winter
11/12/2023	09:10:50	Day	Winter
15/12/2023	07:33:12	Day	Winter
18/12/2023	06:42:55	Day	Winter
19/12/2023	05:54:22	Night	Winter
22/12/2023	05:04:00	Night	Winter
30/12/2023	01:51:32	Night	Winter
14/01/2024	19:24:40	Night	Winter
21/01/2024	17:02:12	Day	Winter
22/01/2024	16:14:22	Day	Winter
04/04/2024	11:38:26	Day	Spring
12/04/2024	08:25:08	Day	Spring
19/04/2024	05:57:57	Night	Spring
28/04/2024	08:18:56	Day	Spring
02/05/2024	06:39:21	Day	Spring
29/05/2024	13:59:24	Day	Spring
13/06/2024	08:10:30	Day	Spring
15/06/2024	13:01:48	Day	Spring
17/06/2024	06:33:56	Day	Spring

Date	Time	Diurnal	Season
23/06/2024	09:52:38	Day	Summer
25/06/2024	03:23:44	Night	Summer
26/06/2024	09:04:19	Day	Summer
27/06/2024	08:16:01	Day	Summer
30/06/2024	07:28:29	Day	Summer
05/07/2024	05:06:09	Night	Summer
25/07/2024	15:23:15	Day	Summer
06/08/2024	10:33:18	Day	Summer
07/08/2024	16:13:47	Day	Summer
14/08/2024	07:20:19	Day	Summer
15/08/2024	13:00:52	Day	Summer
18/08/2024	05:43:19	Night	Summer
19/08/2024	11:23:03	Day	Summer
30/08/2024	07:21:13	Day	Summer
31/08/2024	06:33:30	Day	Summer
31/08/2024	06:34:22	Day	Summer
03/09/2024	05:47:33	Night	Summer
29/09/2024	13:08:48	Day	Autumn
03/10/2024	11:31:52	Day	Autumn
10/10/2024	09:09:56	Day	Autumn
14/10/2024	07:34:15	Day	Autumn
16/10/2024	12:27:19	Day	Autumn
18/10/2024	05:57:25	Night	Autumn
24/10/2024	09:12:54	Day	Autumn
17/11/2024	17:32:16	Day	Autumn
06/12/2024	10:16:46	Day	Winter
10/12/2024	08:38:39	Day	Winter
14/12/2024	06:59:24	Day	Winter
17/12/2024	06:09:08	Day	Winter
17/12/2024	06:10:00	Day	Winter
19/12/2024	11:00:00	Day	Winter
22/12/2024	10:10:31	Day	Winter
23/12/2024	09:21:47	Day	Winter
26/12/2024	08:32:27	Day	Winter

Santa Olalla del Cala

Date	Time	Diurnal	Season
25/05/2019	18:11:36	Night	Spring
05/07/2019	02:05:56	Night	Summer
15/07/2019	21:53:51	Night	Summer
12/11/2019	22:16:00	Night	Autumn
05/01/2020	01:18:57	Night	Winter
12/01/2020	22:08:55	Night	Winter
28/02/2020	04:09:52	Night	Winter
07/03/2020	01:01:45	Night	Winter
14/03/2020	21:54:36	Night	Winter
14/07/2020	21:25:11	Night	Summer
04/09/2020	01:03:09	Night	Summer
26/09/2020	16:32:44	Day	Autumn
29/10/2020	03:17:33	Night	Autumn
09/11/2020	22:33:40	Night	Autumn
03/01/2021	01:31:24	Night	Winter
26/01/2021	16:09:15	Day	Winter
04/03/2021	01:26:20	Night	Winter
07/03/2021	23:53:13	Night	Winter
11/03/2021	22:20:22	Night	Winter
22/03/2021	18:29:46	Night	Spring
02/05/2021	02:15:23	Night	Spring
06/05/2021	00:42:47	Night	Spring
09/05/2021	23:09:37	Night	Spring
21/05/2021	18:30:45	Night	Spring
22/06/2021	06:08:29	Day	Summer
19/07/2021	19:23:08	Night	Summer
23/07/2021	17:50:46	Day	Summer
04/08/2021	13:13:57	Day	Summer
02/09/2021	01:42:13	Night	Summer
02/12/2021	13:42:46	Day	Winter
14/12/2021	09:03:57	Day	Winter
31/12/2021	01:59:56	Night	Winter
03/01/2022	01:11:37	Night	Winter
22/01/2022	17:13:22	Day	Winter
10/02/2022	10:03:13	Day	Winter
06/03/2022	00:28:48	Night	Winter
29/04/2022	02:55:40	Night	Spring
05/06/2022	12:32:09	Day	Spring
05/06/2022	19:02:29	Night	Spring
05/06/2022	19:03:21	Night	Spring
08/06/2022	18:14:44	Night	Spring
10/07/2022	22:06:39	Night	Summer
17/07/2022	19:41:28	Night	Summer
17/08/2022	07:30:58	Day	Summer
25/08/2022	04:15:24	Night	Summer
05/09/2022	00:13:43	Night	Summer
01/10/2022	13:40:00	Day	Autumn
05/10/2022	12:04:07	Day	Autumn
02/11/2022	00:53:39	Night	Autumn
05/11/2022	23:18:19	Night	Autumn
20/11/2022	17:39:32	Day	Autumn
27/11/2022	15:13:10	Day	Autumn

Date	Time	Diurnal	Season
01/12/2022	13:35:52	Day	Winter
29/12/2022	02:23:52	Night	Winter
01/01/2023	01:36:34	Night	Winter
05/01/2023	23:11:46	Night	Winter
31/01/2023	13:27:31	Day	Winter
31/01/2023	13:28:23	Day	Winter
24/02/2023	03:51:08	Night	Winter
26/03/2023	16:04:43	Day	Spring
30/03/2023	14:28:32	Day	Spring
07/04/2023	11:19:22	Day	Spring
07/04/2023	17:49:28	Day	Spring
11/04/2023	09:43:32	Day	Spring
15/04/2023	08:08:03	Day	Spring
19/04/2023	06:31:59	Day	Spring
04/05/2023	00:51:01	Night	Spring
07/05/2023	23:14:30	Night	Spring
08/05/2023	05:44:50	Night	Spring
15/05/2023	20:00:20	Night	Spring
26/05/2023	22:27:04	Night	Spring
26/06/2023	03:38:35	Night	Summer
30/06/2023	02:00:18	Night	Summer
07/07/2023	22:45:00	Night	Summer
22/07/2023	23:39:47	Night	Summer
26/08/2023	03:34:09	Night	Summer
06/09/2023	06:01:09	Day	Summer
10/09/2023	04:25:02	Night	Summer
14/09/2023	19:32:58	Night	Summer
03/11/2023	23:34:22	Night	Autumn
10/11/2023	21:08:19	Night	Autumn
10/11/2023	21:09:10	Night	Autumn
22/11/2023	16:25:55	Day	Autumn
15/12/2023	07:31:28	Day	Winter
04/05/2024	23:18:24	Night	Spring
15/05/2024	19:05:19	Night	Spring
25/05/2024	22:06:58	Night	Spring
29/05/2024	20:28:08	Night	Spring
07/07/2024	05:06:08	Night	Summer
07/07/2024	05:07:00	Night	Summer
07/07/2024	21:47:47	Night	Summer
21/07/2024	23:29:39	Night	Summer
29/07/2024	20:13:42	Night	Summer
18/08/2024	05:41:35	Night	Summer
26/08/2024	02:25:47	Night	Summer
02/09/2024	06:35:22	Day	Summer
10/09/2024	20:11:28	Night	Summer
14/09/2024	18:38:56	Night	Summer
30/09/2024	12:18:33	Day	Autumn
23/10/2024	03:31:05	Night	Autumn
27/10/2024	01:52:27	Night	Autumn
06/11/2024	21:40:31	Night	Autumn
17/11/2024	17:30:32	Day	Autumn
24/11/2024	21:36:32	Night	Autumn
02/12/2024	18:23:03	Night	Winter
06/12/2024	16:45:37	Day	Winter

Puertollano

Date	Time	Diurnal	Season
11/02/2019	10:59:42	Day	Winter
14/02/2019	16:31:06	Day	Winter
17/02/2019	09:04:14	Day	Winter
17/02/2019	09:04:15	Day	Winter
21/02/2019	13:44:01	Day	Winter
21/02/2019	13:44:02	Day	Winter
28/02/2019	10:58:37	Day	Winter
22/05/2019	19:05:29	Night	Summer
27/06/2019	05:29:19	Night	Summer
15/07/2019	21:53:51	Night	Summer
15/07/2019	21:54:43	Night	Summer
22/07/2019	19:22:47	Night	Summer
06/08/2019	13:29:39	Day	Summer
24/09/2019	17:46:26	Day	Summer
01/10/2019	15:20:37	Day	Autumn
05/01/2020	01:18:57	Night	Winter
05/01/2020	01:19:49	Night	Winter
12/01/2020	22:08:55	Night	Winter
28/02/2020	04:09:52	Night	Winter
03/04/2020	14:10:19	Day	Summer
06/05/2020	00:57:58	Night	Summer
20/05/2020	19:28:59	Night	Summer
28/05/2020	16:20:10	Day	Summer
31/08/2020	09:07:40	Day	Summer
29/10/2020	03:17:33	Night	Autumn
18/12/2020	07:47:39	Day	Winter
03/01/2021	01:32:16	Night	Winter
11/03/2021	22:20:22	Night	Winter
06/05/2021	00:42:47	Night	Summer
06/05/2021	07:13:08	Day	Summer
13/05/2021	21:37:33	Night	Summer
22/06/2021	06:09:21	Day	Summer
05/07/2021	00:45:26	Night	Summer
08/10/2021	17:42:00	Day	Autumn
20/10/2021	13:04:54	Day	Autumn
02/12/2021	13:42:46	Day	Winter
15/12/2021	08:17:26	Day	Winter
31/12/2021	01:59:56	Night	Winter
22/01/2022	17:14:14	Day	Winter
10/02/2022	10:04:05	Day	Winter
10/02/2022	16:34:29	Day	Winter
14/02/2022	08:28:21	Day	Winter
29/04/2022	09:25:56	Day	Summer
29/05/2022	14:57:24	Day	Summer
05/06/2022	19:03:21	Night	Summer
13/06/2022	09:20:01	Day	Summer
17/06/2022	07:43:26	Day	Summer
25/06/2022	11:00:47	Day	Summer
17/07/2022	19:42:20	Night	Summer
01/10/2022	13:40:00	Day	Autumn
05/10/2022	12:04:59	Day	Autumn
09/10/2022	10:29:02	Day	Autumn

Date	Time	Diurnal	Season
05/11/2022	23:18:19	Night	Autumn
01/12/2022	13:36:44	Day	Winter
21/12/2022	12:04:47	Day	Winter
29/12/2022	08:54:10	Day	Winter
05/01/2023	23:11:46	Night	Winter
04/02/2023	18:21:49	Night	Winter
12/02/2023	08:39:23	Day	Winter
30/03/2023	14:28:32	Day	Summer
30/03/2023	14:29:24	Day	Summer
03/04/2023	12:54:54	Day	Summer
07/04/2023	11:20:14	Day	Summer
11/04/2023	09:44:24	Day	Summer
15/04/2023	08:08:55	Day	Summer
19/04/2023	06:31:59	Day	Summer
04/05/2023	00:51:53	Night	Summer
07/05/2023	23:15:22	Night	Summer
15/05/2023	20:00:20	Night	Summer
26/06/2023	03:39:27	Night	Summer
30/06/2023	02:00:18	Night	Summer
26/07/2023	22:03:32	Night	Summer
30/07/2023	20:27:26	Night	Summer
30/08/2023	08:27:48	Day	Summer
15/11/2023	02:05:59	Night	Autumn
22/11/2023	16:25:55	Day	Autumn
11/12/2023	09:10:50	Day	Winter
15/12/2023	07:32:20	Day	Winter
19/12/2023	05:53:30	Night	Winter
20/04/2024	05:08:20	Night	Summer
20/04/2024	11:38:04	Day	Summer
08/05/2024	21:39:08	Night	Summer
15/05/2024	19:05:19	Night	Summer
30/06/2024	07:28:29	Day	Summer
07/07/2024	05:07:00	Night	Summer
22/07/2024	22:41:39	Night	Summer
06/08/2024	10:32:26	Day	Summer
18/08/2024	05:42:27	Night	Summer
26/08/2024	02:26:39	Night	Summer
02/09/2024	06:35:22	Day	Summer
03/09/2024	05:47:33	Night	Summer
06/09/2024	05:02:02	Night	Summer
14/09/2024	18:38:56	Night	Summer
23/10/2024	03:31:05	Night	Autumn
03/12/2024	11:05:06	Day	Winter
06/12/2024	16:45:37	Day	Winter
14/12/2024	06:58:32	Day	Winter
14/12/2024	06:59:24	Day	Winter
18/12/2024	05:19:58	Night	Winter
22/12/2024	10:10:31	Day	Winter

Bibliography

Abu-Hamdeh, N. H. (2003). Thermal properties of soils as affected by density and water content [Published online 22 July 2003]. *Biosystems Engineering*, *86*(1), 97–102. https://doi.org/10.1016/S1537-5110(03)00112-0

Afrasiabi Gorgani, S., Panahi, M., & Rezaie, F. (2013). The relationship between NDVI and LST in the urban area of mashhad, iran [27–28 November 2013]. *Proceedings of the International Conference on Civil Engineering, Architecture & Urban Sustainable Development*.

Amato, E., Corradino, C., Torrisi, F., & Del Negro, C. (2023). A deep convolutional neural network for detecting volcanic thermal anomalies from satellite images. *Remote Sensing*, *15*(15), 3718. https://doi.org/10.3390/rs15153718

Ao, S.-I., & Fayek, H. (2023). Continual deep learning for time series modeling [Open access under Creative Commons Attribution (CC BY) license]. *Sensors*, *23*(16), 7167. https://doi.org/10.3390/s23167167

Bai, Y., Chen, M., Zhou, P., Zhao, T., Lee, J., Kakade, S., Wang, H., & Xiong, C. (2021). How important is the train-validation split in meta-learning? In M. Meila & T. Zhang (Eds.), *Proceedings of the 38th international conference on machine learning* (pp. 543–553, Vol. 139). PMLR. https://proceedings.mlr.press/v139/bai21a.html

Belcic, I., & Stryker, C. (2024, November). *What is learning rate in machine learning?* [Accessed: 2025-08-20]. https://www.ibm.com/think/topics/learning-rate#:~:text=The%20learning%20rate%20can%20determine,predictions%20and%20real%2Dworld%20data

Berra, F., D'Ambrogi, C., & Tropeano, M. (2024). Geological maps: Evolution and use of the precious knowledge hidden in a coloured landscape [Special Section – Field Mapping and Stratigraphy: Significant Insights from the Geologic Record]. *Italian Journal of Geosciences*, *143*(1), 5–24. https://doi.org/10.3301/IJG.2024.01

Brevik, E. C., Calzolari, C., Miller, B. A., Pereira, P., Kabala, C., Baumgarten, A., & Jordán, A. (2016). Soil mapping, classification, and pedologic modeling: History and future directions. *Geoderma*, *264*(Part B), 256–274. https://doi.org/10.1016/j.geoderma.2015.05.017

Buda, M., Maki, A., & Mazurowski, M. A. (2018). A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, *106*, 249–259. https://doi.org/10.1016/j.neunet.2018.07.011

Climate, W. (2025). Average monthly rainfall, temperature, sunshine — puertollano, castilla-la mancha, spain [Accessed: 2025-09-29].

Data, C. (2025). Villoslada de cameros, la rioja, spain climate data [Accessed: 2025-09-29].

Dehghani, A., Moazam, H. M. Z. H., Mortazavizadeh, F., Ranjbar, V., Mirzaei, M., Mortezavi, S., Ng, J. L., & Dehghani, A. (2023a). Comparative evaluation of lstm, cnn and convlstm for hourly short-term streamflow forecasting using deep learning approaches. *Ecological Informatics*, *75*, 102119. https://doi.org/10.1016/j.ecoinf.2023.102119

Dehghani, A., Moazam, H. M. Z. H., Mortazavizadeh, F., Ranjbar, V., Mirzaei, M., Mortezavi, S., Ng, J. L., & Dehghani, A. (2023b). Comparative evaluation of lstm, cnn, and convlstm for hourly short-term streamflow forecasting using deep learning approaches. *Ecological Informatics*, *75*, 102119. https://doi.org/10.1016/j.ecoinf.2023.102119

Desai, C. G. (2020). Comparative analysis of optimizers in deep neural networks [Accessed: 2025-07-15]. *International Journal of Innovative Science and Research Technology*, *5*(10), Available online. https://www.researchgate.net/publication/345381779

Dina, A. S., Siddique, A. B., & Manivannan, D. (2023). A deep learning approach for intrusion detection in internet of things using focal loss function [Available online 16 January 2023]. *Internet of Things*, *21*, 100699. https://doi.org/10.1016/j.iot.2023.100699

Engati. (2023). Vanishing gradient problem – definition, causes, solutions [Accessed: 2025-07-15].

Enviroliteracy Team. (2024, April). *How is soil classified?* The Environmental Literacy Council. Retrieved July 17, 2025, from https://enviroliteracy.org/how-is-soil-classified/

European Space Agency. (2013). *Sentinel-1 user handbook* [Accessed: July 2025]. European Space Agency (ESA). https://sentinel.esa.int/documents/247904/685163/Sentinel-1_User_Handbook

FAO & IUSS. (2022). World reference base for soil resources. international soil classification system for naming soils and creating legends for soil maps [4th edition]. https://www.fao.org/3/i3794en/I3794en.pdf

Fridland, V. M. (1986). *Problemy geografii, genezisa i klassifikatsii pochv* [In Russian]. Nauka.

GeeksforGeeks. (2025). What is batch normalization in deep learning? [Last updated May 3, 2025. Accessed: 2025-07-19].

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning* [http://www.deeplearningbook.org]. MIT Press.

Guha, A., & Kumar, K. V. (2014). Potential of thermal emissivity for mapping of greenstone rocks and associated granitoids of Hutti Maski schist belt, Karnataka. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XL-8*, 423. https://doi.org/10.5194/isprsarchives-XL-8-423-2014

Hallsworth, C. R., & Knox, R. W. O. (1999). *Bgs rock classification scheme volume 3: Classification of sediments and sedimentary rocks* (tech. rep. No. RR 99–03) (British Geological Survey Research Report). British Geological Survey. Keyworth, Nottingham, UK. https://www.bgs.ac.uk/downloads/start.cfm?id=2511

Han, W., Zhang, X., Wang, Y., Wang, L., Huang, X., Li, J., Wang, S., Chen, W., Li, X., Feng, R., Fan, R., Zhang, X., & Wang, Y. (2023). A survey of machine learning and deep learning in remote sensing of geological environment: Challenges, advances, and opportunities [Accessed: 2025-07-15]. *ISPRS Journal of Photogrammetry and Remote Sensing*, *202*, 87–113. https://doi.org/10.1016/j.isprsjprs.2023.05.032

Hankare, O. (2023). Convergence in deep learning [Accessed: 2025-07-19]. *Medium*. https://ompramod.medium.com/convergence-in-deep-learning-f96568923d43.

Harlianto, P. A., Adji, T. B., & Setiawan, N. A. (2017a). Comparison of machine learning algorithms for soil type classification. *2017 3rd International Conference on Science and Technology - Computer (ICST)*.

Harlianto, P. A., Adji, T. B., & Setiawan, N. A. (2017b). Comparison of machine learning algorithms for soil type classification. *2017 International Conference on Data and Software Engineering (ICoDSE)*, 1–6. https://doi.org/10.1109/ICODSE.2017.8285859

Hempel, J., Michéli, E., Owens, P., & McBratney, A. (2013). Universal soil classification system: Report from the international union of soil sciences working group. *Soil Horizons*, *54*(2), 1–6. https://doi.org/10.2136/sh13-01-0005

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Huang, S., Tang, L., Hupy, J. P., Wang, Y., & Shao, G. (2021). A commentary review on the use of normalized difference vegetation index (ndvi) in the era of popular remote sensing [© The Author(s) 2020. Corrected publication 2020]. *Journal of Forestry Research*, *32*(1), 1–6. https://doi.org/10.1007/s11676-020-01155-1

Ibáñez, J. J., & Boixadera, J. (2002). The search for a new paradigm in pedology: A driving force for new approaches to soil classification. In E. Michéli, F. O. Nachtergaele, R. J. A. Jones, & L. Montanarella (Eds.), *Soil classification 2001* (pp. 93–110). Office for Official Publications of the European Communities.

IBM Cloud Education. (n.d.). *What are labels in machine learning?* [Accessed: 2025-07-19]. https://www.ibm.com/topics/labels-machine-learning

Instituto Geográfico Nacional (IGN). (2020). Mapa geológico de España 2020 – atlas nacional de españa (ane) [CC-BY 4.0 License. Accessed July 2025].

Kriegler, F. J., Malila, W. A., Nalepka, R. F., & Richardson, W. (1969). Preprocessing transformations and their effect on multispectral recognition. *Proceedings of the Sixth International Symposium on Remote Sensing of Environment*, 97–132.

Laboratory, N. J. P. (2020). Ecostress: Ecosystem spaceborne thermal radiometer experiment on space station [Accessed: 2025-07-25].

Lasaponara, R., Abate, N., Fattore, C., Aromando, A., Cardettini, G., & Di Fonzo, M. (2022). On the use of sentinel-2 ndvi time series and google earth engine to detect land-use/land-cover changes in fire-affected areas [Published: 21 September 2022]. *Remote Sensing*, *14*(19), 4723. https://doi.org/10.3390/rs14194723

Liu, H., Wu, K., Xu, H., & Xu, Y. (2021). Lithology classification using tasi thermal infrared hyperspectral data with convolutional neural networks. *Remote Sensing*, *13*(16), 3117. https://doi.org/10.3390/rs13163117

Liu, S., Qi, Z., Li, X., & Yeh, A. G.-O. (2019). Integration of convolutional neural networks and object-based post-classification refinement for land use and land cover mapping with optical and sar data [Explores the use of RNNs (LSTM) for agricultural classification using SAR data. Emphasizes that incorporating temporal dependencies in neural network models improves classification.]. *Remote Sensing*, *11*(6), 690. https://doi.org/10.3390/rs11060690

Maitre, R. W. L. (Ed.). (2002). *Igneous rocks: A classification and glossary of terms* (2nd) [Includes IUGS nomenclature, TAS and QAPF diagram standards]. Cambridge University Press. https://doi.org/10.1017/CBO9780511535581

Marjoribanks, R. (2010). *Geological methods in mineral exploration and mining* (2nd ed.). Springer-Verlag Berlin Heidelberg. https://doi.org/10.1007/978-3-540-74375-0

MathWorks. (2023). Sequence classification using cnn-lstm network [Accessed: 2025-07-15].

Mei, S., Geng, Y., Hou, J., & Du, Q. (2022). Learning hyperspectral images from RGB images via a coarse-to-fine CNN. *Science China Information Sciences*, *65*(5), 152102. https://doi.org/10.1007/s11432-020-3102-9

Mineo, S., & Pappalardo, G. (2021). Rock emissivity measurement for infrared thermography engineering geological applications. *Applied Sciences*, *11*(9), 3773. https://doi.org/10.3390/app11093773

Ndikumana, E., Minh, D. H. T., Baghdadi, N., Courault, D., & Hossard, L. (2018). Deep recurrent neural network for agricultural classification using multitemporal SAR Sentinel-1 for Camargue, France. *Remote Sensing*, *10*(8), 1217. https://doi.org/10.3390/rs10081217

Nelson, S. A. (2000, March). *Metamorphic rocks—classification, field gradients, and facies* (EENS 212 Lecture Notes / Technical Report) (Last updated 31 March 2000). Tulane University. New Orleans, LA, USA.

Nikiforova, A. A. (2019). Soil classification. *Knowledge Organization*, *46*(4), 467–488. https://doi.org/10.5771/0943-7444-2019-6-467

Norman, J. M., & Becker, F. (1995). Terminology in thermal infrared remote sensing of natural surfaces. *Remote Sensing Reviews*, *12*, 159–173.

Olah, C. (2015). Understanding lstm networks [Accessed: 2025-07-15].

Payra, S., Sharma, A., & Verma, S. (2023). Application of remote sensing to study forest fires. In *Atmospheric remote sensing*. Elsevier. https://doi.org/10.1016/B978-0-323-99262-6.00015-8

Pedrayes, O., Lema, D., García, D., Usamentiaga, R., & Alonso, Á. (2021). Evaluation of semantic segmentation methods for land use with spectral imaging using Sentinel-2 and PNOA imagery [Academic Editor: Saeid Homayouni]. *Remote Sensing*, *13*(12), 2292. https://doi.org/10.3390/rs13122292

Prakash, A. (2000). Thermal remote sensing: Concepts, issues and applications [Working Group WG VII/3, ITC, Geological Survey Division]. *International Archives of Photogrammetry and Remote Sensing*, *33*(Part B1), 239–243. https://www.isprs.org/proceedings/XXXIII/congress/part1/239_XXXIII-part1.pdf

Rani, K., Guha, A., Pal, S. K., & Kumar, K. V. (2018). Comparative analysis of potentials of aster thermal infrared band derived emissivity composite, radiance composite and emissivity–temperature composite in geological mapping of proterozoic rocks in parts of Banswara, Rajasthan. *Journal of the Indian Society of Remote Sensing*, *46*(5), 771–782. https://doi.org/10.1007/s12524-017-0737-z

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 234–241. http://lmb.informatik.uni-freiburg.de/Publications/2015/RFB15a/

Rubio, E., Caselles, V., & Badenas, C. (1997). Emissivity measurements of several soils and vegetation types in the 8–14 μm wave band: Analysis of two field methods. *Remote Sensing of Environment*, *59*(3), 490–521. https://doi.org/10.1016/S0034-4257(96)00124-0

Salisbury, J. W., & Walter, L. S. (1989). Thermal infrared (2.5–13.5 μm) spectroscopic remote sensing of igneous rock types on particulate planetary surfaces. *Journal of Geophysical Research: Solid Earth*, *94*(B7), 9192–9202. https://doi.org/10.1029/JB094iB07p09192

Sharifi, A., & Mahdipour, H. (2024). Utilizing geospatial artificial intelligence for remote sensing applications [Editorial]. *Environmental Earth Sciences*, *83*, 274. https://doi.org/10.1007/s12665-024-11584-4

Shibuya, E., & Hotta, K. (2022). Cell image segmentation by using feedback and convolutional lstm [© The Author(s) 2021. Accepted: 18 June 2021 / Published online: 4 July 2021]. *The Visual Computer*, *38*(10), 3791–3801. https://doi.org/10.1007/s00371-021-02221-3

Shin, H., Lee, J. K., Kim, J., & Kim, J. (2017). Continual learning with deep generative replay. *arXiv preprint arXiv:1705.08690*. https://arxiv.org/abs/1705.08690

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, *15*, 1929–1958.

Srivastava, P., Shukla, A., & Bansal, A. (2021). A comprehensive review on soil classification using deep learning and computer vision techniques [Published under exclusive license to Springer Nature]. *Multimedia Tools and Applications*. https://doi.org/https://doi.org/10.1007/s11042-021-10544-5

Staff, S. S. (1999). *Soil taxonomy: A basic system of soil classification for making and interpreting soil surveys* [2nd edition]. United States Department of Agriculture, Natural Resources Conservation Service.

Streckeisen, A. L. (1974). Classification and nomenclature of plutonic rocks: Recommendations of the iugs subcommission on the systematics of igneous rocks [Introduced QAPF ("Streckeisen") diagram]. *Geologische Rundschau*, *63*(4), 773–785.

Tewari, U. (2021, November). *Regularization — understanding l1 and l2 regularization for deep learning* [Accessed: 2025-08-20]. https://medium.com/analytics-vidhya/regularization-understanding-l1-and-l2-regularization-for-deep-learning-a7b9e4a409bf

U.S. Department of Agriculture, Natural Resources Conservation Service, National Soil Survey Center. (2012, September). *Field book for describing and sampling soils, version 3.0* [Accessed: 2025-07-19]. U.S. Department of Agriculture. https://efotg.sc.egov.usda.gov/references/Public/AZ/Field_Book_for_Describing_and_Sampling_Soils-2013.pdf

U.S. Geological Survey. (2025a). What are igneous rocks? [Accessed: 2025-07-19]. https://www.usgs.gov/faqs/what-are-igneous-rocks

U.S. Geological Survey. (2025b). What are metamorphic rocks? [Accessed: 2025-07-19]. https://www.usgs.gov/faqs/what-are-metamorphic-rocks

U.S. Geological Survey. (2025c). What are sedimentary rocks? [Accessed: 2025-07-19]. https://www.usgs.gov/faqs/what-are-sedimentary-rocks

van Capel, M. (2024, July). *Urban local climate zone classification through deep learning using spatio-temporal thermal imagery* [MSc thesis]. Delft University of Technology [Supervisors: Dr. Azarakhsh Rafiee, Dr. Roderik Lindenbergh. Co-reader: Dr. Vitali Diaz Mercado. http://creativecommons.org/licenses/by/4.0/].

WeatherSpark. (2025). Average weather in santa olalla del cala, spain year-round [Accessed: 2025-09-29].

Wenjie. (2023). *Machine learning in industry: Applications and security challenges* [Dissertation]. Freie Universität Berlin.

Xavier, A. (2019). An introduction to convlstm [Published on Medium, Neuronio. Accessed: 2025-07-15].

Yamashita, R., Nishio, M., Do, R. K., & Togashi, K. (2018). Convolutional neural networks: An overview and application in radiology. *Insights into Imaging*, *9*, 611–629. https://doi.org/10.1007/s13244-018-0639-9

Yan, S. (2016). Understanding lstm and its diagrams [Accessed: 2025-07-15].

Ye, X., Ren, H., Nie, J., Hui, J., Jiang, C., Zhu, J., Fan, W., Qian, Y., & Liang, Y. (2022). Simultaneous estimation of land surface and atmospheric parameters from thermal hyperspectral data using a lstm-cnn combined deep neural network. *IEEE Geoscience and Remote Sensing Letters*, *19*, 5508705. https://doi.org/10.1109/LGRS.2022.3187042

Yin, Q., Zhang, R., & Shao, X. (2019). CNN and RNN Mixed Model for Image Classification. *MATEC Web of Conferences, Proceedings of JCMME 2018*, *277*, 02001. https://doi.org/10.1051/matecconf/201927702001

Ying, X. (2019). An overview of overfitting and its solutions [CISAT 2018]. *Journal of Physics: Conference Series*, *1168*(2), 022022. https://doi.org/10.1088/1742-6596/1168/2/022022

Zhu, X., Gao, Z., Chen, T., Wang, W., Lu, C., & Zhang, Q. (2022). Study on the thermophysical properties and influencing factors of regional surface shallow rock and soil in china. *Frontiers in Earth Science*, *10*, 864548. https://doi.org/10.3389/feart.2022.864548

Further Reading

Akter, M., Miah, M. A., Hassan, M. M., Mobin, M. N., & Baten, M. A. (2015). Textural influence on surface and subsurface soil temperatures under various conditions. *Journal of Environmental Science and Natural Resources*, *8*(2), 147–151.

Grechi, G., Fiorucci, M., Marmoni, G. M., & Martino, S. (2021). 3d thermal monitoring of jointed rock masses through infrared thermography and photogrammetry. *Remote Sensing*, *13*(5), 957. https://doi.org/10.3390/rs13050957

Rockwell, B. W., & Hofstra, A. H. (2008). Identification of quartz and carbonate minerals across northern nevada using aster thermal infrared emissivity data—implications for geologic mapping and mineral resource investigations in well-studied and frontier areas. *Geosphere*, *4*(1), 218–246. https://doi.org/10.1130/GES00126.1

Shirmard, H., Farahbakhsh, E., Heidari, E., Beiranvand Pour, A., Pradhan, B., Müller, D., & Chandra, R. (2022). A comparative study of convolutional neural networks and conventional machine learning models for lithological mapping using remote sensing data. *Remote Sensing*, *14*(4), 819. https://doi.org/10.3390/rs14040819

Colophon

This document was typeset using LaTeX, using the KOMA-Script class `scrbook`. The main font is Palatino.