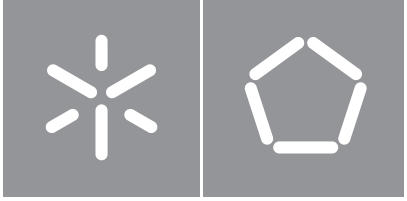


Universidade do Minho
Escola de Engenharia

Hugo Novais

Thematic Forest Species Map In the Iberian Peninsula Using Artificial Intelligence and Remote Sensing Data

October, 2025



Universidade do Minho
Escola de Engenharia

Hugo Novais

Thematic Forest Species Map In the Iberian Peninsula Using Artificial Intelligence and Remote Sensing Data

Master's Dissertation
Master's in Aerospace Engineering

Work supervised by
Paulo Novais
Alexandre Silva

October, 2025

COPYRIGHT AND TERMS OF USE OF THIS WORK BY A THIRD PARTY

This is academic work that can be used by third parties as long as internationally accepted rules and good practices regarding copyright and related rights are respected.

Accordingly, this work may be used under the license provided below.

If the user needs permission to make use of the work under conditions not provided for in the indicated licensing, they should contact the author through the RepositóriUM of Universidade do Minho.

License granted to the users of this work



Creative Commons Atribuição-NãoComercial-CompartilhaIgual 4.0 Internacional
CC BY-NC-SA 4.0

https://creativecommons.org/licenses/by-nc-sa/4.0/deed.pt

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the Universidade do Minho.

_____, _____

(Location) (Date)

_____

(Hugo Novais)

Acknowledgements

I would like to express my gratitude to my supervisor and co-supervisor, Prof. Paulo Novais and Prof. Alexandre Silva, for their continuos support, guidance, and encouragment throughout the development of this thesis.

I would also like to thank GEOSAT for taking me in as an intern, helping me daily, and providing me with the resources necessary to complete this work.

Finally, I am deeply grateful to my family and friends for their encouragement, patience, and unwavering support throughout this journey.

Resumo

Mapeamento de Espécies Florestais na Península Ibérica com Recurso a Inteligência Artificial e Deteção Remota

Os métodos atuais de mapeamento de espécies arbóreas são trabalhosos, dispendiosos e espacialmente limitados, dependendo fortemente de inventários de campo extensos ou de imagens de alta resolução com cobertura reduzida. Este estudo apresenta um modelo híbrido de *deep learning*, desenvolvido para a classificação de espécies arbóreas em grande escala na Península Ibérica, utilizando imagens Sentinel-2 integradas com dados de campo do Inventário Florestal Nacional de Espanha, bem como variáveis topográficas e climáticas.

Entre as configurações testadas, o modelo com oito classes apresentou o melhor desempenho, atingindo uma precisão de 80,49% e um *F1-score* de 80,01%. Estes resultados superam as abordagens tradicionais de *machine learning* e demonstram o potencial do *deep learning* para mapeamento florestal. A avaliação baseada em confiança confirmou ainda que as previsões com maior certeza correspondem fortemente às classificações corretas, permitindo uma filtragem fiável para uso prático.

O modelo treinado foi utilizado para gerar mapas de espécies arbóreas em Espanha e estendido a Portugal sob o mesmo enquadramento, produzindo resultados espacialmente coerentes e ecologicamente consistentes. De modo geral, este trabalho estabelece uma abordagem escalável, reprodutível e economicamente eficiente para a monitorização florestal em larga escala, contribuindo para a avaliação da biodiversidade e para uma gestão florestal mais sustentável.

Palavras-chave: deep learning, machine learning, remote sensing, tree species classification

Abstract

Thematic Forest Species Map In the Iberian Peninsula Using Artificial Intelligence and Remote Sensing Data

Current methods for tree species mapping are labour-intensive, costly, and spatially limited, relying heavily on extensive field surveys or high-resolution imagery. This study presents a hybrid deep learning model designed for large-scale tree species classification across the Iberian Peninsula using Sentinel-2 imagery integrated with ground truth data from the Spanish National Forest Inventory, as well as topographic and climatic variables.

Among the tested configurations, the eight-class model achieved the highest performance, reaching an overall accuracy of 80.49% and an F1-score of 80.01%. These results surpass traditional machine learning benchmarks and demonstrate the potential of hybrid deep learning for operational forest mapping. Confidence-based evaluation further confirmed that predictions with higher certainty closely align with correct classifications, allowing reliable post-filtering for practical use.

The trained model was used to generate tree species maps for Spain and extended to Portugal, producing spatially coherent and ecologically consistent outputs. Overall, this work establishes a scalable, reproducible, and cost-effective approach for continental-scale forest monitoring, contributing to improved biodiversity assessment and sustainable forest management.

Keywords: deep learning, machine learning, remote sensing, tree species classification

Contents

List of Figures x

List of Tables xii

Acronyms xiv

1 Introduction 1

1.1 Contextualization . . . 1
1.2 Problem Statement . . . 2
1.3 Goals and Research Questions . . . 3
1.4 Methodology . . . 4
1.5 Thesis Structure . . . 5

2 Literature Review 7

2.1 Artificial Intelligence and Machine Learning: Foundations and Techniques . . . 7
2.1.1 Artificial Intelligence . . . 7
2.1.2 Machine Learning . . . 8
2.1.3 Deep Learning . . . 9
2.1.4 Types of Models of Machine Learning . . . 10
2.2 Remote Sensing . . . 14
2.2.1 Types of Remote Sensing . . . 14
2.2.2 Applications and Challenges in Remote Sensing . . . 15
2.3 Synergy of Machine Learning and Remote Sensing in Tree Species Classification . . . 16
2.3.1 Review Approach . . . 16
2.3.2 Classification Methods . . . 17
2.3.3 Remote Sensing Systems . . . 19
2.3.4 Literature Trends in Tree Species Classification . . . 21
2.3.5 Limitations and Possible Future Work . . . 23
2.3.6 Summary of the Literature Review . . . 24

3 Materials **25**
3.1 Study Area . . . 25
3.2 Data Understanding . . . 26
3.2.1 Sentinel-2 Images . . . 26
3.2.2 Ground Truth Data . . . 28
3.2.3 Auxiliary Data . . . 28
3.3 Data Preprocessing . . . 29
3.3.1 Retrieve Sentinel-2 Images . . . 29
3.3.2 Filter Images Without Plots . . . 31
3.3.3 Crop Plot Images . . . 31
3.3.4 Extract Data from IFN4 Files . . . 32
3.3.5 Convert Coordinates . . . 32
3.3.6 Filter Forest Regions . . . 33
3.3.7 Add Elevation and Climate Data . . . 34
3.3.8 Remove Incongruencies . . . 34
3.3.9 Add Plots Image Path . . . 34
3.3.10 Portuguese Land Cover Data . . . 34
3.4 Data Analysis . . . 35
3.5 Summary of the Materials Chapter . . . 38

4 Methods **40**
4.1 Evaluation Metrics . . . 40
4.2 Architecture Configuration . . . 41
4.2.1 Data Handling and Preprocessing Pipeline . . . 41
4.2.2 Feature Preparation . . . 42
4.2.3 Model Architecture . . . 42
4.3 Architectures Tested . . . 45
4.4 Benchmarks Configuration . . . 47
4.4.1 Generation of Qualitative Tree Species Maps . . . 48
4.5 Conclusion of the Methods Chapter . . . 49

5 Results **50**
5.1 Optimal Number of Classes . . . 50
5.2 Architecture Evaluation Results . . . 55
5.3 Benchmark Comparison . . . 56
5.4 Ablation Studies . . . 56
5.5 Qualitative Application Across Iberian Regions . . . 57

5.5.1 Application over Spanish Forests . . . 57
5.5.2 Application over Portuguese Forests . . . 60
5.6 Conclusion of the Results Chapter . . . 62

6 Discussion **63**
6.1 Overview of Key Results . . . 63
6.1.1 Multiple Class Models . . . 64
6.1.2 Main Deep Learning Model vs Benchmark Models . . . 64
6.1.3 Discussion of Architecture Evaluation Results . . . 65
6.1.4 The Role of Image and Auxiliary Data in a Hybrid Model . . . 66
6.1.5 Validation with Portuguese Data . . . 66
6.2 Comparison with Existing Literature . . . 67
6.3 Practical Implications . . . 67
6.4 Limitations of the Study . . . 68
6.5 Future Research Directions . . . 69

7 Conclusion **70**
7.1 Summary of Goals and Research Questions . . . 70
7.2 Key Findings and Interpretation . . . 71
7.3 Contributions to the Field . . . 72
7.4 Limitations and Outlook . . . 72
7.5 Future Directions and Broader Impact . . . 72
7.6 Final Remarks . . . 73

Bibliography **74**

List of Figures

1 Methodology fluxogram 5

2 Structure of an artificial neuron (adapted and redrawn from [32]) 9
3 Example of a decision tree 10
4 Example of a 2-dimensional Support Vector Machine (SVM) (reproduced from [34], MDPI, licensed under CC BY 4.0) 11
5 Example of an ensemble learning model (reproduced from [35], IEEE Access, licensed under CC BY 4.0) 11
6 Random forest illustration (reproduced from [37], Hindawi, licensed under CC BY 4.0) . . 12
7 Extreme Gradient Boosting (XGBoost) illustration (reproduced from [39], Asian Journal of Civil Engineering, licensed under CC BY 4.0) 13
8 Example of a Convolutional Neural Network (CNN) architecture (reproduced from [42], Elsevier, licensed under CC BY 4.0) 14
9 PRISMA flow diagram illustrating the study selection process for the systematic review . . 16
10 Percentage of studies using traditional machine learning (TML) methods versus deep learning (DL) methods 17
11 Percentage of studies with different TML algorithms 18
12 Percentage of studies with different DL models 19
13 Frequency of usage of each satellite remote sensing (RS) type 20
14 Percentage of studies using uni-sensor or multi-sensor data 21

15 Land cover of natural forests in the Iberian Peninsula [67] 26
16 Fluxogram of the data preprocessing 30
17 Example of a tile image from the Iberian Peninsula 30
18 Example of the tiling removal strategy results 31
19 Example of a 10x10 pixel plot image 32
20 Example image without spatial filter 33
21 Example image with spatial filter 33
22 Bar plot for the species distribution 36

23 Pie chart of the 10 most common species . . . 36
24 Bar plot of the genus distribution . . . 37
25 Balanced pie chart of the genus distribution . . . 37
26 Plot distribution over topography of Spain . . . 38

27 Hybrid architecture . . . 44
28 Class distribution for the 4k species - 1.5k genus model . . . 46
29 Class distribution for the 4k species - 0.5k genus model . . . 46
30 Class distribution for the 2.5k species - 0.5k genus model . . . 47

31 Class Distribution for the Top 25 Species . . . 51
32 Per-Class overall accuracy (OA) for the 25-Class Model . . . 52
33 Per-Class OA for the 15-Class Model . . . 52
34 Normalized Confusion Matrix for the 15-Class Model . . . 53
35 Per-Class Accuracy Across Confidence Ranges for the 15-Class Model . . . 54
36 Accuracy vs Confidence for the 15-Class Model . . . 54
37 Sentinel-2 Image of the Trasbar region . . . 58
38 LULUCF dense forest filter applied to the Trasbar region . . . 58
39 True Label tree species (TS) map for the Trasbar region . . . 59
40 Predicted TS map for the Trasbar region . . . 59
41 Model confidence map corresponding to the Trasbar region . . . 60
42 Predicted TS map for Serras de Fafe (Portugal). . . . 61
43 Model confidence map for the Serras de Fafe region (Portugal). . . . 61

List of Tables

1 Spectral bands and characteristics of Sentinel-2 imagery . . . 27
2 Summary of vegetation indices used in this study . . . 27
3 Columns used from the 4th Spanish National Forest Inventory (IFN4) database . . . 28
4 Auxiliary features added to the dataset . . . 29
5 Columns present on the dataframe used for training . . . 35
6 Main values of the auxiliary data . . . 35

7 Architecture of the tabular branch . . . 43
8 Architecture of the fusion classifier . . . 44
9 Optimal hyperparameter configurations for the benchmark algorithms . . . 48

10 Results for the species models from 8 to 25 classes . . . 51
11 Seasonal Models Results . . . 55
12 Genus Models Results . . . 55
13 Autonomous Regions Models Results . . . 55
14 Hybrid Genus–Species Models Results . . . 56
15 Results for the benchmark models . . . 56
16 Results for the ablation studies models . . . 57

Acronyms

AI artificial intelligence

ANN artificial neural network

API Application Programming Interface

CNN Convolutional Neural Network

COS Land Use and Occupancy Charter

CRS coordinate reference system

DEM Digital Elevation Model

DGT General Administration for Territory

DL deep learning

EO Earth Observation

HSI hyperspectral images

IFN4 4th Spanish National Forest Inventory

L2A Level-2A

LiDAR Light Detection and Ranging

LULUCF Land Use, Land-Use Change, and Forestry

MFE Spanish forest map

MITECO Spanish Ministry for Ecological Transition and the Demographic Challenge

ML machine learning

MSI multispectral images

NDVI Normalized Difference Vegetation Index

OA overall accuracy

ReLU Rectified Linear Unit

RF Random Forest

RGB red, green, blue

RS remote sensing

SAR Synthetic Aperture Radar

SVM Support Vector Machine

TML traditional machine learning

TS tree species

UM University of Minho

UTM Universal Transverse Mercator

VHR very-high resolution

VI vegetation index

XGBoost Extreme Gradient Boosting

Chapter 1

Introduction

This chapter introduces the research topic, providing context and background information on the classification of TS using RS and machine learning (ML) techniques. It outlines the problem statement, research objectives, and key questions that guide the study. Additionally, the chapter presents an overview of the methodology employed and the structure of the thesis.

1.1 Contextualization

In recent years, RS technology has made significant progress, particularly in the diversification of available sensors, the enhancement of spatial and spectral resolutions, and the expansion of data acquisition capabilities. These advancements have significantly increased the precision and applicability of Earth Observation (EO) technology across various domains, including forestry, agriculture, and environmental monitoring. Modern sensors, such as those on Sentinel-2 and WorldView, now offer multispectral and hyperspectral imagery with unprecedented resolutions [1]. Meanwhile, innovations in active RS technologies, including Light Detection and Ranging (LiDAR) and Synthetic Aperture Radar (SAR), provide detailed information on terrain morphology, vegetation structures, and surface characteristics [2].

With these advancements, the classification of TS has become a key area of interest. This is evident in surveys like Felbermeier et al. [3], where forest professionals ranked TS information as the highest priority among 60 parameters for RS applications. Accurate and up-to-date forest TS maps are essential for numerous ecological, forestry, and environmental applications. Within forestry, TS classification supports forest inventories [4], biodiversity monitoring [5], invasive species management [6], and sustainable forest management. In ecology, spatially explicit TS information deepens the understanding of interactions between TS and their habitats [7, 8]. Likewise, environmental applications rely on TS maps for wildlife

habitat estimation [9], biodiversity monitoring, and the assessment of forest insect abundance [10, 11].

Traditional field-based surveys for TS mapping are labor-intensive, time-consuming, expensive, and often infeasible for large-scale or inaccessible areas [12–14]. Consequently, RS has emerged as a practical and cost-effective alternative for obtaining detailed forest information across extensive regions [15]. Over the last few decades, advancements in RS technology have facilitated the retrieval of diverse forest information, including species composition, structure, and health [2].

The growing interest in TS classification, as reflected in the increasing number of related studies over the past 25 years, stems not only from rapid advancements in RS technologies but also from the development of powerful computational tools [2]. The combination of high-resolution satellite imagery with ML algorithms provides unprecedented potential for generating fine-scale, spatially explicit TS maps. These maps improve ecological understanding and enhance the precision of resource inventories, carbon stock estimation, and forest health monitoring [16].

Despite these advancements, challenges remain. Existing TS maps often rely on coarse vegetation data or species distribution models, which provide potential ranges rather than precise geolocations [16]. The growing demand for accurate TS maps with finer spatial resolutions underscores the critical role of RS in forestry applications. RS offers scalable and robust solutions that can substantially reduce dependency on traditional field-based surveys.

1.2 Problem Statement

The Iberian Peninsula, characterized by its rich and diverse forest ecosystems, plays a critical role in regional biodiversity, carbon sequestration, and ecological balance. Accurate, high-resolution mapping of TS is essential for supporting sustainable forest management, biodiversity conservation, and ecological monitoring. However, existing forest inventories predominantly rely on traditional field-based methods that are labor-intensive, costly, and spatially constrained, making them impractical for large-scale or remote areas.

Although RS has shown great promise in TS classification by leveraging advances in multispectral, hyperspectral, and LiDAR sensors, significant gaps in their application remain. Current approaches often face challenges in achieving the spatial resolution, accuracy, and scalability required to generate comprehensive, species-specific maps across heterogeneous landscapes, like the Iberian Peninsula. These limitations hinder the ability to accurately quantify forest composition, monitor ecosystem health, and inform adaptive forest management practices.

Addressing these challenges is not only critical for advancing forest management and conservation within the Iberian Peninsula but also for contributing to global efforts in biodiversity preservation, climate change mitigation, and sustainable development. The development of innovative, high-resolution species

classification methods using modern RS technologies offers a pathway to bridging these gaps and enabling more effective decisions to be made in forestry and environmental planning.

After analysing current market trends in the satellite industry and identifying the technical skills required for its development, it was agreed, in collaboration with GEOSAT, that the proposed project was both feasible and capable of generating real value for end users. The growing interest in operational applications of EO and artificial intelligence (AI) for forestry and environmental monitoring provided a strong backbone for this research. Consequently, this thesis was designed to address a relevant and emerging challenge within the sector: developing scalable and accurate TS classification methods that leverage open-access satellite data to support sustainable forest management.

1.3 Goals and Research Questions

The primary objective of this study is to develop a DL model capable of classifying and generating TS maps across the entire Iberian Peninsula. This will be achieved using ground truth data from Spain and multispectral satellite imagery from Sentinel-2 to generate accurate, spatially detailed TS maps. To achieve this goal, several sub-objectives have been identified:

1. Satellite Image Retrieval: Acquire satellite images corresponding to all ground truth sample locations for each season (four in total), ensuring minimal cloud cover to maximize data usability.

2. Integrate Tabular Data: Create a dataset that integrates both IFN4 ground truth data and relevant auxiliary data.

3. Model Development: Build and train a DL model capable of classifying more than 90% of the TS present in the IFN4 files with high accuracy and confidence.

4. Assess Model Performance: Compare DL results with TML approaches, identifying weaknesses and potential avenues for model enhancement for future research. Validate the results with quantitative metrics through independent test data by generating TS classification maps using Spain's data as ground truth.

5. Generate Portuguese TS Classification Map: Use the DL models to generate accurate TS maps in Portuguese forests with high confidence levels.

In addition to the goals, several key research questions will be addressed by the conclusion of this study:

- How does RS data improve the accuracy and efficiency of TS classification in the Iberian Peninsula?

- What are the key challenges and limitations in applying Sentinel-2 images for TS classification in ecologically diverse regions like the Iberian Peninsula?
- How do different ML and DL algorithms perform in classifying TS by leveraging satellite imagery and ground truth data?
- How can TS classification maps be generated, ensuring a high overall level of accuracy?

1.4 Methodology

The methodology follows an iterative process of model development, evaluation, and refinement, designed to ensure reliable classification results. Ultimately, this approach will produce probability maps that are both accurate and scalable, suitable for the objectives of this study. Figure 1 provides a fluxogram summarizing the key steps of the research workflow.

- Literature Revision: A comprehensive review of scientific literature, technical documentation, and existing methodologies to identify research gaps, select the algorithm, and define evaluation metrics.
- Ground Truth Data from Spain: The dataset from Spain serves as the reference base for model training, validation, and performance assessment.
- Sentinel-2 Images: High-resolution multispectral data from Sentinel-2 will be processed to extract relevant spectral and spatial features necessary for TS classification.
- Extract Auxiliary Features: Given the dataset's complexity, auxiliary features will be generated and refined to enhance model performance, including necessary data cleaning and validation steps.
- Extract Remote Sensing Features: Preprocessing steps in the satellite data, including band selection, vegetation index (VI) computation, and texture feature extraction to represent spectral and structural properties effectively.
- Train Classification Models: Using the processed datasets, multiple ML and DL models will be developed and trained for TS classification.
- Evaluate Models: Performance will be assessed using standard metrics such as OA, precision, recall, and F1-score.
- Are the Results Acceptable? This decision point determines whether the models meet predefined accuracy thresholds. If the results are not acceptable, adjustments are made to the model parameters or feature selection in the "Refine and Adjust Parameters"step.

- Refine and Adjust Parameters: If the models perform poorly, steps are taken to improve them. This may involve tweaking hyperparameters, using alternative algorithms, adding new features, or incorporating more ground truth data.
- Tree Species Classification Model: The outcome will be a DL model capable of producing reliable, fine-scale classification maps for the Iberian Peninsula.

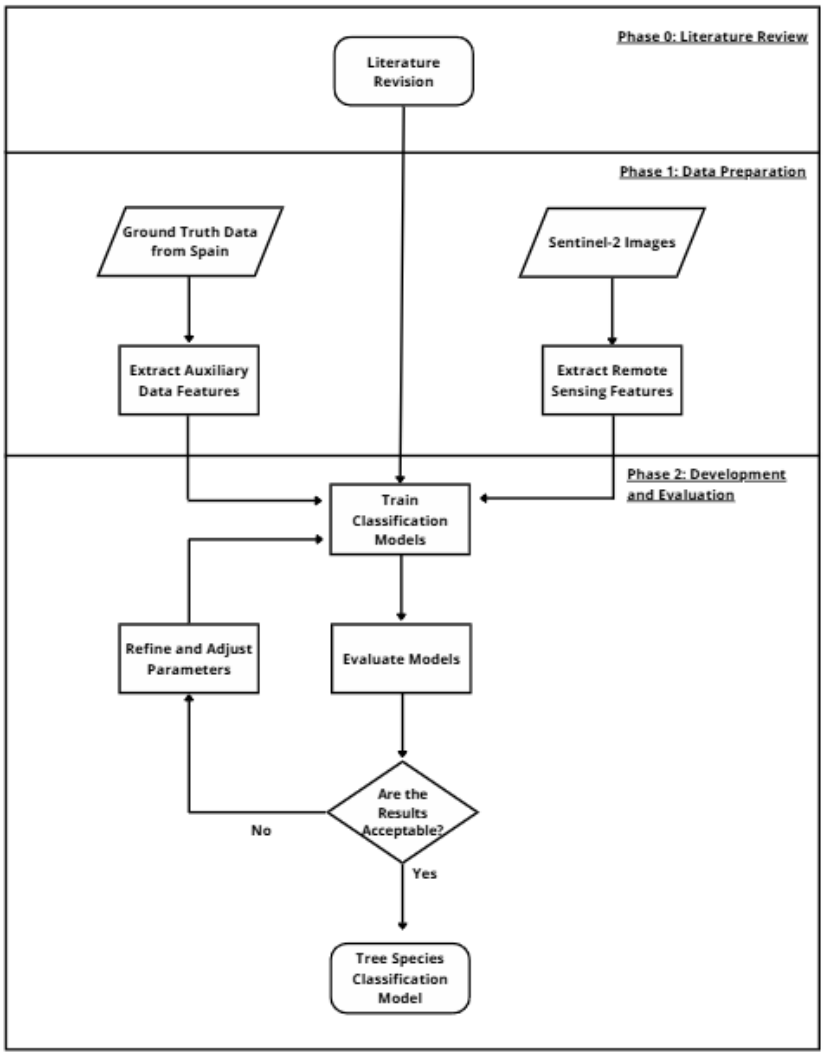


Figure 1: Methodology fluxogram

1.5 Thesis Structure

This thesis is organized into seven chapters, each addressing a key aspect of the research process, from contextual motivation to final conclusions. The structure is as follows:

- Chapter 1 - Introduction: Introduces the research context and motivation, presenting the problem statement, objectives, and research questions. It also provides an overview of the significance of TS classification and sets the stage for the investigation of ML and RS technologies.
- Chapter 2 - Literature Review: Reviews the state of the art in ML and RS for TS classification. It synthesizes the main findings of 33 selected studies, compares methodologies and sensors, and identifies knowledge gaps that guided this research.
- Chapter 3 - Materials: Describes the study area (Iberian Peninsula) and all datasets used, including Sentinel-2 imagery, ground truth data from IFN4, and auxiliary datasets (topography, climate, and Portuguese Land Use and Occupancy Charter (COS) data). It also details preprocessing steps such as image retrieval, filtering, tiling, and data cleaning.
- Chapter 4 - Methods: Explains the methodological framework adopted for the study. It describes the architecture of the hybrid DL model, data handling procedures, feature preparation, training configurations, and evaluation metrics. The chapter also details the benchmark models and the process used to generate TS maps for Spain and Portugal.
- Chapter 5 - Results: Presents the experimental outcomes and analyses. It compares different architectures, evaluates performance metrics, and assesses the influence of single vs. hybrid modalities. The chapter concludes with the generation and interpretation of TS and confidence maps for both Spain and Portugal.
- Chapter 6 - Discussion: Interprets the findings in light of the research objectives and existing literature. It examines model performance trends, cross-border validation results, practical implications, and study limitations, while proposing potential improvements and future directions.
- Chapter 7 - Conclusion: Summarizes the main findings and answers the research questions. It reflects on the study's contributions to RS and forestry applications, outlines recommendations for future research, and emphasizes the practical value of hybrid DL approaches for large-scale ecological monitoring.

Chapter 2

Literature Review

The present chapter provides a comprehensive overview of the foundational concepts and recent advancements in the fields of AI, ML, DL, and RS. It begins by defining AI and ML, elucidating their core principles and methodologies. The chapter then delves into the various types of ML models, highlighting their unique characteristics and applications. Subsequently, it explores the realm of remote sensing, discussing its types, applications, and challenges. Finally, the chapter reviews the synergy between ML and RS in the context of TS classification, emphasizing the latest research trends and methodologies employed in this domain.

2.1 Artificial Intelligence and Machine Learning: Foundations and Techniques

2.1.1 Artificial Intelligence

AI does not have a universally accepted definition. However, one way to define it is through the words of John McCarthy, widely regarded as the father of AI, who described it as "The science and engineering of making intelligent machines"[17].

Alan Turing was the first to consider the question "Can machines think?"[18], but it was only in 1956 that the term "Artificial Intelligence"was first proposed [19]. Since then, several developments have been made in the field, reaching the level that we are used to seeing today.

The traditional goals of AI include natural language processing, which enables machines to understand and generate human language [18–20]; learning, where systems improve their performance based on data or experience [20]; problem-solving, which involves finding solutions to complex or computationally

challenging issues [20]; and decision-making, where systems evaluate options and choose the best course of action [21]. These goals shape the development of AI and contribute to making its behaviors increasingly human-like.

To achieve these objectives, AI is typically divided into specialized fields and methodologies. Among these, the ability to learn from data stands out as one of the most transformative aspects of AI. This is where ML plays a central role, providing tools to develop models that adapt and improve through experience [22]. ML has revolutionized AI by enabling systems to move beyond rigid, pre-programmed rules, allowing them to extract patterns and insights from data. This paradigm shift has greatly increased AI's applicability in a variety of fields, increasing its significance in daily life.

2.1.2 Machine Learning

ML is a subset of AI that focuses on developing algorithms and statistical models to enable machines to learn from data and improve their performance on tasks through experience rather than relying on explicitly programmed instructions [23]. At its core, ML uses that data to train models, which can then make predictions or decisions when presented with new, unseen information.

The significance of ML lies in its ability to process vast amounts of data, uncover patterns and insights, and make decisions without human intervention [24]. This capability has made ML indispensable across numerous fields, including natural language processing, computer vision, healthcare, and environmental monitoring.

ML can be categorized into three primary types based on the nature of the learning process: supervised learning, unsupervised learning, and reinforcement learning. Each type addresses different kinds of tasks and utilizes different approaches to train models.

Supervised learning is the most commonly used type of ML. In this approach, the model is trained on a dataset that includes input-output pairs. Each input (feature) is associated with a corresponding correct output (label), and the model learns to map inputs to outputs by minimizing errors [25]. Once trained, the model can predict outputs for new, unseen data. Supervised learning requires labeled data for training and heavily depends on the quantity and quality of that data [26]. It is typically used for tasks like regression (predicting continuous values) and classification (categorizing data into distinct classes) [27].

Unsupervised learning, on the other hand, deals with datasets where no labels or outputs are provided [28]. The goal is for the model to identify hidden patterns, groupings, or structures in the data without explicit guidance [28]. This approach is often used for exploratory data analysis or for making sense of large, unstructured datasets [29]. Unlike supervised learning, unsupervised learning relies on the intrinsic properties of the data to uncover insights [28]. It is typically used for tasks such as clustering (grouping similar data points) and dimensionality reduction (simplifying datasets while preserving essential information).

Reinforcement learning is a distinct paradigm where an agent learns to interact with its environment

by performing actions and receiving feedback in the form of rewards or penalties [30]. The agent's goal is to maximize cumulative rewards by optimizing its actions over time. This form of sequential decision-making and trial-and-error learning requires an environment where the agent's actions influence future states [30]. The main objective is to learn a policy, which is essentially a map that defines the best action to take in each state.

2.1.3 Deep Learning

DL is a subfield of ML that focuses on algorithms inspired by the structure and function of the human brain [31]. DL models are built using layers of interconnected nodes called artificial neurons. These layers form an artificial neural network (ANN), where each layer learns progressively more abstract features of the input data. The purpose of these models is to automatically extract relevant features from raw data, eliminating the need for manual feature engineering [31]. This makes DL particularly useful in fields where the data is too high-dimensional or unstructured for traditional models, such as images, audio, and text.

The fundamental unit of a DL model is the artificial neuron. It receives inputs, multiplies them by a certain weight, adds a bias term, passes them through an activation function, and in this way produces an output, as can be seen in Figure 2.

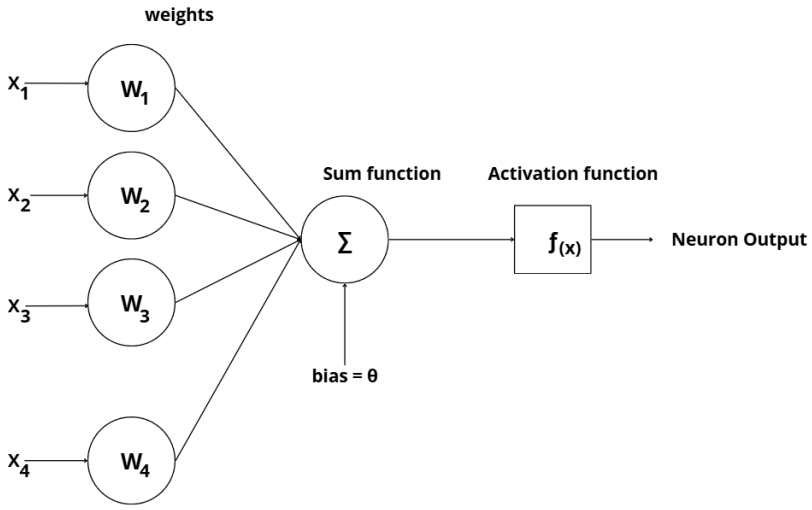


Figure 2: Structure of an artificial neuron (adapted and redrawn from [32])

The artificial neurons are organized into interconnected layers, which typically consists of:

- Input Layer: Receives raw data.
- Hidden Layers: Transform the data.
- Output Layer: Produces the final result.

These connections form the ANN. In each forward pass, data flows from the input layer, through the hidden layers, to the output layer.

2.1.4 Types of Models of Machine Learning

2.1.4.1 Decision Tree

Many traditional ML models are built upon the concept of decision trees. Each node on the tree represents a decision on a specific feature, and each branch represents the result of that test. It repeats this process until reaching a leaf node that represents either a class label, for classification problems, or continuous value for regression problems. An example structure of a decision tree can be seen in Figure 3.

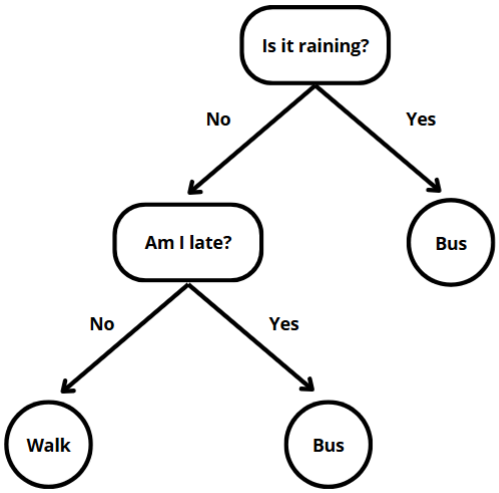


Figure 3: Example of a decision tree

The main problem with decision trees is that a single one can easily lead to overfitting, which fails in generalizing new data. This led to the development of ensemble methods.

2.1.4.2 Support Vector Machine

Founded on statistical learning theory, SVMs operate by identifying a decision boundary, known as a hyperplane, that separates data points of different classes. The optimal hyperplane is the one that maximizes the distance, known as the margin, between itself and the nearest data points from any class. These nearest data points are called support vectors. It is effective in high-dimensional spaces and works particularly well when the number of dimensions is greater than the number of samples [33]. This can be seen as an example in Figure 4.

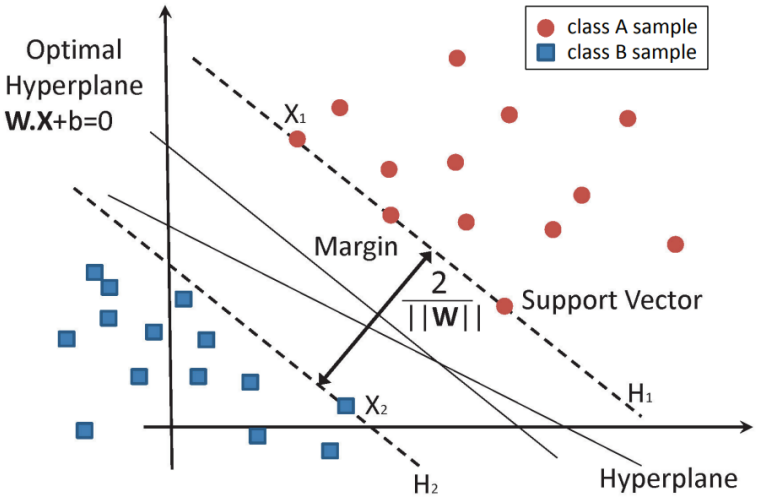


Figure 4: Example of a 2-dimensional SVM (reproduced from [34], MDPI, licensed under CC BY 4.0)

2.1.4.3 Ensemble Learning

Ensemble learning is a higher-performing complex model built from multiple weak learners. Its premise is based on the thinking that multiple experts are better than a single expert.

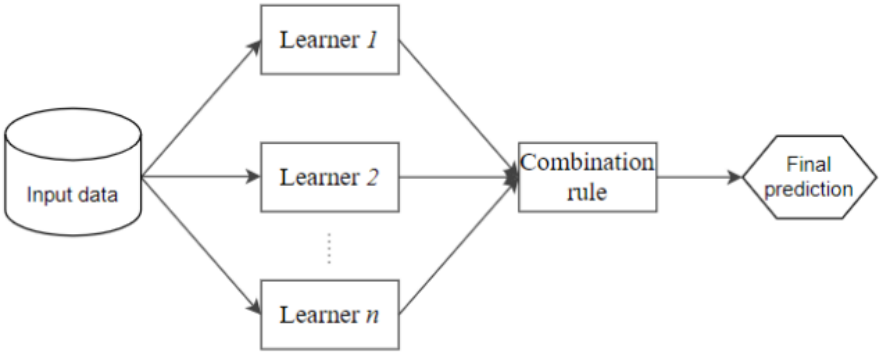


Figure 5: Example of an ensemble learning model (reproduced from [35], IEEE Access, licensed under CC BY 4.0)

To fuse the weak model predictions and obtain a more reliable final output, several techniques can be used, including bagging, boosting, and stacking, leading to new ML models. Bagging primarily reduces variance by averaging decorrelated learners, whereas boosting reduces bias by sequentially correcting errors. In a simple way, ensemble learning can be visualized as it is depicted in Figure 5.

2.1.4.4 Random Forest

One example of ensemble learning is the Random Forest (RF) algorithm. It is an ensemble of decision trees, trained using the bagging technique. It works by training many trees in parallel. Each is trained on a different random subset of the training data, obtained through bootstrapping. At each split, a random subset of features is considered. This decorrelates the trees, ensuring different errors from each. To make a prediction, each tree votes, and the majority vote wins [36]. Figure 6 illustrates an example of this algorithm.

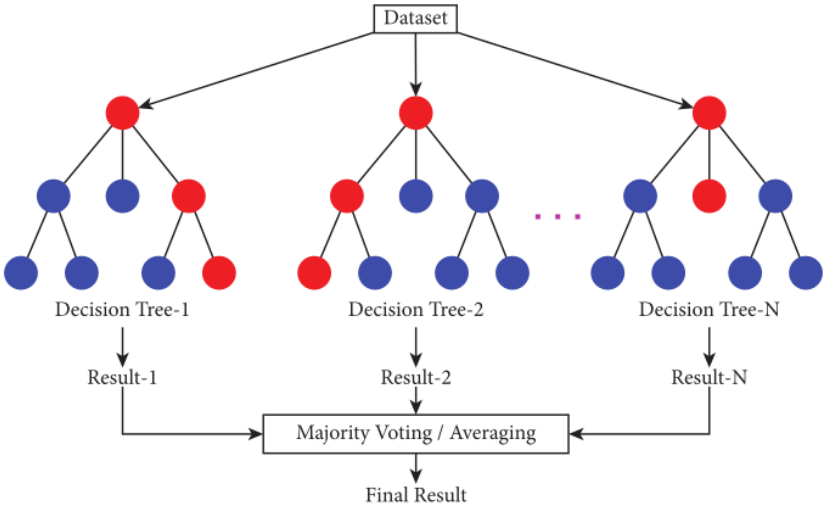


Figure 6: Random forest illustration (reproduced from [37], Hindawi, licensed under CC BY 4.0)

This process reduces overfitting, leading to a better generalization in comparison to a single tree [36].

2.1.4.5 XGBoost

Another example that will be later used is the XGBoost algorithm (eXtreme Gradient Boosting). Like RF, it is an ensemble of decision trees, but trained using the boosting technique. The trees are now trained not in parallel, but sequentially. This allows each new tree to be trained to correct errors made by the previous trees, focusing on previously misclassified data points. The final prediction is a weighted sum of the predictions from all sequential trees [38]. A more complex illustration of the XGBoost algorithm is shown in Figure 7.

Models that use the boosting technique are more prone to overfitting, but if properly regularized, normally achieve better results than those that use the bagging technique. XGBoost is renowned for its speed and performance on tabular data [38].

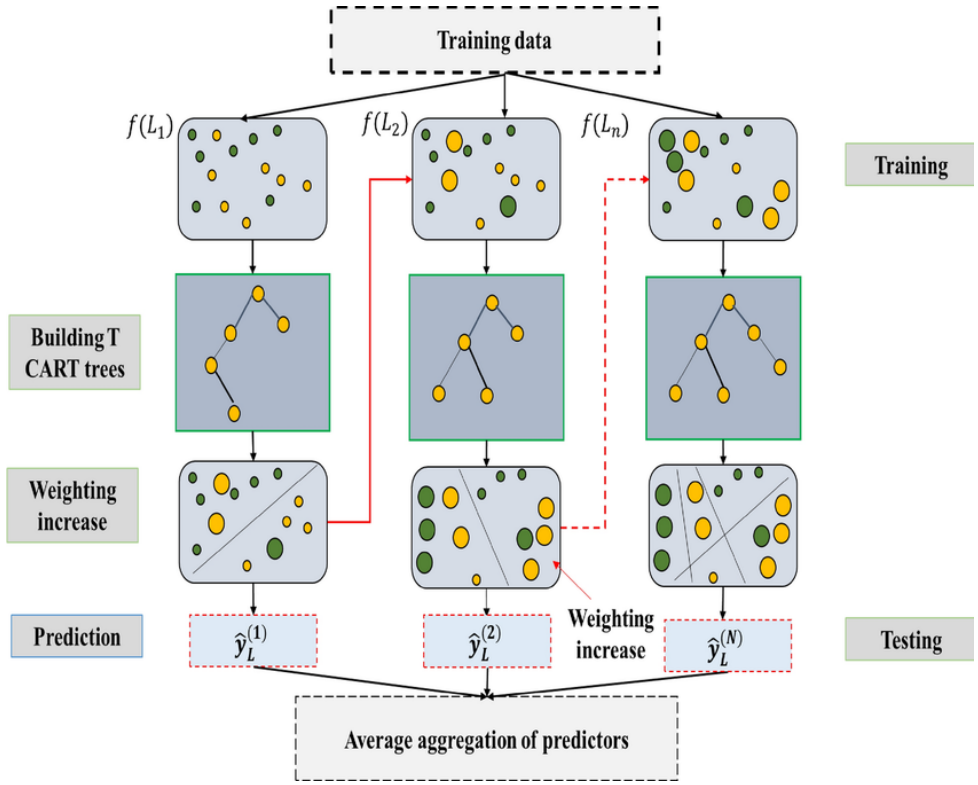


Figure 7: XGBoost illustration (reproduced from [39], Asian Journal of Civil Engineering, licensed under CC BY 4.0)

2.1.4.6 Convolutional Neural Network

One of the most common and impactful types of neural networks used in DL is the CNN. CNNs are a supervised learning method primarily used for processing grid-like data, such as images, and are the cornerstone of modern computer vision tasks like image classification and object detection [40].

The key innovation of this architecture is the use of convolutional layers. They apply filters to input images, which compute element-wise multiplications and sum them to produce a 2-dimensional activation map. This process allows the network to detect relevant patterns such as edges, textures, and objects that are agnostic to their position in the image [41].

A typical CNN architecture is composed of a stack of the following modules:

- Convolutional Layers: This is the core of CNNs. Each layer applies multiple filters to extract hierarchical features. The first layers learn low-level features, like edges and corners, while the deeper layers combine these to form complex patterns and objects [41].
- Activation Function: A non-linear function that is applied to the output of a convolution. This introduces non-linearity, allowing the network to learn complex mappings.
- Pooling Layers: These layers reduce spatial dimensions progressively, resulting in a reduction of the computation load and number of parameters, and also help in reducing overfitting.

- Fully Connected Layers: The end of the network. These layers aggregate the high-level features extracted to perform the final classification.

An example of this architecture can be seen in Figure 8.

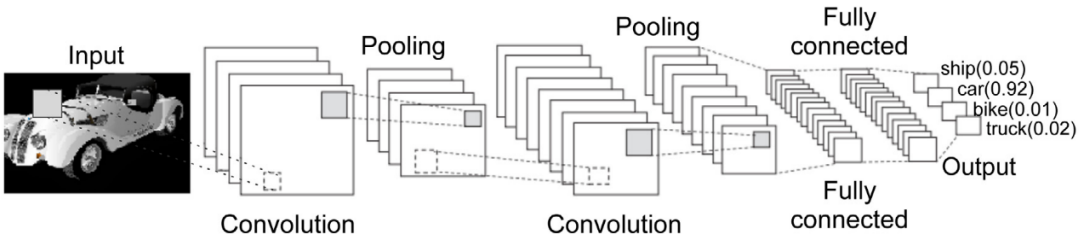


Figure 8: Example of a CNN architecture (reproduced from [42], Elsevier, licensed under CC BY 4.0)

Over the last few years, CNNs have become increasingly popular for identifying objects in images, often preferred over more TML methods like RF and SVM. This is largely due to their ability to automatically learn features directly from raw pixel data, eliminating the need for manual feature engineering [43]. However, these have a black-box nature. This means there is a lack of transparency in these modules, where it is very difficult to understand how the models arrive at a specific prediction, despite the model's ability to produce accurate outputs. Because of its nature, significant improvements in interpretability and computational efficiency are still active areas of research [43].

2.2 Remote Sensing

RS is the process of collecting information about an object, area, or phenomenon without direct physical contact [44]. It is typically achieved through satellites, aircraft, drones, or ground-based sensors that detect electromagnetic radiation reflected or emitted from the Earth's surface [45]. RS plays a crucial role in EO, enabling the collection of data over vast and often inaccessible areas and supporting the monitoring and analysis of environmental, agricultural, and urban changes.

2.2.1 Types of Remote Sensing

RS is commonly classified according to its method of data acquisition: passive RS and active RS, each employing distinct technologies to capture information about the Earth's surface.

Passive RS relies on natural energy sources such as sunlight or thermal emissions to capture data. Sensors of this type detect the reflected or emitted electromagnetic radiation from the Earth's surface, making them dependent on external energy sources and most effective during clear and daylight conditions [46]. Technologies used in passive RS include multispectral sensors, which capture data across multiple wavelength bands, including visible, near-infrared, and shortwave infrared. Hyperspectral sensors take

this capability further by acquiring data in hundreds of narrow spectral bands, enabling detailed analysis of surface materials [47].

In contrast, active RS systems emit their own energy towards the Earth's surface and measure the reflected or scattered signal. Unlike passive systems, they are independent of sunlight or thermal emissions, making them ideal for all weather conditions and nighttime data collection [48]. LiDAR is a prominent active RS technology, using laser pulses to measure distances with exceptional accuracy. It is commonly applied in creating digital elevation models [49], forests structure mapping [50], and to support urban planning [51]. SAR is another key active system, which uses radar waves to penetrate clouds and collect data under any lighting conditions. SAR sensors are particularly valuable for flood mapping, land deformation studies, and vegetation analysis [52].

By leveraging both passive and active systems, RS provides a comprehensive toolkit for analyzing the Earth's surface under diverse conditions, with each approach offering unique advantages depending on the specific requirements of the study or application.

2.2.2 Applications and Challenges in Remote Sensing

RS has a wide range of applications across various scientific, environmental, and industrial fields. It plays a critical role in monitoring vegetation health, mapping land cover, and studying changes in land use, providing valuable data for ecological and agricultural planning. In forestry, RS enables large-scale TS mapping, biomass estimation, and forest health monitoring [53]. Urban planners utilize RS to analyze land development patterns and assess infrastructure, while climate scientists employ it to study atmospheric changes, surface temperatures, and long-term environmental trends [54]. Additionally, RS is essential in disaster management, where it helps track the extent of floods, wildfires, and landslides, providing timely information to mitigate risks and manage emergencies [55]. Oceanographic studies also benefit from RS, with applications in sea level monitoring, wave height measurement, and ocean current detection, all of which are vital for marine conservation and navigation [56].

Despite its extensive utility, RS faces several challenges. One major limitation is the reliance of passive sensors on weather conditions and daylight, which can restrict data availability in regions with frequent cloud cover or during nighttime [57]. The resolution of RS data also presents challenges; while high-resolution imagery provides detailed information, it requires significant computational resources for processing and storage [58]. Furthermore, integrating data from diverse sensors, each with its own unique characteristics, can be difficult and may lead to inconsistencies in analysis [59]. Finally, the interpretation of RS data often demands advanced expertise and robust algorithms, making it necessary to bridge gaps in technical knowledge and computational capacity to fully harness its potential [60].

By addressing these challenges and continually advancing sensor technology and analytical methods, RS remains an indispensable tool for understanding and managing the Earth's dynamic systems.

2.3 Synergy of Machine Learning and Remote Sensing in Tree Species Classification

TS classification has significantly advanced with the integration of AI and ML, particularly through the utilization of RS technologies and hyperspectral imaging. These techniques facilitate large-scale, high-precision mapping of TS, enhancing biodiversity monitoring, forest management, and conservation efforts.

2.3.1 Review Approach

The literature search was conducted in the Scopus and MDPI databases using the query: *remote sensing AND tree species classification AND NOT UAV AND NOT airborne*. As the study aims to use only satellite RS data, the terms "UAV"and "airborne"were excluded to filter out studies using drones or aircraft sensors. The search was restricted to English-language articles published in the last ten years, to reflect rapid technological progress in both hardware and software. This initial query returned 190 articles. From these, papers were retained only if they satisfied all of the following criteria:

1. The study classified more than one TS.
2. The study relied exclusively on satellite RS data.
3. The study reported the classification methods used.
4. The study reported results for at least one of the methods employed.

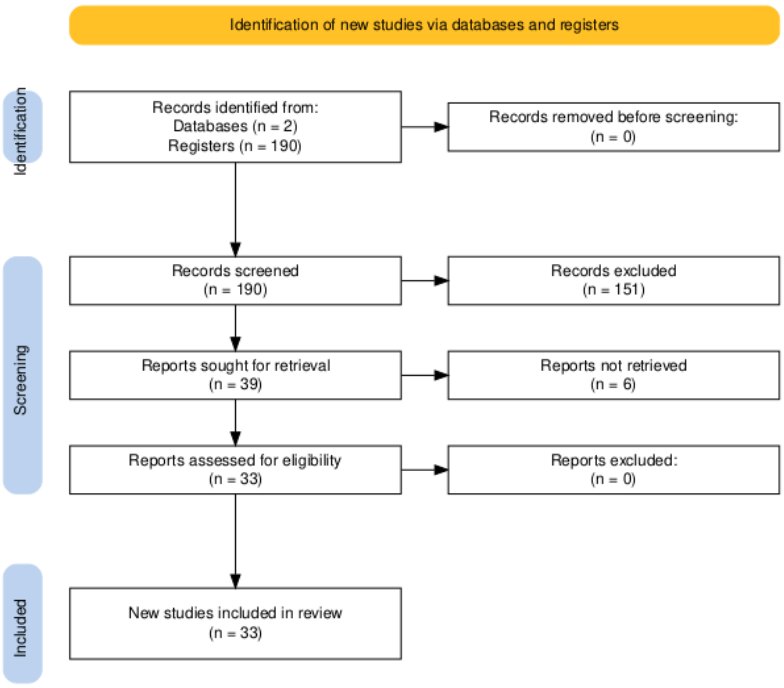


Figure 9: PRISMA flow diagram illustrating the study selection process for the systematic review

The study selection process followed the PRISMA framework [61] to ensure transparency and reproducibility. Figure 9 summarizes the identification, screening, and inclusion steps of the systematic review.

A total of 33 studies met all inclusion criteria and were included in the qualitative synthesis. These studies form the basis of the subsequent analyses presented in this chapter.

2.3.2 Classification Methods

TS classification in the reviewed studies follows two main approaches: TML methods and DL methods.

TML methods, being conceptually simpler, emerged as the initial approach for TS classification and remain valuable for straightforward cases or as benchmark baselines. With the improvements in spatial and spectral resolution on newer RS sensors, DL is rapidly becoming the standard for TS classification.

As shown in Figure 10, TML methods still dominate most studies, but DL methods are gaining traction over the last years.

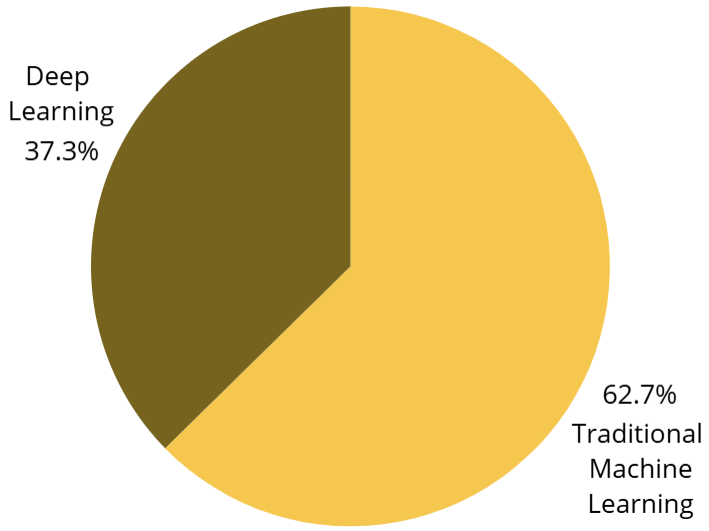


Figure 10: Percentage of studies using TML methods versus DL methods

2.3.2.1 Traditional Machine Learning Methods

TML follows a straightforward workflow. Its relative simplicity makes it an accessible starting point for TS classification studies. Still dominant to this day, it remains effective for low-diversity studies. RF and SVM are the most commonly used algorithms, with RF typically yielding the best results. Other algorithms like XGBoost occasionally outperform alternatives, while methods such as Naive Bayes, k-Nearest Neighbors, Maximum Entropy, CatBoost, Gradient Boosting, and Multiple Linear Regression can also be used for comparative analysis. Figure 11 shows the distribution of algorithm usage across studies.

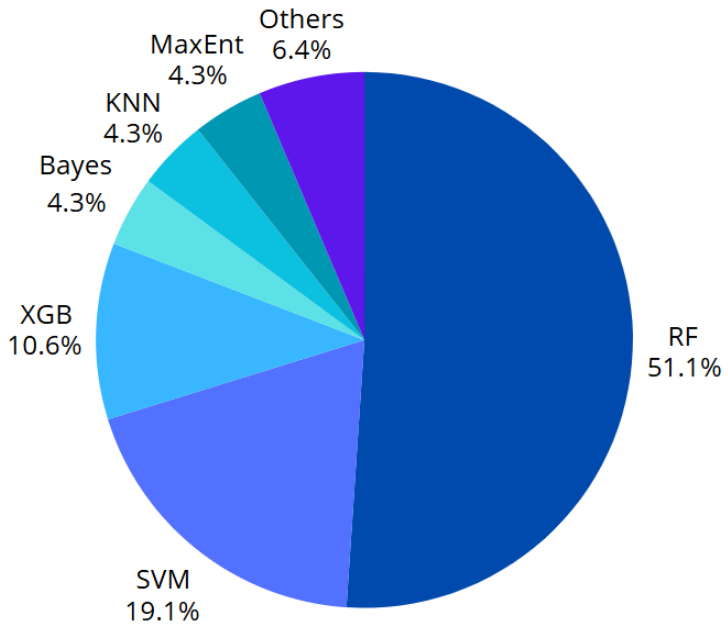


Figure 11: Percentage of studies with different TML algorithms

2.3.2.2 Deep Learning Methods

With the rise of DL over the past five years, CNNs have become prevalent for TS classification. DL offers significant advantages over traditional methods, including the handling of high-dimensional data (e.g., hyperspectral images (HSI)s) and the learning from raw pixels directly, which reduces manual feature engineering requirements. CNNs have enabled studies to classify more TS across broader areas. No single model consistently outperforms others, resulting in diverse model selection tailored to specific case studies (Figure 12).

2.3.2.3 Emerging Hybrid Methods

Despite substantial improvements from newer methods, some cases remain challenging to classify satisfactorily with a single approach. Consequently, recent studies increasingly combine multiple classification methods for different workflow stages. For example, Vaghela et al. [62] used ResNet for feature extraction and RF for classification, obtaining an OA of over 90%.

While no standardized hybrid method has emerged for most applications, proposed combinations have proven effective for enhancing classification accuracy.

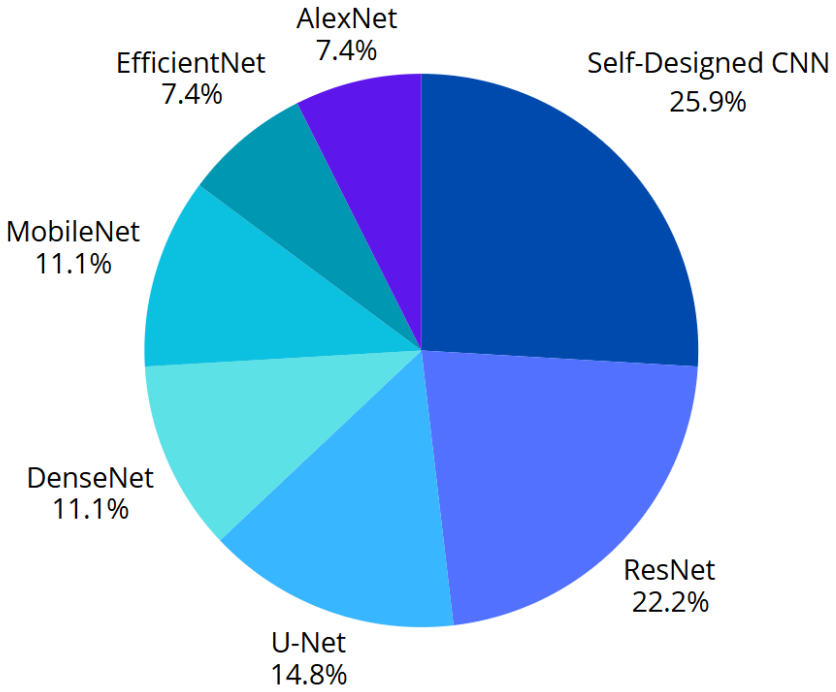


Figure 12: Percentage of studies with different DL models

2.3.3 Remote Sensing Systems

One of the pillars for TS classification is the use of RS data. There are five main types of satellite data used: red, green, blue (RGB); SAR; very-high resolution (VHR); multispectral images (MSI); and HSI.

The efficacy of TS classification is dependent on selecting appropriate RS data, as different sensors capture distinct spectral, spatial, and structural attributes of forests. Modern satellite systems offer a suite of complementary data types, each with unique advantages for TS discrimination. Optical sensors provide spectral signatures critical for species identification, active sensors penetrate clouds and quantify vertical canopy structure, and VHR imagery enables fine-scale crown-level analysis.

This section evaluates the four primary satellite data types used in TS classification: SAR, VHR, MSI, and HSI, highlighting their respective strengths, limitations, and frequency of use (Figure 13).

2.3.3.1 Multispectral Sensors

As seen in Figure 13, MSI is the most commonly used RS data for TS classification.

The spatial resolution of these images ranges from 3 to 30 meters, which is generally sufficient to achieve satisfactory results.

Another advantage of MSI is its spectral resolution, which provides valuable insights into species separability through different spectral bands and VIs.

With a high revisit frequency, MSI helps to minimize issues related to cloud cover and enables the use of multi-temporal data, which is a crucial factor in TS classification. Studies have demonstrated that

incorporating images from different seasons and years significantly improves OA.

The most commonly used MSI satellite is the Sentinel-2, which is included in over half of the studies due to its great spatial and spectral resolution and, most importantly, its free accessibility.

However, MSI has limitations, particularly in distinguishing taxonomically similar species due to its restricted spectral and spatial detail. Studies where the goal is to have very high accuracy results in identifying individual TS need to use other types of data in conjunction with these images.

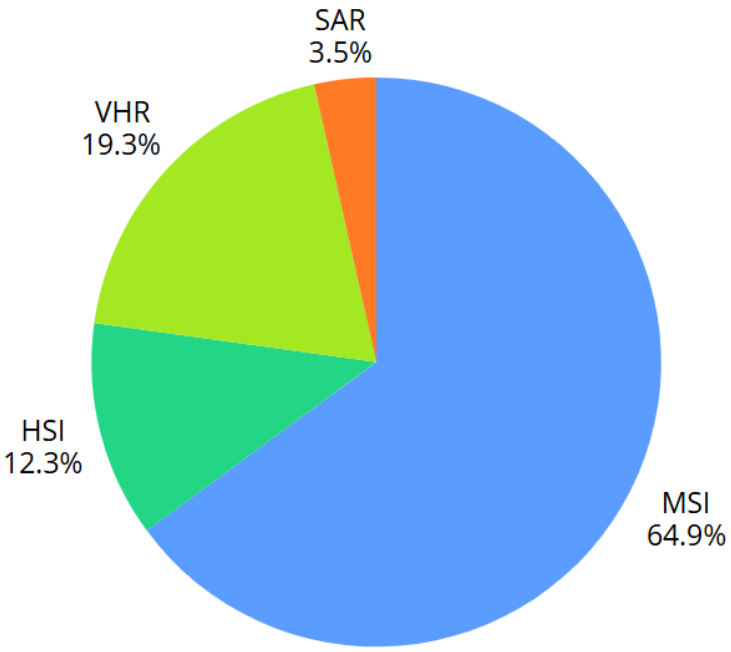


Figure 13: Frequency of usage of each satellite RS type

2.3.3.2 Hyperspectral Sensors

With more than 200 narrow bands for fine spectral signatures, HSIs are increasingly being used in TS classification.

Although it typically has a coarser spatial resolution, with at least 30 meters per pixel, which poses challenges in dense forests, HSI covers the spectral limitation seen in MSI.

Other challenges associated with this type of data include its limited global coverage and high processing complexity. As a result, HSIs are often used with other sensor types, though exceptions exist.

2.3.3.3 Very-High Resolution Imagery

VHR satellites provide images with a resolution finer than one meter. This enables crown delineation and analysis, leading to the detection of specific species traits.

With such resolution, it is a great choice to combine with MSI or HSI to enhance spatial resolution.

Compared to Sentinel-2 images, VHR images are expensive and have limited spectral resolution and a low revisit rate, making it hard to use them alone on dense forests for TS classification.

Despite these constraints, VHR remains a great solution for applications that require fine-scale accuracy.

2.3.3.4 SAR and Multi-Sensor Usage

Satellites equipped with SAR sensors, such as Sentinel-1, offer unique advantages over optical sensors. As an active sensing system, SAR can penetrate clouds and is sensitive to tree structure, making it highly effective in all weather conditions, particularly in cloud-prone regions.

The primary drawback of SAR is its lack of spectral resolution, necessitating fusion with optical data for effective TS classification.

Multi-sensor approaches are increasingly employed to enhance classification accuracy by leveraging the complementary strengths of different sensor types. The number of studies utilizing multi-sensor data has risen in the last years, making it now the more widely used approach, as shown in Figure 14, proving its effectiveness in TS classification.

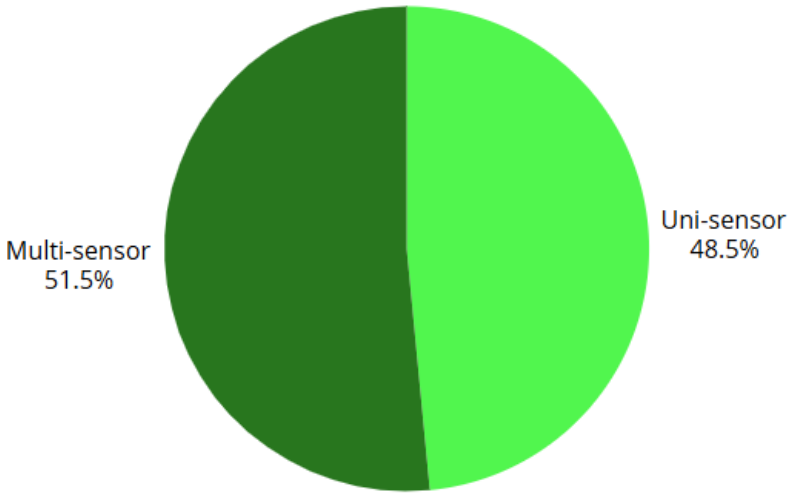


Figure 14: Percentage of studies using uni-sensor or multi-sensor data

2.3.4 Literature Trends in Tree Species Classification

The difficulty of a TS classification study lies in two primary factors: the number of target species and the characteristics of the study area.

As a rule of thumb, increasing the number of species reduces model prediction accuracy, mainly due to spectral overlap between taxonomically similar species. In studies with a large number of target species, genus-level grouping commonly increases OA by approximately 10% [63].

The study area presents greater complexity, as its difficulty depends on multiple factors. The most evident is the size of the area itself. A small area, around 10 km^2, typically yields strong results, whereas a large area of more than 1000 km^2, poses significant challenges. This occurs due to different variables such as topography and ecological uniformity, where flat terrains are studied in conjunction with mountainous terrains, leading to different spectral signatures at different altitudes for the same species, and big changes in biodiversity are also expected in larger regions.

Urban studies also achieve higher accuracy than natural forest analyses. In urban/planted areas, managed tree spacing simplifies crown delineation [4]. Canopy density further impacts accuracy. Dense canopies cause spectral and spatial blending, reducing performance, as He et al. [64] reported, with a 15% accuracy drop in deciduous forests.

2.3.4.1 Temporal Coverage

There are different types of temporal coverage, but the one that is the most used and provides the best results is multi-seasonal images.

Simply adding three more images of different seasons to each sample will enhance results in most studies, as proven by Grabska et al. [63], with a 10% increase in accuracy with seasonal composites.

For studies in non-dense locations and/or with a reduced number of target TS, it might not improve results that much, but for studies in large/dense forests, multi-seasonal images are practically a must-have feature.

2.3.4.2 Auxiliary Data

Training a model with spectral imagery alone provides only part of the solution. Auxiliary data provides the missing piece, giving insight into ways that are not present in spectral bands.

There are several different types of auxiliary data used in all studies, but some of them are used in most.

The simpler ones to add are VIs. They are designed to enhance vegetation properties of the images by transforming image bands into new ones. For TS classification, the most important one is the Normalized Difference Vegetation Index (NDVI). But NDVI is just one of many. Other indices can deliver species separability where NDVI cannot. NDWI, for example, is sensitive to leaf water content, which can help differentiate species with varying water retention capabilities.

Another essential feature is topography data. Normally obtained from a Digital Elevation Model (DEM), it provides data on elevation, slope, and altitude, which boosts the quality of the model in a significant way

in heterogeneous landscapes. Zhang et al. [16] obtained accuracy gains of 20% simply by adding DEM data as a complement to the images.

Climatic data can also be a good addition to most models. Features like precipitation, temperature, and humidity can provide insights that cannot be obtained anywhere else.

Geolocation can be a great way to improve models in large areas, as areas separated by hundreds of kilometers can have different climates and biodiversity.

2.3.4.3 Algorithm Selection

As said before, the algorithms are divided into two main groups: TML and DL. As complexity increases (more species, larger areas, higher-dimensional data), DL models more often outperform TML in the reviewed studies.

TML methods excel in simple scenarios. Low species count, small areas, and moderate dimensional data make it so these methods work as well as or even better than some DL methods.

In turn, DL methods dominate in complex scenarios. They automatically learn subtle spectral-spatial features missed by manual feature engineering and are able to use their convolutional layers to extract relevant features without overfitting.

As an example, Chen et al. [65] used ResU-Net and RF to classify 10+ species in Huangshan, where ResU-Net obtained an OA of 94.3% and RF an OA of 82%.

The hybrid approach can be an option as well, but there is no standard that works for every case, so it needs more complex work to outperform both TML and DL methods. Still, there are a few studies that were able to obtain excellent results by creating a hybrid approach for their specific problem.

2.3.5 Limitations and Possible Future Work

Current TS classification studies face several critical limitations regarding spatial, spectral, and temporal resolution, as well as cloud shading and canopy overlay.

Although improvements in these areas would make it so that better models would be developed, it would still not help the problem of limited ground truth data. Field surveys are time and labor-intensive, leading to a scarce number of samples. For models to be trained as well as possible, more samples are needed, especially of rare species. Studies like Wang et al. [66] are starting to fight this problem by using self-supervised learning, leading to a decrease of 40% in labeled data, but it still isn't standard in the field.

Current studies often choose either TML or DL algorithms and represent the classification with them from start to finish, having a lack of comparison between both methods for different types of studies.

But more importantly, most TS classification studies predict the species of the specific area where the study is being conducted. Whether it is due to the lack of ground data or for the difficulty of the study,

there still isn't a model capable of predicting with precision the TS of a large region, like a big city, country, or even aggregation of countries, without using data from different RS instruments.

2.3.6 Summary of the Literature Review

The reviewed literature demonstrates that the classification of TS using satellite data has advanced considerably in the last decade, largely driven by the integration of ML and DL techniques. TML algorithms such as RF and SVM remain common due to their simplicity and interpretability, while DL approaches, particularly CNN architectures, are increasingly preferred for complex, large-scale studies. Hybrid models that combine the strengths of both paradigms have also emerged as promising alternatives, though they are not yet standardized.

Regarding RS data, MSI imagery, especially from Sentinel-2, dominates current applications owing to its balance between spectral, spatial, and temporal resolution and its open accessibility. However, its moderate resolution limits the discrimination of taxonomically similar species. HSI offers superior spectral detail but remains constrained by coverage and processing complexity, while VHR imagery provides detailed crown-level analysis at higher costs. Recent studies increasingly adopt multi-sensor approaches, fusing data from optical and SAR sensors to enhance classification robustness under diverse conditions.

Across the literature, several consistent trends emerge. Increasing the number of target species typically reduces accuracy due to spectral overlap, leading many studies to perform classification at the genus level or to focus on dominant species. Model performance is also strongly influenced by environmental heterogeneity, topography, canopy density, and the spatial extent of the study area. The use of multi-seasonal imagery and auxiliary variables such as VIs, topographic data, and climatic variables consistently improves model accuracy, often by more than 10–20%.

Despite substantial progress, significant challenges persist. Most studies remain geographically restricted and lack generalization beyond the local scale. The scarcity of ground-truth data continues to limit the development of large-scale, high-diversity models. Few studies have attempted continental or cross-border classification using only freely available satellite imagery. Furthermore, the potential of self-supervised learning, synthetic sample generation, and standardized hybrid workflows remains largely unexplored.

Overall, the literature highlights clear advances in algorithmic capability and data integration, but also reveals a persistent gap in scalable, generalizable TS classification at regional or continental levels. This gap directly motivates the present study, which seeks to develop and evaluate a hybrid DL model for large-scale TS classification across the Iberian Peninsula using freely available MSI data and auxiliary environmental variables.

Chapter 3

Materials

In this third chapter, the study area and the different datasets used in this work are described in detail. The preprocessing steps taken to prepare the data for analysis are outlined, followed by an exploratory data analysis to understand the characteristics and distribution of the data.

3.1 Study Area

Located in southwestern Europe, the Iberian Peninsula is bounded by the Atlantic Ocean to the north and west, the Mediterranean Sea to the east and southeast, and the Pyrenees to the northeast, forming a natural border with France. The peninsula covers approximately 582,000 km^2. The study area coordinates range from 36° 00' N to 43° 47' N in latitude and from 9° 30' W to 3° 19' E in longitude. With a very diverse topography, the Iberian Peninsula includes coastal areas, mountain ranges, plateaus, and river valleys, contributing to its rich biodiversity and varied ecosystems. The average elevation is around 660 meters above sea level, with the highest point being Mulhacén in the Sierra Nevada, which reaches 3,479 meters. The climate is predominantly Mediterranean, characterized by hot, dry summers and mild, wet winters, with the annual average temperatures ranging from values slightly lower than 2.5°C to over 17.5°C. Other climates can also be found in the Peninsula, such as oceanic to the north and northwest, semi-arid to the southeast, alpine in the high mountain areas, and continental Mediterranean along the interior. The average annual daily precipitation varies significantly across the region, with most regions having close to 0mm, while some specific areas receive more than 10mm.

The forests of the Iberian Peninsula are among the most biodiverse in Europe. The combination of climate, soil, and topography has resulted in a mosaic of distinct forest ecosystems. The major forest types include: Atlantic deciduous forests, located in the north; Mediterranean evergreen forests, widespread

across the peninsula; coniferous mountain forests, found at higher elevations; and riparian forests, which can be found along rivers and streams. In Figure 15, a map with the natural forests of the Iberian Peninsula is shown.



Figure 15: Land cover of natural forests in the Iberian Peninsula [67]

3.2 Data Understanding

3.2.1 Sentinel-2 Images

The study used Level-2A (L2A) data from the Sentinel-2 satellites, accessed via the Sentinel Hub Processing Application Programming Interface (API) [68]. This API was configured to return pre-processed images by instructing the backend to apply atmospheric correction using the Sen2Cor algorithm [69]. With this approach, consistency in atmospheric correction across the entire study area was ensured, eliminating the need for computationally heavy pre-processing steps locally, thereby enhancing reproducibility and scalability for a large-scale analysis, as this one.

One of the main problems of using Sentinel-2 data is the different spatial resolutions of the different bands. By utilizing the Processing API, it automatically resampled all the input bands to a common grid of 10 meters per pixel. This meant that every band, whatever its native resolution, was delivered to the pixel function at the same resolution, 10 meters. This on-the-fly resampling ensures that all spectral bands

and subsequently calculated vegetation indices are perfectly aligned pixel-for-pixel, which is a fundamental requirement for accurate pixel-based ML classification.

Within the script given to the API, the surface reflectance values of 12 spectral bands were extracted. These are represented in Table 1

Table 1: Spectral bands and characteristics of Sentinel-2 imagery

Band	Resolution	Central Wavelength	Description
B01	60 m	443 nm	Ultra Blue (Coastal and Aerosol)
B02	10 m	490 nm	Blue
B03	10 m	560 nm	Green
B04	10 m	665 nm	Red
B05	20 m	705 nm	Visible and Near Infrared (VNIR)
B06	20 m	740 nm	Visible and Near Infrared (VNIR)
B07	20 m	783 nm	Visible and Near Infrared (VNIR)
B08	10 m	842 nm	Near Infrared (NIR)
B8A	20 m	865 nm	Narrow NIR (RE4)
B09	60 m	940 nm	Water-vapour absorption
B11	20 m	1610 nm	Short Wave Infrared (SWIR)
B12	20 m	2190 nm	Short Wave Infrared (SWIR)

In addition to the spectral bands, nine scientifically established VIs were calculated directly from the reflectance bands to aid in vegetation characterization. The choice of these VIs was based on the review of the literature mentioned in the last chapter and on the paper by Xue et al. [70], which provides a comprehensive overview of various vegetation indices and their applications. The VIs calculated are displayed on Table 2:

Table 2: Summary of vegetation indices used in this study

Acronym	Name	Formula
NDVI	Normalized Difference Vegetation Index	$\frac{B08-B04}{B08+B04}$
NDWI	Normalized Difference Water Index	$\frac{B03-B08}{B03+B08}$
NDRE	Normalized Difference Red Edge Index	$\frac{B8A-B05}{B8A+B05}$
EVI	Enhanced Vegetation Index	$2.5 \times \frac{B08-B04}{B08+6\times B04-7.5\times B02+1}$
RENDVI	Red Edge Normalized Difference Vegetation Index	$\frac{B8A-B04}{B8A+B04}$
MRENDVI	Modified Red Edge Normalized Difference Vegetation Index	$\frac{B08-B06}{B08+B06}$
ARVI	Atmospherically Resistant Vegetation Index	$\frac{B08-(2\times B04-B02)}{B08+(2\times B04-B02)}$
GCVI	Green Chlorophyll Vegetation Index	$\frac{B08}{B03}-1$
VOG	Vogelmann Red Edge Index	$\frac{B05}{B04}-1$

The primary intended uses of these VIs in this study are:

- NDVI: live green vegetation density and canopy vigor.

- NDWI (Green-NIR variant): surface water mapping; moisture patterns context-dependent.
- NDRE/MRENDVI: chlorophyll content using red-edge sensitivity.
- EVI/ARVI: vegetation monitoring with improved robustness to atmospheric effects.
- GCVI: chlorophyll proxy and sensitivity to green biomass.
- VOG: red-edge/green ratio sensitive to pigment changes.

With this, 21 layers (12 bands + 9 VIs) were generated per pixel. This establishes data consistency and significantly reduces the local computational overhead required for feature engineering.

3.2.2 Ground Truth Data

The ground truth data used to train and validate the models was acquired from the IFN4, conducted by the Spanish Ministry for Ecological Transition and the Demographic Challenge (MITECO) [71].

The IFN is an extensive forest survey carried out across the entire Spanish territory, following a systematic sampling design with a grid of approximately 1 x 1 kilometers. At each intersection point (plot), detailed measurements are taken for all trees within a specified radius, including species, height, and other ecological parameters. This ensures the data is representative of the forest conditions across the entire Spanish territory, making it an ideal source of validated data for large-scale RS studies.

As previously mentioned, the IFN4 database contains more than 100 columns across nine tables. Although a lot of important information is stored on these tables, only a few columns can be used for classification purposes. With that in mind, the attributes shown in Table 3 were extracted for each tree measurement.

Table 3: Columns used from the IFN4 database

Name	Data Type	Description
Provincia	*int64*	Province code (Spain)
Estadillo	*int64*	Plot code within province
Especie	*int64*	Tree species code
CoorX	*int64*	Universal Transverse Mercator (UTM) coordinate X
CoorY	*int64*	UTM coordinate Y

3.2.3 Auxiliary Data

To complement the Sentinel-2 imagery and ground truth data, several auxiliary datasets were incorporated into the study. These datasets provide additional context and, as seen in the literature, can significantly

enhance the accuracy of TS classification models. The auxiliary data include elevation and climate variables. The elevation data was obtained from the Shuttle Radar Topography Mission (SRTM v3, 30 m) digital elevation model produced by NASA and USGS, covering the entire Iberian Peninsula [72]. For the climate data, the E-OBS dataset from the Copernicus Climate Data Store was used [73]. It provides daily gridded observational data derived from a combination of in situ meteorological stations, interpolated to a 0.1° by 0.1° regular grid. For this study, the ensemble mean product was used, covering the period of the Sentinel-2 images with three key variables: mean air temperature (°C), total precipitation (mm), and relative humidity (%). The auxiliary data added to the dataset includes the columns present in Table 4.

Table 4: Auxiliary features added to the dataset

Name	Data Type
Elevation	*int64*
Temperature	*float64*
Humidity	*float64*
Precipitation	*float64*

For each plot location, the elevation is the value in meters above sea level; the temperature is the average annual temperature in degrees Celsius; the humidity is the daily mean air humidity measured near the surface, expressed as a percentage (%) that quantifies the ratio between the actual water vapour content and the maximum possible (saturation) water vapour content at the same temperature; and the precipitation is the total daily accumulation of rainfall, snowfall, and hail, expressed as the height (in millimetres) of the equivalent liquid water over a unit area (1 m^2) [73].

3.3 Data Preprocessing

In the data preprocessing stage, several steps were taken to prepare the data for analysis and training. These steps can be divided into two main categories: processing the plot images and preparing the dataset. Figure 16 shows a flowchart summarizing the different steps taken during this stage.

3.3.1 Retrieve Sentinel-2 Images

Given that the dataset contains over 50,000 plots, a total of four times that number of images was required, one for each season. To manage this, a tiling strategy was employed. A grid of tiles, each measuring 75x75 km, was generated to partition the area of interest. For each tile, a cloud-free composite image was created from the timeframe matching the IFN4 campaign years. This ensured that the satellite imagery corresponded temporally with the ground truth data, minimizing discrepancies due to seasonal changes. A stringent cloud coverage threshold of 10% was applied to reduce atmospheric distortion in the images. An example of a tile image can be seen in Figure 17.

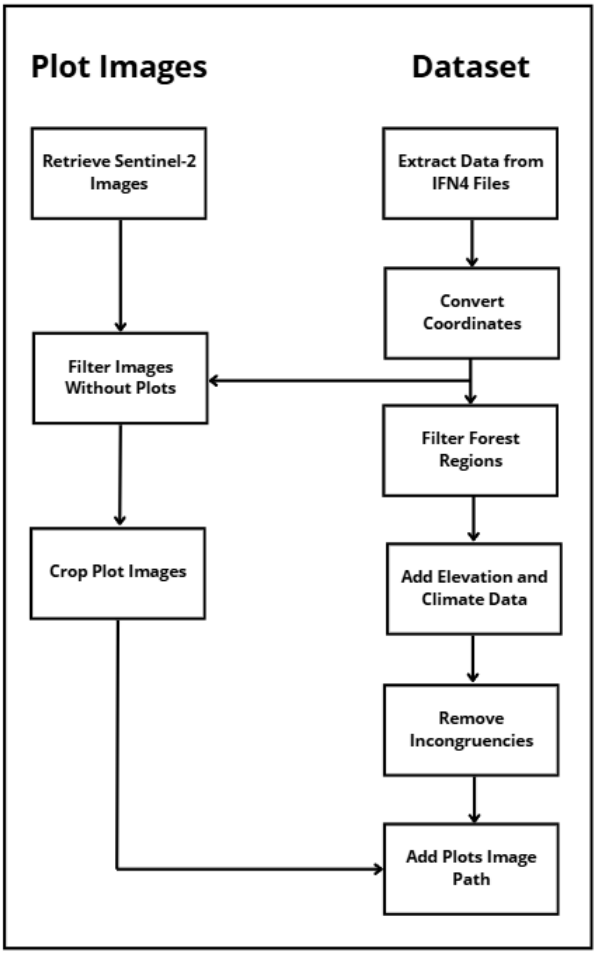


Figure 16: Fluxogram of the data preprocessing

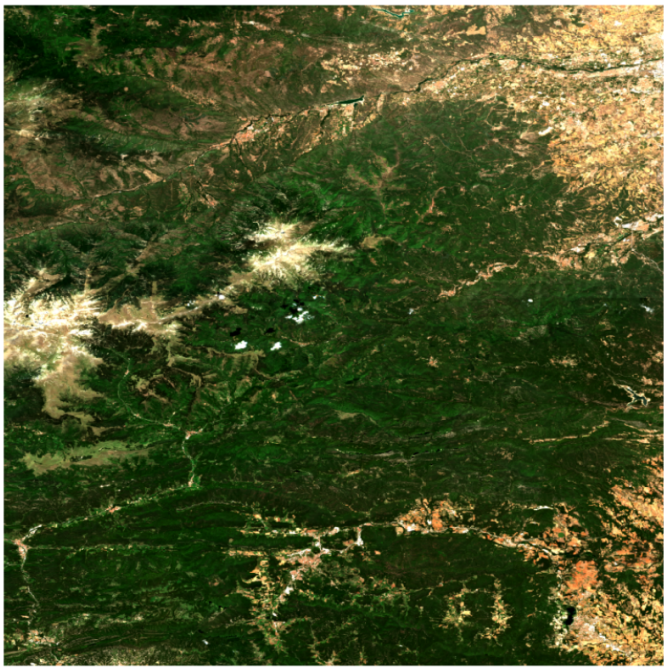


Figure 17: Example of a tile image from the Iberian Peninsula

3.3.2 Filter Images Without Plots

Following the retrieval of the images, an operation was performed to identify and retain only the images that contained at least one IFN4 plot within their boundaries (Figure 18). The coordinates generated in the Section 3.3.5 were used to determine which images did not contain any plot within them. This step significantly reduced computational overhead by eliminating the processing of irrelevant imagery.

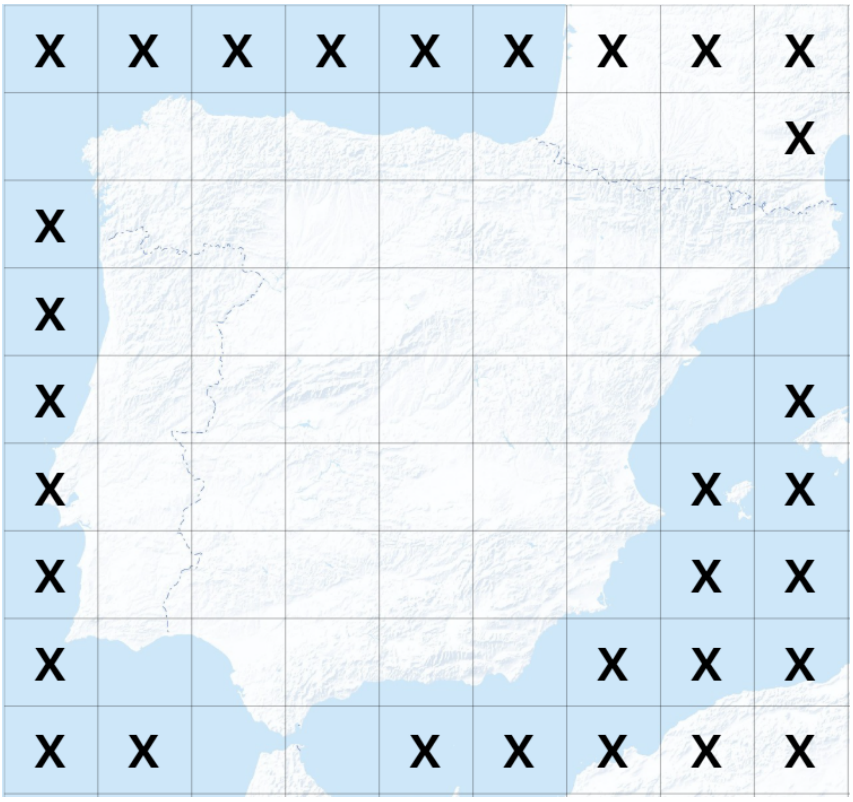


Figure 18: Example of the tiling removal strategy results

3.3.3 Crop Plot Images

The final step involved extracting every individual IFN4 plot located within each tile image. This was achieved by cropping a small image chip centered on each plot's coordinates. A fixed size of 10x10 pixels was chosen, providing a spatial context of approximately 100x100 meters. Figure 19 provides an example of an RGB image of a random plot.

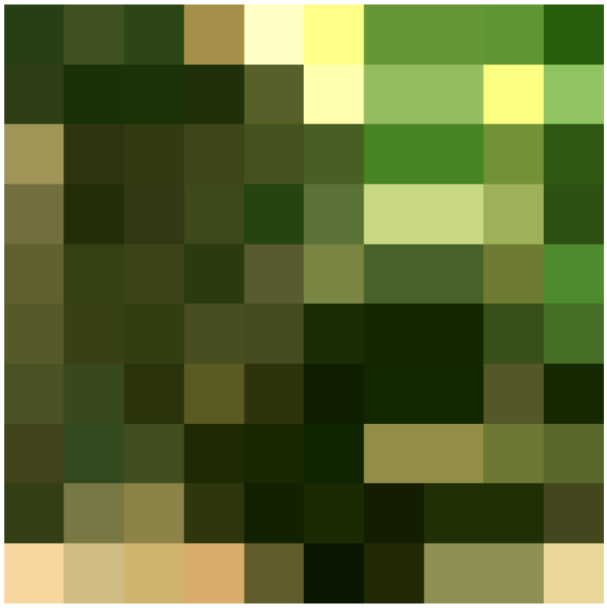


Figure 19: Example of a 10x10 pixel plot image

3.3.4 Extract Data from IFN4 Files

The IFN4 data is divided into a database file per province. A script was developed to loop through each of these files, loading them and extracting the pre-defined set of relevant columns for each plot. The resulting tables were concatenated into a unified national dataframe for consistency.

The following provinces and regions were the ones available in the IFN4 dataset: A Coruña, Albacete, Alicante, Asturias, Ávila, Burgos, Cantabria, Castellón, Cataluña, Ciudad Real, Cuenca, Extremadura, Guadalajara, Huesca, La Rioja, León, Lugo, Madrid, Murcia, Navarra, Ourense, País Vasco, Palencia, Pontevedra, Salamanca, Segovia, Soria, Teruel, Toledo, Valencia, Valladolid, Zamora, and Zaragoza.

3.3.5 Convert Coordinates

The original spatial coordinates in the IFN4 dataset are provided in the UTM coordinate reference system (CRS), with distinct UTM zones for each province. To ensure consistency and uniformity, these coordinates were converted to geographic coordinates, latitude, and longitude, using the appropriate UTM zone for each plot. Special handling was applied to provinces with multiple UTM zones by using a threshold-based logic to select the correct zone. The conversion process utilized geospatial libraries, like *pyproj* [74], to apply the transformation from EPSG:326XX (UTM) to EPSG:4326 (WGS84, latitude/longitude). After conversion, the new latitude and longitude values were integrated into the dataframe.

3.3.6 Filter Forest Regions

To filter the plots located in forested regions, the IFN4 also contains an Spanish forest map (MFE). This dataset has polygons for each autonomous community in Spain, with each representing different land use and land cover types. A spatial filter was applied using the Land Use, Land-Use Change, and Forestry (LULUCF) classification, which includes codes for dense forest types (111, 121, 131). The filter was applied by performing a spatial join between the plot coordinates and the forested regions, ensuring that only plots located within dense forested areas were retained for further analysis.

To better visualize this filter, Figure 20 shows a normal Sentinel-2 image. If the filter is applied to this image, the final result is depicted in Figure 21.



Figure 20: Example image without spatial filter

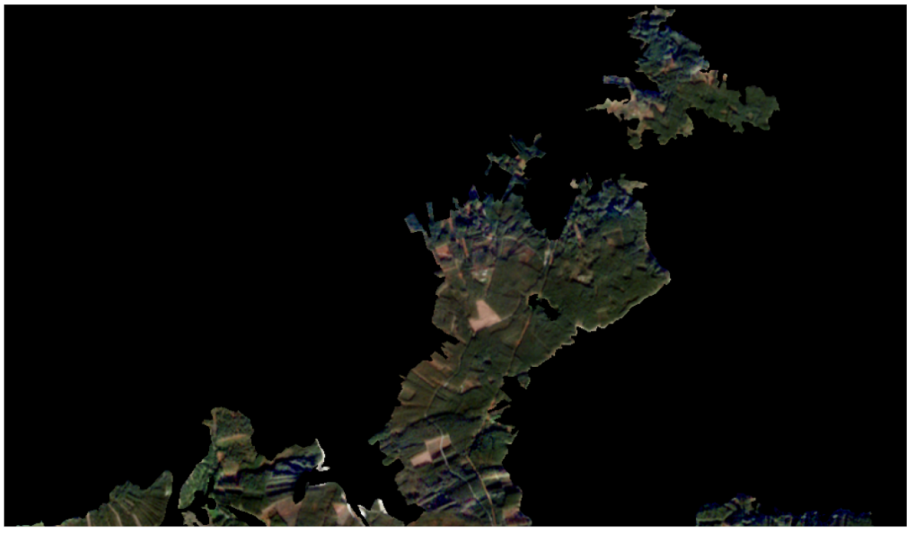


Figure 21: Example image with spatial filter

3.3.7 Add Elevation and Climate Data

Topography is a key factor influencing species distribution. To account for this, a DEM containing Spanish regions was used. For each plot's latitude and longitude, the corresponding elevation value (in meters above sea level) was extracted from the DEM and added as a new feature column to the dataframe.

Climatic variables are fundamental drivers of ecological patterns. High-resolution climate data were sourced from the E-OBS dataset. Using the plot coordinates, the yearly mean values of precipitation, humidity, and temperature were spatially interpolated and extracted for each plot location. These values were then appended to the dataframe attribute table.

3.3.8 Remove Incongruencies

All the previous steps generated a large dataframe with more than 50,000 samples. To ensure data quality, a final cleaning step was performed to remove any inconsistencies. If a sample had missing values in any of the columns or was missing the plot image, it was removed from the dataframe. Duplicate entries were identified and eliminated, ensuring no redundancy. To finish, the plots with images that did not contain any pixel with an NDVI value greater than 0.1 were also removed, as these likely correspond to non-vegetated areas or errors in the data. This threshold was chosen based on the literature review, where an NDVI value of 0.1 is commonly used as a cutoff to distinguish vegetated from non-vegetated surfaces [75]. After these cleaning operations, the dataframe was reset to ensure a continuous index.

3.3.9 Add Plots Image Path

With all the necessary columns added, the final touch was to add the image path of the corresponding image to each plot. By doing so, a multi-modal dataset where each plot is described by its species label, spatial location, elevation, and climatic conditions was created.

3.3.10 Portuguese Land Cover Data

To extend the applicability of the study beyond Spain, the *Carta de Uso e Ocupação do Solo (COS)* dataset [76], published by the General Administration for Territory (DGT), was also incorporated. The COS dataset provides polygon-based delineations of land use and land cover across mainland Portugal, including forested areas, at a comparable scale to Spain's MFE. Its forest-class polygons were used solely as spatial masks to constrain predictions of the trained model to vegetated regions, enabling qualitative analysis of the model's generalization to Portuguese territories.

3.4 Data Analysis

After the preprocessing steps, the final dataset consisted of exactly 41548 samples. 89 unique species were present with 42 unique genera. The columns present in the final dataframe can be seen in Table 5.

Table 5: Columns present on the dataframe used for training

Variable	**Data Type**	**Description**
Province	*object*	Code for the province **Appendix**
Estadillo	*object*	Number of the sample for the current province
LocationID	*object*	ID for the row ({Province}_{Estadillo})
Specie	*object*	Code for the tree specie
Longitude	*float64*	Longitude value
Latitude	*float64*	Latitude value
Elevation (m)	*int64*	Elevation from the sea-level of the plot
Temperature (°C)	*float64*	Annual mean of the temperature of the plot
Humidity (%)	*float64*	Annual mean of the humidity of the plot
Precipitation (mm)	*float64*	Annual mean of the precipitation of the plot
ImagePath	*object*	Path to the corresponding plot image

The minimum, maximum, mean, and standard deviation values for the auxiliary data are displayed in Table 6:

Table 6: Main values of the auxiliary data

Variable	**Min**	**Max**	**Mean**	**Std**
Elevation	0m	2048m	824m	385m
Temperature	0.93°C	18.05°C	11.43°C	2.2°C
Humidity	55.39%	83.18%	71.03%	6.63%
Precipitation	0.30mm	95.85mm	2.42mm	1.87mm

Analysing the species distribution, a lot of information can be extracted. As can be seen on Figure 22, the distribution is highly imbalanced, ranging from the species *Quercus ilex ssp. Ballota*, with 6175 samples to 16 different species, with only 1 sample.

If only the 10 most common species are taken, with the remaining being grouped in *Others*, a pie chart can be made to see their distribution, as is depicted in Figure 23.

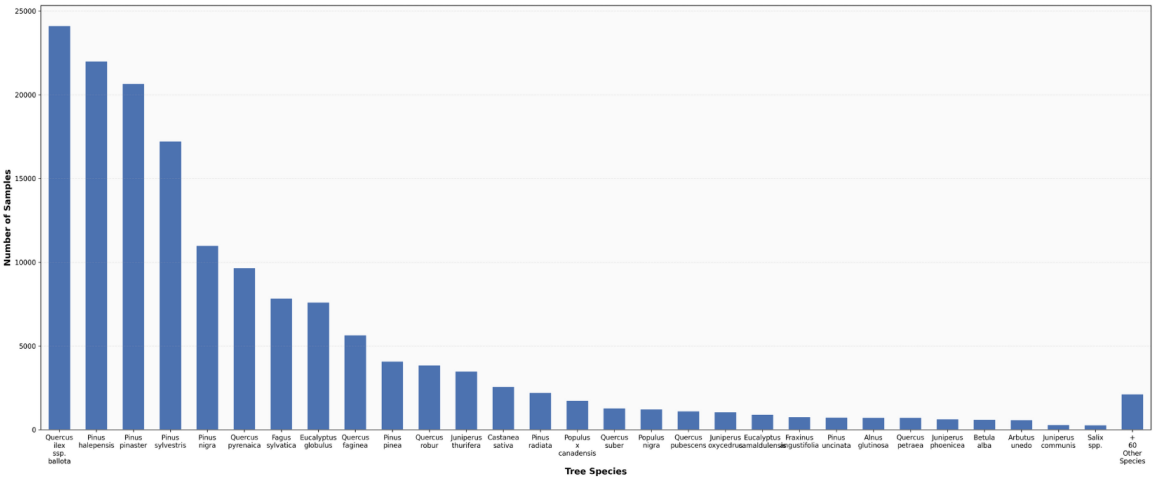


Figure 22: Bar plot for the species distribution

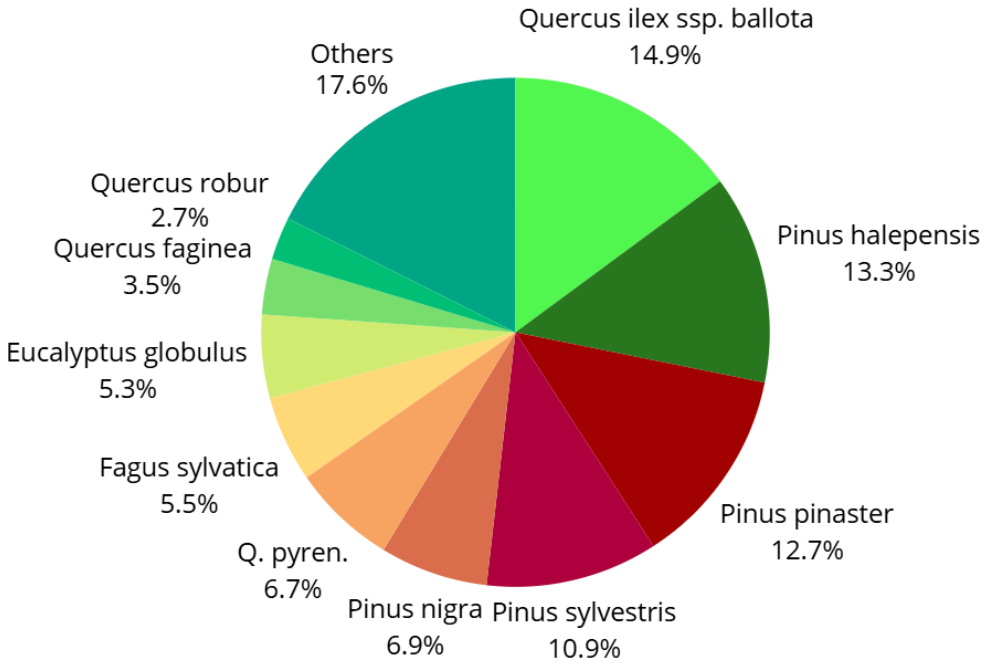


Figure 23: Pie chart of the 10 most common species

By grouping the species by their genus, a more high-level distribution can be made (Figure 24).

Equally, the genus distribution is highly imbalanced, even more than the species distribution. The most common genus is *Pinus* with 20011 samples, and there are 6 genus with only 1 sample. Trying to balance the percentages on a pie chart, the chart is laid out in Figure 25:

As shown, to balance the genus distribution, only 2 genus are needed to represent 78,1% of the samples, with the remaining 21.9% being represented by 31 different genus.

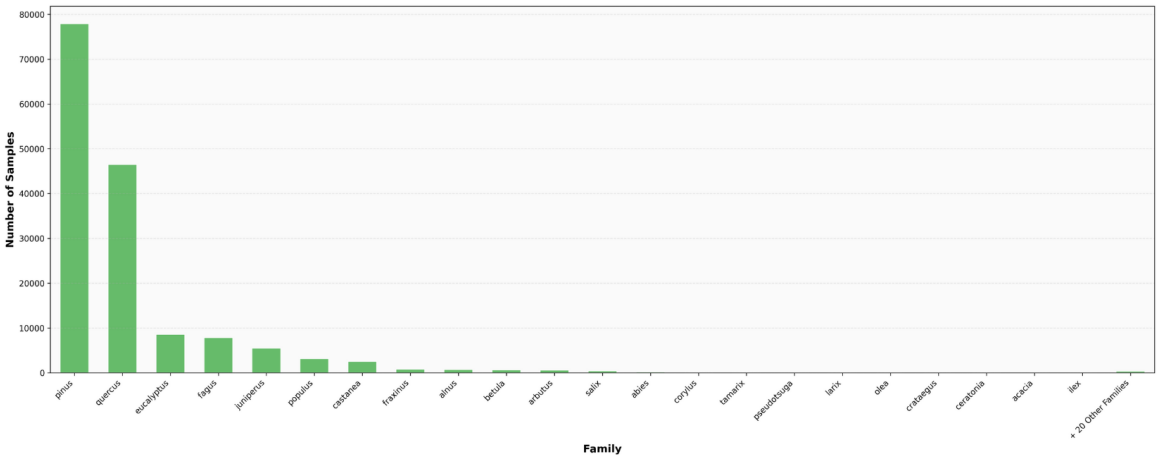


Figure 24: Bar plot of the genus distribution

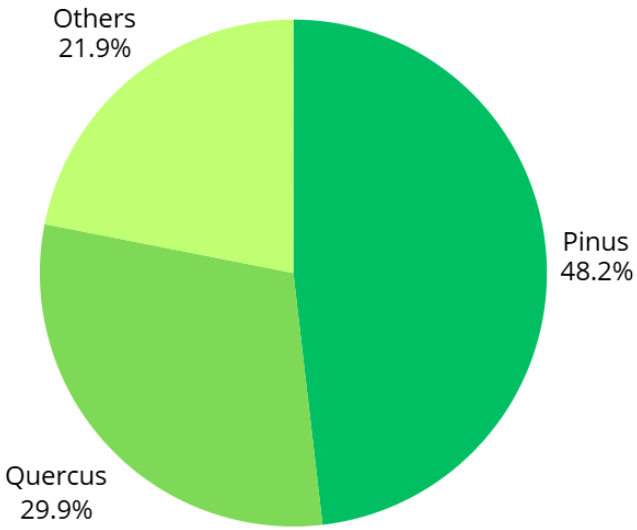


Figure 25: Balanced pie chart of the genus distribution

By taking the longitude and latitude of each plot and adding the elevation, a map can be generated to see the spatial distribution of the samples (Figure 26).

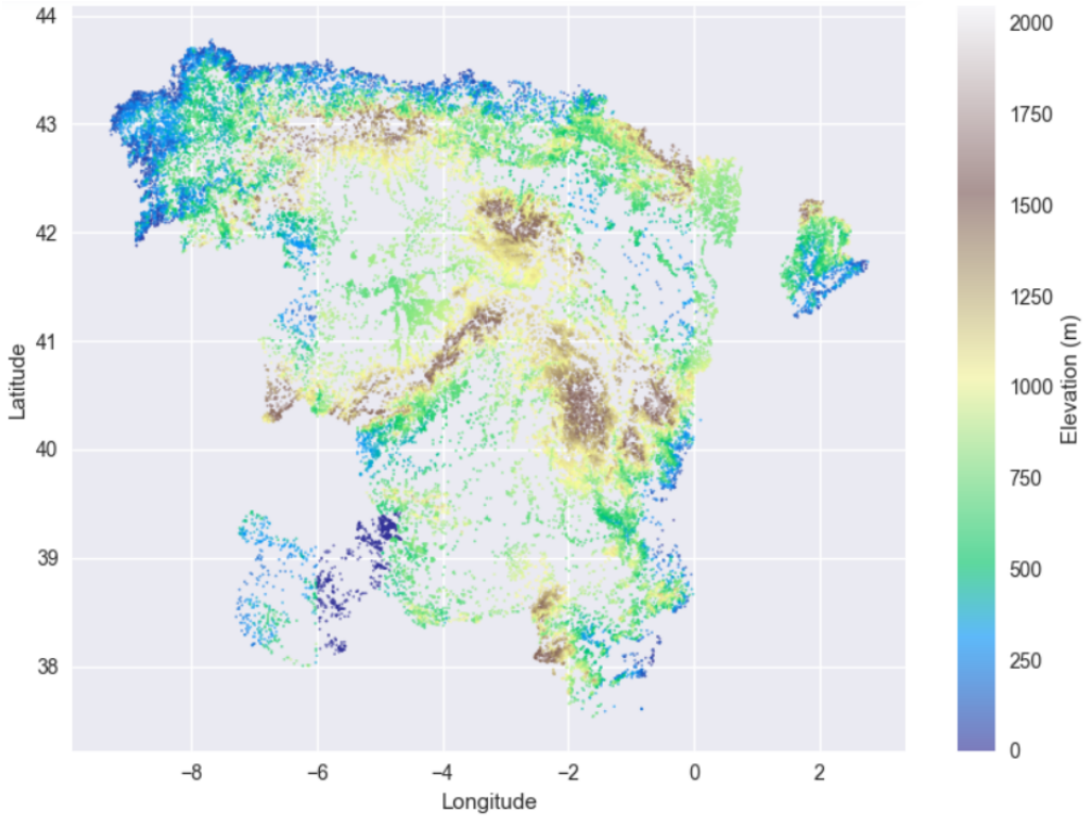


Figure 26: Plot distribution over topography of Spain

Disregarding the provinces without information, a clear map of the Spanish territory can be seen. The samples are well distributed across the country, with a higher density in the north and northwest regions. This is likely due to the higher forest cover in these areas, which is consistent with the known ecological patterns of Spain. The elevation data also shows a wide range of altitudes, from sea level to mountainous regions, indicating that the dataset captures a diverse array of ecological niches. The lack of samples in the central and southern regions corresponds to the more arid and agricultural areas of Spain, where forested plots are less common.

3.5 Summary of the Materials Chapter

This chapter detailed the geographical, environmental, and data foundations underpinning the present study. The Iberian Peninsula was introduced as the study area, characterized by its pronounced climatic and ecological diversity, which provides an ideal natural context for evaluating large-scale tree species classification models.

Multiple complementary datasets were described and integrated. The Sentinel-2 L2A imagery constituted the core remote sensing input, offering 12 spectral bands resampled to 10 meters resolution

and complemented by nine VIs to enhance the spectral representation of vegetation traits. The IFN4 survey served as the principal ground-truth source, ensuring spatially extensive and taxonomically detailed reference data. In addition, auxiliary variables, including topographic and climate information, were incorporated to capture environmental gradients influencing species distribution. Portuguese land-cover data from the COS dataset were also introduced to enable the qualitative extension of the analysis beyond Spain.

Comprehensive preprocessing procedures were applied to ensure consistency and scalability across datasets. These steps included tiling and compositing cloud-free Sentinel-2 imagery, spatially aligning plots with the corresponding satellite tiles, cropping plot-level image chips, converting coordinate systems, filtering non-forested areas using the MFE dataset, and merging all attributes into a unified multimodal dataframe. Rigorous data cleaning ensured the removal of missing, duplicated, or non-vegetated samples.

The final dataset comprised of 41548 samples representing 89 species and 42 genera, covering a wide ecological gradient across Spain. Exploratory analysis revealed substantial class imbalance, with a few dominant genus such as *Quercus* and *Pinus* representing the majority of the samples. The spatial distribution of plots was consistent with known forest patterns, showing higher sample density in the northern regions and lower representation in arid southern areas.

Overall, this chapter established a harmonized and geospatially consistent dataset that integrates multispectral, environmental, and field-based information. These materials constitute the essential foundation for the model development, training, and evaluation described in the following chapter.

Chapter 4

Methods

This chapter describes the methodology used to develop the multi-modal DL model for TS classification. It covers the evaluation metrics, architecture configuration, architectures tested, and benchmark configurations.

4.1 Evaluation Metrics

To evaluate the performance of every model, the metrics used were OA, precision, recall, and F1-score. These metrics are commonly used in classification tasks to assess the performance of ML models, and in the literature review, they were the most commonly used metrics for similar tasks.

The formula for the OA is given by:

$$\text{Overall Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Where:

- True Positives (TP): The number of correct positive predictions.
- True Negatives (TN): The number of correct negative predictions.
- False Positives (FP): The number of incorrect positive predictions.
- False Negatives (FN): The number of incorrect negative predictions.

Although OA is a useful metric, it can be misleading in cases of imbalanced datasets, as a model can simply classify the most predominant species, disregarding the least common ones, obtaining good

accuracy results, but leaving most of the species out of it. Knowing that this is an example of a very imbalanced dataset, precision, recall, and F1-score were also used to provide a more comprehensive evaluation of the models' performance.

Precision is calculated as:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

Recall is calculated as:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

The F1-score is the harmonic mean of precision and recall, and is calculated as:

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

A model is considered effective if it achieves high OA and F1-score, indicating robust classification performance across both common and less frequent species. As the F1-score uses both precision and recall, a good result in this metric also means a good result in precision and recall.

4.2 Architecture Configuration

Once again, the goal of this study is to create a model that can accurately classify TS using a combination of satellite imagery and environmental data. To achieve this, several architectures were tested, and the best one was selected based on its performance on the validation set. In this section, the chosen architecture and its configuration are described.

The model was implemented in *PyTorch*, using the *PyTorch Lightning* framework, which provides a high-level interface that simplifies training, validation, and reproducibility [77].

4.2.1 Data Handling and Preprocessing Pipeline

The first step was to create a *LightningDataModule*. The *LightningDataModule*, a standard component in the *PyTorch Lightning* framework, was used to structure the data pipeline in a reproducible and modular way, handling dataset loading, preprocessing, and splitting for training, validation, and testing. Splitting the dataset into training, validation, and test subsets prevents overfitting and allows the model's performance to be evaluated on unseen data, ensuring that the reported metrics reflect its generalization capability. The batch size, which is the number of samples processed by the model before updating its weights, was set to 64. The values 16, 32, 64, and 128 were tested, and 64 was the one that provided the best results. The training, validation, and test splits were 80%, 15%, and 5%, respectively. Because the dataset

is imbalanced and the study area is very diverse, a bigger training set was chosen to provide the model with more examples to learn from. To divide the dataset, a group shuffle split was used. This method was selected because, on multi-season models, several samples correspond to the same plot. Using group-based splitting prevents data leakage between images of the same location across seasons, ensuring a fair evaluation of the model's generalization.

4.2.2 Feature Preparation

With this done, the creation of the datasets was the next step. The tabular features used to train the model were: elevation, temperature, humidity, precipitation, longitude, and latitude. As the values of these features are all numerical, they were normalized using a standard scaler, which removes the mean and divides by the standard deviation. This ensures all features contribute proportionally to the model and improves training stability.

For the image data, data augmentation techniques were applied to increase the diversity of the training dataset and improve the model's ability to generalize. By generating slightly modified versions of existing images, augmentations simulate the natural variability found in satellite observations, such as differences in illumination, orientation, or sensor noise, without requiring additional data collection. This helps prevent overfitting and improves robustness to real-world conditions.

- Additive Spectral Noise (50%): Gaussian noise was added to simulate real sensor or atmospheric variability. This helps the model learn stable spectral representations instead of memorizing precise pixel values.
- Random Rotation (50%): Images were randomly rotated between 0° and 270° to simulate orientation invariance, since ecological structures appear similar regardless of rotation.
- Random Horizontal and Vertical Flips (50% each): Simple spatial transformations that effectively multiply the apparent dataset size and reduce directional bias.
- Stochastic Channel Dropout (5%): Randomly removes entire spectral bands during training. This technique, known as channel dropout, forces the model to rely on multiple sources of spectral information rather than overfitting to any single band.

With the augmentations applied, the images were then normalized using the mean and standard deviation values calculated from each band and VI.

4.2.3 Model Architecture

The chosen model follows a multi-modal neural network architecture that fuses two data streams:

1. A visual branch processing the Sentinel-2 multi-spectral images.
2. A tabular branch processing the auxiliary environmental data.

4.2.3.1 Visual Branch

The visual component is based on a ResNet-50 CNN pre-trained on ImageNet-1k [78, 79]. The first convolutional layer was modified to accept 21 spectral channels (bands and VIs). The final classification layer was removed, and the resulting feature vector of size 2048 represents high-level spectral-spatial patterns learned from the imagery.

4.2.3.2 Tabular Branch

The auxiliary network processes the six environmental features (elevation, temperature, humidity, precipitation, longitude, and latitude). It consists of two fully connected layers with batch normalization, Rectified Linear Unit (ReLU) activation, and dropout regularization.

- Fully Connected Layer (Dense Layer): A standard neural layer where each neuron is connected to all neurons in the previous layer.
- Batch Normalization: A normalization technique that stabilizes and accelerates training by standardizing the output of each layer.
- ReLU: An activation function defined as $f(x) = \max(0, x)$, introducing non-linearity while preventing vanishing gradients.
- Dropout: A regularization technique that randomly "drops" (sets to zero) a fraction of neurons during training to prevent overfitting.

The auxiliary branch follows the configuration presented in Table 7.

Table 7: Architecture of the tabular branch

Layer	Description / Operation	Output Size
Input	6 environmental features	6
Dense 1	Fully connected layer	64
Batch Normalization	Normalizes activations of previous layer	64
ReLU Activation	$f(x) = \max(0, x)$	64
Dropout 1	Dropout rate = 0.2	64
Dense 2	Fully connected layer	32
Batch Normalization	Normalizes activations of previous layer	32
ReLU Activation	$f(x) = \max(0, x)$	32
Dropout 2	Dropout rate = 0.1	32

4.2.3.3 Fusion and Classification Head

To integrate both data modalities, the 2048-dimensional vector from the ResNet-50 branch was concatenated with the 32-dimensional output of the auxiliary network, producing a combined 2080-dimensional feature representation. This fusion layer captures both spectral and environmental patterns.

The joint representation was then passed through a multi-layer classifier:

Table 8: Architecture of the fusion classifier

Layer	**Description / Operation**	**Output Size**
Concatenation	ResNet-50 features (2048) + auxiliary branch (32)	2080
Dense 1	Fully connected layer	256
Batch Normalization	Normalize activations of previous layer	256
ReLU	$f(x) = \max(0, x)$	256
Dropout	Dropout rate $= 0.3$	256
Dense 2	Fully connected layer	128
Batch Normalization	Normalize activations of previous layer	128
ReLU	$f(x) = \max(0, x)$	128
Output Layer	Fully connected layer to classes	C

ReLU, batch normalization, and dropout work together to enhance training stability and generalization. While ReLU introduces non-linearity while avoiding vanishing gradients, batch normalization ensures stable distributions of activations across layers, allowing faster convergence, and dropout discourages co-dependence among neurons, mitigating overfitting.

This design enables the model to learn complex interactions between visual spectral features and environmental context, mimicking how ecologists integrate multiple data cues for accurate species identification. Simply put, the complete architecture is displayed in Figure 27.

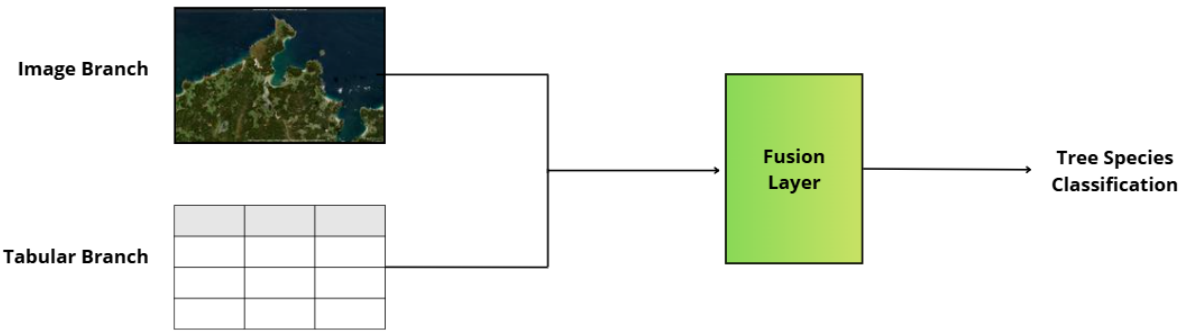


Figure 27: Hybrid architecture

4.3 Architectures Tested

To determine the most suitable model configuration for the TS classification task, several architectures were tested during the experimental phase. This section describes the main configurations evaluated, while the corresponding quantitative results are presented in Chapter 5.2.

The first test came from the assumption, through the literature, that a multi-seasonal approach would provide the best results. For that, 5 different models were created. Having images for the four different seasons, by modifying the *ImagePath* column to the corresponding image, a model for each season was created, trained, and tested. The fifth model was created by repeating each row four times, with the only difference between them being the *ImagePath*, and a new column *Month*. This column corresponded to the month when the image was retrieved. This column was then added to the auxiliary data, but instead of being numerical, it was categorical, and one-hot encoding was used to convert it into a numerical format.

As also mentioned in the literature, genus models could provide finer results, as they would be more abroad than individual species models. Despite the imbalance present in a genus classification task for this dataset, different genus models were created. Unlike the individual species models, instead of removing the rarer classes, a class *Other* was created to encompass those who did not meet the threshold chosen. This could be done here because, once again, the genus is more high-level, and so, the deep-learning model is able to correctly distinguish between them. For the tests, 6 models were tested. The first one has only 3 classes: *Quercus*, *Pinus*, and *Other*. The following models add one more genus, following from most to least common, which can be seen in Figure 12.

As TS are prone to certain regions, a test was made to create region-based models. Three different regions were chosen. These were: *Galicia*, *Castilla-La Mancha*, and *Castilla y León*. For each region, a model was created, trained, and tested. Also, to try and follow the 15-class model base, knowing this model corresponds to 92% of the dataset, the number of classes for each region was chosen to try and reach this coverage, without much imbalance.

The last relevant test made was to try and create a hybrid genus-species model. The way it worked was by giving a threshold to the species. Every species with more than that threshold would be kept as a species, and every remaining species would go to the genus level. On the genus level, another threshold would be given, and every genus with more samples than the threshold would be kept as a genus. The remaining genus would be considered as *Others*. This way, it would be possible to have a model trained with every class available. The thought behind this approach was that the species with more samples would be easier to learn from, and so, the model would be able to learn patterns from them. The rarer species would be grouped into their genus. Knowing that the genus level is broader, with the species that belong to the same genus being similar in some aspects, the model would not need to distinguish between them, and so, it would be easier to classify them. For this approach, three different models were developed, with two different thresholds for the species (4000 and 2500 samples) and genus (1500 and

500 samples). The distribution of classes for each model can be seen in the Figures 28, 29, and 30.

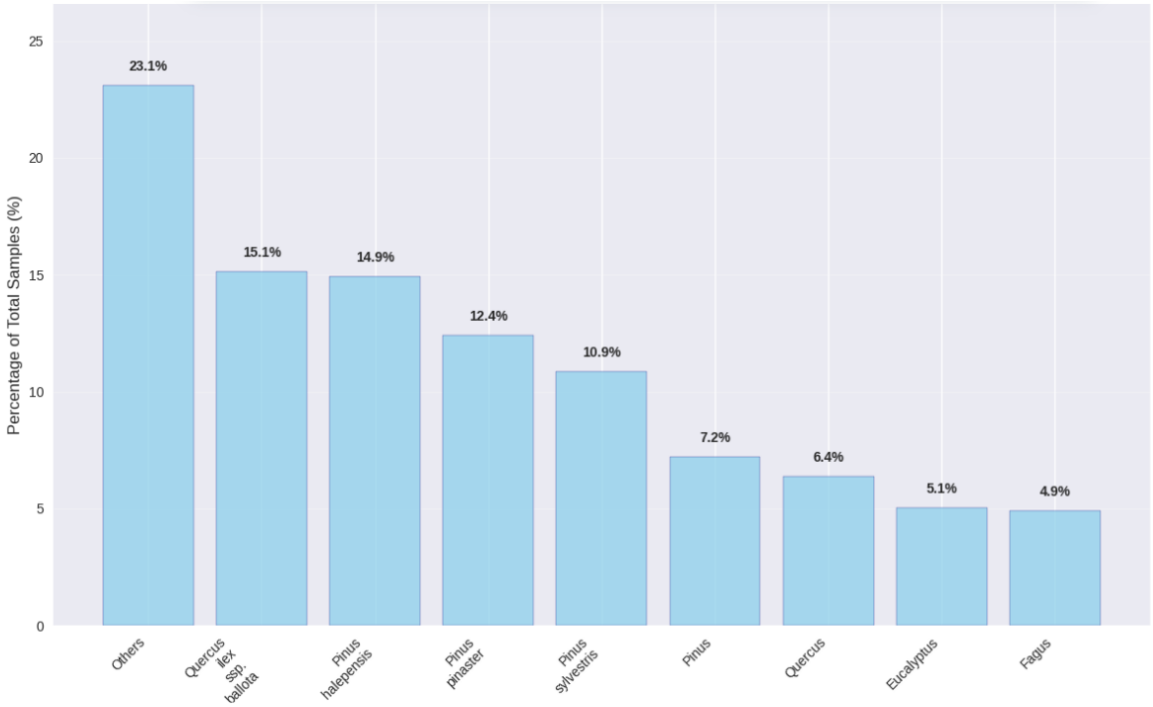


Figure 28: Class distribution for the 4k species - 1.5k genus model

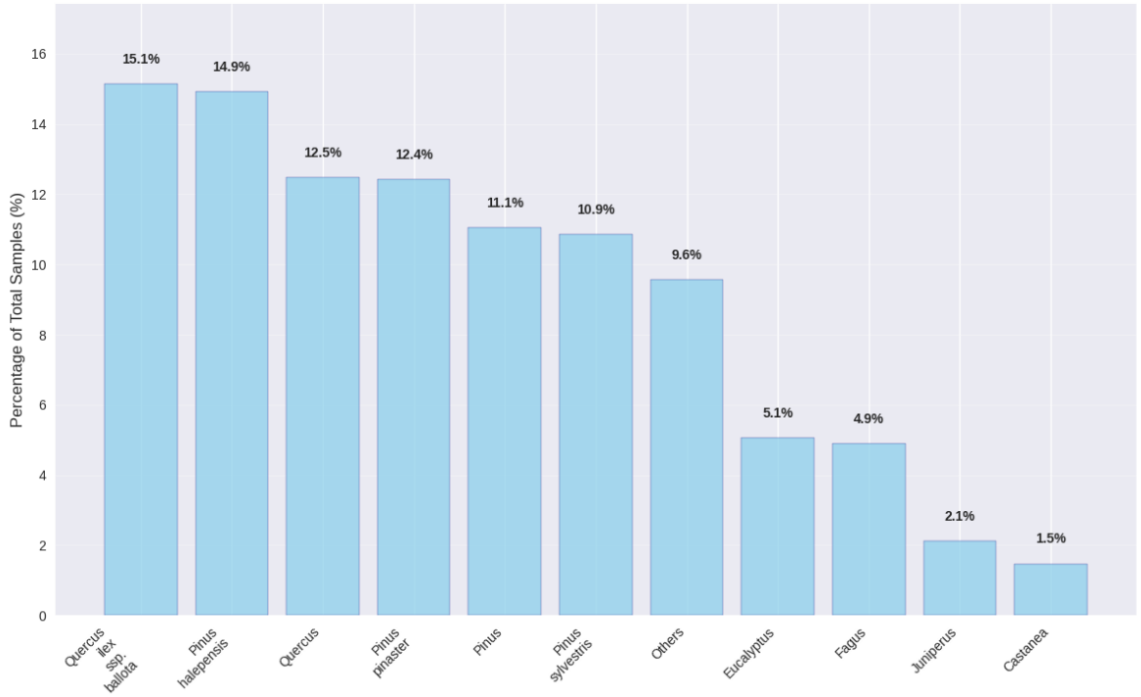


Figure 29: Class distribution for the 4k species - 0.5k genus model

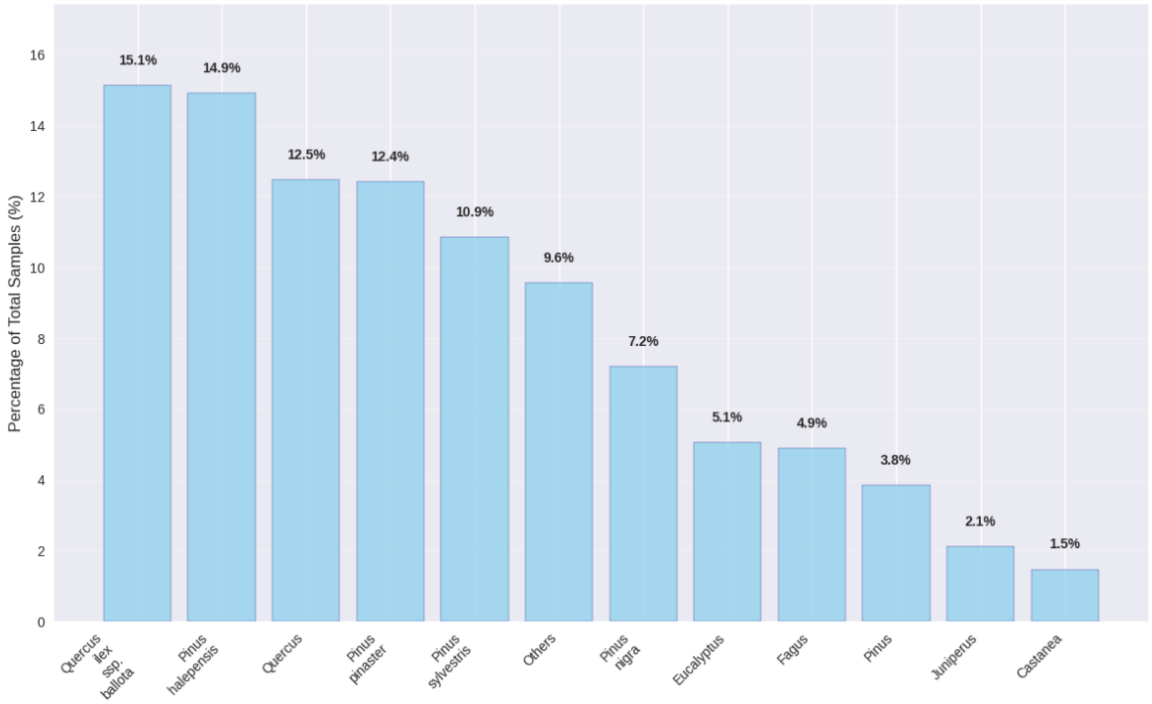


Figure 30: Class distribution for the 2.5k species - 0.5k genus model

These architectural experiments were intended to guide the methodological selection of the final DL configuration rather than to provide standalone scientific results. Each setup was trained and evaluated under the same conditions described in Section 4.2, enabling a consistent comparison across all tested approaches.

4.4 Benchmarks Configuration

A critical component of this study is to establish benchmark performance levels against which the more complex DL multi-modal can be evaluated. To this end, three ML algorithms, as seen in the literature review, widely used, were chosen for this: SVM, RF, and XGBoost.

As TML algorithms do not handle images directly, an extra step to the dataset creation process was added. For each of the 12 bands and 9 VIs, the mean, standard deviation, maximum, and minimum pixel values were calculated, resulting in an extra 84 features. These were then concatenated to the existing dataset.

The same dataset splitting strategy as described in Section 4.2.1 was used: a group shuffle split was used to divide the data into training and testing, using 80% of the data for training and 20% for testing. Using a group shuffle split ensures that plots from the same geographic group do not leak across splits. All features were normalized using a standard scaler fitted solely on the training data to prevent information leakage from the validation or test sets.

Before training each algorithm, a feature correlation analysis was conducted to identify and remove highly correlated features that could introduce redundancy and noise. Features with a correlation coefficient above 0.95 were dropped, resulting in a reduced feature set that improves model interpretability and training efficiency.

For each algorithm, as done for the deep-learning model, a hyperparameter tuning process was performed to identify the optimal configuration. Table 9 presents the training configurations used.

Table 9: Optimal hyperparameter configurations for the benchmark algorithms

Algorithm	**Hyperparameter**	**Value**
5*RF	Number of estimators	300
	Maximum depth	10
	Minimum samples per leaf	10
	Maximum samples	1.0 (entire dataset per tree)
	Maximum features	sqrt
10*XGBoost	Maximum depth	7
	Learning rate	0.01
	Subsample	0.8
	Colsample by tree	0.8
	Colsample by level	0.8
	Colsample by node	0.7
	Regularization α	0.002
	Regularization λ	0.04
	Gamma	1
	Minimum child weight	2
3*SVM	Kernel	RBF
	Regularization parameter C	1.0
	Gamma	scale

4.4.1 Generation of Qualitative Tree Species Maps

To visually assess the model's generalization capacity, qualitative tree species maps were generated for selected regions in Spain and Portugal. For each region, Sentinel-2 imagery was masked using the corresponding national forest map (MFE for Spain and COS for Portugal). The masked images were divided into 10x10-pixel patches, and the auxiliary variables (elevation, temperature, humidity, and precipitation) were assigned to each patch based on their coordinates. The trained hybrid model was then applied to each patch to generate predicted tree species classes and corresponding confidence values derived from the model's softmax output.

4.5 Conclusion of the Methods Chapter

This chapter presented the methodological framework used to develop and evaluate the hybrid deep learning model for large-scale TS classification across the Iberian Peninsula. The selected evaluation metrics, OA, precision, recall, and F1-score, were introduced as the primary indicators of model performance, chosen for their prevalence in the literature and their ability to balance OA with sensitivity to class imbalance.

The architecture of the proposed model was described in detail, highlighting its multi-modal nature that integrates spectral-spatial information from Sentinel-2 imagery with environmental variables derived from auxiliary datasets. The visual branch, based on a modified ResNet-50 network, and the tabular branch, composed of fully connected layers, were fused through a joint classification head designed to learn combined spectral and contextual representations. Data preprocessing, normalization, and augmentation techniques were applied to improve model generalization and robustness.

Several alternative architectures were tested, including multi-seasonal, genus-level, regional, and hybrid genus-species models, to determine the most effective configuration for the problem at hand. In addition, TML algorithms (RF, SVM, and XGBoost) were implemented as benchmarks to provide a comparative baseline against the proposed deep learning approach.

Finally, the methodology for generating qualitative TS maps over Spanish and Portuguese regions was presented, illustrating how the trained model can be applied to real-world remote sensing imagery for large-scale forest monitoring.

In summary, this chapter established the complete methodological foundation for the experimental phase that follows. The next chapter presents and analyzes the quantitative and qualitative results obtained through the implementation of these methods.

Chapter 5

Results

As mentioned in the prior chapter, this project involved various experiments to ensure the final models were optimal. As such, this chapter presents the results of the most significant experiments conducted during the project.

These are divided into three main groups. First, the results of the experiments conducted to determine the optimal number of classes are presented. Then, the comparison of the main deep learning model with the benchmark models is shown. Finally, the results on using only images and auxiliary data are presented to analyse the relative importance of each data source.

5.1 Optimal Number of Classes

Prior work typically reports results for a fixed number of classes. For a study purpose, it can be enough, but for practical applications, not so much. In a large and diverse area like the Iberian Peninsula, end-users may need models of different complexities. For example, one model focuses only on the dominant species, and another covers rarer ones. Thus, this study progressively adds classes to evaluate scalability and practical usability.

Before the experimental results, Figure 31 shows the distribution of samples per class for the 25 species used in the experiments.

The results of this experiment are shown in Table 10. The number of classes is the number of species with which the model was trained. The species were added progressively, starting from the most represented ones. For example, the 8-class model includes the eight most represented species in the dataset. The metrics reported are OA, F1-score, precision, recall, and cumulative percentage of samples covered by the classes. The evaluation metrics were already explained in Section 4.1. The cumulative percentage

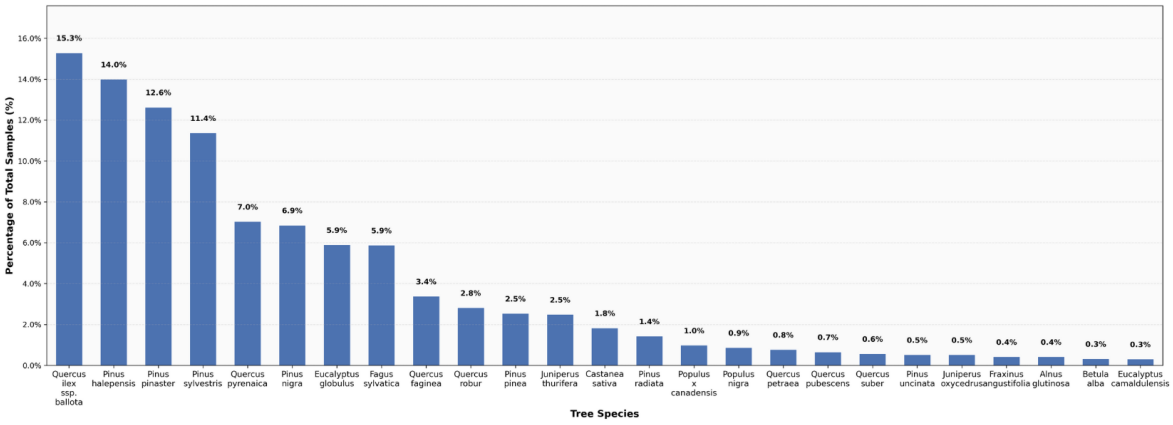


Figure 31: Class Distribution for the Top 25 Species

was added to show the dataset coverage for each number of classes. It consists of the sum of samples of the most represented classes divided by the total number of samples in the dataset. This is important, as adding more classes typically means covering rarer species, which may be less relevant in certain operational contexts.

Table 10: Results for the species models from 8 to 25 classes

Number of Classes	**Overall Accuracy**	**F1-Score**	**Precision**	**Recall**	**Cumulative Percentage (%)**
8	0.8049	0.8001	0.8010	0.8005	79.66
9	0.7792	0.7510	0.7596	0.7514	82.37
10	0.7694	0.7235	0.7312	0.7210	84.87
11	0.7559	0.7228	0.7230	0.7270	86.98
12	0.7504	0.7134	0.7194	0.7114	88.72
13	0.7271	0.6873	0.6890	0.6902	90.23
14	0.7405	0.6817	0.6865	0.6814	91.34
15	0.7309	0.6712	0.6814	0.6645	92.12
16	0.7215	0.6529	0.6680	0.6440	92.90
17	0.7247	0.6275	0.6388	0.6234	93.60
18	0.7232	0.6207	0.6372	0.6131	94.28
19	0.7173	0.6032	0.6244	0.5906	94.92
20	0.7118	0.5830	0.6339	0.5630	95.47
21	0.6969	0.5642	0.6069	0.5456	95.96
22	0.7042	0.5375	0.5543	0.5351	96.43
23	0.6997	0.5315	0.5709	0.5219	96.88
24	0.6970	0.5295	0.5671	0.5243	97.32
25	0.6992	0.5249	0.5450	0.5175	97.71

For the 8-class model, the results show an OA of 80.49% and an F1-score of 80.01%. As the number of classes increases, the performance gradually decreases. The OA drops steadily until it reaches 70% for

the models with more than 20 classes, where it stabilizes there. On the other hand, the F1-score continues to drop, going from 80.01% for 8 classes to 52.49% for 25 classes. Analysing the OA per class for the 25-class model, pictured in Figure 32, it can be seen that OA per class declines sharply beyond the 14th species.

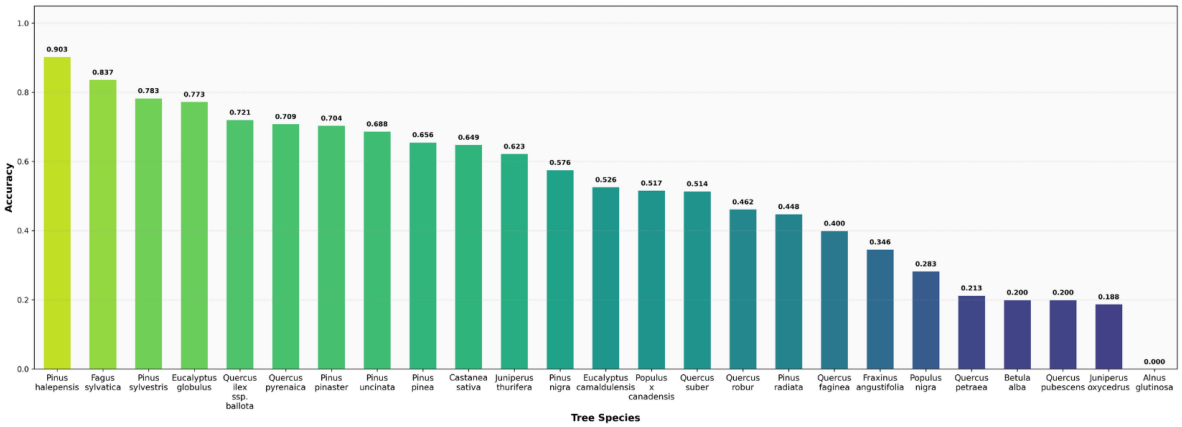


Figure 32: Per-Class OA for the 25-Class Model

From this point, all the results will be shown for the model with 15 classes. It covers 92.12% of the dataset, maintaining a balanced trade-off between dataset coverage and model performance, with an OA of 73.09% and an F1-score of 67.12%.

Figure 33 displays the per-class OA for this configuration. Species 28, 41, and 44 appear as the lowest performing classes by some margin.

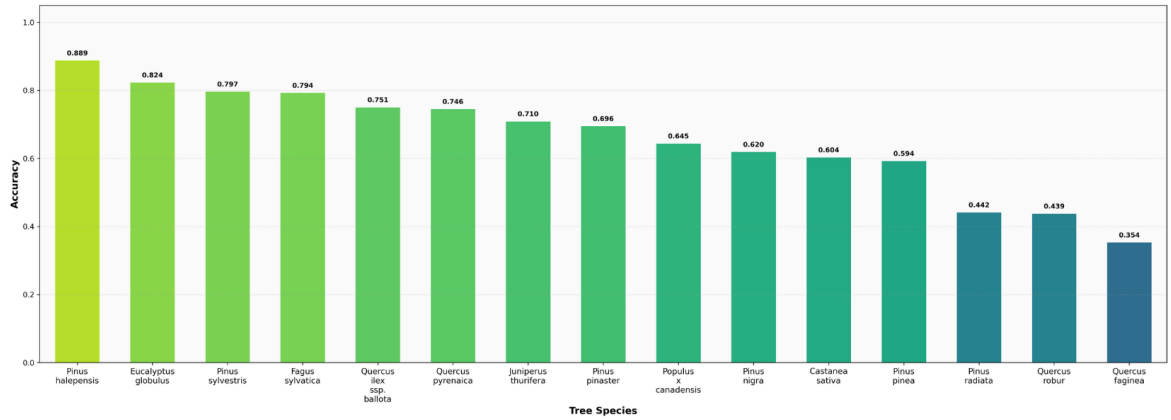


Figure 33: Per-Class OA for the 15-Class Model

The normalized confusion matrix in Figure 34 illustrates the distribution of misclassifications among all classes.

A softmax function is typically used in the final layer of a classification neural network to convert the raw output scores (logits) into probabilities for each class. It works by exponentiating each logit and then normalizing these values by dividing by the sum of all the exponentiated logits. This ensures that the

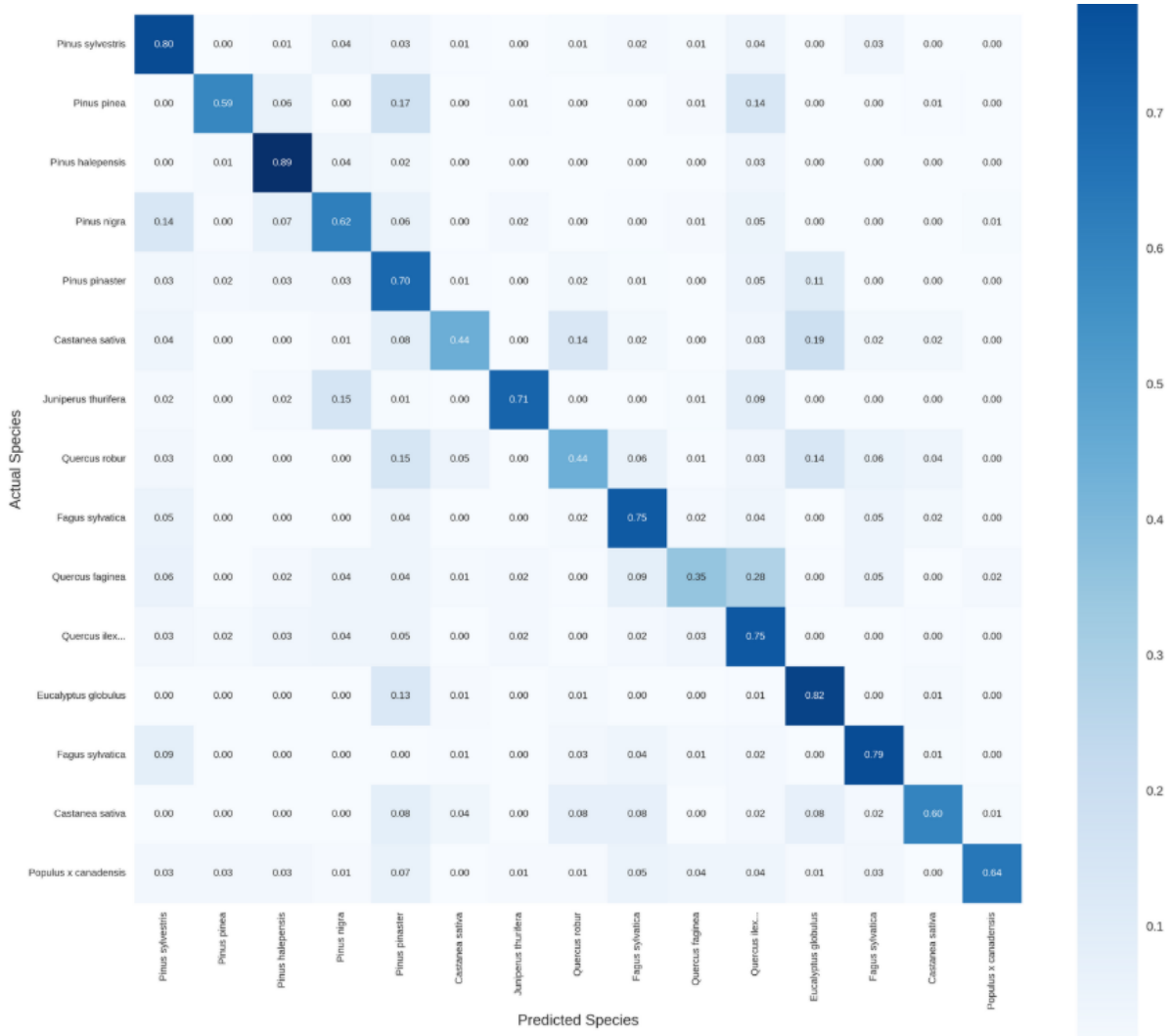


Figure 34: Normalized Confusion Matrix for the 15-Class Model

output probabilities for all classes sum to 1, making them interpretable as confidence scores for each class prediction. This is useful to determine how reliable the predictions are. By storing each prediction's confidence and result, it is possible to analyse how confidence relates to accuracy. Figure 35 shows the class OA for each confidence range in a 0.1 interval for each species and the OA of each.

As can be seen, there is a clear trend of increasing accuracy with increasing confidence. This is further illustrated in Figure 36, which shows the OA against confidence. These quantitative trends are further explored in the Discussion chapter.

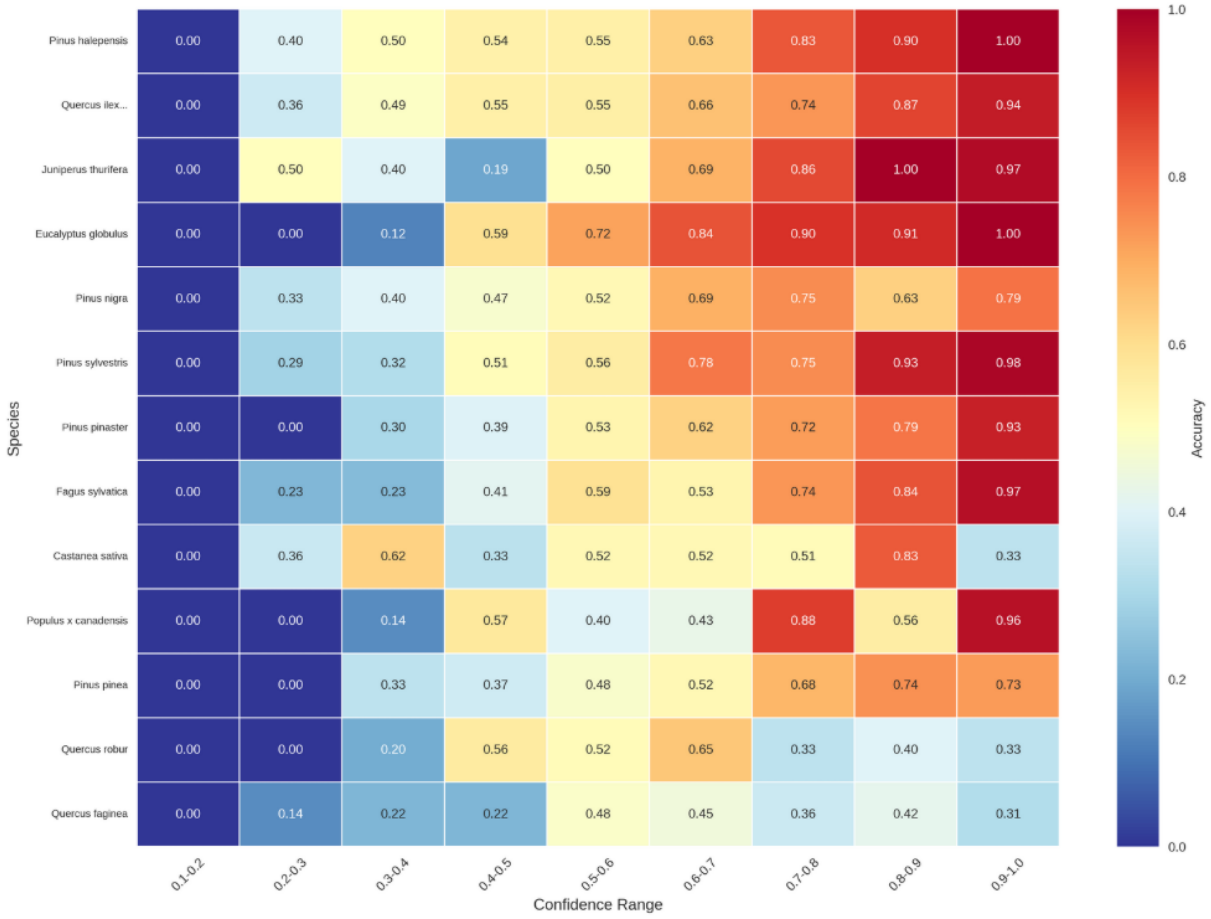


Figure 35: Per-Class Accuracy Across Confidence Ranges for the 15-Class Model

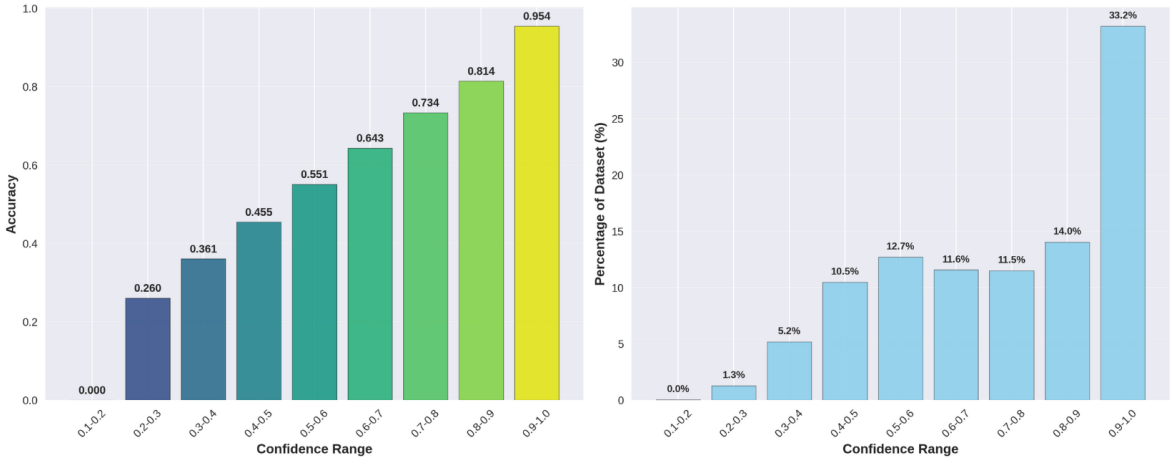


Figure 36: Accuracy vs Confidence for the 15-Class Model

5.2 Architecture Evaluation Results

Several architectural variants were tested to determine which configuration offered the best trade-off between performance and generalization. Table 11 presents the results obtained for the seasonal and multi-seasonal models. Although the literature frequently reports performance gains from multi-seasonal imagery, in this study, the single-season models consistently outperformed it, with the summer model achieving the highest accuracy (73.09%).

Table 11: Seasonal Models Results

Model	**Overall Accuracy**	**F1-Score**	**Precision**	**Recall**
Summer Model	0.7309	0.6712	0.6814	0.6645
Fall Model	0.7161	0.6539	0.6664	0.6463
Winter Model	0.7185	0.6495	0.6623	0.6444
Spring Model	0.7101	0.6441	0.6543	0.6408
Multi-Season Model	0.6523	0.5650	0.6282	0.5466

At the genus level (Table 12), performance surpassed the 15-class species model, but not in a significant way that justified changing to this strategy. This suggests that class imbalance, particularly the highly heterogeneous “Other” class, remains a limiting factor.

Table 12: Genus Models Results

Classes	**Overall Accuracy**	**F1-Score**	**Precision**	**Recall**	**Coverage (%)**
3	0.7933	0.7671	0.7765	0.7607	79.17
4	0.7875	0.7414	0.7493	0.7411	84.63
5	0.7901	0.7364	0.7538	0.7274	89.62
6	0.7863	0.6968	0.7240	0.6802	93.12
7	0.7789	0.6468	0.7015	0.6215	95.13
8	0.7726	0.6042	0.6612	0.5930	96.77

Region-based models (Table 13) yielded competitive results in some areas, particularly in *Castilla y León*, but no clear advantage was observed when compared to the global 15-class model.

Table 13: Autonomous Regions Models Results

Model	**Overall Accuracy**	**F1-Score**	**Precision**	**Recall**
Galicia	0.6860	0.6554	0.6517	0.6639
Castilla-La Mancha	0.7613	0.7403	0.7498	0.7347
Castilla y León	0.7712	0.7206	0.7309	0.7134

Finally, the hybrid genus–species configurations shown in Table 14 produced intermediate results but did not surpass the baseline species-level performance.

Table 14: Hybrid Genus–Species Models Results

Model	Overall Accuracy	F1-Score	Precision	Recall
4k Species – 1.5k Genus	0.7287	0.7248	0.7268	0.7269
4k Species – 0.5k Genus	0.7049	0.6844	0.6784	0.6967
2.5k Species – 0.5k Genus	0.7048	0.6716	0.6690	0.6786

Overall, the results suggest that the seasonal single-image approach, particularly during summer, offered the most stable and accurate performance across experiments. These findings are further discussed in section 6.1.3.

5.3 Benchmark Comparison

To evaluate the performance of the proposed DL model, it was compared against several benchmark models. These included the TML models mentioned in Section 4.4, which are RF, XGBoost, and SVM. The results of this comparison are shown in Table 15. As before, the metrics reported are OA, F1-score, precision, and recall. These models were trained with the 15 most represented classes. For that reason, the model used for comparison was the 15-class model, as explained in the previous section.

Table 15: Results for the benchmark models

Model	Overall Accuracy	F1-Score	Precision	Recall
Hybrid DL Model	0.7309	0.6712	0.6814	0.6645
RF	0.5740	0.4062	0.4984	0.4057
XGBoost	0.6213	0.4996	0.5858	0.4938
SVM	0.5302	0.3581	0.4250	0.3631

From the three benchmark models, XGBoost performed the best, achieving an OA of 62.13% and an F1-score of 49.96%, followed by RF with an OA of 57.40% and an F1-score of 40.62%. SVM performed the worst, with an OA of 53.02% and an F1-score of 35.81%.

The proposed hybrid DL model outperforms all benchmark models by a significant margin, achieving an OA of 73.09% and an F1-score of 67.12%.

These discrepancies between the TML models and the hybrid DL model will be further discussed in the next chapter.

5.4 Ablation Studies

To further understand the contribution of each data modality to the model's performance, ablation studies were conducted. This involved training and evaluating models using only images and only auxiliary data. The results of these studies are shown in Table 16. As before, the metrics reported are OA, F1-score,

precision, and recall. Equally to the benchmark comparison, these models were trained with the 15 most represented classes from the dataset, thus, the model used for comparison was the 15-class model.

Table 16: Results for the ablation studies models

Model	**Overall Accuracy**	**F1-Score**	**Precision**	**Recall**
Images Only Model	0.6689	0.5873	0.6138	0.5720
Auxiliary Data Only Model	0.5646	0.4334	0.4825	0.4374
Hybrid Model	0.7309	0.6712	0.6814	0.6645

The hybrid model, which integrates both imagery and auxiliary data, outperforms the unimodal configurations by a substantial margin, achieving an OA of 73.09% and an F1-score of 67.12%. The image-only model attains an OA of 66.89% and an F1-score of 58.73%, while the auxiliary data-only model performs notably lower, with an OA of 56.46% and an F1-score of 43.34%.

These quantitative results demonstrate a clear performance gap between unimodal and multimodal approaches. Further interpretation of this behaviour is presented in the following chapter.

5.5 Qualitative Application Across Iberian Regions

To demonstrate the spatial consistency and generalization potential of the hybrid DL model, qualitative TS maps were generated for representative forested areas in Spain and Portugal. Although no ground truth data were available for direct quantitative validation outside Spain, the produced maps allow for visual assessment of spatial coherence and ecological plausibility.

5.5.1 Application over Spanish Forests

The region of Trasbar, in northern Galicia, was selected as an example due to its dense forest cover. By masking the Sentinel-2 image (Figure 37) with the LULUCF dense forests filter, present in the MFE, the output is displayed in Figure 38, with the non-forested areas seen in grey.

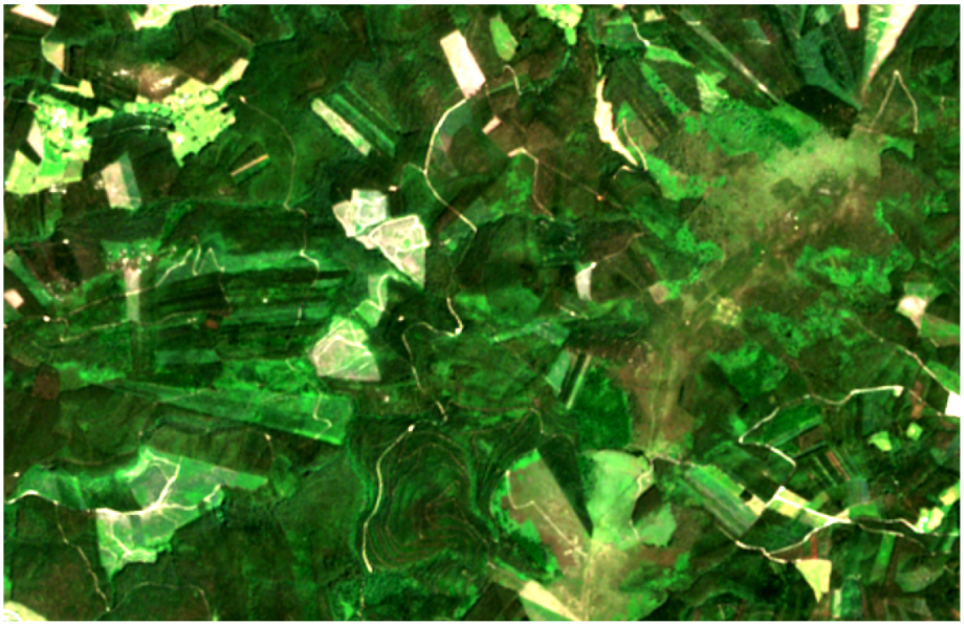


Figure 37: Sentinel-2 Image of the Trasbar region

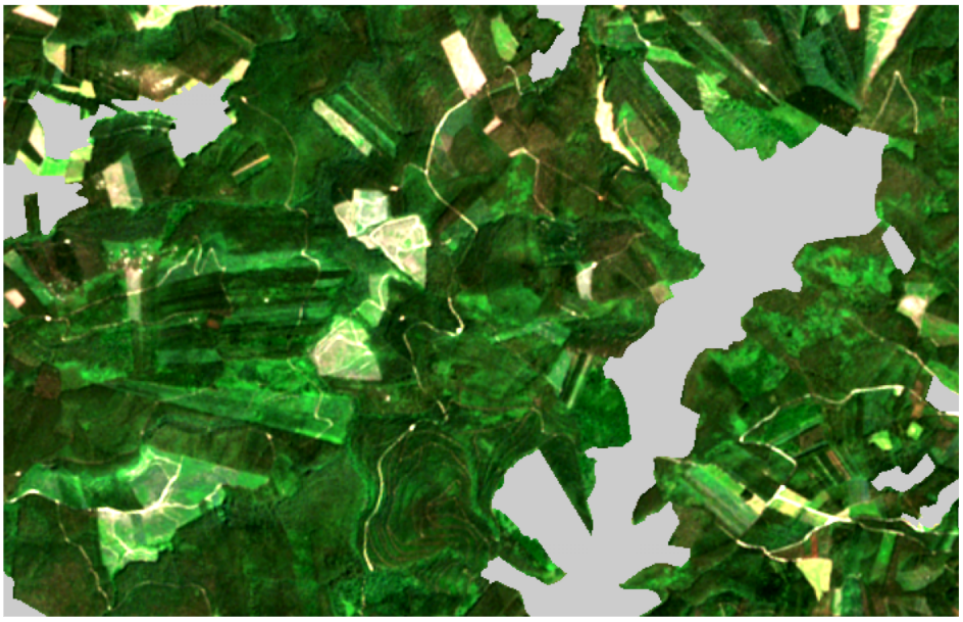


Figure 38: LULUCF dense forest filter applied to the Trasbar region

Using the polygons dominant specie, a true label TS map can be generated (Figure 39) for the dense forest regions.

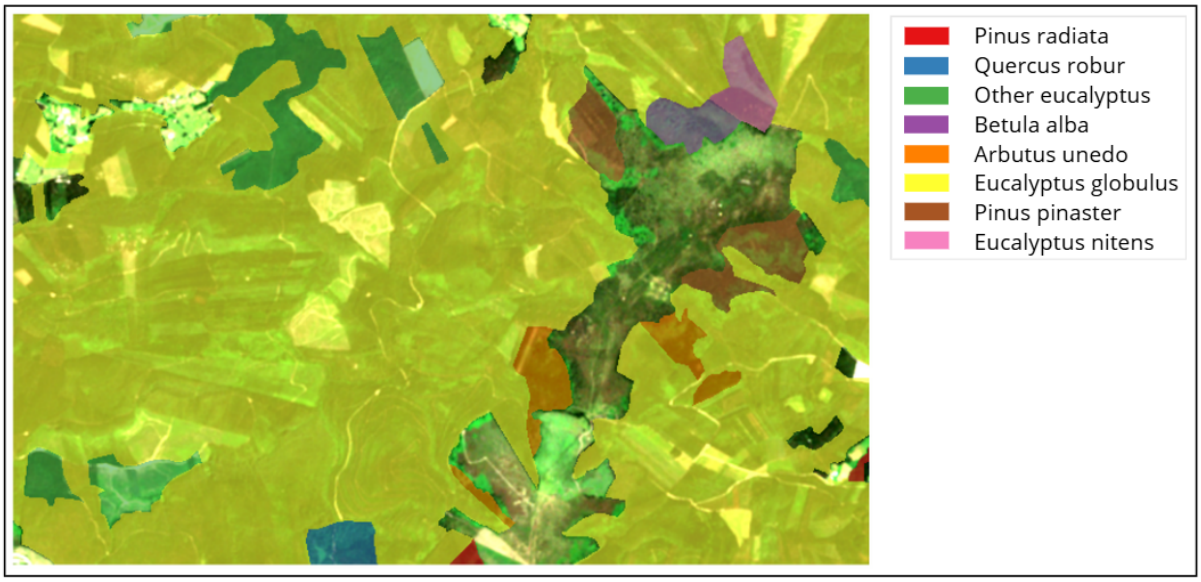


Figure 39: True Label TS map for the Trasbar region

With the trained hybrid DL model trained, applying them to the masked Sentinel-2 image of the Trasbar region, a predicted TS map can be generated. Alongside this, a confidence map is also produced, indicating the model's certainty for each pixel's prediction. These maps are shown in Figures 40 and 41, respectively.

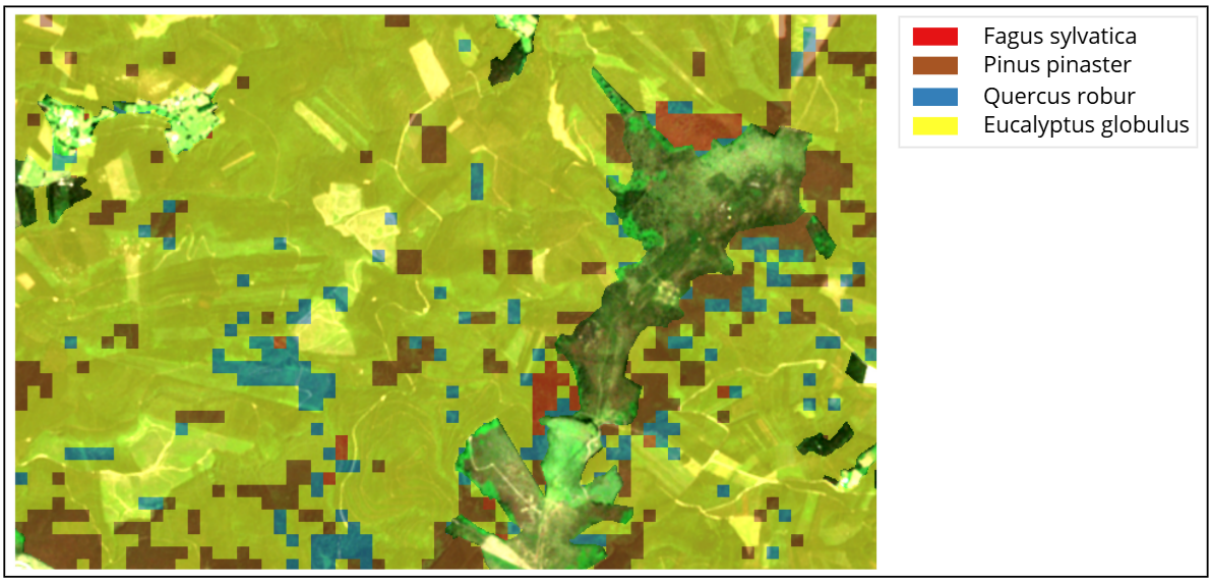


Figure 40: Predicted TS map for the Trasbar region

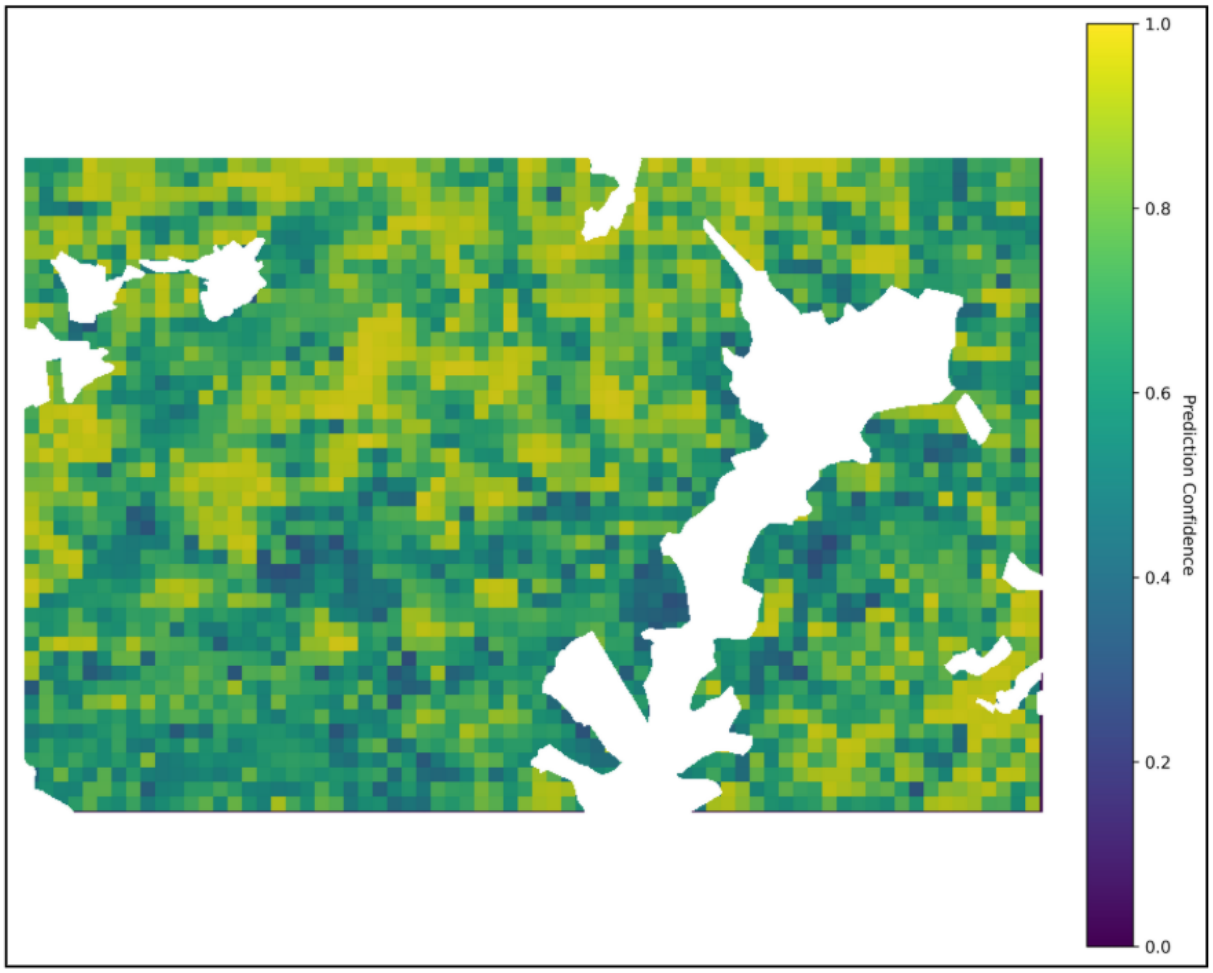


Figure 41: Model confidence map corresponding to the Trasbar region

5.5.2 Application over Portuguese Forests

A similar procedure was applied to the Serras de Fafe region, in northern Portugal, using the COS dataset [76] to filter out non-forested areas. The resulting TS map (Figure 42) and confidence map (Figure 43) show coherent patterns of *Quercus* trees across the forest patches.

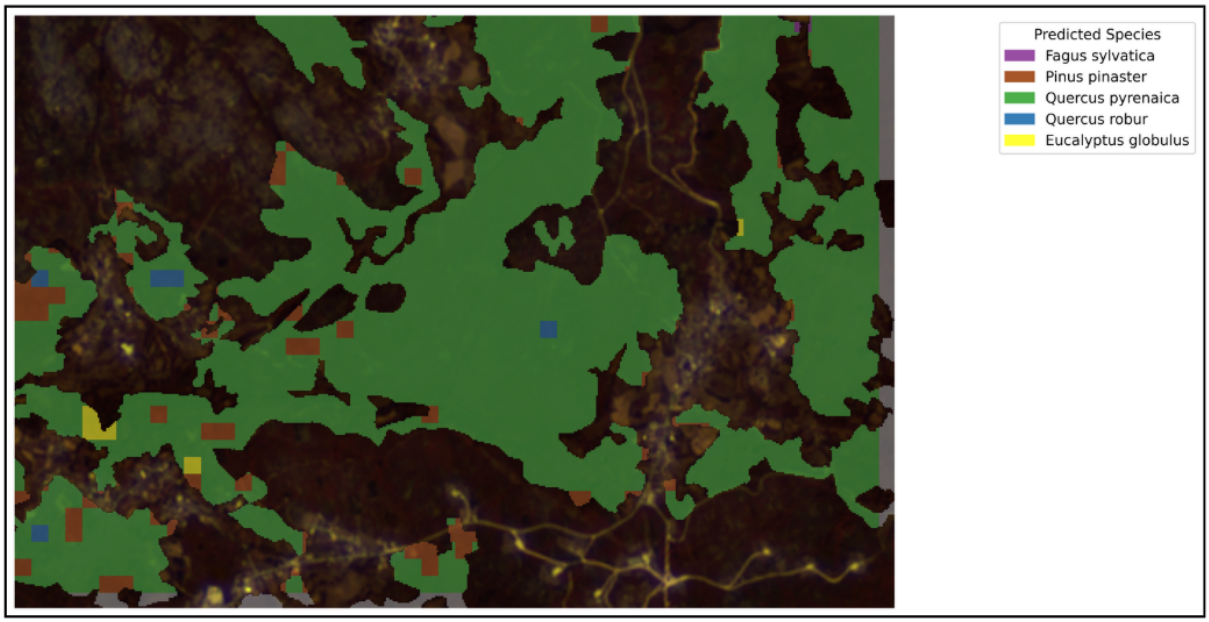


Figure 42: Predicted TS map for Serras de Fafe (Portugal).

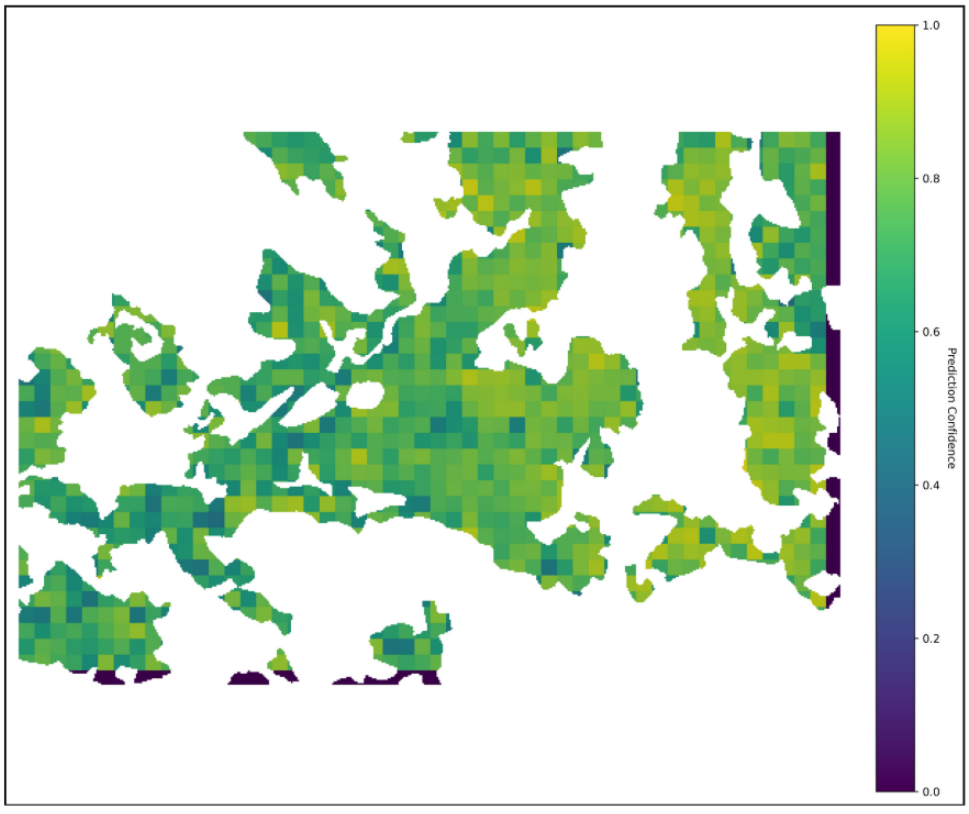


Figure 43: Model confidence map for the Serras de Fafe region (Portugal).

Overall, these qualitative maps illustrate the feasibility of applying the model to the entire Iberian Peninsula, producing visually coherent species predictions even in areas not included in the training dataset.

5.6 Conclusion of the Results Chapter

This chapter presented the experimental and qualitative results obtained through the implementation of the hybrid DL framework for TS classification across the Iberian Peninsula. The progressive class expansion analysis revealed that model performance decreases with increasing taxonomic complexity, with the 8-class model being the one with the best results (OA = 80.49%, F1-score = 80.01%) and the 15-class configuration emerging as the most balanced compromise between coverage and accuracy (OA = 73.09%, F1 = 67.12%).

The architectural experiments showed that single-season imagery, particularly from summer, outperformed multi-seasonal configurations, which was not expected, while genus-based and hybrid genus-species approaches did not yield substantial improvements. Benchmark comparisons further demonstrated the superiority of the proposed hybrid model over traditional TML methods, with clear performance gaps in both OA and F1-score. The ablation studies validated the complementary role of spectral and environmental data, highlighting the advantages of multimodal fusion over unimodal approaches.

Finally, the qualitative applications over Spanish and Portuguese forest regions demonstrated the spatial coherence and generalization ability of the model, producing ecologically plausible species distributions even beyond the training domain. Together, these findings confirm that the proposed hybrid architecture is capable of scalable, robust, and transferable species classification across heterogeneous forest landscapes.

The following chapter interprets these results in the context of existing literature, discussing their implications, limitations, and practical relevance for large-scale forest monitoring and management.

Chapter 6

Discussion

After presenting the results of the experiments in the previous chapter, this chapter aims to discuss and interpret those findings in the context of existing literature and practical applications. The discussion will begin with an overview of the key results, highlighting the main outcomes of the study. This will be followed by a comparison with existing studies in the field, identifying consistencies and novel contributions. The practical implications of the findings will then be explored, considering how they can inform forestry management and ecological monitoring practices. Next, the limitations of the study will be acknowledged, discussing potential sources of bias or uncertainty that may have influenced the results. Finally, suggestions for future research directions will be provided, outlining areas where further investigation could build upon the current work and address its limitations.

6.1 Overview of Key Results

The goal of this study was to develop, evaluate, and validate hybrid DL models for TS classification across the Iberian Peninsula, integrating Sentinel-2 imagery with auxiliary environmental data. In addition to quantitative assessments, the study produced qualitative TS maps for selected forested areas in Spain and Portugal, enabling a spatially explicit evaluation of model performance and generalization. The Portuguese validation, conducted using the COS dataset, provided an important cross-border test of interoperability and practical applicability.

From the experimental and mapping results presented in the previous chapter, several key findings can be drawn. Firstly, the experiments assessing the optimal number of classes are examined. Following this, the performance of the main hybrid DL model is compared with traditional benchmark algorithms. The influence of alternative architectures and data modalities is then discussed, culminating in the evaluation

of the model's transferability through the Portuguese case study.

Overall, the hybrid model combining imagery and auxiliary variables achieved an OA of 80.49% and an F1-score of 80.01% when classifying the 8 most dominant species, which represent around 80% of the dataset, and an OA of 73.09% and an F1-score of 67.12% when classifying the 15 most represented species (≈92 % of the dataset). These values establish the quantitative baseline for the analyses discussed in the subsequent subsections and demonstrate the feasibility of large-scale TS mapping.

6.1.1 Multiple Class Models

The results from the multiple-class models reveal a clear trade-off between model complexity and performance. The 8-class configuration achieved strong results for the scale of this study, but as the number of classes increased, both OA and F1-score declined. This trend is expected, as adding more classes typically introduces rarer species with fewer training samples and greater spectral overlap, increasing classification difficulty. Consequently, the inclusion of these underrepresented species explains the steady reduction in F1-score and the stabilization of OA at higher class counts. In this context, the 15-class model was selected as an optimal balance between coverage and performance, maximizing the number of species while maintaining acceptable accuracy levels for practical use.

Examining the class-level results provides further insight into model behaviour. Species *Pinus radiata*, for instance, exhibits low accuracy due to its minimal representation in the dataset. Conversely, species *Quercus robur* and *Quercus faginea*, which have adequate sample sizes, show reduced accuracy likely caused by spectral similarity or overlapping phenological traits that hinder separability. The confusion matrix supports this interpretation: species *Quercus robur* is misclassified across several unrelated species, while species *Quercus faginea* is primarily confused with species *Quercus ilex ssp. ballota*, a dominant and taxonomically similar species from the same genus. Finally, the analysis of prediction confidence indicates a positive relationship between confidence and accuracy, suggesting that the model's confidence estimates can serve as a reliable indicator of classification reliability. This finding is relevant for practical applications, where only predictions exceeding a defined confidence threshold could be considered operationally trustworthy.

6.1.2 Main Deep Learning Model vs Benchmark Models

The comparison between the hybrid DL model and traditional benchmark algorithms provides important insights into model performance under complex RS conditions. While most studies identify RF as the strongest performer among traditional methods, in this case, XGBoost surpassed it by a clear margin, achieving higher OA and F1-score. This divergence from the common trend reported in the literature suggests that XGBoost's gradient boosting structure may better handle the heterogeneous spectral and

environmental patterns characteristic of the Iberian Peninsula. Nonetheless, the performance of all benchmark models remained moderate, which can be attributed to the large spatial extent of the study area, the limited spatial resolution of Sentinel-2 imagery, and the strong class imbalance present in the dataset. These factors constrain the discriminative power of models that rely solely on aggregated tabular or statistical features.

In contrast, the hybrid DL model demonstrated a marked improvement across all metrics, confirming its capacity to integrate MSI with environmental variables effectively. Its ability to jointly learn spatial, spectral, and contextual patterns likely contributes to its superior generalization, even under challenging conditions of data imbalance and regional variability. This robustness supports the suitability of the hybrid architecture for large-scale TS classification across heterogeneous forest ecosystems and underscores the growing advantage of multi-modal DL approaches over TML techniques in operational RS applications.

6.1.3 Discussion of Architecture Evaluation Results

The comparative analysis of the different architectural configurations provides important insights into model behaviour under varying data and design constraints.

Contrary to common findings in the literature, the multi-seasonal model yielded the lowest accuracy among all tested configurations. While multi-temporal fusion often improves model robustness, in this case, the low spatial resolution of Sentinel-2 (10 m) likely introduced noise when combining images from different seasons. Such temporal variability may have overshadowed subtle spectral cues essential for species discrimination, explaining the superior performance of the single-season (summer) model.

The genus-level experiments demonstrated that coarser taxonomic granularity does not necessarily guarantee big jumps in accuracy. Although genus classification reduces the number of output classes, the heterogeneous nature of the aggregated “Other” class, comprising multiple distinct species, led to inconsistent patterns that the model struggled to learn. This result aligns with previous studies that reported performance degradation when heterogeneous groups are included to balance data distributions.

Region-based models showed promising results in certain autonomous regions, notably in *Castilla y León*, where forest composition is relatively homogeneous. However, the gains were not substantial enough to justify developing independent regional models, especially considering the additional computational cost and reduced generalization capability.

Lastly, the hybrid genus-species approach achieved intermediate performance, confirming that while the concept helps integrate rare species, it also introduces imbalance at the genus level. Some genera encompass highly diverse species with distinct spectral signatures, making generalization difficult at that scale.

Collectively, these findings reinforce the decision to adopt the single-season, species-level architecture as the standard configuration for the remaining experiments. This setup provided the best compromise

between model accuracy, ecological relevance, and computational efficiency.

6.1.4 The Role of Image and Auxiliary Data in a Hybrid Model

The ablation experiments highlight the relative contribution of each data modality to model performance. Both image and auxiliary data provide complementary information, yet the results show that spectral-spatial features derived from Sentinel-2 imagery play a more dominant role in species discrimination. This is consistent with the fact that MSI captures species-specific reflectance signatures and canopy structures directly, while environmental variables such as elevation, temperature, and precipitation add broader contextual cues that support generalization across ecological gradients. The improved performance of the hybrid configuration confirms the value of multimodal fusion for large-scale TS mapping, enabling the model to integrate fine spectral details with environmental context in a way that traditional unimodal systems cannot achieve.

6.1.5 Validation with Portuguese Data

The additional validation using Portuguese data further demonstrates the robustness and generalization capability of the proposed hybrid DL model. By applying the same workflow developed for Spain, combining Sentinel-2 imagery with auxiliary data and forest masks, the model successfully generated coherent TS maps for Portuguese regions. This cross-border validation is particularly important, as it confirms that the model can operate effectively under different land cover mapping standards and geospatial products.

The use of the COS [76] as a forest mask proved essential for this step. Although the COS differs from MFE in structure and classification nomenclature, its polygon-level forest delineations were sufficiently detailed to enable high-quality filtering of forested areas. The results over the Serras de Fafe region were visually consistent, with the *Quercus* genus dominating the results with good confidence levels. With these results, it is possible to conclude that the forest seen in the Sentinel-2 image consists mainly of *Quercus Pyrenaica* trees. This reinforces that the model's learned spectral–environmental relationships are transferable across the Iberian Peninsula.

This finding highlights a key practical implication: large-scale hybrid DL models trained on harmonized datasets can be reused across national boundaries, provided that compatible forest masks or land cover layers (such as COS or MFE) are available. Such interoperability supports the development of continental forest monitoring frameworks and underlines the relevance of open geospatial data initiatives like those of the MITECO.

These outcomes confirm that the proposed framework is not limited to the spatial or semantic boundaries of the training data. The model successfully generalized to Portuguese territories using a different national land-cover reference, highlighting its adaptability to varying geospatial standards. This reinforces

the potential of hybrid DL architectures for continental-scale forest monitoring, where models trained on harmonized datasets could be transferred across administrative borders with minimal adjustment.

6.2 Comparison with Existing Literature

Comparing the results of this study with existing literature reveals several points of consistency and notable new contributions. Many previous studies have demonstrated the effectiveness of DL for TS classification, often achieving high accuracy when the number of target species is limited. The observed decline in performance as the number of classes increases is consistent with these works, which highlight challenges related to class imbalance and spectral similarity among species with comparable canopy structures or phenological traits. Unlike most prior research that typically focuses on a fixed number of species, this study investigated the scalability of classification performance across multiple class configurations, providing valuable insight into the limits of DL when applied to heterogeneous ecosystems such as the Iberian Peninsula.

Another novel contribution of this work is the confidence-based analysis of predictions. While most prior studies report conventional performance metrics such as OA, precision, and recall, few explore the relationship between model confidence and classification reliability. The positive correlation observed in this study between confidence and accuracy offers a practical mechanism for post-classification filtering, where only predictions exceeding a defined confidence threshold may be considered reliable for operational use in ecological monitoring.

In contrast to much of the literature, this study found that the use of multi-seasonal imagery combined with auxiliary data did not improve model performance. Previous studies have often reported significant gains from multi-temporal or multi-seasonal approaches, as they capture phenological variation useful for species discrimination. However, in this case, the combination of multiple Sentinel-2 images led to a degradation in accuracy. This discrepancy may be attributed to the relatively low spatial resolution of Sentinel-2 imagery, which introduces spectral noise and reduces class separability when multi-season composites are used.

Overall, these results reinforce known limitations in large-scale TS classification while contributing novel insights regarding scalability, confidence, reliability, and the potential drawbacks of low-resolution multi-seasonal imagery in multimodal frameworks.

6.3 Practical Implications

The findings of this study carry several practical implications for forestry management and ecological monitoring. For operational applications, forest managers and environmental agencies may employ reduced class models that balance classification coverage and accuracy. Such models can support routine

forest monitoring by providing reliable maps of dominant species without the computational or data burdens associated with highly detailed classifications. Implementing confidence thresholds for predictions would further enhance reliability, ensuring that only highly probable classifications are incorporated into decision-making processes. This approach aligns with the initial goal of the project: to produce dependable, data-driven species maps that can assist management and policy actions.

The superior performance of the hybrid model also suggests that integrating environmental datasets (e.g., elevation, temperature, precipitation) with satellite imagery should be prioritized in future large-scale forest mapping workflows. This integration enables a more holistic characterization of forest ecosystems, improving the accuracy of species identification in heterogeneous regions.

In terms of biodiversity monitoring and carbon stock assessment, focusing on dominant and ecologically significant species may often provide sufficient information for most management and conservation objectives. Attempting fine-grained, species-level mapping with current medium-resolution data may be less reliable, particularly in spectrally complex or mixed-species forests. Nonetheless, the methodological framework established here can serve as a foundation for scalable, hybrid approaches as higher-resolution datasets and improved ground truth data become available.

Overall, this study offers a practical roadmap for leveraging DL in forestry applications while emphasizing the importance of class selection, data balance, and confidence-based decision strategies to ensure robust and operationally relevant outcomes.

6.4 Limitations of the Study

Although this study provides valuable insights into TS classification using DL, several limitations should be acknowledged. The first and most significant limitation concerns the imbalance of the dataset. Many rare species were underrepresented, which likely contributed to the lower performance observed in models trained with a larger number of classes. Future work should therefore focus on collecting more balanced datasets or implementing techniques specifically designed to address data scarcity. While genus-level classification is often used in the literature to mitigate this issue, this approach was not feasible in the present study due to an even greater imbalance at the genus level.

A second limitation relates to the availability and type of RS data. Many recent studies employ multi-sensor approaches that combine datasets with complementary characteristics, for example, using one sensor with high spectral resolution and another with high spatial resolution. In this study, only Sentinel-2 imagery was available, which constrained the spectral and spatial diversity of the input data. The relatively low spatial resolution of Sentinel-2 (10 m) likely limited the model's ability to differentiate between taxonomically similar species, particularly in heterogeneous or mixed forest stands. Future research could benefit from incorporating VHR imagery to improve separability and fine-scale species discrimination.

Computational constraints also limited the exploration of more complex ensemble or transformer-based DL architectures, which could potentially yield improved results. The computational time required to train these DL models is too high, with some more complex approaches reaching more than 12 hours to finish.

Overall, these limitations highlight the challenges of applying DL to large-scale ecological studies and emphasize the need for improved datasets, multi-sensor integration, and advanced learning strategies to further enhance the accuracy and scalability of TS classification models.

6.5 Future Research Directions

There remains significant scope for future research to build upon the findings of this study. Several promising avenues can be pursued to address the limitations identified and further advance TS classification using DL.

First, the adoption of advanced learning paradigms such as self-supervised learning and synthetic sample generation should be investigated to improve model robustness and reduce the impact of class imbalance, particularly for rare species. These methods can enable models to learn more generalized spectral-spatial representations without relying on extensive labeled datasets.

Second, combining Sentinel-2 imagery with complementary RS modalities, such as Sentinel-1 SAR data or VHR optical imagery, warrants further exploration. Such multi-sensor fusion approaches could enhance species separability by leveraging both structural and spectral information. In addition, integrating Portuguese ground-truth data would allow the expansion of the study area to encompass the entire Iberian Peninsula, enabling cross-regional model validation and a more comprehensive assessment of ecological variability.

Third, incorporating techniques from explainable AI (XAI) could help elucidate which spectral bands, vegetation indices, or environmental variables most strongly influence model decisions. Such interpretability is essential not only for improving model transparency and trustworthiness but also for generating ecologically meaningful insights into the spectral behaviour of different species.

Finally, future efforts should focus on the operationalization of the developed models. Transforming them into accessible, user-oriented tools for forestry services would bridge the gap between research and practice. This could involve creating interactive interfaces that enable users to visualize model outputs, adjust confidence thresholds, and customize parameters according to management objectives and risk tolerances.

Overall, these future research directions provide a roadmap for enhancing the accuracy, interpretability, and practical applicability of RS-based TS classification, contributing to more informed and sustainable forest management and ecological monitoring practices across the Iberian Peninsula.

Chapter 7

Conclusion

7.1 Summary of Goals and Research Questions

The primary goal of this study was to develop a hybrid DL model capable of classifying and generating TS maps across the Iberian Peninsula using Sentinel-2 multispectral imagery combined with auxiliary environmental data. This objective was pursued through five specific goals and four guiding research questions, which are revisited and answered below.

Goal 1: Satellite Image Retrieval

High-quality Sentinel-2 L2A imagery was successfully retrieved for all study plots, using the Sentinel Hub Processing API to ensure consistent atmospheric correction and 10 m spatial resolution across seasons. Cloud-free mosaics were generated for four seasonal periods, establishing a harmonized multispectral dataset that supported large-scale model training and validation.

Goal 2: Integrate Tabular and Auxiliary Data

Ground truth data from the Spanish IFN4 inventory were integrated with topographic (SRTM DEM) and climatic variables (E-OBS) to create a multimodal dataset combining 21 spectral and nine environmental features per plot. This integration proved fundamental for contextual learning, enhancing species separability across diverse ecological gradients.

Goal 3: Model Development

A hybrid DL architecture combining a ResNet-50 branch with a fully connected environmental branch was developed. The resulting model achieved an OA of 80.49% and an F1-score of 80.01% when classifying the eight most represented species. This confirmed the feasibility of large-scale, multi-source TS classification and addressed the central research goal of accurate Spanish territory models.

Goal 4: Assess Model Performance

When benchmarked against TML models, RF, SVM, and XGBoost, the hybrid DL model consistently outperformed all alternatives, demonstrating the superiority of multimodal fusion (Research Question 3). The quantitative evaluation and confidence analysis revealed that predictions with high confidence values corresponded closely to correct classifications, allowing reliable post-filtering for operational use. These findings provide a clear methodological answer to how RS and hybrid learning can improve classification accuracy and reliability (Research Questions 1 and 4).

Goal 5: Generate Portuguese Tree Species Maps

Using the Portuguese COS dataset for forest masking, the hybrid model successfully produced high-resolution, coherent TS maps over the Serras de Fafe (northern Portugal). The results, dominated by *Quercus pyrenaica*, confirmed cross-border generalization and interoperability between the Spanish MFE and Portuguese COS frameworks. This outcome demonstrates that hybrid DL models trained on harmonized data can be effectively transferred across national boundaries with minimal adjustments, addressing Research Question 2.

7.2 Key Findings and Interpretation

Overall, the study shows that integrating spectral, spatial, and environmental data enables robust, scalable TS classification across heterogeneous ecosystems. The hybrid approach outperformed unimodal and traditional models, proving that contextual auxiliary data significantly enhance generalization across environmental gradients. The 15-class configuration achieved the best trade-off between accuracy and taxonomic resolution, while experiments with multi-seasonal imagery revealed that increased temporal depth can introduce spectral noise rather than improve separability at Sentinel-2's spatial resolution. Confidence-based interpretation emerged as a practical innovation, providing a mechanism to apply probabilistic filtering in operational mapping workflows.

7.3 Contributions to the Field

This thesis contributes to the state of the art in several ways:

- It establishes a reproducible hybrid DL framework for TS classification using openly available Sentinel-2 and environmental datasets.
- It systematically evaluates class scalability, quantifying the relationship between taxonomic complexity and model accuracy across multiple configurations.
- It introduces a confidence-based validation method that links model certainty to predictive reliability, enabling selective use of outputs for decision-making.
- It demonstrates cross-border transferability by applying the Spanish-trained model to Portuguese forests, providing the first Iberian-wide proof of concept for DL-based TS mapping.

7.4 Limitations and Outlook

Despite these contributions, several limitations were identified. Dataset imbalance remained a major constraint, particularly for rare species, leading to lower F1-scores in high-class models. The 10 m spatial resolution of Sentinel-2 limited fine-scale discrimination in mixed stands, and the computational demands of large DL architectures restricted exploration of more complex networks. Future work should therefore focus on expanding ground truth databases to achieve class balance, incorporating VHR or SAR imagery, and testing self-supervised or transformer-based approaches.

7.5 Future Directions and Broader Impact

Future research should aim to:

- Expand ground truth coverage to include Portuguese inventories, enabling full Iberian calibration and validation.
- Explore multi-sensor fusion (Sentinel-1 + Sentinel-2 + VHR) to improve spectral and structural characterization.
- Integrate explainable AI (XAI) techniques to identify which spectral and environmental variables most influence classification decisions.
- Operationalize the hybrid model through user-oriented mapping platforms that allow real-time visualization, parameter tuning, and confidence-based filtering.

Beyond academic value, the framework developed here offers a practical foundation for large-scale ecological monitoring. It supports data-driven management of the Iberian Peninsula forests, providing tools for species distribution assessment, biodiversity monitoring, and climate adaptation strategies.

7.6 Final Remarks

In conclusion, this study successfully achieved its main objective and answered its guiding research questions. It demonstrates that hybrid DL models integrating Sentinel-2 imagery with environmental data can accurately and efficiently classify dominant TS across the Iberian Peninsula. By validating cross-border applicability through the Portuguese case study, it confirms the potential for scalable, interoperable, and operational AI-based forest monitoring. The methodological innovations and open-data approach established here represent a meaningful step toward continental-scale ecosystem mapping and the integration of AI into sustainable forest management.

Bibliography

[1] M. Drusch, M. Honnayb, D. B. Ferreira, and G Mandorlo. *Sentinel-2 ESA's Optical High-Resolution Mission for GMES Operational Services*. 2012. url: https://www.researchgate.net/publication/272907479.

[2] L. Zhong, Z. Dai, P. Fang, Y. Cao, and L. Wang. *A Review: Tree Species Classification Based on Remote Sensing Data and Classic Deep Learning-Based Methods*. May 2024. doi: 10.3390/f15050852.

[3] B Felbermeier, A Hahn, and T Schneider. *Study on user requirements for remote sensing applications in forestry*. 2010.

[4] F. Fang, B. E. McNeil, T. A. Warner, A. E. Maxwell, G. A. Dahle, E. Eutsler, and J. Li. "Discriminating tree species at different taxonomic levels using multi-temporal WorldView-3 imagery in Washington D.C., USA." In: *Remote Sensing of Environment* 246 (Sept. 2020), p. 111811. issn: 0034-4257. doi: 10.1016/J.RSE.2020.111811.

[5] Y. Sun, J. Huang, Z. Ao, D. Lao, and Q. Xin. "Deep Learning Approaches for the Mapping of Tree Species Diversity in a Tropical Wetland Using Airborne LiDAR and High-Spatial-Resolution Remote Sensing Images." In: *Forests 2019, Vol. 10, Page 1047* 10 (11 Nov. 2019), p. 1047. issn: 1999-4907. doi: 10.3390/F10111047. url: https://www.mdpi.com/1999-4907/10/11/1047/htmhttps://www.mdpi.com/1999-4907/10/11/1047.

[6] M. Arasumani, M. Bunyan, and V. V. Robin. "Opportunities and challenges in using remote sensing for invasive tree species management, and in the identification of restoration sites in tropical montane grasslands." In: *Journal of Environmental Management* 280 (Feb. 2021), p. 111759. issn: 0301-4797. doi: 10.1016/J.JENVMAN.2020.111759.

[7] D. Chambers, C. Périé, N. Casajus, and S. de Blois. "Challenges in modelling the abundance of 105 tree species in eastern North America using climate, edaphic, and topographic variables." In: *Forest Ecology and Management* 291 (Mar. 2013), pp. 20–29. issn: 0378-1127. doi: 10.1016/J.FORECO.2012.10.046.

[8] S. Marconi, B. G. Weinstein, S. Zou, S. A. Bohlman, A. Zare, A. Singh, D. Stewart, I. Harmon, A. Steinkraus, and E. P. White. "Continental-scale hyperspectral tree species classification in the United States National Ecological Observatory Network." In: *Remote Sensing of Environment* 282 (Dec. 2022), p. 113264. issn: 0034-4257. doi: 10.1016/J.RSE.2022.113264.

[9] Ángeles Casas, M. García, R. B. Siegel, A. Koltunov, C. Ramírez, and S. Ustin. "Burned forest characterization at single-tree level with airborne laser scanning for assessing wildlife habitat." In: *Remote Sensing of Environment* 175 (Mar. 2016), pp. 231–241. issn: 0034-4257. doi: 10.1016/J.RSE.2015.12.044.

[10] A. Abdollahnejad and D. Panagiotidis. "Tree Species Classification and Health Status Assessment for a Mixed Broadleaf-Conifer Forest with UAS Multispectral Imaging." In: *Remote Sensing 2020, Vol. 12, Page 3722* 12 (22 Nov. 2020), p. 3722. issn: 2072-4292. doi: 10.3390/RS12223722. url: https://www.mdpi.com/2072-4292/12/22/3722/htmhttps://www.mdpi.com/2072-4292/12/22/3722.

[11] K. M. Potter, M. E. Escanferla, R. M. Jetton, G. Man, and B. S. Crane. "Prioritizing the conservation needs of United States tree species: Evaluating vulnerability to forest insect and disease threats." In: *Global Ecology and Conservation* 18 (Apr. 2019), e00622. issn: 2351-9894. doi: 10.1016/J.GECCO.2019.E00622.

[12] Y. Shi, A. K. Skidmore, T. Wang, S. Holzwarth, U. Heiden, N. Pinnel, X. Zhu, and M. Heurich. "Tree species classification using plant functional traits from LiDAR and hyperspectral data." In: *International Journal of Applied Earth Observation and Geoinformation* 73 (Dec. 2018), pp. 207–219. issn: 1569-8432. doi: 10.1016/J.JAG.2018.06.018.

[13] L. R. Gering, D. M. May, and K. B. Teuber. "The use of aerial photographs and angle-gauge sampling of tree crown diameters for forest inventories." In: *State-of-the-art methodology of forest inventory:a symposium proceedings; 1989 July 30-August 5; Syracuse, NY.* (1990). url: https://research.fs.usda.gov/treesearch/515.

[14] I. Balenović, A. S. Milas, and H. Marjanović. "A Comparison of Stand-Level Volume Estimates from Image-Based Canopy Height Models of Different Spatial Resolutions." In: *Remote Sensing 2017, Vol. 9, Page 205* 9 (3 Mar. 2017), p. 205. issn: 2072-4292. doi: 10.3390/RS9030205. url: https://www.mdpi.com/2072-4292/9/3/205/htmhttps://www.mdpi.com/2072-4292/9/3/205.

[15] M. Deur, M. Gašparović, and I. Balenović. "Tree species classification in mixed deciduous forests using very high spatial resolution satellite imagery and machine learning methods." In: *Remote Sensing* 12 (23 Dec. 2020), pp. 1–18. issn: 20724292. doi: 10.3390/rs12233926.

[16] W. Zhang, X. Liu, B. Xu, J. Liu, H. Li, X. Zhao, X. Luo, R. Wang, L. Xing, C. Wang, and H. Zhao. "Remote Sensing Classification and Mapping of Forest Dominant Tree Species in the Three Gorges Reservoir Area of China Based on Sample Migration and Machine Learning." In: *Remote Sensing* 16 (14 July 2024), p. 2547. issn: 20724292. doi: 10.3390/RS16142547/S1. url: https://www.mdpi.com/2072-4292/16/14/2547/htmhttps://www.mdpi.com/2072-4292/16/14/2547.

[17] J. Mccarthy. *WHAT IS ARTIFICIAL INTELLIGENCE?* Nov. 2007. url: http://www-formal.stanford.edu/jmc/.

[18] A. M. Turing. *COMPUTING MACHINERY AND INTELLIGENCE*. 1950.

[19] J. McCarthy, N. Rochester, M. L. Minsky, and C. E. Shannon. "A PROPOSAL FOR THE DARTMOUTH SUMMER RESEARCH PROJECT ON ARTIFICIAL INTELLIGENCE." In: (Aug. 1955).

[20] M. Minsky. *Steps Toward Artificial Intelligence*. 1961.

[21] A. Newell, J. C. Shaw, and H. A. Simon. "Report On A General Problem-Solving Program." In: (1959).

[22] G Valiant. *A Theory of the Learnable*. 1984.

[23] R. P. França, A. C. B. Monteiro, R. Arthur, and Y. Iano. "An overview of deep learning in big data, image, and signal processing in the modern digital age." In: *Trends in Deep Learning Methodologies: Algorithms, Applications, and Systems* (Jan. 2021), pp. 63–87. doi: 10.1016/B978-0-12-822226-3.00003-9.

[24] K. Santosh, N. Das, and S. Ghosh. "Introduction." In: *Deep Learning Models for Medical Imaging* (Jan. 2022), pp. 1–27. doi: 10.1016/B978-0-12-823504-1.00011-8.

[25] M. M. Miro□czuk and J. Protasiewicz. "A recent overview of the state-of-the-art elements of text classification." In: *Expert Systems with Applications* 106 (Sept. 2018), pp. 36–54. issn: 0957-4174. doi: 10.1016/J.ESWA.2018.03.058.

[26] A. S. Hasan, F. Sohel, D. Diepeveen, H. Laga, and M. G. Jones. "A survey of deep learning techniques for weed detection from images." In: *Computers and Electronics in Agriculture* 184 (May 2021), p. 106067. issn: 0168-1699. doi: 10.1016/J.COMPAG.2021.106067.

[27] M. R. M. Talabis, R. McPherson, I. Miyamoto, J. L. Martin, and D. Kaye. "Analytics Defined." In: *Information Security Analytics* (Jan. 2015), pp. 1–12. doi: 10.1016/B978-0-12-800207-0.00001-0.

[28] U. Ewuzie, O. P. Bolade, and A. O. Egbedina. "Application of deep learning and machine learning methods in water quality modeling and prediction: a review." In: *Current Trends and Advances in Computer-Aided Intelligent Environmental Data Engineering* (Jan. 2022), pp. 185–218. doi: 10.1016/B978-0-323-85597-6.00020-3.

[29] A. Goyal, V. Gupta, and M. Kumar. "Recent Named Entity Recognition and Classification techniques: A systematic review." In: *Computer Science Review* 29 (Aug. 2018), pp. 21–43. issn: 1574-0137. doi: 10.1016/J.COSREV.2018.06.001.

[30] D. Daoun, F. Ibnat, Z. Alom, Z. Aung, and M. Azim. "Reinforcement Learning: A Friendly Introduction." In: Aug. 2021, pp. 134–146. isbn: 978-3-030-84336-6. doi: 10.1007/978-3-030-84337-3_11.

[31] C. Mishra and D. L. Gupta. "Deep Machine Learning and Neural Networks: An Overview." In: *IAES International Journal of Artificial Intelligence (IJ-AI)* 6 (2 June 2017), p. 66. issn: 2089-4872. doi: 10.11591/ijai.v6.i2.pp66-73.

[32] J. A. Yacim and D. G. B. Boshoff. "Impact of artificial neural networks training algorithms on accurate prediction of property values." In: *Journal of Real Estate Research* 40 (3 2018), pp. 375–418. issn: 26911175. doi: 10.1080/10835547.2018.12091505.

[33] C. Cortes, V. Vapnik, and L. Saitta. *Support-Vector Networks Editor*. Tech. rep. 1995, pp. 273–297.

[34] E. García-Gonzalo, Z. Fernández-Muñiz, P. J. G. Nieto, A. B. Sánchez, and M. M. Fernández. "Hard-rock stability analysis for span design in entry-type excavations with learning classifiers." In: *Materials* 9 (7 July 2016). issn: 19961944. doi: 10.3390/ma9070531.

[35] I. D. Mienye and Y. Sun. "A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects." In: (2022). doi: 10.1109/ACCESS.2017.DOI.

[36] L. Breiman. *Random Forests*. Tech. rep. 2001, pp. 5–32.

[37] M. Y. Khan, A. Qayoom, M. S. Nizami, M. S. Siddiqui, S. Wasi, and S. M. K. U. R. Raazi. "Automated Prediction of Good Dictionary EXamples (GDEX): A Comprehensive Experiment with Distant Supervision, Machine Learning, and Word Embedding-Based Deep Learning Techniques." In: *Complexity* 2021 (2021). issn: 10990526. doi: 10.1155/2021/2553199.

[38] T. Chen and C. Guestrin. "XGBoost: A scalable tree boosting system." In: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Vol. 13-17-August-2016. Association for Computing Machinery, Aug. 2016, pp. 785–794. isbn: 9781450342322. doi: 10.1145/2939672.2939785.

[39] Z. Ali and A. Burhan. "Hybrid machine learning approach for construction cost estimation: an evaluation of extreme gradient boosting model." In: *Asian Journal of Civil Engineering* 24 (Apr. 2023), pp. 1–16. doi: 10.1007/s42107-023-00651-z.

[40] I. H. Sarker. *Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions*. Nov. 2021. doi: 10.1007/s42979-021-00815-1.

[41] Y. Lecun, Y. Bengio, and G. Hinton. *Deep learning*. May 2015. doi: 10.1038/nature14539.

[42] M. Abbasi, A. Shahraki, and A. Taherkordi. *Deep Learning for Network Traffic Monitoring and Analysis (NTMA): A Survey*. Mar. 2021. doi: 10.1016/j.comcom.2021.01.021.

[43] Z. Hao. "Deep learning review and discussion of its future development." In: *MATEC Web of Conferences* 277 (2019), p. 02035. doi: 10.1051/matecconf/201927702035.

[44] U. A. C. of Engineers. "EM 1110-1-2910." In: (2021).

[45] R. DeFries. "Remote Sensing and Image Processing." In: *Encyclopedia of Biodiversity: Second Edition* (Jan. 2013), pp. 389–399. doi: 10.1016/B978-0-12-384719-5.00383-X.

[46] D. F. Marble and D. J. Peuquet. "Geographic Information Systems and Remote Sensing." In: *Manual of remote sensing, 2nd edition. Vol. I* (Jan. 2017), pp. 71–85. issn: 0007-4632. doi: 10.1016/B978-0-08-100021-2.00004-3.

[47] Z. Zhang and J. C. Moore. "Remote Sensing." In: *Mathematical and Physical Fundamentals of Climate Change* (Jan. 2015), pp. 111–124. doi: 10.1016/B978-0-12-800066-3.00004-8. url: https://linkinghub.elsevier.com/retrieve/pii/B9780128000663000048.

[48] S. Gupta, R. K. Kohli, V. K. Garg, and A. Dhawan. *Remote Sensing GIS Applications in Environmental Sciences*. 2018. url: https://www.researchgate.net/publication/325382226.

[49] G. Saritha, T. Saravanan, K. Anbumani, and J. Surendiran. "Digital elevation model and terrain mapping using LiDAR." In: *Materials Today: Proceedings* 46 (Jan. 2021), pp. 3979–3983. issn: 2214-7853. doi: 10.1016/J.MATPR.2021.02.525.

[50] A. Michez, S. Bauwens, S. Bonnet, and P. Lejeune. "Characterization of Forests with LiDAR Technology." In: *Land Surface Remote Sensing in Agriculture and Forest* (Jan. 2016), pp. 331–362. doi: 10.1016/B978-1-78548-103-1.50008-X.

[51] J. Jia, H. Sun, C. Jiang, K. Karila, M. Karjalainen, E. Ahokas, E. Khoramshahi, P. Hu, C. Chen, T. Xue, T. Wang, Y. Chen, and J. Hyyppä. "Review on Active and Passive Remote Sensing Techniques for Road Extraction." In: *Remote Sensing 2021, Vol. 13, Page 4235* 13 (21 Oct. 2021), p. 4235. issn: 2072-4292. doi: 10.3390/RS13214235. url: https://www.mdpi.com/2072-4292/13/21/4235/htmhttps://www.mdpi.com/2072-4292/13/21/4235.

[52] A. Tsokas, M. Rysz, P. M. Pardalos, and K. Dipple. "SAR data applications in earth observation: An overview." In: *Expert Systems with Applications* 205 (Nov. 2022), p. 117342. issn: 0957-4174. doi: 10.1016/J.ESWA.2022.117342.

[53] A. M. Lechner, G. M. Foody, and D. S. Boyd. "Applications in Remote Sensing to Forest Ecology and Management." In: *One Earth* 2 (5 May 2020), pp. 405–412. issn: 2590-3322. doi: 10.1016/J.ONEEAR.2020.05.001.

[54] T. Wellmann, A. Lausch, E. Andersson, S. Knapp, C. Cortinovis, J. Jache, S. Scheuer, P. Kremer, A. Mascarenhas, R. Kraemer, A. Haase, F. Schug, and D. Haase. "Remote sensing in urban planning: Contributions towards ecologically sound policies?" In: *Landscape and Urban Planning* 204 (Dec. 2020), p. 103921. issn: 0169-2046. doi: 10.1016/J.LANDURBPLAN.2020.103921.

[55] C. V. Westen. *REMOTE SENSING FOR NATURAL DISASTER MANAGEMENT*.

[56] M. Amani, A. Moghimi, S. M. Mirmazloumi, B. Ranjgar, A. Ghorbanian, S. Ojaghi, H. Ebrahimy, A. Naboureh, M. E. Nazari, S. Mahdavi, S. H. A. Moghaddam, R. M. Asiyabi, S. A. Ahmadi, S. Mehravar, F. Mohseni, and S. Jin. *Ocean Remote Sensing Techniques and Applications: A Review (Part I)*. Nov. 2022. doi: 10.3390/w14213400.

[57] A. S. Almohsen. "Challenges Facing the Use of Remote Sensing Technologies in the Construction Industry: A Review." In: *Buildings 2024, Vol. 14, Page 2861* 14 (9 Sept. 2024), p. 2861. issn: 2075-5309. doi: 10.3390/BUILDINGS14092861. url: https://www.mdpi.com/2075-5309/14/9/2861/htmhttps://www.mdpi.com/2075-5309/14/9/2861.

[58] H. Wu and Z. L. Li. "Scale Issues in Remote Sensing: A Review on Analysis, Processing and Modeling." In: *Sensors (Basel, Switzerland)* 9 (3 Mar. 2009), p. 1768. issn: 14248220. doi: 10.3390/S90301768. url: https://pmc.ncbi.nlm.nih.gov/articles/PMC3345842/.

[59] O. Dubovik, G. L. Schuster, F. Xu, Y. Hu, H. Bösch, J. Landgraf, and Z. Li. "Grand Challenges in Satellite Remote Sensing." In: *Frontiers in Remote Sensing* 2 (Feb. 2021), p. 619818. issn: 26736187. doi: 10.3389/FRSEN.2021.619818/BIBTEX. url: www.frontiersin.org.

[60] H. Han, Z. Liu, J. Li, and Z. Zeng. "Challenges in remote sensing based climate and crop monitoring: navigating the complexities using AI." In: *Journal of Cloud Computing* 13 (1 Dec. 2024), pp. 1–14. issn: 2192113X. doi: 10.1186/S13677-023-00583-8/FIGURES/3. url: https://journalofcloudcomputing.springeropen.com/articles/10.1186/s13677-023-00583-8.

[61] N. R. Haddaway, M. J. Page, C. C. Pritchard, and L. A. McGuinness. "PRISMA 2020: An updated checklist for systematic reviews and meta-analyses." In: *Nature Protocols* 17.3 (2022), pp. 384–403. doi: 10.1038/s41596-021-00645-8.

[62] H. P. Vaghela and R. A. A. Raja. "Automatic Identification of Tree Species From Sentinel-2A Images Using Band Combinations and Deep Learning." In: *IEEE Geoscience and Remote Sensing Letters* 21 (2024). issn: 15580571. doi: 10.1109/LGRS.2024.3354814.

[63] E. Grabska, D. Frantz, and K. Ostapowicz. “Evaluation of machine learning algorithms for forest stand species mapping using Sentinel-2 imagery and environmental data in the Polish Carpathians.” In: *Remote Sensing of Environment* 251 (Dec. 2020), p. 112103. issn: 00344257. doi: 10.1016/j.rse.2020.112103. url: https://www.scopus.com/record/display.uri?eid=2-s2.0-85091593190&doi=10.1016%2fj.rse.2020.112103&origin=inward&txGid=7ab563c071d6afb6d3a578d78c347d84.

[64] Y. He, J. Yang, J. Caspersen, and T. Jones. “An operational workflow of deciduous-dominated forest species classification: Crown delineation, gap elimination, and object-based classification.” In: *Remote Sensing* 11 (18 Sept. 2019), p. 2078. issn: 20724292. doi: 10.3390/rs11182078. url: https://www.scopus.com/record/display.uri?eid=2-s2.0-85072634921&doi=10.3390%2frs11182078&origin=inward&txGid=9098e470013cbb4e737ee5ba01b17abd.

[65] C. Chen, L. Jing, H. Li, and Y. Tang. “A new individual tree species classification method based on the resu-net model.” In: *Forests* 12 (9 Sept. 2021). issn: 19994907. doi: 10.3390/F12091202.

[66] X. Wang, J. Wang, Z. Lian, and N. Yang. “Semi-Supervised Tree Species Classification for Multi-Source Remote Sensing Images Based on a Graph Convolutional Neural Network.” In: *Forests* 14 (6 June 2023). issn: 19994907. doi: 10.3390/F14061211.

[67] G. F. Watch. *Global Forest Watch – Natural Forests Map Layer*. https://www.globalforestwatch.org/. World Resources Institute (WRI), accessed 26 October 2025. 2025.

[68] S. Hub. *Processing API — Sentinel Hub*. https://docs.sentinel-hub.com/api/latest/api/process/. Accessed: YYYY-MM-DD.

[69] M. Main-Knorn, B. Pflug, J. Louis, F. Gascon, and U. Müller-Wilm. “Sen2Cor for Sentinel-2.” In: *Remote Sensing* (2017). Abstract only; atmospheric correction processor.

[70] J. Xue and B. Su. *Significant remote sensing vegetation indices: A review of developments and applications*. 2017. doi: 10.1155/2017/1353691.

[71] M. for the Ecological Transition and the Demographic Challenge (MITECO). *Cuarto Inventario Forestal Nacional (IFN4)*. https://www.miteco.gob.es/en/biodiversidad/temas/inventarios-nacionales/inventario-forestal-nacional/cuarto_inventario.html. Accessed: YYYY-MM-DD.

[72] N. JPL. *NASA Shuttle Radar Topography Mission Global 1 arc second (SRTMGL1) Version 003*. Void-filled ”SRTM Plus”; elevations in meters relative to WGS84/EGM96 geoid; accessed 2025-10-14. 2013. doi: 10.5067/MEASURES/SRTM/SRTMGL1.003. url: https://doi.org/10.5067/MEASURES/SRTM/SRTMGL1.003.

[73] R. C. Cornes, G. van der Schrier, E. J. van den Besselaar, and P. D. Jones. “An ensemble version of the E-OBS temperature and precipitation datasets.” In: *Earth System Science Data* 10.1 (2018), pp. 233–247. doi: 10.5194/essd-10-233-2018. url: https://doi.org/10.5194/essd-10-233-2018.

[74] P. D. Team. *pyproj: Transformer documentation*. https://pyproj4.github.io/pyproj/stable/api/transformer.html. Accessed: YYYY-MM-DD.

[75] N. Pettorelli, J. O. Vik, A. Mysterud, J.-M. Gaillard, C. J. Tucker, and N. C. Stenseth. “Using the satellite-derived NDVI to assess ecological responses to environmental change.” In: *Trends in Ecology Evolution* 20.9 (2005), pp. 503–510. issn: 0169-5347. doi: https://doi.org/10.1016/j.tree.2005.05.011. url: https://www.sciencedirect.com/science/article/pii/S016953470500162X.

[76] Direção-Geral do Território (DGT). *Carta de Uso e Ocupação do Solo (COS) - Série 2*. https://dados.gov.pt/pt/datasets/carta-de-uso-e-ocupacao-do-solo-cos-serie-2-nova/. Versão COS2023v1. Dados oficiais do Sistema de Monitorização da Ocupação do Solo (SMOS). Acesso em 26 de outubro de 2025. 2025.

[77] L. AI. *PyTorch Lightning*. https://github.com/Lightning-AI/pytorch-lightning. Accessed: YYYY-MM-DD.

[78] K. He, X. Zhang, S. Ren, and J. Sun. “Deep Residual Learning for Image Recognition.” In: *CVPR*. 2016.

[79] O. Russakovsky, J. Deng, H. Su, and et al. “ImageNet Large Scale Visual Recognition Challenge.” In: *International Journal of Computer Vision* 115.3 (2015), pp. 211–252. doi: 10.1007/s11263-015-0816-y.