

Integração de Inteligência Artificial com Deteção Remota para a Monitorização do Uso e da Ocupação do Solo

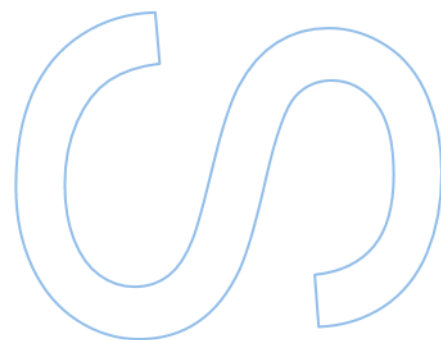
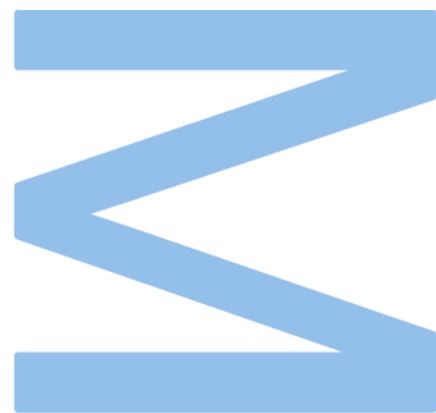
Gonçalo Mendonça Afonso Arantes

Mestrado em Engenharia Geoespacial

Departamento de Geociências, Ambiente e Ordenamento do Território

Faculdade de Ciências da Universidade do Porto

2025



Integração de Inteligência Artificial com Detecção Remota para a Monitorização do Uso e da Ocupação do Solo

Gonçalo Mendonça Afonso Arantes

Relatório de Estágio realizado no âmbito do Mestrado em
Engenharia Geoespacial

Departamento de Geociências, Ambiente e Ordenamento do Território
2025

Orientador

José Tadeu Marques Aranha, Professor Catedrático, Universidade
de Trás-os-Montes e Alto Douro

Coorientador

Maria Clara Gomes Quadros Lázaro da Silva, Professora Auxiliar,
Faculdade de Ciências da Universidade do Porto

Supervisor externo na Entidade de Acolhimento

André Dias, Diretor do Espaço, CEiiA



Agradecimentos

Gostaria de expressar a minha gratidão a todos aqueles que, de forma direta ou indireta, contribuíram para a realização deste estágio.

Agradeço aos Professores José Aranha e Clara Lázaro, orientadores por parte da faculdade, pelo acompanhamento e orientações ao longo de todo o processo.

Manifesto igualmente o meu agradecimento ao André Dias, orientador por parte do CEiiA, e a toda a equipa Space Digital pela forma como me receberam e integraram no grupo. Em particular à Cristina Errea, ao António Linhares e ao Tomás Silva, pela disponibilidade diária, pelo apoio técnico e pelo contributo para um ambiente de trabalho positivo, que muito favoreceu o meu crescimento profissional e pessoal.

Agradeço ainda ao CEiiA e à Universidade do Porto pela oportunidade concedida para a realização deste estágio.

De igual forma, agradeço a todo o corpo docente e a todos os meus colegas da Licenciatura e do Mestrado em Engenharia Geoespacial, por me terem permitido adquirir uma base sólida de conhecimento ao longo dos cinco anos do curso, que se revelou importante no desenvolvimento do trabalho.

Por fim, expresso a minha gratidão aos meus familiares e amigos, que sempre me apoiaram e acompanharam ao longo da minha trajetória académica.

Resumo

Num contexto atual marcado pelas alterações climáticas e pelos crescentes desafios da sustentabilidade, torna-se essencial monitorizar de forma rigorosa o uso e a ocupação do solo (*Land Use and Land Cover*, LULC). Os avanços recentes na deteção remota e na inteligência artificial (IA), em particular no *machine learning* e *deep learning*, oferecem ferramentas poderosas para enfrentar este desafio. Assim, este estágio, desenvolvido no **CEiiA**, teve como principal objetivo a aplicação destas técnicas, visando o desenvolvimento de soluções inovadoras que contribuam para a monitorização e gestão sustentável do território. Neste âmbito, foram conduzidos dois projetos complementares.

O primeiro consistiu no desenvolvimento de um modelo de IA para deteção de mudanças de LULC na Península Ibérica. O processo envolveu a criação de *datasets* anotados com polígonos das várias classes de LULC, o processamento de imagens Sentinel-1 e Sentinel-2 e o treino de um classificador *Random Forest* com mosaicos multiespetrais. Os resultados evidenciam um desempenho satisfatório para uma fase inicial, embora persistam limitações, como a classificação incorreta de áreas com sombra como corpos de água e, também, a dificuldade na distinção entre classes de vegetação. Entre as soluções apontadas, destacam-se o reforço do treino com polígonos adicionais, a integração de variáveis auxiliares e a exploração futura de abordagens avançadas de *deep learning*.

O segundo projeto focou-se na preparação de *datasets* de vinhas do concelho de Arcos de Valdevez a partir de imagens GEOSAT-2, incluindo o pré-processamento necessário para o treino de modelos de *deep learning* com segmentação semântica. Este terá como principal objetivo a deteção automática de vinhas. Este projeto abre caminho não apenas para o cadastro automático de vinhas, mas também para aplicações em agricultura de precisão.

Em suma, o trabalho desenvolvido permitiu consolidar uma base robusta para a monitorização do LULC com resultados encorajadores, ao mesmo tempo que assegurou a preparação rigorosa dos dados necessários para a futura deteção automática de vinhas, abrindo perspetivas promissoras para aplicações mais avançadas neste domínio.

Palavras-chave: Uso e Ocupação do Solo, Detecção Remota, *Machine Learning*, *Random Forest*, Sentinel, GEOSAT, Vinhas.

Abstract

In the current context marked by climate change and the growing challenges of sustainability, it is essential to rigorously monitor land use and land cover (LULC). Recent advances in remote sensing and artificial intelligence (AI), particularly in machine learning and deep learning, offer powerful tools to address this challenge. Thus, this internship, carried out at **CEiiA**, had as its main objective the application of these techniques, aiming at the development of innovative solutions that contribute to the monitoring and sustainable management of the territory. Within this scope, two complementary projects were conducted.

The first consisted of the development of an AI model for LULC change detection in the Iberian Peninsula. The process involved the creation of annotated datasets with polygons of the various LULC classes, the processing of Sentinel-1 and Sentinel-2 images, and the training of a Random Forest classifier with multispectral composites. The results show a satisfactory performance for an initial stage, although some limitations remain, such as the misclassification of shadowed areas as water bodies and the difficulty in distinguishing between vegetation classes. Possible solutions include reinforcing the training dataset with additional polygons, integrating auxiliary variables, and exploring advanced deep learning approaches in the future.

The second project focused on the preparation of vineyard datasets from the municipality of Arcos de Valdevez using GEOSAT-2 imagery, including the preprocessing required for the training of deep learning models with semantic segmentation. The main goal is the automatic detection of vineyards. This project paves the way not only for automatic vineyard mapping but also for applications in precision agriculture.

Ultimately, the work developed consolidated a robust foundation for LULC monitoring with encouraging results, while also ensuring the rigorous preparation of data for the future automatic detection of vineyards, opening promising perspectives for more advanced applications in this domain.

Keywords: Land Use and Land Cover, Remote Sensing, Machine Learning, Random Forest, Sentinel, GEOSAT, Vineyards.

Índice

Lista de Tabelas	vi
Lista de Figuras	vii
Lista de Abreviaturas	ix
1. Introdução	1
1.1. Contextualização Teórica	1
1.1.1. Importância da Monitorização do Uso e Ocupação do Solo	1
1.1.2. Detecção Remota	4
1.1.3. Inteligência Artificial e Machine Learning	6
1.2. Enquadramento do Estágio e Objetivos	7
1.3. Estrutura do Relatório	8
2. Métodos de Detecção Remota e IA	9
2.1. Aplicação de ML e DL para a Monitorização do LULC	9
2.2. Classificação Supervisionada e <i>Random Forest</i>	13
3. Metodologia	15
3.1. Metodologia-LULC	15
3.1.1. Área de Estudo	15
3.1.2. Dados Utilizados	17
3.1.3. Pré-processamento das Imagens de Satélite	18
3.1.4. Anotações Manuais	22
3.1.5. Metodologia de Treino e de Avaliação do Modelo	27
3.1.6. Treino do Modelo	29
3.2. Metodologia-Vinhas	30
3.2.1. Área de Estudo	30
3.2.2. Dados utilizados	31
3.2.3. Preparação do Dataset de Treino	32
4. Resultados e Discussão	36
4.1. Resultados-LULC	36

4.1.1. Desempenho do modelo	36
4.1.2. Resultados das Inferências	39
4.1.3. Discussão	49
4.2. Resultados-Vinhas	51
4.2.1. Resultados	51
4.2.2 Discussão e Perspetivas Futuras	54
5. Conclusão	55
Referências Bibliográficas	57

Lista de Tabelas

Tabela 1-Revisão da Literatura relativamente à aplicação de ML e DL ao LULC;	10
Tabela 2-Lista das bandas e índices que compõem os mosaicos;	21
Tabela 3-Sistema de Classificação LULC adotado;	24
Tabela 4-Estatísticas do <i>dataset</i> após o processo de atualização;	52

Lista de Figuras

Figura 1- Evolução da temperatura média global e da temperatura à superfície terrestre desde o período pré-industrial (" <i>Special Report on Climate Change and Land</i> ", IPCC, 2019);	2
Figura 2-Evolução das emissões globais de CO ₂ , CH ₄ e N ₂ O provenientes de atividades do setor AFOLU (" <i>Special Report on Climate Change and Land</i> ", [4]);	3
Figura 3- Sensores passivos e ativos (NASA EarthData);.....	5
Figura 4-Mapa Climático da Península Ibérica baseado na classificação de Köppen-Geiger [30];.....	16
Figura 5-Exemplo dos mosaicos Ótico, SAR e Ótico-SAR de janeiro de 2021;.....	20
Figura 6-Polígonos gerados sobre Sevilha;	22
Figura 7- Variabilidade sazonal de um campo agrícola;.....	26
Figura 8-Polígonos LULC 01/2021 Sevilha;	26
Figura 9-Mapa climático com as áreas de teste do modelo [30];.....	30
Figura 10-Localização e Freguesias do Concelho de Arcos de Valdevez [65];	31
Figura 11-Polígono mal classificado como vinha;	33
Figura 12- Matriz de Confusão obtida após o treino do modelo RF com os <i>datasets</i> do Porto e de Palência;	37
Figura 13-Gráfico da importância das bandas obtido após o treino do modelo RF com os <i>datasets</i> do Porto e de Palência;.....	37
Figura 14-Gráfico das métricas de precisão, <i>recall</i> e <i>F1</i> -score por classe obtido após o treino do modelo RF com os <i>datasets</i> do Porto e de Palência;	38
Figura 15-Mapa LULC de Zamora obtido a partir de uma inferência realizada com o modelo RF, referente a agosto de 2024;.....	39
Figura 16-Mapa LULC com valores <i>NoData</i> obtido em Zamora, referente a dezembro de 2024;	40
Figura 17-Efeito causado pela sombra na classificação do modelo (Canfranc, janeiro 2024);	41
Figura 18-Classificação do modelo sem o efeito da sombra (Canfranc, maio 2024); ..	42
Figura 19-Efeito da sombras nas imagens Sentinel-2 ao longo do ano (Canfranc, 2024);	42
Figura 20-Classificação incorreta de <i>Bare Soil</i> como <i>Urban</i> (Maravillas, janeiro 2024);	44
Figura 21-Classificação incorreta de <i>Railways</i> como <i>Urban</i> (Maravillas, setembro 2024);	44

Figura 22-Erros de classificação (Zamora, janeiro 2024);	45
Figura 23-Mapa LULC (Zamora, fevereiro de 2024);.....	47
Figura 24-Mosaico de Zamora (fevereiro, 2024);	48
Figura 25-Composição do <i>dataset</i> de treino segundo as décadas de plantação das vinhas;	52
Figura 26-Dados pré-processados (<i>Tile</i> GEOSAT-2; Máscara binária, em que as áreas a preto representam as vinhas; Sobreposição da máscara sobre a imagem);	53

Lista de Abreviaturas

AFOLU	AGRICULTURE, FORESTRY, AND OTHER LAND USES
ANN	ARTIFICIAL NEURAL NETWORK
CART	CLASSIFICATION AND REGRESSION TREE
CEIIA	CENTRO DE ENGENHARIA E DESENVOLVIMENTO DO PRODUTO
C-SAR	C-BAND SYNTHETIC APERTURE RADAR
CNIG	CENTRO NACIONAL DE INFORMACIÓN GEOGRÁFICA
COS	CARTA DE OCUPAÇÃO DO SOLO
CPU	CENTRAL PROCESSING UNIT
CVRVV	COMISSÃO DE VITICULTURA DA REGIÃO DOS VINHOS VERDES
DEM	DIGITAL ELEVATION MODEL
DR	DETEÇÃO REMOTA
DGT	DIREÇÃO GERAL DO TERRITÓRIO
DT	DECISION TREE
ESA	EUROPEAN SPACE AGENCY
FAO	FOOD AND AGRICULTURE ORGANIZATION
FUZZY ARTMAP	FUZZY ADAPTIVE RESONANCE THEORY MAPPING
G-LIHT	GODDARD'S LIDAR HYPERSPECTRAL AND THERMAL
GNSS	GLOBAL NAVIGATION SATELLITE SYSTEM
GPU	GRAPHICS PROCESSING UNIT
GTB	GRADIENT-BOOSTED TREES
HIRAS	HIGH RESOLUTION ADVANCED MULTISPECTRAL SENSOR
IA	INTELIGÊNCIA ARTIFICIAL
IPCC	INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE
IVV	INSTITUTO DA VINHA E DO VINHO
KNN	K-NEAREST NEIGHBORS
LSAT	LAND SURFACE AIR TEMPERATURE
LSTM	LONG SHORT-TERM MEMORY
LULC	LAND USE AND LAND COVER
MD	MINIMUM DISTANCE
ML	MACHINE LEARNING

MODIS	MODERATE RESOLUTION IMAGING SPECTRORADIOMETER
MSI	MULTISPECTRAL IMAGER
OLI	OPERATION LAND IMAGER
RF	RANDOM FOREST
RNN	RECURRENT NEURAL NETWORK
SAM	SEGMENT ANYTHING MODEL
SAR	SYNTHETIC APERTURE RADAR
SID	SPECTRAL INFORMATION DIVERGENCE
SIOSE	SISTEMA DE INFORMACIÓN SOBRE OCUPACIÓN DEL SUELO DE ESPAÑA
SVM	SUPPORT VECTOR MACHINE
VANT	VEÍCULO AÉREO NÃO TRIPULADO

1. Introdução

1.1. Contextualização Teórica

1.1.1. Importância da Monitorização do Uso e Ocupação do Solo

Atualmente, o conceito de uso e ocupação do solo (*Land Use and Land Cover*, LULC) constitui uma base fundamental para a compreensão da interação entre as atividades humanas e os processos naturais. Embora os termos *Land Cover* e *Land Use* sejam frequentemente utilizados em conjunto, os seus significados são distintos. O primeiro refere-se às características biofísicas observáveis à superfície terrestre, como florestas, corpos de água, áreas agrícolas e zonas urbanas [1]. Por outro lado, o termo *Land Use* descreve a forma como os seres humanos utilizam essas superfícies, como por exemplo, através da agricultura, da expansão urbana ou da exploração florestal, enfatizando sobretudo as suas dimensões económicas e sociais [2]. Desta forma, alterações no LULC assumem uma elevada relevância a nível global, regional e nacional, uma vez que podem provocar efeitos profundos nos sistemas ecológicos, ambientais e socioeconómicos [3].

Nas últimas décadas, o papel do LULC tem ganho cada vez mais peso no contexto das alterações climáticas e da sustentabilidade ambiental. De acordo com o Painel Intergovernamental sobre Alterações Climáticas (*Intergovernmental Panel on Climate Change*, IPCC), as dinâmicas de uso do solo estão profundamente interligadas com o sistema climático, atuando simultaneamente como fator de pressão, através das emissões de gases com efeito de estufa, e como elemento vulnerável às consequências do aquecimento global. Segundo o *Special Report on Climate Change and Land*, publicado pelo IPCC [4], a superfície continental tem registado um aquecimento mais rápido do que os oceanos. A Figura 1 apresenta a evolução da temperatura média global desde o período pré-industrial até 2018, evidenciando que a temperatura do ar à superfície terrestre (*Land Surface Air Temperature*, LSAT) aumentou a um ritmo superior ao da temperatura média global combinada (oceano e continentes). Entre 1850–1900 (período pré-industrial) e 2006–2015, a superfície continental aqueceu, em média, 1.53 °C, enquanto a média global se situou em 0.87 °C, o que reforça a vulnerabilidade dos ecossistemas terrestres ao aquecimento global.

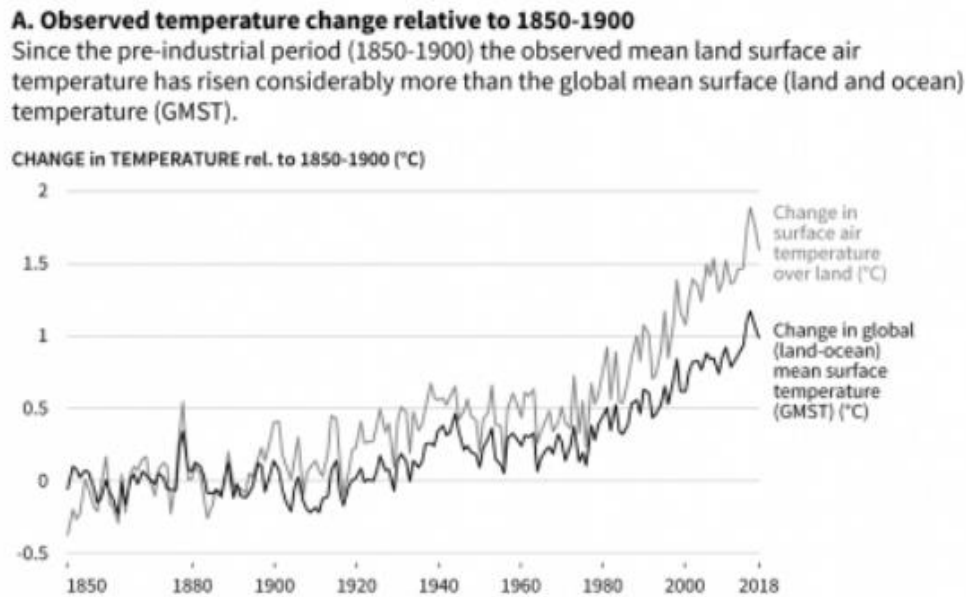


Figura 1- Evolução da temperatura média global e da temperatura à superfície terrestre desde o período pré-industrial ("Special Report on Climate Change and Land", IPCC, 2019);

Uma das principais causas deste fenómeno está nas emissões antropogénicas associadas ao setor da Agricultura, Florestas e Outros Usos do Solo (*Agriculture, Forestry and Other Land Uses*, AFOLU). Este setor foi responsável, no período de 2007–2016, por aproximadamente 13% das emissões globais de dióxido de carbono (CO₂), 44% das emissões de metano (CH₄) e 81% das emissões de óxido nitroso (N₂O), o que corresponde a cerca de 23% das emissões globais de gases com efeito de estufa (12,0 ± 2,9 GtCO₂-eq/ano) [4]. A Figura 2 apresenta a evolução histórica destas emissões desde 1961, evidenciando a sua contribuição significativa para as alterações climáticas. Observa-se uma tendência de crescimento contínuo na agricultura, associada ao aumento da produção alimentar e ao uso de fertilizantes, enquanto as emissões ligadas às florestas revelam maior variabilidade, refletindo episódios de desflorestação, degradação e políticas de conservação.

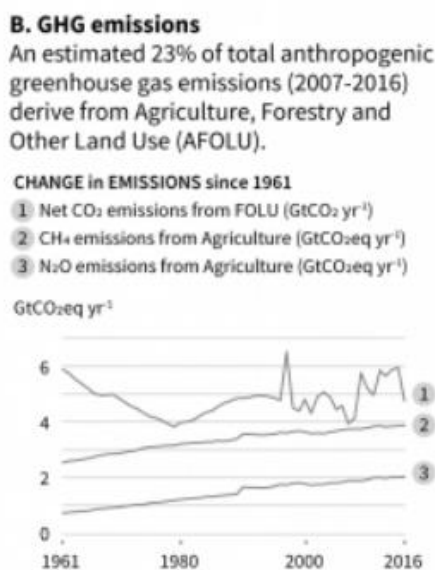


Figura 2-Evolução das emissões globais de CO₂, CH₄ e N₂O provenientes de atividades do setor AFOLU (“*Special Report on Climate Change and Land*”, [4]);

Dados recentes da Organização das Nações Unidas para a Alimentação e para Agricultura (*Food and Agriculture Organization*, FAO) relativos ao período entre 2001 e 2023 evidenciam mudanças relevantes na distribuição global dos usos do solo [5]. Em 2023, a área agrícola mundial correspondia a cerca de 4 800 milhões de hectares, representando mais de um terço da superfície terrestre sem gelo. Dentro desta categoria, aproximadamente 1 600 milhões de hectares correspondiam a terras de cultivo, enquanto prados e pastagens permanentes abrangiam 3 200 milhões de hectares. Simultaneamente, as florestas ocupavam 4 050 milhões de hectares e as restantes superfícies, como desertos, áreas geladas, terrenos áridos ou edificadas, cerca de 4 150 milhões de hectares. Entre 2001 e 2023, registou-se um aumento de cerca de 80 milhões de hectares de terras de cultivo, em contraste com uma redução de 150 milhões de hectares de pastagens, resultando numa ligeira diminuição (cerca de 2%) da área agrícola global. Apesar desta tendência, verificou-se uma intensificação considerável da atividade agrícola, traduzida num aumento de mais de 60% da produtividade por hectare.

Estas transformações socioambientais e climáticas, incluindo o aquecimento global e as alterações na distribuição e intensidade das áreas agrícolas e outros usos do solo, originam impactos significativos nos ecossistemas terrestres e no regime climático. O aquecimento global tem induzido deslocamentos das zonas climáticas, aumentando a frequência e intensidade de fenómenos extremos como ondas de calor, secas prolongadas e incêndios florestais. Adicionalmente, processos como a degradação dos

solos e a desertificação têm-se intensificado, especialmente em regiões tropicais e de alta latitude, onde os ecossistemas se tornam mais vulneráveis a condições ambientais para as quais não estão adaptados [4].

Neste enquadramento, a monitorização sistemática do LULC assume uma relevância crítica. Por um lado, permite compreender os impactos diretos e indiretos das alterações climáticas sobre os ecossistemas terrestres. Por outro, constitui uma ferramenta essencial para avaliar e apoiar estratégias de gestão sustentável do solo, cada vez mais reconhecidas como instrumentos de mitigação e adaptação às alterações climáticas.

Para além do contexto climático, a monitorização do LULC é igualmente crucial em ambientes urbanos, na gestão de recursos naturais, planeamento territorial e conservação da biodiversidade, reforçando o seu papel como ferramenta estratégica para a implementação de políticas ambientais e de desenvolvimento sustentável [6].

1.1.2. Detecção Remota

Antes da evolução das tecnologias espaciais e da consequente disponibilização sistemática de imagens de satélite, a monitorização do LULC recorria a métodos tradicionais, tais como levantamentos de campo, fotointerpretação de ortofotos, cartografia topográfica, levantamentos cadastrais e inquéritos locais. Embora estes métodos pudessem fornecer dados precisos a escalas locais, revelavam-se inviáveis para monitorização regional ou global, devido ao elevado tempo de execução, custos associados, intensa necessidade de mão de obra e suscetibilidade a erros de escala e desatualização, limitando a capacidade de acompanhar de forma eficaz as rápidas alterações no território [7].

A evolução das tecnologias de observação da Terra transformou profundamente a monitorização do LULC, superando as limitações dos métodos tradicionais. Neste contexto, a Detecção Remota (DR), entendida como o conjunto de técnicas de observação e monitorização que permitem obter informação sobre a superfície terrestre a partir da aquisição e interpretação de dados recolhidos por sensores instalados em satélites, aviões ou Veículos Aéreos Não Tripulados (VANT), surge como uma ferramenta essencial [8].

Os sensores utilizados na DR dividem-se em passivos e ativos (Figura 3). Os primeiros utilizam a radiação solar como fonte de energia, medindo a radiação eletromagnética refletida pelos objetos localizados na superfície da Terra e operam, maioritariamente,

nas regiões do visível e do infravermelho do espectro eletromagnético. Em contrapartida, os sensores ativos emitem a sua própria energia e registam a radiação eletromagnética refletida pela superfície terrestre, operando sobretudo na banda das micro-ondas, o que lhes permite operar em condições de elevada nebulosidade e durante a noite, contornando as limitações associadas aos sensores passivos.



Figura 3- Sensores passivos e ativos (NASA EarthData);

É importante salientar que a DR fornece, sobretudo, informação relativa ao *land cover*, isto é, às características biofísicas observáveis da superfície terrestre. A distinção entre os diferentes usos do solo (*land use*) requer, frequentemente, a integração desta informação com dados auxiliares, como registos cadastrais, planos de ordenamento do território, bases de dados ambientais, entre outros .

Como mencionado anteriormente, a deteção remota baseia-se na reflexão da radiação eletromagnética pela superfície terrestre. Cada material reflete, absorve ou transmite energia de forma característica, criando uma impressão digital espectral única, permitindo a identificação de diferentes tipos de solo, vegetação e corpos de água, bem como de formações rochosas e minerais. A resolução espectral de um sensor, ou seja, a capacidade que este possui de diferenciar comprimentos de onda específicos do espectro eletromagnético, influencia a capacidade de discriminar diferentes materiais e tipos de ocupação do solo [9].

Desta forma, a deteção remota permite mapear áreas extensas com elevada resolução temporal e detetar alterações na superfície terrestre que seriam impraticáveis de acompanhar por métodos tradicionais. Entre os satélites mais utilizados destacam-se os sistemas das missões Landsat e Sentinel, que fornecem dados multiespectrais com resoluções espaciais e temporais adequadas para análises detalhadas do LULC. Esta capacidade de observação sistemática torna a deteção remota uma ferramenta indispensável para avaliar os impactos das alterações climáticas, apoiar a gestão

sustentável dos recursos naturais e fornecer informações consistentes para o planeamento territorial e implementação de políticas ambientais.

1.1.3. Inteligência Artificial e Machine Learning

A Inteligência Artificial (IA) é um ramo da ciência da computação que pretende criar sistemas capazes de realizar tarefas que, normalmente, requerem inteligência humana, como o reconhecimento de padrões, a tomada de decisões e a previsão de resultados. Estes sistemas conseguem processar grandes volumes de dados, perceber o ambiente através de sensores, adaptar o seu comportamento com base em experiências passadas e agir autonomamente para atingir objetivos específicos. Embora algumas tecnologias de IA existam há várias décadas, os recentes avanços na capacidade computacional, na disponibilidade de grandes volumes de dados e no desenvolvimento de novos algoritmos impulsionaram progressos significativos, consolidando a IA como uma ferramenta central na transformação digital [10]. Atualmente, a IA é aplicada em diversas áreas, desde a saúde, com diagnósticos médicos assistidos por algoritmos, até à economia, com a previsão de mercados financeiros, passando pela robótica, transportes autónomos e, naturalmente, a monitorização ambiental e a deteção remota.

O *Machine Learning* (ML) é uma subárea da IA que se centra no desenvolvimento de algoritmos capazes de aprender padrões a partir de dados, sem necessidade de programação explícita para cada tarefa, melhorando progressivamente o desempenho do sistema à medida que novos dados são disponibilizados. No contexto da deteção remota, os modelos de ML assumem-se, atualmente, como uma das principais técnicas analíticas para mapear e quantificar o LULC, bem como monitorizar a sua evolução temporal. Algoritmos como *k-Nearest Neighbour* (kNN), *Support Vector Machines* (SVM), *Artificial Neural Networks* (ANN) e *Random Forest* (RF) têm sido amplamente aplicados com sucesso na classificação de diferentes tipos de LULC em múltiplas escalas geográficas, permitindo a classificação automática de imagens multiespectrais, a deteção de alterações temporais e a previsão de tendências de ocupação do solo [11]. A aplicação conjunta destas técnicas com dados de satélite possibilita análises consistentes e de elevada precisão, superando as limitações inerentes aos métodos tradicionais, que eram mais demorados, dispendiosos e menos adaptáveis a alterações rápidas no território. É importante ainda referir o crescente papel do *Deep Learning*, uma vertente do ML baseada em arquiteturas de redes neuronais profundas, que, embora

muito mais exigente em termos computacionais e de dados, tem demonstrado resultados particularmente promissores na classificação do LULC [12] [13].

1.2. Enquadramento do Estágio e Objetivos

O presente estágio curricular, realizado no âmbito do Mestrado em Engenharia Geoespacial, foi realizado no **CEiiA** (Centro de Engenharia e Desenvolvimento do Produto), no setor do espaço **downstream**, integrando a equipa **Space Digital** durante oito meses. Este teve como principal objetivo a aplicação de técnicas de DR e IA, com foco no *machine learning*, para o desenvolvimento de soluções inovadoras que contribuam para a monitorização e gestão sustentável do território.

Neste contexto, a integração na equipa **Space Digital** visou apoiar o cumprimento dos objetivos estratégicos do **CEiiA** através da participação em dois projetos distintos, mas interligados pela utilização de metodologias semelhantes e pela partilha de um mesmo enquadramento tecnológico.

O primeiro projeto centrou-se na monitorização do LULC na Península Ibérica, com o objetivo de desenvolver um modelo de IA capaz de detetar automaticamente alterações no território. Para esse fim, foram utilizadas imagens multiespectrais mensais provenientes dos satélites Sentinel-1 e Sentinel-2. Este trabalho envolveu o pré-processamento de dados geoespaciais, incluindo a preparação de *datasets* através de anotações manuais e semiautomáticas, bem como a implementação de *scripts* em **Python**, recorrendo a bibliotecas como *GDAL*, *GeoPandas* e *Rasterio*. Posteriormente, estes dados foram utilizados no treino do modelo, permitindo avaliar a sua performance e a capacidade de identificar alterações relevantes no LULC. Até ao momento de término do estágio, não existia ainda um volume elevado de resultados, sendo que os testes realizados se centraram mais em regiões circundantes a caminhos de ferro, de acordo com necessidades específicas de um cliente. Este projeto apresenta aplicações práticas diversas, como o ordenamento do território, a gestão de recursos naturais e o acompanhamento de dinâmicas ambientais e socioeconómicas.

O segundo projeto incidiu sobre a análise de um *dataset* de vinhas do concelho de Arcos de Valdevez, que incluía os registos históricos desta cultura. Este trabalho foi desenvolvido através da utilização de imagens de alta resolução do satélite **GEOSAT-2** e envolveu o pré-processamento dos dados com recurso a *scripts* em *Python*, incluindo a criação de máscaras e a geração de *tiles*. O objetivo final consistiu em preparar a

informação necessária para treinar um modelo de IA orientado à identificação automática de vinhas, com potencial para apoiar práticas de monitorização agrícola de elevada precisão.

Importa referir que este estágio foi desenvolvido na mesma equipa e no mesmo projeto em que participou igualmente o colega Tomás Silva. Dada a natureza conjunta do trabalho e a sobreposição de funções, algumas figuras e resultados apresentados neste relatório são comuns a ambos, refletindo a partilha do mesmo conjunto de dados e procedimentos relativamente ao projeto do LULC.

1.3. Estrutura do Relatório

O presente documento encontra-se organizado em cinco capítulos principais, apresentando de forma clara e estruturada o trabalho realizado ao longo do estágio.

O primeiro capítulo corresponde à introdução, na qual são apresentados os conceitos mais relevantes no contexto deste estágio, bem como os objetivos do mesmo e a estrutura do relatório.

Relativamente ao segundo capítulo, este apresenta o estado da arte, focando-se nos avanços recentes nas áreas da deteção remota e da inteligência artificial e nas suas aplicações relativas à classificação do LULC.

No terceiro capítulo é descrita a metodologia adotada ao longo do estágio, detalhando os dados utilizados e todos os procedimentos aplicados, organizados em duas secções distintas, correspondentes aos dois projetos desenvolvidos.

O quarto capítulo expõe e analisa os resultados obtidos em ambos os projetos.

Por fim, o último capítulo apresenta uma síntese das conclusões alcançadas neste estágio, bem como recomendações para trabalhos futuros relacionados com a continuidade e o desenvolvimento destes projetos.

2. Métodos de Detecção Remota e IA

2.1. Aplicação de ML e DL para a Monitorização do LULC

Atualmente, a extração dos limites de propriedade constitui uma componente fundamental no mapeamento do LULC, sobretudo em contextos que exigem a delimitação rigorosa de diferentes classes territoriais, como áreas urbanas, agrícolas ou florestais. A identificação precisa destes limites é indispensável para aplicações como o ordenamento do território, a gestão agrícola ou a administração cadastral [14].

Neste sentido, diversas abordagens têm sido propostas e desenvolvidas, baseadas em técnicas de DR e em métodos computacionais de processamento e análise de imagens. Entre estas, destacam-se cinco grandes categorias:

- 1. Métodos de deteção remota baseados em imagens de satélite e dados adquiridos por drones:** Estas técnicas recorrem a imagens de satélite de alta resolução espacial ou a dados captados por drones, aplicando processos de segmentação e deteção de contornos para definir limites visíveis. A principal vantagem reside na capacidade de cobrir grandes áreas com rapidez e a custos relativamente reduzidos quando comparados com levantamentos tradicionais, além de poderem ser automatizadas com recurso a técnicas de *machine learning* e *deep learning*. Contudo, dependem fortemente da presença de fronteiras físicas claras (estradas, muros, linhas de vegetação), sendo que a sua eficácia pode ser limitada em contextos em que os limites não sejam facilmente distinguíveis [15].
- 2. Segmentação semântica:** Recorre a redes neuronais avançadas, como *UNet* ou *DeepLabV3*, para identificar e extrair fronteiras, aprendendo simultaneamente padrões espectrais e espaciais presentes nas imagens. Este tipo de abordagem é especialmente eficaz em ambientes complexos, como áreas urbanas densas ou áreas agrícolas, apresentando elevado potencial de automatização e precisão [16]. Contudo, a necessidade de grandes volumes de dados anotados e de elevado esforço computacional constitui um entrave significativo à sua adoção generalizada [17].
- 3. Abordagens Semiautomáticas:** Combinam a extração automática de elementos com a intervenção do utilizador para refinar os resultados e resolver ambiguidades, permitindo acelerar os processos e reduzir custos, sem abdicar da supervisão humana [18].
- 4. Levantamentos Tradicionais:** Métodos baseados em levantamentos de campo, através do uso de estações totais e de recetores de sistemas globais de navegação por satélite (*Global Navigation Satellite System*, GNSS) continuam a ser a referência em termos de fiabilidade. Todavia, apresentam desvantagens óbvias: são processos morosos, dispendiosos e pouco adequados para monitorizações periódicas de larga escala.

5. **Algoritmos Avançados:** Algoritmos computacionais inovadores, como modelos de contornos ativos ou triangulações de Delaunay, têm demonstrado elevado rigor geométrico e eficiência na deteção de fronteiras. No entanto, a sua aplicabilidade é frequentemente limitada pela necessidade de parametrizações específicas e pela dificuldade em lidar com cenários complexo, onde diferentes objetos apresentam assinaturas espectrais muito próximas [19].

De forma transversal, todas estas abordagens têm beneficiado significativamente da integração de tecnologias emergentes, como a fusão de dados multissensoriais (imagens SAR e óticas) e o recurso a algoritmos de inteligência artificial cada vez mais avançados. Ainda assim, permanece o desafio de desenvolver soluções totalmente automáticas, capazes de generalizar entre diferentes regiões e condições ambientais, reduzindo simultaneamente a dependência de dados proprietários ou do trabalho humano. A necessidade de elevados recursos computacionais e de grandes volumes de dados constitui um dos principais entraves à generalização do uso de segmentação semântica e de outros algoritmos avançados, uma vez que estes conseguem atingir níveis de precisão bastante elevados.

Complementarmente aos métodos de deteção remota, a integração com técnicas de ML e DL tem vindo a revelar-se fundamental para melhorar a monitorização do LULC, permitindo explorar de forma mais eficaz o potencial dos dados de alta resolução. A Tabela 1 sintetiza alguns avanços recentes na aplicação de ML e DL para a monitorização do LULC, destacando os métodos utilizados, os *datasets* explorados e as principais conclusões dos estudos.

Tabela 1-Revisão da Literatura relativamente à aplicação de ML e DL ao LULC;

Referência	Objetivo	Dataset e Área de estudo	Conclusões
Swapan Talukdar et al. [20]	Comparação do desempenho de 6 algoritmos de ML: RF, SVM, ANN, <i>Fuzzy Adaptive Resonance Theory Mapping</i> (Fuzzy ARTMAP), <i>Segment Anything Model</i> (SAM), <i>Minimum Distance</i> (MD).	Landsat 8; Índia.	RF destacou-se como o melhor classificador; sublinha a necessidade de testar em diferentes contextos climáticos.
Sonia Sharma Banjade et al. [21]	Comparação do desempenho dos algoritmos SVM, <i>Spectral Information Divergence</i> (SID) e SAM em tarefas de classificação supervisionada LULC.	Imagem hiperespectral do G-LiHT Airborn Imager; Tennessee, USA.	SVM apresentou a maior exatidão (92,03%); resultados consistentes em áreas correspondentes a corpos de água e florestas; maior variação em áreas urbanas.

Yisa Ginath Yuh et al. [22]	Avaliação do desempenho dos algoritmos kNN, SVM, ANN e RF para monitorização do LULC.	Landsat 7 ETM+; OLI; região norte dos Camarões.	Todos os algoritmos superaram os 80% de exatidão; RF superou os 90%, obtendo a melhor performance.
Getachew Bayable et al. [23]	Avaliação do desempenho dos algoritmos <i>Classification and Regression Trees</i> (CART), RF e SVM na classificação do LULC.	Sentinel-1 (SAR); Sentinel-2 (MSI); Landsat 8; região norte da Etiópia.	RF e SVM obtiveram excelentes desempenhos (90% de exatidão).
Hafsa Ouchra et al. [24]	Avaliação do desempenho dos algoritmos de classificação supervisionada: SVM, RF, CART, MD, <i>Decision Tree</i> (DT), <i>Gradient Tree Boost</i> (GTB).	Landsat 8 OLI; Casablanca.	RF obteve o melhor desempenho, alcançando 95.42% de exatidão; SVM foi o pior com 83%.
Ganji Tejasree et al. [25]	Propor um modelo <i>Long Short-Term Memory</i> (LSTM) de DL para a classificação LULC usando imagens hiperespectrais.	Pavia University, Kennedy Space Center, Indian Pines.	Exatidão muito elevada (99,33%); eficaz para dados hiperespectrais; exige alto poder computacional.
Mengmeng Hao et al. [26]	Comparação entre modelos de DL (<i>FCN-8s</i> , <i>SegNet</i> , <i>UNet</i> , <i>Swin-Unet</i>).	Imagens UAV.	Swin-UNet obteve o melhor desempenho com uma exatidão global de 96,01%.
Shengyu Zhao et al. [27]	Revisão de métodos DL (CNNs, RNNs, Transformers) vs. técnicas de classificação tradicionais de ML (RF, SVM).	No Data.	RF e SVM perdem exatidão quando aplicados a novos dados ou áreas de estudo e exigem sempre a intervenção manual do utilizador; métodos de DL eliminam a necessidade de intervenção manual, aprendendo automaticamente as características relevantes e adaptam-se mais facilmente a novos <i>datasets</i> e regiões; DL exige maior volume e qualidade de dados e requer maiores recursos computacionais.

<p>Wuttichai Boonpook et al. [28]</p>	<p>Propor uma nova arquitetura de rede LoopNet; comparação com outros modelos de DL (SegNet, UNet, PSPNet, ResUNet, DeeplabV3 e UNet++).</p>	<p>Landsat 8.</p>	<p>Apresentou melhor desempenho do que métodos <i>pixel-based</i> e do que outras arquiteturas de modelos de DL convencionais.</p>
---------------------------------------	--	-------------------	--

As técnicas de ML afirmaram-se nas últimas décadas como ferramentas centrais na classificação do LULC, sobretudo quando aplicadas a imagens multiespetrais e hiperespetrais. A literatura demonstra que os algoritmos clássicos, com destaque para o RF e o SVM, têm produzido resultados consistentes e robustos em diferentes cenários do LULC. Estudos comparativos, como os de Swapan Talukdar et al. [20], Yisa Ginath Yuh et al. [22], Getachew Bayable et al. [23] e Hafsa Ouchra et al. [24], confirmam esta tendência, evidenciando o bom desempenho destes algoritmos em contextos diversos, desde áreas urbanas a regiões agrícolas e a ecossistemas naturais. De forma sistemática, o RF surge como o classificador mais estável e preciso, superando alternativas como SVM, kNN, ANN, SAM, CART e outros. O SVM, por sua vez, tem revelado especial eficácia em cenários com elevada diversidade de classes, onde a separação entre diferentes tipos de cobertura do solo é mais exigente [21].

É importante salientar que a maioria das abordagens de classificação supervisionada em ML é *pixel-based*, ou seja, cada pixel é atribuído a uma classe exclusivamente com base nos seus valores espectrais. Este princípio permite capturar relações diretas entre as assinaturas espectrais e as diferentes classes do LULC, mas apresenta limitações em contextos de elevada semelhança espectral entre diferentes classes (por exemplo, culturas agrícolas e matos). Para além disso, a transferibilidade dos modelos permanece um desafio recorrente, uma vez que classificadores treinados num determinado conjunto de dados ou área tendem a perder desempenho quando aplicados a novas regiões ou sobre outros sensores, isto é, a instrumentos de observação da Terra com diferentes resoluções espaciais, espectrais e radiométricas, o que obriga à construção de *datasets* de treino específicos para cada caso [27].

Nos últimos anos, o DL emergiu como uma alternativa promissora, introduzindo vantagens significativas graças à sua capacidade de extrair automaticamente características relevantes das imagens, eliminando em grande medida a necessidade de definição manual de atributos por parte do utilizador. Modelos baseados em redes convolucionais (*Convolutional Neural Networks*, CNNs), como *UNet* ou *SegNet*, continuam a ser amplamente utilizados e a alcançar desempenhos competitivos.

Contudo, observa-se uma clara tendência para a adoção de arquiteturas mais avançadas, incluindo modelos baseados em transformadores (*Swin-UNet*, *Vision Transformers*), arquiteturas como LSTM, e variantes de CNN, como o *LoopNet* [25,26,28]. Estas abordagens têm-se destacado particularmente no tratamento de dados de alta resolução, alcançando níveis de exatidão superiores e melhorando a capacidade de generalização, ou seja, de transferir modelos treinados num conjunto de dados e área de estudo para outros contextos geográficos ou de sensores, com menor perda de desempenho. No entanto, exigem grandes volumes de dados e elevados recursos computacionais, o que limita a sua aplicação prática em muitos contextos.

Em síntese, a literatura evidencia que os algoritmos de ML supervisionados, com destaque para o RF, oferecem uma solução equilibrada entre desempenho e exigência de recursos, razão pela qual continuam a ser amplamente aplicados. O *deep learning*, por sua vez, representa o estado da arte em termos de precisão e adaptabilidade, mas a sua adoção generalizada permanece condicionada por exigências computacionais e pela necessidade de bases de dados extensas e anotadas.

2.2. Classificação Supervisionada e *Random Forest*

No âmbito deste projeto optou-se pela utilização da classificação supervisionada com o algoritmo *Random Forest* (RF). Esta decisão assenta em múltiplos fatores. Em primeiro lugar, a literatura tem demonstrado de forma consistente que o RF apresenta elevada robustez e fiabilidade na classificação de imagens de satélite para fins de mapeamento do LULC, mostrando desempenhos competitivos em diferentes contextos territoriais e tipos de dados [20,22,23,24]. Outro aspeto determinante para a sua escolha prende-se com os requisitos computacionais moderados quando comparados com modelos mais avançados como os de *deep learning*. Enquanto estes últimos exigem grandes volumes de dados e elevada capacidade de processamento, o RF permite alcançar resultados sólidos com menores *datasets* de treino e recursos computacionais, o que se revela particularmente vantajoso numa fase inicial como a deste projeto.

Neste contexto, a classificação supervisionada é um processo em que o algoritmo de classificação aprende a distinguir diferentes classes LULC a partir de exemplos previamente fornecidos pelo utilizador. Para tal, é necessário criar *datasets* de treino, constituídos por polígonos anotados manualmente sobre as imagens de satélite, que representam as classes de interesse. Cada polígono funciona como uma amostra de referência, a partir da qual o algoritmo extrai características, como os valores espectrais dos píxeis, permitindo, desta forma, construir um modelo de classificação. O seu

desempenho depende significativamente da qualidade e representatividade do conjunto de treino, sendo crucial garantir a diversidade espacial e espectral dos polígonos anotados.

O RF, proposto por Breiman (2001) [29], é um método de *ensemble learning* que combina múltiplas árvores de decisão para melhorar a exatidão da classificação e a capacidade de generalização. Cada árvore é construída a partir de subconjuntos aleatórios de observações e, em cada nó de decisão, a variável de divisão é selecionada de um subconjunto aleatório de atributos. Esta dupla aleatoriedade diminui a correlação entre as árvores, reduzindo o risco de *overfitting* e aumentando a robustez do modelo face a ruído. A previsão final resulta do voto maioritário entre todas as árvores, o que estabiliza o desempenho e aumenta a fiabilidade.

Entre as suas principais vantagens, destaca-se a capacidade de lidar com dados de elevada dimensionalidade sem necessidade de pressupostos estatísticos rígidos sobre a distribuição das variáveis. O RF é também menos dependente de um processo exaustivo de seleção manual de atributos, uma vez que a diversidade de árvores tende a explorar diferentes combinações de atributos relevantes. Adicionalmente, incorpora mecanismos internos de avaliação de desempenho que permitem aferir a qualidade do modelo sem recorrer obrigatoriamente a um conjunto de validação externo. O algoritmo disponibiliza ainda métricas de importância de variáveis, úteis para interpretar quais os atributos mais determinantes na separação entre classes.

Apesar destas vantagens, importa referir algumas limitações práticas. O RF aplicado de forma direta em imagens de satélite opera tipicamente de forma *pixel-based*, baseando-se sobretudo em valores espectrais individuais. Isto pode dificultar a discriminação de classes com assinaturas espectrais semelhantes. Relativamente à desvantagem da menor adaptabilidade dos modelos face a novos sensores e áreas de estudo, mencionada anteriormente, neste caso não assume particular relevância dado que as aplicações previstas do modelo se restringem à Península Ibérica e o treino foi realizado com datasets da mesma região.

Em síntese, o RF apresenta-se como uma solução equilibrada para o desenvolvimento deste projeto, combinando precisão, eficiência computacional e facilidade de implementação, ao mesmo tempo que oferece ferramentas interpretativas valiosas. Estas características justificam a sua ampla adoção em estudos de LULC e a sua seleção como ponto de partida para o presente trabalho.

3. Metodologia

O presente capítulo encontra-se organizado em duas secções distintas, correspondentes aos dois projetos desenvolvidos ao longo do estágio: o projeto de monitorização do LULC e o projeto de preparação de dados para identificação automática de vinhas. Em cada secção é descrita a metodologia seguida, com especial foco nas tarefas desenvolvidas individualmente dentro da equipa, de forma a contribuir para o alcance dos objetivos da mesma.

No caso do projeto do LULC, o trabalho realizado visou contribuir para o desenvolvimento de um modelo de inteligência artificial capaz de detetar alterações no uso e ocupação do solo a partir de imagens de satélite, envolvendo tarefas como anotações manuais, pré-processamento de dados geoespaciais, entre outras. Já o projeto das vinhas, centrou-se na análise e no pré-processamento de um *dataset* com o registo histórico das vinhas do concelho de Arcos de Valdevez, criando as bases necessárias para que, numa fase posterior, seja possível treinar um modelo de IA capaz de identificar vinhas de forma automática.

3.1. Metodologia-LULC

3.1.1. Área de Estudo

Considerando que, numa fase inicial, as aplicações previstas para o modelo se encontram circunscritas ao contexto da Península Ibérica, optou-se por delimitar a área de estudo a esta região. Esta opção garante uma maior consistência entre o conjunto de dados de treino e as condições reais de aplicação do modelo, assegurando a sua relevância para o contexto em análise.

De modo a considerar a diversidade ambiental e climática da Península Ibérica, a seleção das áreas de treino foi baseada na classificação climática de Köppen-Geiger [30]. Este sistema, amplamente utilizado a nível global, baseia-se em parâmetros como a temperatura média e a precipitação acumulada mensais e as suas respetivas distribuições sazonais, permitindo distinguir zonas climáticas de forma objetiva e comparável. Esta abordagem classifica os climas em cinco categorias principais (Tropical, Árido, Temperado, Continental e Polar) que se dividem em diversos subtipos determinados por padrões específicos de temperatura e precipitação. A sua adoção neste projeto justifica-se pela capacidade de representar de forma consistente a heterogeneidade climática do território, fator determinante para assegurar que os

datasets de treino capturam uma diversidade adequada de condições ambientais e espectrais.

Desta forma, foram seleccionadas seis áreas de treino distintas, representativas das diferentes zonas climáticas da Península Ibérica, conforme apresentado na Figura 4.

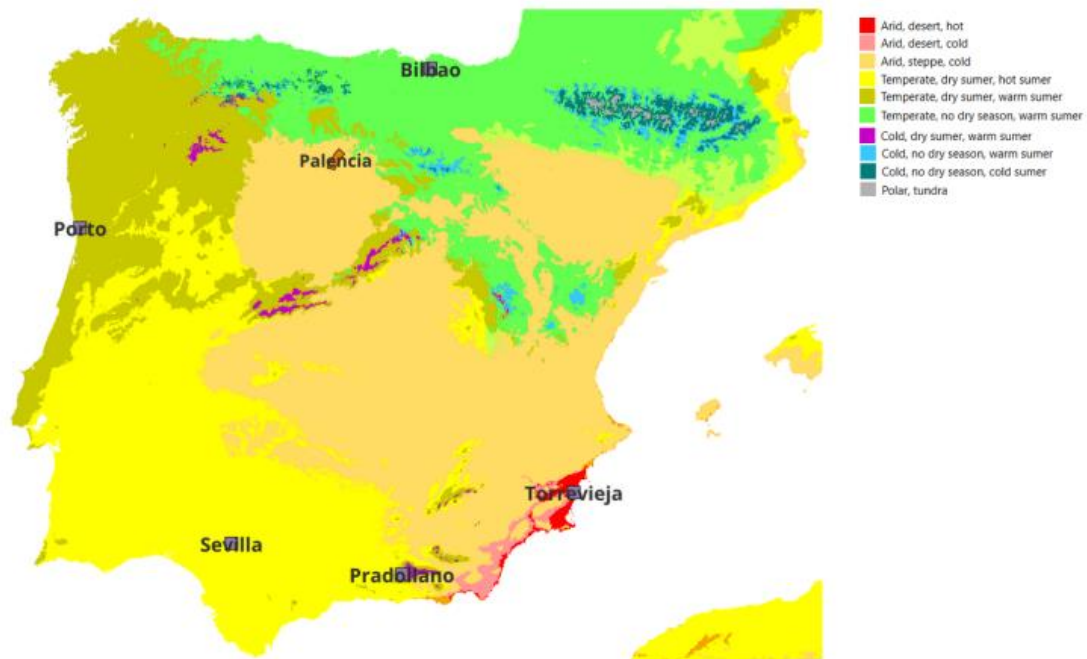


Figura 4-Mapa Climático da Península Ibérica baseado na classificação de Köppen-Geiger [30];

As áreas de treino seleccionadas foram as seguintes:

- **Porto** – clima temperado com verão seco e moderado;
- **Sevilla** – clima temperado com verão seco e quente;
- **Bilbao** – clima temperado sem estação seca e com verão moderado;
- **Torre Vieja** – clima árido com temperaturas elevadas;
- **Pradollano** – clima continental com verão seco e quente;
- **Palência** – clima árido e frio;

Como foi referido anteriormente, a escolha destas localizações teve como principal objetivo a criação de *datasets* heterogéneos, capazes de representar a diversidade climática e ambiental do território ibérico. Esta heterogeneidade é essencial para o processo de treino do modelo, uma vez que confere maior robustez e fiabilidade à classificação, permitindo que o algoritmo aprenda a lidar com diferentes condições espectrais associadas a distintos contextos geográficos. Assim, é possível aumentar o potencial de generalização do modelo e a sua aplicabilidade prática em diferentes cenários dentro da região de estudo.

3.1.2. Dados Utilizados

No contexto deste projeto, foram utilizadas imagens multiespectrais do Sentinel-2 e imagens SAR (*Synthetic Aperture Radar*) do Sentinel-1, correspondentes às áreas de treino previamente definidas. As missões Sentinel-1 e Sentinel-2 constituem pilares do programa *Copernicus*, promovido pela Agência Espacial Europeia (*European Space Agency*, ESA), cujo objetivo é disponibilizar informação sistemática, gratuita e de elevada qualidade para aplicações de monitorização ambiental, ordenamento do território e gestão de recursos [31].

A missão Sentinel-2 consiste numa constelação de satélites óticos de alta resolução dedicados à aquisição de imagens multiespectrais. Esta constelação é constituída por dois satélites gémeos operacionais: o Sentinel-2B, lançado em 2017, e o Sentinel-2C, lançado em 2024, que assumiu as operações do Sentinel-2A a partir de janeiro de 2025. Ambos os satélites estão equipados com o sensor MSI (*Multispectral Instrument*), que permite a aquisição de imagens multiespectrais em 13 bandas espectrais distintas, com resoluções espaciais de 10, 20 e 60 metros. A largura da faixa de observação é de 290 km, garantindo uma ampla cobertura a cada passagem. Os dois satélites seguem uma órbita heliosíncrona, desfasados em 180°, o que assegura uma elevada resolução temporal (aproximadamente cinco dias). A elevada resolução espacial e temporal, aliada à sua capacidade espectral, confere ao Sentinel-2 a competência necessária para a monitorização do LULC.

A missão Sentinel-1 consiste numa constelação de dois satélites (Sentinel-1A e Sentinel-1C) equipados com o instrumento C-SAR (*C-band Synthetic Aperture Radar*), que permite adquirir imagens em quaisquer condições meteorológicas, de dia ou de noite. Estes satélites operam numa órbita quase polar heliosíncrona, desfasados de 180°, o que lhes permite obter uma resolução temporal de, aproximadamente, 6 dias. O sistema oferece diversos modos de aquisição, com resoluções espaciais que podem alcançar 5 metros, larguras de faixa até 400 km e capacidade de dupla polarização. Esta flexibilidade torna o Sentinel-1 particularmente relevante para estudos do LULC, ao fornecer dados que registam propriedades morfológicas e texturais complementares às imagens óticas do Sentinel-2, melhorando significativamente a discriminação de algumas classes com valores espectrais semelhantes.

Complementarmente às imagens de satélite, recorreu-se a cartografias oficiais da ocupação do solo para apoio à construção dos *datasets* de treino supervisionados. Para as áreas de treino localizadas em Espanha, foi utilizada a SIOSE (*Sistema de*

Información sobre Ocupación del Suelo de España), disponibilizada pelo *Centro Nacional de Información Geográfica* (CNIG), baseada em fotointerpretação de imagens aéreas. Para a área de treino portuguesa, utilizou-se a COS (Carta de Ocupação do Solo), produzida pela Direção-Geral do Território (DGT), que segue uma metodologia semelhante. Estes dados constituíram a base inicial para a criação dos polígonos de treino.

3.1.3. Pré-processamento das Imagens de Satélite

O pré-processamento das imagens de satélite constituiu uma etapa fundamental para a construção dos *datasets* utilizados na classificação do LULC. Este procedimento possibilitou a integração consistente dos dados óticos do Sentinel-2 com os dados SAR do Sentinel-1, originando mosaicos mensais para todas as áreas de treino, no período compreendido entre janeiro de 2021 e outubro de 2024.

3.1.3.1. Aquisição das Imagens

A aquisição das imagens foi realizada através do *Copernicus Open Access Hub* [32], utilizando acessos automatizados à sua API integrados em *workflows* em *Python*. Este procedimento assegurou o *download* dos dados brutos para todas as áreas de interesse e para todo o período temporal em análise. Para cada missão (Sentinel-1 e Sentinel-2), foi aplicada uma seleção prévia baseada em parâmetros como área geográfica, intervalo temporal e a qualidade dos produtos (cobertura das nuvens), de forma a garantir que apenas os dados relevantes eram transferidos para o *pipeline* de processamento subsequente.

3.1.3.2. Produção dos Mosaicos Mensais do Sentinel-2

As imagens óticas do Sentinel-2 foram processadas para gerar mosaicos mensais para cada uma das áreas de treino. Durante a fase de aquisição das imagens foi aplicado um filtro de nuvens de 60%, assegurando que imagens com elevada cobertura atmosférica não fossem utilizadas. Cada composição foi calculada tendo em conta os valores da mediana, de modo a assegurar uma maior robustez face a *outliers* e variações abruptas.

No total, cada mosaico mensal resultou num *raster* único com 39 bandas:

- **12 bandas espectrais** originais do Sentinel-2;
- **27 índices espectrais** derivados, calculados a partir das bandas, incluindo índices de vegetação, índices de humidade, entre outros.

Para evitar inconsistências no cálculo dos índices espectrais e, conseqüentemente, erros de classificação, foi estabelecida a regra de que, caso alguma das bandas

principais (B2, B3, B4, B5, B6, B7, B8, B8A, B11 ou B12) apresentasse valores *NoData*, todas as restantes bandas seriam igualmente assinaladas como *NoData* no mesmo pixel. Importa salientar que as bandas B1 e B9 não foram incluídas neste processo, por se tratarem de bandas auxiliares destinadas principalmente a correções atmosféricas, pelo que não desempenham um papel relevante no cálculo dos índices espectrais utilizados.

3.1.3.3. Produção dos Mosaicos Mensais do Sentinel-1

Os dados SAR do Sentinel-1 foram alvo de um *pipeline* de pré-processamento mais complexo, dada a sua natureza. Este processo foi automatizado com recurso à biblioteca PyroSAR [33], que inclui um conjunto de operações fundamentais:

- **Aplicação dos ficheiros de órbita:** utilização dos dados precisos das órbitas dos satélites para corrigir e georreferenciar (WGS84, EPSG:32629) com exatidão as imagens adquiridas;
- **Remoção de ruído térmico:** eliminação de interferências causadas pelo ruído gerado pelo próprio sistema do sensor;
- **Calibração radiométrica:** conversão dos valores dos píxeis das imagens em unidades físicas, permitindo uma comparação quantitativa entre diferentes imagens e áreas;
- **Filtragem do *speckle*:** redução do ruído inerente aos sensores radar, que confere um aspeto granular às imagens e dificulta a sua interpretação e análise;
- **Correção de terreno:** ajuste das distorções geométricas causadas pelas variações de relevo no terreno, garantindo que as imagens representem corretamente a posição dos objetos na superfície da Terra;
- **Conversão dos valores de escala linear para decibéis (dB):** transformação dos valores do sinal de uma escala linear para uma escala logarítmica (decibéis), que é mais apropriada para a análise e visualização dos dados, uma vez que reflete melhor as variações no sinal.

Assim, foram gerados os mosaicos mensais correspondentes às áreas de treino e ao intervalo temporal em análise, incluindo as bandas de polarização VV e VH, a razão VV/VH e o índice RVI (*Radar Vegetation Index*).

3.1.3.4. Fusão dos Dados Óticos com SAR

Para o treino do modelo de classificação, foi imprescindível a integração dos dados provenientes das missões Sentinel-1 e Sentinel-2. Nesse sentido, foi desenvolvido um script em *Python* especificamente orientado para o alinhamento e junção dos dados óticos e de SAR. Este procedimento assegurou a correspondência temporal precisa

entre todos os mosaicos obtidos, a uniformização do sistema de coordenadas de referência e a reamostragem necessária para garantir a compatibilidade da resolução e da extensão espacial dos diferentes conjuntos de dados.

O resultado consistiu em mosaicos mensais compostos por quarenta e três bandas, todas com resolução espacial de 10 metros, incluindo as doze bandas multiespectrais do Sentinel-2, vinte e sete índices espectrais e quatro bandas SAR do Sentinel-1. No total, foram produzidos quarenta e seis mosaicos mensais por área de treino, abrangendo o período de observação compreendido entre janeiro de 2021 e outubro de 2024. Estes mosaicos constituíram, em conjunto com os ficheiros vetoriais dos polígonos que serão abordados na secção seguinte, a base utilizada para o treino do modelo de classificação.

A Figura 5 ilustra esquematicamente o processo de integração dos dados, apresentando, em primeiro lugar, o mosaico obtido a partir das imagens Sentinel-2, seguido pelo mosaico correspondente ao Sentinel-1 e, por fim, o resultado da fusão de ambos num único produto. Este procedimento requer, de forma estritamente obrigatória, que os mosaicos a integrar se encontrem no mesmo sistema de referência geodésico (WGS84), projetados no mesmo sistema de coordenadas (EPSG:32629) e possuam a mesma extensão espacial, de modo a garantir a correta sobreposição e a consistência dos dados. Complementarmente, a Tabela 2 enumera todas as bandas e índices que compõem o mosaico final, constituindo a base de dados multiespectral e SAR utilizada para o treino do modelo.

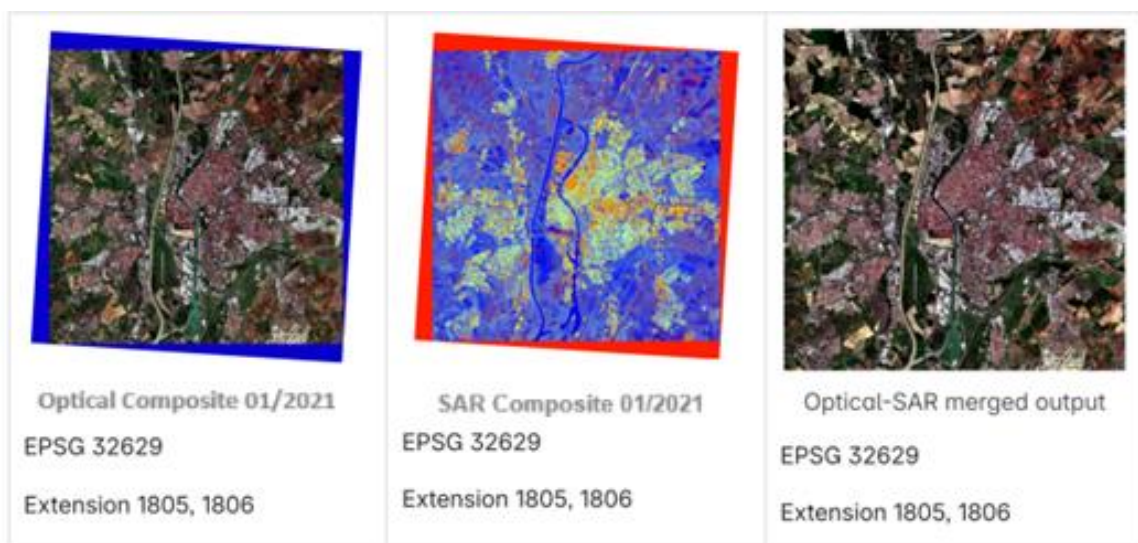


Figura 5-Exemplo dos mosaicos Ótico, SAR e Ótico-SAR de janeiro de 2021;

Tabela 2-Lista das bandas e índices que compõem os mosaicos;

Band Names	Band Labels
B01	Coastal aerosol (Sen-2)
B02	Blue band (Sen-2)
B03	Green band (Sen-2)
B04	Red band (Sen-2)
B05	Red Edge 1 (Sen-2)
B06	Red Edge 2 (Sen-2)
B07	Red Edge 3 (Sen-2)
B08	Near Infrared (Sen-2)
B8A	Red Edge 4 (Sen-2)
B09	Water Vapour (Sen-2)
B11	Shortwave Infrared 1 (Sen-2)
B12	Shortwave Infrared 2 (Sen-2)
ARVI	Atmospherically Resistant Vegetation Index [34]
AWEINSH	Automated Water Extraction Index (Shadow) [35]
AWEISH	Automated Water Extraction Index (High) [35]
BAI	Burn Area Index [36]
BAIS2	Burn Area Index S2 [37]
BUI	Built-Up Index [38]
CIRE	Chlorophyll Index Red Edge [39]
EVI	Enhanced Vegetation Index [40]
GCI	Green Chlorophyll Index [41]
GEMI	Global Environment Monitoring Index [42]
GNDVI	Green Normalized Difference Vegetation Index [43]
LAI	Leaf Area Index [44]
MSAVI2	Modified SoilAdjusted Vegetation Index 2 [45]
MTVI2	Modified Triangular Vegetation Index 2 [46]
NBR	Normalized Burn Ratio [47]
NDBI	Normalized Difference Bare Soil Index [48]
NDMI	Normalized Difference Moisture Index [49]
NDRE	Normalized Difference Red Edge [50]
NDVI	Normalized Difference Vegetation Index [51]
NDWI	Normalized Difference Water Index [52]
OSAVI	Optimized SoilAdjusted Vegetation Index [53]
RI	Redness Index [54]
RTVI	Ratio Vegetation Index [55]
SAVI	Soil-Adjusted Vegetation Index [56]
SIPI	Structure Insensitive Pigment Index [57]
SRRE	Simple Ratio Red Edge [58]
VARI	Visible Atmospherically Resistant Index [59]
VV	Single co-polarization, vertical transmit/vertical receive (Sen-1)
VH	Single co-polarization vertical transmit/horizontal receive (Sen-1)
VV_VH_ratio	Ratio (division) of VV with VH (Sen-1)
RVI	Radar Vegetation Index [60]

3.1.4. Anotações Manuais

Como referido anteriormente, os dados de referência utilizados para a definição das classes de LULC corresponderam, no caso das áreas de treino espanholas, à cartografia da SIOSE [61] e, no caso do Porto, à COS [62]. Os ficheiros vetoriais destas cartografias foram descarregados a partir das respetivas plataformas institucionais e, posteriormente, recortados para as áreas de interesse através da ferramenta *Clip* do QGIS [63]. Cada uma destas áreas corresponde a uma extensão de 17 km × 17 km, coincidente com a dimensão dos mosaicos gerados anteriormente. Desta operação resultaram, assim, mapas LULC de cada região de interesse, que serviram de base para a geração inicial dos polígonos de treino.

De seguida, a partir destes mapas recortados, procedeu-se à geração automática de polígonos representativos das diferentes classes. Este procedimento, automatizado através de *scripts* (implementados em QGIS/Python), incluiu, entre outras operações, a desagregação de polígonos de grande dimensão, a definição de limites mínimos (1000 m²) e máximos (500000 m²) para as áreas dos mesmos e a aplicação de um *buffer* (1000 m) destinado a evitar a seleção de polígonos adjacentes da mesma classe, ou seja, qualquer polígono cuja sua geometria interseste o buffer de outro polígono da mesma classe previamente selecionado. Estas regras foram impostas com o objetivo de assegurar uma distribuição espacial mais equilibrada e reduzir potenciais enviesamentos na amostragem. Adicionalmente, foi garantido um número equilibrado de polígonos por classe, de modo a evitar desequilíbrios entre categorias. A Figura 6 ilustra o resultado deste processo para a área de Sevilha, onde foram gerados 1079 polígonos classificados de acordo com a SIOSE.

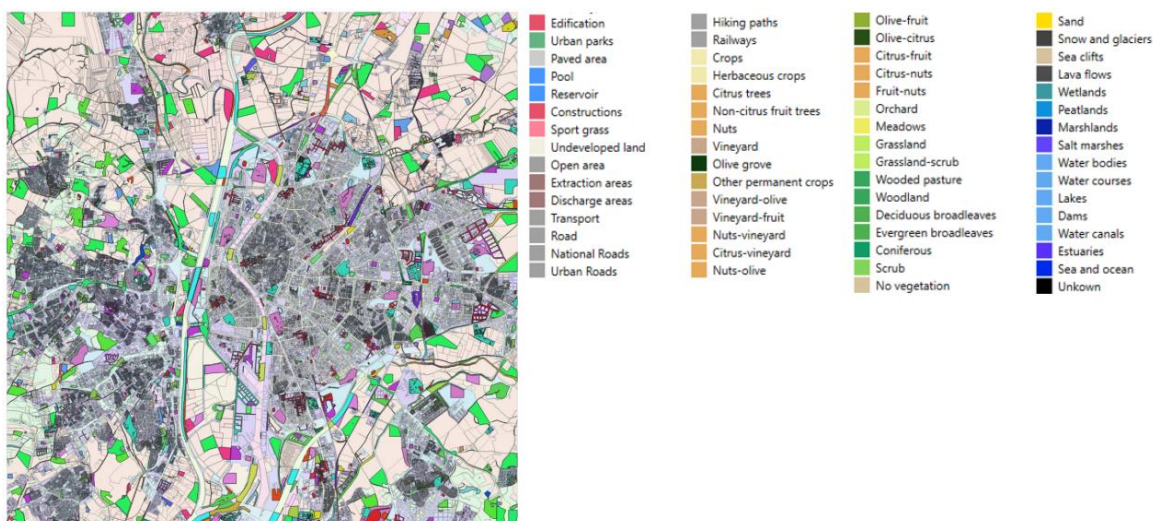


Figura 6-Polígonos gerados sobre Sevilha;

Embora os dados da SIOSE e da COS utilizados se tenham revelado fundamentais para a elaboração do *dataset* inicial, a sua data de produção (2017 e 2018, respetivamente) implicou a necessidade de uma etapa adicional de refinamento, de modo a atualizar os polígonos à realidade atual. Esse processo consistiu na verificação manual e individual de cada polígono, conduzindo, em diversos casos, à sua reclassificação para a categoria correspondente no presente ou, quando a classe não era clara ou apresentava inconsistências significativas, à sua eliminação.

Um dos aspetos mais relevantes nesta fase foi garantir que cada polígono incluísse apenas píxeis homogéneos, isto é, correspondentes à mesma classe. Este cuidado visou evitar a introdução de ruído no processo de treino do modelo. Por exemplo, em áreas urbanas, alguns polígonos inicialmente abrangiam simultaneamente edifícios e zonas verdes adjacentes, mas após o refinamento passaram a incluir exclusivamente a parte edificada. De forma semelhante, registaram-se diversas inconsistências recorrentes: na SIOSE, extensas áreas pavimentadas estavam por vezes classificadas como pastagens; na COS, polígonos atribuídos a estradas ou edifícios continham vegetação dispersa. Além disso, a SIOSE não contemplava uma classe correspondente a áreas industriais, agrupando-as no tecido urbano contínuo. No entanto, no âmbito deste projeto, considerou-se fundamental distinguir explicitamente as áreas industriais, pelo que foi criada uma classe adicional dedicada a esse uso.

Este processo de refinamento foi realizado no QGIS, recorrendo às ferramentas da *Advanced Digitalization Toolbar*, e resultou em *datasets* ajustados à realidade atual, classificados segundo a nomenclatura base da SIOSE ou da COS.

Apesar da sua utilidade, ambos os sistemas apresentam um número elevado de classes, 62 no caso da SIOSE e 83 no caso da COS, o que introduz uma complexidade excessiva para o treino de modelos de inteligência artificial, aumentando o risco de confusões entre categorias ou da presença de classes pouco representadas. Para mitigar esse problema, foi implementado um processo de reclassificação, através do qual as múltiplas classes originais foram agregadas em categorias mais gerais e adaptadas aos objetivos do projeto.

O novo sistema, ilustrado na Tabela 3, organiza-se em três níveis hierárquicos:

- **Nível 1:** agrupa categorias muito abrangentes;
- **Nível 2:** reflete classes mais ajustadas às necessidades práticas da monitorização do LULC;

- **Nível 3:** mantém maior detalhe na diferenciação das classes.

Tabela 3-Sistema de Classificação LULC adotado;

LULC Classification System		
Level-1	Level-2	Level-3
1. Artificial Areas	1.1. Urban	1.1.1. Urban
		1.1.2. Industrial
	1.2. Roads	1.2.1. Roads
	1.3. Railways	1.3.1. Railways
2. Vegetated Areas	2.1. Agriculture	2.1.1. Growing Crops
		2.1.2. Dry crops
	2.2. Forest	2.2.1. Forest
	2.3. Shrubland	2.3.1. Shrubland
3. Bare Areas	3.1. Bare Soil	3.1.1 Bare Soil
	3.2. Sand	3.2.1. Sand
	3.3. Others	3.3.1 Volcanic
		3.3.2. Burnt
		3.3.3. Snow
		3.3.4. Rock
4. Water Bodies	4.1. Water	4.1.1. River
		4.1.2. Sea
		4.1.3. Lake

Relativamente ao nível 3, a classe *Urban* corresponde a edifícios residenciais e tecido urbano consolidado, enquanto a categoria *Industrial* abrange edifícios industriais e unidades fabris, frequentemente caracterizados por elevados níveis de refletância em tons brancos. A classe *Roads* inclui estradas, autoestradas e parques de estacionamento, enquanto a categoria *Railways* diz respeito às infraestruturas ferroviárias. Relativamente às áreas agrícolas, distinguem-se dois grupos: *Growing Crops*, correspondentes a campos de cultivo ativos e vegetação agrícola em

crescimento, com tonalidades verdes, e *Dry Crops*, que representam culturas em estado seco ou após a colheita, evidenciando tons acastanhados. A categoria *Forest* integra as áreas florestais densas, ao passo que *Shrubland* corresponde a matos e vegetação arbustiva mais esparsa. A classe *Bare Soil* agrega superfícies de solo exposto ou áreas de construção, em tons amarelados ou castanhos-claros, enquanto *Sand* identifica zonas litorais de praia. A classe *Rocks* refere-se a afloramentos rochosos. No que respeita às massas de água, foram discriminadas em três subcategorias distintas: *Rivers*, *Lakes* e *Ocean*, de modo a refletir a diversidade hidrológica presente no território. Importa salientar que as categorias *Volcanic*, *Snow* e *Burnt* não foram incluídas no sistema aplicado neste projeto até ao momento, pelo que não existem polígonos correspondentes a estas três classes.

Apesar do nível 3 assegurar uma caracterização detalhada e diferenciada das classes, este grau de especificidade poderá revelar-se excessivo para a maioria das aplicações. Assim, prevê-se a implementação de etapas de pós-processamento que permitirão a agregação das classes do nível 3 em categorias mais amplas do nível 2, alinhadas com as necessidades típicas de potenciais utilizadores.

Após a definição do novo sistema, os polígonos foram reclassificados no QGIS, dando origem aos *datasets* de treino final, estruturados de acordo com as especificações necessárias para um treino mais robusto e eficiente do modelo.

Posteriormente, foi realizada uma verificação temporal dos polígonos para os 46 mosaicos mensais gerados entre janeiro de 2021 e outubro de 2024, em cada uma das áreas de treino. Este procedimento consistiu numa análise manual e individual de todos os polígonos, com o objetivo de identificar possíveis alterações ocorridas ao longo do período de estudo. Sempre que se verificavam mudanças relevantes, a classe atribuída era ajustada de modo a refletir a nova realidade observada. A necessidade desta etapa justifica-se, em grande medida, pela variabilidade sazonal característica dos mapas LULC. Os campos agrícolas constituem um exemplo desse fator: em determinados meses apresentam-se cobertos de vegetação ativa, enquadrando-se na classe *Growing Crops*, enquanto em outros períodos do ano se encontram secos, passando a corresponder à classe *Dry Crops*. A Figura 7 ilustra este fenómeno, mostrando um campo inicialmente classificado como *Growing Crops* que, devido às alterações sazonais, foi posteriormente reclassificado como *Dry Crops*.



Figura 7- Variabilidade sazonal de um campo agrícola;

Para além das variações sazonais, a análise de um período de quatro anos permitiu ainda identificar transformações estruturais significativas no território, resultantes sobretudo da ação humana. Em vários casos, polígonos anteriormente classificados como *Shrubland* foram reclassificados como *Urban*, devido à construção de novos edifícios. Foram igualmente identificadas alterações associadas a processos de desflorestação.

Este processo foi aplicado sistematicamente a todas as áreas de treino, originando um extenso conjunto de dados para o treino do modelo. No total, foram produzidos 46 ficheiros vetoriais de polígonos para cada área, correspondentes aos mosaicos mensais do período analisado. A Figura 8 apresenta um exemplo desta integração, exibindo os polígonos classificados sobre o mosaico de janeiro de 2021 da área de Sevilha.

January 2021 Sevilha Composite

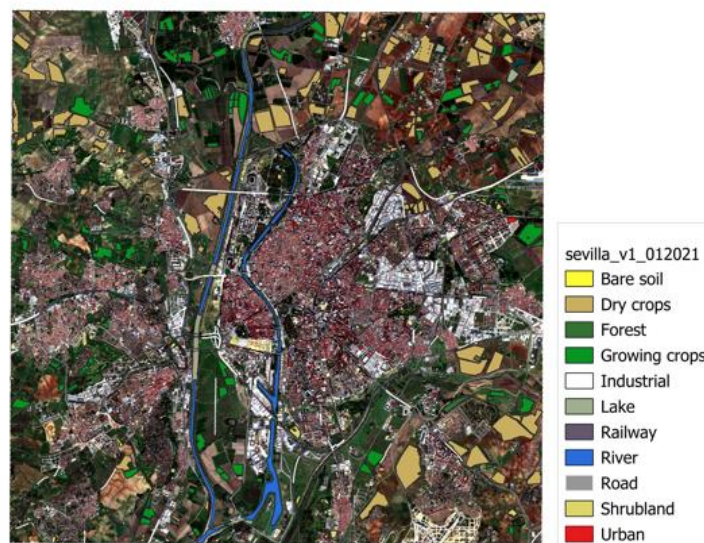


Figura 8-Polígonos LULC 01/2021 Sevilha;

Em suma, todo este processo de anotações manuais permitiu a construção de um conjunto de dados completo, robusto e representativo de todas as épocas do ano, constituindo assim uma base de treino de elevada qualidade, capaz de potenciar a obtenção de resultados consistentes e precisos no modelo desenvolvido.

3.1.5. Metodologia de Treino e de Avaliação do Modelo

O treino e a posterior avaliação do modelo foram realizados recorrendo a um repositório de código desenvolvido pela equipa, que constitui uma *framework* completa para a classificação do LULC de imagens de satélite. Esta estrutura integra todas as etapas necessárias ao desenvolvimento de modelos robustos, desde a preparação e gestão dos dados, passando pelo treino e afinação de classificadores, até à execução de inferências e deteção de alterações temporais.

O processo inicia-se com o módulo *DataLoader*, responsável por organizar e preparar os dados de entrada. Este módulo permite selecionar as bandas espectrais dos mosaicos mensais a utilizar no treino, excluir aquelas que não são relevantes e assegurar o carregamento eficiente tanto dos *rasters* (mosaicos) como dos ficheiros vetoriais correspondentes. Além disso, garante a integração consistente com vários modelos de ML.

O módulo *ModelTrainer* permite selecionar o classificador (RF, neste caso) e definir se o treino será executado utilizando a *Central Processing Unit* (CPU) ou a *Graphics Processing Unit* (GPU), sendo que a utilização desta última proporciona um processamento mais rápido e adequado de grandes volumes de dados, enquanto a CPU é mais flexível, mas menos eficiente quando sujeita a cargas computacionais intensivas. Além disso, este módulo produz diversas métricas fundamentais para a avaliação do desempenho do modelo, nomeadamente a exatidão, a precisão, o *recall* e o *F1-score*. A precisão expressa a proporção de amostras corretamente identificadas como positivas em relação ao total de amostras classificadas como positivas, enquanto o *recall* indica a proporção de amostras positivas corretamente identificadas em relação ao total de amostras positivas existentes. O *F1-score* corresponde à média entre a precisão e o *recall*, constituindo assim uma métrica robusta para avaliar o equilíbrio entre ambos. Para além destas métricas, o módulo produz também matrizes de confusão e a *feature importance*, que quantifica o contributo relativo de cada banda espectral para a aprendizagem do modelo.

O módulo *ModelFinetune* visa a otimizar o processo, permitindo afinar os parâmetros do classificador através de estratégias como *grid search* e validação cruzada. O *grid search*

consiste em testar de forma sistemática diferentes combinações de hiperparâmetros até encontrar a configuração mais adequada dos *datasets*, enquanto a validação cruzada assegura que os resultados obtidos não são específicos de uma única porção dos dados, garantindo maior generalização. Este módulo possibilita ainda a seleção das características mais relevantes (*top-k feature selection*), eliminando possíveis redundâncias e melhorando a eficiência do modelo.

A interpretação dos resultados é apoiada pelo módulo *Visualizers*, que disponibiliza um conjunto de ferramentas gráficas para compreender o comportamento do modelo e a qualidade do treino. Entre estas funcionalidades, destacam-se a visualização da distribuição de classes no *dataset*, gráficos de desempenho e representações da importância das variáveis, o que facilita a análise crítica do modelo.

Uma vez concluído o treino, o modelo é aplicado na fase de execução de inferências, suportada pelo módulo *InferenceEngine*. Esta etapa permite aplicar o classificador previamente treinado a novas áreas geográficas, produzindo previsões do LULC sob a forma de *rasters* classificados. O processo integra igualmente uma componente de pós-processamento, na qual podem ser realizados procedimentos de remapeamento no QGIS ou a fusão de classes correlacionadas, de modo a reduzir o ruído e aumentar a consistência dos mapas gerados. Um exemplo desta etapa consiste na agregação das classes *Growing Crops* e *Dry Crops* numa única categoria de agricultura, correspondente ao nível 2 do sistema de classificação desenvolvido. Os resultados obtidos são organizados em diretórios estruturados e, subsequentemente, avaliados no QGIS através da comparação visual com as imagens de satélite de referência.

Paralelamente, o repositório disponibiliza o módulo *ChangeDetection* especificamente concebido para a análise temporal das alterações no LULC. Este módulo oferece diferentes metodologias, incluindo comparações *pixel-to-pixel* entre imagens de diferentes datas, abordagens baseadas em *rasters* de referência e análises de séries temporais. Entre as funcionalidades mais relevantes destaca-se a geração de máscaras de mudança e de severidade, que consistem em *rasters* que são capazes de mapear as alterações ocorridas, bem como quantificar a sua intensidade. Este tipo de análise revela-se particularmente útil para compreender a intensidade e a distribuição espacial das transformações verificadas ao longo do tempo.

3.1.6. Treino do Modelo

Após a conclusão do processo de anotações manuais, encontravam-se reunidos todos os dados necessários para o treino do modelo, nomeadamente os mosaicos mensais com 43 bandas espectrais, referentes ao período compreendido entre janeiro de 2021 e outubro de 2024, bem como os respetivos ficheiros vetoriais com os polígonos da LULC.

O processo de treino iniciou-se com o carregamento dos dados através do módulo *DataLoader*. Devido a limitações computacionais, optou-se por utilizar, nesta fase inicial, apenas os *datasets* de Palência e do Porto. Além disso, foram consideradas apenas 41 bandas, tendo sido excluídas a B1 e a B9, correspondentes a correções atmosféricas, uma vez que o objetivo era focar exclusivamente nos valores de refletância espectral à superfície.

Com o *dataset* devidamente preparado, procedeu-se ao treino do modelo com recurso ao *ModelTrainer*. Para este efeito, foi selecionado o algoritmo *Random Forest*, dada a sua robustez e capacidade de lidar com dados de elevada dimensionalidade, como foi evidenciado anteriormente. O treino foi realizado com recurso a CPU e GPU, confirmando-se que a execução em GPU oferece vantagens significativas na gestão de grandes volumes de dados e na redução do tempo de treino. Os polígonos de referência foram divididos em três subconjuntos: 70% para treino, 20% para teste e 10% para validação, garantindo assim uma avaliação equilibrada e fiável da performance do modelo. Esta fase de treino gerou um conjunto de métricas de avaliação que caracterizam o desempenho do classificador, as quais serão apresentadas e discutidas na secção de resultados.

Com o modelo treinado, foram então executadas inferências, através das quais se produziram mapas LULC para áreas de interesse específicas. Este trabalho foi realizado em resposta às necessidades de um cliente, incidindo em zonas adjacentes a linhas ferroviárias, para as quais foram definidos *buffers* de 150 metros ao longo do traçado. A análise centrou-se em quatro áreas da Península Ibérica (Corunha, Santander, Canfranc e Maravillas), representadas no mapa climático da Figura 9, bem como numa área mais ampla localizada em Zamora. Para estas regiões foi necessário gerar mosaicos compatíveis com o processo de treino, assegurando, desta forma, a mesma resolução espacial e a mesma composição espectral. Foram realizadas inferências mensais para todo o ano de 2024, cujos resultados foram posteriormente avaliados em ambiente QGIS, comparando-se a classificação automática com a observação direta

das imagens de satélite. Destaca-se o caso da região da Corunha, que, partilhando o mesmo tipo climático do Porto (um dos *datasets* de treino), se apresentava como a área com maior potencial para a obtenção de resultados consistentes e robustos.

Por fim, explorou-se a componente de deteção de alterações (*ChangeDetection*), que à data ainda se encontrava em fase inicial de desenvolvimento. Embora tenham sido testados alguns dos métodos disponíveis, os resultados preliminares não apresentaram a qualidade desejada, razão pela qual não foi possível incluí-los como *outputs* finais do presente trabalho.

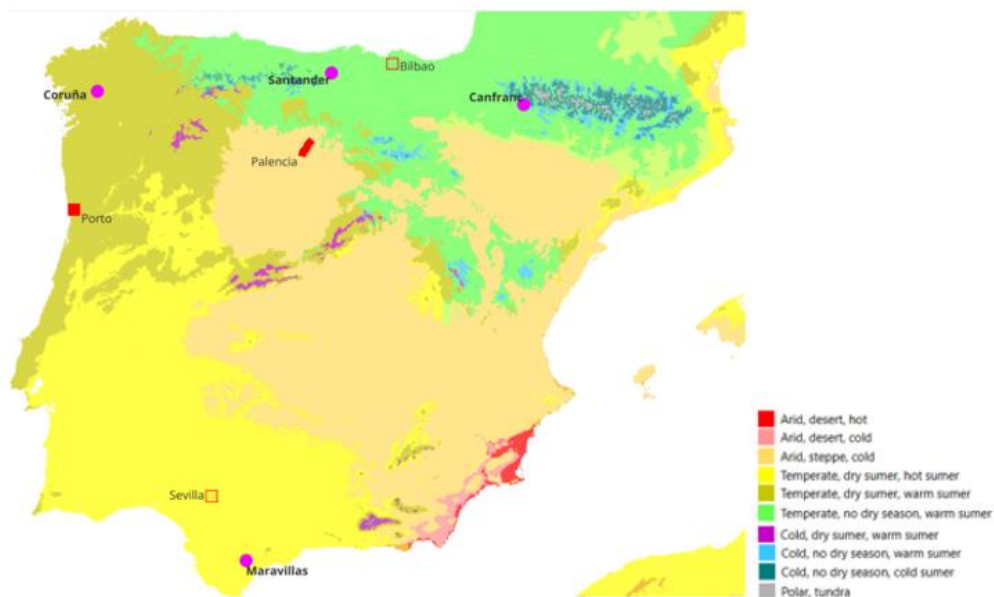


Figura 9-Mapa climático com as áreas de teste do modelo [30];

3.2. Metodologia-Vinhas

3.2.1. Área de Estudo

Este projeto visa, em última instância, desenvolver um modelo de inteligência artificial capaz de detetar automaticamente vinhas através de imagens de satélite de alta resolução. Contudo, nesta fase inicial, a aplicação centra-se especificamente no concelho de Arcos de Valdevez, funcionando como projeto-piloto para aferir a viabilidade do processo e identificar os métodos mais robustos para segmentação de vinhas. O objetivo final passa pela realização de um cadastro completo das vinhas do município, isto é, assegurar o registo atualizado e sistemático de todas as parcelas vitícolas existentes.

O concelho de Arcos de Valdevez, situado no norte de Portugal e composto por 36 freguesias (Figura 10), integra-se na região dos Vinhos Verdes, uma das mais emblemáticas do panorama vitivinícola nacional. A Comissão de Viticultura da Região dos Vinhos Verdes (CVRVV) reconhece esta área como de elevada importância vitícola, destacando-se pelas castas tradicionais, pela topografia diversificada e pelas condições climáticas particularmente favoráveis ao cultivo da vinha. O relevo acidentado, característico do Minho, a presença de escarpas fluviais e a proximidade do Atlântico conferem ao território elevada pluviosidade, verões amenos e invernos suaves, fatores que potenciam o vigor das vinhas [64].



Figura 10-Localização e Freguesias do Concelho de Arcos de Valdevez [65];

A viticultura em Arcos de Valdevez transcende a sua dimensão económica, assumindo um papel estrutural no património cultural e paisagístico local. A necessidade de registar, conservar e mapear com rigor estas vinhas revela-se, assim, fundamental não apenas para a gestão agrícola e para o ordenamento do território, mas também para a valorização do seu património.

3.2.2. Dados utilizados

No âmbito deste projeto, foram utilizados dois conjuntos de dados principais: uma *shapefile* fornecida pelo Instituto da Vinha e do Vinho (IVV) [66] e imagens de satélite de alta resolução disponibilizadas ao **CEiiA** pela empresa **GEOSAT** [67].

Relativamente ao ficheiro vetorial, este incluía um total de 5241 polígonos, cada um representativo de uma vinha registada no concelho de Arcos de Valdevez. Para além

da geometria, este ficheiro incluía diversos atributos relevantes para a caracterização das parcelas vitícolas, nomeadamente:

- **Geocódigo:** identificador administrativo único de cada parcela, associado à freguesia e ao sistema oficial de cadastro nacional;
- **Casta:** espécie da videira;
- **Tipo:** classificação das castas como brancas ou tintas;
- **Área:** dimensão em metros quadrados de cada polígono;
- **Ano de plantação:** registo histórico da data em que cada vinha foi plantada, desde a década de 1930 até ao presente.

Este ficheiro assumiu-se como uma fonte de referência fundamental para o projeto. Contudo, dada a sua natureza histórica e o facto de se encontrar desatualizado, tornou-se evidente a necessidade de desenvolver um método automático de deteção de vinhas. A tarefa de identificar manualmente parcelas inexistentes ou corrigir as já obsoletas revelar-se-ia excessivamente demorada e pouco prática, justificando assim a aposta em soluções baseadas em inteligência artificial.

Complementarmente, recorreu-se a imagens de satélite adquiridas pelo GEOSAT-2, um satélite ótico de observação da Terra lançado em 2014, que orbita a uma altitude de, aproximadamente, 620 km numa órbita heliosíncrona. Equipado com o sensor *High Resolution Advanced Multispectral Sensor* (HiRAS), o GEOSAT-2 disponibiliza imagens pancromáticas e imagens multiespetrais de alta resolução. Para este projeto foram utilizadas 14 imagens multiespetrais, fornecidas diretamente pela própria **GEOSAT** ao **CEiiA**, cobrindo diferentes épocas do ciclo vegetativo da vinha entre os anos de 2023 e 2025. Estas imagens, com resolução espacial de 1 metro e compostas pelas bandas do visível (RGB) e do infravermelho próximo (NIR), revelaram-se particularmente adequadas para a observação de parcelas agrícolas de pequena dimensão, como é o caso de muitas das vinhas.

3.2.3. Preparação do Dataset de Treino

3.2.3.1. Processo de Atualização do Dataset

A *shapefile* fornecida pelo IVV, apesar da sua relevância, apresentava um conjunto significativo de limitações, uma vez que correspondia a um registo histórico das vinhas do concelho de Arcos de Valdevez, incluindo parcelas cuja data de plantação remontava à década de 1930. Naturalmente, muitas dessas vinhas já não se encontravam presentes, tendo sido substituídas por construções, áreas industriais, infraestruturas

rodoviárias ou novas ocupações agrícolas. Assim, tornou-se essencial proceder a uma atualização e limpeza do *dataset*, garantindo que apenas os polígonos correspondentes a vinhas efetivamente existentes fossem mantidos.

Este processo iniciou-se com a aplicação de um critério de seleção baseado na área mínima das parcelas. Considerando que vinhas de reduzida dimensão não eram facilmente distinguíveis nas imagens de satélite, optou-se, inicialmente, por excluir todos os polígonos com área inferior a 1000 m². Após esta filtragem inicial, o *dataset* foi reduzido de 5241 para 1096 polígonos.

De seguida, para acelerar o processo de verificação individual de cada polígono, foi integrada a COS. Foi realizado um *clip* no QGIS para o concelho de Arcos de Valdevez e, posteriormente, foi imposta a condição de que os polígonos das vinhas coincidentes com classes não agrícolas fossem automaticamente sinalizados. Esta etapa permitiu selecionar de imediato um número relevante de polígonos desatualizados e, conseqüentemente, eliminá-los. Os restantes polígonos foram então validados manualmente, recorrendo ao *plugin* do *Google Satellite* como imagem de referência. Cada polígono foi inspecionado individualmente, sendo removidos aqueles que já não correspondiam a vinhas. Este trabalho incluiu também a correção de casos em que o polígono abrangia não apenas vinhas, mas também áreas adjacentes de vegetação espontânea, árvores isoladas ou elementos não agrícolas, assegurando que os dados finais representassem exclusivamente píxeis de vinha. A Figura 11 ilustra um exemplo de polígono inicialmente classificado como vinha no *shapefile* original, mas que, à data da verificação, já não apresentava qualquer cultura vitícola.

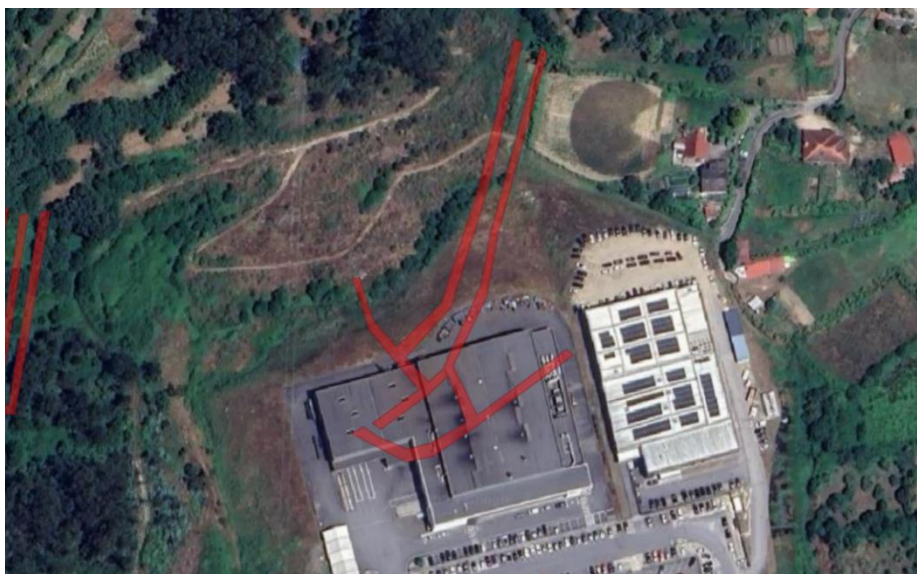


Figura 11-Polígono mal classificado como vinha;

No final deste processo, obteve-se um *dataset* significativamente mais fiável e representativo da realidade atual do concelho de Arcos de Valdevez. Para efeitos de acompanhamento e comunicação do progresso ao longo do projeto, foram geradas estatísticas descritivas que caracterizam a constituição final do *dataset*. Estas incluem a distribuição das vinhas de acordo com o ano de plantação, bem como o número e a percentagem de polígonos eliminados em diferentes décadas de referência. Estes resultados, apresentados sob a forma de gráficos e tabelas, encontram-se detalhados na secção de resultados.

3.2.3.2. Validação do *Dataset* com Imagens GEOSAT-2

De seguida, com o *dataset* atualizado, tornou-se necessário reforçar a sua robustez e aumentar o número de exemplos representativos, assegurando que os diferentes estágios fenológicos da vinha estivessem contemplados. Para tal, recorreu-se a um conjunto de 14 imagens multiespectrais do satélite GEOSAT-2, adquiridas entre 2023 e 2025, assegurando a presença das várias estações do ano. Estas imagens, disponibilizadas com uma resolução espacial de 1 metro, permitiram observar as vinhas com elevado detalhe, constituindo um suporte fundamental para a validação do *dataset*.

Dado que várias imagens não abrangiam a totalidade do concelho de Arcos de Valdevez, foi necessário realizar uma verificação individualizada para cada *footprint*. Assim, em cada imagem, foram eliminados os polígonos que se encontravam fora da sua área de cobertura. Os restantes polígonos foram analisados manualmente, de modo a confirmar se correspondiam de facto a vinhas na época em causa. Este processo foi particularmente relevante para casos de plantações recentes, que em imagens de 2023 ainda não apresentavam características visíveis de vinha, exigindo a sua exclusão. Outro fator de correção resultou da presença de nuvens em algumas imagens, que comprometiam a identificação visual das vinhas. Nestes casos, os polígonos localizados nas áreas afetadas foram removidos, de forma a garantir a integridade e a fiabilidade dos dados.

O resultado final deste processo consistiu na criação de 14 pares de *datasets*, cada um composto por uma imagem de satélite GEOSAT-2 e o respetivo ficheiro vetorial com os polígonos validados.

3.2.3.3. Pré-processamento do *Dataset*

Face às dificuldades em diferenciar vinhas apenas com base nos valores espectrais, tornou-se evidente que o modelo a desenvolver deverá recorrer a técnicas de DL e segmentação semântica. Ao contrário dos métodos clássicos, que se apoiam sobretudo

em informação espectral, esta abordagem permite explorar também características espaciais e morfológicas, mais adequadas para identificar vinhas face a outras formas de vegetação. Assim, com o *dataset* já validado, procedeu-se a uma etapa de pré-processamento, essencial para preparar os dados para o treino do modelo de segmentação. Este processo envolveu duas operações principais: a conversão dos ficheiros vetoriais em máscaras binárias e a subdivisão das imagens em *tiles* de dimensão fixa.

A criação de máscaras binárias consistiu em transformar a camada vetorial dos polígonos das vinhas em *rasters*, onde cada pixel assume o valor 1 se corresponder a uma área de vinha e 0 no caso contrário. Estas máscaras constituem o *ground truth* do processo de treino, ou seja, a referência a partir da qual as previsões do modelo são comparadas. Ao fornecer ao algoritmo pares de dados, a imagem multiespectral como *input* e a máscara binária como *output* esperado, o modelo conseguirá aprender não só as diferenças espectrais, mas também os padrões geométricos e espaciais característicos das vinhas.

Após a criação das máscaras, avançou-se para o processo de *tiling*, no qual as imagens de satélite e as respetivas máscaras foram subdivididas em parcelas de 256×256 píxeis. Este procedimento é indispensável, uma vez que permite normalizar o tamanho dos *inputs* fornecidos à rede neuronal, requisito essencial na maioria das arquiteturas de DL, ao mesmo tempo que multiplica o número de amostras disponíveis, aumentando a diversidade de situações observadas pelo modelo e, conseqüentemente, a sua capacidade de generalização. Para além disso, possibilita o processamento eficiente de imagens de muito alta resolução, reduzindo a carga computacional e tornando o treino mais viável. Finalmente, ao segmentar as imagens em pequenas parcelas, o modelo poderá concentrar-se em padrões locais e estruturais, cruciais para distinguir vinhas de outros tipos de vegetação.

Importa salientar que, quando do término do estágio, o modelo ainda se encontrava em fase de desenvolvimento, pelo que não foi possível obter resultados. Nesta etapa, os *outputs* obtidos correspondem exclusivamente às máscaras binárias aplicadas sobre os *tiles*, constituindo a base sólida para os futuros avanços do projeto. Exemplos destes *outputs* serão apresentados e discutidos na secção de resultados, de modo a ilustrar o processo de preparação do *dataset* e a sua adequação para treino de modelos de segmentação semântica.

4. Resultados e Discussão

Este capítulo está estruturado em duas secções principais, cada uma dedicada a um dos projetos desenvolvidos durante o estágio. Em cada secção, serão apresentados os resultados obtidos, seguidos de uma análise crítica que permitirá efetuar uma avaliação global do desempenho e dos resultados obtidos até à data. Adicionalmente, serão discutidas possíveis soluções para os problemas identificados, promovendo uma reflexão sobre eventuais melhorias e otimizações a implementar.

4.1. Resultados-LULC

4.1.1. Desempenho do modelo

Tal como referido anteriormente, o modelo foi treinado com os *datasets* de Palência e do Porto, recorrendo às 41 bandas disponíveis, com exceção da 1 e da 9. Para a fase de treino, foi utilizado o classificador RF, executado com recurso a GPU, o que permitiu reduzir significativamente o tempo de processamento e garantir uma gestão mais eficiente do grande volume de dados em análise.

A avaliação do desempenho do treino do modelo foi feita através de diferentes métricas. Entre elas, destaca-se a matriz de confusão (Figura 11), que ilustra a capacidade do modelo em distinguir as diferentes classes do LULC. De uma forma geral, o modelo apresentou uma boa performance na separação da maioria das classes. Contudo, observou-se maior dificuldade na diferenciação das categorias de vegetação, nomeadamente entre *Shrubland* e *Forest*, verificando-se que muitos píxeis correspondentes a *Shrubland* foram classificados como *Forest*. Este resultado era esperado, dado que estas duas classes apresentam assinaturas espectrais bastante semelhantes. Para mitigar este tipo de erro, poderão ser exploradas diferentes estratégias, tais como o refinamento dos polígonos de treino, a introdução de índices espectrais mais específicos para distinguir vegetação arbustiva de vegetação arbórea ou, ainda, a integração de dados auxiliares que capturem características estruturais ou biofísicas complementares, como a biomassa e o carbono.

A importância das bandas espectrais na classificação é apresentada na Figura 12. Esta apresenta duas medidas de importância, a *Feature Importance*, que indica o quanto cada banda foi utilizada pelo modelo para tomar decisões, e a *Permutation Importance*, que avalia o impacto de cada banda nas previsões do modelo ao permutar os seus valores de forma aleatória, mostrando assim o quanto o modelo, realmente, depende

A Figura 13 permite observar o comportamento do modelo para cada categoria, através das métricas de precisão, *recall* e F1-score. Classes bem representadas no treino tendem a apresentar valores elevados e equilibrados nestas métricas, enquanto classes mais complexas ou espectralmente próximas a outras, como é o caso de *Shrubland*, registam uma ligeira redução de desempenho.

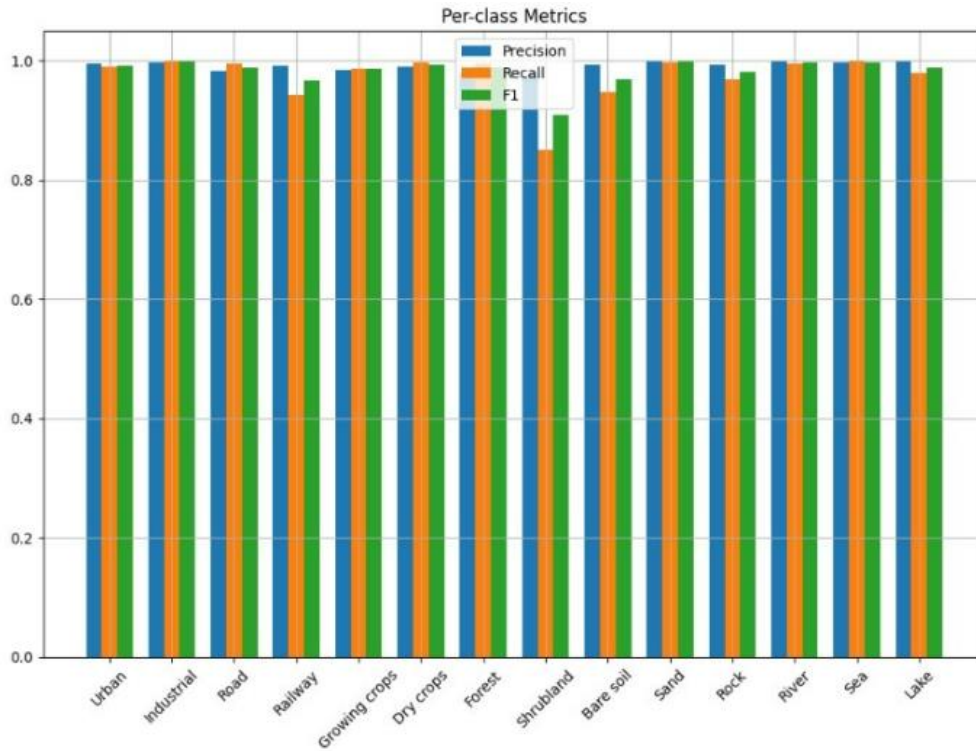


Figura 14-Gráfico das métricas de precisão, *recall* e F1-score por classe obtido após o treino do modelo RF com os *datasets* do Porto e de Palência;

Em suma, os resultados obtidos demonstram que o modelo apresenta uma performance globalmente sólida, conseguindo distinguir de forma robusta a maioria das classes LULC. As principais limitações prendem-se com a diferenciação entre classes de vegetação estruturalmente semelhantes, aspeto que poderá ser melhorado em versões futuras através da introdução de novos índices, dados complementares ou metodologias de classificação mais avançadas. Importa salientar que os resultados aqui apresentados dizem respeito exclusivamente ao desempenho do modelo nos dados utilizados para treino e validação. No capítulo seguinte serão analisados os *outputs* gerados pelo modelo em áreas externas ao treino, permitindo avaliar a sua capacidade de generalização e a robustez das classificações obtidas em novos contextos.

4.1.2. Resultados das Inferências

Com o modelo já treinado, foram realizadas inferências para áreas adjacentes a linhas de caminho-de-ferro em Canfranc, Maravillas, Corunha e Santander, aplicando-se *buffers* de 150 metros em torno da linha, bem como para uma área de maior dimensão em Zamora. A Figura 15 apresenta o mapa resultante da classificação LULC para esta última área, seguido do mosaico correspondente.

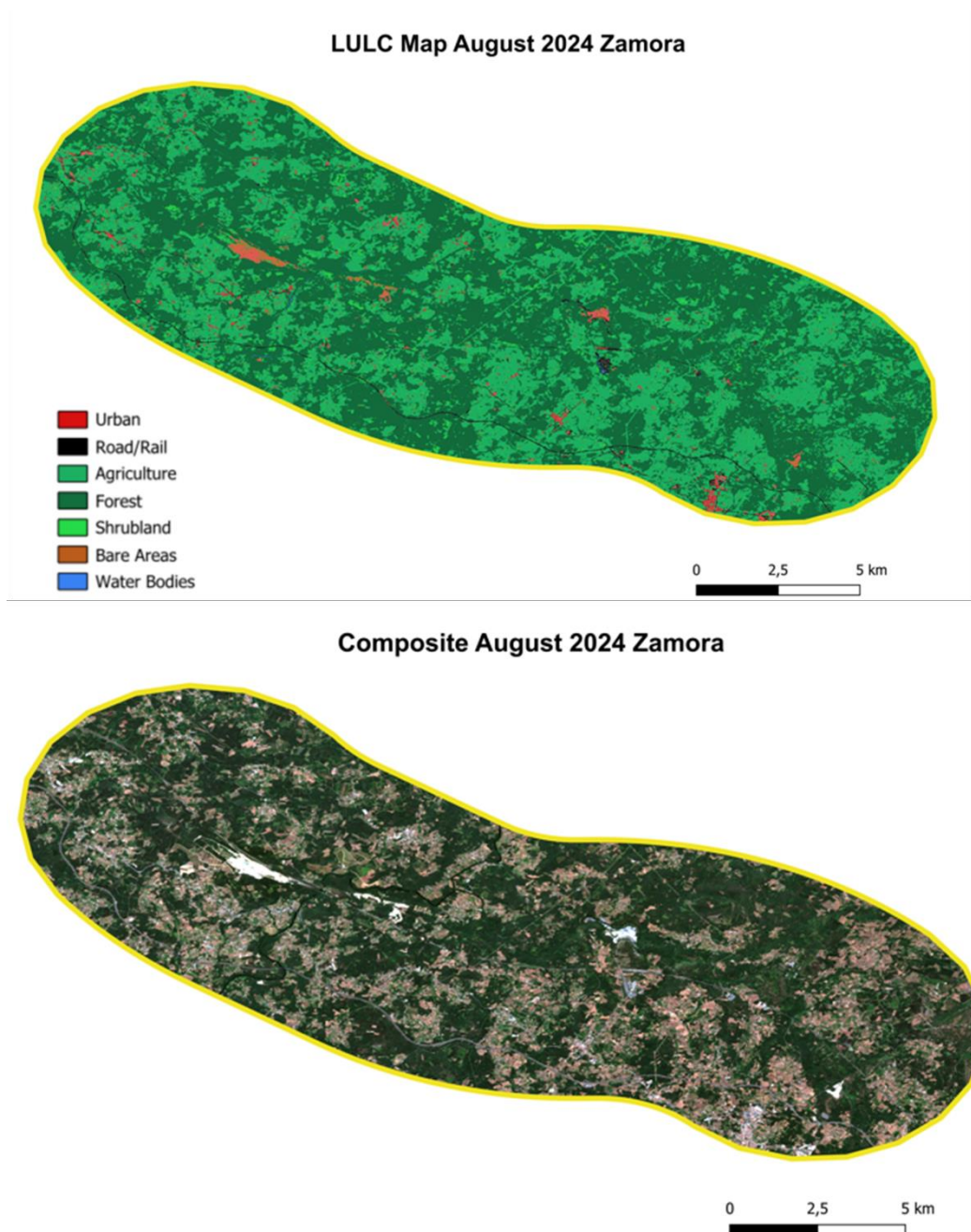


Figura 15-Mapa LULC de Zamora obtido a partir de uma inferência realizada com o modelo RF, referente a agosto de 2024;

De uma análise inicial, realizada a partir da Figura 15, observa-se que o modelo procede à classificação da totalidade dos pontos presentes no interior do *buffer*, produzindo um resultado que, globalmente, aparenta enquadrar-se de forma consistente com o que se observa no mosaico de referência. Importa, no entanto, salientar que, apesar de neste caso o território se encontrar integralmente classificado, existem outros *outputs* em que determinados píxeis permanecem sem classe atribuída.

Esta ausência de classificação não se deve a falhas do modelo em si, mas sim às características do *raster* de entrada. Concretamente, um pixel não é classificado quando ocorre uma ausência total de informação espectral no *raster* correspondente ou, alternativamente, quando pelo menos uma das bandas apresenta valores *NoData*. Esta condição foi previamente definida no processamento de forma a evitar a introdução de ruído na classificação, uma vez que a ausência de dados numa única banda comprometeria o cálculo de diversos índices derivados, conduzindo a resultados incorretos.

A Figura 16 ilustra um destes casos, evidenciando áreas em que, pela inexistência de valores espectrais nos píxeis ou no caso de pelo menos uma banda apresentar valores *NoData*, o modelo não atribui qualquer classe, representando-as a branco.

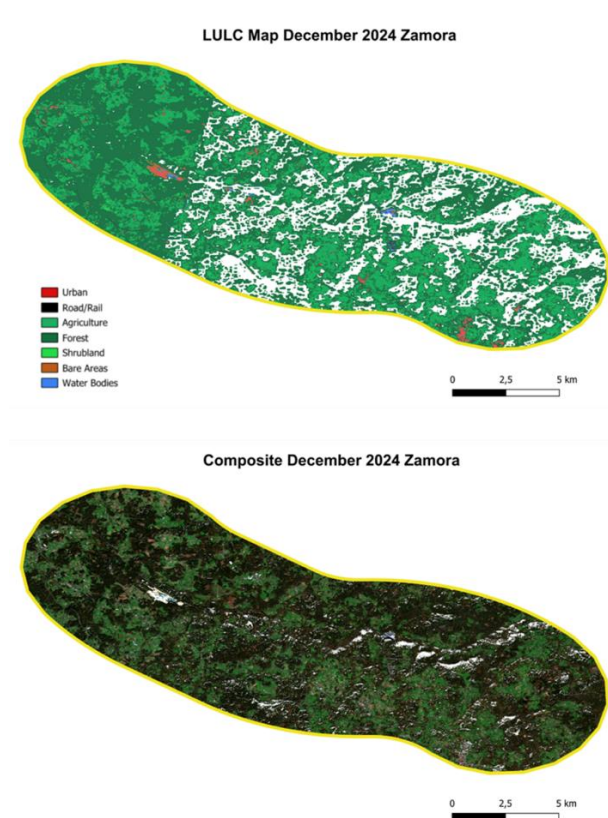


Figura 16-Mapa LULC com valores *NoData* obtido em Zamora, referente a dezembro de 2024;

Apesar de, à primeira vista, os resultados apresentados na Figura 15 parecerem consistentes com o mosaico de referência, uma análise mais detalhada de todas as áreas evidenciou a existência de um conjunto de erros sistemáticos. Um dos problemas mais recorrentes, particularmente destacado em Canfranc, está associado à presença de sombras, que o modelo tende a interpretar como corpos de água.

Este fenómeno pode ser explicado pelas condições topográficas da região. Canfranc localiza-se num vale montanhoso, onde a topografia condiciona significativamente a incidência solar. Nesta região, as imagens do Sentinel-2 são adquiridas por volta das 11h (hora local), mas, devido à altitude das montanhas circundantes, o sol ainda não se encontra suficientemente elevado para iluminar plenamente toda a área. Como consequência, formam-se zonas de sombra em que a informação espectral captada é distorcida, deixando de corresponder aos padrões das classes presentes no treino. Nessas circunstâncias, o modelo tende a atribuir incorretamente a classe *Water Bodies* a píxeis que, na realidade, pertencem a áreas terrestres.

Este problema manifesta-se de forma mais expressiva durante os meses de inverno, entre novembro e fevereiro, quando a altura solar é significativamente menor. Durante o verão, há uma melhoria notável nas condições, devido à maior altura solar, que proporciona uma iluminação mais uniforme e consistente. As Figuras 17 e 18 ilustram claramente este efeito, mostrando a classificação LULC de janeiro, onde as sombras provocam uma falsa deteção de água, comparativamente a maio, mês em que a maior incidência solar reduz significativamente este erro. Para reforçar esta análise, a Figura 19 apresenta ainda uma sequência mensal do mosaico ao longo de 2024, evidenciando que esta questão se concentra essencialmente nos meses de inverno.

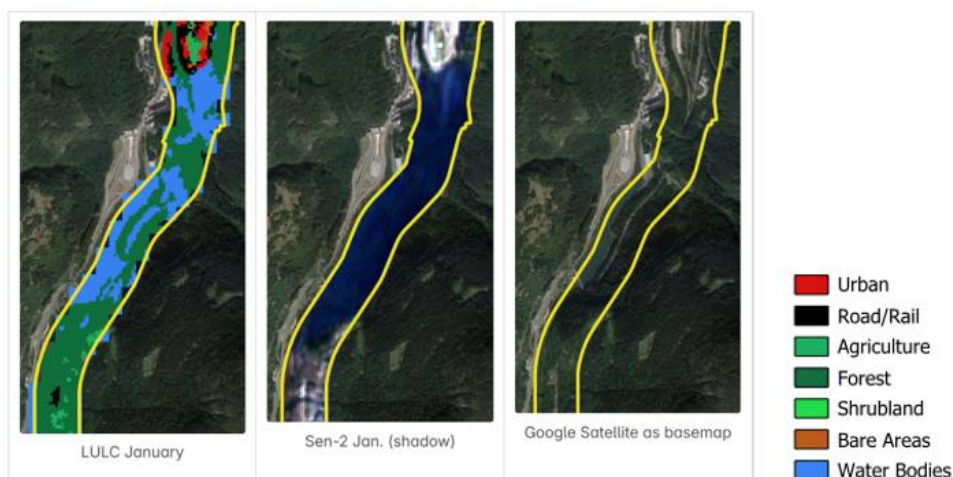


Figura 17-Efeito causado pela sombra na classificação do modelo (Canfranc, janeiro 2024);



Figura 18-Classificação do modelo sem o efeito da sombra (Canfranc, maio 2024);

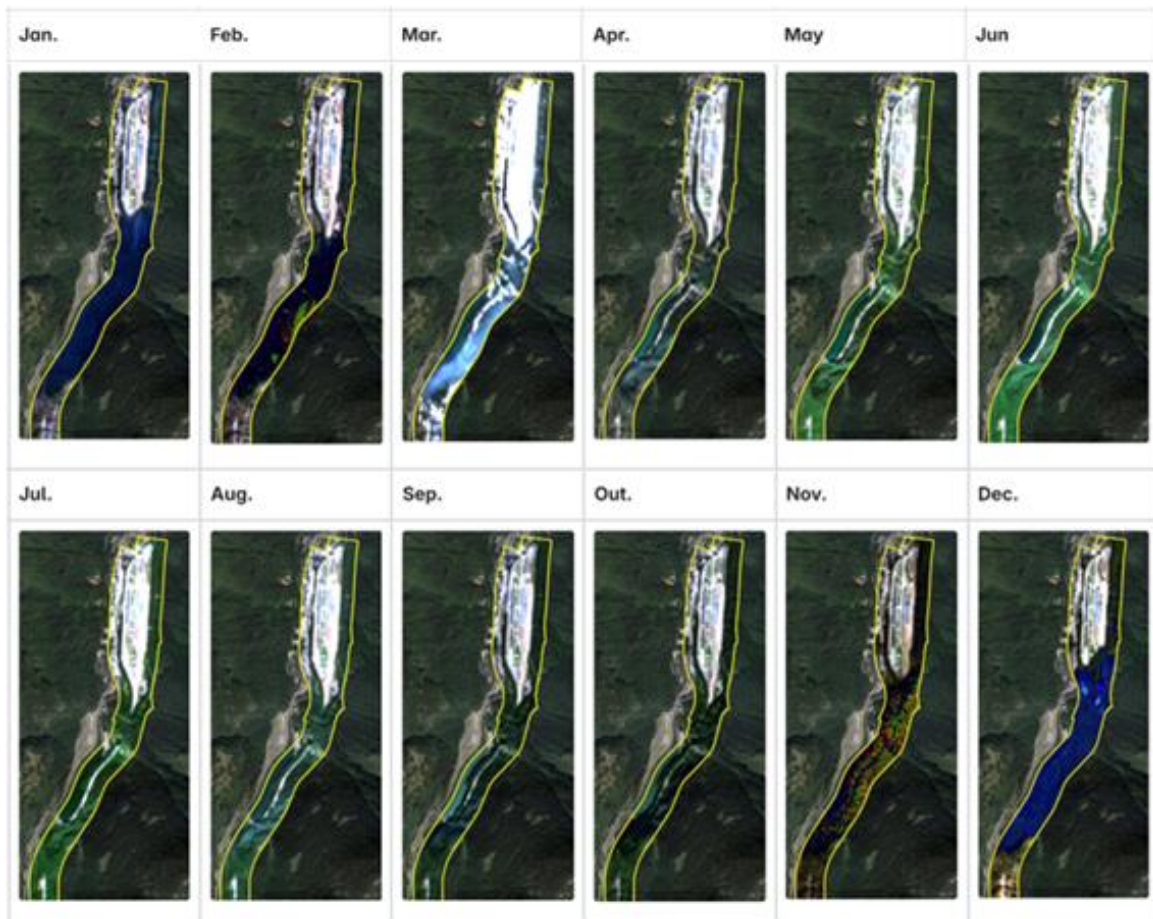


Figura 19-Efeito das sombras nas imagens Sentinel-2 ao longo do ano (Canfranc, 2024);

Este erro poderá ser mitigado através de diferentes abordagens complementares. Uma primeira estratégia passa pela incorporação, durante o treino, de áreas sujeitas a sombra, de modo a permitir que o modelo aprenda a reconhecer e a lidar com este tipo

de condição. Em alternativa, ou de forma adicional, poderá ser integrada informação topográfica proveniente de modelos digitais de elevação (DEM), nomeadamente variáveis como o *aspect*, que descreve a orientação das vertentes e fornece contexto espacial relevante para melhorar a precisão da classificação. Uma terceira possibilidade consiste na identificação direta dos píxeis afetados por sombra e na sua marcação como *NoData*, evitando assim a atribuição incorreta de classes em zonas onde a informação espectral se encontra comprometida. Embora nenhuma destas soluções assegure, por si só, a eliminação completa do problema, a sua aplicação poderá aumentar de forma significativa o desempenho do modelo em cenários semelhantes.

Nas Figuras 17 e 18, e também noutros casos analisados, verificam-se píxeis classificados como água junto às bordas do *buffer*, por vezes ultrapassando mesmo o limite definido. Este comportamento está associado ao processo de reamostragem das bandas do Sentinel-2. Como se sabe, nem todas as bandas possuem a mesma resolução espacial (10, 20 e 60 metros no caso do Sentinel-2). Para harmonizar os dados e possibilitar a sua utilização conjunta no modelo, é necessário proceder à reamostragem das bandas de 20 e 60 metros para os 10 metros. No entanto, quando esta operação é realizada apenas após o recorte do *raster* para o *buffer* em estudo, perde-se a informação de vizinhança presente no *raster* completo, levando a que os algoritmos de interpolação produzam valores artificiais junto às margens. Estes valores alteram a assinatura espectral dos píxeis de fronteira, originando classificações incorretas, neste caso, a atribuição à classe correspondente aos cursos de água. Assim, é fundamental realizar a reamostragem previamente ao recorte para a área de estudo, garantindo que todos os cálculos de interpolação dispõem da totalidade da informação, evitando assim distorções nas bordas do *buffer*.

Outro erro identificado em alguns *outputs* do modelo corresponde à classificação incorreta de áreas de *Bare Soil* como áreas urbanas. Esta situação poderá estar relacionada com as características espectrais destes píxeis, que em determinados contextos apresentam valores muito elevados de refletância, assumindo tonalidades próximas do branco. Este comportamento aproxima-os visualmente das áreas industriais, que integram a classe das áreas urbanas. A Figura 20 ilustra um exemplo deste erro observado em Maravillas.

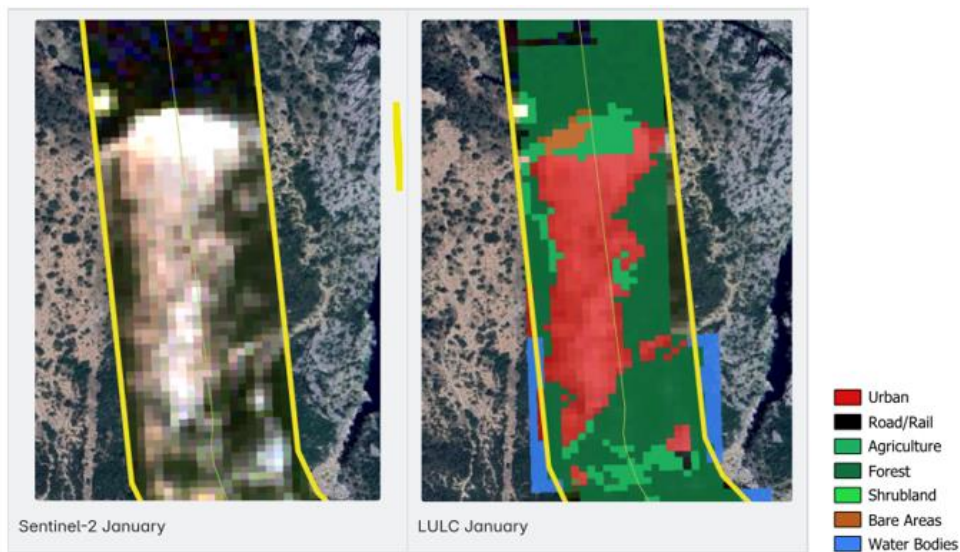


Figura 20-Classificação incorreta de *Bare Soil* como *Urban* (Maravillas, janeiro 2024);

Ainda no que diz respeito à classe *Urban*, verificou-se que determinados troços de estradas e linhas de caminho de ferro foram classificados como áreas urbanas. A Figura 21 ilustra um desses erros, detetado em Maravillas. Esta confusão é compreensível, uma vez que, tal como evidenciado pela matriz de confusão, existe uma forte correlação entre as classes *Urban* e *Roads/Railways*. No entanto, no âmbito específico deste projeto, a distinção entre estas duas categorias não assume grande relevância prática. Consequentemente, na fase de *Change Detection*, será implementado um critério de seleção que desconsidera as mudanças entre estas classes, de modo a evitar a geração de falsos positivos em áreas onde a mudança é pouco significativa.

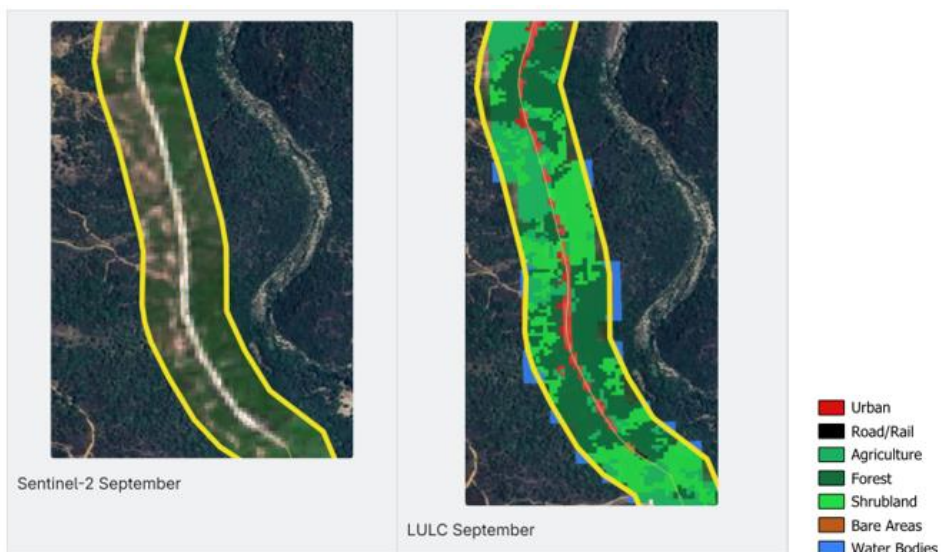


Figura 21-Classificação incorreta de *Railways* como *Urban* (Maravillas, setembro 2024);

Para além dos erros mais recorrentes anteriormente descritos, foram também identificadas algumas incorreções pontuais. Em determinados casos, registou-se a presença de píxeis classificados como água em estradas, bem como a classificação como *Urban* de parcelas agrícolas secas. A Figura 22 exemplifica estas situações detetadas em Zamora.

Relativamente às classes de vegetação, os resultados demonstraram um desempenho globalmente satisfatório. O modelo conseguiu distinguir de forma consistente áreas florestais de superfícies agrícolas, revelando boa capacidade de generalização nestas categorias. Contudo, a classe *Shrubland* mantém-se como um dos principais desafios. A sua classificação surge frequentemente fragmentada, com blocos dispersos de pequenos píxeis atribuídos a esta classe no seio de áreas predominantemente florestais ou agrícolas, como ilustrado na Figura 23. Deste modo, embora os resultados obtidos constituam uma base sólida, a classe *Shrubland* persiste como um dos maiores desafios a superar em desenvolvimentos futuros.

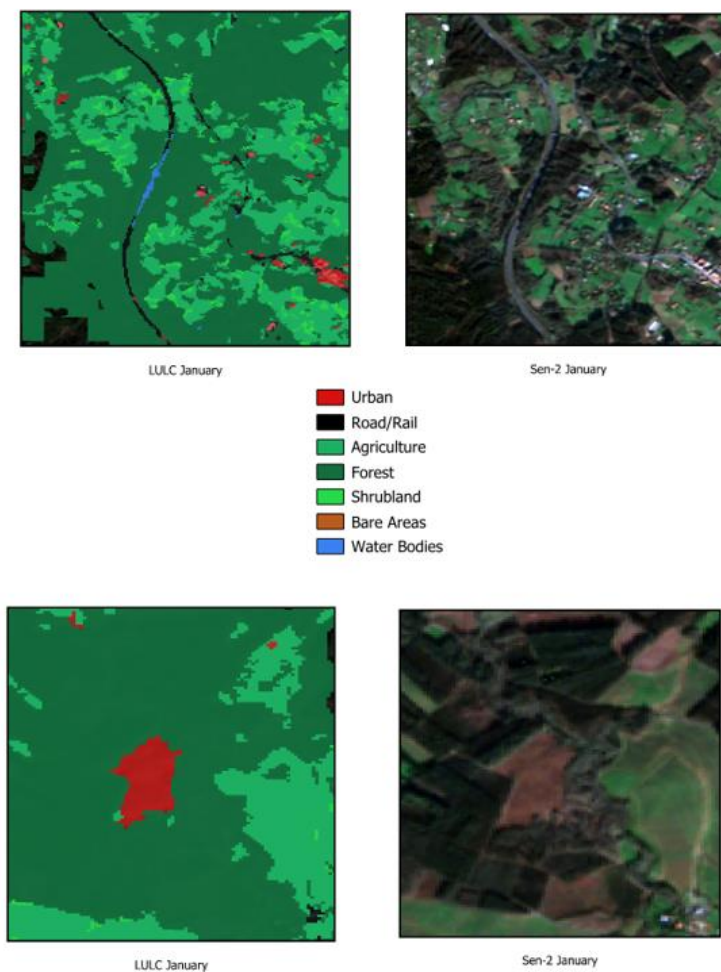


Figura 22- Erros de classificação (Zamora, janeiro 2024);

LULC February

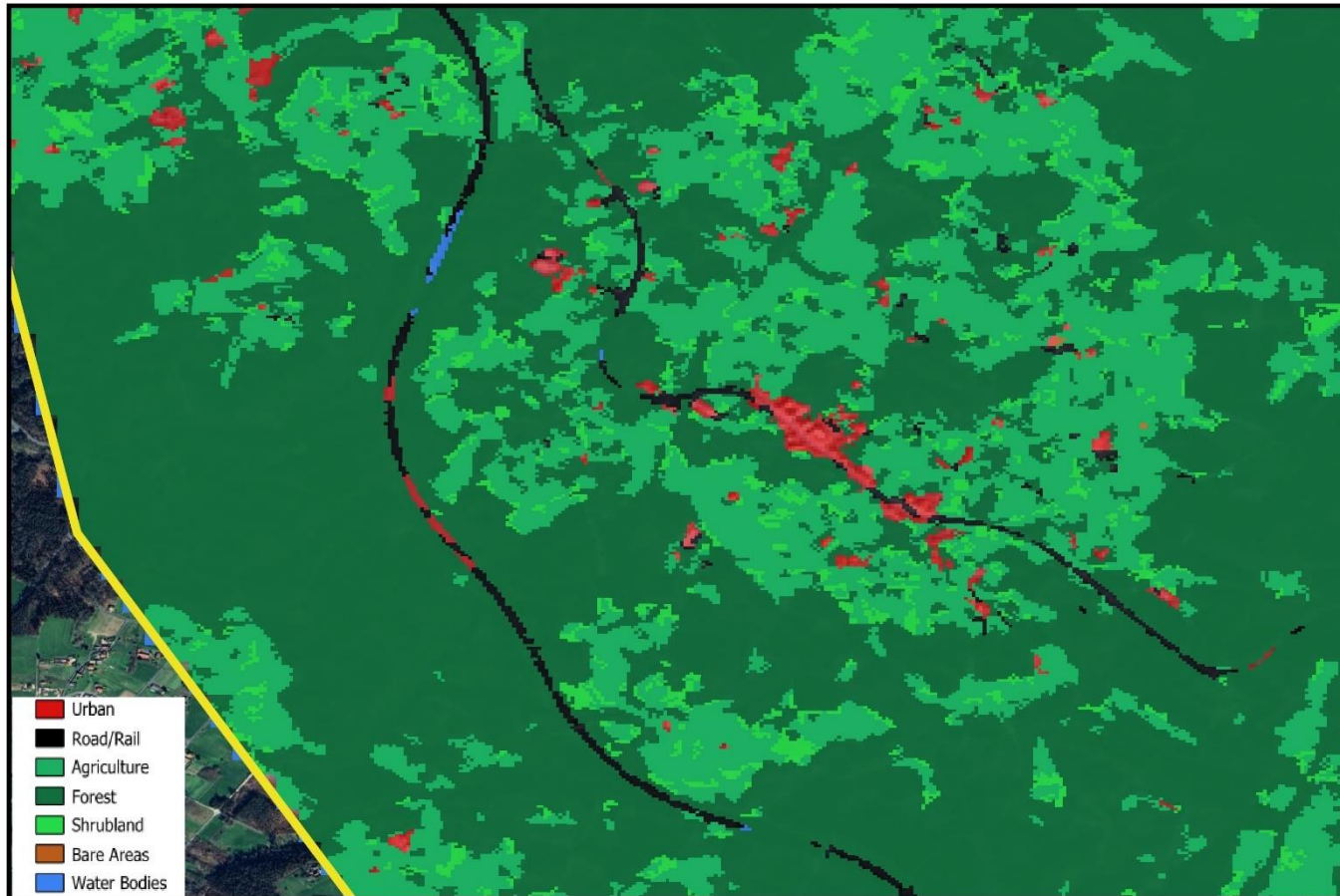


Figura 23-Mapa LULC (Zamora, fevereiro de 2024);

Composite February

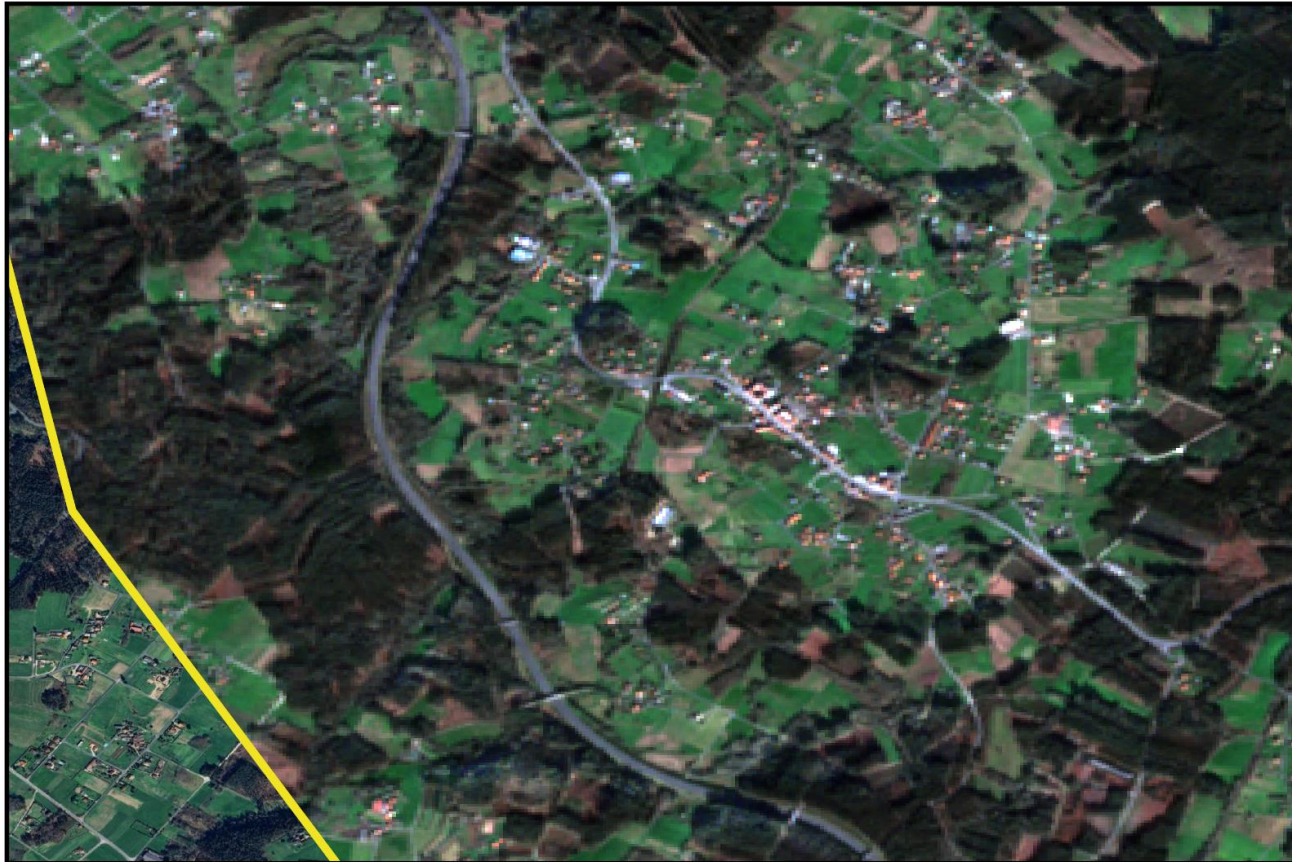


Figura 24-Mosaico de Zamora (fevereiro, 2024);

A análise detalhada das Figuras 23 e 24, relativas a uma área de Zamora, permite confirmar que o modelo apresenta já resultados consistentes na classificação das diferentes classes LULC. Verifica-se uma distinção globalmente correta entre áreas agrícolas e florestais, bem como uma boa identificação de estradas e áreas urbanas. Embora o exemplo apresentado corresponda ao mês de fevereiro, o comportamento do modelo mantém-se estável nos meses de verão, conseguindo igualmente diferenciar áreas agrícolas mais secas das restantes classes. Ainda assim, em algumas situações específicas, determinados campos secos são confundidos com áreas urbanas, revelando limitações do classificador.

Apesar de constituírem uma base sólida para uma fase ainda inicial de desenvolvimento, os resultados estão longe de ser perfeitos. Na Figura 23 verificam-se erros já anteriormente discutidos, como a presença de píxeis de água ou urbanos em estradas, bem como a fragmentação da classe *Shrubland*, que surge em pequenos blocos dispersos no meio de florestas e áreas agrícolas, continuando a representar um dos principais desafios do modelo. Observa-se ainda que as estradas surgem por vezes interrompidas por outras classes.

Outro aspeto a destacar é a dificuldade do modelo em representar elementos de reduzida dimensão, como habitações isoladas ou estradas estreitas. Verifica-se que, em muitos casos, estes elementos não surgem representados no *output*. Nestes casos, o modelo tende a reatribuir píxeis individuais à classe predominante na sua vizinhança, resultando na eliminação de píxeis isolados. Embora este procedimento possa comprometer a identificação de objetos de pequena escala, a sua ocorrência não constitui um problema para o objetivo principal do projeto. Pelo contrário, esta generalização contribui para reduzir a ocorrência de falsos positivos e minimizar o ruído na geração das máscaras de mudança e de severidade, assegurando resultados mais consistentes e interpretáveis.

4.1.3. Discussão

A análise realizada ao longo deste capítulo permitiu identificar tendências relevantes e limitações significativas no desempenho do modelo de classificação LULC. De forma geral, os resultados obtidos mostram que o modelo é capaz de distinguir adequadamente a maioria das classes, conseguindo uma boa separação entre áreas agrícolas, florestais e urbanas. Contudo, também se verificaram erros sistemáticos, dos quais se destacam a dificuldade em lidar com zonas de sombra e a dificuldade em diferenciar a classe *Shrubland* em relação às áreas agrícolas e às florestas. Para além

destes erros mais recorrentes, foram também observados alguns problemas pontuais como a classificação incorreta de algumas classes. Apesar destas limitações, os *outputs* obtidos constituem já uma base sólida de trabalho, revelando o potencial do modelo numa fase ainda inicial de desenvolvimento.

Face a estes resultados, torna-se evidente a necessidade de implementar medidas que visem melhorar a robustez e a capacidade de generalização do modelo. Uma primeira abordagem passa pelo treino com um conjunto mais alargado de *datasets*, incluindo os das regiões de Sevilha, Bilbao, Torre Vieja e Pradollano que já se encontram preparados. Este procedimento, para além de aumentar a quantidade de dados disponíveis, permitirá introduzir exemplos provenientes de diferentes contextos climáticos, potenciando a capacidade de adaptação do modelo a condições mais diversas.

Outra medida fundamental consiste na integração de dados auxiliares que possam complementar a informação espectral de base. A utilização de variáveis topográficas derivadas de DEM revela-se particularmente promissora para mitigar os erros causados pelas sombras em áreas montanhosas. Além disso, o treino de polígonos correspondentes a áreas de sombras pode vir a ser uma solução viável. De igual modo, a incorporação de dados relacionados com a biomassa poderá melhorar a capacidade de diferenciar entre tipos de vegetação, aspeto especialmente relevante para enfrentar as dificuldades verificadas na classificação da classe *Shrubland*.

Adicionalmente, importa explorar a afinação de parâmetros do modelo de *Random Forest*. Pequenos ajustes neste nível poderão ter um impacto significativo na precisão da classificação. A médio prazo, poderá ainda ser vantajosa a adoção de metodologias mais avançadas baseadas em DL e segmentação semântica, as quais apresentam maior capacidade para explorar padrões espaciais e contextuais de forma mais detalhada do que os algoritmos baseados apenas em informação espectral, como está comprovado pela literatura.

Em suma, os resultados obtidos até ao momento, embora já evidenciem alguns indicadores promissores, revelam que o modelo ainda apresenta limitações significativas, o que é natural, dado que se encontra numa fase inicial de desenvolvimento. Com o reforço e diversificação do treino, a integração de dados complementares e a evolução para técnicas mais avançadas, é expectável que a performance do modelo seja substancialmente melhorada, consolidando a sua aplicabilidade prática no contexto de monitorização e deteção de mudanças no uso e na ocupação do solo.

4.2. Resultados-Vinhas

4.2.1. Resultados

No caso específico da componente dedicada à deteção automática de vinhas, os resultados obtidos encontram-se ainda numa fase preliminar, dado que o modelo de classificação baseado em DL e segmentação semântica não se encontrava totalmente desenvolvido no final do estágio. Assim, nesta secção apresentam-se apenas os *outputs* referentes ao processo de construção e preparação do *dataset*, que constituirá a base para o treino do modelo numa fase subsequente.

Após o processo de limpeza do ficheiro vetorial inicialmente disponibilizado pelo IVV, verificou-se uma redução do número total de parcelas de vinha. Dos 1096 polígonos originais, apenas 793 se mantiveram válidos para a construção do *dataset*, o que corresponde a cerca de 72,4% do total inicial. Os restantes 27,6% foram eliminados por já não corresponderem a vinhas ativas ou por apresentarem inconsistências que inviabilizavam a sua utilização. A Tabela 4 apresenta a distribuição dos polígonos por década de plantação, antes e depois do processo de limpeza, bem como a percentagem de polígonos eliminados em cada período. Como seria de esperar, a década de 1930, a mais antiga no *dataset*, apresenta a maior taxa de exclusão, com cerca de 55% das vinhas inicialmente registadas já inexistentes na atualidade. Este resultado reflete a dinâmica de substituição e abandono de vinhas mais antigas, contribuindo para um *dataset* final mais atualizado e representativo da realidade vitivinícola do concelho de Arcos de Valdevez.

Além disso, foi também avaliada a constituição do *dataset* final de treino em termos de representatividade por década. Observou-se que as décadas de 1930 e 2010 se destacam como as mais representadas, correspondendo a mais de metade do total de vinhas disponíveis. Esta distribuição encontra-se representada no diagrama da Figura 25.

Tabela 4-Estatísticas do *dataset* após o processo de atualização;

Dataset Refinado			
Ano de Plantação	Nº Polígonos Iniciais	Nº Polígonos Finais	Polígonos eliminados (%)
1930-1940	495	223	<u>55%</u>
1950-1959	2	0	100%
1980-1989	20	17	15%
1990-1999	96	85	11%
2000-2009	118	115	3%
2010-2019	280	270	4%
2020-2025	85	83	2%
Total	1096	793	

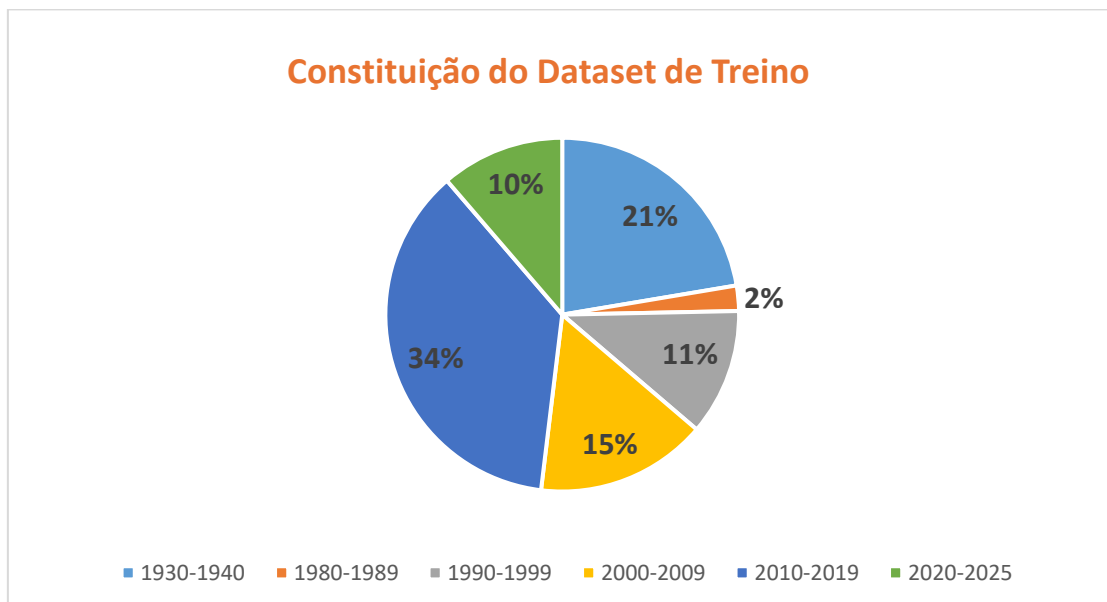


Figura 25-Composição do *dataset* de treino segundo as décadas de plantação das vinhas;

Após a validação dos polígonos e respetiva correspondência com as imagens do satélite GEOSAT-2, foi desenvolvido o processo de pré-processamento dos dados, tal como

descrito na metodologia. Este processo consistiu na conversão dos ficheiros vetoriais em máscaras binárias e na divisão das imagens em *tiles* regulares de 256 × 256 píxeis. A Figura 26 ilustra alguns exemplos dos *outputs* produzidos nesta fase. Estes resultados constituem a base estruturada a partir da qual será treinado o modelo de deteção automática de vinhas.

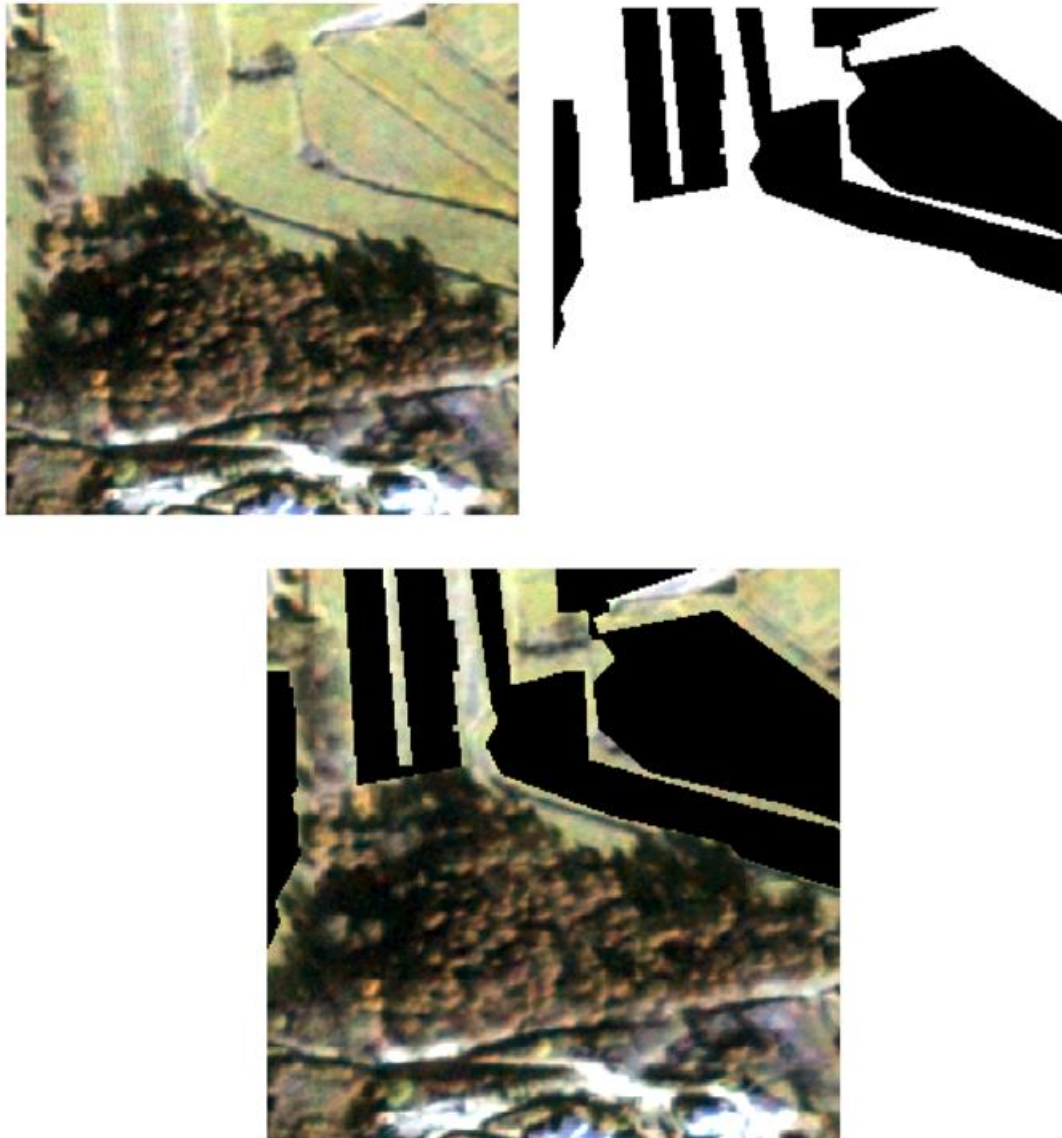


Figura 26-Dados pré-processados (*Tile* GEOSAT-2; Máscara binária, em que as áreas a preto representam as vinhas; Sobreposição da máscara sobre a imagem);

4.2.2 Discussão e Perspetivas Futuras

Os dados necessários para o treino inicial do modelo foram preparados com sucesso, permitindo a criação de um *dataset* validado e devidamente estruturado. As etapas de limpeza, validação e pré-processamento asseguraram a construção de uma base sólida para o desenvolvimento de um modelo de deteção automática de vinhas.

Numa fase posterior, o modelo será desenvolvido com recurso a técnicas de *deep learning* e segmentação semântica, abordagens particularmente adequadas a este tipo de problema, uma vez que exploram não apenas os padrões espectrais, mas também as características espaciais e morfológicas que distinguem as vinhas de outras coberturas do solo. Após o treino, os resultados serão avaliados quantitativa e qualitativamente, recorrendo a métricas de desempenho e a análises visuais. É expectável que, para melhorar a robustez do modelo, seja necessário aumentar o número de polígonos de treino, de forma a capturar a diversidade de padrões que caracterizam as vinhas em diferentes fases fenológicas. Uma das principais incertezas prende-se, contudo, com a própria resolução espacial das imagens GEOSAT-2 (1 metro). Apesar de garantir já um nível de detalhe bastante elevado, permanece a dúvida sobre se será o suficiente para treinar um modelo que seja capaz de discriminar com clareza todas as vinhas. Neste sentido, poderá ser necessário recorrer a imagens de muito alta resolução adquiridas por drones. Paralelamente, será importante explorar e testar índices espectrais que possam contribuir para uma melhor diferenciação entre vinhas e outros tipos de vegetação.

Embora o presente projeto tenha como objetivo inicial a deteção e o cadastro automático das vinhas no concelho de Arcos de Valdevez, a sua eventual consolidação poderá abrir caminho a múltiplas aplicações no domínio da agricultura de precisão. Para além da cartografia automatizada, modelos desta natureza poderão apoiar, por exemplo, a monitorização do estado vegetativo das vinhas, a avaliação de situações de stress hídrico e a deteção precoce de pragas e doenças.

Deste modo, este projeto-piloto poderá não só contribuir para a modernização da gestão vitivinícola na região dos Vinhos Verdes, como também servir de base para a criação de soluções mais abrangentes de monitorização agrícola, escaláveis a outros territórios e culturas.

5. Conclusão

O trabalho desenvolvido ao longo deste estágio permitiu demonstrar o potencial da integração entre dados de Detecção Remota e técnicas de Inteligência Artificial, em particular o *machine learning*, para a monitorização do LULC e para a exploração de possíveis novas abordagens aplicadas ao setor agrícola. Num contexto em que a pressão das alterações climáticas e a necessidade de uma gestão sustentável do território são cada vez maiores, a investigação realizada neste projeto constitui um contributo relevante para a criação de soluções inovadoras neste domínio.

O primeiro projeto incidiu sobre o desenvolvimento de um modelo de classificação e deteção de mudanças no LULC, treinado com dados Sentinel-1 e Sentinel-2 e recorrendo ao algoritmo *Random Forest*. Os resultados obtidos revelaram uma capacidade promissora de diferenciação entre diversas classes, sobretudo nas categorias de maior contraste espectral. Contudo, foram igualmente identificadas limitações significativas, como a dificuldade em lidar com áreas de sombra, frequentemente confundidas com a classe correspondente aos corpos de água, bem como em distinguir diferentes tipos de vegetação, em particular entre *shrubland* e floresta. Estes desafios sublinham a necessidade de melhorar o processo de treino, quer pela introdução de novos polígonos anotados em diferentes contextos climáticos, quer pela integração de variáveis auxiliares, como modelos digitais de elevação ou indicadores de biomassa. No futuro, a transição para metodologias baseadas em *deep learning* poderá potenciar ganhos substanciais de desempenho e generalização.

O segundo projeto focou-se no caso de estudo das vinhas, com a preparação de *datasets* a partir de imagens GEOSAT-2 e respetivo pré-processamento, incluindo a criação de máscaras binárias e a divisão das imagens em *tiles*. Embora o modelo de deteção automática ainda não tenha sido implementado, o trabalho desenvolvido assegurou uma base sólida para as fases seguintes. Para além do cadastro automático de vinhas, perspetivam-se aplicações relevantes na agricultura de precisão.

Em conclusão, o trabalho desenvolvido ao longo do estágio permitiu consolidar uma base robusta para a monitorização do LULC com resultados encorajadores, ao mesmo tempo que assegurou a preparação rigorosa dos dados necessários para a futura deteção automática de vinhas. Embora ambos os projetos se encontrem ainda numa fase preliminar, evidenciam já a viabilidade e relevância das abordagens exploradas. A continuidade deste trabalho, com recurso a metodologias mais avançadas e à

integração de novos dados, poderá contribuir de forma decisiva para o desenvolvimento de ferramentas operacionais de apoio à gestão ambiental e territorial.

Referências Bibliográficas

[1] Di Gregorio, A., & Jansen, L. J. M., “Land Cover Classification System: Classification Concepts and User Manual”, 2000.

<https://www.fao.org/4/y7220e/y7220e06.htm#bm06.2>

[2] Yanzhao Wang, Yonghua Sun, Xuyue Cao, Yihan Wang, Wangkuan Zhang, Xinglu Cheng, “A review of regional and Global scale Land Use/Land Cover (LULC) mapping products generated from satellite remote Sensing”, Vol. 206, 2023, Pages 311-334.

<https://doi.org/10.1016/j.isprsjprs.2023.11.014>.

[3] Maria José Vale; Rui Reis; Bruno M. Meneses, “A caracterização do uso e ocupação do solo de Portugal Continental.”, DGT, 2018, Cap.1.

https://www.dgterritorio.gov.pt/sites/default/files/projetos/LandynL1-Livro_Uso_Ocupacao_Solo_Cap1.pdf

[4] IPCC, *Special Report on Climate Change and Land (2019)*.

<https://www.ipcc.ch/srccl/>

[5] FAO, *Land Statistics 2001–2023*, Food and Agriculture Organization of the United Nations, 2025.

<https://doi.org/10.4060/cd5765en>

[6] K.C. Seto, B. Güneralp, & L.R. Hutyra, “Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools”, Proc. Natl. Acad. Sci. U.S.A. 109 (40) 16083-16088, 2012.

<https://doi.org/10.1073/pnas.1211658109>

[7] Lu, Dengsheng & Mausel, Paul & Brondízio, Eduardo & Moran, Emilio. (2004). Change Detection Techniques. International Journal of Remote Sensing.

https://www.researchgate.net/publication/235245895_Change_Detection_Techniques

[8] NASA EarthData. *Earth Observation Data Basics: Remote Sensing*.

<https://www.earthdata.nasa.gov/learn/earth-observation-data-basics/remote-sensing>

[9] U.S. Geological Survey (USGS). *What is remote sensing and what is it used for?*.

<https://www.usgs.gov/faqs/what-remote-sensing-and-what-it-used>

[10] European Parliament. *What is artificial intelligence and how is it used?* Topics – Artificial intelligence.

<https://www.europarl.europa.eu/topics/en/article/20200827STO85804/what-is-artificial-intelligence-and-how-is-it-used>

[11] Yisa Ginath Yuh, Wiktor Tracz, H. Damon Matthews, Sarah E. Turner, Application of machine learning approaches for land cover monitoring in northern Cameroon, *Ecological Informatics*, Volume 74, 2023, 101955, ISSN 1574-9541.

<https://doi.org/10.1016/j.ecoinf.2022.101955>

[12] Zhao S, Tu K, Ye S, Tang H, Hu Y, Xie C. Land Use and Land Cover Classification Meets Deep Learning: A Review. *Sensors*. 2023; 23(21):8966.

<https://doi.org/10.3390/s23218966>

[13] Vali A, Comai S, Matteucci M. Deep Learning for Land Use and Land Cover Classification Based on Hyperspectral and Multispectral Earth Observation Data: A Review. *Remote Sensing*. 2020; 12(15):2495.

<https://doi.org/10.3390/rs12152495>

[14] Grift, J., Persello, C., & Koeva, M. N. 2023. Cadastral boundary delineation using deep learning and remote sensing imagery: state of the art and future developments. FIG Working Week 2023: Protecting our world, conquering new frontiers

https://fig.net/fig2023/technical_program.htm

[15] Fetai, B., Oštir, K., Kosmatin Fras, M., & Lisec, A. 2019. Extraction of visible boundaries for cadastral mapping based on UAV imagery. *Remote sensing*, 1113, 1510.

https://www.researchgate.net/publication/334044635_Extraction_of_Visible_Boundaries_for_Cadastral_Mapping_Based_on_UAV_Imagery

[16] Sheng J, Sun Y, Huang H, Xu W, Pei H, Zhang W, Wu X. HBRNet: Boundary Enhancement Segmentation Network for Cropland Extraction in High-Resolution Remote Sensing Images. *Agriculture*. 2022; 12(8):1284.

<https://doi.org/10.3390/agriculture12081284>

[17] Wang, J., Ding, N., & He, G. (2023). A boundary enhancement loss function for semantic segmentation of land cover. *International Journal of Remote Sensing*, 44(12), 3637–3659.

<https://doi.org/10.1080/01431161.2023.2224101>

[18] Luo X, Bennett RM, Koeva M, Lemmen C. Investigating Semi-Automated Cadastral Boundaries Extraction from Airborne Laser Scanned Data. *Land*. 2017; 6(3):60.

<https://doi.org/10.3390/land6030060>

[19] Ahmadi, S., Zoej, M. V., Ebadi, H., Moghaddam, H. A., & Mohammadzadeh, A. 2010. Automatic urban building boundary extraction from high resolution aerial images using an innovative model of active contours. *International Journal of Applied Earth Observation and Geoinformation*, 123, 150157

https://www.researchgate.net/publication/220492022_Automatic_urban_building_boundary_extraction_from_high_resolution_aerial_images_using_an_innovative_model_of_active_contours

[20] Talukdar S, Singha P, Mahato S, Shahfahad, Pal S, Liou Y-A, Rahman A. Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations—A Review. *Remote Sensing*. 2020; 12(7):1135.

<https://doi.org/10.3390/rs12071135>

[21] Sharma Banjade S, Rai N, Subedi B. Comparison of Supervised Classification Algorithms Using a Hyperspectral Image for Land Use/Land Cover Classification. *Environmental Sciences Proceedings*. 2024; 29(1):59.

<https://doi.org/10.3390/ECRS2023-16702>

[22] Yisa Ginath Yuh, Wiktor Tracz, H. Damon Matthews, Sarah E. Turner, Application of machine learning approaches for land cover monitoring in northern Cameroon, *Ecological Informatics*, Volume 74, 2023, 101955, ISSN 1574-9541.

<https://doi.org/10.1016/j.ecoinf.2022.101955>.

[23] Getachew Bayable, Getie Gebrie, Tadele Melese, Alebel Melaku, Land use/cover classification using machine learning algorithms and their impacts on land surface temperature and soil moisture in the Alawuha Watershed, Ethiopia, *Environmental and Sustainability Indicators*, Volume 27, 2025, 100797, ISSN 2665-9727.

<https://doi.org/10.1016/j.indic.2025.100797>.

[24] Ouchra, Hafsa & Belangour, Abdessamad & Erraissi, Allae. (2024). Supervised Machine Learning Algorithms for Land Cover Classification in Casablanca, Morocco. *Ingénierie des systèmes d information*.

https://www.researchgate.net/publication/378519468_Supervised_Machine_Learning_Algorithms_for_Land_Cover_Classification_in_Casablanca_Morocco

[25] Ganji Tejasree, L. Agilandeewari, Land use/land cover (LULC) classification using deep-LSTM for hyperspectral images, *The Egyptian Journal of Remote Sensing and Space Sciences*, Volume 27, Issue 1, 2024, Pages 52-68, ISSN 1110-9823.

<https://doi.org/10.1016/j.ejrs.2024.01.004>.

[26] Hao, Mengmeng & Dong, Xiaohan & Jiang, Dong & Yu, Xianwen & Ding, Fangyu & Zhuo, Jun. (2024). Land-use classification based on high-resolution remote sensing imagery and deep learning models.

https://www.researchgate.net/publication/379920472_Land-use_classification_based_on_high-resolution_remote_sensing_imagery_and_deep_learning_models

[27] Zhao S, Tu K, Ye S, Tang H, Hu Y, Xie C. Land Use and Land Cover Classification Meets Deep Learning: A Review. *Sensors*. 2023; 23(21):8966.

<https://doi.org/10.3390/s23218966>

[28] Boonpook W, Tan Y, Nardkulpat A, Torsri K, Torteeka P, Kamsing P, Sawangwit U, Pena J, Jainaen M. Deep Learning Semantic Segmentation for Land Use and Land Cover Types Using Landsat 8 Imagery. *ISPRS International Journal of Geo-Information*. 2023; 12(1):14.

<https://doi.org/10.3390/ijgi12010014>

[29] Breiman, L. Random Forests. *Machine Learning* **45**, 5–32 (2001).

<https://doi.org/10.1023/A:1010933404324>

[30] Beck, H.E., N.E. Zimmermann, T.R. McVicar, N. Vergopolan, A. Berg, E.F. Wood: Present and future Köppen-Geiger climate classification maps at 1-km resolution, *Scientific Data* 5:180214, doi:10.1038/sdata.2018.214 (2018).

<https://doi.org/10.1038/sdata.2018.214>

[31] Copernicus, Sentinel Mission

<https://sentiwiki.copernicus.eu/web/sentinel-1>

[32] ESA, Copernicus Access Hub

<https://www.copernicus.eu/en/access-data/conventional-data-access-hubs>

[33] Repositório PyroSAR

<https://github.com/johntruckenbrodt/pyroSAR>

[34] Y. J. Kaufman and D. Tanre, "Atmospherically resistant vegetation index (ARVI) for EOS-MODIS," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 30, no. 2, pp. 261-270, March 1992

[10.1109/36.134076](https://doi.org/10.1109/36.134076)

[35] Feyisa, G. L., Meilby, H., Fensholt, R., & Proud, S. R. (2014). Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery. *Remote Sensing of Environment*, 140, 23–35.

<https://doi.org/10.1016/j.rse.2013.08.029>

[36] García, Mariano & Chuvieco, Emilio. (2004). Assessment of the potential of SAC-C/MMRS imagery for mapping burned areas in Spain. *Remote Sensing of Environment - REMOTE SENS ENVIRON*. 92. 10.1016/j.rse.2004.04.011.

https://www.researchgate.net/publication/228860943_Assessment_of_the_potential_of_SAC-CMMRS_imagery_for_mapping_burned_areas_in_Spain

[37] Filipponi, F. (2018). BAIS2: Burned Area Index for Sentinel-2. *Proceedings*, 2(7), 364. <https://doi.org/10.3390/ecrs-2-05177>

[38] Zha, Y., Gao, J. and Ni, S. (2003) Use of Normalized Difference Built-Up Index in Automatically Mapping Urban Areas from TM Imagery. *International Journal of Remote Sensing*, 24, 583-594.

<https://doi.org/10.1080/01431160304987>

[39] Anatoly A. Gitelson, Yuri Gritz †, Mark N. Merzlyak, Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves, *Journal of Plant Physiology*, Volume 160, Issue 3, 2003, Pages 271-282, ISSN 0176-1617

<https://doi.org/10.1078/0176-1617-00887>.

[40] Huete, A. R., et al. (2002). Overview of the radiometric and biophysical performance of MODIS vegetation indices. *Remote Sensing of Environment*, 83(1–2), 195–213.

[https://doi.org/10.1016/S0034-4257\(02\)00096-2](https://doi.org/10.1016/S0034-4257(02)00096-2)

[41] Chaoyang Wu, Zheng Niu, Shuai Gao, The potential of the satellite derived green chlorophyll index for estimating midday light use efficiency in maize, coniferous forest and grassland, *Ecological Indicators*, Volume 14, Issue 1, 2012, Pages 66-73, ISSN 1470-160X

<https://doi.org/10.1016/j.ecolind.2011.08.018>.

[42] Pinty, Bernard & Verstraete, Michel. (1992). GEMI: A non-linear index to monitor global vegetation from satellites. *Vegetatio*. 101. 15-20. 10.1007/BF00031911.

https://www.researchgate.net/publication/226985098_GEMI_A_non-linear_index_to_monitor_global_vegetation_from_satellites

[43] Mangewa, L. J., Ndakidemi, P. A., Alward, R. D., Kija, H. K., Bukombe, J. K., Nasolwa, E. R., & Munishi, L. K. (2022). Comparative Assessment of UAV and Sentinel-2 NDVI and GNDVI for Preliminary Diagnosis of Habitat Conditions in Burunge Wildlife Management Area, Tanzania

<https://doi.org/10.3390/earth3030044>

[44] Chen, J. M., & Black, T. A. (1992). Defining leaf area index for non-flat leaves. *Plant, Cell & Environment*, 15(4), 421–429.

<https://doi.org/10.1111/j.1365-3040.1992.tb00992.x>

[45] J. Qi, A. Chehbouni, A.R. Huete, Y.H. Kerr, S. Sorooshian, A modified soil adjusted vegetation index, *Remote Sensing of Environment*, Volume 48, Issue 2, 1994, Pages 119-126, ISSN 0034-4257

[https://doi.org/10.1016/0034-4257\(94\)90134-1](https://doi.org/10.1016/0034-4257(94)90134-1).

[46] Xing, N., Huang, W., Xie, Q., Shi, Y., Ye, H., Dong, Y., Wu, M., Sun, G., & Jiao, Q. (2020). A Transformed Triangular Vegetation Index for Estimating Winter Wheat Leaf Area Index. *Remote Sensing*, 12(1), 16.

<https://doi.org/10.3390/rs12010016>

[47] Key, Carl & Benson, Nate. (2006). Landscape Assessment: Ground measure of severity, the Composite Burn Index; and Remote sensing of severity, the Normalized Burn Ratio.

https://www.researchgate.net/publication/241687027_Landscape_Assessment_Ground_measure_of_severity_the_Composite_Burn_Index_and_Remote_sensing_of_severity_the_Normalized_Burn_Ratio

[48] Zha, Y., Gao, J. and Ni, S. (2003) Use of Normalized Difference Built-Up Index in Automatically Mapping Urban Areas from TM Imagery. International Journal of Remote Sensing, <https://doi.org/10.1080/01431160304987>

[49] Gao, B. C. (1996). NDMI – Normalized difference moisture index for monitoring vegetation water content. *Remote Sensing of Environment*, 58(3), 257–266.

[https://doi.org/10.1016/S0034-4257\(96\)00067-3](https://doi.org/10.1016/S0034-4257(96)00067-3)

[50] Buschmann, C. and Nagel, E. (1993) In Vivo Spectroscopy and Internal Optics of Leaves as Basis for Remote Sensing of Vegetation. International Journal of Remote Sensing, <http://dx.doi.org/10.1080/01431169308904370>

[51] Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1974). Monitoring vegetation systems in the Great Plains with ERTS. *NASA/GSFC, Final Report*.

<https://ntrs.nasa.gov/citations/19740022614>

[52] McFeeters, S.K. (1996) The Use of the Normalized Difference Water Index (NDWI) in the Delineation of Open Water Features. International Journal of Remote Sensing, 17, 1425-1432, <http://dx.doi.org/10.1080/01431169608948714>

[53] Rondeaux, G., Steven, M., & Baret, F. (1996). Optimization of soil-adjusted vegetation indices. *Remote Sensing of Environment*, 55(2), 95–107.

[https://doi.org/10.1016/0034-4257\(95\)00186-7](https://doi.org/10.1016/0034-4257(95)00186-7)

[54] Barron, V. and Torrent, J. (1986), Use of the Kubelka-Munk theory to study the influence of iron oxides on soil colour. *Journal of Soil Science*, 37: 499-510.

[55] Lei Ji, Albert J. Peters, Performance evaluation of spectral vegetation indices using a statistical sensitivity function, *Remote Sensing of Environment*, Volume 106, Issue 1, 2007, Pages 59-65, ISSN 0034-4257, <https://doi.org/10.1016/j.rse.2006.07.010>.

- [56] Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, 25(3), 295–309. [https://doi.org/10.1016/0034-4257\(88\)90106-X](https://doi.org/10.1016/0034-4257(88)90106-X)
- [57] Peñuelas, J., Baret, F., & Filella, I. (1995). Semi-empirical indices to assess carotenoids/chlorophyll a ratio from leaf spectral reflectance. *Photosynthetica*, 31(2), 221–230
- [58] Blank, Veronika, Roman Skidanov, Leonid Doskolovich, and Nikolay Kazanskiy. 2021. "Spectral Diffractive Lenses for Measuring a Modified Red Edge Simple Ratio Index and a Water Band Index" <https://doi.org/10.3390/s21227694>
- [59] Gitelson, A., Kaufman, Y. J., Stark, R., and Rundquist, D., 2002, Novel algorithms for remote estimation of vegetation fraction. *Remote Sensing of Environment*, 80, 76–87. [https://doi.org/10.1016/S0034-4257\(01\)00289-9](https://doi.org/10.1016/S0034-4257(01)00289-9)
- [60] Y. Kim, T. Jackson, R. Bindlish, H. Lee and S. Hong, "Radar Vegetation Index for Estimating the Vegetation Water Content of Rice and Soybean," in *IEEE Geoscience and Remote Sensing Letters*, vol. 9, no. 4, pp. 564-568, July 2012 doi: 10.1109/LGRS.2011.2174772
- [61] SIOSE Dataset, Sistema de Información de Ocupación del Suelo de España
<https://www.siose.es/web/guest/descripcion-ar>
- [62] COS Dataset, Direção Geral do Território
<https://dados.gov.pt/pt/datasets/carta-de-uso-e-ocupacao-do-solo-2018/>
- [63] QGIS <https://qgis.org/>
- [64] Comissão de Viticultura da Região dos Vinhos Verdes, CVRVV
<https://portal.vinhoverde.pt/pt/>
- [65] Mapa de Ruído do Concelho de Arcos de Valdevez, APA
https://apambiente.pt/sites/default/files/Ar_Ruido/Ruido/SituacaoNacional/MapasRuidoMunicipais/Arcos_de_Valdevez_2022_RNT.pdf
- [66] Instituto da Vinha e do Vinho, IVV
<https://www.ivv.gov.pt/np4/home.html>
- [67] GEOSAT
<https://geosat.space/pt/>