

Mestrado em
**Ciência e Sistemas
de Informação Geográfica**

Especialização em
Ciência e Sistemas de Informação Geográfica

**Metodologia SIG para
Previsão da Qualidade
da Cortiça com Machine
Learning**

Beatriz Mariano Veloso Seco Fonseca

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

Metodologia SIG para Previsão da Qualidade da Cortiça com *Machine Learning*

por

Beatriz Mariano Veloso Seco Fonseca

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre em Ciência e Sistemas de Informação Geográfica, com especialização em Ciência e Sistemas de Informação Geográfica.

Orientador:

Professor André Duarte, Escola Superior Agrária de Coimbra

Coorientador:

Dr. Marco Octávio Trindade Painho, NOVA Information Management School

Julho, 2026



Declaração de Integridade

Declaro ter realizado o presente trabalho académico com integridade. Confirmando que não recorri à prática de plágio ou de qualquer outra forma de utilização indevida de informação ou de falsificação de resultados durante o processo de elaboração deste trabalho. Declaro ainda que tenho conhecimento das Regras de Conduta e do Código de Honra da NOVA Information Management School.

Lisboa, 15 de julho de 2026

Beatriz Fonseca

UTILIZAÇÃO DE INTELIGÊNCIA ARTIFICIAL GENERATIVA

Tarefas	NÃO	SIM	Ferramentas de Inteligência Artificial Generativa
Compreender melhor as questões relacionadas com a investigação		x	Gemini/ChatGPT
Resumir texto de bibliografia / recursos		x	Notebook KML
Resumo do(s) método(s) utilizado(s)	x		
Tradução de texto		x	DeepL Translate
Verificação gramatical		x	Gemini/ChatGPT
Parafrasear ou reescrever texto de outras pessoas/recursos		x	Gemini/ChatGPT
Codificação em R, Python, etc.	x		
Obter ajuda sobre um software		x	Gemini/ChatGPT
Criação e edição de imagens, mapas, vídeos, etc.	x		
Análise de dados		x	Gemini/ChatGPT
Especificar abaixo outras tarefas não mencionadas acima:	x		



DEDICATÓRIA

Dedico este trabalho aos meus pais, Catarina Veloso e Luís Fonseca e avós, Lúcia Veloso e António Veloso, pelo amor e apoio incondicional.



AGRADECIMENTOS

A minha mais profunda gratidão aos meus pais e avós, pelo apoio incondicional, pela paciência e perseverança constantes. Foram eles que me transmitiram os valores do trabalho, da dedicação e da crença de que, com esforço e determinação, é possível alcançar qualquer objetivo.

Quero igualmente agradecer ao Professor José Gaspar docente da Escola Superior Agrária de Coimbra, por me ter incentivado e motivado a ingressar neste mestrado. Sem o seu apoio e confiança, este percurso não teria sido possível.

Ao meu orientador, o Professor André Duarte, manifesto o meu sincero agradecimento pela orientação científica, pelo acompanhamento atento e pela disponibilidade ao longo de todo o desenvolvimento desta dissertação.

Manifesto ainda o meu agradecimento ao meu colega de curso e agora amigo, Paulo Fernandes, cuja presença foi fundamental nos momentos de maior dificuldade e desmotivação. O seu apoio constante, otimismo e amizade foram uma verdadeira força estabilizadora. Para além disso, partilhámos um percurso de crescimento académico desde o primeiro semestre, através de trabalhos conjuntos e troca de conhecimentos, o que enriqueceu significativamente a minha formação.

Por fim, agradeço à NOVA Information Management School – Instituto Superior de Estatística e Gestão de Informação, da Universidade NOVA de Lisboa, por me ter proporcionado uma formação sólida e exigente, que contribuiu de forma decisiva para o desenvolvimento das minhas competências académicas e pessoais. Agradeço igualmente a todos os docentes pelo rigor académico e pela exigência, que me fizeram reconhecer o meu próprio potencial e acreditar nas minhas capacidades.



RESUMO

O sobreiro (*Quercus suber*) é uma espécie fundamental para a sustentabilidade ecológica e económica de Portugal, enfrentando, todavia, ameaças crescentes devido a problemas fitossanitários, que em conjugação com os efeitos das alterações climáticas, conduzem ao declínio da espécie. Esta dissertação propõe uma metodologia baseada em Sistemas de Informação Geográfica (SIG) e técnicas de *Machine Learning* para a previsão do estado fitossanitário e da qualidade potencial da cortiça. O estudo incidu sobre uma área no concelho da Chamusca, onde foram recolhidos dados de inventário de 137 sobreiros, integrando variáveis dendrométricas medidas *in situ* com dados provenientes de deteção remota e variáveis biofísicas do terreno. A metodologia envolveu a utilização de dados LiDAR (*Light Detection and Ranging*) aerotransportados da Direção-Geral do Território para a caracterização estrutural e imagens multiespectrais do satélite Sentinel-2 para o cálculo de diversos índices de vegetação. Foram testadas duas abordagens de modelação na plataforma KNIME, o modelo completo, que integrou todas as variáveis, e o modelo de satélite, focado exclusivamente em variáveis espectrais e as biofísicas do terreno. Para a classificação, aplicaram-se os algoritmos *Random Forest* (RF), *Support Vector Machine* (SVM) e o *Multilayer Perceptron* (RProp MLP). Os resultados revelaram que, no Modelo completo, o SVM apresentou o melhor desempenho com uma exatidão de 81%. Notavelmente, no Modelo Satélite, o algoritmo RProp MLP superou os restantes, alcançando uma exatidão 83% e um índice Kappa de 67%. A análise demonstrou que os índices espectrais, nomeadamente o CI, CVI, LAI e GNDVI, são os principais determinantes do diagnóstico fitossanitário, enquanto as métricas dendrométricas individuais têm um contributo marginal. O estudo mostrou que a integração de tecnologias de Sistemas de Informação Geográfica (SIG) e algoritmos de *Machine Learning* para o diagnóstico fitossanitário não invasivo dos montados de sobreiro, permite explicar, de forma robusta, o estado fitossanitário das árvores com base em índices de vegetação obtidos a partir de imagens de satélite Sentinel-2, reduzindo a necessidade de realização de inventário florestal tradicional.

PALAVRAS-CHAVE

Quercus suber; *Machine Learning*; SIG; Deteção Remota; Índices de Vegetação; Estado Fitossanitário.



ABSTRACT

The cork oak (*Quercus suber*) is a fundamental species for the ecological and economic sustainability of Portugal. However, it faces increasing threats due to phytosanitary problems which, together with the effects of climate change, lead to the decline of the species. This dissertation proposes a methodology based on Geographic Information Systems (GIS) and Machine Learning techniques for predicting phytosanitary status and potential cork quality. The study focused on an area in the municipality of Chamusca, where inventory data were collected from 137 cork oak trees, integrating dendrometric variables measured *in situ* with data obtained from remote sensing and terrain biophysical variables. The methodology involved the use of airborne LiDAR (Light Detection and Ranging) data from the Portuguese Directorate-General for Territory (Direção-Geral do Território) for structural characterization and multispectral Sentinel-2 satellite imagery for the calculation of several vegetation indices. Two modelling approaches were tested in the KNIME platform: the Full Model, which integrated all variables, and the Satellite Model, focused exclusively on spectral and terrain biophysical variables. For the classification, the Random Forest (RF), Support Vector Machine (SVM), and Multilayer Perceptron (RProp MLP) algorithms were applied. The results revealed that, in the Full Model, the SVM achieved the best performance, with an accuracy of 81%. Notably, in the Satellite Model, the RProp MLP algorithm outperformed the others, achieving an accuracy of 83% and a Kappa index of 67%. The analysis demonstrated that the spectral indices, namely CI, CVI, LAI, and GNDVI, are the main determinants of phytosanitary diagnosis, while the individual dendrometric metrics have only a marginal contribution. The study showed that the integration of Geographic Information Systems (GIS) technologies and Machine Learning algorithms for the non-invasive phytosanitary diagnosis of cork oak stands makes it possible to explain, in a robust manner, the phytosanitary status of trees based on vegetation indices obtained from Sentinel-2 satellite imagery, reducing the need for carrying out traditional forest inventories.

PALAVRAS-CHAVE

Quercus suber; Machine Learning; GIS; Remote Sensing; Vegetation Indices; Phytosanitary Status.

Objetivos do Desenvolvimento Sustentável (ODS):





ÍNDICE

1. Introdução	9
1.1. Objetivo da investigação	10
2. Revisão da Literatura	11
2.1. Enquadramento Geral	11
2.2. Importância Económica do sobreiro	13
2.3. Problemas fitossanitários do sobreiro	15
2.4. LiDAR no Inventário Florestal	20
2.5. Índices de Vegetação	25
3. Metodologia	27
3.1. Área de Estudo	28
3.2. Recolha de Dados	30
3.2.1. Inventário Florestal	31
3.2.2. Dados LiDAR	35
3.2.3. Imagens Sentinel- 2	37
3.2.3. Detecção Remota	39
3.3. Processamento dos Dados no KNIME	41
3.3.1 Importação e pré processamento dos dados	41
3.3.2 Partição da Amostra (treino e validação)	44
3.3.3 Definição das Abordagens de Modelação e Algoritmos Aplicados	44
3.3.4 Avaliação do Desempenho Preditivo dos Modelos	45
4. Resultados	47
4.1. Análise da Correlação entre variáveis para os modelos	47
4.2. Análise Exploratória das Variáveis	50
4.3. Modelos de Aprendizagem Máquina (<i>Machine Learning</i>)	51
4.3.1. Abordagem 1: Modelo Completo	51
4.3.2. Abordagem 2: Modelo Satélite	54
5. Discussão	56
6. Conclusões e Investigação Futura	58
Referências Bibliográficas	60



ÍNDICE DE FIGURAS

Figura 2.1- Distribuição de Área de Sobreiro em Portugal com dados da Direção-Geral do Território (DGT), (COS 2023).	12
Figura 2.2 - Ocupação do solo de Florestas de sobreiro em Portugal (IFN6, 2015). 13	
Figura 2.3 - Balança Comercial dos Produtos florestais, 2023 e 2024 (INE 2025)... 14	
Figura 2.4 - Cortiça com massas heterogéneas VS Massas Homogéneas, respetivamente (UNAC, 2021).	15
Figura 2.5 - <i>Biscogniauxia mediterranea</i> em <i>Quercus rubra</i> : A; B) superfície do estroma carbonáceo; C; D) ascósporos imaturos e maduros (Georgieva et al., 2024).	17
Figura 2.6 - Sintomas observados em sobreiro causados pela doença do cancro <i>Diplodia cortícola</i> (Muñoz-Adalia et al., 2022).	18
Figura 2.7 – <i>Coroebus undatus</i> (Observation.org, 2026).	18
Figura 2.8 – Inseto <i>Platypus cylindrus</i> macho e fêmea (Muñoz-Adalia et al., 2022) 19	
Figura 2.9 - Esquema do procedimento de estimativa da altura das árvores: (A) individualização das árvores na nuvem de pontos; (B) determinação da altura em árvores verticais após filtragem do ruído; e (C) aplicação do método em árvores inclinadas C.....	23
Figura 2.10- Fotografia de uma sessão de demonstração da tecnologia <i>Mobile Laser Scanning</i> (MLS).....	24
Figura 3.1 – Diagrama do Fluxo de Trabalho.	27
Figura 3.2 - Localização da Área de Estudo.....	28
Figura 3.3 - Carta da Altimetria na Área de Estudo.	29
Figura 3.4 - Carta de Declive da Área de Estudo.	29
Figura 3.5 - Carta da Exposição da Área de Estudo.	30
Figura 3.6 - Medidas dendrométricas do sobreiro. Foto retirada de Tomé, Barreiro e Soares (2019).....	32
Figura 3.7 - Formulário digital na aplicação ArcGIS <i>Field Maps</i> utilizado para a recolha de dados de inventário florestal.....	33
Figura 3.8 - Sobreiro com presença de carvão do entrecasco, observado durante os inventários florestais.....	34
Figura 3.9 - Sobreiro com o entrecasco com cobrilha.	34
Figura 3.10 - Representação da Contagem de Árvores pelo Estado Fitossanitário . 35	
Figura 3.11 - Raster nDSM.....	36
Figura 3.12 - Resultado da segmentação das copas e correspondência com os pontos do inventário florestal.	37
Figura 3.13 - Ferramentas de cálculo de índices espectrais disponíveis no SNAP.. 39	
Figura 3.14 - Metodologia realizada no KNIME	41



Figura 3.15 – Diagrama do fluxo de trabalho em KNIME.	42
Figura 3.16 - <i>Box plot</i> das variáveis utilizadas no estudo para identificação de valores	43
Figura 3.17 - Fluxo de Trabalho no KNIME do Modelo Satélite.....	44
Figura 4.1 - Estado Fitossanitários das Árvores.	53
Figura 4.2 - Resultado da Previsão do Modelo SVM.....	53

ÍNDICE DE TABELAS

Tabela 3.1 - Base de dados da investigação.....	31
Tabela 3.2 - Imagens Satélite Sentinel-2 usadas para o processamento.....	39
Tabela 3.3 - Fórmulas dos Índices de Vegetação realizados para a investigação. ...	40
Tabela 4.1 - Variáveis do Estudo.....	47
Tabela 4.2 - Matriz de Correlação dos índices de vegetação.....	48
Tabela 4.3 - VIF obtido do GLR.....	51
Tabela 4.4 - Métricas de Desempenho preditivo dos algoritmos do Modelo Completo.	52
Tabela 4.5 - Métricas de Desempenho preditivo dos algoritmos no Modelo Satélite.	55



1. INTRODUÇÃO

O sobreiro (*Quercus suber* L.) constitui uma das espécies florestais de maior relevância ecológica, económica e social da região mediterrânica ocidental, apresentando uma distribuição natural concentrada em Portugal, Espanha e Norte de África. Em Portugal, esta espécie assume particular importância, não apenas por sustentar a liderança mundial na produção de cortiça (Mata & Dos-Santos, 2024), mas também por integrar os ecossistemas de montado, sistemas agroflorestais de elevado valor ambiental reconhecidos pela sua elevada biodiversidade e pela capacidade de fornecer múltiplos bens e serviços dos ecossistemas (Lopes-Fernandes et al., 2024), incluindo sequestro de carbono, conservação do solo, regulação hidrológica e suporte às atividades agro-silvo-pastoris (Aroson et al., 2004).

Entre os produtos associados ao sobreiro, a cortiça destaca-se pelo seu elevado valor económico e pelas propriedades físicas únicas que justificam a sua utilização em diversos setores industriais, particularmente na produção de rolhas naturais de elevada qualidade. Consequentemente, a qualidade da cortiça constitui um dos principais fatores de valorização económica dos montados de sobreiro. Contudo, esta apenas pode ser avaliada diretamente após o processo de descortiçamento e respetiva classificação industrial, procedimento que implica elevados custos operacionais e que está condicionado pelo longo ciclo produtivo da espécie, realizado aproximadamente de nove em nove anos (Mata & Dos-Santos, 2024). A qualidade da cortiça, constitui, por isso, um fator determinante para a valorização económica dos povoamentos de sobreiro.

Esta limitação evidencia a necessidade de desenvolver metodologias capazes de antecipar, de forma indireta, o potencial produtivo das árvores antes da extração da cortiça. Embora a qualidade da cortiça dependa de diversos fatores genéticos, ambientais e de gestão, o estado fitossanitário do sobreiro constitui um importante indicador da sua condição fisiológica e do seu vigor vegetativo. De um modo geral, árvores que apresentam melhores condições de saúde tendem a possuir maior capacidade de crescimento e, consequentemente, maior potencial para produzir cortiça de melhor qualidade (Díez-Méndez et al., 2026). Assim, a monitorização do estado fitossanitário assume um papel fundamental na gestão sustentável dos montados, permitindo identificar precocemente situações de declínio e apoiar a tomada de decisão antes da realização do descortiçamento.

Tradicionalmente, a avaliação do estado fitossanitário dos sobreiros baseia-se em inventários florestais realizados em campo. Apesar da elevada fiabilidade destes métodos, a sua aplicação é limitada pelo tempo necessário na recolha de dados. Assim, torna-se fundamental o desenvolvimento de metodologias complementares que permitam uma monitorização mais eficiente e espacialmente contínua.

Neste contexto, a deteção remota tem vindo a afirmar-se como uma ferramenta de elevado potencial para a monitorização dos ecossistemas florestais, permitindo



adquirir informação multitemporal em grande extensões territoriais de forma rápida e economicamente mais eficiente do que os métodos exclusivamente baseados em trabalho de campo (Fassnacht et al., 2024). A utilização de imagens multiespectrais, como as da missão Sentinel-2, bem como de dados LiDAR, permite caracterizar o vigor da vegetação, a estrutura do coberto arbóreo e outros fatores ambientais relevantes para o desenvolvimento dos povoamentos (Bambagioni et al., 2025). Apesar destes avanços, a integração de deteção remota, Sistemas de Informação Geográfica (SIG) e técnicas de *Machine Learning* aplicada aos montados de sobreiro ainda é relativamente pouco explorada (Bambagioni et al. 2025). A maioria dos estudos tem-se centrado na caracterização físico-química da cortiça ou na avaliação isolada do estado fitossanitário das árvores, persistindo uma lacuna no desenvolvimento de metodologias preditivas capazes de integrar múltiplas fontes de informação espacial, tal como evidenciado por Navarro et al., 2019.

Em resposta a estes desafios, os algoritmos de *Machine Learning* têm demonstrado elevada capacidade para modelar fenómenos ambientais complexos, permitindo identificar relações entre variáveis biofísicas e espaciais e detetar padrões associados ao vigor e ao declínio da vegetação. Modelos como o *Random Forest* (RF) e o *Support Vector machine* (SVM) têm sido aplicados com sucesso na monitorização da saúde florestal e na identificação de fatores relacionados com o stress hídrico e problemas fitossanitários (Sebastiani et al., 2024; Lahssini et al., 2015).

1.1. Objetivo da investigação

O presente estudo teve como objetivo principal desenvolver uma metodologia baseada em Sistemas de Informação Geográfica (SIG) e técnicas de *Machine Learning*, para a previsão indireta da qualidade potencial da cortiça através da modelação do estado fitossanitário dos sobreiros. Para esse efeito, foram integradas múltiplas fontes de informação espacial e ambiental:

- Inventário florestal tradicional com avaliação do estado fitossanitário e localização de indivíduos com recurso a tecnologia GNSS-RTK, garantindo elevada precisão geoespacial;
- Imagens multiespectrais Sentinel-2 do Programa *Copernicus* da Agência Espacial Europeia, para o cálculo dos índices de vegetação;
- Dados LiDAR aerotransportados disponibilizados pela Direção-Geral do Território (DGT), para a caracterização estrutural do coberto vegetal;
- Modelação preditiva com algoritmos de *Machine Learning* com o objetivo de identificar padrões capazes de distinguir indivíduos em bom estado vegetativo de indivíduos em declínio.

Pretende-se, desta forma, contribuir para o desenvolvimento de metodologias de monitorização mais eficientes, não invasivas e espacialmente contínuas dos montados de sobreiro, bem como para a avaliação indireta da qualidade potencial da cortiça.



2. REVISÃO DA LITERATURA

2.1. Enquadramento Geral

O sobreiro (*Quercus suber* L.) é uma árvore de folha persistente que pertence à família das Fagáceas e ao género *Quercus*. Este género possui mais de 600 espécies distintas dispersas em todo o mundo. Segundo vários estudos, o sobreiro é uma árvore muito antiga, existindo desde o período Terciário, uma altura em que o clima era mais quente e húmido do que o atual, na Península Ibérica. Ao longo das eras o clima tem mudado para o padrão mediterrânico que temos hoje, onde os verões são secos e os invernos muito chuvosos e húmidos. O sobreiro conseguiu adaptar-se bem a estas alterações climáticas, mostrando uma grande capacidade de resistência (Silva, J.S., 2007).

A distribuição geográfica do sobreiro hoje em dia, é bastante restrita, em comparação com outras espécies do mesmo género *Quercus* (Kim et al., 2017). É uma árvore que pode viver até muitos anos, podendo chegar aos 200 a 205 anos, atingindo facilmente os 25 metros e mais de 2 metros de diâmetro à altura do peito (Silva et al., 2023).

O sobreiro caracteriza-se pela sua capacidade de se adaptar a condições de acesso ilimitado de água a grandes profundidades durante quase todo o ano, principalmente devido ao seu sistema radicular composto por uma raiz profunda e ramificações laterais que se expandem horizontalmente que permitem uma maior absorção de todos os nutrientes e água (Silva, J.S., 2007), mas mesmo assim, o sobreiro sofre com a falta de água prolongada e com o calor extremo dos verões. Estes fatores, agravados pelas alterações climáticas, prejudicam o crescimento e a qualidade da cortiça, tornando-a mais densa e fazendo que os anéis de crescimento fiquem mais estreitos (Costa et al., 2022). Apesar deste problema, a espécie está classificada como “Pouco Preocupante” na Lista Vermelha da IUCN - União Internacional para a Conservação da Natureza e dos Recursos Naturais das espécies ameaçadas.

Em Portugal, o sobreiro está espalhado de norte a sul. No norte do país, é normal vê-lo misturado com carvalhos, nas zonas perto do mar, aparece muitas vezes ao lado de pinheiros, e no sul com azinheira. Nas zonas de serra é menos comum, já que não se dá tão bem com o frio (Silva, J.S., 2007).

Em suma, esta espécie assume um papel vital na floresta nacional, uma vez que Portugal detém mais de um terço da área mundial e lidera as exportações globais de cortiça (Silva, J.S., 2007). Trata-se de um recurso com uma vasta gama de aplicações, sendo utilizado tanto no fabrico de rolhas e aglomerados como no desenvolvimento de fragrâncias e produtos farmacêuticos (Marcos-Romero et al., 2025).



De acordo com a Carta de Uso e Ocupação do Solo (COS) 2023, a floresta de sobreiro ocupa uma área de 742 290,53 hectares em Portugal, incluindo áreas classificadas como sistemas silvo-pastoris e zonas de pastagem arborizadas. A distribuição espacial destes povoamentos é representada na Figura 2.1. A COS constitui a principal fonte cartográfica nacional para caracterizar a distribuição espacial do sobreiro. A interpretação destes valores deve ter em consideração a metodologia utilizada na sua produção, baseada na fotointerpretação e digitalização manual de imagens aéreas e de satélite. Esta abordagem permite uma representação detalhada da ocupação do solo, mas não fornece informação sobre a estrutura interna dos povoamentos nem sobre a sua evolução biológica.

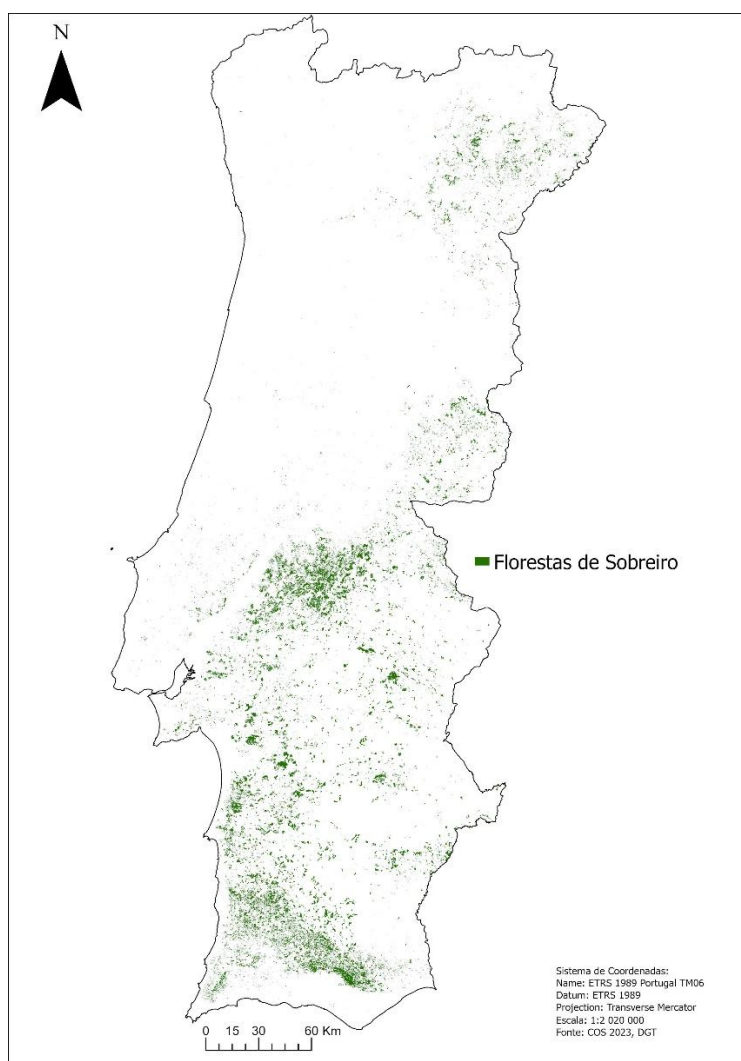


Figura 2.1- Distribuição de Área de Sobreiro em Portugal com dados da Direção-Geral do Território (DGT), (COS 2023).

Em contraste, o IFN6 – Inventário Florestal Nacional 6, assenta numa metodologia de amostragem por parcelas distribuídas sistematicamente pelo território nacional, produzindo estimativas estatísticas da área florestal e dos atributos estruturais dos povoamentos. Esta diferença metodológica limita a comparabilidade direta entre as duas fontes, mas torna o IFN6 particularmente relevante para análises temporais. Os



dados harmonizados do IFN6 indicam uma tendência de redução da área de sobreiro ao longo das últimas décadas (Figura 2.2).

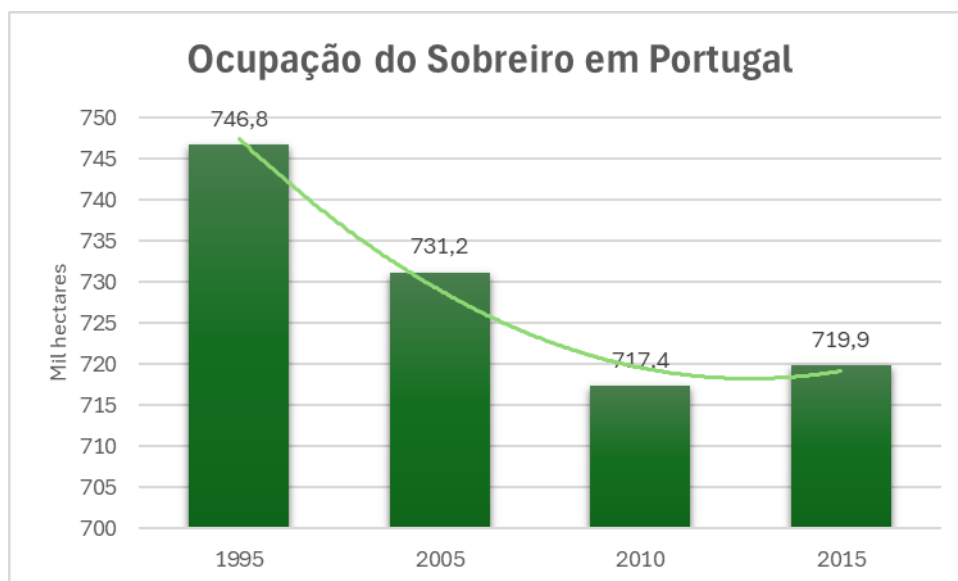


Figura 2.2 - Ocupação do solo de Florestas de sobreiro em Portugal (IFN6, 2015).

2.2. Importância Económica do sobreiro

Ao contrário da maioria das espécies florestais, o sobreiro possui uma casca espessa de cortiça que pode ser removida periodicamente do tronco e dos ramos sem comprometer a sobrevivência da árvore. Após o descortiçamento, forma-se uma nova camada de cortiça que cresce continuamente em espessura, permitindo a exploração sustentável deste recurso ao longo de várias décadas. Esta capacidade de recuperação tem sido aproveitada desde em tempos antigos e constitui, ainda hoje, a base da importância económica do sobreiro e dos montados de sobreiro (Silva, J.S., 2007).

A cortiça representa um produto florestal de elevado valor económico, sendo a sua qualidade um fator determinante para a valorização da produção. Neste contexto, Portugal destaca-se como líder mundial no setor da cortiça. De acordo com a Associação Portuguesa da Cortiça (APCOR), as exportações atingiram 1.232 milhões de euros em 2023, representando um novo recorde para o setor, sendo as rolhas de cortiça o principal produto exportado (Boletim Estatístico – Mercados da Cortiça, 2023).

Segundo o Instituto Nacional de Estatística (2023), a produção nacional de cortiça revelou-se insuficiente para satisfazer as necessidades da indústria transformadora, nomeadamente na produção de rolhas e outros produtos derivados, conduzindo a níveis significativos de importação. Esta situação evidencia a importância de promover estratégias que permitam aumentar a produtividade e a qualidade da cortiça produzida em território nacional, contribuindo para reduzir a dependência externa e reforçar a competitividade do setor.



De acordo com a ficha técnica das Estatísticas Agrícolas de 2024, as exportações de cortiça continuaram a representar um dos principais contributos da fileira florestal para a economia portuguesa. Nesse ano, cerca de 50 mil toneladas de rolhas geraram aproximadamente 835 milhões de euros, colocando a cortiça entre os produtos florestais com maior valor de exportação (Figura 2.3). Estes resultados demonstram a elevada capacidade produtiva e a relevância económica do setor corticeiro em Portugal.

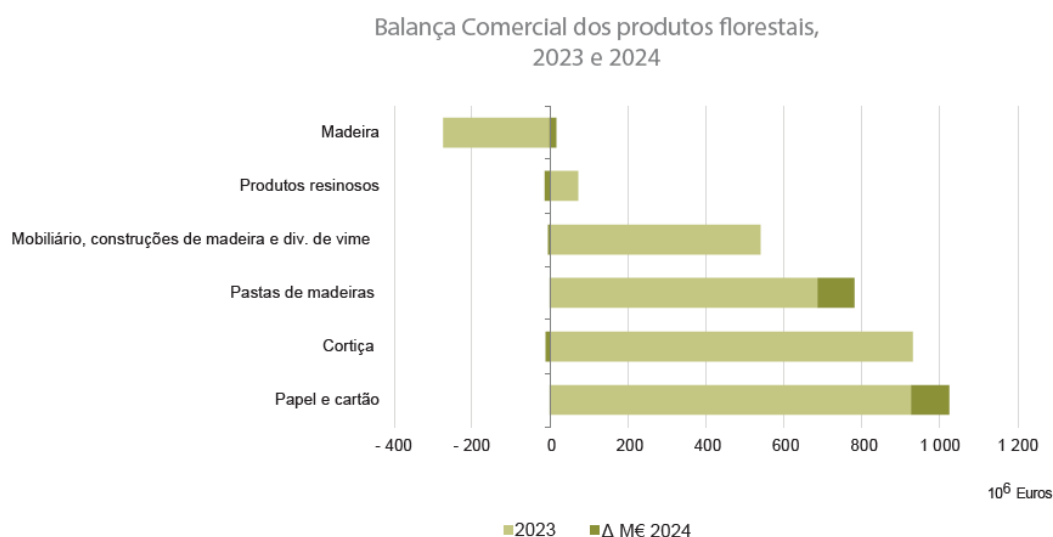


Figura 2.3 - Balança Comercial dos Produtos florestais, 2023 e 2024 (INE 2025).

Foram igualmente estabelecidos diversos mecanismos legais para a proteção do sobreiro, reconhecendo a sua importância económica, ambiental e social. Entre estas medidas incluem-se a proibição do corte sem justificação legal, a proteção contra danos que comprometam a sua vitalidade e regulamentação da extração da cortiça, que apenas pode ocorrer quando as árvores atingem determinados requisitos de dimensão do seu diâmetro do tronco. Paralelamente, a produção de cortiça é monitorizada através de manifestos estatísticos obrigatórios (Silva, J.S., 2007).

A rentabilidade da exploração do sobreiro depende, em grande medida, da quantidade e da qualidade da cortiça produzida. Segundo o *Guia de Comercialização da cortiça* da UNAC (2021), os principais parâmetros de valorização incluem a espessura da cortiça (calibre), a uniformidade da sua estrutura e a reduzida presença de defeitos, como porosidade excessiva, fendas ou inclusões lenhosas. Cortiças com maior calibre e estrutura homogénea são particularmente valorizadas pela indústria, uma vez que permitem a produção de rolhas naturais, o produto de maior valor acrescentado da fileira corticeira. Em contraste, cortiças de menor espessura ou com defeitos estruturais são geralmente destinadas ao fabrico de aglomerados, materiais de isolamento e outros produtos industriais com menor valor comercial, como ilustrado na Figura 2.4.



Figura 2.4 - Cortiça com massas heterogéneas VS Massas Homogéneas, respetivamente (UNAC, 2021).

Para além da importância económica diretamente associada à produção de cortiça, os montados de sobreiro desempenham também um papel crescente na valorização territorial e na dinamização de atividades complementares. Um desses contributos indiretos resulta da expansão do turismo de natureza em regiões onde o montado predomina, particularmente no Alentejo, cuja paisagem aberta, baixa densidade populacional e elevada qualidade ambiental constituem fatores de atração reconhecidos. Segundo a Ficha Técnica – Estatísticas do Turismo 2024, esta região tem registado um aumento consistente da atividade turística, associado à procura por destinos com forte componente natural.

Neste contexto, os sistemas agros-silvo-pastoris característicos do montado desempenham uma dupla função: por um lado contribuem para a conservação dos ecossistemas e, por outro, reforçam a atratividade económica das regiões onde predominam ao integrarem práticas tradicionais, biodiversidade e paisagem natural. Esta valorização multifuncional do montado sustenta modelos de turismo sustentável assentes na identidade ecológica e sociocultural do território, conforme referido por (Coelho et al. 2019).

2.3. Problemas fitossanitários do sobreiro

Os primeiros registos de problemas fitossanitários em povoamentos de sobreiro remontam à década de 1950, período em que cerca de 200 mil árvores foram abatidas devido ao seu estado de declínio. Entre 1958 e 1987, a área afetada manteve-se relativamente estável. No entanto, a partir da década de 1990, verificou-se um agravamento significativo da situação na Península Ibérica. Este fenómeno foi associado ao aparecimento de uma nova estirpe de patogénico responsável por elevados níveis de mortalidade em sobreiros e azinheiras. Os sintomas observados caracterizam-se pela desfoliação progressiva da copa, perda de vigor e morte das árvores, estando frequentemente associados a organismos do género *Phytophthora*, transmitidos pelo solo e pela água, que provocam necroses radiculares e a destruição das raízes finas (De Sampaio E Paiva Camilo-Alves et al., 2013).

A vitalidade do sobreiro (*Quercus suber*) depende fortemente da disponibilidade hídrica e das condições edafoclimáticas onde se desenvolve. Nas últimas décadas, os povoamentos de sobreiro em Portugal têm registado uma diminuição progressiva



da sua área e do seu estado fitossanitário, fenómeno frequentemente associado às alterações climáticas, nomeadamente ao aumento das temperaturas e à redução da precipitação (Kim et al., 2017). Paralelamente, a ausência de regeneração natural tem sido apontada como um dos principais fatores responsáveis pela degradação dos montados, em conjunto com a ocorrência de surtos de doenças e pragas (Sousa et al., 2007).

Para além dos fatores climáticos, diversos fatores de origem antrópica contribuem para a reduzida capacidade de regeneração do sobreiro e para o enfraquecimento dos montados. O sobre pastoreio, determinadas práticas agrícolas e a degradação dos solos limitam o estabelecimento de novas plantas e comprometem a renovação natural dos povoamentos. Adicionalmente, a intensa atividade humana nos montados, nomeadamente o pisoteio frequente e compactação do solo (calcamento). Em muitas situações, a ocorrência de solos pobres e pouco férteis agrava ainda mais este processo, favorecendo o enfraquecimento das árvores e aumentando a sua suscetibilidade e agentes bióticos prejudiciais (Costa et al., 2010). Segundo Sousa et al., (2007), a desfolha do sobreiro não apresenta um padrão uniforme, podendo resultar da ação simultânea de múltiplos fatores, incluindo condições climáticas adversas, elevada exposição solar, défices hídricos e nutricionais, pressão antrópica, compactação dos solos, poluição e degradação edáfica.

De acordo com o Plano de Observação da Sanidade Florestal (ICNF, 2023), a identificação de problemas fitossanitários inicia-se através da observação direta das árvores em campo. A avaliação fitossanitária contempla diferentes parâmetros, começando pela análise da copa, onde se observa a presença ou ausência de rebentação nova, a existência de ramos secos ou envelhecidos e a produção de frutos (bolotas), de acordo com a época do ano.

Posteriormente, procede-se à avaliação do tronco e da casca, verificando-se a presença de deformações, feridas, fissuras ou raízes expostas. A desfolha e a descoloração da copa são igualmente classificadas segundo o respetivo grau de severidade. Numa fase mais detalhada, identificam-se as possíveis causas dos danos observados, permitindo associar os sintomas a agentes bióticos ou abióticos específicos (ICNF, 2023).

Nos montados de sobreiro, as árvores descortçadas permitem, frequentemente, uma observação mais detalhada dos danos, sendo possível identificar sinais como galerias de insetos, presença de carvão no entrecasco, cobrilha e feridas resultantes de práticas inadequadas durante a extração da cortiça. Estas lesões podem favorecer a instalação de pragas e doenças, sobretudo em árvores já debilitadas ou sujeitas a condições de *stress* (Cabral et al., 1992). A identificação destes sintomas é fundamental para compreender os agentes responsáveis pelo declínio dos sobreiros. De acordo com Sousa et al. (2007), a maioria dos agentes patogénicos associados ao sobreiro corresponde a fungos endófitos, organismos que permanecem latentes nos



tecidos da árvore e apenas manifestam sintomas quando o hospedeiro se encontra enfraquecido ou numa fase avançada de declínio.

Principais doenças e pragas associadas ao sobreiro:

I. Declínio e podridão radicular por *Phytophthora*.spp

Esta doença é causada por várias espécies do género *Phytophthora*, incluindo *P.citricola*, *P.cactorum*, *P.cambivora* e a *P.quercina*, estando associada ao declínio crónico e progressivo de espécies do género *Quercus* na Europa. Os sintomas incluem a formação de tufos de foliares, abscisão de ramos, rebentação epicórmica, amarelecimento e murchidão das folhas, conduzindo à progressiva transparência da copa. O patogénico ataca principalmente o sistema radicular, provocando a destruição das raízes finas e lesões necróticas em raízes de maior calibre. A deterioração das raízes e a obstrução dos vasos reduzem a condução de água e nutrientes, originando stress hídrico e défices nutricionais na planta (Jung et al., 1996).

II. Doença do Carvão

A doença do carvão é causada pelo fungo *Biscogniauxia mediterranea* e constitui uma das patologias mais frequentes associadas ao declínio do sobreiro na região mediterrânica. Trata-se de um fungo oportunista que se desenvolve preferencialmente em árvores sujeitas a condições de *stress* ambiental (D. Costa et al., 2020). Os sintomas refletem alterações nas relações hídricas e fotossintéticas da árvore causando o fechamento dos estomas. O sinal mais característico é a presença de estruturas negras semelhantes ao carvão, que dão origem à designação comum “carvão no entrecasco” (Santos, 2003).

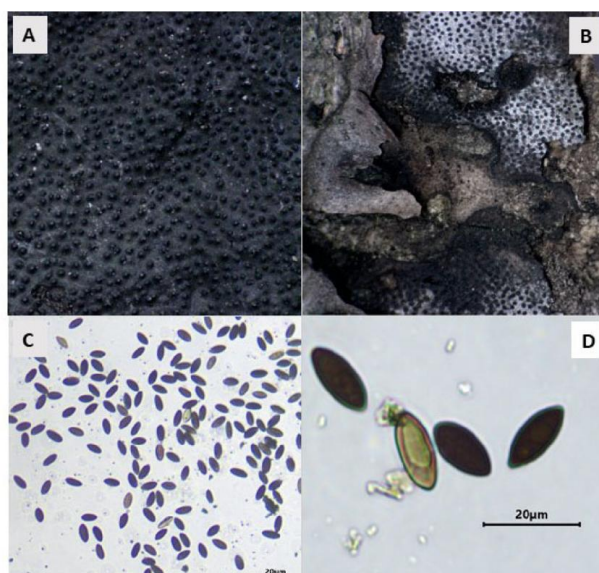


Figura 2.5 - *Biscogniauxia mediterranea* em *Quercus rubra*: A; B) superfície do estroma carbonáceo; C; D) ascósporos imaturos e maduros (Georgieva et al., 2024).

III. Cancro de Diplodia



A espécie *Diplodia corticola* é o agente causal do “cancro de bot”, uma doença cuja incidência tem aumentado significativamente nas últimas décadas (Figura 2.6). Esta patologia provoca o desenvolvimento de cancrios no tronco que afetam gravemente a produção e a qualidade da cortiça, podendo conduzir à morte da árvore em situações severas (D. Costa et al., 2020).



Figura 2.6 - Sintomas observados em sobreiro causados pela doença do cancro *Diplodia corticola* (Muñoz-Adalia et al., 2022).

IV. Cobrilha do Sobreiro

A cobrilha do sobreiro, causada por *Coroebus undatus* (Figura 2.7), é considerada uma das pragas mais importantes desta espécie devido aos elevados prejuízos económicos que provoca. As fêmeas depositam os ovos nas fissuras da cortiça e, após a eclosão, as larvas escavam galerias sinuosas que podem atingir cerca de 1,8 m de comprimento e entre 3 e 5 mm de largura. O ataque reduz significativamente o valor comercial da cortiça, podendo causar perdas de qualidade próximas dos 90%. Os sinais visíveis incluem manchas cloróticas, exsudações escuras na casca e, em situações mais severas, desfolha provocada pela alimentação dos adultos (Fürstenau et al., 2015).



Figura 2.7 – *Coroebus undatus* (Observation.org, 2026).

V. Ataque de *Platypus cylindrus*



O *Platypus cylindrus* é um escaravelho xilófago (Figura 2.8) da qual constitui um dos principais agentes bióticos associados ao declínio e mortalidade dos sobreiros. Este escaravelho perfura o lenho da árvore, instalando galerias no tronco que comprometem a circulação da seiva e favorecem a colonização por fungos associados ao processo de degradação. A sua ocorrência é particularmente frequente em povoamentos sujeitos a desequilíbrios ecológicos, nomeadamente em áreas com drenagem deficiente ou após períodos prolongados de seca, condições que reduzem o vigor das árvores e aumentam a sua suscetibilidade ao ataque. Embora possa afetar indivíduos de diferentes idades, os ataques são especialmente comuns em árvores descortiçadas, podendo levar à morte destes (Muñoz-Adalia et al., 2022).



Figura 2.8 – Inseto *Platypus cylindrus* macho e fêmea (Muñoz-Adalia et al., 2022)

Outro fator de grande relevância para a sanidade do sobreiro é o descortiçamento. Embora esta operação seja indispensável para a exploração económica do sobreiro, constitui, simultaneamente, uma importante fonte *stress* quando realizada de forma inadequada. Durante a extração da cortiça, a utilização de machados pode provocar ferimentos profundos que atingem o felogénio ou mesmo o câmbio vascular. Estes danos desencadeiam processos de cicatrização que comprometem a uniformidade da cortiça produzida e podem interromper a sua formação durante vários anos, bem como constituem portas de entrada para fungos oportunistas, incluindo *Biscogniauxia mediterranea*, cuja incidência aumenta significativamente em árvores sujeitas a *stress* hídrico. Os efeitos deste *stress* refletem-se também na fenologia da árvore (A. Costa, 2004). Segundo o mesmo autor, as árvores debilitadas apresentam incrementos diamétricos anuais inferiores aos observados em árvores saudáveis (8,5 mm/ano e 9,8 mm/ano, respetivamente). Paralelamente, registou-se uma redução da espessura da cortiça produzida entre 13% e 16%, evidenciando o impacto do enfraquecimento fisiológico no crescimento da árvore e na qualidade da produção.

Em indivíduos enfraquecidos, o início do crescimento primaveril e os principais picos de crescimento anual podem sofrer atrasos de aproximadamente um mês relativamente a árvores saudáveis, prolongando o período de vulnerabilidade a condições ambientais adversas (A. Costa, 2004).



Assim, o declínio do sobreiro resulta frequentemente da interação entre fatores de gestão, condições ambientais adversas e agentes bióticos, mais do que da ação isolada de um único agente patógeno (A. Costa, 2004).

Perante esta complexidade diagnóstica e a necessidade de monitorização sistemática do estado fitossanitário dos sobreiros, tecnologias de deteção remota emergem como ferramentas complementares aos inventários florestais tradicionais. A integração de dados LiDAR oferece a capacidade de caracterizar simultaneamente a estrutura florestal e o vigor vegetativo, conforme analisado em seguida.

2.4. LiDAR no Inventário Florestal

A tecnologia LiDAR (*Light Detection and Ranging*) funciona fundamentalmente através da emissão de impulsos de luz laser que medem a distância entre o sensor e as superfícies refletoras, gerando conjuntos massivos de coordenadas tridimensionais conhecidos como nuvens de pontos (Laino et al., 2024). Esta tecnologia funciona através de medição do tempo que cada impulso demora a percorrer o trajeto de ida e volta, permitindo o cálculo de coordenadas tridimensionais (X,Y,Z) extremamente precisas que, em conjunto, formam o que se designa por nuvem de pontos (Laino et al., 2024).

O LiDAR é utilizado para a criação de modelos geoespaciais de alta resolução, sendo essencial em áreas como a gestão florestal, a hidrologia para simulação de cheias e a engenharia civil (Zhang et al., 2016). A sua execução divide-se principalmente em dois tipos de plataformas: sistemas aerotransportados tripulados e UAVs (drones com LiDAR incorporado), que permitem cobrir vastas áreas geográficas e sistemas terrestres, focados na precisão e detalhe ao nível da parcela (Laino et al., 2024; Xi et al., 2024).

1.1.1. LiDAR Aerotransportado (ALS)

O Sistema LiDAR aerotransportado (ALS) é amplamente reconhecido pela sua eficiência em capturar variações na estrutura florestal e topográfica em escalas regionais e de paisagem (Xi et al., 2024). O seu funcionamento científico assenta na capacidade do laser em penetrar parcialmente a copa das árvores, permitindo obter retornos tanto da vegetação como do solo (Xi et al., 2024; Zhang et al., 2016). Contudo, uma limitação crítica do ALS é a oclusão do sinal em povoamentos densos, onde a vegetação pode ocultar mais de 25% da copa, dificultando a extração de métricas precisas do subcoberto e dos troncos (Xi et al., 2024). Para converter os dados brutos em produtos úteis como o Modelo Digital do Terreno (MDT), é necessário aplicar algoritmos de filtragem que separem os pontos do solo dos objetos acima deste (Zhang et al., 2016).

Um método estudado e desenvolvido é a aplicação do *Cloth Simulation Filtering* (CSF), proposto por Zhang et al., (2016), um algoritmo de filtragem/processamento



para dados LiDAR. Este método permite filtrar nuvens de pontos provenientes de LiDAR, de forma a separar os pontos de terreno (solo) de objetos (não solo), como edifícios e vegetação. O objetivo principal é facilitar a geração de modelos digitais de terreno com um algoritmo que exige poucos parâmetros e de fácil utilização. Este método inverte a nuvem de pontos e simula o comportamento mecânico de um tecido virtual rígido a “cair” sobre a superfície invertida, permitindo identificar de forma robusta a geometria do terreno mesmo em topografias complexas.

Recentemente, Xi et al., (2024) demonstraram o potencial da integração de arquiteturas de aprendizagem profunda, nomeadamente transformadores *SegFormer* e redes neuronais convolucionais (CNN), para ultrapassar as limitações associadas à oclusão de sinal de levantamentos ALS. No estudo, os dados TLS (*Terrestrial Laser Scanning*), capazes de captar com elevada precisão os detalhes dos troncos ao nível do solo, foram utilizados como “verdade terreno” para calibrar modelos de inteligência artificial aplicados aos dados ALS. A metodologia incluiu a utilização de dados ALS brutos, aos quais foi aplicada uma classificação binária dos pontos, separando-os em classes de árvore e não árvore. Posteriormente, foi implementada uma rede baseada em transformadores, denominada *SegFormer* e adaptada para o processamento tridimensional, que apresentou um desempenho superior aos métodos tradicionais de classificação, como *LASTools* e o *TerraScan*. Para a segmentação individual das árvores na nuvem de pontos, os autores desenvolveram um novo método *Fuzzy Center Segmentation* (FCS). Esta abordagem identifica inicialmente o centro difuso de cada copa e, em seguida, agrupa os restantes pontos através de algoritmos baseados em grafos. Os resultados obtidos revelaram uma taxa de deteção de árvores de 79%, superando significativamente os métodos convencionais baseados em imagens bidimensionais, sobretudo na deteção de árvores de menor porte ou ocultas sob o dossel dominante. O diâmetro à altura do peito (DAP) foi extraído manualmente a partir das nuvens de pontos TLS para servir como referência na calibração dos modelos preditivos. Foram avaliadas diferentes abordagens de regressão profunda, destacando-se o modelo CNN-GPR (Rede Neuronal Convolucional com Processo Gaussiano), que alcançou uma precisão elevada na estimativa do DAP, apresentando um erro médio de apenas 4,2. No que respeita ao mapeamento da biomassa individual de mais de 30 milhões de árvores identificadas nas áreas de estudo. Esta abordagem demonstrou ser uma solução económica e escalável para converter levantamento ALS, de âmbito regional, em inventários florestais detalhados ao nível da árvore individual. Os autores concluem que a utilização de técnicas de *Deep Learning* permite transferir atributos detalhados dos troncos, obtidos através do TLS, para a escala da paisagem recorrendo a dados ALS, automatizando um processo que tradicionalmente dependia de medições de campo manuais e, por vezes, destrutivas. Com o objetivo de promover o avanço da monitorização florestal e facilitar a replicação dos resultados, o código desenvolvido no âmbito do estudo foi disponibilizado publicamente (Xi et al., 2024).



De forma complementar, estudos baseados em UAV-LiDAR, têm reforçado a importância da otimização dos processos de segmentação de copas em cenários de elevada complexidade topográfica. Em particular, investigações recentes realizadas em plantações *Camellia oleifera* (um arbusto de porte médio explorado em sistemas de produção agrícola florestal) demonstraram que a combinação de segmentação em nuvem de pontos 3D, métodos baseados em Canopy Height Model (CHM) e abordagens híbridas pode atingir elevados níveis de precisão na delimitação de copas individuais, mesmo em terrenos montanhosos e com forte sobreposição de copas. Os resultados indicam desempenhos entre 84% e 93%, sendo o *clustering* direto da nuvem de pontos o método mais robusto. Verificou-se ainda que o pré-processamento com filtragem do solo (*Cloth Simulation Filter*) é determinante para o aumento da precisão, uma vez que reduz significativamente o ruído associado à vegetação rasteira e aos pontos do terreno. Adicionalmente, a resolução do CHM, com valores ótimos próximos de 0,5 metros, influencia diretamente a ocorrência de sobre-segmentação ou sub-segmentação, afetando a delimitação correta das copas individuais (Farajelahi et al., 2023).

1.1.2. LiDAR Terrestre (TLS) e LiDAR Móvel (MLS)

Diferente dos sistemas LiDAR aerotransportados, o LiDAR terrestre destaca-se pela capacidade de adquirir dados com resolução subcentimétrica, permitindo a reconstrução tridimensional detalhada de ramos e fustes sem recorrer a métodos de amostragem destrutivos (Prendes et al., 2021). Esta tecnologia subdivide-se em duas categorias principais: o *Terrestrial Laser Scanning* (TLS), baseado em estações fixas, e o *Mobile Laser Scanning* (MLS), que utiliza plataformas móveis para a aquisição de dados (Laino et al., 2024).

O TLS opera a partir de posições estacionárias e é particularmente adequado para inventários florestais que exigem elevada precisão na caracterização estrutural das árvores e na modelação do volume comercial da madeira (Prendes et al., 2021). Neste contexto, Prendes et al., (2021) desenvolveram uma metodologia destinada à extração automática de variáveis dendrométricas, nomeadamente o diâmetro à altura do peito (DAP) e a altura total das árvores, com o objetivo de ajustar equações de volume sem necessidade de intervenção manual intensiva. A metodologia proposta compreende várias etapas fundamentais de processamento de nuvem de pontos. Inicialmente, é gerado um Modelo Digital de Terreno (MDT), que serve de referência para a determinação das alturas relativamente ao solo. Seguidamente, é selecionada uma faixa horizontal da nuvem de pontos onde são aplicados algoritmos de agrupamento (*clustering*) para identificar potenciais troncos. Após esta fase, recorre-se à Análise de Componentes Principais (PCA) para determinar o eixo principal de cada tronco, permitindo distinguir de forma eficaz os elementos pertencentes à copa e aos ramos. Posteriormente, procede-se à individualização das árvores, associando cada ponto da nuvem ao eixo da árvore mais próximo. A altura total é então estimada a partir do ponto mais elevado localizado no interior de um cilindro virtual definido em



torno do eixo do tronco. Segundo os autores, este procedimento apresenta elevada robustez face ao ruído presente nos dados e mantém um bom desempenho mesmo em árvores inclinadas. Para a estimativa dos diâmetros ao longo do fuste, o tronco é dividido em secções horizontais de espessura uniforme, sendo ajustado um círculo a cada secção para determinar o respetivo diâmetro (Figura 2.9).

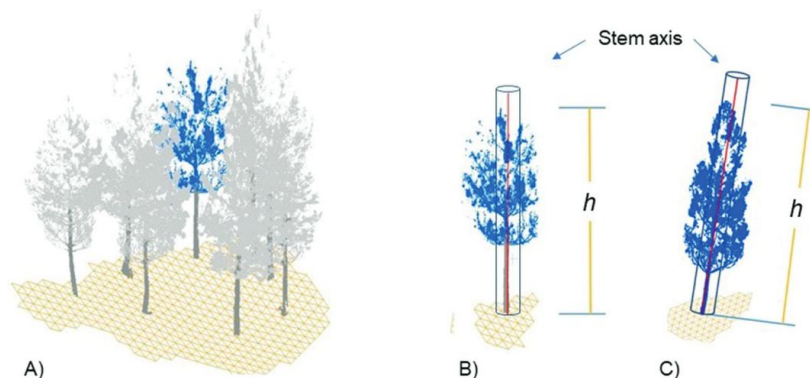


Figura 2.9 - Esquema do procedimento de estimativa da altura das árvores: (A) individualização das árvores na nuvem de pontos; (B) determinação da altura em árvores verticais após filtragem do ruído; e (C) aplicação do método em árvores inclinadas

O método foi validado numa parcela florestal com aproximadamente 5700 m², contendo 428 árvores, localizada numa área com declive acentuado e vegetação densa. Apesar da complexidade das condições de aquisição, o algoritmo identificou corretamente cerca de 97 % das árvores presentes. Relativamente ao DAP, foi obtido um erro médio quadrático (RMSE) de apenas 1,14 cm, enquanto a estimativa da altura apresentou um RMSE de 1,52 m, verificando-se uma ligeira tendência para a subestimação devido à oclusão das copas provocada pela densidade da folhagem. O processamento necessário para estimar as variáveis dendrométricas foi de aproximadamente três minutos por árvore, evidenciando uma elevada eficiência quando comparado com os métodos tradicionais de inventário florestal (Prendes et al., 2021).

De forma geral, os resultados demonstram que o TLS constitui uma alternativa eficaz aos métodos convencionais de inventário florestal, possibilitando a criação de reconstruções tridimensionais permanentes da floresta, realizar análises multitemporais e extrair novas variáveis estruturais sem necessidade do inventário tradicional. Embora a precisão diminua nas regiões superiores da copa devido aos efeitos de oclusão causados pelos ramos e pela folhagem, os autores concluem que a incorporação de informação supervisionada contribui para melhorar significativamente a qualidade das estimativas finais de volume (Prendes et al., 2021).

Por sua vez, o *Mobile Laser Scanning* (MLS) utiliza sensores montados em plataformas móveis (Figura 2.10), que permitem adquirir dados enquanto o operador percorre a área de estudo. Esta abordagem possibilita a observação dos troncos a partir de múltiplos ângulos, reduzindo significativamente os efeitos de oclusão e



melhorando a estimativa de variáveis dendrométricas, nomeadamente o DAP, quando comparada com o *Terrestrial Laser Scanning* (TLS) (Laino et al., 2024).



Figura 2.10- Fotografia de uma sessão de demonstração da tecnologia *Mobile Laser Scanning* (MLS)

O processamento das nuvens de pontos provenientes do MLS é frequentemente realizado através do *software 3DFin*, uma ferramenta gratuita e de código aberto desenvolvida para automatizar inventários florestais tridimensionais (Laino et al., 2024). O software integra algoritmos de normalização do terreno baseados no *Cloth Simulation Filtering* (CSF), filtros de verticalidade e técnicas de agrupamento que permitem identificar automaticamente os troncos, remover o ruído associado à vegetação e calcular parâmetros dendrométricos como o DAP, a altura total e a localização espacial das árvores. O 3DFin encontra-se disponível como *plugin* para QGIS e *CloudCompare*, bem como aplicação autónoma para *Windows* e pacote *Python*, suportando ficheiros LAS e LAZ com exportação dos resultados em formato XLSX. Segundo Laino et al., (2024), o algoritmo apresenta taxas de deteção próximas de 100%, com RMSE inferiores a 2 centímetros na estimativa do DAP e tempos de processamento compreendidos entre 2 e 7 minutos por parcela. No entanto, os autores referem que o método não é adequado para dados exclusivamente provenientes de LiDAR aerotransportado ou de sensores de satélite, uma vez que estes não permitem observar diretamente os troncos e o terreno. Como perspetiva futura, é proposta a integração de técnicas de *Deep Learning* para melhorar a segmentação da madeira e a estimativa do volume das árvores.

Em síntese, enquanto o ALS permite caracterizar a estrutura da floresta à escala da paisagem, os sistemas TLS e MLS proporcionam uma caracterização estrutural muito mais detalhada ao nível da árvore individual. A complementaridade entre estas



tecnologias permite melhorar a precisão dos inventários florestais, calibrar modelos alométricos e validar produtos derivados de detecção remota, constituindo uma base sólida para aplicações em monitorização florestal e modelação espacial (Xi et al., 2024).

2.5. Índices de Vegetação

Os índices de vegetação constituem métricas espectrais amplamente utilizadas em teledeteção para caracterizar a condição, estrutura e dinâmica da vegetação a partir da resposta espectral registada por sensores remotos. Estes índices resultam de combinações matemáticas entre diferentes bandas espectrais, sobretudo no vermelho, infravermelho próximo (NIR) e, mais recentemente, no *red-edge*. A sua utilização permite estimar de forma indireta propriedades biofísicas como a biomassa, o teor de clorofila, área foliar (LAI), o vigor vegetativo, *stress* hídrico e estado fitossanitário (Chrysafis et al., 2020b).

O princípio físico subjacente aos índices de vegetação baseia-se no comportamento espectral das plantas. A clorofila absorve intensamente a radiação na região do vermelho (Red), utilizada no processo fotossintético, enquanto a estrutura interna das folhas provoca uma elevada reflectância na região do infravermelho próximo (NIR). A diferença entre estas duas bandas constitui, assim, um indicador da presença e do estado fisiológico da vegetação, sendo que a vegetação saudável apresenta valores elevados de reflectância no NIR e reduzidos na banda vermelho (Fabijańczyk & Zawadzki, 2022).

Com o desenvolvimento de sensores multiespectrais, como o Sentinel-2, tornou-se possível calcular diversos índices de vegetação, cada um desenvolvido para responder a limitações específicas ou para realçar determinadas características da cobertura vegetal. Entre os mais utilizados destaca-se o NDVI (*Normalized Difference Vegetation Index*), amplamente aplicado na monitorização da densidade e vigor da vegetação devido à sua simplicidade e robustez. No entanto, em áreas onde existe elevada exposição do solo, como zonas urbanas, agrícolas ou com vegetação esparsa, o NDVI pode ser influenciado pelo brilho do solo, conduzindo a estimativas menos precisas.

Para minimizar esta limitação foram desenvolvidos índices ajustados ao efeito do solo, como o MSAVI2, que reduz a influência do solo exposto sem necessidade de parâmetros empíricos adicionais. Este índice apresenta melhor desempenho em áreas de baixa cobertura vegetal, sendo particularmente útil em ambientes urbanos e em fases iniciais do desenvolvimento das culturas (Fabijańczyk & Zawadzki, 2022).

Os índices de vegetação são calculados a partir de imagens previamente sujeitas a processamento radiométrico e atmosférico, utilizando, preferencialmente, reflectâncias de superfície. No caso das imagens Sentinel-2, este processamento



pode ser realizado no *software* SNAP (*Sentinels Application Platform*), desenvolvido pela Agência Espacial Europeia (ESA).

Para além do cálculo direto de índices espectrais, o SNAP disponibiliza algoritmos destinados à estimativa de variáveis biofísicas, como o *Leaf Area Index* (LAI), *Fraction of Absorbed Photosynthetically Active Radiation* (FAPAR), *Fractional Vegetation Cover* (FCOVER) e Cab (conteúdo de clorofila). Segundo Jay (s.d.), estes produtos são obtidos através de redes neuronais artificiais (*Artificial Neural Networks* – ANN), previamente treinadas com dados simulados por modelos de transferência radiativa acoplados. Estes modelos incluem o PROSPECT, responsável pela simulação das propriedades óticas das folhas, e o SAIL (*Scattering by Arbitrarily Inclined Leaves*), que descreve a interação da radiação com a estrutura do dossel vegetal. A rede neuronal utiliza como variáveis de entrada as reflectâncias de superfície de bandas específicas do Sentinel-2, bem como a geometria de aquisição da imagem, produzindo estimativas de parâmetros biofísicos de forma automática (Jay, s.d.).

A literatura demonstra que os índices de vegetação podem igualmente ser integrados com métodos de análise espacial para caracterizar padrões da paisagem. Fabijańczyk e Zawadzki (2022) combinaram os índices NDVI e MSAVI2 com técnicas geoestatísticas, recorrendo à análise de variogramas para avaliar a continuidade espacial da vegetação. Os autores analisaram parâmetros como o alcance da correlação espacial e a semivariância, verificando que as florestas naturais apresentam maior continuidade espacial, refletida por alcances de correlação mais elevados, enquanto prados e pastagens evidenciam maior heterogeneidade devido à variabilidade sazonal e às práticas de gestão. O estudo demonstrou ainda que o MSAVI2 apresenta melhor desempenho na caracterização de vegetação baixa e fragmentada em ambientes urbanos, reduzindo significativamente a influência do solo relativamente ao NDVI.

Em síntese, os índices de vegetação representam ferramentas fundamentais na monitorização remota da vegetação, permitindo obter informação quantitativa sobre o seu estado e funcionamento a diferentes escalas espaciais e temporais. A disponibilidade de imagens multiespectrais de elevada resolução, como as do Sentinel-2, associada a plataformas de processamento como o SNAP, tornou possível a extração sistemática destes índices e de variáveis biofísicas derivadas, constituindo um importante suporte para aplicações nas áreas da gestão florestal, agricultura de precisão, planeamento urbano e monitorização ambiental. A complementaridade entre índices de vegetação e estrutura florestal demonstrada pela literatura sugere que a integração destas variáveis em modelos de aprendizagem automática contribui, significativamente, a previsão do estado fitossanitário. A presente investigação explora esta hipótese através de uma metodologia que integra múltiplas fontes de dados geoespaciais.



3. METODOLOGIA

A metodologia, expressa na Figura 3.1, deste estudo teve como objetivo analisar e modelar padrões associados ao estado fitossanitário e à qualidade potencial da cortiça, através da integração de dados de inventário florestal, informação geoespacial e técnicas de *Machine Learning*.

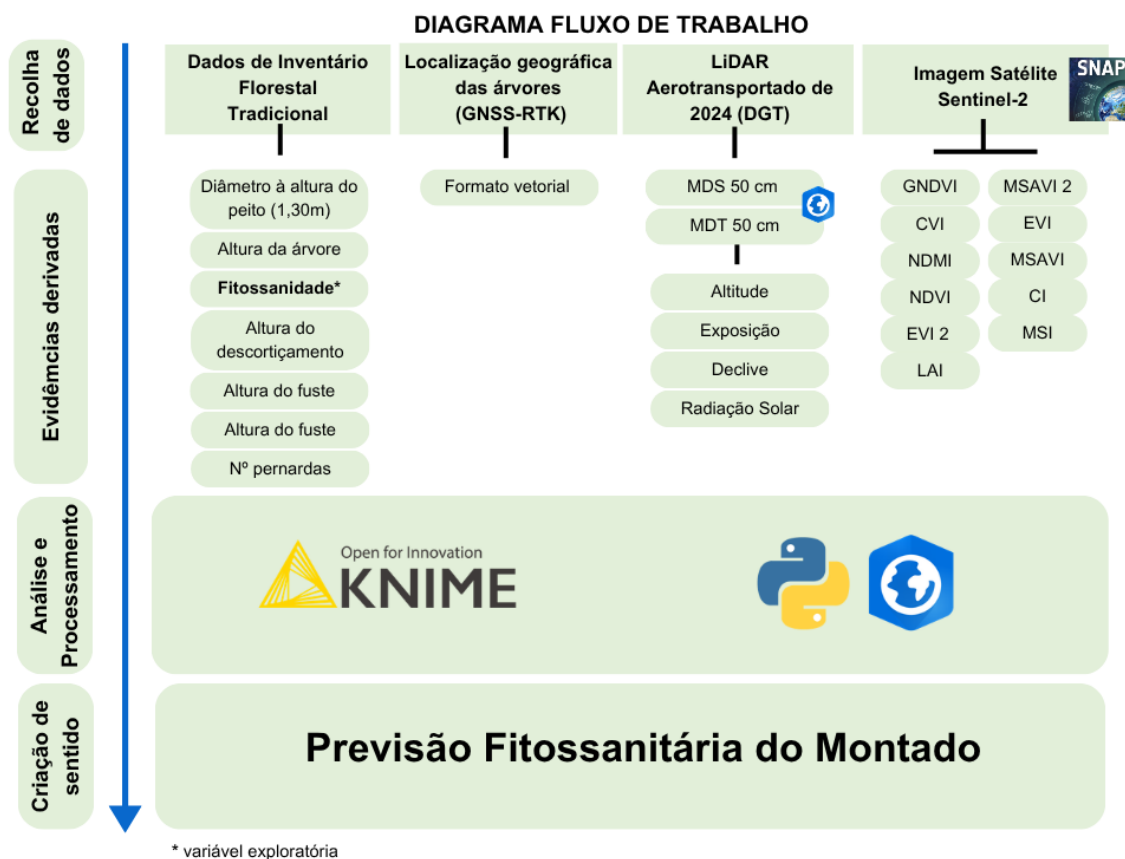


Figura 3.1 – Diagrama do Fluxo de Trabalho.

Foram recolhidos dados dendrométricos e de avaliação fitossanitária em inventário de campo, com georreferenciação precisa das árvores recorrendo a GNSS-RTK, constituindo a base de validação do estudo.

Inicialmente, a estratégia de investigação contemplava a recolha de dados LiDAR móvel terrestre. No entanto, optou-se pelos dados do LiDAR aerotransportados da Direção-Geral do Território (DGT), permitindo a caracterização estrutural da vegetação, nomeadamente altura das árvores e diâmetro das copas.

Foram igualmente utilizadas imagens Sentinel-2 recolhidas no *Copernicus Data Space Ecosystem*, a partir das quais foram calculados índices de vegetação para avaliação do vigor e deteção de possíveis sinais de declínio. Estas variáveis foram complementadas com informação ambiental e espacial derivada de SIG.

Todos os dados foram integrados em ambiente SIG, permitindo a individualização das árvores e a construção de uma base de dados multivariada. Posteriormente, no



KNIME *Analytics Platform* (v. 5.9.0), foram aplicados modelos de *Machine Learning* após etapas de pré-processamento e seleção de variáveis, com o objetivo de distinguir indivíduos saudáveis de indivíduos em declínio (Figura 3.1).

Para validar a viabilidade de uma abordagem baseada em SIG e *Machine Learning* para diagnóstico fitossanitário não-invasivo, foi selecionada uma área piloto no interior alentejano.

3.1. Área de Estudo

A área de estudo localiza-se no interior do Alentejo Central, nas proximidades do município da Chamusca, entre os concelhos de Abrantes, Ponte de Sor e Coruche (Figura 3.2). Esta zona foi selecionada devido à elevada concentração de sobreiros em exploração para extração de cortiça e à identificação de exemplares com problemas fitossanitários.

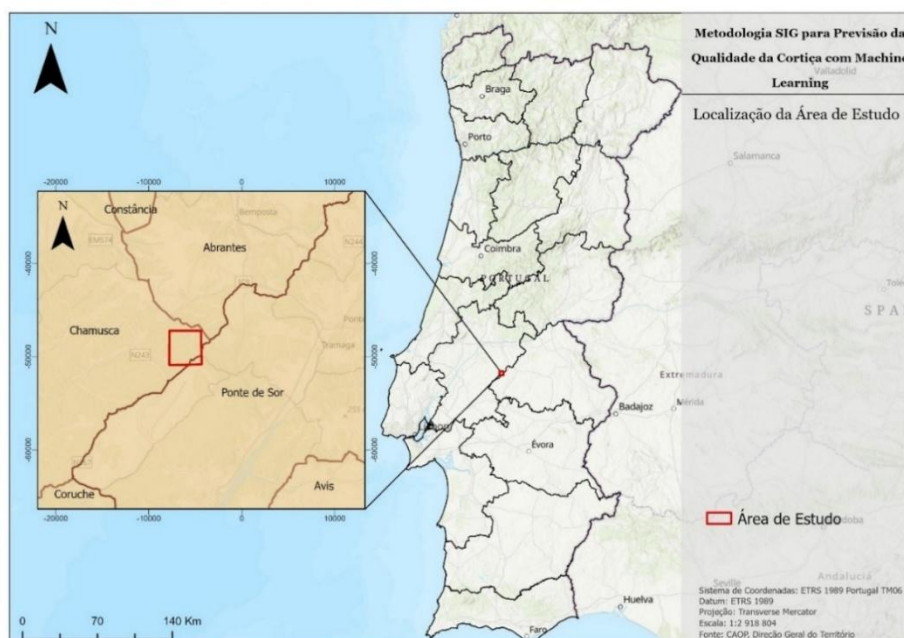


Figura 3.2 - Localização da Área de Estudo.

A caracterização topográfica foi realizada em ambiente ArcGis Pro (ESRI), com base no Modelo Digital do Terreno (MDT), a partir do qual foram derivados os principais parâmetros morfológicos utilizados na análise espacial, nomeadamente altitude, declive e exposição das vertentes. Para a produção cartográfica, recorreu-se a ferramentas de geoprocessamento, designadamente *Slope* e *Aspect*, seguidas da respetiva reclassificação, o que permitiu a sua representação temática e posterior integração com a informação contida na *shapefile* do inventário florestal.

As cartas temáticas produzidas permitiram não só caracterizar a área piloto, mas também contextualizar espacialmente os exemplares amostrados, servindo de suporte à análise da relação entre condições biofísicas do terreno e estado fitossanitário dos sobreiros.



Em relação à carta de altimetria, a área apresenta um relevo caracterizado por colinas suaves e vales abertos, elevação média de 171 metros, conforme ilustrado na Figura 3.3.

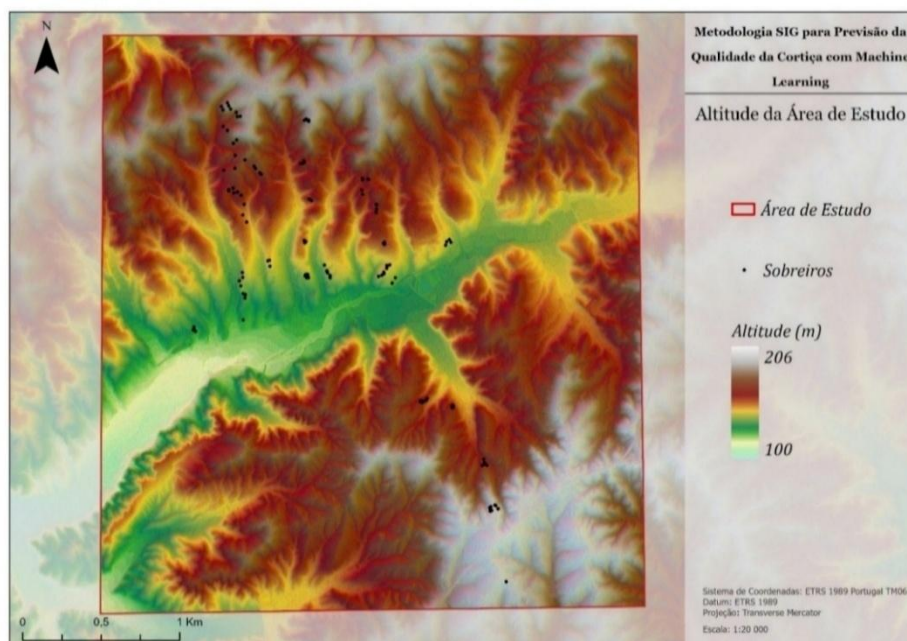


Figura 3.3 - Carta da Altimetria na Área de Estudo.

Relativamente ao declive, verifica-se o predomínio da classe 10–15 %, embora existam áreas com inclinações mais acentuadas nas margens da linha de água que atravessa o perímetro da área em análise (Figura 3.4). Os sobreiros inventariados distribuem-se maioritariamente em zonas de declive moderado.

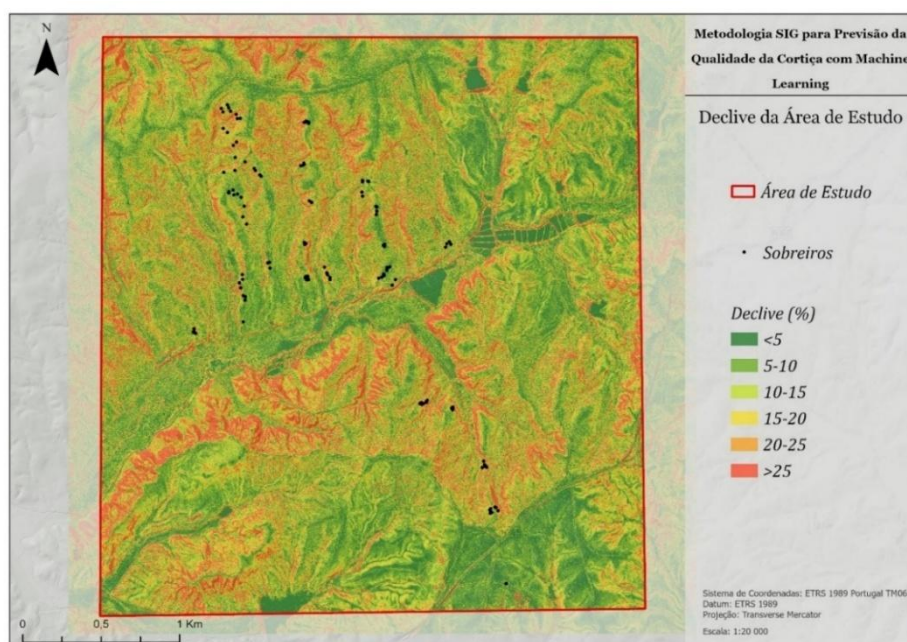


Figura 3.4 - Carta de Declive da Área de Estudo.



No que respeita à exposição das vertentes (Figura 3.5), observa-se, na amostra recolhida, uma maior concentração de sobreiros em áreas orientadas a Sudeste e Sul. Estas exposições tendem a receber uma maior quantidade de radiação solar ao longo do dia, influenciando as condições de temperatura e disponibilidade hídrica do local, fatores que podem condicionar o crescimento e a distribuição da espécie.

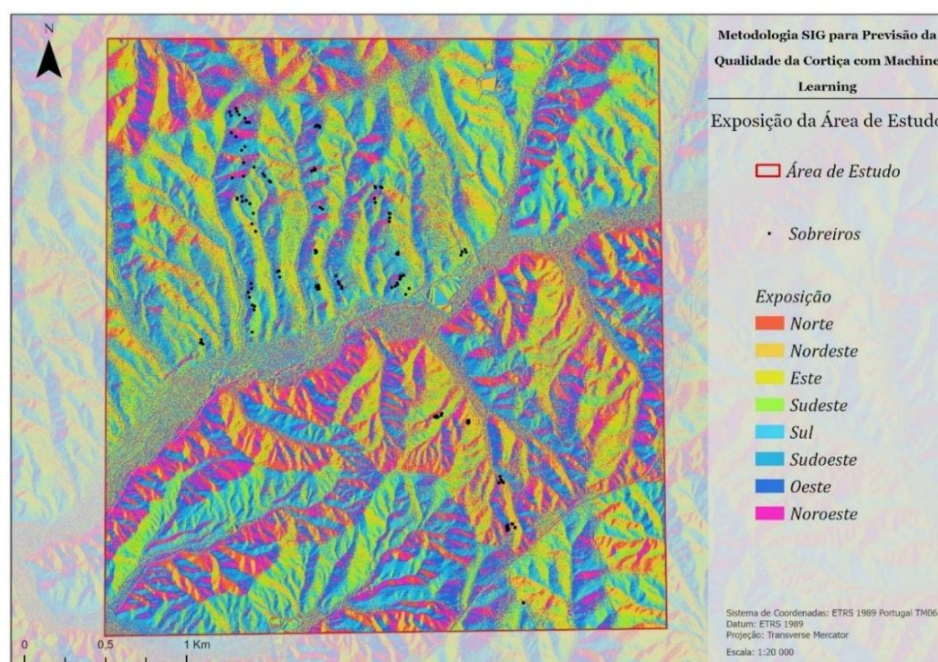


Figura 3.5 - Carta da Exposição da Área de Estudo.

3.2. Recolha de Dados

A base de dados da presente investigação envolveu a integração de diferentes fontes de informação matricial e vetorial, de forma a caracterizar os sobreiros e as condições biofísicas da área de estudo. Todos os dados geográficos foram harmonizados no sistema de referência ETRS89 / PT-TM06 (EPSG:3763), garantindo a consistência espacial e a compatibilidade entre as diferentes camadas de informação utilizadas nas análises.

A Tabela 3.1 apresenta a síntese das variáveis consideradas, suas respectivas fontes, resolução espacial e formato dos dados.



Tabela 3.1 - Base de dados da investigação

Variáveis	Dados	Resolução/ Escala	Fonte	Formato
Variável Dendrométrica	DAP (Diâmetro à altura do Peito)	-	Própria	Vetorial
	Altura total da árvore			
	Espessura da cortiça			
	Diâmetro da Copa			
	Ano de descortiçamento			
	Estado do Fitossanitário (Cobrilha, Carvão e Feridas)			
Variável Espetral	Índices de Vegetação (Tabela 3.3)	10 metros	Imagens de Satélite (Sentinel-2)	Raster
Variáveis Biofísicas do Terreno	Altitude	0,5 metros	Modelo Digital do Terreno (DGT, 2024)	Raster
	Exposição			
	Declive			
	Radiação Solar (anual)			

3.2.1. Inventário Florestal

O inventário florestal foi realizado recorrendo a GPS- RTK acoplado com antena GNSS, de forma a georreferenciar individualmente cada árvore do inventário, gerando um ficheiro vetorial de pontos.

Neste total de idas ao campo foram recolhidas 137 amostras para o modelo, seguindo um protocolo de inventário de Tomé, Barreiro e Soares (2019), realizado a sobreiros, na Companhia das Lezírias, para a recolha das variáveis dendrométricas.

Foram recolhidas todas as variáveis dendrométricas (Figura 3.6) relevantes para o estudo, nomeadamente:

- DAP (diâmetro à altura do peito)
- Altura total da árvore
- Espessura da cortiça
- Diâmetro da copa
- Estado Fitossanitário (cobrilha, feridas, carvão) - **Variável Dependente**
- Registo do ano de descortiçamento

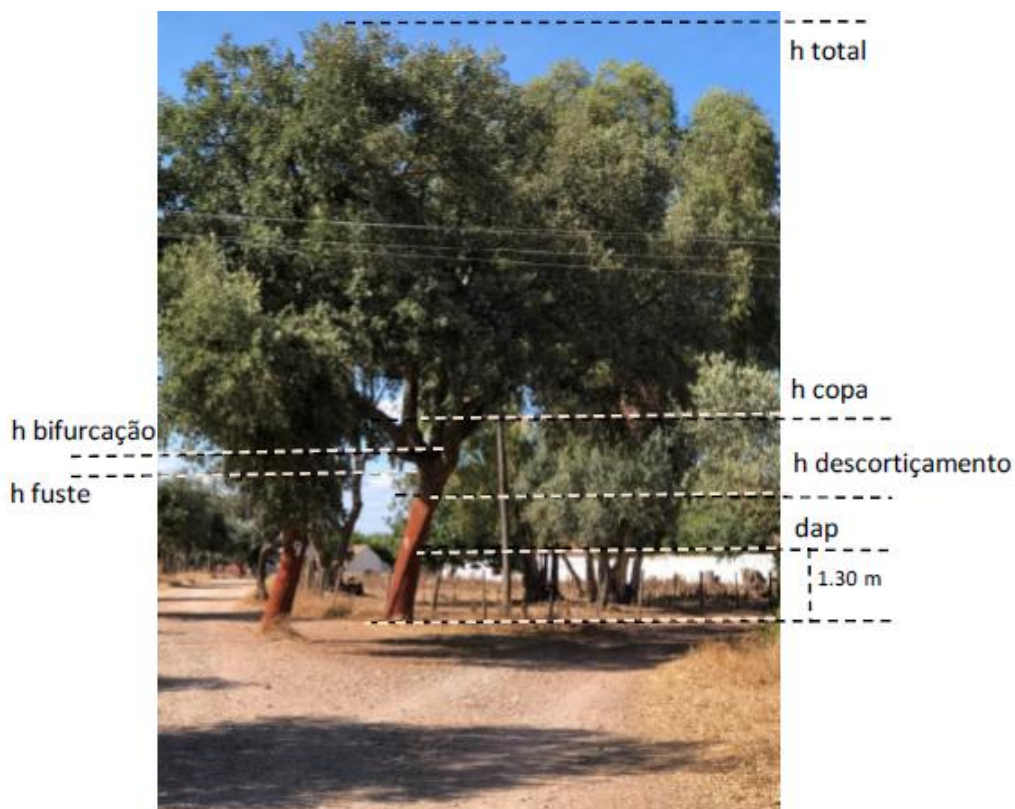


Figura 3.6 - Medidas dendrométricas do sobreiro. Foto retirada de Tomé, Barreiro e Soares (2019).

Para a recolha dos dados de inventário florestal em campo, foi desenvolvido um formulário digital onde foi utilizada a aplicação ArcGIS *Field Maps*. Este formulário (Figura 3.7), permitiu o registo estruturado e georreferenciado das observações dendrométricas e do estado fitossanitário dos sobreiros, garantindo a uniformização da informação recolhida e a sua integração direta em ambiente SIG.



Figura 3.7 - Formulário digital na aplicação ArcGIS *Field Maps* utilizado para a recolha de dados de inventário florestal.

A avaliação fitossanitária foi realizada em simultâneo com o inventário dendrométrico, mediante observação visual detalhada de cada indivíduo amostrado. Esta avaliação teve como objetivo identificar a presença de sintomas e sinais associados às principais patologias e fatores de degradação que afetam os povoamentos de sobreiro, tomando como referência os critérios no documento de *Diagnóstico das causas da anormal mortalidade dos sobreiros a Sul do Tejo* (Cabral et al., 1992).

Durante o trabalho de campo, foram observados e registados diversos indicadores do estado sanitário das árvores, nomeadamente a presença de ramos secos ou mortos, desfolha anormal, redução da densidade da copa e outros sintomas associados a situações de declínio.

Para efeitos de modelação, a classificação fitossanitária centrou-se na deteção de três tipos de anomalias consideradas relevantes para o estudo (as mais comuns na área de estudo):

- Carvão do entrecasco, caracterizado pela presença de manchas negras ou escurecimento da superfície do entrecasco, associado ao desenvolvimento de fungos oportunistas e a processos de degradação dos tecidos corticais (Figura 3.8).



Figura 3.8 - Sobreiro com presença de carvão do entrecasco, observado durante os inventários florestais.

- Cobrilha, identificada através da observação de galerias, perfurações e outros sinais de atividade de insetos xilófagos na cortiça ou no lenho superficial (Figura 3.9).



Figura 3.9 - Sobreiro com o entrecasco com cobrilha.

- Feridas no tronco, resultantes de operações de descortiçamento, danos mecânicos, constituindo um ponto de entrada para agentes patogénicos.

Para cada árvore foi registada no formulário a presença ou ausência destas anomalias, tendo essa informação sido posteriormente agregada numa variável dependente binária, designada FITOSSA. Assim considerou-se FITOSSA= 0, quando o indivíduo não apresentava quaisquer sintomas visíveis, e FITOSSA= 1, quando apresentava pelo menos uma das três anomalias consideradas, como apresentado no gráfico da Figura 3.10.

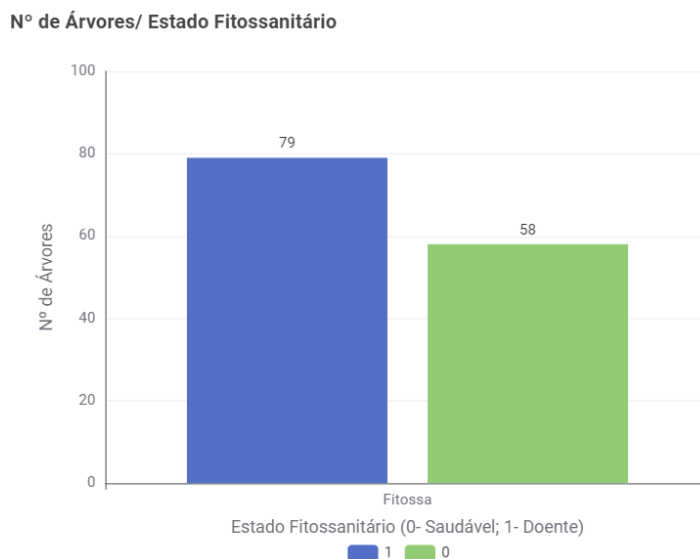


Figura 3.10 - Representação da Contagem de Árvores pelo Estado Fitossanitário

3.2.2. Dados LiDAR

Foram utilizados os dados de LiDAR, disponibilizados pela Direção Geral do Território (DGT, 2024), com cobertura integral de Portugal Continental, incluindo produtos derivados da nuvem de pontos, nomeadamente o Modelo Digital do Terreno (MDT) e o Modelo Digital de Superfície (MDS), ambos com resolução espacial de 0,5 metros.

Estes dados, foram descarregados a partir do Portal da DGT para a área de estudo, com o objetivo de complementar a informação proveniente do inventário florestal tradicional e apoiar a extração das variáveis topográficas. A partir do Modelo Digital do Terreno (MDT) foram derivadas as variáveis altitude, declive, exposição solar das vertentes e a radiação solar anual, uma vez que estas condicionam fatores ambientais como a disponibilidade hídrica, a temperatura, a acumulação de humidade e a exposição à radiação, influenciando o desenvolvimento das árvores e a sua suscetibilidade ao declínio (A. Costa et al., 2010). A inclusão destas variáveis teve como objetivo avaliar a existência de padrões espaciais associados ao estado fitossanitário dos sobreiros e testar o seu contributo para os modelos preditivos.

O processo foi automatizado através de um *script* desenvolvido em *Python* com *ArcPy* (ArcGIS Pro), garantindo a reprodutibilidade metodológica. Os *rasters* do MDT e do MDS foram preparados e uniformizados no sistema de referência espacial ETRS89/PT-TM06 (EPSG:3763). A partir destes modelos foi gerado o Modelo Digital Normalizado de Superfície (*Normalized Digital Surface Model*- nDSM), também frequentemente designado na literatura como *Canopy Height Model* (CHM) (Farajelahi et al., 2023), obtido através da diferença entre o MDS e o MDT (Figura 3.11).

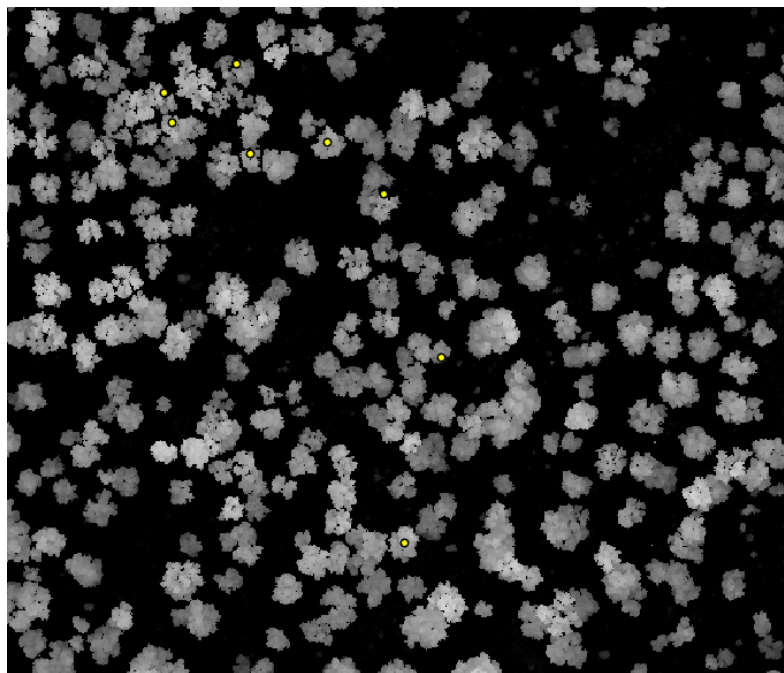


Figura 3.11 - Raster nDSM.

Este procedimento permitiu estimar a altura relativa à vegetação acima do solo, constituindo a base para a análise da estrutura vertical do povoamento. A partir do nDSM foi reclassificado de forma binária considerando-se como vegetação arbórea apenas os pixéis com altura superior a 1 metro, enquanto os restantes valores correspondentes a vegetação rasteira, foram eliminados (Farajelahi et al., 2023). Esta etapa permitiu reduzir o ruído existente nos dados antes da etapa de segmentação, constituindo uma abordagem amplamente utilizada em estudos de deteção automática de copas (Zhen et al., 2016).

A segmentação das copas foi realizada com recurso à ferramenta *Region Group*, que agrupa células contíguas em regiões espacialmente contínuas, interpretadas como copas individuais ou, em alguns casos, como conjuntos de copas adjacentes (Esri, 2024). O princípio deste método assenta na continuidade espacial da vegetação arbórea representada no nDSM, permitindo delimitar unidades de copa a partir da estrutura vertical do coberto.

Após a segmentação inicial, foi aplicada uma etapa de filtragem com a ferramenta *Set Null* do ArcGis Pro, eliminando automaticamente todas as regiões com dimensão inferior ao limiar definido ($COUNT < 20$). Esta operação permitiu remover pequenas regiões resultantes de ruído, artefactos de processamento ou vegetação de reduzida dimensão, reduzindo situações de sobre-segmentação e melhorando a representação das copas individuais (Zhen et al., 2016).

As regiões segmentadas foram posteriormente convertidas para formato vetorial, possibilitando a análise geométrica das copas e a sua integração com a informação proveniente do inventário florestal. A validação foi efetuada através da ferramenta *Select By Location*, selecionando apenas os polígonos que apresentavam interseção



espacial com pelo menos um ponto correspondente aos sobreiros inventariados em campo. Os polígonos selecionados foram exportados para uma nova camada vetorial, correspondente às copas segmentadas com interseção espacial com os pontos do inventário florestal (Figura 3.12). Este procedimento permitiu eliminar regiões segmentadas sem correspondência com árvores reais, aumentando a fiabilidade dos resultados obtidos (Zhen et al., 2016).

Por fim, para cada copa validada foi calculado o respetivo diâmetro equivalente, assumindo-se uma geometria circular com área igual à do polígono correspondente. Este procedimento foi posteriormente utilizado na comparação entre os resultados obtidos por deteção remota e as medições recolhidas durante o inventário florestal.



Figura 3.12 - Resultado da segmentação das copas e correspondência com os pontos do inventário florestal.

3.2.3. Imagens Sentinel- 2

A integração de imagens satélite constituiu um dos elementos centrais desta investigação. As imagens óticas utilizadas foram obtidas através da plataforma *online* de acesso aberto, nomeadamente do *Copernicus Data Space Ecosystem*, que disponibiliza gratuitamente os produtos da missão *Copernicus Sentinel-2*.

Foram utilizadas 3 imagens adquiridas pelo sensor multiespectral (MSI), a bordo dos satélites Sentinel-2A (S2A), Sentinel-2B (S2B) e Sentinel-2C (S2C). A utilização desta constelação justifica-se pela sua elevada resolução espacial e espectral, tal como em estudos anteriores, estas vantagens motivaram igualmente a sua utilização (Laurentino, 2019).

Atualmente, a constelação Sentinel-2 é composta por três satélites que operam de forma complementar, permitindo visitar a mesma área aproximadamente a cada



2 a 3 dias e assegurando uma elevada disponibilidade temporal de imagens. Segundo ESA (2025), os satélites Sentinel-2B e Sentinel-2C operam na mesma órbita, separados por 180°, enquanto o Sentinel-2A se encontra temporariamente deslocado cerca de 36° relativamente ao Sentinel-2B.

O sensor MSI instalado em cada satélite adquire dados em 13 bandas espectrais, distribuídas por três resoluções espaciais distintas (Delwart, 2015):

- 10 metros: B2 (azul), B3 (vermelho), B8 (Infravermelho próximo - NIR);
- 20 metros: B5, B6, B7 (*Red-edge*), B8A (NIR estreito), B11 e B12 (Infravermelho de ondas curtas - SWIR);
- 60 metros: B1, B9 e B10.

Estas bandas abrangem comprimentos de onda compreendidos entre 443 e 2190 nm, cobrindo as regiões do visível, do infravermelho próximo (NIR) e infravermelho de ondas curtas (SWIR), permitindo uma caracterização detalhada vegetação, do solo e das massas de água (ESA, 2025).

A escolha das imagens Sentinel-2 justificou-se pela sua elevada resolução espacial e pela disponibilidade de um maior número de bandas espectrais quando comparadas com a missão *Landsat*, bem como pela resolução significativamente superior à dos dados MODIS (Isbaex et al., 2021). Estas características permitiram calcular diversos índices de vegetação, fundamentais para a caracterização biofísica da área de estudo.

A seleção das imagens Sentinel-2 utilizadas nesta investigação baseou-se dois critérios principais: a proximidade temporal relativamente aos dados de inventário florestal realizados em campo e a minimização da cobertura de nuvens, de forma a garantir a qualidade radiométrica das imagens. Embora a literatura, indique que os valores de reflectância apresentam maior robustez na discriminação do estado fitossanitário das árvores durante a estação seca, em particular devido à menor interferência da humidade do solo e da vegetação herbácea, a seleção temporal foi condicionada pela disponibilidade de imagens e pela sua correspondência com os momentos de inventário (Laurentino, 2019).

As três imagens selecionadas, apresentam uma percentagem de cobertura de nuvens inferior a 0.002%, assegurando que praticamente toda a área de estudo se encontra livre de cobertura das nuvens, garantindo assim a fiabilidade e a consistência dos dados espectrais utilizados nas análises subsequentes.

A Tabela 3.2 apresenta a identificação das imagens Sentinel-2 selecionadas para o processamento.



Tabela 3.2 - Imagens Satélite Sentinel-2 usadas para o processamento.

ID da Imagem	Data	% de nuvens	Satélite
S2A_MSIL2A_20241224T112501_N0511_R037_T29SND_20241224T153355	24 de dezembro de 2024	0.001642%	Sentinel 2 ^a (S2A)
S2B_MSIL2A_20250627T112119_N0511_R037_T29SND_20250627T133415	27 de junho de 2025	0.001098%	Sentinel 2B (S2B)
S2C_MSIL2A_20251229T112511_N0511_R037_T29SND_20251229T125809	29 de dezembro de 2025	0.001802%	Sentinel 2C (S2C)

3.2.3. Detecção Remota

Com base nas imagens Sentinel-2 previamente selecionadas, procedeu-se ao processamento dos dados óticos e ao cálculo de índices espectrais de vegetação. Esta etapa teve como objetivo extrair variáveis associadas ao vigor vegetativo, teor de clorofila e stress hídrico, de forma a apoiar a identificação de padrões relacionados com o estado fitossanitário dos sobreiros, tal como a metodologia utilizada realizado nos estudos de Aubard et al. (2019) e Chrysafis et al. (2020).

Relativamente ao processamento, as bandas originalmente disponibilizadas com 20 m de resolução espacial foram reamostradas para 10 m, recorrendo à plataforma SNAP da ESA. Este procedimento assegurou a uniformização espacial das bandas utilizadas no cálculo dos índices e permitiu uma análise mais consistente.

Posteriormente, foram calculados diversos índices espectrais, recorrendo à plataforma SNAP da ESA (Figura 3.13). Estes índices são amplamente utilizados na monitorização do vigor da vegetação, do teor da clorofila e do stress hídrico, permitindo caracterizar o estado fisiológico do coberto vegetal e identificar alterações estruturais ou funcionais relevantes para a análise fitossanitária.

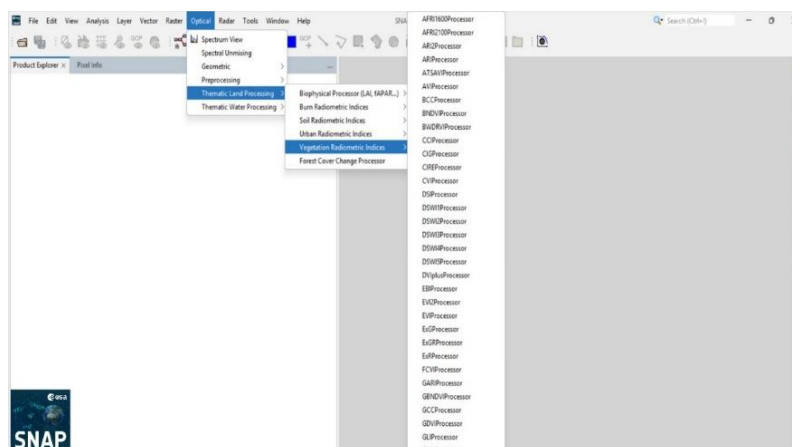


Figura 3.13 - Ferramentas de cálculo de índices espectrais disponíveis no SNAP.

A Tabela 3.3. apresenta os índices de vegetação calculados, as respetivas fórmulas e as aplicações de cada índice utilizado.



Tabela 3.3- Fórmulas dos Índices de Vegetação realizados para a investigação.

Índices de Vegetação	Equação	Aplicação	Referências
Normalized Difference Vegetation Index (NDVI)	$\frac{(NIR - Red)}{(NIR + Red)}$	Monitorização da densidade e vigor da vegetação	(Aubard et al., 2019) (Catalão et al., 2022) (Navarro et al., 2019b) (Laurentino, 2019)
Green Normalized Difference Vegetation index (GNDVI)	$\frac{(NIR - Green)}{(NIR + Green)}$	Estimativa do teor de clorofila foliar, mais sensível que o NDVI para detetar em lugares com vegetação mais densa	(Navarro et al., 2019b)
Modified Soil Adjusted Vegetation Index (MSAVI)	$\frac{(2NIR + 1 - \sqrt{((2NIR + 1)^2 - 8(NIR - RED))})}{2}$	Avaliação da estrutura florestal e da densidade do coberto vegetal	(Estrada et al., 2023)
Modified Soil Adjusted Vegetation Index 2 (MSAVI 2)	$\frac{(2NIR + 1 - \sqrt{((2NIR + 1)^2 - 8(NIR - RED))})}{2}$	Avaliação da estrutura florestal e da densidade do coberto vegetal	(Fabijańczyk & Zawadzki, 2022)
Enhanced Vegetation Index (EVI)	$2.5 \cdot ((NIR - Red)(NIR + 6Red - 7.5Blue + 1))$	Previsão do teor de humidade foliar e monitorização da vegetação	(Laurentino, 2019)
Enhanced Vegetation Index 2 (EVI2)	$2.5 \cdot ((NIR - Red)(NIR + 2.4Red + 1))$	Previsão do teor de humidade foliar e monitorização da vegetação	(Jiang et al., 2008)
Leaf Area Index (LAI)	$LAI = 3.618 \times EVI - 0.118$	Estimativa da área foliar total por unidade de superfície de solo	(Chrysafis et al., 2020b)
Chlorophyll Vegetation Index (CVI)	$NIR \left(\frac{Red}{Green^2} \right)$	Avaliação do teor de clorofila na vegetação	(Vincini et al., 2008)
Moisture Stress Index (MSI)	$\left(\frac{SWIR1}{NIR1} \right)$	Avaliação do stress hídrico na vegetação, valores mais elevados indicam maior défice de água	(Grabska et al., 2020)
Normalized Difference Moisture Index (NDMI)	$\frac{(NIR - SWIR)}{(NIR + SWIR)}$	Monitorização do conteúdo de água na vegetação e no solo	(Navarro et al., 2019b) (Isbaex et al., 2021)
Chlorophyll Index Red-edge (CI)	$\left(\frac{NIR}{(Red - edge)} \right) - 1$	Avaliação precisa do teor de clorofila	(Abdollahnejad & Panagiotidis, 2020)



3.3. Processamento dos Dados no KNIME

Após a preparação dos dados em ambiente SIG, todas as variáveis selecionadas foram integradas num único ficheiro vetorial (*shapefile*), contendo a localização georreferenciada dos 137 sobreiros e os respetivos atributos dendrométricos, índices de vegetação e variáveis biofísicas do terreno.

Este ficheiro constituiu o conjunto de dados de entrada do fluxo de trabalho desenvolvido na plataforma KNIME *Analytics Platform* (v. 5.9.0) (Figura 3.14), onde foram realizadas as etapas de pré-processamento, modelação e avaliação dos algoritmos de *Machine Learning*. A escolha desta plataforma deveu-se à sua capacidade de integrar diferentes fontes de dados e disponibilizar fluxos de trabalho visuais, proporcionando um ambiente flexível para análise de dados e desenvolvimento de modelos de *Machine Learning* (Kalathas et al., 2021).

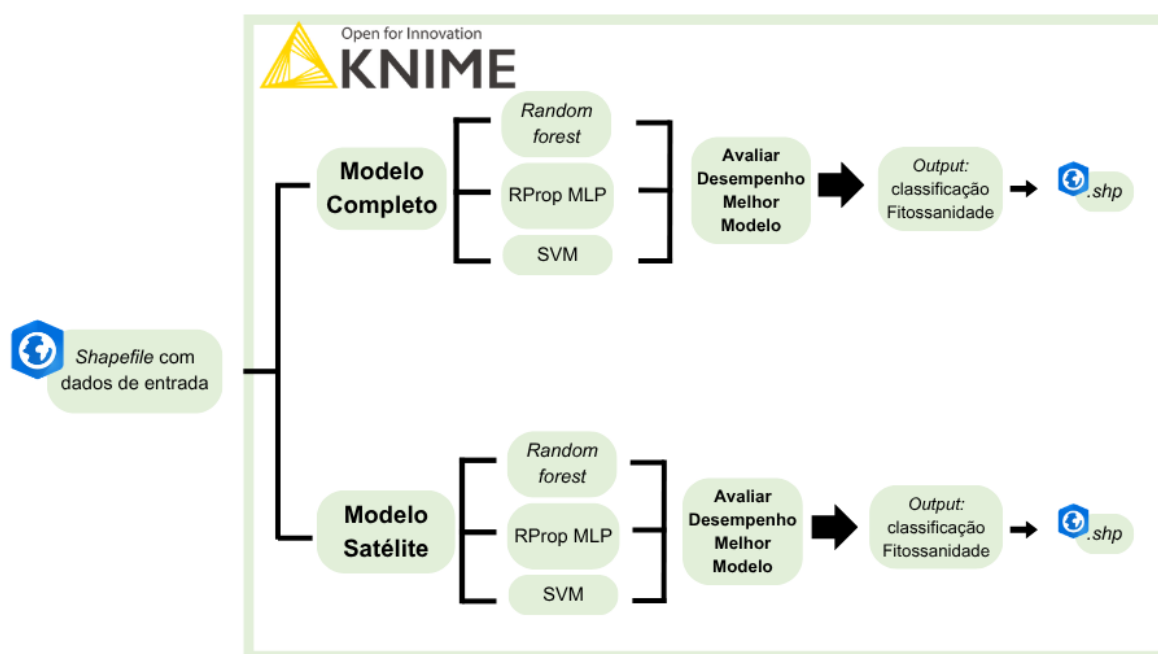
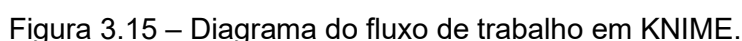
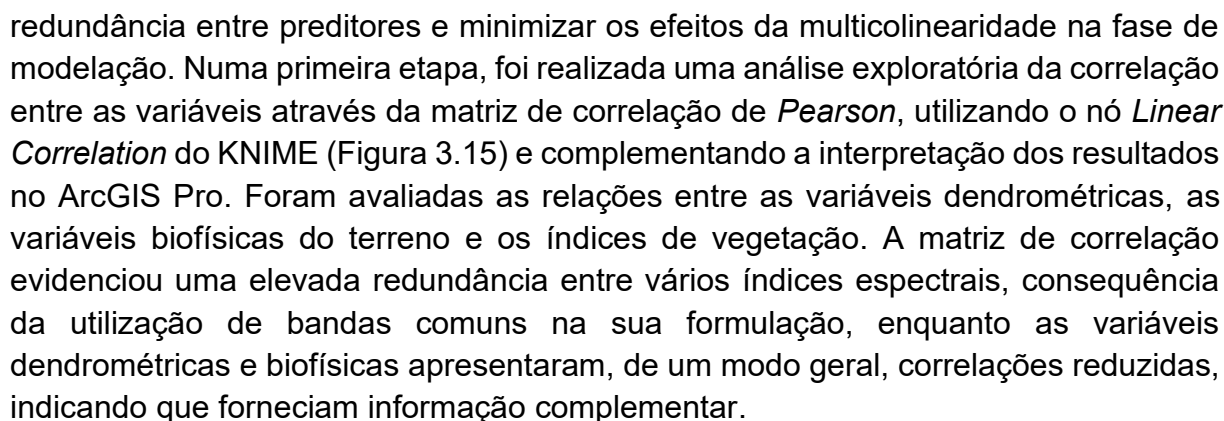


Figura 3.14 - Metodologia realizada no KNIME

3.3.1 Importação e pré processamento dos dados

Os dados foram importados para a plataforma KNIME *Analytics Platform* v.5.9.0 através do nó *Geofile Reader*. Este procedimento permitiu a leitura direta da *shapefile* contendo a localização georreferenciada de cada sobreiro e os respetivos atributos dendrométricos, índices de vegetação e variáveis biofísicas do terreno.

Após a importação, iniciou-se o pré-processamento dos dados, começando pela seleção e validação das variáveis explicativas, com o objetivo de reduzir a



Com base nesta análise, procedeu-se à eliminação das variáveis espectrais com maior redundância, mantendo-se os índices CI, CVI, LAI e GNDVI, por constituírem o conjunto de preditores espectrais mais informativo sem perda significativa de capacidade preditiva. Como etapa complementar, foi efetuado um diagnóstico da multicolinearidade através do Fator de Inflação da Variância (*Variance Inflation Factor* – VIF), recorrendo à ferramenta *Generalized Linear Regression* (GLR) do ArcGIS Pro. Adotou-se como critério operacional um valor de $VIF \leq 10$ (O'Brien, 2007), removendo-se iterativamente as variáveis que excediam este limiar até à obtenção de um conjunto final de preditores estável. Este procedimento é importante, pois permitiu reduzir a redundância e controlar a multicolinearidade residual, preservando a complementaridade entre a informação espectral, as características dendrométricas e os fatores biofísicos do terreno.

Definido o conjunto final de variáveis explicativas, procedeu-se à identificação de valores extremos através do nó *Numeric Outliers*, recorrendo à representação gráfica por diagramas de caixa (*boxplots*). Embora tenham sido identificadas observações



atípicas, estas foram mantidas, por corresponderem a medições obtidas durante a o inventário de campo e refletirem a variabilidade natural da amostra estudada, não existindo indícios de erros de medição ou de registo.

A Figura 3.16 apresenta a distribuição das variáveis através de diagramas de caixa. Observa-se que as variáveis dendrométricas, nomeadamente o DAP, a altura e o diâmetro da copa, apresentam maior dispersão e um número mais elevado de valores atípicos. O DAP evidencia a maior amplitude de valores, refletindo a elevada variabilidade estrutural entre os sobreiros amostrados. Em contraste, os índices de vegetação apresentam distribuições mais concentradas e menor variabilidade. Importa ainda referir que as diferenças de escala entre as variáveis dendrométricas e os índices espectrais fazem com que estes últimos surjam visualmente comprimidos quando representados no mesmo gráfico, sem que tal represente uma menor relevância estatística.

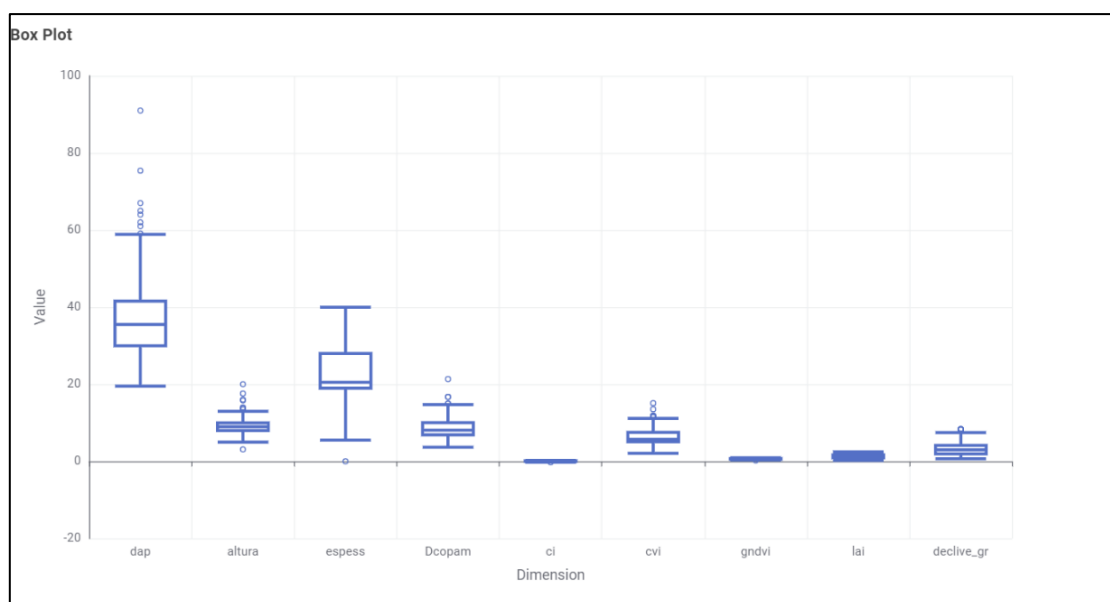


Figura 3.16 - *Box plot* das variáveis utilizadas no estudo para identificação de valores

Posteriormente, a variável dependente *Fitossa* foi convertida do formato numérico para categórico através do nó *Number to String*, passando o valor 1 a representar indivíduos doentes e o valor 0 indivíduos saudáveis. Esta transformação permitiu que a variável fosse corretamente reconhecida como nominal pelos algoritmos de classificação utilizados nas etapas seguintes.

Por fim, todas as variáveis explicativas foram normalizadas através do nó *Normalizer*, transformando os respetivos valores para uma escala comum e reduzindo o impacto das diferenças de magnitude entre os preditores. Este procedimento é particularmente importante para algoritmos sensíveis à escala das variáveis, contribuindo para melhorar o desempenho do processo de aprendizagem e evitando que preditores com amplitudes numéricas superiores exerçam uma influência desproporcionada na modelação (KNIME, 2021).



3.3.2 Partição da Amostra (treino e validação)

De forma a avaliar a capacidade de generalização dos modelos e mitigar o fenómeno de sobreajustamento (*overfitting*), a amostra total de 137 observações foi cindida em subconjuntos independentes de treino e validação através do nó *Table Partitioner*, adotando-se uma proporção clássica de 70% para o treino dos algoritmos e 30% para a validação do desempenho preditivo (Ordenes & Silipo, 2021).

Com o intuito de garantir a comparabilidade estrita entre as abordagens testadas, utilizou-se rigorosamente a mesma partição de dados em todos os fluxos de trabalho, gerados no KNIME. Esta padronização assegurou que os subconjuntos de treino (70%) e validação (30%) fossem matematicamente idênticos em ambos cenários de modelação. Desta forma a discrepância observada no desempenho final dos modelos poderá ser atribuída exclusivamente à inclusão ou ausência das variáveis dendrométricas, eliminando o viés de variações aleatórias na divisão dos dados.

3.3.3 Definição das Abordagens de Modelação e Algoritmos Aplicados

O desenho experimental estruturou-se em torno de duas abordagens de modelação preditiva distintas:

- **Modelo Completo:** Integrou a totalidade das variáveis seleccionadas e validadas nas fases anteriores, nomeadamente os índices de vegetação (CI, CVI, LAI e GNDVI), as variáveis dendrométricas recolhidas *in situ* (DAP, altura, espessura da cortiça, diâmetro da copa e ano de descortiçamento) e as características biofísicas do terreno (altitude, declive, exposição solar e radiação solar).

- **Modelo Satélite:** Construído exclusivamente com variáveis obtidas por deteção remota e derivadas de Sistemas de Informação Geográfica (SIG), prescindindo de dados recolhidos diretamente em campo. Esta abordagem incluiu os índices de vegetação e as variáveis biofísicas derivadas do Modelo Digital do Terreno (MDT). O objetivo desta segunda abordagem consistiu em avaliar o potencial da deteção remota na previsão autónoma da fitossanidade dos sobreiros, testando a viabilidade de dispensa de inventários florestais. Os fluxos de processamento foram implementados na plataforma KNIME, conforme ilustrado na Figura 3.17.

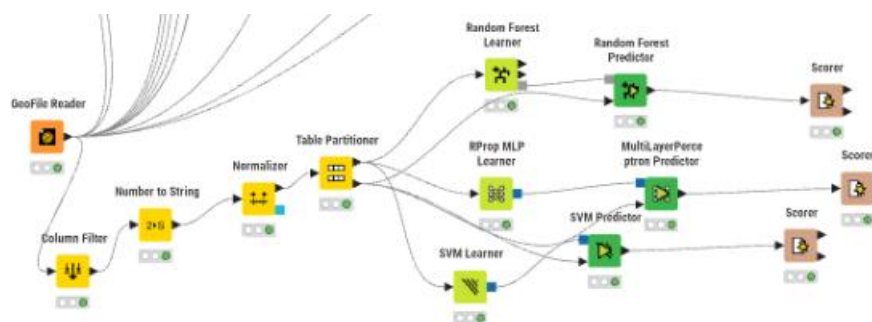


Figura 3.17 - Fluxo de Trabalho no KNIME do Modelo Satélite.



Para a estimativa e classificação da fitossanidade em ambas as abordagens, foram aplicados três algoritmos de aprendizagem automática (*Machine Learning*):

1. *Random Forest (RF)*: Algoritmo baseado na construção de múltiplas árvores de decisão independentes, treinadas com amostras aleatórias (*bootstrap*) e subconjuntos restritos de variáveis. A classificação final resulta da agregação e votação maioritária das decisões individuais de cada árvore individual (Tripathi et al., 2023), o que confere ao modelo uma elevada robustez contra o *overfitting* e um ótimo desempenho perante variáveis categóricas e dados ruidosos (Navarro et al., 2019; Nasiri et al., 2022).

2. *RProp MLP*: Consiste numa rede neuronal artificial do tipo *feedforward*, composta por uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída. O processo de aprendizagem é realizado através do algoritmo *Resilient Backpropagation* (RProp), que atualiza os pesos da rede com base apenas no sinal do gradiente da função de erro, ajustando individualmente o tamanho dos passos de atualização para cada peso. Esta estratégia permite acelerar a convergência durante o treino, reduzir a influência da magnitude do gradiente e melhorar a estabilidade da otimização. As redes MLP (*Multilayer Perceptron*) são particularmente eficazes na modelação de relações complexas e não lineares entre as variáveis independentes e a variável resposta, sendo amplamente utilizadas em problemas de classificação e previsão (Kazanskiy et al., 2025).

3. *Support Vector Machine (SVM)*: Procura determinar o hiperplano ótimo que maximiza a margem de separação entre classes (saudável vs doente) no espaço multidimensional de variáveis independentes. Através da utilização de funções *kernel*, o algoritmo pode projetar dados para espaços de maior dimensão, permitindo separar classes que não são linearmente separáveis no espaço original. O SVM destaca-se pelo seu elevado desempenho em problemas de classificação com elevada dimensionalidade e conjuntos de dados de dimensão reduzida ou moderada (Duarte et al., 2021).

3.3.4 Avaliação do Desempenho Preditivo dos Modelos

A validação cruzada e a comparação do desempenho preditivo dos algoritmos de aprendizagem máquina basearam-se em técnicas de validação cruzada e em métricas de avaliação derivadas da matriz de confusão, seguindo abordagens amplamente utilizadas na literatura (Tripathi et al., 2023; Carvacho, s.d.). Estas metodologias permitem uma avaliação robusta da capacidade de generalização dos modelos, reduzindo o risco de *overfitting* e mitigando o viés associado à divisão única dos dados em treino e teste.

Para avaliar o comportamento de cada modelo, foram calculados e analisados os seguintes indicadores estatísticos:



- Exatidão Global (*Accuracy*, Q): representa a proporção de previsões corretas (classes positivas e negativas) relativamente ao total de observações. Embora seja uma métrica intuitiva, pode ser influenciada por desequilíbrios entre classes;
- Validação Cruzada (*k-fold Cross-Validation*): técnica de reamostragem em que o conjunto de dados é dividido em k subconjuntos (*folds*). O modelo é treinado k vezes, utilizando $k-1$ *folds* para treino e um *fold* para validação, sendo o desempenho final obtido pela média dos resultados. Neste estudo foram utilizadas variantes de 5 e 10 *folds*, reconhecidas por fornecer um bom equilíbrio entre viés e variância da estimativa (Kohavi, 1995);
- Precisão (*Precision*, PR): mede a proporção de verdadeiros positivos entre todas as observações classificadas como positivas pelo modelo, sendo especialmente relevante na redução de falsos positivos;
- Sensibilidade (*Recall*, RC): quantifica a capacidade do modelo em identificar corretamente os casos positivos reais, minimizando falsos negativos, sendo particularmente importante em problemas de deteção de estados de doença;
- *F- Measure* (F1- Score): Média harmónica entre a precisão e a sensibilidade, funcionando como um indicador robusto do equilíbrio geral do modelo, especialmente útil face a distribuições de classes potencialmente não equilibradas.



4. RESULTADOS

4.1. Análise da Correlação entre variáveis para os modelos

Para a análise da correlação entre as variáveis biofísicas e o estado fitossanitário dos sobreiros, recorreu-se ao ambiente ArcGis Pro e ao ambiente KNIME. Utilizou-se o ficheiro em formato vetorial dos sobreiros contendo a geolocalização individual de cada exemplar, à qual foram indexados os respetivos atributos dendrométricos, ambientais e os índices de vegetação calculados, conforme sistematizado na Tabela 4.1.

Tabela 4.1- Variáveis do Estudo.

Categoria	Variáveis Incluídas no Modelo	Unidades
Variável Dependente	Fitossanidade (FITOSSA)	Categórica Binária (1-Doente; 0- Saudável)
Dendrométricas e Fisiológicas	DAP, Altura, Espessura da Cortiça (ESPESS), Diâmetro da Copa (DCOPAM), Ano de descortiçamento (ANODESC)	Centímetros (cm), Milímetros (mm), Metros (m), Anos
Sintomas Visuais de Doença	Cobrilha, Carvão, Feridas	Categórica Binária (1-Sim; 0- Não)
Índices de Vegetação	CVI, EVI, EVI2, GNDVI, LAI, MSAVI, MSAVI2, MSI, NDMI, NDVI, CI	Contínua/Numérica (adimensional)
Biofísicas do Terreno	Altitude, Exposição Solar (ASPECT_50), Radiação Solar (S_RADIATIO), DECLIVE_GR	Contínua/Numérica (m, °, Wh/m2, °)

Com o intuito de avaliar o comportamento conjunto das variáveis e evitar multicolineariedade (redundância de dados) nos modelos subsequentes, realizou-se uma análise de correlação preliminar na plataforma *KNIME Analytics Platform*, recorrendo ao nó *Linear Correlation*. Nesta etapa, cruzaram-se todas as variáveis do estudo: as dendrométricas (DAP, altura, diâmetro da copa, espessura da cortiça), as topográficas e ambientais (altitude, declive, exposição solar e radiação solar) e os diferentes índices de vegetação.

Enquanto as variáveis dendrométricas e as biofísicas do terreno demonstraram fraca correlação entre si, a análise da matriz de correlação de *Pearson* revelou uma redundância severa e expetável no bloco dos índices de vegetação, com coeficientes de correlação linear frequentemente superiores a 0,90, como apresentada na Tabela 4.2. Este comportamento decorre do facto destes índices partilharem as mesmas bandas espectrais na sua fórmula matemática.

Face a este cenário, e em linha com as recomendações de (Tripathi et al., 2023), procedeu-se a uma seleção criteriosa destas variáveis para simplificar a arquitetura do modelo sem comprometer o desempenho preditivo.



Tabela 4.2 - Matriz de Correlação dos índices de vegetação.

Índice	CI	CVI	EVI2	EVI	GNDVI	LAI	MSAVI	MSAVI 2	MSI	NDMI	NDVI
CI	1,000	-0,180	-0,491	-0,323	-0,636	-0,677	-0,417	-0,460	0,711	-0,731	-0,828
CVI	-0,180	1,000	0,148	-0,004	0,786	0,405	0,079	0,113	-0,400	0,419	0,649
EVI2	-0,491	0,148	1,000	0,972	0,351	0,896	0,994	0,998	-0,746	0,733	0,510
EVI	-0,323	-0,004	0,972	1,000	0,147	0,785	0,992	0,982	-0,618	0,595	0,300
GNDVI	-0,636	0,786	0,351	0,147	1,000	0,665	0,257	0,304	-0,587	0,614	0,930
LAI	-0,677	0,405	0,896	0,785	0,665	1,000	0,851	0,877	-0,790	0,796	0,776
MSAVI	-0,417	0,079	0,994	0,992	0,257	0,851	1,000	0,998	-0,693	0,677	0,415
MSAVI2	-0,400	0,113	0,998	0,982	0,304	0,877	0,998	1,000	-0,725	0,712	0,465
MSI	0,711	-0,400	-0,746	-0,618	-0,587	-0,790	-0,693	-0,725	1,000	-0,988	-0,760
NDMI	-0,731	0,419	0,733	0,595	0,614	0,796	0,677	0,712	-0,988	1,000	0,781
NDVI	-0,828	0,649	0,510	0,300	0,930	0,776	0,415	0,465	-0,760	0,781	1,000

O processo de seleção orientou-se pela eliminação da redundância estatística e pela maximização da complementaridade da informação biológica. Uma vez que os índices de vegetação eram os que registavam os maiores níveis de colinearidade global, o critério de exclusão baseou-se na identificação dos pares com maior dependência linear. Consequentemente, dado que o NDVI e o EVI2 exibiram uma correlação extremamente elevada com o MSAVI e o MSAVI2, optou-se pela exclusão destes indicadores. Assim, o alinhamento final de preditores espectrais ficou circunscrito aos índices CI, CVI, LAI e GNDVI, os quais em conjunto com as variáveis dendrométricas e as características biofísicas do terreno descritas anteriormente, integram a arquitetura dos modelos definidos.

Posteriormente, e com o objetivo de detetar problemas latentes de colinearidade que pudessem ter resistido à análise bivariada inicial, utilizou-se a ferramenta *Generalized Linear Regression* (GLR) em ambiente ArcGIS Pro como uma ferramenta matemática de diagnóstico. O critério teórico adotado para a retenção ou exclusão definitiva das variáveis baseou-se no cálculo do Fator de Inflação da Variância (VIF). Como critério de decisão, estabeleceu-se que as variáveis que manifestassem redundância informacional severa ou que pudessem introduzir viés de causalidade no modelo seriam progressivamente removidas.



4.1.1 Análise da Autocorrelação Espacial

Complementarmente à análise da multicolinearidade, foi realizada a análise de autocorrelação espacial, também em ambiente ArcGis Pro, com base no procedimento apresentado por Costa (2024). Esta foi realizada com recurso à avaliação do modelo *Multiscale Geographically Weighted Regression* (MGWR), com o objetivo de verificar se os resíduos apresentavam dependência espacial significativa (Yu et al., 2020).

O MGWR permite estimar cada variável explicativa a diferentes escalas espaciais através da atribuição de um *bandwidth* específico. Desta forma, considera-se que os processos espaciais podem atuar em diferentes escalas, proporcionando uma representação mais realista da heterogeneidade espacial (Yu et al., 2020). A definição da vizinhança espacial foi realizada através do método *Number of Neighbors*, com determinação do número ótimo de vizinhos pelo método *Golden Search*, que identifica iterativamente a escala espacial que maximiza o desempenho do modelo.

Após a estimação do MGWR, foram analisados os resíduos padronizados através do teste *Spatial Autocorrelation (Global Moran's I)*, de forma a verificar se existiam padrões espaciais nos erros. Da qual se obteve os seguintes resultados, apresentados na figura infra.

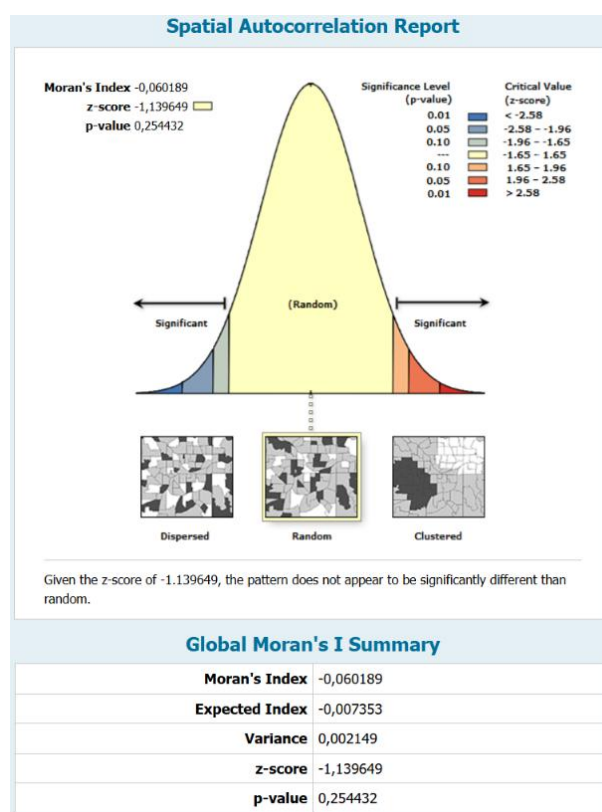


Figura 4.1 - Resultados do Índice *Moran* no ArcGis Pro



O Índice de *Moran*, com valor de -0,060, indicou uma ligeira tendência para autocorrelação espacial negativa, mas próxima de zero. Contudo, o valor de *p-value* = 0,254, é superior a 0,05 e o *z-score* encontra-se dentro do intervalo esperado para uma distribuição aleatória.

4.2. Análise Exploratória das Variáveis

Procedeu-se à modelação através da ferramenta *Generalized Linear Regression* (GLR) em ambiente ArcGIS Pro, definindo a fitossanidade como variável dependente binária, com base nas 137 observações recolhidas em campo. Esta etapa foi realizada exclusivamente para verificar o Fator de Inflação de Variância (VIF) de forma a diagnosticar a potencial multicolineariedade entre os preditores, servindo apenas como procedimento diagnóstico para aferir a adequabilidade e a qualidade do ajustamento do modelo global. A multicolinearidade global pode ser avaliada pelo Fator de Inflação de Variância (VIF) de cada variável explicativa. A regra geral de 10 é um limiar comumente usado para identificar a presença de multicolinearidade grave, mas alguns autores usam uma regra mais conservadora de 4 (O'brien, 2007). Neste projeto é considerado um limiar de 10, portanto preditores com $VIF > 10$ devem ser removidos, um a um, até que os valores de VIF para todas as variáveis explicativas restantes estejam abaixo desse limiar.

No refinamento do modelo através da exclusão dos índices com elevada redundância informacional – nomeadamente EVI, EVI2, MSAVI, MSAVI2, NDMI, NDVI e MSI. Foram igualmente removidas as variáveis COBRILHA e CARVAO, uma vez que a sua inclusão introduziria viés no modelo ao indicarem presença de doença. O modelo resultante reteve treze variáveis explicativas, todas com $VIF < 10$, como demonstra a Tabela 4.3. O coeficiente de determinação ajustado indica que o modelo explica aproximadamente 20,5% da variância total da “Fitossa” valor que, não obstante ser moderado, é coerente com a complexidade ecológica inerente à modelação de variáveis biofísicas em contexto florestal. A análise dos coeficientes individuais, revelou uma distinção clara entre o impacto dos índices de vegetação e o das métricas dendrométricas estruturais.



Tabela 4.3 - VIF obtido do GLR.

Variável	Coeficiente	Erro Padrão	Estatística t	VIF
DAP	0,005	0,005	0,903	2,431
ALTURA	0,031	0,024	1,279	2,057
ESPESS	0,010	0,006	1,696	1,253
DCOPAM	-0,019	0,020	-0,944	2,080
ANODESC	0,012	0,011	1,095	1,463
CI	-1,453	0,822	-1,768	4,414
CVI	-0,024	0,041	-0,591	5,328
GNDVI	1,969	1,013	1,943	8,354
LAI	-0,308	0,134	-2,305	2,405
ALTITUDE	0,000	0,003	-0,066	1,281
ASPECT_50	0,000	0,000	0,149	1,247
S_RADIATIO	-0,003	0,002	-1,813	1,551
DECLIVE_GR	-0,005	0,024	-0,209	1,242

Entre as treze variáveis independentes incluídas no modelo, destacam-se o LAI, CI, GNDVI com estatísticas t mais elevados, atingindo o limiar de significância a $p < 0,05$.

Posteriormente, as variáveis selecionadas (as variáveis da Tabela 4.1), foram utilizadas como *inputs* nos modelos de aprendizagem automática desenvolvidos no KNIME, descritos no subcapítulo seguinte.

4.3. Modelos de Aprendizagem Máquina (*Machine Learning*)

4.3.1. Abordagem 1: Modelo Completo

A análise comparativa dos três algoritmos de aprendizagem máquina testados — *Support Vector Machine* (SVM), *RProp MLP* e *Random Forest* (RF) — revelou distinções significativas na capacidade de generalização e robustez preditiva entre os modelos, quando aplicados ao conjunto completo de variáveis explicativas, nomeadamente os índices de vegetação CI, CVI, LAI e GNDVI, combinados com as métricas dendrométricas e biofísicas do terreno.

Os resultados detalhados encontram-se apresentados na Tabela 4.4.



Tabela 4.4 - Métricas de Desempenho preditivo dos algoritmos do Modelo Completo.

Modelo	Classe	Precision	F-measure	Accuracy	Cohen's Kappa
SVM	0- Não doente	0.778	0.778	0.81	0.611
	1- Doente	0.833	0.833		
RProp MLP	0- Não doente	0.611	0.667	0.738	0.454
	1- Doente	0.833	0.784		
Random Forest	0- Não doente	0.778	0.757	0.786	0.566
	Classe	0.792	0.826		

O modelo *Support Vector Machine* (SVM), obteve o melhor desempenho global entre os algoritmos avaliados, alcançando uma Accuracy de 81,0% e o índice de *Cohen's Kappa* de 0,61. Adicionalmente, apresentou valores equilibrados de Precision e F-measure para ambas as classes, destacando-se especialmente na identificação da classe Doente com uma Precision e F-measure de 0,833.

De forma a complementar a avaliação quantitativa do desempenho do modelo SVM, procedeu-se a uma análise espacial, comparando a distribuição geográfica do estado fitossanitário observado em campo (Figura 4.2) com a distribuição prevista pelo modelo (Figura 4.3).

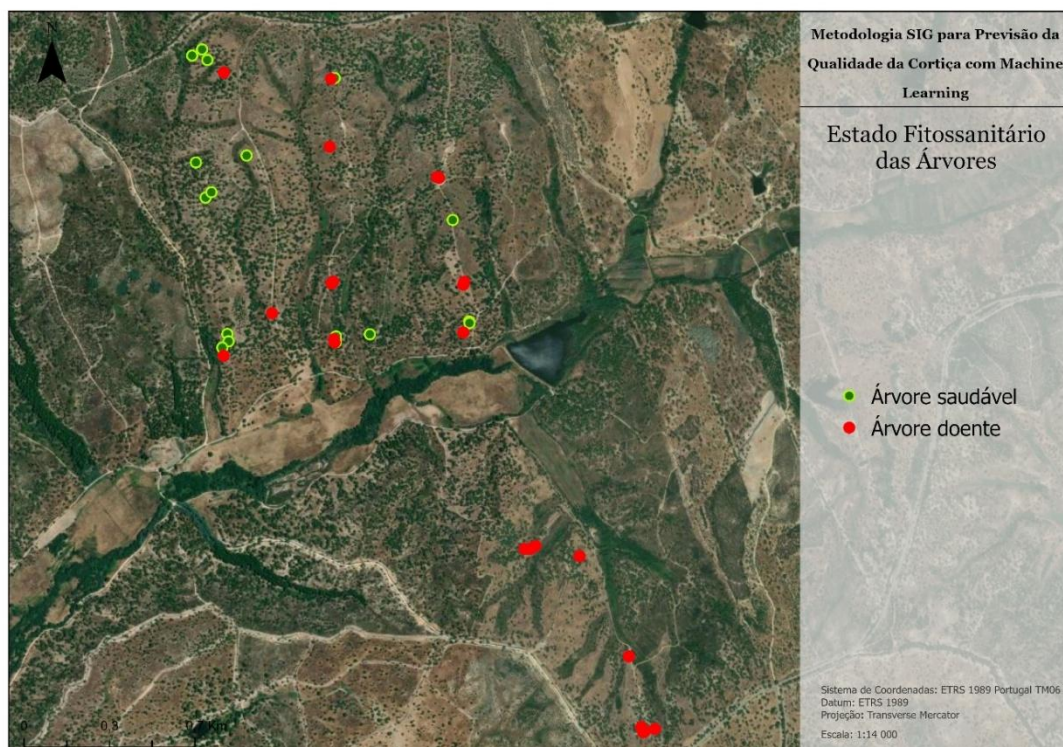


Figura 4.2 - Estado Fitossanitários das Árvores.

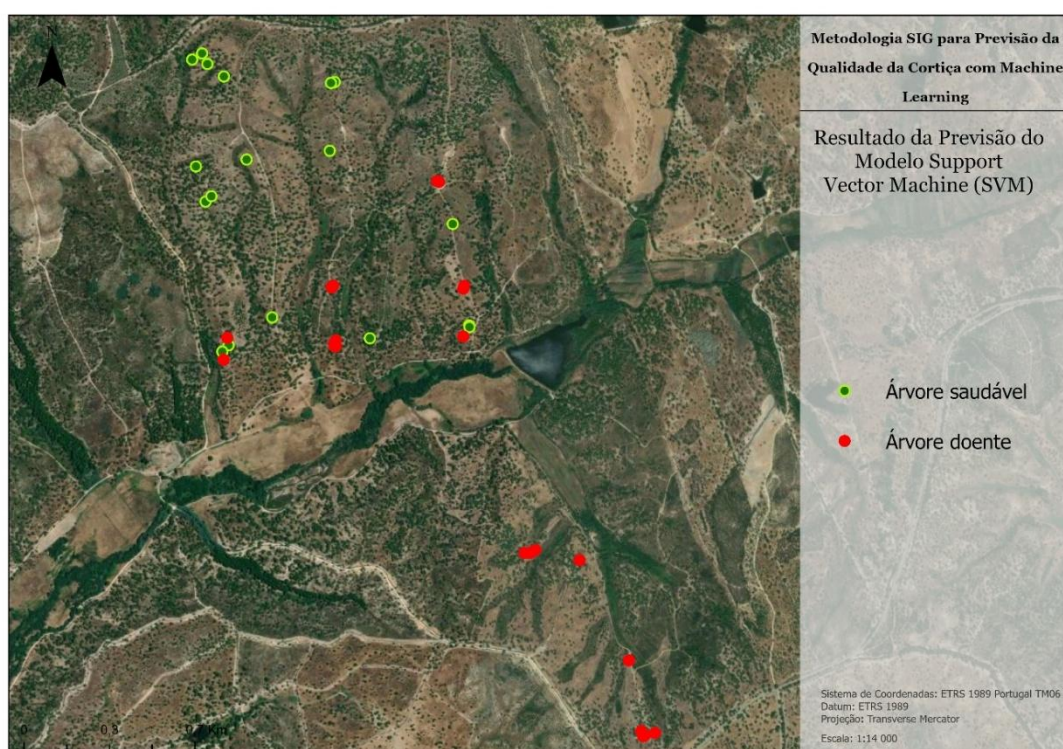


Figura 4.3 - Resultado da Previsão do Modelo SVM.

No algoritmo *Support Vector Machine* (SVM) foi utilizada a função *Kernel Radial Basis Function* (RBF), tendo sido seleccionado um valor de σ (*sigma*) igual a 0,3. A



escolha deste parâmetro resultou de um processo experimental em que foram testados diferentes valores de sigma. Os resultados mostraram que valores de sigma superiores ou inferiores a 0,3 conduziram a um desempenho significativamente inferior, apresentando valores de *Cohen's Kappa* próximos de 0, evidenciando uma reduzida capacidade de generalização.

O algoritmo *Random Forest* (RF) apresentou o segundo melhor desempenho entre os modelos avaliados, registando uma *Accuracy* de 78,6% e um índice de *Cohen's Kappa* de 0,566, indicando uma concordância moderada e substancial entre as classificações previstas e as observadas. Embora tenha apresentado um desempenho ligeiramente inferior ao SVM, o RF demonstrou um comportamento bastante consistente. Relativamente às métricas por classe, o modelo obteve uma *Precision* de 0,778 e uma *F-measure* de 0,757 para a classe “Não Doente”, enquanto para a classe “Doente” alcançou uma *Precision* de 0,792 e uma *F-measure* de 0,826. Estes resultados evidenciam um desempenho equilibrado entre ambas as classes, demonstrando uma boa capacidade de identificação dos indivíduos doentes sem comprometer significativamente a classificação dos indivíduos não doentes.

4.3.2. Abordagem 2: Modelo Satélite

Complementarmente ao modelo completo, foi desenvolvida uma abordagem alternativa designada “Modelo Satélite”, com o objetivo de avaliar a capacidade preditiva dos algoritmos na ausência absoluta de dados recolhidos *in situ*. Neste cenário, as variáveis dendrométricas de medição direta (DAP, altura, ano de descortçamento e espessura da cortiça) foram excluídas, sendo o modelo alimentado exclusivamente por variáveis derivadas de deteção remota e de Sistemas de Informação Geográfica (SIG).

Foram integrados os índices espectrais (CI, CVI, LAI e GNDVI) e as variáveis biofísicas do terreno derivadas do Modelo Digital de Terreno (altitude, declive, exposição solar e radiação solar). O diâmetro da copa (DCOPAM) foi mantido, uma vez que pode ser estimado por dados LiDAR, sem necessidade de inventário florestal tradicional.

Os três algoritmos de aprendizagem automática foram replicados sob o mesmo protocolo experimental, incluindo pré-processamento e partição amostral (70% treino/ 30% validação), garantindo comparabilidade direta com o modelo completo. Os resultados obtidos (Tabela 4.5) demonstram que, mesmo na ausência de variáveis *in situ* os modelos mantêm um desempenho preditivo globalmente sólido.



Tabela 4.5 - Métricas de Desempenho preditivo dos algoritmos no Modelo Satélite.

Modelo	Classe	Precision	F-measure	Accuracy	Cohen's Kappa
SVM	0- Não doente	0.556	0.606	0.69	0.355
	1- Doente	0.792	0.745		
RProp MLP	0- Não doente	0.889	0.821	0.833	0.667
	1- Doente	0.792	0.844		
Random Forest	0- Não doente	0.722	0.703	0.738	0.469
	Classe	0.7	0.766		

O algoritmo *Support Vector Machine* (SVM), apresentou uma performance moderada, com um Accuracy de 0,69 e índice de *Cohen's Kappa* de 0,355, evidenciando uma ligeira melhoria na classe “Não doente” com um *F-measure* de 0,606, embora mantendo maior robustez na classe “Doente” com um *F-measure* de 0,745.

O RProp MLP destacou-se como modelo com melhor desempenho global neste cenário, atingindo uma Accuracy de 0,833 e índice de *Cohen's Kappa* de 0,667, evidenciando um comportamento mais equilibrado entre classes, com especial destaque para a classe “Doente” com um *F-measure* de 0,844.

O *Random Forest* apresentou desempenho intermédio, com uma Accuracy de 0,738 e índice de *Cohen's Kappa* de 0,469, relevando um comportamento estável e relativamente equilibrado entre classes.

Globalmente, os resultados indicam que a informação dos índices de vegetação (CI, CVI, LAI e GNDVI) e as variáveis biofísicas são os principais determinantes da classificação da fitossanidade do montado, sendo que as variáveis dendrométricas individuais contribuem de forma marginal para este objetivo específico.

Do ponto de vista operacional, o Modelo Satélite demonstra elevado potencial como ferramenta custo-eficaz de monitorização florestal, permitindo estimar o estado fitossanitário do sobreiro com base exclusivamente em dados geoespaciais de acesso livre, viabilizando monitorização contínua e de larga escala.



5. DISCUSSÃO

Os resultados obtidos evidenciaram diferenças na capacidade preditiva dos três algoritmos de aprendizagem automática em função do conjunto de variáveis utilizado. No Modelo Completo, que integrou variáveis dendrométricas, dados LiDAR e índices de vegetação derivados do Sentinel-2, o algoritmo *Support Vector Machine* (SVM) apresentou o melhor desempenho global, alcançando uma exatidão de 81%.

O bom desempenho do SVM no Modelo Completo pode ser explicado pelas características do próprio algoritmo. O SVM foi desenvolvido para maximizar a separação entre classes, apresentando elevada capacidade de generalização em conjuntos de dados de dimensão reduzida e elevada dimensionalidade (Duarte et al., 2021). Considerando que esta investigação utilizou apenas 137 árvores, das quais 58% são doentes e um conjunto relativamente diversificado de variáveis biofísicas do terreno, espectrais e dendrométricas, é expectável que o modelo SVM tenha conseguido construir fronteiras de decisão mais robustas, reduzindo o risco de sobreajuste e proporcionando uma classificação equilibrada entre as classes "Doente" e "Não doente".

O Modelo Satélite, baseado exclusivamente em variáveis espectrais derivadas de deteção remota, o RProp MLP obteve o melhor desempenho, atingindo uma exatidão de 83,3%. Este resultado sugere que as relações entre os índices espectrais derivados do Sentinel-2 e o estado fitossanitário dos sobreiros apresentam um comportamento fortemente não linear, favorecendo modelos baseados em redes neuronais artificiais. Segundo Kazanskiy et al. (2025), as redes MLP treinadas através do algoritmo *Resilient Backpropagation* são particularmente eficazes na modelação de relações complexas entre a variável dependente e as variáveis explicativas, uma vez que ajustam os pesos da rede com base apenas no sinal do gradiente da função de erro, acelerando a convergência e melhorando a estabilidade do processo de optimização. A superioridade do RProp MLP no Modelo Satélite poderá indicar que os índices de vegetação captam padrões fisiológicos subtis, cuja relação com o estado fitossanitário dificilmente é representada por fronteiras de decisão mais simples.

Um dos resultados mais relevantes desta investigação foi o facto de o Modelo Satélite apresentar um desempenho ligeiramente superior ao Modelo Completo, contrariando a expectativa inicial de que a integração de variáveis dendrométricas medidas em campo melhoraria a capacidade preditiva. Este comportamento sugere que as variáveis dendrométricas individuais contribuíram com informação relativamente reduzida para a classificação do estado fitossanitário, podendo inclusive introduzir alguma redundância ou ruído estatístico. Pelo contrário, os índices espectrais derivados do Sentinel-2 revelaram-se suficientemente informativos para descrever as alterações fisiológicas associadas ao stress das árvores.

Esta interpretação é reforçada pela análise da importância das variáveis, na qual os índices CI, CVI, LAI e GNDVI surgiram como os principais preditores do estado



fitossanitário. Estes índices encontram-se diretamente relacionados com propriedades fisiológicas da vegetação, nomeadamente a concentração de clorofila, o teor de azoto, a atividade fotossintética e a densidade foliar, parâmetros que tendem a responder precocemente a situações de stress hídrico ou fitossanitário. Em contraste, variáveis dendrométricas como o diâmetro ou a altura representam características estruturais relativamente estáveis ao longo do tempo, refletindo alterações apenas em fases mais avançadas do declínio das árvores, podendo estas serem detetadas ao fim de vários anos.

Embora a proposta inicial contemplasse a utilização de LiDAR móvel (MLS), a fase exploratória revelou limitações práticas que condicionaram a sua aplicação nesta investigação. Para realizar um levantamento suficientemente robusto e com uma quantidade de árvores suficientes, seriam necessários uma grande quantidade de recursos, tanto humanos como em tempo, tornando inviável a sua integração de forma robusta no âmbito deste trabalho. Estas limitações estão de acordo com a literatura, que identifica o MLS como uma tecnologia de elevada precisão, mas também exigente em termos de processamento e de definição de procedimentos adequados, sobretudo em ambientes florestais complexos (Laino et al., 2024).

A substituição do MLS pelo LiDAR aerotransportado (ALS) revelou-se metodologicamente adequada aos objetivos desta investigação. Enquanto o MLS privilegia a caracterização detalhada de troncos e da estrutura ao nível da árvore individual, o ALS oferece uma representação mais abrangente e contínua da estrutura do dossel, permitindo uma caracterização mais consistente das copas dos sobreiros em grandes áreas. Uma vez que este estudo incidiu na análise do estado fitossanitário através da informação espectral e estrutural das copas, os dados ALS constituíram uma fonte de informação particularmente relevante, proporcionando uma melhor contextualização espacial e facilitando a integração com os índices de vegetação derivados das imagens Sentinel-2 (Xi et al., 2024).

Os resultados do modelo MGWR evidenciaram uma ligeira significância estatística evidenciando que os resíduos podem apresentar alguma correlação espacial, no entanto, quase insignificante, apesar, do diagnóstico da multicolineariedade e da eliminação das variáveis com VIF superior a 10. Este resultado poderia eventualmente ser melhorado caso se admitisse um limite de VIF inferior a 7,5 (O'Brien, 2007).

Em suma, esta investigação apresentou algumas limitações, nomeadamente a resolução espacial de 10 m das imagens Sentinel-2, que pode ser insuficiente para análises à escala da árvore individual, devido ao efeito de mistura espectral entre a copa, vegetação envolvente e solo. Esta limitação poderá reduzir a capacidade discriminativa dos índices de vegetação e afetar o desempenho dos modelos preditivos. Além disso, o reduzido número de árvores inventariadas condicionou o conjunto de treino, podendo limitar a generalização dos modelos para outras áreas de montado.



6. CONCLUSÕES E INVESTIGAÇÃO FUTURA

O presente trabalho demonstrou a viabilidade da integração de tecnologias de Sistemas de Informação Geográfica (SIG) e algoritmos de *Machine Learning* para o diagnóstico fitossanitário não invasivo de montados de sobro, constituindo uma abordagem com elevado potencial para apoiar a previsão indireta da qualidade da cortiça. Os resultados obtidos evidenciaram que o estado de saúde das árvores pode ser explicado de forma consistente através de índices de vegetação derivados de satélites, os quais apresentaram maior capacidade preditiva do que as variáveis dendrométricas recolhidas em campo.

Os índices analisados, o CI, CVI, LAI e GNDVI revelaram-se os preditores mais relevantes, pela sua elevada importância estatística e capacidade para discriminar árvores saudáveis de árvores em declínio. Estes resultados confirmam o potencial da informação espectral para a monitorização do estado fitossanitário dos montados.

Relativamente aos modelos de *Machine Learning*, verificou-se que o algoritmo *Support Vector Machine* (SVM) apresentou o melhor desempenho no Modelo Completo, atingindo uma exatidão de 81%. Contudo, quando apenas foram utilizadas variáveis provenientes da deteção remota (Modelo Satélite), o algoritmo RProp MLP revelou um desempenho superior, alcançando uma exatidão de 83%, enquanto o SVM obteve 69%. Este resultado demonstra que a utilização exclusiva de índices de vegetação constitui uma abordagem robusta para a classificação do estado fitossanitário, reduzindo simultaneamente a necessidade de recolha de dados dendrométricos em campo, frequentemente mais demorados e dispendiosos.

Os resultados obtidos reforçam igualmente o potencial da utilização de dados gratuitos, como as imagens Sentinel-2 e os Modelos Digitais de Terreno/Superfície disponibilizados pela Direção-Geral do Território, para o desenvolvimento de sistemas de monitorização contínua e de baixo custo aplicáveis à gestão florestal. Esta abordagem poderá facilitar a implementação de programas de acompanhamento fitossanitário em larga escala, contribuindo para uma gestão mais eficiente dos montados.

A relevância destes resultados ultrapassa a componente metodológica. A produtividade e a qualidade da cortiça dependem diretamente do vigor vegetativo dos sobreiros, sendo afetadas por fatores como o declínio, o stress hídrico, pragas e doenças. Assim, a deteção precoce de alterações no estado fitossanitário constitui uma ferramenta importante para apoiar a tomada de decisão na gestão florestal, permitindo planejar intervenções atempadas que contribuam para preservar a produtividade, a qualidade da cortiça e a sustentabilidade económica e ecológica dos montados.

Como linhas de investigação futura, recomenda-se, a integração sistemática de dados LiDAR, explorando a complementaridade entre LiDAR aerotransportado (ALS)



e LiDAR móvel ou terrestre (MLS/TLS). Enquanto o ALS permite caracterizar a estrutura das copas e realizar monitorização em grandes áreas, o MLS/TLS poderá fornecer informação estrutural detalhada ao nível da árvore individual, permitindo validar e complementar os modelos desenvolvidos.

Recomenda-se a utilização de imagens de satélite de maior resolução espacial, uma vez que a resolução de 10 m do Sentinel-2 constitui uma limitação na análise à escala da árvore individual. Sensores com pixels de menor dimensão poderão reduzir os efeitos de mistura espectral e melhorar a extração de informação relativa às copas dos sobreiros, contribuindo para um aumento da precisão dos modelos de previsão do estado fitossanitário. Pois os resultados obtidos para efeitos de monitorização fitossanitária, constituem um indicador mais sensível às condições fisiológicas dos sobreiros, do que as métricas dendrométricas tradicionalmente utilizadas em inventários florestais.

Sugere-se igualmente o desenvolvimento de metodologias avançadas em software especializado em processamento de nuvens de pontos, com vista à segmentação automática de árvores individuais e à extração de métricas estruturais diretamente das nuvens de pontos, um procedimento menos moroso, ainda assim com mais detalhe, permitindo a obtenção de mais variáveis de inventário (como por exemplo o volume da árvore, a densidade dos ramos e das folhas, etc). Paralelamente, será importante aumentar a quantidade de dados, incluindo de montados localizados em diferentes regiões, condições edafoclimáticas e regimes de gestão, permitindo melhorar a capacidade de generalização dos modelos de *Machine Learning*.

A integração de séries temporais de imagens de satélite, dados LiDAR e informação climática poderá contribuir para o desenvolvimento de sistemas operacionais de monitorização fitossanitária, modelos com capacidade de distinguir diferentes tipos de doenças, permitindo ainda uma deteção mais precoce de situações de vulnerabilidade e promovendo uma utilização mais eficiente dos recursos florestais.

Em suma, esta investigação demonstrou que a combinação de dados de deteção remota, tecnologias SIG e algoritmos de *Machine Learning* permite avaliar de forma eficaz o estado fitossanitário dos sobreiros, reduzindo a necessidade de campanhas de campo extensivas. Embora a qualidade da cortiça não tenha sido estimada diretamente, os resultados evidenciam que o estado fitossanitário constitui um indicador relevante do potencial qualitativo da cortiça, reforçando o interesse desta metodologia como ferramenta de apoio à gestão florestal e abrindo caminho para o desenvolvimento de modelos futuros direcionados para a previsão direta da qualidade da cortiça.



REFERÊNCIAS BIBLIOGRÁFICAS

- Abdollahnejad, A., & Panagiotidis, D. (2020). Tree Species Classification and Health Status Assessment for a Mixed Broadleaf-Conifer Forest with UAS Multispectral Imaging. *Remote Sensing*, 12(22), 3722. <https://doi.org/10.3390/rs12223722>
- Aroson, J., S. Pereira, João, & G. Pausas, J. (2004). *Cork Oak Woodlands On the Edge*.
- Aubard, V., Paulo, J. A., & Silva, J. M. N. (2019). Long-Term Monitoring of Cork and Holm Oak Stands Productivity in Portugal with Landsat Imagery. *Remote Sensing*, 11(5), 525. <https://doi.org/10.3390/rs11050525>
- Bambagioni, E., Anzilotti, S., Borghi, C., Chirici, G., Salbitano, F., Marchetti, M., & Francini, S. (2025). Satellite Remote Sensing for Monitoring Cork Oak Woodlands—A Comprehensive Literature Review. *Diversity*, 17(6), 420. <https://doi.org/10.3390/d17060420>
- Boletim Estatístico – Mercados da Cortiça. (2023). APCOR.
- Cabral, M. T., Ferreira, M. C., Moreira, T., Carvalho, E. C., & Diniz, A. C. (1992). Diagnóstico das causas da anormal mortalidade dos sobreiros a sul do Tejo. *Scientia Gerundensis*, 18, 205–214.
- Carvacho, J. M. C. (s.d.). *Prompt2Map: Map Generation from Natural Language Queries Using Large Language Models (LLM)*.
- Catalão, J., Navarro, A., & Calvão, J. (2022). Mapping Cork Oak Mortality Using Multitemporal High-Resolution Satellite Imagery. *Remote Sensing*, 14(12), 2750. <https://doi.org/10.3390/rs14122750>
- Chrysafis, I., Korakis, G., Kyriazopoulos, A. P., & Mallinis, G. (2020b). Retrieval of Leaf Area Index Using Sentinel-2 Imagery in a Mixed Mediterranean Forest Area. *ISPRS International Journal of Geo-Information*, 9(11), 622. <https://doi.org/10.3390/ijgi9110622>
- Coelho, C., et al. (2019). Turismo e montado: contributos socioeconómicos.
- Costa, A. (2004). The effect of cork-stripping damage on diameter growth of *Quercus suber* L. *Forestry*, 77(1), 1–8. <https://doi.org/10.1093/forestry/77.1.1>
- Costa, A. C. (2024). ArcGIS Pro Tutorial: *Multiscale Geographically Weighted Regression (MGWR) with the Georgia dataset*. NOVA Information Management School.
- Costa, A., Graça, J., Barbosa, I., & Spiecker, H. (2022). *Effect of climate on cork-ring width and density of Quercus suber L. in Southern Portugal*. *Trees*, 36(6), 1711–1720. <https://doi.org/10.1007/s00468-022-02321-0>



- Costa, A., Pereira, H., & Madeira, M. (2010). Analysis of spatial patterns of oak decline in cork oak woodlands in Mediterranean conditions. *Annals of Forest Science*, 67(2), 204–204. <https://doi.org/10.1051/forest/2009097>
- Costa, D., Tavares, R. M., Baptista, P., & Lino-Neto, T. (2020). Cork Oak Endophytic Fungi as Potential Biocontrol Agents against *Biscogniauxia mediterranea* and *Diplodia corticola*. *Journal of Fungi*, 6(4), 287. <https://doi.org/10.3390/jof6040287>
- Costa, A., & Pereira, H. (s.d.). Montados e sobreirais: uma espécie, duas perspetivas (pp. 17–36).
- COS – Carta de Uso e Ocupação do Solo. (2023). Direção-Geral do Território.
- De Sampaio E Paiva Camilo-Alves, C., Da Clara, M. I. E., & De Almeida Ribeiro, N. M. C. (2013). Decline of Mediterranean oak trees and its association with *Phytophthora cinnamomi*: A review. *European Journal of Forest Research*, 132(3), 411–432. <https://doi.org/10.1007/s10342-013-0688-z>
- Delwart, S. (2015). *ESA Standard Document*. (1).
- Díez-Méndez, A., Díez, J. J., & Poveda, J. (2026). Biocontrol Agents for Disease Management in Mediterranean Agroforestry Species Within the Genus *Quercus*: Holm, Cork, Lusitanian and Pyrenean Oaks. *Agriculture*, 16(4), 409. <https://doi.org/10.3390/agriculture16040409>
- Direção-Geral do Território. (2024). LiDAR – *Modelos Digitais do Relevo* [Conjunto de dados]. Centro de Dados de Portugal.
- Duarte, A., Borralho, N., & Caetano, M. (2021). A Machine Learning Approach to Detect Dead Trees Caused by Longhorned Borer in Eucalyptus Stands Using UAV Imagery. *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, 5818–5821. <https://doi.org/10.1109/IGARSS47720.2021.9554947>
- Esri. (2024). *Region Group* (Spatial Analyst). ArcGIS Pro 3.4 Tool Reference. <https://pro.arcgis.com/en/pro-app/3.4/tool-reference/spatial-analyst/region-group.htm>
- ESA – European Space Agency. (s.d.). Sentinel-2. <https://dataspace.copernicus.eu>
- Estrada, J. S., Fuentes, A., Reszka, P., & Auat Cheein, F. (2023). Machine learning assisted remote forestry health assessment: A comprehensive state of the art review. *Frontiers in Plant Science*, 14, 1139232. <https://doi.org/10.3389/fpls.2023.1139232>
- Fabijańczyk, P., & Zawadzki, J. (2022). Spatial correlations of NDVI and MSAVI2 indices of green and forested areas of urban agglomeration, case study Warsaw, Poland. *Remote Sensing*



- Farajelahi, B., Eya, F. F., & Arefi, H. (2023). FOREST MODELING AND INVENTORY ESTIMATION USING LIDAR DATA. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, X-4/W1-2022, 159–164. <https://doi.org/10.5194/isprs-annals-X-4-W1-2022-159-2023>
- Fassnacht, F. E., White, J. C., Wulder, M. A., & Næsset, E. (2024). Remote sensing in forestry: Current challenges, considerations and directions. *Forestry: An International Journal of Forest Research*, 97(1), 11–37. <https://doi.org/10.1093/forestry/cpad024>
- Fürstenau, B., Quero, C., Riba, J. M., Rosell, G., & Guerrero, A. (2015). Field trapping of the flathead oak borer *Coroebus undatus* (Coleoptera: Buprestidae) with different traps and volatile lures. *Insect Science*, 22(1), 139–149. <https://doi.org/10.1111/1744-7917.12138>
- Georgieva, M., Georgiev, G. T., Ivanov, V., & Hristova, M. (2024). First record of *Biscogniauxia mediterranea* on *Quercus rubra* in Bulgaria. *Historia naturalis bulgarica*, 46(10), 265–271. <https://doi.org/10.48027/hnb.46.102>
- Grabska, E., Hawryło, P., & Socha, J. (2020). Continuous Detection of Small-Scale Changes in Scots Pine Dominated Stands Using Dense Sentinel-2 Time Series. *Remote Sensing*, 12(8), 1298. <https://doi.org/10.3390/rs12081298>
- ICNF – Instituto da Conservação da Natureza e das Florestas. (2023). Plano de Observação da Sanidade Florestal (POSF). <https://www.icnf.pt>
- INE – Instituto Nacional de Estatística. (2023). *Contas Económicas da Silvicultura – 2021*.
- INE – Instituto Nacional de Estatística. (2025). *Estatísticas Agrícolas 2024*.
- INE – Instituto Nacional de Estatística. (2025). *Estatísticas do Turismo 2024*.
- International Union for Conservation of Nature. (2021). *The IUCN Red List of Threatened Species*. <https://www.iucnredlist.org>
- Isbaex, C., Sousa, A. M. O., & Gonçalves, A. C. (2021). *Forest Species Mapping using Sentinel 2A images for the Central Alentejo Region (Portugal)*.
- Jay, S. (s.d.). *Book Captain: Marie Weiss (INRAE)*.



- Jiang, Z., Huete, A., Didan, K., & Miura, T. (2008). Development of a two-band enhanced vegetation index without a blue band. *Remote Sensing of Environment*, 112(10), 3833–3845. <https://doi.org/10.1016/j.rse.2008.06.006>
- Jung, T., Blaschke, H., & Neumann, P. (1996). Isolation, identification and pathogenicity of *Phytophthora* species from declining oak stands. *European Journal of Forest Pathology*, 26(5), 253–272. <https://doi.org/10.1111/j.1439-0329.1996.tb00846.x>
- Kalathas, I., Papoutsidakis, M., & Drosos, C. (2021). Business Intelligence and Machine Learning Methods for Predictive Maintenance in Greek railways. *Open Journal of Applied Sciences*, 11(01), 20–35. <https://doi.org/10.4236/ojapps.2021.111A003>
- Kazanskiy, N., Khabibullin, R., Nikonorov, A., & Khonina, S. (2025). A Comprehensive Review of Remote Sensing and Artificial Intelligence Integration: Advances, Applications, and Challenges. *Sensors*, 25(19), 5965. <https://doi.org/10.3390/s25195965>
- Kim, H. N., Jin, H. Y., Kwak, M. J., Khaine, I., You, H. N., Lee, T. Y., Ahn, T. H., & Woo, S. Y. (2017). Why does *Quercus suber* species decline in Mediterranean areas? *Journal of Asia-Pacific Biodiversity*, 10(3), 337–341. <https://doi.org/10.1016/j.japb.2017.05.004>
- KNIME. (2021). *Steps in Data Preparation: An In-Depth Guide to Data Prep*.
- Lahssini, S., Lahlaoui, H., Alaoui, H. M., Hlal, E. A., Bagaram, M., & Ponette, Q. (2015). Predicting Cork Oak Suitability in Maâmora Forest Using Random Forest Algorithm. *Journal of Geographic Information System*, 07(02), 202–210. <https://doi.org/10.4236/jgis.2015.72017>
- Laino, D., Cabo, C., Prendes, C., Janvier, R., Ordóñez, C., Nikonovas, T., Doerr, S., & Santin, C. (2024). 3DFin: A software for automated 3D forest inventories from terrestrial point clouds. *Forestry: An International Journal of Forest Research*, 97(4), 479–496. <https://doi.org/10.1093/forestry/cpae020>
- Laurentino, J. R. S. (2019). *Condição fitossanitária do montado de sobreiro com recurso a imagens multiespectrais de alta resolução*.
- Lopes-Fernandes, M., Martínez-Fernández, E., Alves, R., Boa-Nova, D., Branquinho, C., Bugalho, M. N., Campos-Mardones, F., Coca-Pérez, A., Frazão-Moreira, A., Marques, M., Moreno-Ortiz, J., Paulo, O., Príncipe, A., Quintero, V., Sendim, A., Sobral, H., & Escalera-Reyes, J. (2024). Cork oak woodlands and decline: A social-ecological review and future transdisciplinary approaches. *Agroforestry Systems*, 98(6), 1927–1944. <https://doi.org/10.1007/s10457-024-00999-4>



- Marcos-Romero, J. C., Poveda, J., & Diez, J. J. (2025). Relation of the soil microbiota of cork oak groves and surrounding grasslands to tree decline. *Applied Soil Ecology*, 212, 106165. <https://doi.org/10.1016/j.apsoil.2025.106165>
- Mata, F., & Dos-Santos, M. (2024). Modeling Cork Yield, Thickness, Price, and Gross Income in the Portuguese Cork Oak Montado. *Forests*, 15(12), 2163. <https://doi.org/10.3390/f15122163>
- Muñoz-Adalia, E. J., Ahmed, J., & Colinas, C. (2022). Microclimatic conditions drive summer flight phenology of *Platypus cylindrus* in managed cork oak stands. *Journal of Applied Entomology*, 146(8), 964–974. <https://doi.org/10.1111/jen.13025>
- Nasiri, V., Darvishsefat, A. A., Arefi, H., Griess, V. C., Sadeghi, S. M. M., & Borz, S. A. (2022). Modeling Forest Canopy Cover: A Synergistic Use of Sentinel-2, Aerial Photogrammetry Data, and Machine Learning. *Remote Sensing*, 14(6), 1453. <https://doi.org/10.3390/rs14061453>
- Navarro, A., Catalao, J., & Calvao, J. (2019a). Assessing the Use of Sentinel-2 Time Series Data for Monitoring Cork Oak Decline in Portugal. *Remote Sensing*, 11(21), 2515. <https://doi.org/10.3390/rs11212515>
- O'Brien, R. M. (2007). A Caution Regarding Rules of Thumb for Variance Inflation Factors. *Quality & Quantity*, 41(5), 673–690. <https://doi.org/10.1007/s11135-006-9018-6>
- Van Hoogstraten, B. (2019, July 26). *Coraebus undatus* [Fotografia]. Observation.org. <https://observation.org/species/556643/>
- Ordenes, F. V., & Silipo, R. (2021). Machine Learning for Marketing on the KNIME Hub: The Development of a Live Repository for Marketing Applications. *Journal of Business Research*.
- Prendes, C., Cabo, C., Ordoñez, C., Majada, J., & Canga, E. (2021). An algorithm for the automatic parametrization of wood volume equations from Terrestrial Laser Scanning point clouds: Application in *Pinus pinaster*. *GIScience & Remote Sensing*, 58(7), 1130–1150. <https://doi.org/10.1080/15481603.2021.1972712>
- Sebastiani, A., Bertozzi, M., Vannini, A., Morales-Rodriguez, C., Calfapietra, C., & Vaglio Laurin, G. (2024). Monitoring ink disease epidemics in chestnut and cork oak forests in central Italy with remote sensing. *Remote Sensing Applications: Society and Environment*, 36, 101329. <https://doi.org/10.1016/j.rsase.2024.101329>



- Silva, J., Araújo, S. D. S., Sales, H., Pontes, R., & Nunes, J. (2023). *Quercus suber* L. Genetic Resources: Variability and Strategies for Its Conservation. *Forests*, 14(9), 1925. <https://doi.org/10.3390/f14091925>
- Silva, J. S. (2007). *Os Montados—Muito para além das árvores* [Livro]. Público, Comunicação Social, SA.
- Sousa, E. M. R., Santos, M. N. S., Varela, M. C., & Henriques, J. (2007). Perda de vigor dos montados de sobreiro e azinho: Análise da situação e perspectivas.
- Tomé, M., Barreiro, S., & Soares, P. (2019). Inventário florestal de sobreiro na Companhia das Lezírias: Protocolo de campo (Relatório Técnico n.º 01/2019). ForChange.
- Tripathi, N., Bhardwaj, N., Kumar, S., & Jain, S. K. (2023). A machine learning-based KNIME workflow to predict VEGFR -2 inhibitors. *Chemical Biology & Drug Design*, 102(1), 38–50. <https://doi.org/10.1111/cbdd.14250>
- UNAC – União da Floresta Mediterrânica. (2021). Guia de comercialização da cortiça.
- Vincini, M., Frazzi, E., & D'Alessio, P. (2008). A broad-band leaf chlorophyll vegetation index at the canopy scale. *Precision Agriculture*, 9(5), 303–319. <https://doi.org/10.1007/s11119-008-9075-z>
- Xi, Z., Hopkinson, C., & Chasmer, L. (2024). Supervised terrestrial to airborne laser scanner model calibration for 3D individual-tree attribute mapping using deep neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 209, 324–343. <https://doi.org/10.1016/j.isprsjprs.2024.02.010>
- Yu, H., Fotheringham, A. S., Li, Z., Oshan, T., Kang, W., & Wolf, L. J. (2020). *Inference in multiscale geographically weighted regression*. *Geographical Analysis*, 52(1), 87–106. <https://doi.org/10.1111/gean.12189>
- Zhang, W., Qi, J., Wan, P., Wang, H., Xie, D., Wang, X., & Yan, G. (2016). An Easy-to-Use Airborne LiDAR Data Filtering Method Based on Cloth Simulation. *Remote Sensing*, 8(6), 501. <https://doi.org/10.3390/rs8060501>
- Zhen, Z., Quackenbush, L., & Zhang, L. (2016). Trends in Automatic Individual Tree Crown Detection and Delineation—Evolution of LiDAR Data. *Remote Sensing*, 8(4), 333. <https://doi.org/10.3390/rs8040333>

Data with Purpose.

