

UNIVERSIDADE DO ALGARVE

FACULDADE DE CIÊNCIAS E TECNOLOGIA

INSTITUTO SUPERIOR DE ENGENHEIRIA

IDENTIFICAÇÃO DA OCUPAÇÃO E USO DO SOLO COM BASE EM IMAGENS PROVENIENTES DE DETEÇÃO REMOTA E EM ALGORITMOS DE *MACHINE* *LEARNING*:

a Reserva da Faia Brava como caso de estudo

Paula Maria de Fraga Borges Pacheco

**Dissertação efetuada sob a orientação de:
Joaquim Luís – Professor Auxiliar
Nuno de Santos Loureiro – Professor Auxiliar**

2025

UNIVERSIDADE DO ALGARVE

FACULDADE DE CIÊNCIAS E TECNOLOGIA

INSTITUTO SUPERIOR DE ENGENHEIRIA

IDENTIFICAÇÃO DA OCUPAÇÃO E USO DO SOLO COM BASE EM IMAGENS PROVENIENTES DE DETEÇÃO REMOTA E EM ALGORITMOS DE *MACHINE* *LEARNING*:

a Reserva da Faia Brava como caso de estudo

Paula Maria de Fraga Borges Pacheco

**Dissertação efetuada sob a orientação de:
Joaquim Luís – Professor Auxiliar
Nuno de Santos Loureiro – Professor Auxiliar**

2025

IDENTIFICAÇÃO DA OCUPAÇÃO E USO DO SOLO COM BASE EM IMAGENS PROVENIENTES DE DETEÇÃO REMOTA E EM ALGORITMOS DE *MACHINE* *LEARNING*:

a Reserva da Faia Brava como caso de estudo

Declaração de Autoria do Trabalho

Declaro ser a autora deste trabalho, que é original e inédito. Autores e trabalhos consultados estão devidamente citados no texto e constam da listagem de referências incluída.

Paula Maria de Fraga Borges Pacheco

© Copyright

A Universidade do Algarve tem o direito, perpétuo e sem limites geográficos, de arquivar e publicitar este trabalho através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, de o divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objetivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.



DEDICATÓRIA:

Dedico este trabalho àqueles que acreditam na preservação do Planeta como um compromisso inadiável.

“A Terra não pertence ao homem; o homem pertence à Terra.” – Chefe Seattle

Que esta dissertação contribua, ainda que modestamente, para o entendimento e proteção das paisagens naturais que nos sustentam.



RESUMO

Esta dissertação procura identificar a ocupação e uso do solo com base em imagens de deteção remota e algoritmos de *Machine Learning* (ML), utilizando como estudo de caso a Reserva da Faia Brava, situada no vale do Côa, distrito da Guarda, Portugal. O estudo avalia a exequibilidade de ferramentas de código aberto, como o QGIS e o *plugin Orfeo Toolbox*, para implementar fluxos de trabalho de classificação supervisionada. Foram processadas imagens de alta resolução obtidas por UAV (2,7 cm/pixel) e pelo satélite Pléiades-Neo (30 cm/pixel), integrando índices de vegetação (GLI, NDVI, SAVI e MSAVI) e métricas texturais de *Haralick*. O treino dos modelos foi realizado numa quadrícula com 500 metros de lado, selecionada pela sua diversidade ecológica, e posteriormente testada em outras áreas da reserva para avaliar a capacidade de generalização. Dois algoritmos de ML, *Random Forest* (RF) e *Support Vector Machine* (SVM), foram testados, com desempenhos avaliados através de matrizes de confusão, *F1-scores* e coeficientes *Kappa*. Os resultados evidenciaram a superioridade dos ortofotomosaicos UAV face às imagens de satélite, especialmente quando combinados com análise textural, embora tenham sido identificadas limitações relacionadas com variações sazonais da vegetação e a interoperabilidade entre sensores. O algoritmo RF mostrou maior consistência enquanto o SVM revelou sensibilidade à complexidade espectral. O estudo destaca a aplicabilidade prática destes métodos para a monitorização ambiental, sublinhando a importância das soluções *open source* para a democratização das tecnologias de deteção remota. Como produto final foi produzido, com base nos modelos de ML, uma carta temática para a área total da Reserva da Faia Brava, abrangendo quatro classes: árvores e arbustos, vegetação herbácea, afloramentos rochosos e outras ocupações e usos do solo.

Palavras-Chave: deteção remota, *Machine Learning*, QGIS, *Orfeo Toolbox*, ocupação e uso do solo, Faia Brava.



ABSTRACT

This dissertation uses the Faia Brava Reserve, located in the Côa Valley in the Guarda district of Portugal, as a case study to identify land use land cover based on remote sensing images and Machine Learning (ML) algorithms. The study evaluates the feasibility of using open-source tools such as QGIS and the Orfeo Toolbox plugin to implement supervised classification workflows. High-resolution images obtained from a UAV (2.7 cm/pixel) and a Pléiades-Neo satellite (30 cm/pixel) were processed and integrated with vegetation indices (GLI, NDVI, SAVI and MSAVI) and Haralick textural metrics. The models were trained on a 500-metre square area selected for its ecological diversity and then tested in other areas of the reserve to evaluate their generalisation capacity. Two ML algorithms, Random Forest (RF) and Support Vector Machine (SVM), were tested and their performance was assessed using confusion matrices, F1-scores and Kappa coefficients. The results demonstrated the superiority of UAV orthophotomosaics over satellite images, particularly when combined with textural analysis. However, limitations relating to seasonal variations in vegetation and interoperability between sensors were also identified. The RF algorithm demonstrated greater consistency, whereas the SVM was sensitive to spectral complexity. The study highlights the practical applicability of these methods for environmental monitoring and underlines the importance of open-source solutions in democratising remote sensing technologies. The final product, based on ML models, was a thematic map of the entire Faia Brava Reserve area, divided into four categories: trees and shrubs, herbaceous vegetation, rocky outcrops, and other land uses land covers.

Keywords: remote sensing, Machine Learning, QGIS, Orfeo Toolbox, land use land cover, Faia Brava.



ÍNDICE

1. INTRODUÇÃO.....	1
1.1. APRESENTAÇÃO E RELEVÂNCIA DO TEMA	1
1.2. CARACTERIZAÇÃO E JUSTIFICAÇÃO DA ESCOLHA DA ÁREA DE ESTUDO.....	2
1.3. PROBLEMA DE INVESTIGAÇÃO	5
1.4. OBJETIVOS DO ESTUDO.....	5
1.5. ESTRUTURA DA DISSERTAÇÃO.....	6
2. ENQUADRAMENTO TEÓRICO	8
2.1. DETEÇÃO REMOTA NO CONTEXTO DA MONITORIZAÇÃO AMBIENTAL.....	9
2.2. ÍNDICES DE VEGETAÇÃO	10
2.2.1. GLI (<i>Green Leaf Index</i>).....	11
2.2.2. NDVI (<i>Normalized Difference Vegetation Index</i>)	11
2.2.3. SAVI (<i>Soil-Adjusted Vegetation Index</i>).....	12
2.2.4. MSAVI (<i>Modified Soil-Adjusted Vegetation Index</i>)	13
2.3. TEXTURA.....	13
2.4. FOSS (<i>FREE AND OPEN SOURCE SOFTWARE</i>).....	14
2.4.1. QGIS	14
2.4.2. <i>Orfeo Toolbox</i>	14
2.5. FUNDAMENTOS DE <i>MACHINE LEARNING</i> APLICADOS À DETEÇÃO REMOTA	15
2.5.1. Aprendizagem Supervisionada.....	15
2.5.2. Aprendizagem Não Supervisionada	16
2.5.3. Aprendizagem Por Reforço	16
2.6. ALGORITMOS DE <i>MACHINE LEARNING</i>	16
2.6.1. <i>Random Forest</i>	16
2.6.2. <i>Support Vector Machine</i>	17
2.7. MÉTRICAS DE VALIDAÇÃO	17
2.7.1. Matriz de Confusão.....	18
2.7.2. <i>F1-score</i>	18
2.7.3. Coeficiente <i>Kappa</i>	18
2.8. SÍNTESE	19
3. MATERIAIS, MÉTODOS E FLUXO DE TRABALHO	20
3.1. MATERIAIS	20
3.1.1. Fontes de dados.....	20
3.1.1.1. Imagens de Satélite	20
3.1.1.2. Imagens UAV.....	21
3.1.1.3. Dados Auxiliares	22
3.1.2. <i>Softwares</i> e Ferramentas	23
3.2. MÉTODOS.....	24
3.2.1. Pré-Processamento	24
3.2.1.1. Correção Geométrica e Alinhamento Espacial	24
3.2.1.2. Reamostragem das Imagens UAV	27
3.2.1.3. Remoção da Banda Alfa.....	27

3.2.1.4.	Normalização Radiométrica.....	27
3.2.1.5.	Valorização dos Dados UAV	28
3.2.2.	Derivação de Variáveis	29
3.2.2.1.	Índices de Vegetação	29
3.2.2.2.	Texturas de <i>Haralick</i>	30
3.2.3.	Concatenação.....	33
3.2.4.	Amostragem e Definição das Classes.....	34
3.2.4.1.	Amostragem	34
3.2.4.2.	Configuração e Seleção de Amostras.....	35
3.2.4.3.	Seleção Espacial de Pixels	36
3.2.4.4.	Integração no Fluxo de Classificação	36
3.2.5.	Modelos de Classificação	37
3.2.5.1.	<i>Random Forest</i>	37
3.2.5.2.	<i>Support Vector Machine</i>	38
3.2.6.	Validação dos Modelos	39
3.3.	FLUXO DE TRABALHO.....	40
3.3.1.	Preparação inicial dos dados.....	40
3.3.2.	Integração e Processamento de Imagens	40
3.3.3.	Geração de Variáveis Derivadas	41
3.3.4.	Treino dos Modelos de Classificação	41
3.3.5.	Aplicação e Classificação Final	42
4.	RESULTADOS E DISCUSSÃO.....	43
4.1.	AVALIAÇÃO QUANTITATIVA	43
4.1.1.	OrtoFotoUAV	44
4.1.1.1.	RF.....	44
4.1.1.2.	SVM	45
4.1.2.	OrtoSat2023 - Cor Verdadeira.....	46
4.1.2.1.	RF.....	46
4.1.2.2.	SVM	47
4.1.3.	OrtoSat2023 Falsa Cor	48
4.1.3.1.	NDVI.....	48
4.1.3.1.1.	RF.....	48
4.1.3.1.2.	SVM	49
4.1.3.2.	SAVI.....	50
4.1.3.2.1.	RF.....	50
4.1.3.2.2.	SVM	51
4.1.3.3.	MSAVI	52
4.1.3.3.1.	RF.....	53
4.1.3.3.2.	SVM	54
4.1.4.	Análise do <i>F1-score</i> entre Modelos.....	55
4.1.5.	Análise do Coeficiente Kappa entre Modelos	58
4.1.6.	Análise geral de valores de <i>Kappa</i>	59
4.2.	AVALIAÇÃO QUALITATIVA.....	60
4.2.1.	OrtoFotoUAV	61
4.2.2.	OrtoSat2023 (Cor Verdadeira)	62
4.2.3.	OrtoSat2023 (Falsa Cor)	63
4.3.	ANÁLISE CRÍTICA/DISCUSSÃO.....	65



4.4.	APLICABILIDADE PRÁTICA	66
5.	CONCLUSÕES E TRABALHOS FUTUROS.....	68
5.1.	PRINCIPAIS CONCLUSÕES.....	68
5.2.	LIMITAÇÕES DO ESTUDO.....	69
5.3.	RECOMENDAÇÕES E MELHORIAS.....	71
5.4.	PROPOSTAS PARA TRABALHOS FUTUROS.....	72
5.5.	CONTRIBUIÇÕES PARA OS OBJETIVOS DE DESENVOLVIMENTO SUSTENTÁVEL (ODS)	74
BIBLIOGRAFIA	75	
ANEXOS.....	79	



ÍNDICE DE FIGURAS

FIGURA 1.1 - LOCALIZAÇÃO DA ÁREA DE ESTUDO	2
FIGURA 1.2 - PAISAGENS DA RESERVA DA FAIA BRAVA (CRÉDITOS: JOÃO DE CASTRO, 2024).....	3
FIGURA 1.3 - CARTA HIPSOMÉTRICA DA RESERVA DA FAIA BRAVA	3
FIGURA 1.4 - CARTA DE DECLIVES DA RESERVA DA FAIA BRAVA.....	4
FIGURA 3.1 - BLOCOS DAS COMPOSIÇÕES ORTOsat2023.....	20
FIGURA 3.2 - ÁREA DE COBERTURA DO ORTOFOTOUAV	21
FIGURA 3.3 - GRELHAS UTILIZADAS NO ESTUDO	22
FIGURA 3.4 - QUADRICULA 21	23
FIGURA 3.5 - PONTOS DE CONTROLO	25
FIGURA 3.6- ÍNDICES DE VEGETAÇÃO	30
FIGURA 3.7 - TEXTURAS SIMPLES DE <i>HARALICK</i>	32
FIGURA 3.8 - <i>SHAPEFILES</i> DE TREINO E VALIDAÇÃO - Q21	34
FIGURA 4.1 - DESEMPENHO DA CLASSIFICAÇÃO DO ORTOFOTOUAV COM O ALGORITMO RF.....	45
FIGURA 4.2 -DESEMPENHO DA CLASSIFICAÇÃO DO ORTOFOTOUAV COM O ALGORITMO SVM.....	46
FIGURA 4.3 - DESEMPENHO DA CLASSIFICAÇÃO DO ORTOsat2023 COR VERDADEIRA COM O ALGORITMO RF.....	47
FIGURA 4.4 - DESEMPENHO DA CLASSIFICAÇÃO DO ORTOsat2023 COR VERDADEIRA COM O ALGORITMO SVM.....	48
FIGURA 4.5 - DESEMPENHO DA CLASSIFICAÇÃO DO ORTOsat2023 FALSA COR COM O ALGORITMO RF UTILIZANDO O NDVI.....	49
FIGURA 4.6 - DESEMPENHO DA CLASSIFICAÇÃO DO ORTOsat2023 FALSA COR COM O ALGORITMO SVM UTILIZANDO O NDVI.	50
FIGURA 4.7 - DESEMPENHO DA CLASSIFICAÇÃO DO ORTOsat2023 FALSA COR COM O ALGORITMO RF UTILIZANDO O SAVI.....	51
FIGURA 4.8 - DESEMPENHO DA CLASSIFICAÇÃO DO ORTOsat2023 FALSA COR COM O ALGORITMO SVM UTILIZANDO O SAVI.	52
FIGURA 4.9 - DESEMPENHO DA CLASSIFICAÇÃO DO ORTOsat2023 FALSA COR COM O ALGORITMO RF UTILIZANDO O MSAVI.	53
FIGURA 4.10 - DESEMPENHO DA CLASSIFICAÇÃO DO ORTOsat2023 FALSA COR COM O ALGORITMO SVM UTILIZANDO O MSAVI.....	54
FIGURA 4.11 - F1 SCORE DA CLASSE 1 (ÁRVORES E ARBUSTOS) COM RF (ESQUERDA) E SVM (DIREITA)	55
FIGURA 4.12 - F1 SCORE DA CLASSE 2 (VEGETAÇÃO HERBÁCEA) COM RF (ESQUERDA) E SVM (DIREITA).	56
FIGURA 4.13 - F1 SCORE DA CLASSE 3 (AFLORAMENTOS ROCHOSOS) COM RF (ESQUERDA) E SVM (DIREITA).	57
FIGURA 4.14 - F1 SCORE DA CLASSE 4 (OUTRAS OCUPAÇÕES E USOS DO SOLO) COM RF (ESQUERDA) E SVM (DIREITA).....	58
FIGURA 4.15 – COEFICIENTE <i>KAPPA</i> COM RF (ESQUERDA) E SVM (DIREITA).....	59
FIGURA 4.16 -RESULTADOS DA CLASSIFICAÇÃO DA Q41 COM O RF (A), SVM (B) E O ORTOFOTOUAV (C).	61
FIGURA 4.17- RESULTADOS DA CLASSIFICAÇÃO NA Q41 COM O RF (A), SVM (B) E ORTOsat2023 COR VERDADEIRA (C).....	62
FIGURA 4.18- RESULTADOS DA CLASSIFICAÇÃO NA Q41 COM O RF (A), SVM (B) E ORTOsat2023 FALSA COR (C).	63
FIGURA 4.19 - EXEMPLOS DE ZONAS CLASSIFICADAS NA R-FB COM O MODELO FINAL (RF + RGB + GLI + ENTROPIA)	64



ÍNDICE DE TABELAS

TABELA 3.1 - TABELA DE ATRIBUTOS DA <i>SHAPEFILE</i> DOS PONTOS DE CONTROLO	26
TABELA 3.2- VALORES UTILIZADOS PARA TREINO E VALIDAÇÃO	35

ÍNDICE DE ANEXOS

ANEXO 1- DESEMPENHO DAS COMPOSIÇÕES TESTADAS COM O ALGORITMO RANDOM FOREST (RF) APLICADO AO ORTOFOTOUAV.	
APRESENTAM-SE OS VALORES DE PRECISÃO, REVOCAÇÃO, <i>F1-SCORE</i> POR CLASSE E O COEFICIENTE <i>KAPPA</i> TOTAL. A COMPOSIÇÃO RGB+GLI+B2 OBTVEU O MELHOR DESEMPENHO GLOBAL (ASSINALADO A VERDE), ENQUANTO RGB+GLI+HS APRESENTOU O PIOR RESULTADO (ASSINALADO A VERMELHO).....	80
ANEXO 2 - DESEMPENHO DAS COMPOSIÇÕES TESTADAS COM O ALGORITMO SUPPORT VECTOR MACHINE (SVM) APLICADO AO ORTOFOTOUAV.	
APRESENTAM-SE OS VALORES DE PRECISÃO, REVOCAÇÃO, <i>F1-SCORE</i> POR CLASSE E O COEFICIENTE <i>KAPPA</i> TOTAL. A COMPOSIÇÃO RGB+GLI+B2 OBTVEU O MELHOR DESEMPENHO GLOBAL (ASSINALADO A VERDE), ENQUANTO RGB+GLI+HS APRESENTOU O PIOR RESULTADO (ASSINALADO A VERMELHO).....	80
ANEXO 3 - DESEMPENHO DAS COMPOSIÇÕES TESTADAS COM O ALGORITMO RF APLICADO ÀS IMAGENS ORTOsat2023 (COR VERDADEIRA).	
APRESENTAM-SE OS VALORES DE PRECISÃO, REVOCAÇÃO, <i>F1-SCORE</i> POR CLASSE E O COEFICIENTE <i>KAPPA</i> TOTAL. A COMPOSIÇÃO RGB OBTVEU O MELHOR DESEMPENHO GLOBAL (ASSINALADO A VERDE), ENQUANTO RGB+GLI+HS APRESENTOU O PIOR RESULTADO (ASSINALADO A VERMELHO).....	81
ANEXO 4 - DESEMPENHO DAS COMPOSIÇÕES TESTADAS COM O ALGORITMO SVM APLICADO ÀS IMAGENS ORTOsat2023 (COR VERDADEIRA).	
APRESENTAM-SE OS VALORES DE PRECISÃO, REVOCAÇÃO, <i>F1-SCORE</i> POR CLASSE E O COEFICIENTE <i>KAPPA</i> TOTAL. A COMPOSIÇÃO RGB+GLI+HS OBTVEU O MELHOR DESEMPENHO GLOBAL (ASSINALADO A VERDE), ENQUANTO RGB+GLI+B3 APRESENTOU O PIOR RESULTADO (ASSINALADO A VERMELHO).....	81
ANEXO 5 - DESEMPENHO DAS COMPOSIÇÕES TESTADAS COM O ALGORITMO RF APLICADO ÀS IMAGENS ORTOsat2023 (FALSA COR – NDVI).	
APRESENTAM-SE OS VALORES DE PRECISÃO, REVOCAÇÃO, <i>F1-SCORE</i> POR CLASSE E O COEFICIENTE <i>KAPPA</i> TOTAL. A COMPOSIÇÃO IRG+NDVI+B7 OBTVEU O MELHOR DESEMPENHO GLOBAL (ASSINALADO A VERDE), ENQUANTO IRG+NDVI+B4 APRESENTOU O PIOR RESULTADO (ASSINALADO A VERMELHO).....	82
ANEXO 6 - DESEMPENHO DAS COMPOSIÇÕES TESTADAS COM O ALGORITMO SVM APLICADO ÀS IMAGENS ORTOsat2023 (FALSA COR – NDVI).	
APRESENTAM-SE OS VALORES DE PRECISÃO, REVOCAÇÃO, <i>F1-SCORE</i> POR CLASSE E O COEFICIENTE <i>KAPPA</i> TOTAL. A COMPOSIÇÃO IRG+NDVI+HS OBTVEU O MELHOR DESEMPENHO GLOBAL (ASSINALADO A VERDE), ENQUANTO IRG+NDVI+B3 APRESENTOU O PIOR RESULTADO (ASSINALADO A VERMELHO).....	82
ANEXO 7 - DESEMPENHO DAS COMPOSIÇÕES TESTADAS COM O ALGORITMO RF APLICADO ÀS IMAGENS ORTOsat2023 (FALSA COR – SAVI).	
APRESENTAM-SE OS VALORES DE PRECISÃO, REVOCAÇÃO, <i>F1-SCORE</i> POR CLASSE E O COEFICIENTE <i>KAPPA</i> TOTAL. A COMPOSIÇÃO IRG+SAVI+B7 APRESENTOU O MELHOR DESEMPENHO GLOBAL (ASSINALADO A VERDE), ENQUANTO IRG+SAVI+B4 REGISTOU O PIOR RESULTADO (ASSINALADO A VERMELHO).....	83
ANEXO 8 - DESEMPENHO DAS COMPOSIÇÕES TESTADAS COM O ALGORITMO SVM APLICADO ÀS IMAGENS ORTOsat2023 (FALSA COR – SAVI).	
APRESENTAM-SE OS VALORES DE PRECISÃO, REVOCAÇÃO, <i>F1-SCORE</i> POR CLASSE E O COEFICIENTE <i>KAPPA</i> TOTAL. A COMPOSIÇÃO IRG+SAVI+HS APRESENTOU O MELHOR DESEMPENHO GLOBAL (ASSINALADO A VERDE), ENQUANTO IRG+SAVI+B3 REGISTOU O PIOR RESULTADO (ASSINALADO A VERMELHO).....	83
ANEXO 9 - DESEMPENHO DAS COMPOSIÇÕES TESTADAS COM O ALGORITMO RF APLICADO ÀS IMAGENS ORTOsat2023 (FALSA COR – MSAVI).	
APRESENTAM-SE OS VALORES DE PRECISÃO, REVOCAÇÃO, <i>F1-SCORE</i> POR CLASSE E O COEFICIENTE <i>KAPPA</i> TOTAL. A COMPOSIÇÃO IRG+MSAVI+B7 APRESENTOU O MELHOR DESEMPENHO GLOBAL (ASSINALADO A VERDE), ENQUANTO IRG+MSAVI+B4 REGISTOU O PIOR RESULTADO (ASSINALADO A VERMELHO).....	84
ANEXO 10 - DESEMPENHO DAS COMPOSIÇÕES TESTADAS COM O ALGORITMO SVM APLICADO ÀS IMAGENS ORTOsat2023 (FALSA COR – MSAVI).	
APRESENTAM-SE OS VALORES DE PRECISÃO, REVOCAÇÃO, <i>F1-SCORE</i> POR CLASSE E O COEFICIENTE <i>KAPPA</i> TOTAL. A COMPOSIÇÃO IRG+MSAVI+HS APRESENTOU O MELHOR DESEMPENHO GLOBAL (ASSINALADO A VERDE), ENQUANTO IRG+MSAVI+B3 REGISTOU O PIOR RESULTADO (ASSINALADO A VERMELHO).....	84
ANEXO 11 -MAPA FINAL DE CLASSIFICAÇÃO DA RESERVA DA FAIA BRAVA.	
O MAPA REPRESENTA A DISTRIBUIÇÃO ESPACIAL DAS QUATRO CLASSES DE OCUPAÇÃO DO SOLO: ÁRVORES E ARBUSTOS, VEGETAÇÃO HERBÁCEA, AFLORAMENTOS ROCHOSOS E OUTRAS OCUPAÇÕES E USOS DO SOLO. A CLASSIFICAÇÃO FOI REALIZADA COM O ALGORITMO <i>RANDOM FOREST</i> APLICADO AO ORTOFOTOMOSAICO RGB, COMBINADO COM O ÍNDICE GLI E A MÉTRICA DE TEXTURA ENTROPIA. A GRELHA UTM DE 500 METROS FOI SOBREPOSTA PARA FACILITAR A REFERÊNCIA ESPACIAL E ANÁLISE POR QUADRÍCULA.....	85



ANEXO 12 - DISTRIBUIÇÃO ESPACIAL DA CLASSE "ÁRVORES E ARBUSTOS" NA RESERVA DA FAIA BRAVA. MAPA OBTIDO A PARTIR DO MODELO FINAL DE CLASSIFICAÇÃO, DESTACANDO EXCLUSIVAMENTE A MANCHA DE VEGETAÇÃO ARBÓREA E ARBUSTIVA IDENTIFICADA NO TERRITÓRIO.	86
ANEXO 13 - DISTRIBUIÇÃO ESPACIAL DA CLASSE "VEGETAÇÃO HERBÁCEA" NA RESERVA DA FAIA BRAVA. MAPA RESULTANTE DA CLASSIFICAÇÃO FINAL, EVIDENCIANDO AS ÁREAS DOMINADAS POR VEGETAÇÃO HERBÁCEA NO TERRITÓRIO.	87
ANEXO 14 - DISTRIBUIÇÃO ESPACIAL DA CLASSE "AFLORAMENTOS ROCHOSOS" NA RESERVA DA FAIA BRAVA. REPRESENTAÇÃO ISOLADA DESTA CLASSE, CONFORME IDENTIFICADA NO MODELO FINAL DE CLASSIFICAÇÃO, EVIDENCIANDO A SUA CONCENTRAÇÃO NAS ZONAS DE RELEVO MAIS EXPOSTO OU PEDREGOSO.....	88
ANEXO 15 - DISTRIBUIÇÃO ESPACIAL DA CLASSE "OUTRAS OCUPAÇÕES E USOS DO SOLO" NA ÁREA DE ESTUDO. ESTA CATEGORIA AGRUPA VÁRIOS ELEMENTOS, INCLUINDO CAMINHOS, SOLO NU, CONSTRUÇÕES, CORPOS DE ÁGUA E ZONAS DE SOMBRA. A SUA REPRESENTAÇÃO PERMITE OBSERVAR A LOCALIZAÇÃO E EXTENSÃO DESTES ELEMENTOS NO TERRITÓRIO.....	89

ÍNDICE DE ABREVIATURAS

R-FB – Reserva da Faia Brava

CNES – *Centre National d'Études Spatiales*

FOSS – *Free and Open Source Software*

GLCM – *Gray Level Co-occurrence Matrix*

GLI – *Green Leaf Index*

GNSS - *Global Navigation Satellite System*

ICNF – Instituto de Conservação da Natureza e das Florestas

ML – *Machine Learning*

MSAVI – *Modified Soil-Adjusted Vegetation Index*

NDVI – *Normalized Difference Vegetation Index*

NIR – *Near Infrared*

ODS - Objetivos de Desenvolvimento Sustentável

ONGA – Organização Não-Governamental de Ambiente

OTB – *Orfeo Toolbox*

RF – *Random Forest*

RTK - *Real Time Kinematic*

SAVI – *Soil-Adjusted Vegetation Index*

SMOS – Sistema de Monitorização da Ocupação do Solo

SVM – *Support Vector Machine*

TIN - *Triangulated Irregular Network*

UAV – *Unmanned Aerial Vehicle*

1. INTRODUÇÃO

1.1. Apresentação e Relevância do Tema

A monitorização da ocupação e uso do solo é fundamental para a gestão sustentável dos recursos naturais, sobretudo em áreas com elevado valor ecológico e/ou económico. Atualmente, o rápido desenvolvimento das tecnologias de deteção remota permite o acesso a imagens de alta resolução, facilitando análises detalhadas sobre as alterações no território e a distinção entre as diversas classes de vegetação. Em conjunto com técnicas de *Machine Learning* (ML), tais avanços têm revolucionado a produção de cartografia temática, tornando o processo de classificação mais automatizado e preciso (Maxwell et al., 2018).

Neste contexto, a Reserva da Faia Brava (R-FB), localizada no vale do rio Côa (distrito da Guarda), constitui um laboratório natural para a aplicação de metodologias inovadoras de deteção remota. Gerida por uma Organização Não-Governamental de Ambiente, a Faia Brava - Associação de Conservação da natureza, esta área protegida enfrenta desafios decorrentes da desertificação ambiental e humana, os quais reforçam a necessidade de estratégias de conservação sustentadas por dados rigorosos. A paisagem caracteriza-se por um mosaico de habitats, dominado por matos de giesta-branca (*Cytisus multiflorus*), por vezes arborescentes, com presença de sobreiro (*Quercus suber*) e azinheira (*Quercus rotundifolia*) e de afloramentos rochosos, além de áreas de solo nu (ICNF, 2025), tornando a Faia Brava um exemplo relevante e desafiante para avaliar o sucesso da aplicação de métodos de classificação da ocupação e uso do solo.

Nos últimos anos algoritmos de ML como o *Random Forest* (RF) e o *Support Vector Machine* (SVM) têm demonstrado elevada precisão na discriminação entre diferentes coberturas (Pal, 2005; Belgiu & Drăguț, 2016). A adoção dessas técnicas exige, frequentemente, conhecimentos de programação em linguagens como o *Python*, R ou *Julia*, o que pode constituir uma barreira para utilizadores menos experientes. É nesse ponto que sobressaem soluções de *software* livre (*open source*), como o QGIS, que podem integrar diversos *plugins* específicos para a análise espacial, como é o *Orfeo Toolbox* (OTB). Esses *plugins*, por sua vez, oferecem interfaces

gráficas intuitivas, permitindo executar algoritmos de ML sem a necessidade de linhas de comando.

Desta forma, o presente estudo propõe-se avaliar a exequibilidade desse *plugin* na classificação da ocupação e uso do solo, verificando se ferramentas como o OTB conseguem assegurar todo o processo de modelação sem exigir conhecimentos avançados de codificação e, ainda assim, produzir resultados robustos. Para tal, recorrer-se-á a métricas de validação consolidadas, como a matriz de confusão e o coeficiente *Kappa* (Maxwell et al., 2018), bem como o *F1-score* (Ferreira et al., 2003).

Ao possibilitar a cartografia detalhada e a monitorização contínua da Faia Brava através de ferramentas acessíveis e processos automatizados, este trabalho contribui para a democratização das tecnologias de deteção remota e para a valorização da ciência aplicada a áreas de reconhecido valor ambiental. Em paralelo, reforça a capacidade de técnicos e investigadores na conservação da natureza, consolidando o QGIS enquanto plataforma confiável para a aplicação de algoritmos de ML sem recurso a linguagens de programação mais complexas.

1.2. Caracterização e Justificação da Escolha da Área de Estudo

Segundo o Instituto de Conservação da Natureza e das Florestas (ICNF), a Reserva da Faia Brava (R-FB ou Faia Brava) insere-se no vale do rio Côa, num troço de cerca de 5 km de extensão, com orientação sul-norte (Fig. 1.1).



Figura 1.1 - Localização da área de estudo

Dominam na paisagem as rochas graníticas, por vezes escarpadas (Fig. 1.2). O clima é continental seco, com grandes amplitudes térmicas e fraca pluviosidade (IPMA, 2025a; IPMA, 2025b).



Figura 1.2 - Paisagens da Reserva da Faia Brava (Créditos: João de Castro, 2024)

A área de estudo, revela um relevo marcado por contrastes altimétricos significativos (Fig. 1.3). A topografia da Reserva da Faia Brava é dominada por um vale central, de cotas mais baixas, ladeado por encostas elevadas e muito declivosas. As linhas de água e as curvas de nível, sobrepostas à carta hipsométrica, evidenciam uma paisagem profundamente modelada pela erosão

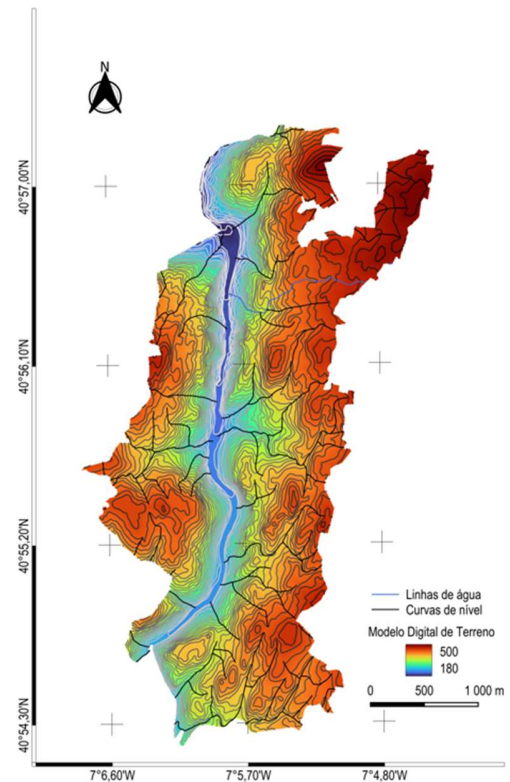
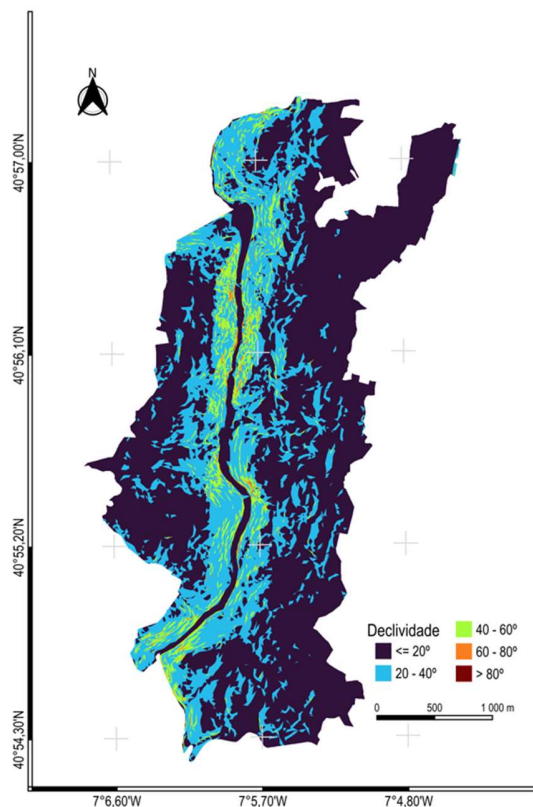


Figura 1.3 - Carta hipsométrica da Reserva da Faia Brava



O contexto físico, da área de estudo, é marcado por um gradiente topográfico significativo, com áreas com declives inferiores a 20°, intercaladas com setores mais acentuados que ultrapassam pontualmente os 80° (Fig. 1.4)

Figura 1.4 - Carta de declives da Reserva da Faia Brava

A R-FB foi selecionada pela sua importância ecológica e pelos desafios ambientais que enfrenta. Classificada como a primeira Área Protegida Privada em Portugal, distingue-se pela conservação da biodiversidade e pela recuperação de habitats naturais.

Como foi já referido, a paisagem da Faia Brava caracteriza-se por alguma heterogeneidade, e essa diversidade de coberturas torna-a particularmente adequada à aplicação de metodologias de deteção remota de elevada resolução, provenientes de veículos aéreos não tripulados (UAV / drone) e de satélites, como o Pléiades-Neo, permitindo uma avaliação pormenorizada dos métodos de classificação da ocupação e do uso do solo.

Neste quadro, a Faia Brava representa um caso de estudo de excelência para aferir do potencial do QGIS, com o *plugin* OTB, para classificação da ocupação e uso do solo, em virtude da sua multiplicidade de coberturas, dos desafios prementes de conservação e da possibilidade de se aplicarem metodologias inovadoras de forma acessível e cientificamente fundamentada.

1.3. Problema de Investigação

A crescente adoção de técnicas de ML na classificação da ocupação e uso do solo tem permitido análises mais rigorosas, eficientes e automatizadas (Pal, 2005; Maxwell et al., 2018). Contudo, muitas destas abordagens implicam a utilização de linguagens de programação, o que pode constituir uma barreira ao acesso a utilizadores sem formação específica na área da informática.

Neste enquadramento surge a necessidade de se encontrarem soluções que tornem exequível a aplicação de algoritmos robustos, como o RF e o SVM, através de interfaces gráficas, dispensando o uso de linhas de comando. O QGIS, *software* livre amplamente utilizado, pode integrar *plugins* como o OTB, o qual possibilita métodos de ML em *workflows* essencialmente visuais (Coetzee et al., 2020).

O objeto de estudo (problema de investigação) consiste em avaliar até que ponto esse *plugin* providencia um fluxo de trabalho completo, rigoroso e reproduzível para a classificação de imagens, sem necessidade de programação e com resultados comparáveis às abordagens tradicionais. Para responder a esta questão recorrer-se-á a imagens de satélite (Pléiades-Neo) e a ortofotomosaicos de UAV (drones), bem como a métricas de validação reconhecidas (Ferreira et al., 2003; Maxwell et al., 2018).

1.4. Objetivos do Estudo

Objetivo Geral:

- Avaliar a exequibilidade do QGIS, e em particular do *plugin* OTB, para a classificação da ocupação e uso do solo, garantindo um fluxo de trabalho acessível, replicável e tecnicamente robusto.

Objetivos Específicos:

- Comparar o desempenho dos algoritmos RF e SVM, quando implementados através do *plugin* OTB, no QGIS.
- Avaliar a exatidão dos resultados de classificação, com recurso a métricas de validação reconhecidas na literatura - matriz de confusão, precisão, revocação, *F1-score* e coeficiente *Kappa*.

- Identificar vantagens e limitações decorrentes do uso exclusivo de interfaces gráficas, sem recorrer a linhas de comando, em fluxos de trabalho de detecção remota.
- Propor melhorias e boas práticas na utilização do QGIS para cartografia e gestão ambiental em áreas de elevado valor ambiental, reforçando a aplicabilidade e a eficiência das metodologias adotadas.

1.5. Estrutura da Dissertação

A dissertação encontra-se organizada em cinco capítulos, articulados de forma lógica e progressiva:

- Capítulo 1 - Introdução: Apresenta o enquadramento geral do tema, destacando a relevância do QGIS enquanto *software* livre, a caracterização e justificação da área de estudo, o problema de investigação, os objetivos do estudo e a estrutura do documento.
- Capítulo 2 – Enquadramento Teórico: Explora os conceitos fundamentais da detecção remota, os índices de vegetação, as texturas de *Haralick* e os algoritmos de *Machine Learning* (RF e SVM). Introduce o QGIS e o *plugin* OTB, bem como as principais métricas de avaliação da classificação supervisionada, apresentando ainda uma revisão do estado da arte.
- Capítulo 3 – Metodologia: Descreve o fluxo de trabalho adotado no QGIS, abrangendo desde a aquisição e pré-processamento das imagens até à implementação dos algoritmos RF e SVM, à configuração do *plugin* OTB e à definição dos critérios de validação.
- Capítulo 4 – Resultados e Discussão: Apresenta os resultados obtidos, incluindo a produção de mapas temáticos, e compara o desempenho dos algoritmos. Discute, também, a fiabilidade e aplicabilidade das ferramentas OTB, considerando a R-FB como caso de estudo.
- Capítulo 5 – Conclusões e Trabalhos Futuros: Sintetiza as principais conclusões do estudo, evidencia as limitações encontradas e propõe melhorias na aplicação do QGIS a projetos de detecção remota. Sugere, ainda, direções para investigação futura, de modo a consolidar e expandir a utilização de técnicas de ML em *software* livre. O capítulo destaca, também, as contribuições do estudo para os Objetivos de Desenvolvimento



Sustentável (ODS), nomeadamente nos domínios da ação climática, conservação da biodiversidade, inovação tecnológica e educação.

A dissertação apoia-se em referências científicas atualizadas, assegurando uma abordagem metodológica rigorosa e em conformidade com as boas práticas na deteção remota e na classificação supervisionada.

2. ENQUADRAMENTO TEÓRICO

A classificação automática da ocupação e uso do solo com base em imagens de satélite tem sido exaustivamente estudada, dada a capacidade das técnicas de ML para lidar com grandes áreas e diferentes coberturas, muitas vezes sujeitas a transformações rápidas (Belgiu & Drăguț, 2016; Maxwell et al., 2018; Pal, 2005). A crescente disponibilidade de imagens de satélite e de UAV, com resoluções espaciais cada vez mais elevadas, tem contribuído para melhorar a precisão das classificações em ecossistemas heterogêneos, como se observa na Faia Brava (Müllerová et al., 2016). Se, por um lado, os satélites asseguram coberturas globais e sistemáticas, por outro, os UAV permitem aquisições de imagens em momentos específicos, ajustados aos ciclos fenológicos da vegetação ou a eventos pontuais, como demonstrado por Müllerová et al. (2016) no caso da deteção da *Robinia pseudoacacia* fora da fase de floração, possibilitando assim uma análise multiescala mais detalhada.

No âmbito da deteção remota, a adoção de ferramentas de código aberto tem ganho destaque devido ao potencial de redução de custos, à flexibilidade e à possibilidade de ajustar soluções às necessidades específicas dos utilizadores (Lourenço et al., 2021). *Softwares* como o QGIS, em conjunto com *plugins* como o OTB, constituem opções robustas para a realização de fluxos de trabalho completos — desde o pré-processamento até à classificação e validação (Coetzee et al., 2020). Ao competirem diretamente com *software* proprietário, incentivam igualmente a democratização da tecnologia e o desenvolvimento colaborativo (Dias et al., 2024).

O OTB, desenvolvido pelo *Centre National d'Études Spatiales* (CNES, França), é uma biblioteca *open source* voltada para o processamento de imagens de deteção remota, destacando-se pela capacidade para lidar com grandes volumes de dados e pela diversidade de ferramentas específicas disponíveis. A sua integração no QGIS permite realizar tarefas como as correções geométricas e radiométricas, o cálculo de índices de vegetação, a análise de texturas e a classificação supervisionada com algoritmos como os RF e SVM, sem necessidade de utilizar linhas de código (Grizonnet et al., 2017; OTB Development Team, 2023). Trata-se, portanto, de uma solução acessível e robusta para projetos de cartografia e monitorização ambiental.

No que diz respeito aos algoritmos de classificação em deteção remota, RF e SVM são duas das opções mais utilizadas (Belgiu & Drăguț, 2016; Maxwell et al., 2018). O RF recorre ao método

de agregação (*bagging*) de múltiplas árvores de decisão, elevando a robustez do modelo e diminuindo a probabilidade de *overfitting* (fenómeno em que o modelo se ajusta demasiado aos dados de treino perdendo a capacidade de generalizar para novas áreas). O *bagging* (do inglês *bootstrap aggregating*) consiste em gerar vários subconjuntos dos dados de treino, com reposição, para treinar diferentes modelos — neste caso, árvores de decisão — cujos resultados são depois combinados (por exemplo, por votação), o que reduz a variância do modelo e melhora a sua estabilidade. Além disso, o RF permite avaliar a importância relativa de cada variável. O SVM, por sua vez, concentra-se na maximização da margem entre as classes, podendo recorrer a *kernels* adequados para lidar com fronteiras não lineares. Um *kernel* é uma função que transforma os dados de entrada para um espaço de maior dimensão, onde as classes se tornam mais facilmente separáveis, o que se revela vantajoso em cenários com elevada complexidade espectral. Quando aliados a abordagens OBIA (*Object-Based Image Analysis*), que fazem a segmentação das imagens em objetos homogéneos, os algoritmos tendem a alcançar níveis de precisão ainda mais elevados (Lourenço et al., 2021; Cini et al., 2024). Estudos como os de Pal (2005), Belgiu & Drăguț (2016) e Phan Thanh Noi & Kappas (2018) demonstram a alta precisão de ambos os algoritmos na classificação de dados multiespectrais, nomeadamente em tarefas de deteção e monitorização da ocupação e uso do solo. Em todas estas abordagens as métricas de validação consolidadas na literatura científica, como a matriz de confusão, o *F1-score* e o coeficiente *Kappa*, são cruciais para aferir da fiabilidade dos resultados (Ferreira et al., 2003; Maxwell et al., 2018).

2.1. Deteção Remota no Contexto da Monitorização Ambiental

A deteção remota consiste na recolha de dados sobre a superfície terrestre através de sensores instalados em satélites, aeronaves ou drones, registando-se a radiação refletida ou emitida pelos elementos do terreno em diversas faixas do espectro eletromagnético (Maxwell et al., 2018). Para fins de monitorização ambiental, as imagens resultantes (em formato *raster*) permitem detetar persistências ou alterações no coberto vegetal, no solo e nos recursos hídricos, graças à resolução espacial e às múltiplas bandas espectrais disponíveis.

Nos últimos anos, a utilização de veículos aéreos não tripulados (UAV) tem vindo a complementar e, em certos casos, substituir parcialmente as imagens de satélite, sobretudo em

escalas locais ou regionais (Anderson & Gaston, 2013). Os drones possibilitam a recolha de dados em momentos específicos, alinhados com as dinâmicas sazonais da vegetação e/ou com a ocorrência de fenómenos pontuais (por exemplo, após incêndios florestais). Permitem também resoluções espaciais muito superiores às obtidas através da maioria dos satélites. Além disso, a fotogrametria de UAV permite também a geração de Modelos Digitais de Superfície (DSM), alargando as perspetivas de análise, mormente em habitats que requerem um exame pormenorizado (de Castro et al., 2018).

No presente trabalho, recorre-se a imagens de satélite Pléiades-Neo, com resolução de 30 cm/pixel, fornecidas pela Direção-Geral do Território, e a imagens de UAV em RGB, com resolução de 2,76 cm/pixel, recolhidas por uma equipa do Mestrado em Sistemas de Informação Geográfica - Universidade do Algarve. Contudo, antes da sua análise aprofundada é imprescindível efetuar correções geométricas e radiométricas, de modo a minimizar distorções e assegurar a qualidade das medições espaciais (Maxwell et al., 2018). Tal rigor no pré-processamento é determinante em contextos ambientais complexos, como é o da R-FB.

2.2. Índices de Vegetação

Os índices de vegetação assumem um papel essencial na análise de dados de deteção remota. Permitem destacar contrastes e padrões que poderiam passar despercebidos numa simples observação de combinações-padrão das bandas espectrais (Pettorelli et al., 2005; Huang et al., 2021). Em termos gerais, cada índice é obtido por combinações específicas de bandas (por exemplo, vermelho e infravermelho próximo), permitindo estimar, de forma indireta, parâmetros como o vigor ou a densidade da vegetação. Assim, tornam-se ferramentas de grande utilidade para a cartografia e monitorização do território e dos ecossistemas, fornecendo indicadores tanto para estudos de biodiversidade como para aplicações agrícolas e florestais. Ao longo das últimas décadas, diversos índices foram propostos para responder a diferentes condições ambientais e requisitos operacionais, o que levou à consolidação de vários deles na prática da deteção remota, com destaque para o GLI, NDVI, SAVI e MSAVI (Xue & Su, 2017). Porém, há que ter consciência de que a simplificação numérica dos índices de vegetação pode omitir alguma complexidade ambiental do território (Butt, 2018). Por esse motivo, é

recomendável complementá-los com outras variáveis, como informações espaciais ou texturais, para que a classificação atinja níveis mais elevados de detalhe e precisão (Radočaj et al., 2023).

2.2.1. GLI (*Green Leaf Index*)

O GLI destaca a componente verde da vegetação utilizando apenas as bandas de vermelho (*Red*), verde (*Green*) e azul (*Blue*), dispensando o infravermelho próximo (*NIR*) e sendo, por isso, particularmente adequado a câmaras simples, como as instaladas em muitos drones. A possibilidade de detetar sinais de vigor vegetativo apenas a partir de RGB amplia o alcance de análises rápidas e de baixo custo, sobretudo em áreas agrícolas ou ecossistemas suscetíveis a mudanças pontuais (Butt, 2018). Ainda assim, à semelhança de outros índices, o GLI deve ser interpretado com precaução, pois qualquer simplificação numérica pode não refletir a complexidade total dos fatores que influenciam a vegetação - desde condições ambientais até variáveis relacionadas com a intervenção humana, como uso do solo e gestão agrícola - exigindo uma análise criteriosa do contexto local (Butt, 2018; Radočaj et al., 2023).

$$GLI = \frac{(2 * Green) - Red - Blue}{(2 * Green) + Red + Blue}$$

2.2.2. NDVI (*Normalized Difference Vegetation Index*)

O NDVI foi originalmente desenvolvido por Rouse et al. (1973), no *Texas A&M Remote Sensing Center*, como um índice capaz de quantificar a cobertura vegetal com base na resposta espectral da vegetação em imagens de satélite. O NDVI é um dos índices de vegetação mais difundidos, resultado da razão entre a diferença e a soma das bandas do infravermelho próximo (*NIR*) e do vermelho (*Red*).

$$NDVI = \frac{(NIR - R)}{(NIR + R)}$$

Assume valores entre -1 e +1. Os mais próximos de +1 indicam vegetação verdejante, os próximos de 0 podem representar solo nu ou superfícies construídas, e os valores negativos tendem a associar-se aos planos de água (Pettorelli et al., 2005).

A popularidade do NDVI decorre da forte correlação com parâmetros vegetativos, como a biomassa e o índice de área foliar, e da simplicidade de cálculo a partir de sensores multiespectrais. Ainda assim, está sujeito a limitações, como a saturação em cobertos vegetais muito densos - onde variações adicionais de biomassa deixam de ser detetadas, e a diferenças entre sensores - que exigem calibração prévia para garantir a comparabilidade dos dados (Huang et al., 2021). Quando se utilizam sensores distintos, como de drones e de satélites, a calibração direta dos valores de NDVI pode ser inviável devido a diferenças técnicas, sendo comum que os dados sejam tratados separadamente, priorizando análises internas consistentes (Anderson & Gaston, 2013; Müllerová et al., 2016; Huang et al., 2021).

2.2.3. SAVI (*Soil-Adjusted Vegetation Index*)

Proposto por Huete (1988), o SAVI segue a mesma abordagem do NDVI ao combinar as bandas do vermelho (*Red*) e do infravermelho próximo (*NIR*), mas introduz um fator *L* para atenuar a influência do solo exposto. Deste modo, a sua aplicação mostra-se particularmente vantajosa em áreas de vegetação pouco densa, como regiões semiáridas ou nas fases iniciais de crescimento das culturas (Radočaj et al., 2023). Assim, o SAVI tende a oferecer valores mais consistentes do que o NDVI quando há grande proporção de solo visível.

$$SAVI = \left(\frac{NIR - Red}{NIR + Red + L} \right) * (1 + L)$$

O fator *L* é um parâmetro de ajuste utilizado no cálculo do índice SAVI, com o objetivo de minimizar a influência do solo exposto na estimativa da vegetação, especialmente em áreas com cobertura vegetal esparsa. Trata-se de um valor empírico que varia entre 0 e 1, sendo adotado conforme o grau de exposição do solo. O valor padrão mais comum é 0,5, amplamente utilizado como um compromisso para diferentes tipos de cobertura, por proporcionar um equilíbrio entre a correção do solo e a sensibilidade à vegetação (Huete, 1988). O valor $L = 0$ torna o SAVI equivalente ao NDVI, ou seja, sem qualquer correção para o solo, enquanto o valor $L = 1$ corresponde à correção máxima, recomendada para áreas com vegetação extremamente escassa e elevada exposição do solo (Huete, 1988). Esta parametrização permite, eventualmente, ajustar

o índice às características específicas da paisagem, aumentando a fiabilidade dos resultados em contextos heterogéneos.

2.2.4. MSAVI (*Modified Soil-Adjusted Vegetation Index*)

O MSAVI é uma evolução do SAVI, dispensando a atribuição de um valor fixo ao fator de correção e adaptando-se automaticamente às condições de solo exposto. Tal característica promove uma maior precisão em cenários de vegetação esparsa ou em mosaico, onde a fração de solo nu pode alterar significativamente a resposta espectral (Radočaj et al., 2023). Em aplicações de agricultura, silvicultura ou ordenamento do território, o MSAVI contribui para melhor discriminar níveis intermédios de cobertura vegetal, especialmente quando se utilizam imagens de alta resolução.

$$MSAVI = \frac{2 * NIR + 1 - \sqrt{(2 * NIR + 1)^2 - 8 * (NIR - Red)}}{2}$$

2.3. Textura

A introdução de parâmetros texturais na análise de imagens de deteção remota tem assumido importância crescente, já que apenas os valores espectrais podem não ser suficientes para diferenciar classes aparentemente semelhantes (Wang et al., 2021). Entende-se por textura o padrão de variação espacial dos tons de cinzento ou valores de pixel numa imagem, refletindo características como rugosidade, repetição, uniformidade ou irregularidade das superfícies representadas. Para quantificar essa informação utilizam-se os chamados parâmetros texturais, que traduzem numericamente essas características. Um dos métodos mais utilizados para extrair estes parâmetros é a *Gray Level Co-occurrence Matrix* (GLCM), uma técnica estatística de segunda ordem amplamente reconhecida pela sua capacidade para descrever relações espaciais entre pares de pixels com diferentes níveis de cinzento, indo além da simples análise dos valores individuais de cada pixel. Esta matriz analisa a ocorrência conjunta de tons de cinzento entre pixels vizinhos, segundo direções e distâncias pré-definidas, gerando um conjunto de indicadores relevantes para a análise espacial (Haralick, 1986).

Entre os principais indicadores extraídos da GLCM incluem-se o contraste (variação entre tons vizinhos), a homogeneidade (uniformidade dos tons), a entropia (grau de desordem ou

complexidade do padrão) e a correlação (nível de dependência entre os valores dos pixels vizinhos). Estes parâmetros introduzem uma dimensão espacial adicional ao modelo de classificação, permitindo captar padrões de organização que enriquecem a interpretação espectral tradicional. A conjugação entre índices de vegetação e características texturais tende, assim, a produzir classificações mais robustas, sobretudo em contextos com elevada heterogeneidade espacial ou com coberturas com fortes contrastes espectrais (Wang et al., 2021).

2.4. FOSS (*Free and Open Source Software*)

A adoção de soluções de código aberto (FOSS - *Free and Open Source Software*) tem vindo a crescer na área da deteção remota, atendendo à flexibilidade que oferecem e à comunidade de utilizadores que colabora ativamente no seu desenvolvimento (Manomba-Mbadinga et al., 2023). O QGIS, *software* livre bastante difundido, suporta um vasto conjunto de formatos *raster* e vetoriais, bem como a instalação de *plugins* especializados em processamento de imagens de deteção remota.

2.4.1. QGIS

O QGIS é um *software open source* amplamente reconhecido pela sua modularidade e pela comunidade global de utilizadores e programadores (Coetzee et al., 2020; Rosas-Chavoya et al., 2022; Dias et al., 2024). Através de *plugins* e da integração com bibliotecas de processamento, viabiliza a realização de análises avançadas em deteção remota. Constitui, portanto, uma plataforma inclusiva e flexível, adaptável a diferentes contextos geoespaciais (Rosas-Chavoya et al., 2022).

2.4.2. Orfeo Toolbox

O *Orfeo Toolbox*, desenvolvido pelo CNES, é uma biblioteca de processamento de imagens de deteção remota focada na eficiência e na escalabilidade (*OTB Development Team*, 2023). A escalabilidade refere-se à capacidade de um sistema manter ou melhorar o seu desempenho à medida que aumenta o volume de dados ou a complexidade das operações, sendo, por isso, uma característica essencial em contextos de processamento massivo de imagens. A sua integração no QGIS ocorre através de um *plugin* próprio, oferecendo um conjunto amplo de ferramentas,

desde correções radiométricas e geométricas até algoritmos de fusão de imagens e classificação supervisionada (Grizonnet et al., 2017). A capacidade de lidar com grandes volumes de dados torna o OTB particularmente atrativo para projetos de monitorização ambiental em larga escala (Lourenço et al., 2021; Dias et al., 2024). Além disso, contempla métodos de OBIA que permitem segmentar a imagem em objetos, agregando informação de forma mais coerente do que a análise por pixels isolados (Blaschke et al., 2014).

2.5. Fundamentos de *Machine Learning* Aplicados à Detecção Remota

O recurso a ML na deteção remota tem-se consolidado como uma estratégia fundamental para lidar com a crescente disponibilidade de dados fornecidos por satélites e drones, permitindo classificar extensas áreas com elevada diversidade de cobertura (Maxwell et al., 2018; Holloway & Mengersen, 2018). Embora o ML inclua diferentes técnicas, a sua aplicação em deteção remota pode ser agrupada em três paradigmas principais: i. aprendizagem supervisionada, ii. aprendizagem não supervisionada e iii. aprendizagem por reforço.

2.5.1. Aprendizagem Supervisionada

Na aprendizagem supervisionada o algoritmo recebe exemplos rotulados, nos quais cada pixel ou conjunto de pixels conhece a classe de ocupação e uso do solo a que pertence (Maxwell et al., 2018). O modelo aprende a associar as assinaturas espectrais - padrões únicos de resposta da radiação eletromagnética em diferentes comprimentos de onda, característicos de cada tipo de cobertura do solo - e os índices de vegetação às respetivas classes, possibilitando a classificação de novas regiões desconhecidas. Em deteção remota, esse processo implica selecionar amostras de treino fiáveis, ou seja, conjuntos de dados representativos, corretos e bem distribuídos geograficamente que reflitam a variabilidade das classes a identificar, bem como escolher algoritmos como o *Random Forest* ou o *Support Vector Machine* e otimizar parâmetros para reduzir o *overfitting* (Pal, 2005; Belgiu & Drăguț, 2016). A abordagem supervisionada (classificação) tende a ser a preferida em contextos com boa separação espectral e imagens de muito alta resolução espacial, geralmente com pixels inferiores a 1 metro, o que permite detetar objetos de pequena dimensão - como árvores isoladas, construções dispersas ou caminhos estreitos - e produzir mapas finais detalhados e representativos da realidade do terreno.

2.5.2. Aprendizagem Não Supervisionada

Na aprendizagem não supervisionada (ou *clustering*), algoritmos como o *K-means* ou o *ISODATA* agrupam os pixels pela sua semelhança espectral, sem que haja rótulos pré-definidos. Posteriormente, o analista interpreta e atribui uma categoria a cada *cluster*. Embora esta abordagem seja útil em análises exploratórias ou em contextos com dados de terreno limitados, tende a ser menos precisa em áreas de elevada heterogeneidade ou classes espectralmente semelhantes (Maxwell et al., 2018).

2.5.3. Aprendizagem Por Reforço

A aprendizagem por reforço (*Reinforcement Learning*) é um paradigma distinto, no qual um agente (isto é, a entidade que aprende e toma decisões) interage com o ambiente e recebe recompensas ou penalizações pelas ações realizadas (Li et al., 2018). Muito aplicada em robótica e em jogos, é pouco difundida na deteção remota, sobretudo devido às exigências computacionais e à dificuldade de conceber funções de recompensa adequadas. Contudo, abre caminho para abordagens dinâmicas onde o sistema se adapta continuamente a variações ambientais ou temporais.

2.6. Algoritmos de *Machine Learning*

Entre os algoritmos mais populares para a classificação de dados de deteção remota destacam-se o RF e o SVM (Belgiu & Drăguț, 2016; Maxwell et al., 2018). O RF integra o grupo de métodos de *ensemble learning*, enquanto o SVM se baseia em princípios de maximização da margem entre classes.

2.6.1. *Random Forest*

O *Random Forest* recorre a múltiplas árvores de decisão criadas a partir de amostras *bootstrap* do conjunto de treino e de subconjuntos aleatórios de variáveis (Breiman, 2001; Hastie et al., 2009; Pal, 2005; Belgiu & Drăguț, 2016). Cada árvore vota na classificação final e a agregação dessas previsões reduz a variância do modelo e o risco de *overfitting* (Mahesh, 2020; Holloway & Mengersen, 2018). O RF destaca-se pela facilidade de parametrização, pela robustez face a dados 'ruidosos' e pela possibilidade de calcular a importância de cada variável. Em estudos de deteção remota demonstrou elevada eficácia ao integrar bandas multiespectrais e índices de

vegetação (Pelletier et al., 2020). Além disso, a técnica *out-of-bag* (OOB) permite uma estimativa interna de erro, dispensando um conjunto de validação separado (Breiman, 2001).

2.6.2. *Support Vector Machine*

O SVM mostra grande capacidade de generalização, mesmo em cenários de alta dimensionalidade, ou seja, quando há muitas variáveis ou atributos por amostra (Srivastava & Bhambhu, 2006; Maxwell et al., 2018). O princípio fundamental consiste em encontrar um hiperplano ótimo que separe as classes de forma a maximizar a margem, que é a distância entre o hiperplano e os exemplos de treino mais próximos de cada classe (os chamados vetores de suporte) (Srivastava & Bhambhu, 2006). O hiperplano é então a 'fronteira' que separa as classes, podendo ser uma linha (2D), um plano (3D) ou, em espaços com mais dimensões, uma superfície matemática que desempenha o mesmo papel de separação. Essa abordagem tende a reduzir o risco de *overfitting*, mesmo em espaços com muitas dimensões.

Quando os dados não são linearmente separáveis no espaço original o SVM recorre ao chamado *kernel trick*, uma técnica que permite projetar os dados num espaço de dimensão superior onde a separação entre as classes se torna possível (Mountrakis et al., 2011). Um dos *kernels* mais utilizados é o *radial basis function* (RBF), particularmente eficaz em problemas com fronteiras não lineares. No entanto, em situações em que os dados apresentam separabilidade próxima do linear pode ser mais adequado utilizar o *kernel* linear, que aplica o algoritmo diretamente no espaço original dos atributos, com menor complexidade computacional. A eficácia do SVM, independentemente do *kernel* escolhido, depende da seleção criteriosa dos seus parâmetros: o C , comum a todos os *kernels*, controla a penalização por erro de classificação, enquanto o γ (*gamma*), específico do *kernel* RBF, define o alcance da influência dos vetores de suporte (Pal, 2005). Por esse motivo é habitual recorrer-se a validação cruzada para otimizar esses parâmetros. Na área da deteção remota o SVM tem demonstrado resultados competitivos em diversas aplicações, incluindo a utilização de índices de vegetação ou de imagens de elevada resolução (Manomba-Mbadinga et al., 2023).

2.7. Métricas de Validação

A avaliação rigorosa dos resultados de classificação obtidos por ML em deteção remota é essencial para aferir a qualidade dos mapas produzidos e sustentar eventuais tomadas de decisão

(Foody, 2002). Entre as métricas mais recorrentes destacam-se a matriz de confusão, o *F1-score* e o coeficiente *Kappa* (Belgiu & Drăguț, 2016; Maxwell et al., 2018; Hossin & Sulaiman, 2015). Cada uma contribui de modo específico para caracterizar a eficácia do modelo.

2.7.1. Matriz de Confusão

A matriz de confusão organiza os resultados da classificação num quadro de dupla entrada: as linhas representam as classes preditas e as colunas as classes de referência (Foody, 2002). Na diagonal principal encontram-se os acertos, ao passo que as restantes células indicam os erros de classificação. Esta ferramenta não só mostra o total de acertos como também revela quais as classes que tendem a ser confundidas, proporcionando assim pistas para o aperfeiçoamento do modelo (Hossin & Sulaiman, 2015).

2.7.2. *F1-score*

O *F1-score* é calculado como a média harmónica entre a precisão (*precision*) e a revocação (*recall*), focando-se simultaneamente em falsos positivos e falsos negativos. Dado que a precisão mede a proporção de previsões positivas corretas e a revocação avalia o quão bem o modelo identifica todos os positivos, o *F1-score* torna-se particularmente útil em cenários multi-classe, pois equilibra o desempenho entre diferentes classes (Hossin & Sulaiman, 2015). Além disso, ao fornecer uma análise mais detalhada das classes minoritárias, contribui para avaliações mais rigorosas em problemas de classificação complexos, oferecendo uma visão conjunta dos verdadeiros positivos e das instâncias mal classificadas (Grandini et al., 2020).

2.7.3. Coeficiente *Kappa*

O coeficiente *Kappa* (κ) ajusta a exatidão global ao descontar os acertos que ocorreriam ao acaso (Foody, 2002). O valor de κ varia entre -1 e 1, sendo que κ próximo de 1 indica elevada concordância entre as classes previstas e as classes verdadeiras, enquanto κ próximo de 0 sugere que a classificação não supera a aleatoriedade. A partir de valores de κ superiores a 0,41 já se considera uma concordância moderada, sendo que valores acima de 0,61 representam concordância substancial e, acima de 0,81, uma concordância quase perfeita (Landis & Koch, 1977). Ainda que seja amplamente utilizado em estudos de deteção remota, deve ser aplicado com cuidado em ambientes com grande desbalanceamento de classes, pois a distribuição marginal das classes pode influenciar significativamente a interpretação dos resultados

(Delgado & Tibau, 2019). Em contrapartida, quando associado a outras métricas, o *Kappa* é útil para aferir o nível de concordância e contribuir para uma avaliação mais sólida do modelo (Figueiredo & Vieira, 2007).

2.8. Síntese

Este enquadramento teórico sublinha a utilidade dos índices de vegetação (GLI, NDVI, SAVI e MSAVI) e da informação textural (GLCM) na classificação da ocupação e uso do solo, evidenciando o contributo de algoritmos como o RF e a SVM em ambientes de deteção remota. A integração com *software* livre, nomeadamente o QGIS, e a adoção do *plugin* OTB permite estabelecer fluxos de trabalho completos e acessíveis, sustentados por métricas de validação credíveis.

3. MATERIAIS, MÉTODOS E FLUXO DE TRABALHO

3.1. Materiais

3.1.1. Fontes de dados

Para assegurar uma análise geoespacial robusta e multiescala da ocupação e uso do solo da R-FB foi utilizado um conjunto diversificado de dados, integrando imagens de satélite, levantamentos por UAV e dados auxiliares derivados da cartografia oficial.

3.1.1.1. Imagens de Satélite

Uma das fontes de informação espectral consistiu nas imagens provenientes da constelação Pléiades-Neo (satélites 3 e 4), disponibilizadas através do produto OrtoSat2023 do Sistema de Monitorização da Ocupação do Solo (SMOS) (Direção-Geral do Território, 2023). Estas imagens, adquiridas entre abril e outubro de 2023, têm resolução espacial de 30 cm, permitindo uma representação detalhada da superfície terrestre. Foram disponibilizadas em duas composições espectrais:

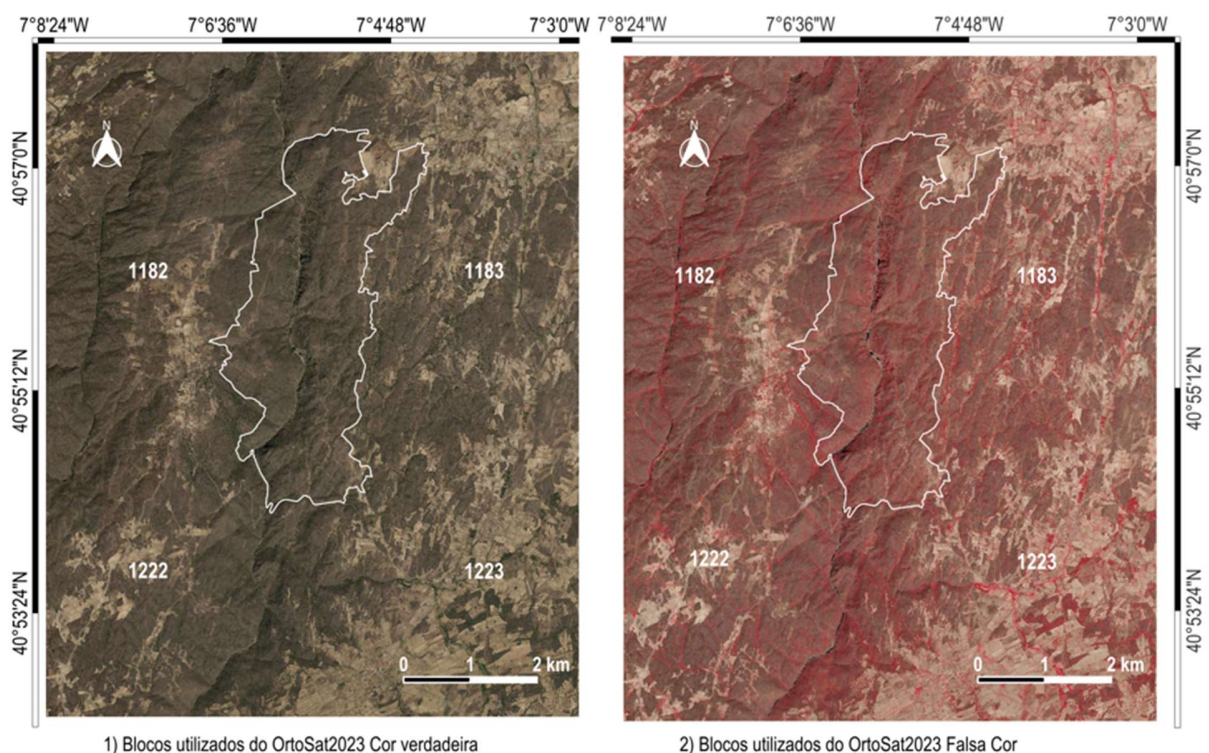


Figura 3.1 - Blocos das composições OrtoSat2023

(1) cor verdadeira (RGB), combinando as bandas do vermelho, verde e azul para uma visualização natural da paisagem; e (2) falsa cor (IRG), que substitui a banda azul pelo infravermelho próximo (*NIR*), maximizando o contraste da vegetação e facilitando a discriminação de cobertos naturais. As imagens utilizadas corresponderam aos blocos 1182, 1183, 1222 e 1223 (Fig. 3.1), posteriormente unidos e recortados segundo os quadrados da grelha UTM da R-FB, utilizando as ferramentas *Merge* e *Clip raster by mask layer* do QGIS.

3.1.1.2. Imagens UAV

Complementarmente, foi realizada uma campanha de aquisição de imagens por UAV entre agosto e setembro de 2024, com recurso a um drone DJI Mavic 3M¹, equipado com sistema GNSS RTK (*Real-Time Kinematic*) para aumentar a precisão posicional. Os voos foram planeados para garantir uma cobertura uniforme da área, com sobreposição frontal e lateral otimizadas (80 frontal e 70 lateral), a uma altitude de 102,4 m e velocidade de 10 m/s, compatíveis com os requisitos fotogramétricos. As imagens obtidas, com resolução espacial de 2,76 cm, foram georreferenciadas no sistema ETRS89/PT-TM06, assegurando a coerência com os restantes dados do projeto (Fig. 3.2).

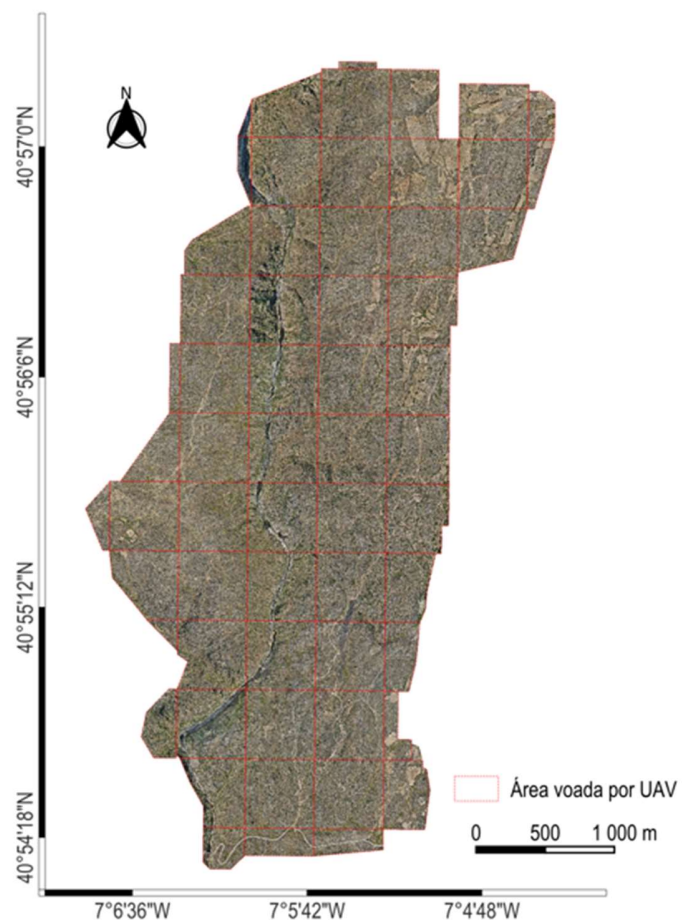


Figura 3.2 - Área de cobertura do OrtoFotoUAV

¹ <https://ag.dji.com/mavic-3-m>

3.1.1.3. Dados Auxiliares

Para enriquecer a análise foram incorporados dados auxiliares de relevância espacial e temática. Entre eles destacam-se os Modelos Digitais de Terreno (MDT), gerados a partir das curvas de nível da Carta Militar de Portugal à escala 1:25.000, com equidistância de 10 metros, permitindo caracterização do relevo da área de estudo.

Os limites oficiais da R-FB, fornecidos pela Faia Brava - Associação de Conservação da Natureza, foram essenciais para a delimitação precisa da área de estudo e para a organização dos dados num Sistema de Informação Geográfica (SIG) coerente. Foram disponibilizados: (i) uma grelha UTM inicial, utilizada para estruturar e organizar espacialmente os dados e facilitar operações de recorte e mosaico; (ii) uma segunda grelha UTM, recortada com base no limite do voo do drone, permitindo a análise mais direcionada dos produtos UAV; e (iii) um polígono com a delimitação oficial da R-FB, que foi utilizado como base para a aplicação e validação final do modelo de ML (Fig. 3.3).



Figura 3.3 - Grelhas utilizadas no estudo

Para operacionalizar a análise espacial foi selecionada a quadrícula 21 (Q21) dentro da R-FB, devido à sua diversidade paisagística e representatividade ecológica (Fig. 3.4). Essa escolha permitiu o treino de um modelo robusto de classificação, com potencial para ser generalizado e aplicado às restantes quadrículas da R-FB.

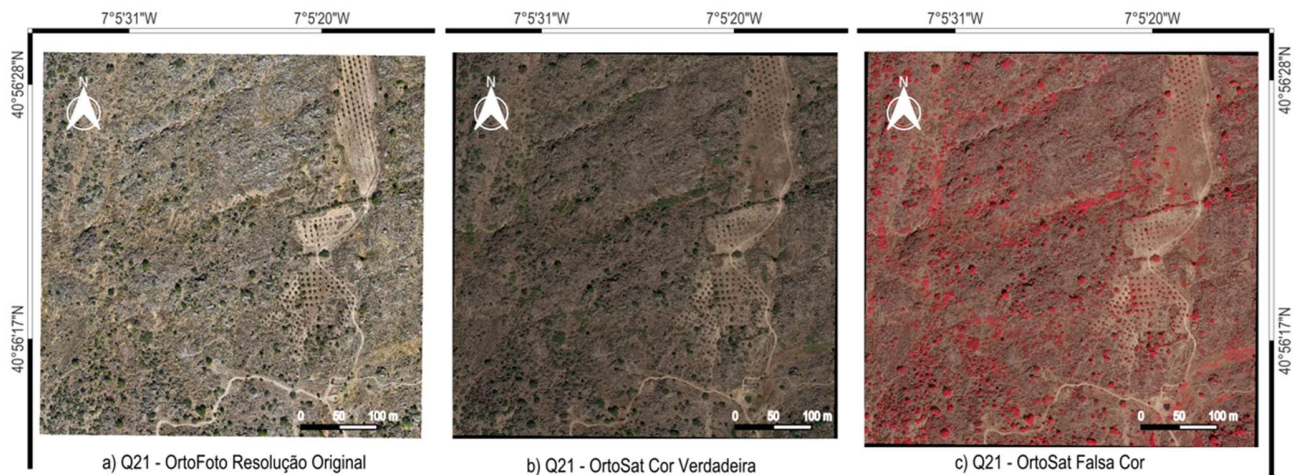


Figura 3.4 - Quadrícula 21

3.1.2. Softwares e Ferramentas

Para a realização da presente investigação, recorreu-se exclusivamente a *software* livre e de código aberto, assegurando acessibilidade, reprodutibilidade e alinhamento com os princípios da democratização tecnológica. As ferramentas utilizadas, bem como as respectivas funcionalidades e etapas metodológicas em que foram aplicadas, são descritas em seguida:

- QGIS 3.34.13 ‘Prizren’: Constituiu o ambiente SIG central de trabalho, sendo utilizado em diversas fases do processo, nomeadamente na integração e visualização dos dados, georreferenciação de imagens, recorte espacial com base nos limites da Reserva da Faia Brava, vetorização de pontos de controlo, cálculo de índices de vegetação, amostragem de áreas de treino e validação, e análise estatística básica dos desvios espaciais (através da funcionalidade *Statistical Summary*).
- OTB: Funcionou como extensão do QGIS, sendo fundamental nas fases de concatenação de variáveis raster, extração de métricas texturais com base na GLCM (*Gray-Level Co-occurrence Matrix*) e, sobretudo, na classificação supervisionada das imagens, através

da implementação dos algoritmos RF e SVM. O OTB foi ainda responsável pela normalização automática das variáveis de entrada no momento do treino dos modelos, dispensando a necessidade de reescalagem manual para o intervalo [0-1].

- GDAL (*Geospatial Data Abstraction Library*, acessível via QGIS): Utilizado essencialmente nas tarefas de reamostragem espacial (redução da resolução das imagens UAV para 30 cm/pixel), normalização radiométrica (através da expressão `-scale`, no parâmetro avançado na ferramenta *Translate*) e conversão de formatos raster. Estas operações garantiram a compatibilidade entre produtos de diferentes fontes e a integridade dos dados ao longo do processo.

Estas ferramentas, articuladas entre si, permitiram o desenvolvimento de um fluxo de trabalho robusto, flexível e totalmente baseado em *software* livre.

3.2. Métodos

3.2.1. Pré-Processamento

Um pré-processamento rigoroso foi essencial para garantir a compatibilidade espacial e radiométrica entre os diferentes produtos utilizado, nomeadamente os ortofotomosaicos UAV (OrtoFotoUAV) e as imagens OrtoSat2023, assegurando que todas as camadas estivessem devidamente alinhadas e prontas para integração em modelos de classificação supervisionada.

Embora o OTB disponibilize um vasto conjunto de ferramentas para operações de pré-processamento, em algumas etapas específicas optou-se pelo uso de funcionalidades nativas do QGIS. Essa escolha deve-se, sobretudo, à maior flexibilidade, simplicidade operacional ou visualização interativa que essas ferramentas oferecem, alinhando-se com o propósito deste trabalho: promover fluxos de trabalho acessíveis, eficientes e adaptados às necessidades práticas de técnicos e investigadores com diferentes níveis de literacia digital.

3.2.1.1. Correção Geométrica e Alinhamento Espacial

O OrtoSat2023 foi inicialmente fornecido com ortorretificação e correções geométricas aplicadas com base no modelo digital de elevação WorldDEM4Ortho (dados.gov, s.d.²).

² <https://dados.gov.pt/pt/datasets/ortosat-30-cm-portugal-continental-2023/>

Contudo, a precisão posicional média (~2,2 m) revelou-se insuficiente para a sua integração com o OrtoFotoUAV, que apresenta uma precisão submétrica teórica, garantida por GNSS RTK.(AMA, 2023)

Para mitigar essa discrepância o OrtoSat2023 foi regeorreferenciado no QGIS, utilizando a ferramenta **Georeferencer** (*Layer > Georeferencer...*). Para utilizar o **Georeferencer** foi necessário encontrar os desvios médios entre pontos de controlo do OrtoFotoUAV e do OrtoSat2023, extraídos em duas camadas vetoriais de pontos. O OrtoFotoUAV foi utilizado como referência geométrica, sendo definidos 13 pontos de controlo visíveis em ambas as imagens, aleatoriamente distribuídos na área de estudo (Fig. 3.5.).

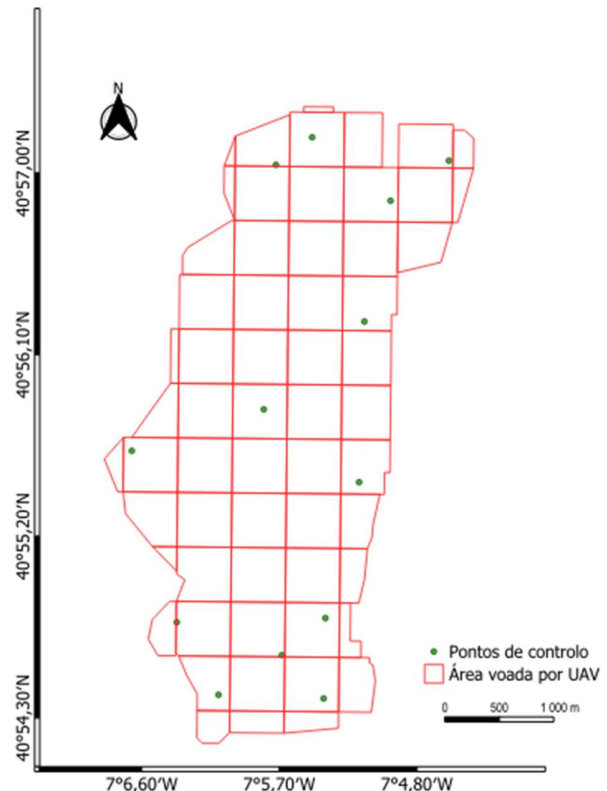


Figura 3.5 - Pontos de controlo

Após a digitalização dos pontos foram preenchidos os campos das coordenadas X e Y nas respetivas tabelas de atributos (para o OrtoSat2023 e o OrtoFotoUAV). As duas camadas vetoriais foram depois unidas com base num campo de identificação comum (*ID*), através da ferramenta **Join Attributes by Field Value**, consolidando, assim, os pares de coordenadas correspondentes e permitindo a análise direta das discrepâncias posicionais entre os pontos homólogos extraídos das duas imagens.

Foram então criados dois campos - **Dif_X** e **Dif_Y** - utilizando a ferramenta **Field Calculator**. Como complemento à análise, e por curiosidade técnica, foi criado um campo adicional denominado **dist_m**, com o objetivo de calcular a distância euclidiana em metros entre cada par

de pontos correspondentes. Com base nos valores obtidos na tabela 3.1, procedeu-se à análise estatística das diferenças, recorrendo à funcionalidade *Statistical Summary*, disponível em *View > Statistical Summary*, permitindo calcular os desvios médios nas direções X e Y.

Tabela 3.1 - Tabela de atributos da shapefile dos pontos de controlo

id	OrtoSat_X	OrtoSat_Y	OrtoFoto_X	OrtoFoto_Y	Dif_X	Dif_Y	Dist_m
1	86533,577550	138705,964700	86534,922540	138708,371500	-1,345	-2,407	2,757
2	86916,633720	138040,378000	86916,159780	138042,558100	0,474	-2,180	2,231
3	87498,542870	138403,906400	87498,295110	138405,109900	0,248	-1,204	1,229
4	87898,922190	138744,368800	87898,827410	138746,643700	0,095	-2,275	2,277
5	87882,535710	138006,183200	87882,820070	138008,363300	-0,284	-2,180	2,199
6	88497,618370	142574,222600	88497,896390	142575,564700	-0,278	-1,342	1,371
7	87444,773730	142904,047800	87445,348940	142905,946000	-0,575	-1,898	1,983
8	87332,914620	140660,384900	87332,416100	140662,033800	0,499	-1,649	1,723
9	86120,238380	140279,500200	86119,663170	140281,417600	0,575	-1,917	2,002
10	88257,065680	141466,023600	88257,008160	141467,672500	0,058	-1,649	1,650
11	89033,560410	142942,433500	89033,713800	142943,277100	-0,153	-0,844	0,857
12	88208,594670	139991,262600	88208,517980	139993,180000	0,077	-1,917	1,919
13	87777,053190	143156,833300	87777,264100	143158,185000	-0,211	-1,352	1,368

A partir destes valores, realizou-se a retificação geométrica do OrtoSat2023. No módulo *Georeferencer* selecionaram-se dois pontos de controlo ajustados segundo os desvios médios identificados. A transformação escolhida foi do tipo linear, e a interpolação utilizada foi *Nearest Neighbour* (NN). O novo raster foi exportado com um nome distinto, garantindo a preservação do original. Optou-se ainda por guardar os pontos de controlo, assegurando uma consistência geométrica para os restantes blocos da série OrtoSat2023.

Concluída a regeorreferenciação, as imagens (em cor verdadeira e falsa cor) foram recortadas com base nos limites oficiais da Reserva da Faia Brava. Para garantir o alinhamento espacial entre os diferentes produtos foi desmarcada a opção *Match the extent of the clipped raster to the extent of the mask layer*, assegurando a preservação da extensão do raster de referência (o OrtoFotoUAV).

Após concluídas estas etapas todos os produtos raster passaram a apresentar alinhamento espacial, resolução e extensão uniformes, estando prontos para serem utilizados nas fases seguintes de concatenação de variáveis e classificação supervisionada.

3.2.1.2. Reamostragem das Imagens UAV

A resolução nativa dos OrtoFotoUAV era de 2,7 cm/pixel, o que, embora altamente vantajoso para análises locais detalhadas, poderia causar sobrecarga computacional nos processos seguintes. Para mitigar essa limitação computacional e garantir a compatibilidade com a resolução espacial do OrtoSat2023 foi realizada uma reamostragem espacial (*down sampling*) para a mesma resolução do OrtoSat2023, de 30 cm/pixel.

Este processo foi executado no QGIS com recurso à ferramenta *Warp (Reproject)* do GDAL, acessível em *Raster > Projections > Warp (Reproject...)*, utilizando o método de interpolação *cubic (4×4 kernel)*. Este método foi escolhido por ser particularmente adequado a imagens contínuas, uma vez que preserva a suavidade das transições tonais e a fidelidade radiométrica dos dados originais. A reprojeção foi aplicada individualmente a cada uma das 60 quadrículas cobertas pelo voo do drone, procedendo-se em seguida à sua união num único mosaico contínuo, recorrendo à ferramenta *Merge* no QGIS (*Raster > Miscellaneous > Merge...*). Esta operação permitiu consolidar toda a cobertura UAV num único raster, correspondente à área total da zona de estudo, garantindo uniformidade espacial e facilitando as etapas subsequentes de análise e classificação.

3.2.1.3. Remoção da Banda Alfa

Durante o pré-processamento do OrtoFotoUAV identificou-se a presença de uma banda alfa (canal de transparência). Esta banda, por não conter informação espectral relevante, representava um risco potencial para o processo de classificação supervisionada, uma vez que poderia introduzir valores nulos ou artefactos nas composições multibanda, comprometendo a integridade dos dados.

A remoção foi realizada com recurso à ferramenta *Translate (Convert format)* do GDAL, adicionando nos parâmetros avançados a expressão `-b 1 -b 2 -b 3`, que instruem o programa a manter apenas as bandas RGB válidas. O resultado foram imagens de três bandas com conteúdo espectral consistente e pronto para análise.

3.2.1.4. Normalização Radiométrica

Após o recorte do OrtoSat2023 verificou-se que os valores espectrais apresentavam intervalos reduzidos em todas as bandas, o que poderia comprometer tanto a análise espectral como o

desempenho dos algoritmos de classificação. Para ultrapassar esta limitação foi aplicada uma normalização linear ao intervalo [0–255], utilizando o parâmetro adicional `-scale` na ferramenta *Translate (Convert format)*.

Quando utilizado sem argumentos explícitos o parâmetro `-scale` deteta automaticamente os mínimos e máximos de cada banda e reescala-os para o intervalo desejado, preservando as proporções radiométricas internas.

Importa sublinhar que não foi feita uma reescalagem adicional para o intervalo [0-1] visto que o OTB, utilizado para a concatenação e classificação das imagens, realiza internamente a normalização dos dados de entrada. Tal funcionalidade assegura a uniformização automática das variáveis durante o treino e classificação dos modelos.

A normalização das bandas espectrais é especialmente relevante no cálculo de determinados índices de vegetação, sobretudo os que envolvem expressões com potências ou radicais, como o SAVI e o MSAVI, que são altamente sensíveis à escala dos valores de reflectância. A utilização de dados não normalizados nestes casos pode originar distorções nos resultados, conduzindo a valores fora do intervalo teórico ou a interpretações erradas da cobertura vegetal.

Por outro lado, índices como o NDVI e o GLI, que se baseiam em razões entre bandas, são menos influenciados pela escala absoluta, desde que todas as bandas estejam coerentemente reescaladas. Ainda assim, para garantir a comparabilidade entre sensores distintos, sobretudo em análises temporais, a normalização é recomendada para todos os índices, assegurando consistência matemática e maior fiabilidade aos resultados obtidos.

3.2.1.5. Valorização dos Dados UAV

O OrtoFotoUAV, de muito alta resolução, desempenhou um papel estratégico em três vertentes principais: (i) validação visual e pontual dos resultados da classificação automática, sobretudo em áreas com elevada complexidade espectral; (ii) caracterização detalhada da estrutura da vegetação, permitindo identificar variações não perceptíveis em imagens de menor resolução; e (iii) análise de elementos paisagísticos de pequenas dimensões, que exigem um nível de detalhe superior ao proporcionado pelas imagens de satélite.

A integração sistemática desses dados permitiu uma abordagem multiescala robusta, combinando a abrangência espacial dos dados orbitais com a precisão submétrica dos levantamentos por UAV.

3.2.2. Derivação de Variáveis

Nesta etapa foram extraídas variáveis espectrais e texturais com o objetivo de enriquecer os dados de entrada nos modelos de classificação supervisionada, maximizando a capacidade discriminativa das diferentes classes de uso e ocupação do solo na R-FB.

3.2.2.1. Índices de Vegetação

Após finalizada a preparação das bases de trabalho foram feitos os cálculos para toda a extensão da Faia Brava, dos índices de vegetação a utilizar no trabalho, GLI, NDVI, SAVI e MSAVI, usando o *Raster Calculator* do QGIS.

Os índices NDVI, SAVI e MSAVI foram calculados a partir do OrtoSat2023 de falsa cor (*NIR-Red-Green*), enquanto o índice GLI foi aplicado às imagens RGB do OrtofotoUAV e do OrtoSat2023.

Os resultados foram mantidos em formato *raster* de vírgula flutuante (*float32*) para manter a precisão dos valores, aquando do concatenamento.

Na figura 3.6 apresenta-se um extrato da área de estudo (Q21), com o objetivo de representar visualmente cada um dos índices de vegetação. Para o SAVI foi utilizado o valor de 0,5 como fator de correção do solo, L, conforme recomendado na literatura científica.

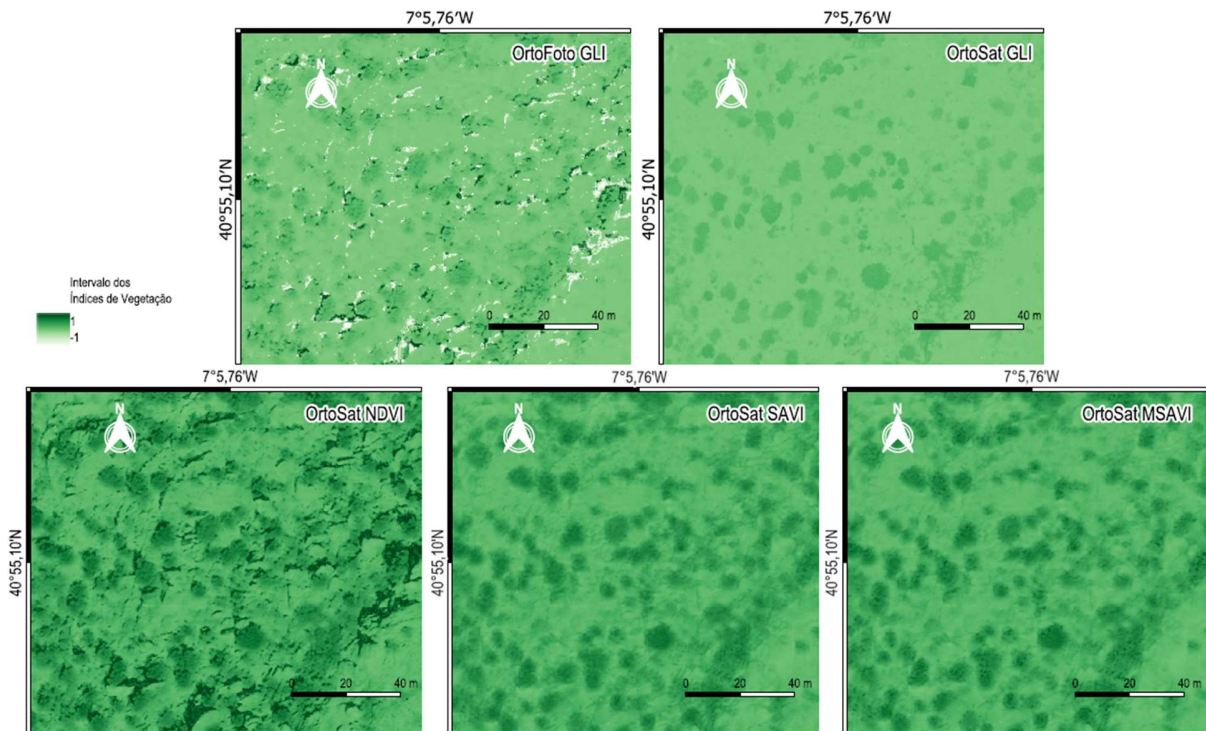


Figura 3.6- Índices de vegetação

3.2.2.2. Texturas de Haralick

A introdução de variáveis texturais na classificação supervisionada visa complementar a informação espectral, especialmente em contextos onde diferentes tipos de cobertura do solo apresentam assinaturas espectrais semelhantes. Na área de estudo da Faia Brava, marcada por elevada heterogeneidade espacial, estruturas como arbustos, herbáceas e afloramentos rochosos podem beneficiar de uma análise mais detalhada baseada em padrões de disposição dos pixels. Nesse sentido, optou-se pela utilização da metodologia baseada na GLCM.

Espera-se que a inclusão destas métricas texturais contribua positivamente para a performance dos algoritmos de classificação, nomeadamente na capacidade de distinguir entre classes com resposta espectral semelhante, mas padrões espaciais distintos. A verificação da sua utilidade efetiva será determinada nos resultados obtidos a partir da aplicação dos algoritmos RF e SVM, analisados posteriormente com base nas métricas de validação, calculadas internamente pelo OTB.

A extração das texturas foi realizada através do módulo *HaralickTextureExtraction*, disponível no OTB. Foram selecionadas duas bandas como base para o cálculo das texturas: a banda vermelha (*Red*), proveniente da imagem RGB, e a banda do infravermelho próximo (*NIR*), da composição em falsa cor. A escolha destas bandas decorreu da sua relevância para a diferenciação da vegetação e da sua complementaridade na resposta espectral. A janela de cálculo foi definida como 15×15 pixels (equivalente a um raio de 7 pixels), equilibrando a necessidade de suavização de variações locais com a da preservação de padrões estruturais significativos. O deslocamento (*offset*) foi estabelecido em 1 pixel, de forma a capturar relações de vizinhança imediata, e o número de níveis de cinzento foi fixado em 32, permitindo detalhar as transições tonais sem comprometer o desempenho computacional.

Foram extraídas oito métricas de textura (*Simple*³) (HS), cada uma com um papel específico na caracterização dos padrões espaciais. A métrica *Energy* (B1) mede a uniformidade da textura, sendo elevada em regiões homogéneas; *Entropy* (B2) avalia o grau de aleatoriedade, indicando complexidade estrutural; *Correlation* (B3) expressa a relação linear entre pixels vizinhos, refletindo regularidade; *Homogeneity* (B4), também conhecida como *Inverse Difference Moment* (IDM), quantifica a suavidade da imagem; *Contrast* (ou *Inertia*) (B5) mede a variação local, sendo sensível a mudanças abruptas de intensidade; *Cluster Shade* (B6) avalia a assimetria da distribuição tonal; *Cluster Prominence* (B7) mede o grau de acentuação (*kurtosis*) e destaca texturas extremas; por fim, *Haralick Correlation* (B8) fornece uma medida ajustada da correlação espacial, reforçando a deteção de padrões repetitivos.

A extração destas métricas foi efetuada antes do recorte final pela área de estudo, para evitar distorções nos bordos provocadas por janelas de cálculo incompletas. Todos os produtos foram guardados em formato de virgula flutuante (*float32*), assegurando a precisão dos valores contínuos e evitando perdas de informação durante o processamento subsequente. Após a geração, as texturas foram recortadas segundo os limites da Faia Brava e integradas nas composições multibanda juntamente com as bandas espectrais e os índices de vegetação, formando conjuntos de dados mais completos para a modelação.

³ https://www.orfeo-toolbox.org/CookBook/Applications/app_HaralickTextureExtraction.html

A fim de ilustrar visualmente as métricas texturais extraídas, apresenta-se abaixo (Fig.3.7) um exemplo representativo das oito bandas de textura de *Haralick*, calculadas a partir do OrtoFotoUAV para o Q21. Esta amostra permite observar a variação espacial de cada métrica, evidenciando o seu potencial para distinguir padrões estruturais associados a diferentes tipos de cobertura do solo.

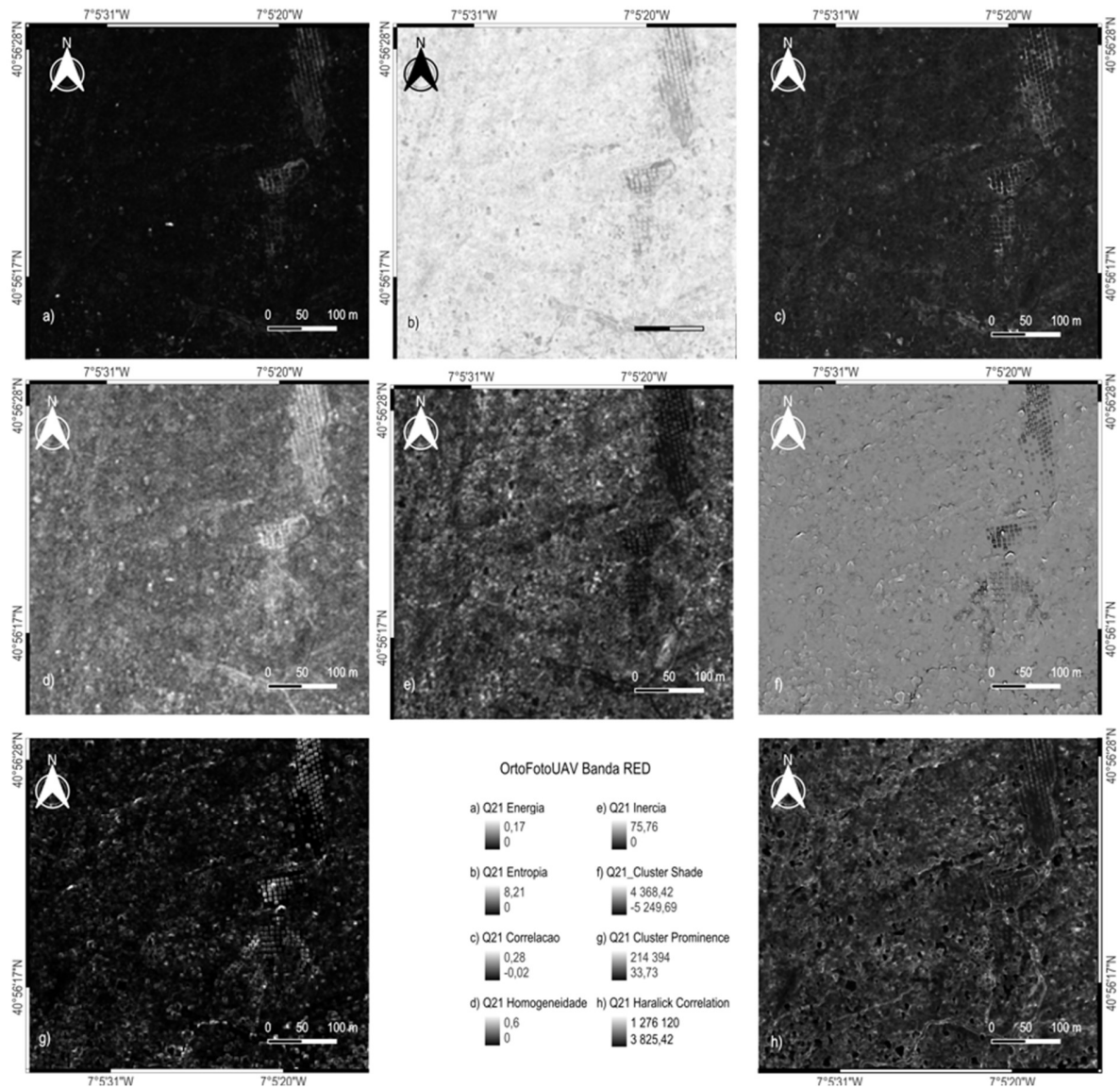


Figura 3.7 - Texturas Simples de Haralick

3.2.3. Concatenação

Com as diferentes camadas previamente tratadas, uniformizadas e normalizadas quanto à sua geometria e resolução espacial, iniciou-se a fase de construção dos conjuntos multibanda a utilizar na modelação supervisionada. Esta etapa foi realizada com o auxílio da ferramenta *ConcatenateImages* do OTB, a qual permite agregar diversas bandas raster numa única imagem composta, facilitando o treino de algoritmos sobre múltiplas variáveis.

O processo seguiu uma lógica incremental e exploratória. Numa primeira fase foram concatenadas as três bandas visíveis do OrtoFotoUAV com o índice de vegetação GLI, criando uma composição base. Em seguida foram integradas camadas de textura de forma isolada, permitindo testar o impacto individual de cada métrica sobre o desempenho da classificação. Finalmente foram adicionadas em conjunto as oito texturas *Haralick*, derivadas da matriz de coocorrência de níveis de cinzento (GLCM), reconhecidas pelo seu potencial para capturar padrões espaciais e variações locais, características particularmente relevantes em mosaicos de ocupação do solo complexos como o da Faia Brava. O mesmo procedimento foi feito para cada índice de vegetação.

Após cada combinação de bandas foi feito um recorte pela quadrícula 21 e utilizou-se o módulo *ComputeImageStatistics* para gerar um ficheiro *.xml* contendo estatísticas descritivas por banda (mínimo, máximo, média e desvio padrão), da área de treino do modelo. Esse ficheiro tem um papel fundamental: permite que o algoritmo aplique automaticamente a normalização interna das variáveis, assegurando que todas as bandas, independentemente da sua escala original - como valores espectrais entre 0 e 255 ou índices de vegetação entre -1 e 1 - sejam tratadas de forma equilibrada. Esta funcionalidade é especialmente importante para algoritmos sensíveis à amplitude dos dados, como o SVM, evitando a necessidade de normalização manual e assegurando que todas as variáveis contribuem corretamente para o modelo.

3.2.4. Amostragem e Definição das Classes

3.2.4.1. Amostragem

A definição das amostras de treino e validação foi uma etapa crítica na implementação da classificação supervisionada. Com base na tipologia dominante da paisagem da Reserva da Faia Brava e na interpretação visual combinada dos OrtoSat2023 e do OrtoFotoUAV foram definidas quatro classes temáticas principais:

- árvores e arbustos
- vegetação herbácea
- afloramentos rochosos
- outras ocupações e usos do solo (como caminhos, solo nu, construções e água)

As amostras foram desenhadas manualmente no QGIS, com base na interpretação combinada do OrtoFotoUAV e do OrtoSat2023, apoiada pelo ortofotomosaico na sua resolução original e baseadas na quadrícula 21. O processo resultou em duas *shapefiles* distintas: uma destinada ao treino e outra à validação. Ambas foram construídas com distribuição espacial representativa e sem sobreposição entre si, assegurando a independência entre os dados de treino e de validação (ver Fig. 3.8).

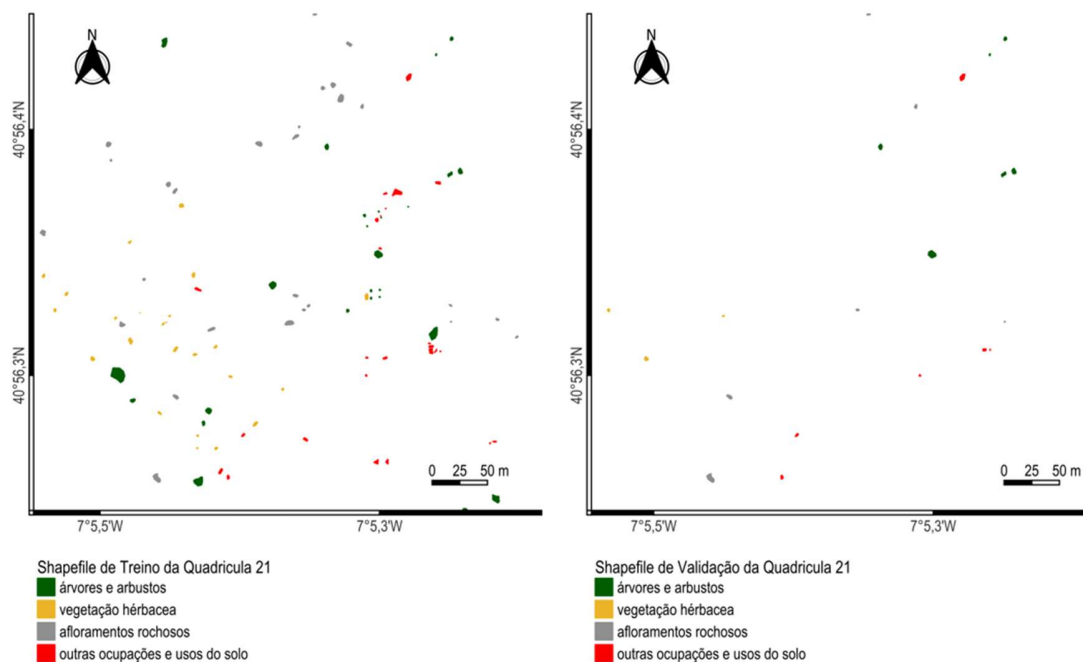


Figura 3.8 - Shapefiles de Treino e Validação - Q21

Embora tenha sido realizada uma análise preliminar da área ocupada por cada classe, não foi necessário balanceamento manual das amostras, pois o módulo *TrainImagesClassifier* do OTB integra a funcionalidade *bound sample by minimum*, a qual ajusta automaticamente o número de amostras com base na classe com menor representação.

3.2.4.2. Configuração e Seleção de Amostras

Na interface do *TrainImagesClassifier* foram definidos os seguintes parâmetros principais:

- Imagem multibanda de entrada (já concatenada)
- *Shapefiles* separadas de treino e validação
- Ficheiro *.xml* com estatísticas da imagem, utilizadas para normalização automática
- Número máximo de amostras por classe (valor introduzido por defeito: 1000)

Apesar do valor 1000 ter sido definido como limite, o OTB ultrapassou automaticamente esse número ao aplicar o critério da classe menos representada. No caso em análise foram utilizadas 1081 amostras por classe no treino (total de 4324) e 233 por classe na validação (total de 932), conforme detalhado na Tabela seguinte:

Tabela 3.2- Valores utilizados para Treino e Validação

Classe	Total disponível (treino)	Amostras usadas (treino)	% utilizada (treino)	Total disponível (validação)	Amostras usadas (treino)	% utilizada (validação)
1	4155	1081	26.0%	778	233	29.9%
2	1081	1081	100%	233	233	100%
3	2164	1081	49.9%	560	233	41.6%
4	1090	1081	99.2%	354	233	65.8%

Este processo garantiu uma amostragem equilibrada entre categorias, evitando a sobre-representação de classes dominantes e assegurando um treino mais justo e robusto. A opção de fornecer *shapefiles* distintas para treino e validação também permitiu ignorar o parâmetro

Training and validation sample ratio. Esta configuração garante um controlo total e explícito sobre as fases de treino e de validação, promovendo maior rigor e replicabilidade no processo.

3.2.4.3. Seleção Espacial de Pixels

A seleção dos pixels dentro de cada amostra é realizada automaticamente pelo OTB, com base em estratégias espaciais periódicas e assegurando uma distribuição homogênea dos dados dentro de cada classe. Esta abordagem reforça a representatividade espacial e contribui para a coerência do treino.

A imagem multibanda de entrada continha aproximadamente 2,8 milhões de pixels, mas o número de amostras utilizado foi considerado suficiente, visto que o objetivo não é classificar todos os pixels manualmente mas sim fornecer exemplos representativos que permitam ao modelo reconhecer padrões espectrais e espaciais característicos de cada classe.

3.2.4.4. Integração no Fluxo de Classificação

Após a definição e configuração das amostras procedeu-se ao treino dos modelos RF e SVM (ver Secção 3.2.5), cujos parâmetros foram ajustados com base em testes iterativos. O módulo **TrainImagesClassifier** do OTB foi utilizado para este fim, integrando as seguintes operações de forma automatizada:

- Normalização das variáveis com base no ficheiro *.xml*
- Seleção interna dos pixels das amostras com distribuição espacial equilibrada
- Aplicação do critério da classe com menor número de amostras para uniformizar o treino
- Treino dos algoritmos com os dados fornecidos

O modelo com melhor desempenho foi então aplicado à totalidade da imagem multibanda através do módulo **ImageClassifier**. A *shapefile* de validação foi utilizada para calcular métricas de desempenho internas, nomeadamente matriz de confusão *F1-score* e coeficiente *Kappa*, permitindo uma avaliação objectiva da qualidade da classificação.

3.2.5. Modelos de Classificação

Dando seguimento ao processo iniciado com o módulo *TrainImagesClassifier*, após a definição dos dados de entrada (imagem multibanda, *shapefiles* de treino e validação e ficheiro de estatísticas), passou-se à etapa seguinte do quadro da ferramenta, dedicada à escolha e parametrização dos algoritmos de classificação. Foram testados dois algoritmos supervisionados: RF e SVM, com o objetivo de comparar a sua eficácia na distinção das quatro classes de uso e ocupação do solo definidas previamente.

A implementação de cada algoritmo foi cuidadosamente ajustada com base nos parâmetros disponibilizados pelo OTB, respeitando as especificidades de cada método e explorando as opções que maximizassem o seu desempenho dentro do contexto dos dados disponíveis.

3.2.5.1. Random Forest

O modelo foi treinado através do módulo *TrainImagesClassifier*, o que permitiu realizar todo o processo sem necessidade de utilizar linhas de comando.

A configuração dos parâmetros foi precedida de uma série de testes empíricos com o objetivo de identificar os valores mais adequados para o equilíbrio entre precisão, generalização e tempo de processamento. O número máximo de árvores na floresta foi fixado em 500, garantindo estabilidade estatística sem comprometer o desempenho computacional. A profundidade máxima das árvores foi limitada a 20 níveis, evitando modelos excessivamente complexos. Para permitir uma segmentação detalhada das classes, definiu-se um mínimo de 2 amostras por nó terminal. O número de variáveis aleatórias consideradas em cada divisão foi deixado a 0, o que permite ao OTB aplicar a regra padrão, adaptada automaticamente à dimensionalidade dos dados. O critério de terminação baseado no erro OOB (*out-of-bag*) foi fixado em 0,01, funcionando como limite mínimo de melhoria para a continuação do treino. O valor da semente aleatória (*Random Seed*) foi fixado em 42, por ser um valor simbólico, imortalizado na obra de Douglas Adams – *Hitch-Hikers Guide to the Galaxy* - e por razões de reprodutibilidade, permitindo obter resultados consistentes ao repetir o mesmo processo com os mesmos dados. Os parâmetros relativos a modelos de regressão (como o critério de terminação) e a dados altimétricos (como DEM, ficheiros geoidais ou elevação padrão) não foram utilizados, por não serem relevantes neste contexto.

O processo gerou automaticamente dois ficheiros principais: o modelo treinado (*.model*), posteriormente utilizado para classificar a imagem completa, e a matriz de confusão (*.csv*), que permitiu avaliar o desempenho com base em métricas como precisão, revocação, *F1-score* e coeficiente *Kappa*. Esta metodologia demonstrou ser eficaz e acessível, possibilitando a execução de uma classificação supervisionada rigorosa e replicável, ancorada numa plataforma de código aberto e sem dependência de linhas de comando.

3.2.5.2. Support Vector Machine (SVM)

No algoritmo SVM, também implementado através do módulo *TrainImagesClassifier*, optou-se pelo *kernel* linear, após se verificar que outras funções, como o radial (RBF), não apresentavam vantagens em termos de desempenho para os dados em questão *versus* custo computacional. A escolha do *kernel* linear simplifica consideravelmente a configuração do modelo, uma vez que apenas o parâmetro *C* é efetivamente considerado - os restantes, como o γ (*gamma*), o grau, o coeficiente ou o parâmetro *Nu*, são ignorados pelo OTB por não influenciarem o funcionamento do modelo com esse tipo de *kernel*.

Todos os parâmetros, exceto três, foram mantidos nos valores por defeito: foi definida uma *random seed* (42) para garantir a reprodutibilidade dos resultados; ativou-se a opção *Parameters optimization*, permitindo ao OTB ajustar automaticamente os parâmetros relevantes dentro de intervalos predefinidos, sem necessidade de calibração manual; ativou-se a *Probability estimation*, o que possibilitou obter não apenas a classe atribuída, mas também a probabilidade associada a cada predição, enriquecendo a interpretação dos resultados.

Importa referir, novamente, que o OTB ignora automaticamente todos os parâmetros que não são relevantes para o algoritmo selecionado. Esta gestão automática evita conflitos e facilita a configuração do fluxo de trabalho.

O processo também gerou automaticamente o modelo treinado (*.model*) e a matriz de confusão (*.csv*), com as métricas de avaliação. Esta configuração demonstrou-se eficaz e prática, mantendo o fluxo de trabalho acessível e coerente com os princípios de reprodutibilidade e automatização que orientaram a metodologia deste estudo.

Após o treino, os melhores modelos foram aplicados sobre a imagem multibanda utilizando o módulo *ImageClassifier*.

3.2.6. Validação dos Modelos

A avaliação do desempenho dos modelos de classificação foi realizada com base em métricas consolidadas na literatura, cuja fundamentação teórica já foi detalhada no Capítulo 2. No contexto do presente estudo essas métricas foram aplicadas para aferir a qualidade das classificações supervisionadas, considerando diferentes combinações de variáveis, algoritmos e fontes de dados.

Após o treino dos modelos o modulo *TrainImagesClassifier* do OTB gerou automaticamente um ficheiro *.csv* contendo a matriz de confusão, calculada a partir da *shapefile* de validação externa fornecida. A partir dessa matriz foram obtidos os seguintes indicadores, por classe e de forma global: precisão, revocação, *F1-score* e coeficiente *Kappa*.

A combinação destas métricas foi adotada para garantir uma avaliação abrangente e robusta pois a matriz de confusão permite identificar diretamente erros e acertos de classificação entre pares de classes; o *F1-score* e a revocação são essenciais para detetar falhas em classes minoritárias ou com maior sobreposição espectral; a precisão indica a confiança do modelo nas previsões efetuadas; o coeficiente *Kappa* adiciona robustez estatística à análise.

Além das métricas principais o OTB apresenta o número de amostras (pixels) por classe (tanto no treino como na validação), o que possibilita verificar o equilíbrio amostral e garantir a fiabilidade estatística da avaliação.

A análise comparativa focou-se nos seguintes eixos:

- Conjuntos de variáveis: imagens base (RGB e IRG), índices de vegetação, texturas de *Haralick* e composições multibanda combinadas.
- Fontes de dados: UAV vs. constelação de satélites Pléiades-Neo.
- Algoritmos de classificação: RF e SVM.

Foram ainda testadas diferentes áreas da R-FB, externas ao treino, para avaliar a robustez espacial dos modelos e a sua capacidade de generalização. Esta abordagem permitiu identificar não apenas o desempenho numérico mais elevado, mas também a consistência entre diferentes

zonas da área de estudo, assegurando uma escolha fundamentada do modelo final a ser aplicado à totalidade da R-FB.

3.3. Fluxo de Trabalho

Todos os procedimentos de carregamento dos dados de base, de preparação dos mesmos para a análise, a análise em si mesma, a obtenção dos resultados, a sua interpretação e a produção de figuras e tabelas finais foram levados a cabo no QGIS, versão 3.34.13 ‘Prizren’ — uma versão LTR (*Long Term Release*), estável e amplamente utilizada. Foram instalados no QGIS módulos do OTB⁴ para as operações de classificação e análise espectral. Todos os procedimentos seguiram fluxos estruturados e iterativos, que combinaram o tratamento de dados *raster* com a modelação supervisionada, culminando numa abordagem de ML para a classificação final.

3.3.1. Preparação inicial dos dados

O processo começou com o carregamento dos dados vetoriais de altimetria e hidrografia. Para validar a coerência das curvas de nível e das linhas de água foi utilizada a Carta Militar de Portugal, folhas 151 e 161 (2.^a edição), como fonte de referência topográfica auxiliar.

A partir das curvas de nível validadas foi gerado um Modelo Digital Topográfico (MDT) da área de estudo, utilizando o método TIN (*Triangulated Irregular Network*). Com base no MDT foi criada uma carta de declives, com o objetivo de detalhar a caracterização do relevo e fornecer o enquadramento físico da zona analisada.

3.3.2. Integração e Processamento de Imagens

Na etapa seguinte foi carregado no QGIS o ortofotomosaico obtido com drone, o qual passou por um processo de *resampling* para atingir uma resolução espacial de 30 cm, garantindo compatibilidade com as imagens OrtoSat2023.

Foram então importados os quatro blocos do OrtoSat2023 correspondentes à área de estudo. As imagens foram unidas com a ferramenta *Merge* e recortadas com base na quadrícula UTM da Reserva da Faia Brava, respeitando os seus limites oficiais.

⁴ <https://www.orfeo-toolbox.org/CookBook-9.0/QGISInterface.html#plugin-installation>

Após o recorte foi realizada a normalização radiométrica das bandas das imagens OrtoSat2023 para o intervalo [0–255], utilizando ferramentas de conversão raster. As imagens foram georreferenciadas e alinhadas espacialmente ao OrtoFotoUAV — adotado como referência de maior precisão absoluta — utilizando o módulo *Georeferencer* do QGIS. Esse alinhamento foi essencial para permitir a comparação e eventual fusão de dados provenientes de sensores distintos.

3.3.3. Geração de Variáveis Derivadas

Foram calculados os principais índices de vegetação (GLI, NDVI, SAVI e MSAVI) com a ferramenta *Raster Calculator*. Em paralelo foram extraídas texturas de *Haralick* por meio do módulo *HaralickTextureExtraction* do OTB, utilizando duas bandas: a vermelha (*Red*) da imagem RGB e a infravermelha (*NIR*) da imagem de falsa cor. Para cada banda foram geradas oito camadas texturais distintas.

3.3.4. Treino dos Modelos de Classificação

Para cada composição multibanda foi extraída a Q21, para serem calculadas estatísticas descritivas usando o módulo *ComputeImagesStatistics*, etapa necessária para permitir a normalização automática dos dados durante o treino supervisionado.

O treino dos modelos foi realizado com o módulo *TrainImagesClassifier*, utilizando como entrada as composições multibanda concatenadas da Q21 e as *shapefiles* de treino e de validação previamente definidas. Foram testados dois algoritmos de aprendizagem supervisionada: RF e SVM.

Como saída, foram gerados automaticamente:

- o arquivo de modelo treinado (*.model*)
- a matriz de confusão (*.csv*), contendo as principais métricas de desempenho: precisão, revocação, *F1-score* e coeficiente *Kappa*

Com base na comparação dos resultados selecionaram-se os melhores modelos para cada combinação de variáveis e algoritmos.

3.3.5. Aplicação e Classificação Final

Após a comparação dos desempenhos obtidos com os diferentes conjuntos de variáveis e algoritmos selecionaram-se os melhores modelos resultantes de cada teste individual. Cada um desses modelos foi, então, aplicado a uma área distinta e não utilizada no treino, mas ainda localizada dentro dos limites da Reserva da Faia Brava, com o objetivo de avaliar através da generalização a robustez do desempenho para zonas não estudadas previamente. Com base nesses resultados identificou-se o modelo com melhor desempenho global - em termos de precisão, coerência espacial e estabilidade entre áreas - o qual foi, por fim, utilizado para realizar a classificação supervisionada da totalidade da área da Faia Brava. Esta etapa final foi executada com o módulo *ImageClassifier* do OTB, resultando na produção do mapa temático definitivo de uso e ocupação do solo.

4. RESULTADOS E DISCUSSÃO

Apresentam-se os resultados, quantitativos e qualitativos, da classificação supervisionada recorrendo aos algoritmos RF e SVM, considerando diferentes composições espectrais e texturais. A avaliação dos resultados foi levada a cabo com base na precisão, revocação, *F1-score* e coeficiente *Kappa*, permitindo comparar o desempenho entre classes, composições e fontes de dados. Apenas se apresentam resultados com textura de *Haralick* extraída em raio 7, por terem demonstrado desempenho superior em todos os cenários testados, comparativamente a outros raios menores. As composições espectrais incluíram o GLI, para as imagens RGB, e os NDVI, SAVI e MSAVI para a imagem IRG, selecionados devido à sua relevância na deteção de vegetação. As camadas foram sendo adicionadas de forma cumulativa, iniciando com a imagem original (RGB ou IRG) e progredindo com a inserção sucessiva de índices espectrais e bandas específicas, até se alcançar o composto mais completo, com oito métricas texturais derivadas da GLCM. Esse processo permitiu avaliar, de forma controlada, o impacto incremental de cada grupo de variáveis na acurácia da classificação.

4.1. Avaliação Quantitativa

Os resultados quantitativos estão organizados em três grupos, de acordo com a origem e a composição das imagens utilizadas: (1) OrtoFotoUAV (RGB), (2) OrtoSat2023 em cor verdadeira (RGB) e (3) OrtoSat2023 em falsa cor (IRG). Em cada grupo, os desempenhos por classe foram analisados e avaliados considerando a adição progressiva de variáveis, desde a imagem original até às composições enriquecidas com índices espectrais, bandas isoladas de textura e a composição completa das texturas *Haralick*.

Os resultados são apresentados por meio de gráficos de linhas, as quais ilustram as métricas de precisão, revocação, *F1-score* por classe e coeficiente *Kappa* para cada configuração testada. Embora a precisão e revocação estejam representadas graficamente, a análise principal dos resultados concentrou-se no *F1-score* e no coeficiente *Kappa*, por oferecerem uma visão mais sintética e equilibrada do desempenho dos modelos. O *F1-score*, ao integrar precisão e revocação, foi considerado mais adequado para avaliar a eficácia real dos algoritmos em cada classe, enquanto o coeficiente *Kappa* permitiu uma comparação geral entre composições.

Adicionalmente, foi realizada uma análise comparativa dos *F1-score* e dos coeficientes *Kappa* obtidos em cada combinação de variáveis e para ambos os algoritmos (RF e SVM), permitindo avaliar com maior profundidade o impacto das diferentes composições de entrada no desempenho global e por classe.

Em anexo poderão ser consultadas tabelas que reúnem os valores numéricos correspondentes, permitindo a análise mais detalhada e facilitando a comparação entre as diferentes composições testadas.

As métricas de avaliação evidenciam o impacto das distintas combinações de variáveis para a acurácia dos algoritmos RF e SVM, destacando-se os efeitos das composições mais simples em contraste com as multivariadas, e permitindo observar como cada algoritmo responde à crescente complexidade espectral e estrutural dos dados.

4.1.1. OrtoFotoUAV

4.1.1.1. RF

Os resultados revelaram elevada robustez, sobretudo na classe 1 (árvores e arbustos), que manteve o *F1-score* acima de 0,98 em todas as composições, incluindo as mais simples. A classe 3 (afloramentos rochosos) também apresentou um desempenho consistente, com *F1-score* superior a 0,90 na maioria das combinações. A introdução de bandas texturais teve impacto positivo, em particular na classe 2 (vegetação herbácea), cuja separação beneficiou da informação espacial adicional, nomeadamente a Entropia (B2). Por outro lado, a inclusão do conjunto completo de texturas *Haralick* (HS) provocou uma queda acentuada na performance, em especial na classe 4 (outras ocupações e usos do solo), evidenciando indícios de sobreajuste. De forma geral, o algoritmo RF demonstrou desempenho estável e eficiente ao longo das diferentes composições testadas, adaptando-se bem tanto a conjuntos simples como a mais complexos de variáveis, o que confirma a sua versatilidade e capacidade de generalização (Fig. 4.1).

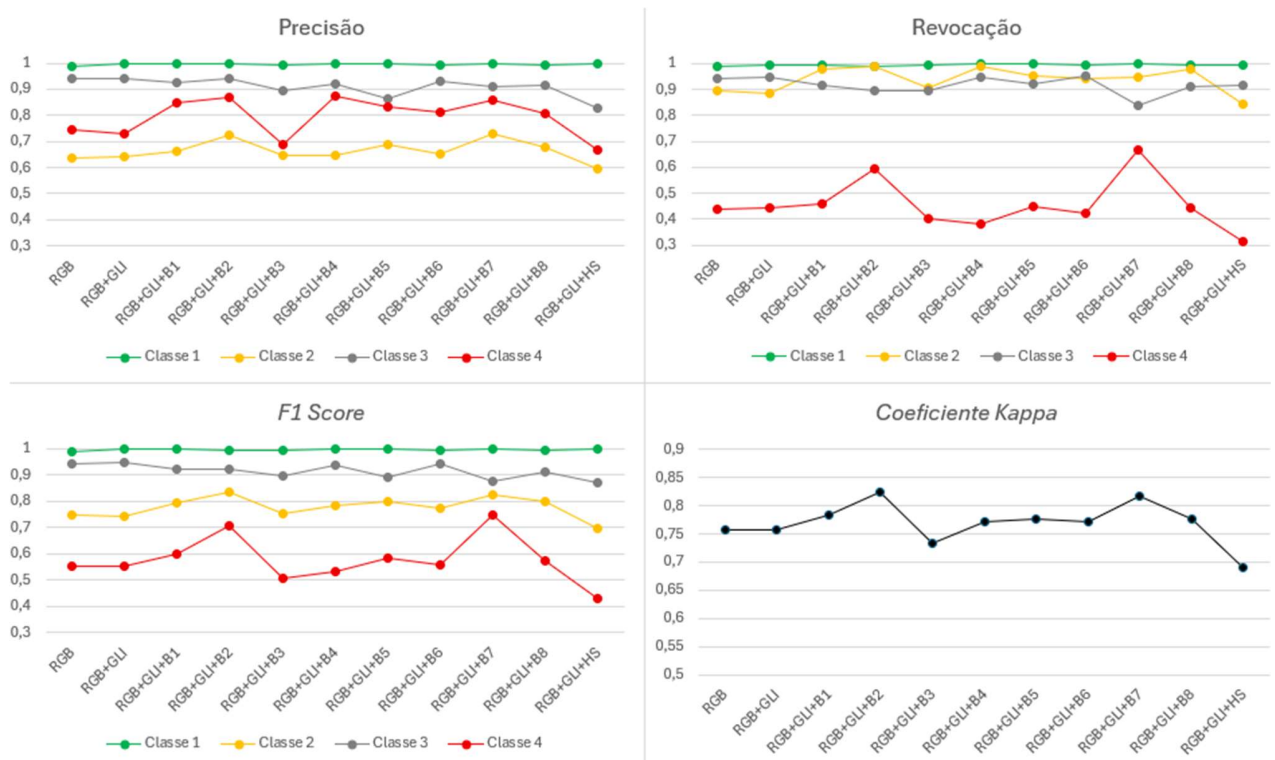


Figura 4.1 - Desempenho da classificação do OrtoFotoUAV com o algoritmo RF.

4.1.1.2. SVM

O algoritmo SVM apresentou desempenho elevado nas classes 1 (árvores e arbustos) e 3 (afloramentos rochosos), com valores de *F1-score* consistentemente superiores a 0,98 e 0,95, respetivamente, em várias composições. No entanto, mostrou maior sensibilidade às variações nas combinações de variáveis de entrada, com oscilações notáveis nos resultados das classes 2 (vegetação herbácea) e 4 (outras ocupações e usos do solo). A inclusão de algumas métricas texturais, como as bandas B2 e B5, contribuiu positivamente para a classe 2, atingindo *F1-scores* acima de 0,89 e *Kappa* global até 0,884. Por outro lado, a combinação completa das texturas *Haralick* (HS) resultou em decréscimos marcados, sobretudo na classe 4, cujo *F1-score* caiu para 0,464. Apesar dessas flutuações, o SVM mostrou-se responsivo a composições multivariadas, tirando partido da complexidade espectral e estrutural presente em algumas combinações, e superou o RF em determinados cenários, especialmente no desempenho global de configurações com índices e métricas texturais bem ajustadas (Fig 4.2).

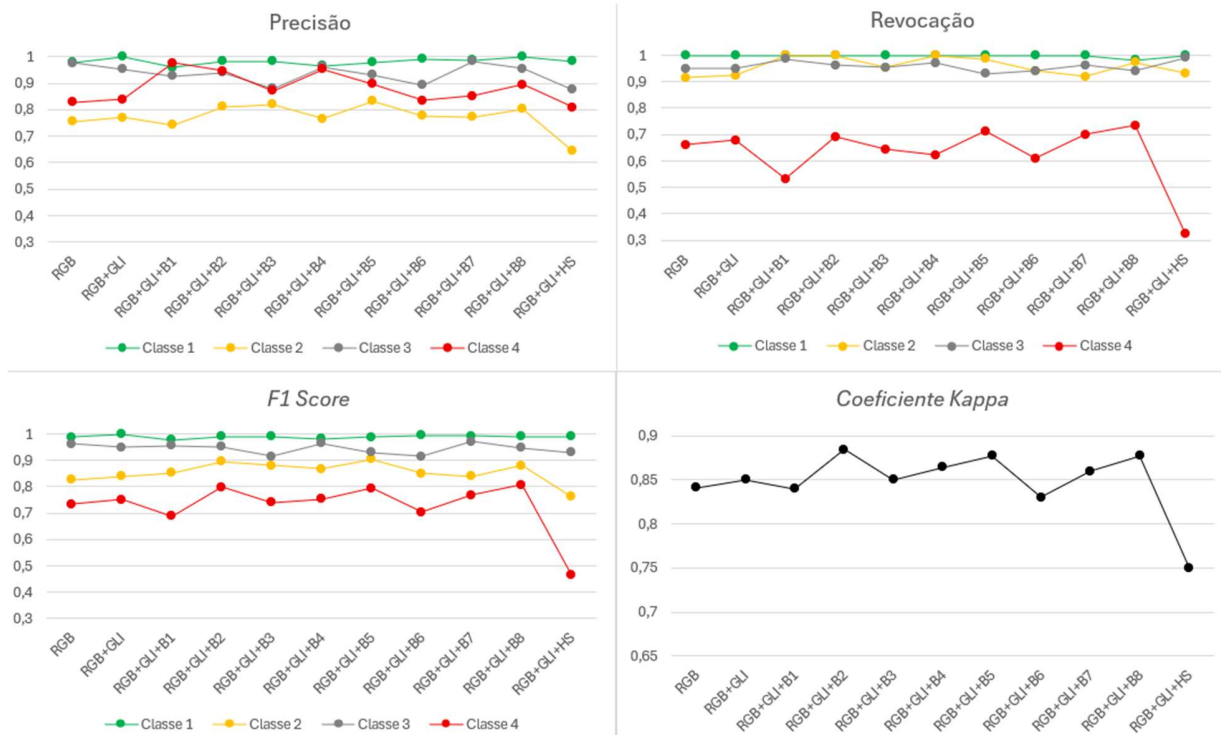


Figura 4.2 - Desempenho da classificação do OrtoFotoUAV com o algoritmo SVM

4.1.2. OrtoSat2023 - Cor Verdadeira

A aplicação dos algoritmos RF e SVM à imagem OrtoSat2023 (cor verdadeira) revelou padrões de desempenho distintos, especialmente nas classes mais sensíveis às variações espectrais.

4.1.2.1. RF

A classe 1 (árvores e arbustos) manteve resultados consistentes, com valores de *F1-score* elevados mesmo em composições simples. A classe 3 (afloramentos rochosos) também apresentou estabilidade, com destaque para a configuração RGB. Por outro lado, as classes 2 (vegetação herbácea) e 4 (outras ocupações e usos do solo) mostraram maior instabilidade, principalmente quando determinadas bandas (como B1 ou HS) foram utilizadas, o que reduziu significativamente a performance. A melhor composição global foi para o RGB, com o maior valor de *Kappa* (0,798), enquanto RGB+GLI+HS teve o pior desempenho (0,578) (Fig.4.3).

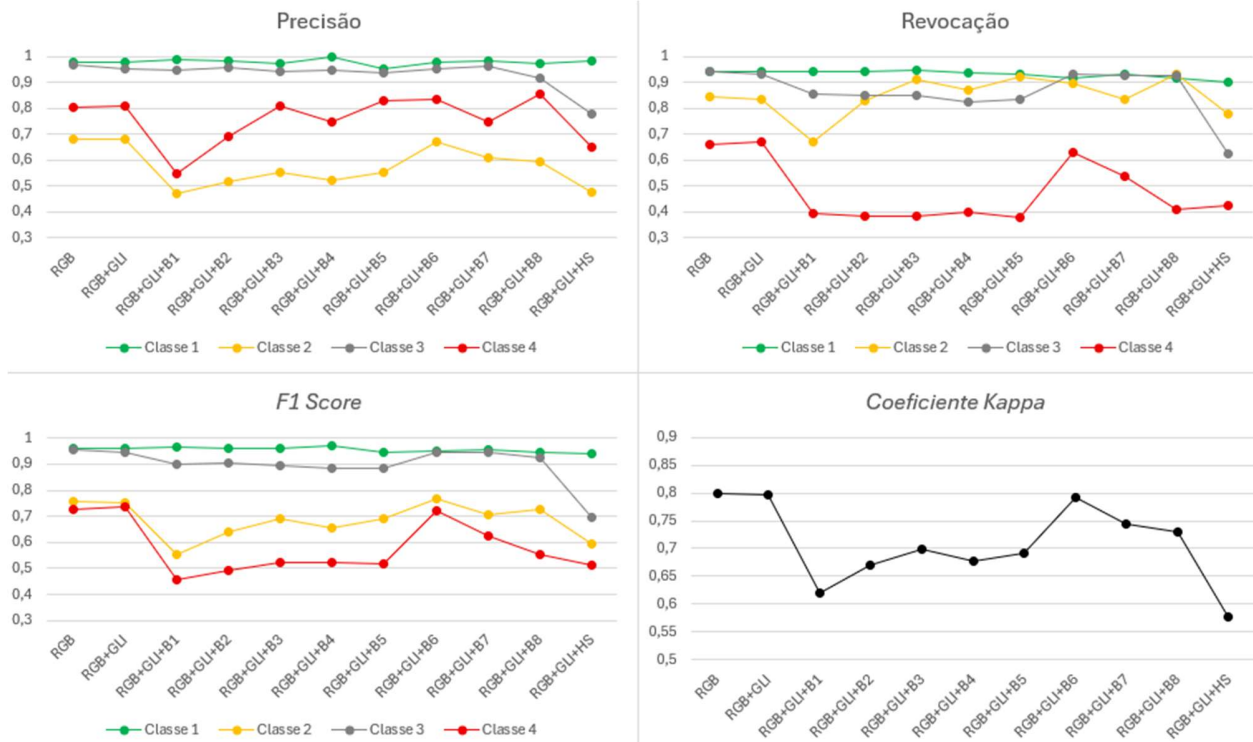


Figura 4.3 - Desempenho da classificação do OrtoSat2023 Cor Verdadeira com o algoritmo RF.

4.1.2.2. SVM

Observou-se maior sensibilidade à escolha das composições. Embora as classes 1 e 3 tenham mantido elevada performance, o desempenho das classes 2 e 4 oscilou bastante, melhorando apenas com composições enriquecidas, como a RGB+GLI+HS. A classe 3 respondeu bem ao aumento da complexidade espectral, obtendo bons resultados com RGB+GLI+HS. O coeficiente *Kappa* confirmou essa tendência, variando de 0,592 (RGB+GLI+B3) a 0,780 (RGB+GLI+HS) (Fig. 4.4).

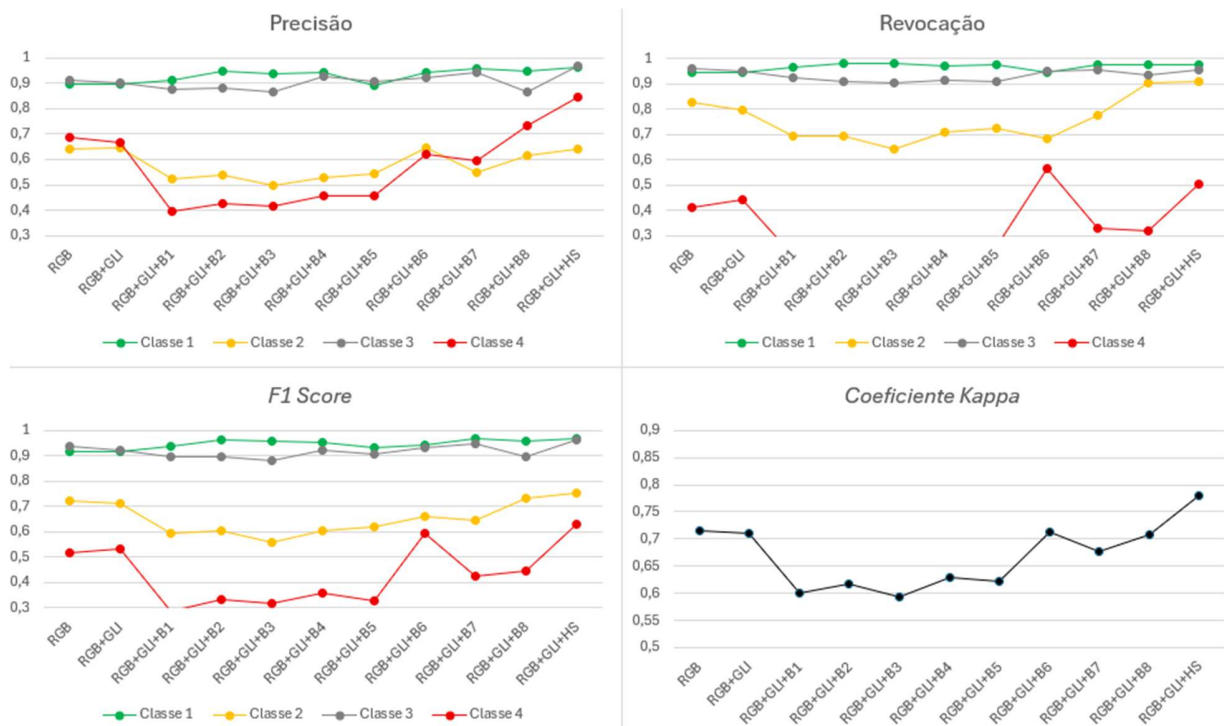


Figura 4.4 - Desempenho da classificação do OrtoSat2023 Cor Verdadeira com o algoritmo SVM.

4.1.3. OrtoSat2023 Falsa Cor

4.1.3.1. NDVI

A imagem de satélite OrtoSat2023 em falsa cor (composição IRG), combinada com o NDVI e texturas *Haralick*, apresentou um desempenho geral mais robusto em comparação com as composições RGB, do mesmo sensor.

4.1.3.1.1. RF

A composição IRG isolada forneceu bons resultados, com destaque para as classes 1 (árvores e arbustos) e 3 (afloramentos rochosos), que atingiram valores de *F1-score* de 0,983 e 0,965, respetivamente. A classe 4 (outras ocupações e usos do solo) também beneficiou da informação espectral em falsa cor, atingindo um *F1-score* máximo de 0,830 com a composição IRG+NDVI+B7 (valor superior ao registado nas composições RGB). Já a classe 2 (vegetação

herbácea) revelou maior instabilidade, com quebras expressivas em determinadas combinações, como IRG+NDVI+B4, onde o *F1-score* caiu para 0,581. O coeficiente *Kappa* seguiu tendência semelhante, com um pico de 0,828 (IRG+NDVI+B7) e um mínimo de 0,660 (IRG+NDVI+B4), evidenciando o impacto direto da seleção de variáveis sobre o desempenho do modelo (Fig. 4.5).

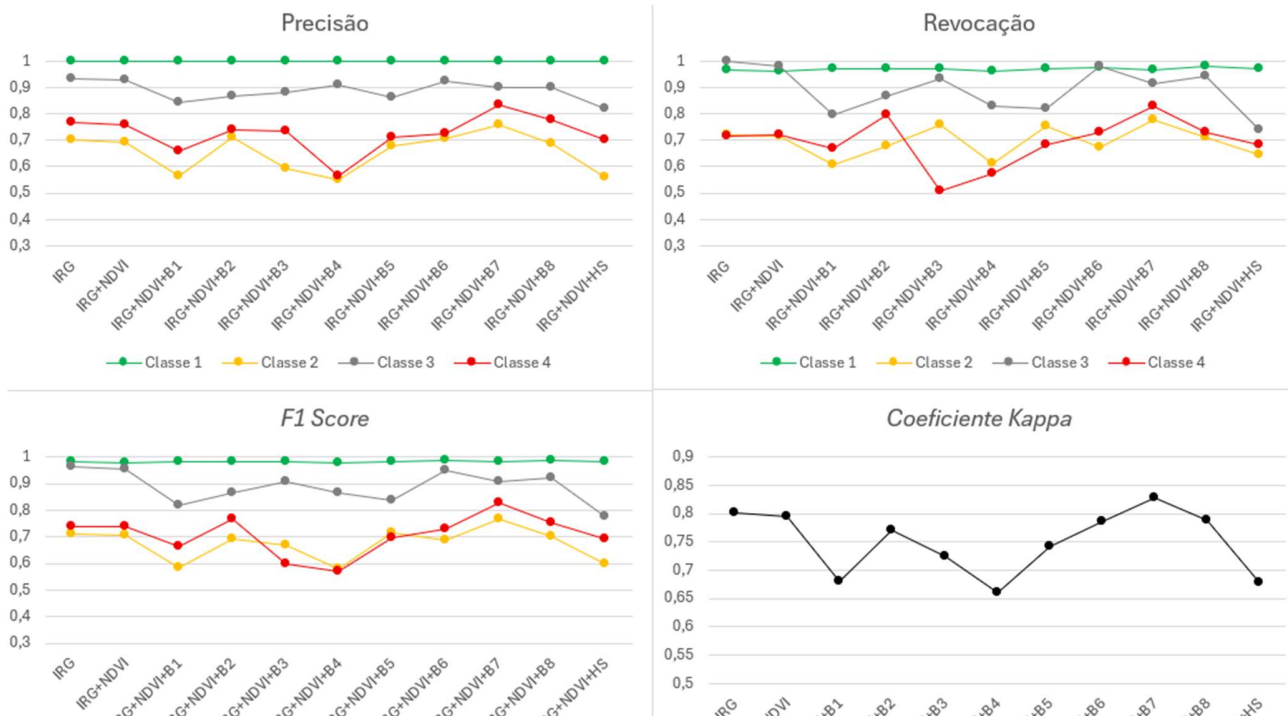


Figura 4.5 - Desempenho da classificação do OrtoSat2023 Falsa Cor com o algoritmo RF utilizando o NDVI.

4.1.3.1.2. SVM

A classe 1 (árvores e arbustos) manteve desempenho excelente em todas as combinações, com valores de *F1-score* entre 0,983 e 0,994, sendo particularmente favorecida por combinações como IRG+NDVI+B2, B6 e HS. A classe 3 (afloramentos rochosos) também apresentou resultados sólidos, com destaque para a composição IRG+NDVI+B7 e B8, que ultrapassou 0,960, enquanto o a composição com HS ficou ligeiramente abaixo do limiar, com 0,948. A classe 2 (vegetação herbácea) mostrou maior variabilidade: teve o pior desempenho com a banda B3 (*F1-score* de 0,609) e melhorou com a composição B8, atingindo 0,722. A classe 4 (outras ocupações e usos do solo) continuou a ser a mais difícil de classificar, embora tenha alcançado o melhor resultado com a composição IRG isolada (*F1-score* de 0,701). O coeficiente *Kappa*

variou entre 0,682 (IRG+NDVI+B3) e 0,771 (IRG+NDVI+HS), refletindo a influência direta das variáveis de entrada sobre o desempenho global do modelo (Fig. 4.6).

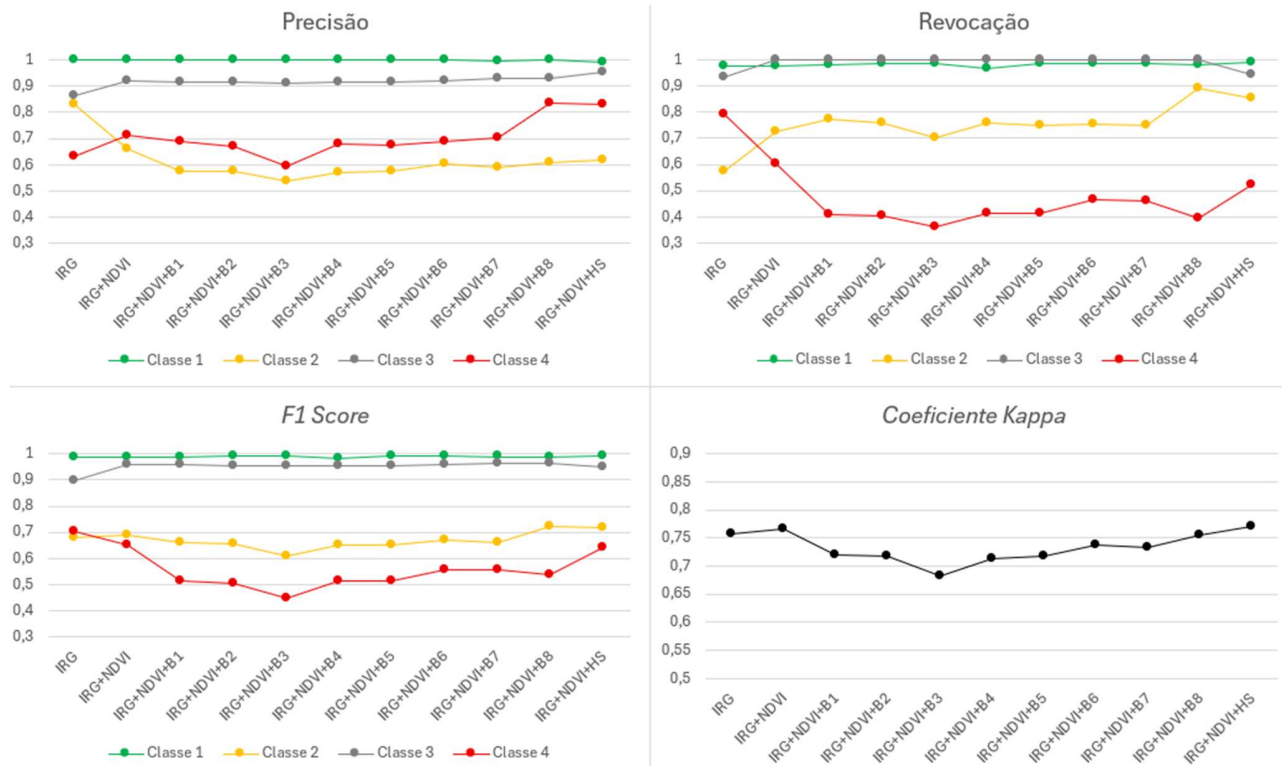


Figura 4.6 - Desempenho da classificação do OrtoSat2023 Falsa Cor com o algoritmo SVM utilizando o NDVI.

4.1.3.2. SAVI

A inclusão do SAVI nas composições espectrais da imagem OrtoSat2023 (falsa cor IRG) permitiu testar o seu potencial discriminativo em diferentes configurações.

4.1.3.2.1. RF

A inclusão do SAVI não trouxe ganhos expressivos em relação à composição IRG isolada. A classe 1 (árvores e arbustos) manteve desempenho excelente em todas as combinações, com *F1-scores* acima de 0,97, sendo ligeiramente favorecida pela composição IRG+SAVI+B6, que atingiu 0,987. Em contrapartida, a classe 2 (vegetação herbácea) revelou maior instabilidade, com queda acentuada em IRG+SAVI+B4 (*F1-score* de 0,561) e desempenho apenas substancial mesmo nas melhores composições, como B7 (*F1-score* de 0,745). A classe 3 (afloramentos

rochosos) teve seu melhor desempenho com o IRG isolado (*F1-score* de 0,964), não sendo claramente beneficiada pelo índice SAVI ou pelas bandas adicionais. Já a classe 4 (outras ocupações e usos do solo) apresentou o seu melhor resultado com IRG+SAVI+B7 (*F1-score* de 0,822), enquanto a banda B4 comprometeu fortemente a classificação (*F1-score* de 0,547). O coeficiente *Kappa* variou entre 0,644 (IRG+SAVI+B4) e 0,817 (IRG+SAVI+B7), acompanhando o padrão observado nas métricas por classe (Fig. 4.7).

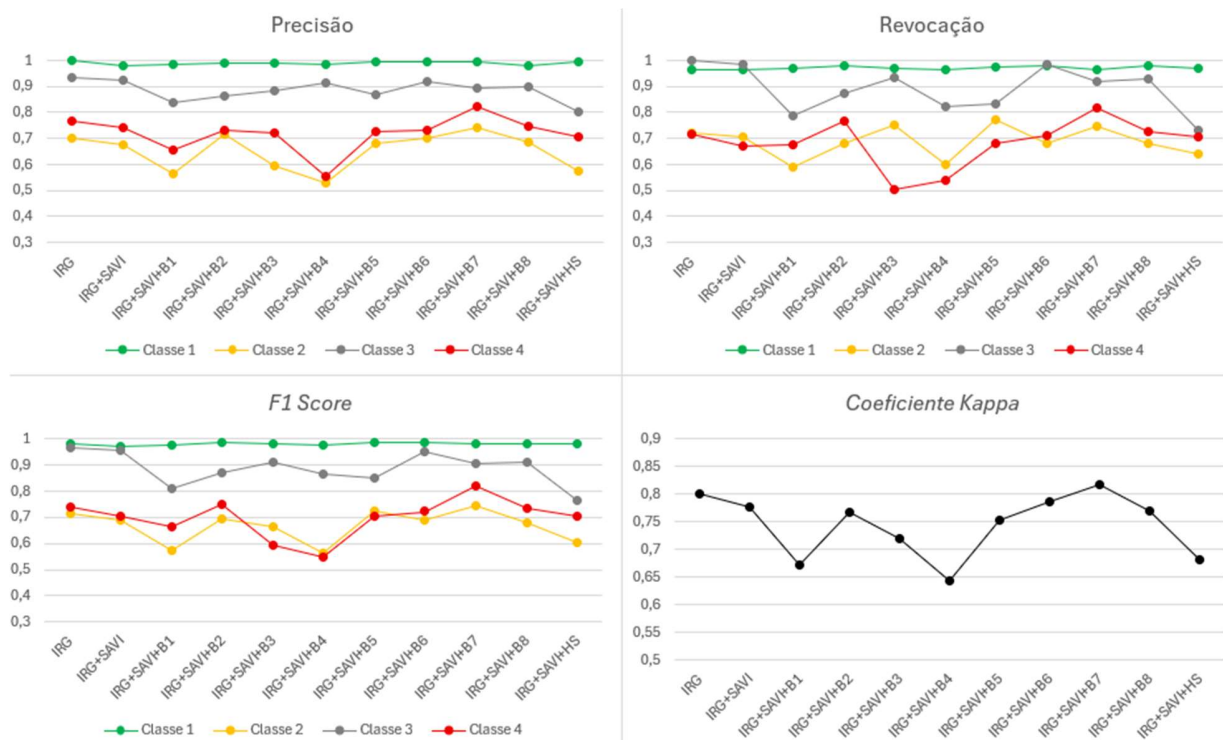


Figura 4.7 - Desempenho da classificação do OrtoSat2023 Falsa Cor com o algoritmo RF utilizando o SAVI.

4.1.3.2.2. SVM

O impacto da introdução do SAVI foi relativamente estável, especialmente quando associado a composições enriquecidas com métricas texturais. A classe 1 (árvores e arbustos) manteve desempenho excelente em todas as configurações, com *F1-scores* entre 0,985 e 0,994, incluindo composições com precisão próxima da perfeição. A classe 2 (vegetação herbácea) foi mais sensível, tendo o seu pior desempenho com IRG+SAVI+B3 (*F1-score* de 0,594), mas recuperando com a adição das texturas *Haralick* completas (HS), alcançando 0,720. A classe 3 (afloramentos rochosos) manteve-se em patamares elevados, com melhor desempenho na

composição IRG+SAVI+B8 (*F1-score* de 0,973), evidenciando o benefício da banda textural. Por outro lado, a classe 4 (outras ocupações e usos do solo) continuou a apresentar os resultados mais baixos, com *F1-scores* abaixo de 0,660 em quase todas as combinações testadas, sendo o melhor desempenho obtido com o IRG isolado (*F1-score* de 0,701), o que sugere que nem o SAVI nem as bandas adicionais foram eficazes para essa classe. O coeficiente *Kappa* acompanhou esse padrão, variando entre 0,674 (IRG+SAVI+B3) e 0,771 (IRG+SAVI+HS) (Fig. 4.8).

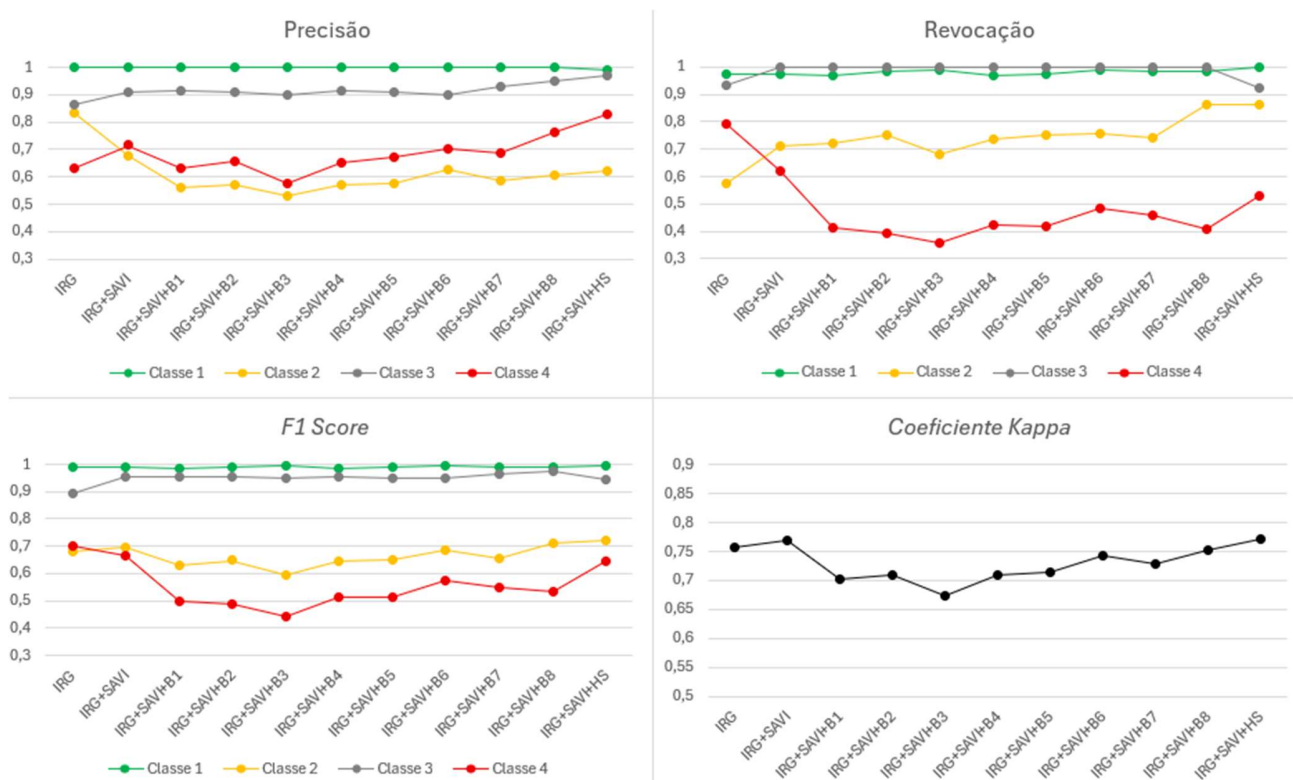


Figura 4.8 - Desempenho da classificação do OrtoSat2023 Falsa Cor com o algoritmo SVM utilizando o SAVI.

4.1.3.3. MSAVI

A combinação da imagem IRG com o MSAVI foi avaliada, visando testar o seu impacto na separação das quatro classes temáticas. Assim como nas abordagens anteriores, diferentes bandas texturais foram adicionadas para observar os efeitos combinatórios sobre o desempenho do modelo.

4.1.3.3.1. RF

O MSAVI mostrou-se eficaz em diversas configurações, destacando-se especialmente quando associado à banda B7, que contribuiu de forma significativa para a melhoria da classificação em classes mais ambíguas. A classe 1 (árvores e arbustos) manteve desempenho elevado e estável em todas as composições, com *F1-scores* entre 0,970 e 0,983, confirmando a sua fácil delimitação em imagens IRG. A classe 2 (vegetação herbácea) revelou maior sensibilidade, com um dos piores desempenhos na composição IRG+MSAVI+B1 (*F1-score* de 0,538) e uma recuperação expressiva com IRG+MSAVI+B7 (*F1-score* de 0,748), indicando que o MSAVI só é vantajoso quando complementado com variáveis discriminativas. A classe 3 (afloramentos rochosos) manteve estabilidade, embora sua melhor performance tenha ocorrido com IRG puro (*F1-score* de 0,965), sugerindo que o MSAVI não teve impacto positivo nessa classe. A classe 4 (outras ocupações e usos do solo) teve o seu melhor resultado com IRG+MSAVI+B7 (*F1-score* de 0,833), enquanto a composição com a banda B4 comprometeu significativamente o desempenho (*F1-score* de 0,568). O coeficiente *Kappa* variou entre 0,644 (IRG+MSAVI+B4) e 0,820 (IRG+MSAVI+B7), acompanhando os padrões de desempenho por classe (Fig. 4.9).

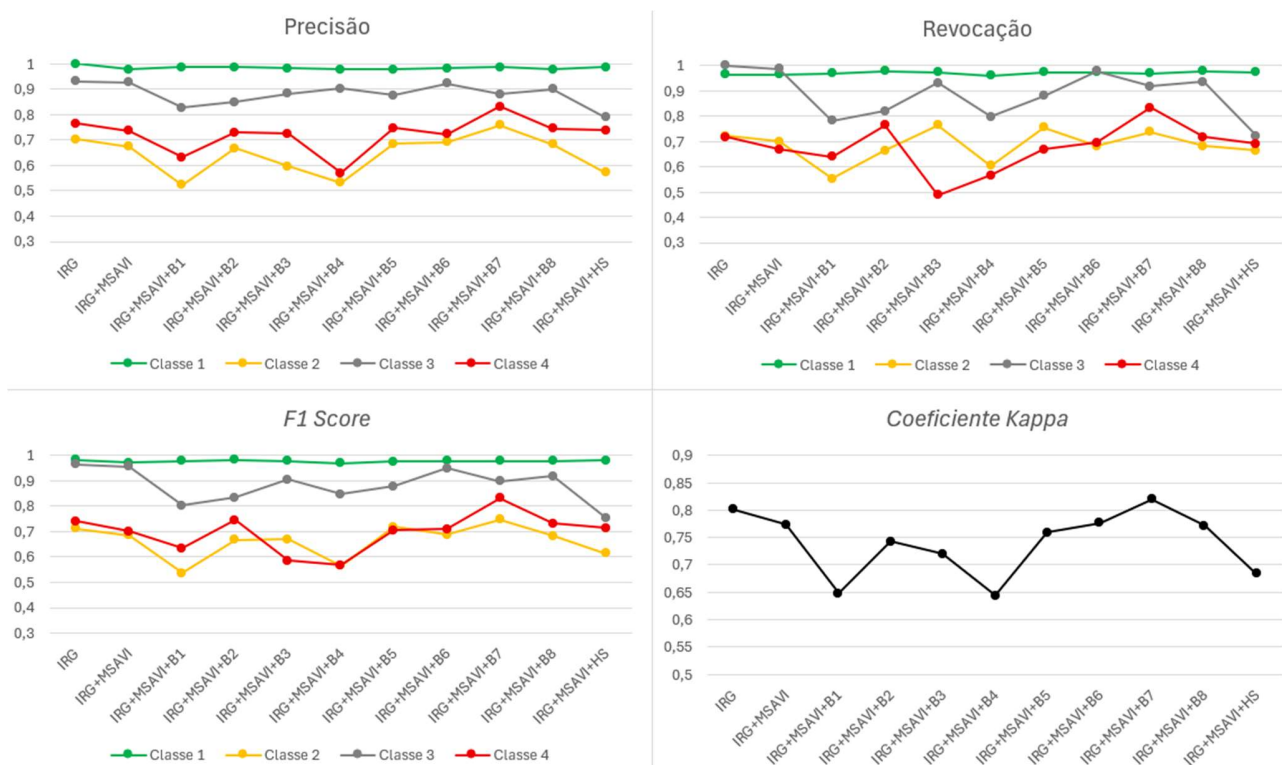


Figura 4.9 - Desempenho da classificação do OrtoSat2023 Falsa Cor com o algoritmo RF utilizando o MSAVI.

4.1.3.3.2. SVM

A introdução do MSAVI apresentou efeitos oscilantes conforme as composições utilizadas. A classe 1 (árvores e arbustos) manteve desempenho excelente em todas as configurações, com valores de *F1-score* entre 0,985 e 0,994, demonstrando separabilidade consistente, mesmo com diferentes combinações de variáveis. A classe 2 (vegetação herbácea) foi a mais sensível às variações: o pior desempenho foi observado com a composição IRG+MSAVI+B3 (*F1-score* de 0,584), enquanto a inclusão das texturas totais de *Haralick* (HS) proporcionou uma melhoria significativa (*F1-score* de 0,719). A classe 3 (afloramentos rochosos) manteve desempenho elevado e estável, com destaque para as composições IRG+MSAVI+B7 e B8, ambas superando 0,960. A classe 4 (outras ocupações e usos do solo) permaneceu com os menores valores de F1, com destaque negativo para IRG+MSAVI+B3 (*F1-score* de 0,429) e um desempenho máximo com IRG+MSAVI isolado (*F1-score* de 0,725). Embora o índice MSAVI tenha beneficiado algumas combinações, especialmente com texturas, ele não foi suficiente para superar as limitações associadas à classe 4 (Fig. 4.10).

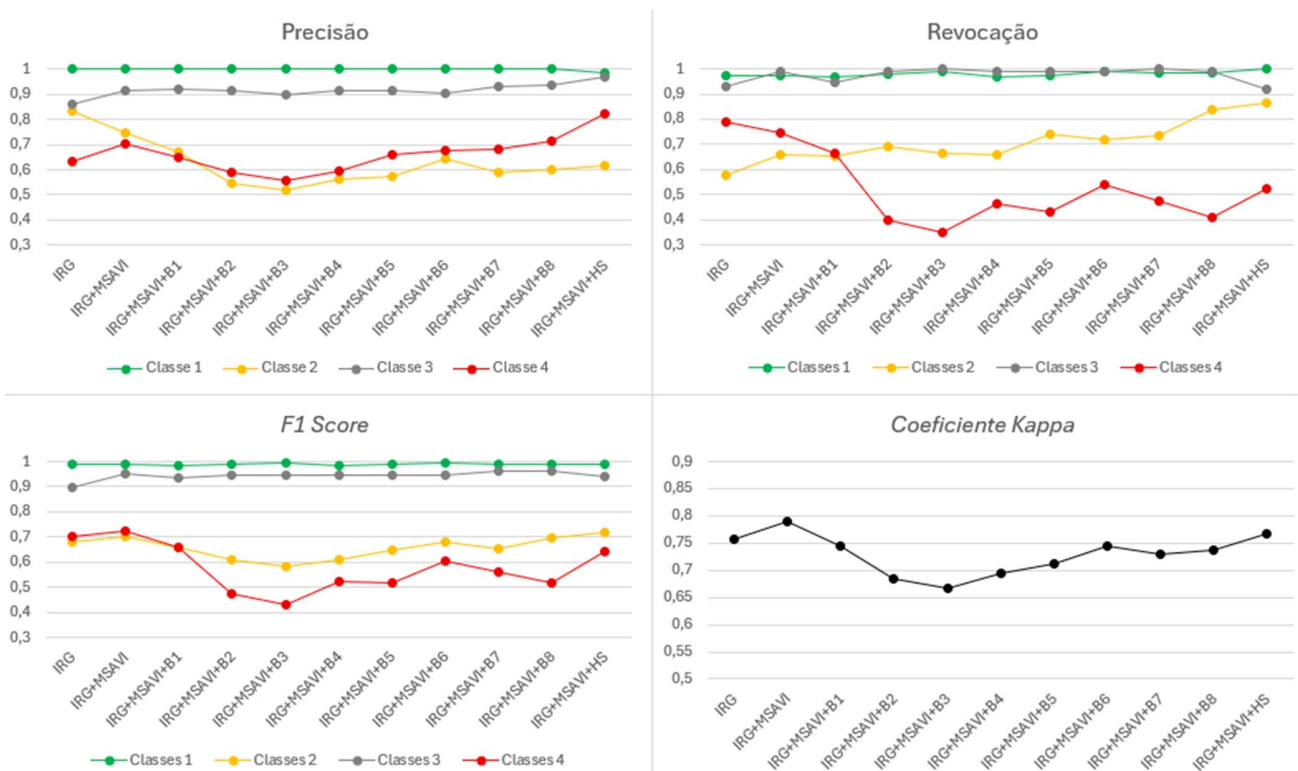


Figura 4.10 - Desempenho da classificação do OrtoSat2023 Falsa Cor com o algoritmo SVM utilizando o MSAVI.

4.1.4. Análise do *F1-score* entre Modelos

A análise dos valores de *F1-score* entre os algoritmos RF e SVM revelou padrões distintos para as quatro classes temáticas, refletindo a complexidade espectral e estrutural de cada categoria, além da influência das diferentes composições de dados.

A classe 1 – árvores e arbustos destacou-se como a categoria com melhor desempenho global, apresentando valores de *F1-score* frequentemente superiores a 0,960 para ambos os algoritmos. Esta classe demonstra elevada separabilidade espectral, o que lhe garante desempenho elevado mesmo em composições mais simples, como RGB. A inclusão de índices de vegetação (NDVI, SAVI, MSAVI) e de bandas do *NIR* trouxe ganhos marginais, sobretudo no SVM. O SVM atingiu os valores máximos de *F1-score*, muitas vezes próximos de 1, mas com maior variabilidade entre composições. Já o RF revelou-se mais estável, com desempenho consistentemente alto em praticamente todas as configurações, independentemente da complexidade das variáveis. Estes resultados confirmam a clareza espectral da classe 1 e a sua fácil distinção face às restantes coberturas (Fig. 4.11).

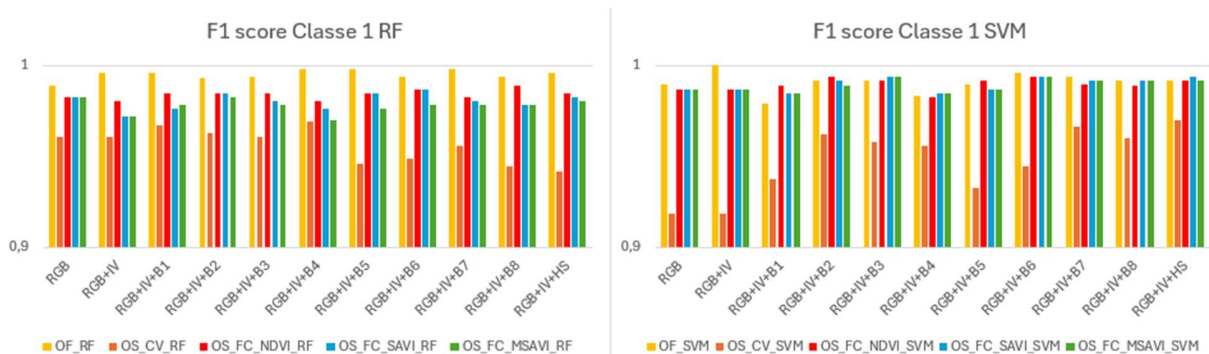


Figura 4.11 - *F1 score* da Classe 1 (árvores e arbustos) com RF (esquerda) e SVM (direita)

A classe 2 – vegetação herbácea revelou desempenho substancialmente superior nas classificações com o OrtoFotoUAV, em comparação com o OrtoSat2023, tanto em cor verdadeira (RGB) quanto em falsa cor (IRG). No modelo SVM, os valores de *F1-score* nas composições com GLI atingiram valores próximos ou superiores a 0,900 (GLI+B5 igual a 0,904), enquanto as melhores composições das imagens de satélite, mesmo com índices de vegetação e bandas adicionais, raramente ultrapassaram 0,730 (NDVI+B7 igual a 0,734). No

RF, o padrão manteve-se, com o UAV (GLI+B7 igual a 0,825) superando tanto o RGB (0,746) quanto as melhores composições com MSAVI ou SAVI nas imagens IRG (até 0,748 com MSAVI+B7). As composições RGB e IRG mostraram-se insuficientes para captar a complexidade estrutural e espectral da vegetação herbácea, que beneficiou visivelmente da maior resolução espacial e detalhe estrutural do OrtoFotoUAV. Essa diferença foi particularmente acentuada no SVM, cujo desempenho com imagens RGB caiu para 0,827, e com IRG ficou em torno de 0,680, evidenciando sensibilidade à limitação espectral das imagens de satélite e à necessidade de variáveis mais descritivas para modelar essa classe. Em síntese, os resultados demonstram que a classe 2 depende fortemente de dados com elevada resolução e variabilidade espacial para alcançar uma separação confiável, sendo as imagens provenientes do drone mais eficazes nesse contexto do que as composições multiespectrais orbitais (Fig 4.12).

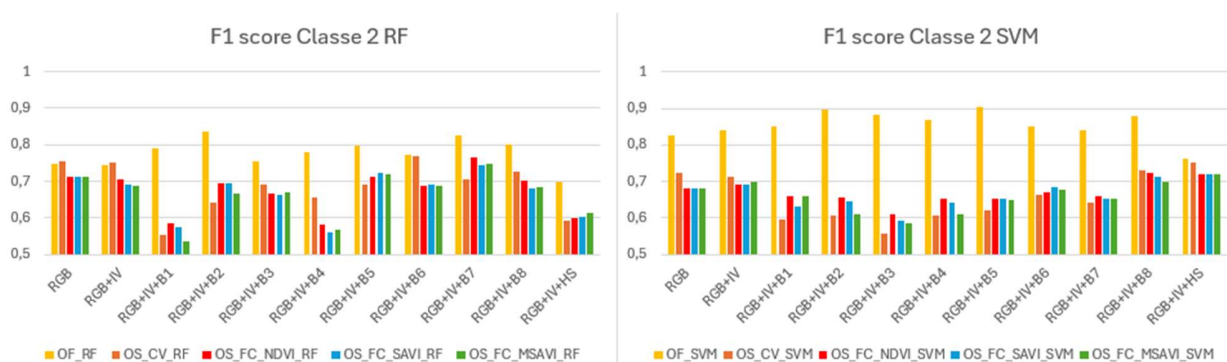


Figura 4.12 - F1 score da Classe 2 (vegetação herbácea) com RF (esquerda) e SVM (direita).

A classe 3 – afloramentos rochosos apresentou desempenho elevado e estável, sobretudo no modelo SVM, com valores de *F1-score* superiores a 0,950 em praticamente todas as composições, tanto nas imagens UAV como nas imagens de satélite, em cor verdadeira ou falsa. Este padrão de precisão evidencia a elevada separabilidade espectral e estrutural da classe, que favorece a generalização dos modelos mesmo em composições mais simples, como o RGB. O RF também obteve resultados competitivos, especialmente em composições bem ajustadas, como RGB+GLI e RGB+IV+B6, mas com maior variabilidade em outras combinações. Apesar disso, o RF mostrou-se particularmente eficaz com imagens UAV, atingindo até 0,944 em composições simples. Já o SVM demonstrou superioridade global, com desempenho mais

consistente entre composições e sensores, superando o RF em quase todas as situações, especialmente em contextos multivariados. A classe 3, portanto, representa um caso de cobertura com excelente discriminabilidade, onde ambos os algoritmos são eficazes, embora com vantagens distintas: o RF beneficia de combinações bem otimizadas, enquanto o SVM oferece maior robustez frente à diversidade de composições (Fig. 4.13).

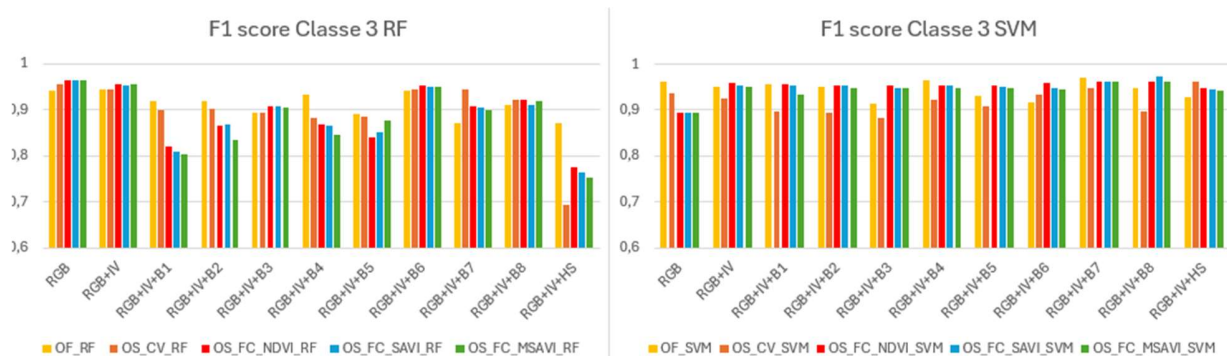


Figura 4.13 - F1 score da Classe 3 (afloramentos rochosos) com RF (esquerda) e SVM (direita).

Já a classe 4 – outras ocupações e usos do solo revelou-se a mais desafiante de todo o processo classificatório, com valores de *F1-score* que oscilaram entre aproximadamente 0,420 e 0,830 com o RF, e entre cerca de 0,290 e 0,800 com o SVM. Esta classe, composta por elementos muito heterogéneos (como sombra, construções, solo nu ou água) e com reduzida representatividade, demonstrou elevada sensibilidade às composições espectrais e texturais utilizadas.

O algoritmo RF apresentou um desempenho mais robusto e estável, destacando-se particularmente em composições enriquecidas com a bandas B7 e índices espectrais baseados no infravermelho próximo (*NIR*), como o NDVI. Em contraste, o SVM revelou maiores dificuldades, com quedas acentuadas do desempenho em composições multivariadas, refletindo limitações na modelação de classes estruturalmente complexas ou internamente ruidosas.

Estes resultados reforçam a importância de estratégias complementares, como a redefinição das classes temáticas, o aumento do número de amostras por subclasse e a aplicação de técnicas de balanceamento ou amostragem estratificada, para melhorar a performance em categorias compostas e sub-representadas em ambientes heterogéneos (Fig. 4.14).



Figura 4.14 - F1 score da Classe 4 (outras ocupações e usos do solo) com RF (esquerda) e SVM (direita).

4.1.5. Análise do Coeficiente *Kappa* entre Modelos

A análise do coeficiente *Kappa* evidenciou diferenças claras na consistência global dos modelos de classificação. O algoritmo RF revelou-se particularmente robusto e estável, com valores de *Kappa* a oscilar entre 0,580 e 0,830. Os melhores desempenhos foram registados em composições multivariadas que integraram bandas RGB do ortofotomosaico ou IRG com índices espectrais (NDVI, SAVI, MSAVI) e texturas derivadas da GLCM, destacando-se a Entropia (B2) e o *Cluster Prominence* (B7). Estas variáveis mostraram-se eficazes na captação da estrutura espacial da vegetação e dos elementos do terreno, melhorando substancialmente a coerência da classificação. Entre os índices, o NDVI destacou-se pela sua estabilidade e eficácia, superando o SAVI e o MSAVI, cujos efeitos foram mais limitados, provavelmente devido à reduzida presença de solo mineral exposto na área de estudo.

O algoritmo SVM, por seu lado, apresentou valores de *Kappa* entre 0,590 e 0,880, com o melhor desempenho observado nas composições simples provenientes do drone, onde a elevada resolução espacial e a menor complexidade dimensional facilitaram a separação entre classes. No entanto, a inclusão de múltiplas texturas e índices espectrais resultou frequentemente em quedas de desempenho, evidenciando as limitações do *kernel* linear na modelação de relações não lineares entre atributos (Pal, 2005). Apesar da aplicação de estratégias de balanceamento interno das classes, o SVM mostrou-se menos eficaz em contextos de elevada complexidade estrutural e espectral (Fig. 4.15).

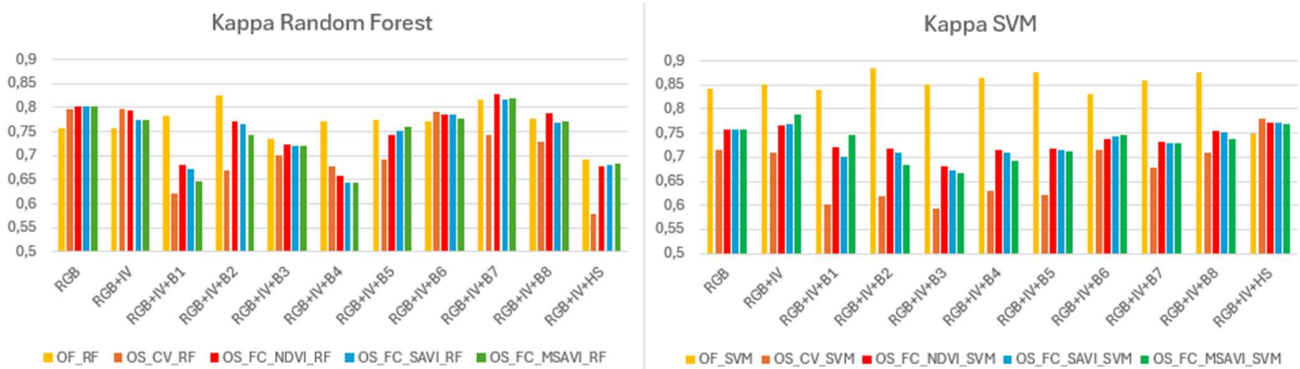


Figura 4.15 – Coeficiente Kappa com RF (esquerda) e SVM (direita).

4.1.6. Análise geral de valores de Kappa

A classe 4 destacou-se como o principal fator de redução dos valores de *Kappa*, devido à sua natureza heterogénea e representação reduzida. A sobreposição espectral com a classe 2, sobretudo em zonas de transição, contribuiu igualmente para o aumento da ambiguidade classificatória.

Outro aspecto relevante prende-se com a extração de texturas a partir de uma única banda por sensor (*Red* no OrtoFotoUAV e OrtoSat2023 cor verdadeira; *NIR* no OrtoSat2023 falsa cor). Esta abordagem, embora prática, pode ter limitado a capacidade dos modelos para captar padrões espaciais relevantes, sobretudo em classes não vegetadas, como afloramentos rochosos ou superfícies artificiais. Adicionalmente, diferenças de alinhamento espacial entre produtos e variações nos limites de amostragem podem ter introduzido ruído espectral em áreas de transição, afetando a pureza e coerência dos dados de treino.

Apesar destas limitações, os modelos apresentaram resultados metodologicamente válidos. O RF demonstrou uma maior tolerância à colinearidade e ao ruído nos dados enquanto o SVM, apesar de eficaz em cenários bem definidos, foi mais sensível à variabilidade interna das classes. As métricas utilizadas confirmaram a coerência global dos resultados (Foody, 2002; Grandini et al., 2020).

Conclui-se que a eficácia dos modelos supervisionados depende fortemente da qualidade espectral dos dados, da definição temática das classes, da seleção das variáveis e do desenho do processo amostral. Estes fatores devem ser criteriosamente considerados em estudos futuros, especialmente em ambientes com elevada heterogeneidade paisagística ou com categorias temáticas de natureza composta.

4.2. Avaliação Qualitativa

Com base na análise quantitativa realizada no ponto anterior, foram identificadas, para cada combinação de dados e algoritmo, as configurações com melhor desempenho segundo métricas consolidadas como a precisão, revocação, *F1-score* e coeficiente *Kappa*. Para garantir maior rigor e eficiência na modelação, optou-se por aplicar os algoritmos apenas às composições espectrais e texturais com desempenho ótimo em cada caso. Esta abordagem permitiu concentrar os esforços nos cenários mais promissores, reduzindo a complexidade computacional e maximizando a fiabilidade dos mapas produzidos.

A avaliação visual das classificações geradas revelou, contudo, que os melhores resultados estatísticos nem sempre se traduziram em classificações espacialmente mais coerentes. Em particular, o algoritmo RF, ainda que pontualmente superado pelo SVM em algumas métricas, apresentou classificações mais uniformes, com menor fragmentação e melhor preservação da estrutura das feições no espaço geográfico. Já o SVM, apesar da elevada sensibilidade espectral, mostrou tendência à criação de ruído e sobresegmentação, sobretudo em zonas de transição entre coberturas.

Para ilustrar essas diferenças de forma clara, foi utilizada a quadrícula Q41 como exemplo visual. Esta foi selecionada aleatoriamente dentro da Reserva da Faia Brava e representa um setor com elevada diversidade de coberturas, sendo, por isso, ideal para demonstrar o comportamento dos modelos. A disposição das figuras segue uma lógica constante em todos os conjuntos: à esquerda encontra-se o resultado da classificação com RF, à direita com SVM, e na base a imagem original de referência. Foram analisados três conjuntos de dados, correspondentes às seguintes composições espectrais:

4.2.1. OrtoFotoUAV

No primeiro conjunto (Fig. 4.16), referente aos ortofotomosaicos RGB obtidos com drone, ambos os algoritmos conseguiram mapear as principais classes de uso e ocupação do solo. O RF gerou mapas mais suaves e contínuos, com menor fragmentação entre polígonos adjacentes. Por contraste, o SVM apresentou classificações mais angulares e segmentadas, sendo visíveis alguns artefactos em zonas com mistura de vegetação e solo exposto, bem como o não reconhecimento de sombras. Neste cenário, a composição com inclusão de texturas (GLI e B2) revelou-se particularmente eficaz para ambas as abordagens, aumentando o contraste entre classes e melhorando a separação em áreas mais homogéneas.

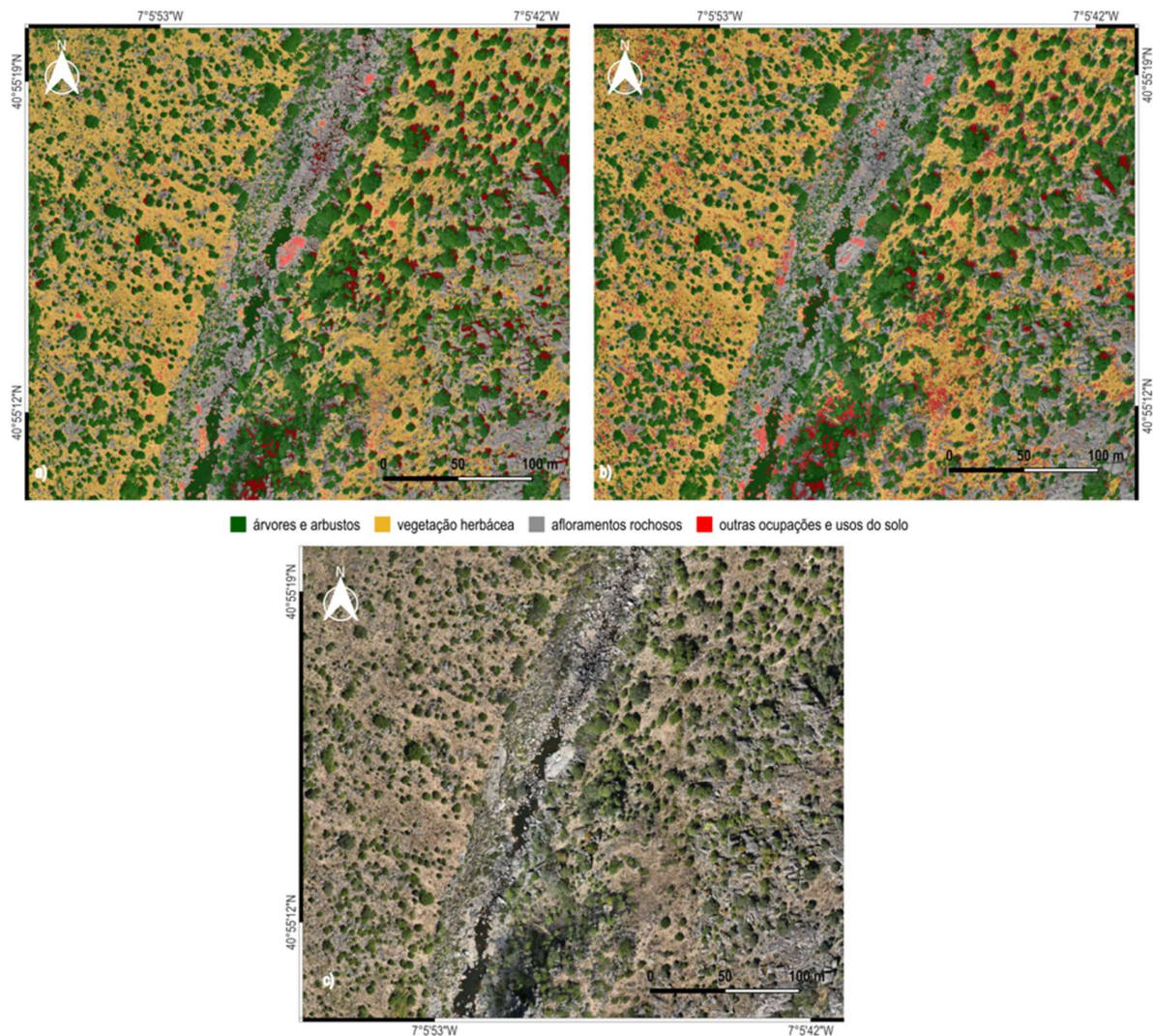


Figura 4.16 - Resultados da classificação da Q41 com o RF (a), SVM (b) e o OrtoFotoUAV (c).

4.2.2. OrtoSat2023 (Cor Verdadeira)

O segundo conjunto (Fig. 4.17) corresponde às imagens Pléiades-Neo em cor verdadeira (RGB), de resolução espacial inferior ao ortofotomosaico original. Essa limitação refletiu-se numa redução da precisão espacial das classificações, especialmente nas margens dos elementos geográficos. Ainda assim, o RF manteve um desempenho estável, apresentando mapas mais homogêneos. O SVM, por sua vez, mostrou maior sensibilidade a pequenas variações de reflectância, o que resultou em zonas com segmentação excessiva. Composições enriquecidas com índices como o GLI e o conjunto das componentes texturais demonstraram melhorar a robustez da classificação, atenuando parte do ruído gerado pelo SVM com outras bandas.

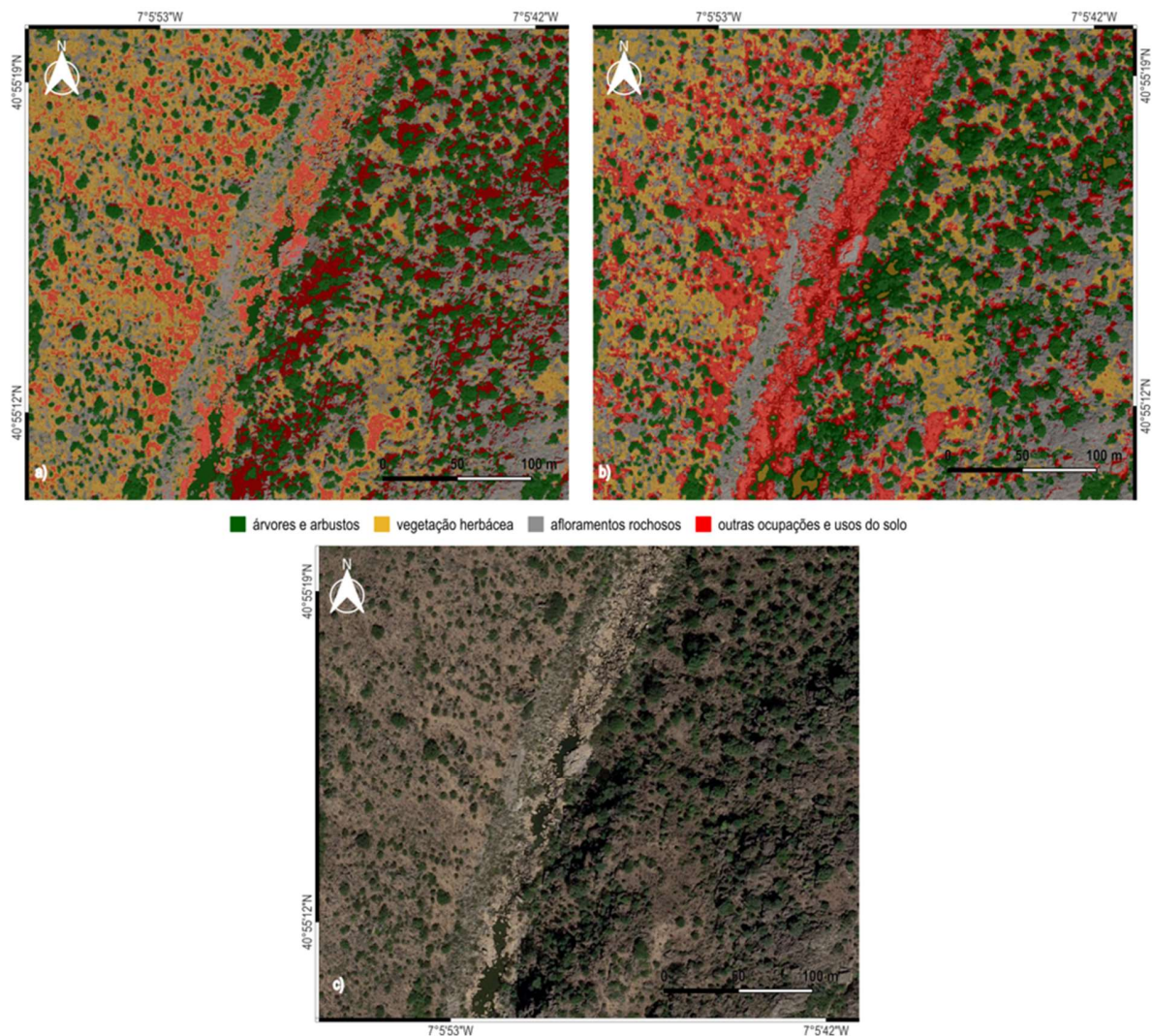


Figura 4.17- Resultados da classificação na Q41 com o RF (a), SVM (b) e OrtoSat2023 Cor Verdadeira (c).

4.2.3. OrtoSat2023 (Falsa Cor)

No terceiro conjunto (Fig.4.18), correspondente à imagem de falsa cor (*NIR-Red-Green*), era expectável uma melhoria na distinção entre coberturas vegetais, dada a presença da banda dos infravermelhos próximos. De facto, os resultados obtidos indicam um desempenho visual superior relativamente à imagem em cor verdadeira do satélite. Tanto o RF como o SVM apresentaram classificações com menor fragmentação e maior coerência espacial. A classificação com o SVM, revelou-se ruidosa, com contornos mais definidos, mas com tendência a sobresegmentar áreas de transição. Ainda que algumas composições com variáveis texturais tenham contribuído para melhorar a separação espectral, o resultado final demonstrou que a falsa cor, nesta configuração, não superou a imagem do OrtoFotoUAV em termos de legibilidade e consistência visual das classes.

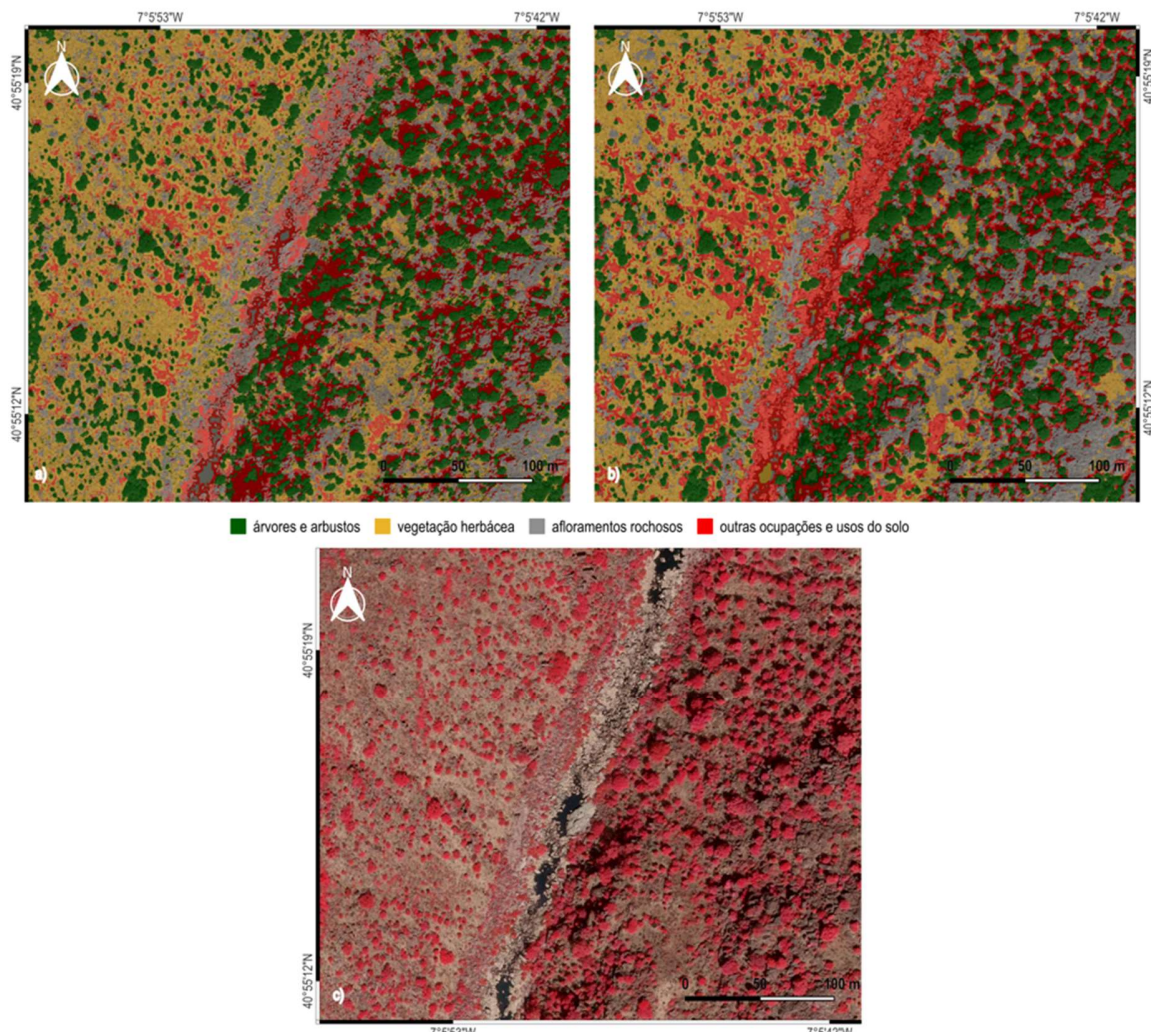


Figura 4.18- Resultados da classificação na Q41 com o RF (a), SVM (b) e OrtoSat2023 Falsa Cor (c).

Em síntese, os mapas de classificação demonstram que a escolha do algoritmo e da composição de dados influencia significativamente o resultado final. O RF revelou-se mais consistente na preservação da coerência espacial, enquanto o SVM mostrou bom desempenho em ambientes espectralmente complexos, embora com maior suscetibilidade a ruído. A composição e combinação adequada de bandas, índices e texturas foram decisivas para aumentar a exatidão e a utilidade prática dos produtos gerados, evidenciando a importância de integrar múltiplas fontes de informação na classificação da ocupação e uso do solo.

Com base nos resultados obtidos, foi então gerado o mapa final de classificação da R-FB. Este mapa foi produzido com o modelo que apresentou melhor desempenho global, utilizando o algoritmo RF aplicado ao ortofotomosaico RGB, combinado com o índice GLI e a métrica de textura Entropia. Esta composição mostrou-se especialmente eficaz na distinção entre coberturas, mantendo uma boa coerência espacial e reduzindo a fragmentação. Para facilitar a leitura visual do resultado, o mapa da figura 4.19 apresenta várias zonas ampliadas, escolhidas aleatoriamente, que permitem observar com mais detalhe a qualidade da classificação. O mapa completo da R-FB, em formato A3, encontra-se disponível nos anexos, com o objetivo de proporcionar uma visão integral e em escala adequada da classificação final.

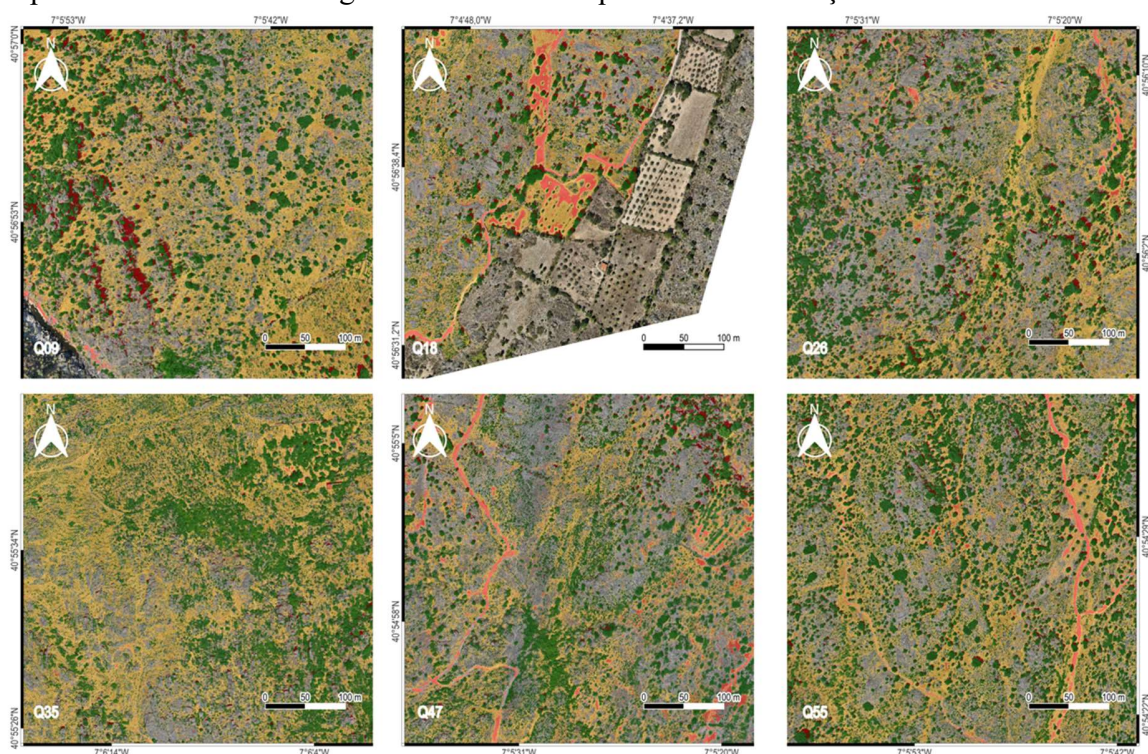


Figura 4.19 - Exemplos de zonas classificadas na R-FB com o modelo final (RF + RGB + GLI + Entropia)

4.3. Análise Crítica/Discussão

A avaliação crítica dos resultados obtidos evidencia a consistência e o mérito científico da abordagem metodológica adotada, que se destacou pela integração eficaz de dados multiescalares, ferramentas de código aberto e algoritmos de ML. A articulação entre imagens de alta resolução e variáveis espectrais e texturais permitiu alcançar classificações robustas e coerentes com a complexidade ecológica da Reserva da Faia Brava, refletindo a adaptabilidade do método a contextos ambientais heterogéneos.

A superior performance das classificações baseadas nos OrtoFotoUAV está de acordo com a sua elevada resolução espacial, que permite captar com grande detalhe elementos estruturais do território. Embora seja plausível que outros fatores, como o controlo da data de aquisição ou a homogeneidade radiométrica, também tenham contribuído para esses resultados, tais efeitos não foram isoladamente quantificados. Assim, embora se reconheça o potencial desses elementos, a sua influência direta deve ser interpretada com cautela e seria recomendável explorá-los em estudos dedicados.

A introdução de índices de vegetação, nomeadamente GLI, NDVI, SAVI e MSAVI, revelou-se particularmente benéfica na separação de coberturas com assinaturas espectrais semelhantes, enquanto as métricas texturais de *Haralick* aumentaram a capacidade de diferenciação em contextos com elevada variabilidade estrutural. No entanto, a decisão técnica de extrair texturas com base numa única banda por imagem (*Red* nas imagens RGB e *NIR* nas de falsa cor) poderá ter limitado o seu potencial discriminativo em classes como rochas, solo nu e estruturas antrópicas, cuja variabilidade se expressa noutras faixas do espectro visível. A utilização de múltiplas bandas para extração textural poderá representar uma melhoria relevante em aplicações futuras.

Do ponto de vista algorítmico, os modelos RF e SVM apresentaram comportamentos diferenciados. O RF demonstrou estabilidade notável em composições mais simples, mantendo desempenho elevado mesmo com conjuntos reduzidos de variáveis. O SVM, por sua vez, respondeu de forma mais sensível à complexidade dos dados, alcançando, por vezes, métricas superiores em composições multivariadas, mas com maior oscilação entre classes e

composições. Estes padrões sublinham a importância de adequar o algoritmo ao tipo e à dimensionalidade dos dados utilizados.

Importa ainda reconhecer limitações decorrentes do desenho experimental. A amostragem realizada sobre as três imagens analisadas assegurou coerência espacial e comparabilidade direta, mas poderá ter introduzido contaminação entre treino e validação, sobretudo em zonas de transição ou com pequenas discrepâncias de alinhamento espacial. Este fator poderá ter afetado a pureza de algumas amostras e, conseqüentemente, influenciado as métricas de avaliação.

A complexidade computacional das composições mais ricas, particularmente aquelas que combinam múltiplas bandas e texturas, revelou-se uma limitação operacional a considerar. Ainda que o OTB tenha demonstrado boa escalabilidade e integração com o QGIS, os tempos de processamento e a necessidade de capacidade computacional adequada podem constituir um entrave em projetos de maior escala ou com infraestrutura limitada.

Em síntese, a metodologia desenvolvida evidencia um equilíbrio notável entre robustez científica, flexibilidade técnica e viabilidade operacional. A conjugação de dados ortofotográficos de alta resolução, variáveis espectrais/texturais relevantes e algoritmos de ML acessíveis via *software* livre posiciona esta abordagem como uma solução promissora para aplicações em conservação, ordenamento do território e monitorização ambiental. Ainda que algumas limitações devam ser consideradas, os resultados obtidos validam a coerência, a eficácia e a aplicabilidade do modelo proposto.

4.4. Aplicabilidade Prática

A presente investigação evidencia de forma clara o valor prático de integrar algoritmos de ML em fluxos de trabalho acessíveis, visuais e cientificamente robustos, com particular relevância para contextos de gestão e conservação ambiental. A utilização de métodos supervisionados, nomeadamente RF e SVM, permitiu alcançar classificações de elevada precisão, reforçando o potencial destas técnicas no apoio à tomada fundamentada de decisão.

Um aspeto distintivo desta abordagem reside na forma como a análise foi integralmente operacionalizada em ambiente gráfico, através do QGIS e do *plugin* OTB, sem recurso a linhas

de comando. Esta característica torna a metodologia simultaneamente sofisticada do ponto de vista técnico e acessível a equipas multidisciplinares, mesmo na ausência de formação especializada em ciência de dados, o que representa uma mais-valia relevante em projetos locais com recursos humanos e financeiros limitados.

A comparação entre o OrtoFotoUAV e o OrtoSat2023 (Pléiades-Neo) revelou implicações operacionais significativas. Embora o OrtoSat2023 apresente resolução razoável (30 cm/pixel), a sua disponibilização pública em Portugal é esporádica e condicionada, o que limita a sua utilidade em contextos que exijam atualizações frequentes ou capacidade de resposta rápida.

Por contraste, os drones oferecem elevada flexibilidade de aquisição e resolução submétrica, permitindo recolhas sob pedido e em datas estrategicamente definidas. No presente estudo, essa liberdade possibilitou a aquisição durante o verão, período que se revelou particularmente vantajoso para a separação de coberturas vegetais. A secura sazonal das herbáceas mitigou a interferência espectral entre classes, favorecendo a diferenciação de vegetação perene, solo nu e elementos antrópicos. Este resultado sublinha o papel crítico da época de aquisição na eficácia dos processos classificatórios, sobretudo em contextos mediterrânicos ou semiáridos.

Para a gestão de reservas naturais como a Faia Brava, a conjugação entre dados geoespaciais de alta resolução, algoritmos de ML validados e ferramentas *open source* constitui uma solução concreta e eficiente. A capacidade de produzir cartografia temática detalhada, atualizável e adaptada às dinâmicas locais representa um recurso estratégico para planeamento, monitorização e conservação, com forte potencial de replicação noutros territórios com desafios semelhantes.

Em síntese, a metodologia aqui desenvolvida confirma a viabilidade técnica e a relevância prática da aplicação de técnicas de ML em ambientes operacionais reais. A sua estrutura modular, o uso de dados acessíveis e a integração com software livre conferem-lhe escalabilidade, robustez e aplicabilidade imediata em múltiplos contextos de gestão territorial e ambiental.

5. CONCLUSÕES E TRABALHOS FUTUROS

5.1. Principais Conclusões

O presente estudo teve como objetivo central a avaliação da exequibilidade e aplicabilidade de uma abordagem baseada em *software* livre — concretamente o QGIS com o *plugin* OTB — para a classificação da ocupação e uso do solo, com recurso a algoritmos de ML. A análise foi realizada na Reserva da Faia Brava, um território de elevada complexidade ecológica, utilizando dados obtidos a partir de imagens de satélite (OrtoSat2023) e ortofotomosaicos UAV (drone), e integrando outras variáveis espectrais e texturais.

Os principais resultados confirmam o cumprimento integral dos objetivos gerais e específicos da investigação:

O ambiente QGIS + OTB mostrou-se plenamente funcional e fiável para a implementação de fluxos de trabalho completos de classificação supervisionada, desde o pré-processamento das imagens até à geração dos mapas temáticos e à avaliação dos modelos. A possibilidade de executar algoritmos avançados como RF e SVM sem recorrer a linhas de comando constitui uma mais-valia considerável, tornando-a uma metodologia acessível a técnicos e investigadores não especializados em informática e geoestatística.

A incorporação de variáveis texturais derivadas da GLCM trouxe ganhos substanciais na qualidade da classificação. Estas métricas captaram padrões espaciais que os dados espectrais, por si só, não conseguiam representar, contribuindo para uma melhor diferenciação entre coberturas vegetais semelhantes e elementos estruturais da paisagem.

O OrtoFotoUAV demonstrou superioridade inequívoca face ao OrtoSat2023. Para além da resolução substancialmente mais elevada (2,7 cm/pixel), nos seus dados originais, os UAVs oferecem a possibilidade de aquisição consoante a necessidade, permitindo um controlo mais preciso sobre o momento e as condições de recolha. A realização dos voos no verão revelou-se uma decisão estratégica acertada, dado que a secura das herbáceas reduziu a confusão espectral entre estas e outras coberturas, favorecendo a performance dos algoritmos.

A validação estatística, através de métricas como os coeficientes *F1-score* e o *Kappa*, indicou níveis elevados de precisão e coerência interna dos modelos, reforçando a fiabilidade dos resultados obtidos. Estes indicadores confirmam que a metodologia adotada é adequada para a caracterização e monitorização de áreas ambientalmente sensíveis.

Em conclusão, esta dissertação valida a eficácia de integrar algoritmos de ML em ambientes de *software* livre com dados multiespectrais e de alta resolução espacial. A abordagem utilizada demonstrou ser cientificamente robusta, operacionalmente viável e tecnicamente acessível, destacando-se como uma solução de grande utilidade para a gestão territorial em áreas de reservas e para outras aplicações de monitorização ambiental com recursos limitados.

5.2. Limitações do Estudo

Apesar dos resultados satisfatórios obtidos neste estudo, importa reconhecer um conjunto de limitações que condicionaram, em certa medida, a abrangência e a generalização dos mesmos. Estas restrições estão relacionadas com aspetos técnicos, metodológicos e operacionais, e devem ser tidas em consideração em futuros desenvolvimentos.

Uma das principais limitações prende-se com a resolução e a disponibilidade temporal das imagens de satélite OrtoSat2023. Embora apresentem uma resolução aceitável (30 cm/pixel), estas imagens só são disponibilizadas gratuitamente, em território nacional, de forma pontual e com cobertura limitada no tempo. Esta restrição temporal compromete a atualização regular dos dados e limita o seu uso em estratégias de monitorização contínua. Além disso, a resolução espacial, ainda que adequada para análises gerais, mostrou-se insuficiente para distinguir elementos de menor escala, quando comparada com os OrtoFotoUAV.

No que diz respeito às imagens UAV, apesar da elevada qualidade espacial e flexibilidade de aquisição, a área coberta foi limitada, o que impõe constrangimentos à extrapolação do modelo para áreas extensas. Acresce que o uso de drones requer condições meteorológicas favoráveis, autorizações legais específicas e recursos técnicos adicionais para planificação de voo, processamento e armazenamento de dados, o que pode dificultar a sua utilização em contextos operacionais mais restritivos.

Do ponto de vista metodológico, embora os algoritmos RF e SVM tenham sido bem-sucedidos, a sua parametrização no ambiente gráfico do OTB apresenta algumas limitações, nomeadamente no controlo fino de hiperparâmetros e no acesso a funcionalidades mais avançadas que estariam disponíveis através de implementação por código. A ausência de afinação detalhada pode influenciar o desempenho dos modelos em situações mais complexas ou heterogêneas.

Em termos computacionais, o processamento de composições multibanda integrando bandas espectrais, índices de vegetação e métricas texturais revelou-se exigente, tanto em tempo de execução como capacidade de armazenamento. Esta limitação poderá dificultar a aplicação da metodologia em ambientes com infraestrutura computacional mais modesta, sobretudo quando se pretende alargar a análise a áreas mais vastas ou incluir dados multitemporais.

Por fim, destaca-se a necessidade de uma validação mais extensa no terreno. Embora tenham sido utilizados pontos de controlo e dados de referência com base em ortofotomosaicos de muito alta resolução, a confirmação sistemática das classes através de visitas de campo georreferenciadas teria reforçado a robustez da validação e permitido identificar potenciais ambiguidades na classificação automática. Esta limitação pode ser superada em trabalhos futuros com a integração de campanhas de campo específicas, ajustadas aos períodos de aquisição das imagens.

Adicionalmente, foi necessário garantir que as amostras utilizadas no treino e validação fossem espacialmente coincidentes nas três imagens analisadas (OrtoFotoUAV, OrtoSat2023 cor verdadeira e falsa cor). Para isso, os polígonos foram desenhados de forma reduzida e centrada em áreas homogêneas, evitando sombras e bordos. Apesar do alinhamento espacial realizado, persistiram pequenos desajustes entre as imagens que, associados ao tamanho limitado dos polígonos, podem ter originado a contaminação de alguns pixels com classes adjacentes. As diferenças temporais entre os produtos (cerca de um ano) também podem ter influenciado a consistência espectral das amostras. Trabalhos futuros poderão contornar esta limitação através da extração de amostras específicas para cada imagem.

5.3. Recomendações e Melhorias

Tendo como princípio orientador a utilização exclusiva de ferramentas de código aberto, as recomendações aqui apresentadas visam melhorar e aprofundar a metodologia aplicada, mantendo a acessibilidade e a replicabilidade do processo. As melhorias sugeridas baseiam-se nas observações práticas feitas ao longo do estudo, bem como nas limitações detetadas em termos de desempenho computacional, otimização dos modelos e qualidade visual dos resultados.

Uma primeira recomendação fundamental é a integração de Modelos Digitais de Superfície (MDS) obtidos através da fotogrametria dos dados UAV. Esta adição permitiria representar a dimensão vertical da vegetação — fator crítico para a distinção entre coberturas herbáceas, arbustivas e arbóreas. A informação altimétrica derivada, como altura relativa ou densidade estrutural, pode ser usada como variável adicional nos modelos de classificação e melhorar substancialmente a capacidade de discriminar classes com respostas espectrais semelhantes. Importa ainda referir que os MDS e MDT provenientes do Levantamento LiDAR de Portugal Continental, disponibilizado pela DGT, não foram integrados neste estudo, uma vez que apenas se tornaram acessíveis numa fase final da elaboração da tese. Apesar disso, representam um recurso valioso, com elevado potencial para complementar e enriquecer futuras aplicações desta metodologia, sobretudo no que diz respeito à caracterização estrutural da vegetação em larga escala.

Relativamente ao SVM, embora o algoritmo tenha apresentado métricas de validação robustas, com elevada exatidão e bons valores de *F1-score* e *Kappa*, os mapas gerados evidenciaram ruído visual considerável, com fragmentação excessiva e contornos artificiais entre as classes. Esta discrepância entre desempenho estatístico e qualidade visual poderá estar relacionada com o uso exclusivo do *kernel* linear, que, apesar de otimizado automaticamente pelo OTB, pode não ser o mais adequado para padrões espaciais não lineares. Recomenda-se, por isso, a exploração controlada do *kernel* radial (RBF), conhecido pela sua capacidade de generalização em cenários complexos. No entanto, importa reconhecer que este *kernel* é computacionalmente exigente, o que pode limitar a sua aplicação em computadores pessoais. Como solução intermédia, sugere-se avaliar estratégias de redução de dimensionalidade (como PCA) ou restrição da área de

análise, para viabilizar o uso do *kernel* RBF mesmo em ambientes de processamento mais modestos.

Em relação ao RF, os parâmetros utilizados, como o número de árvores, profundidade máxima e variáveis por divisão, foram definidos manualmente e aplicados de forma uniforme ao longo do estudo. Embora essa abordagem tenha produzido bons resultados, recomenda-se, para trabalhos futuros, avaliar sistematicamente a sensibilidade do modelo a diferentes combinações de parâmetros. A criação de conjuntos de treino e validação específicos para testes de afinação poderá permitir identificar configurações mais eficientes, adaptadas à complexidade de cada composição multibanda (por exemplo, apenas bandas, bandas + índices, bandas + texturas).

É igualmente relevante verificar se os mesmos parâmetros utilizados no RF são apropriados para o SVM, uma vez que os dois algoritmos possuem lógicas internas distintas e podem responder de forma divergente à mesma estrutura de dados. A utilização automática de parâmetros no SVM, embora prática, pode beneficiar de algum controlo externo, mesmo em ambiente gráfico, caso sejam disponibilizadas futuras atualizações do OTB com opções mais avançadas.

Por fim, recomenda-se o desenvolvimento de fluxos de trabalho documentados e reutilizáveis dentro do QGIS e OTB, incluindo *templates* de composição multibanda, conjuntos de parâmetros testados e recomendações para pré-processamento e normalização. Esta prática contribuirá para reforçar a reprodutibilidade dos resultados e facilitará a adoção desta abordagem por outros técnicos e investigadores que operem com recursos limitados, mas que pretendam aplicar metodologias robustas de classificação supervisionada em contextos ecológicos. Estas melhorias permitirão consolidar a qualidade técnica da abordagem atual, respeitando os princípios de acessibilidade, transparência e inovação que orientaram todo o desenvolvimento desta dissertação.

5.4. Propostas para Trabalhos Futuros

Tendo como base os resultados e limitações identificados nesta dissertação delineiam-se algumas propostas de continuidade que podem reforçar a robustez metodológica, ampliar a aplicabilidade da abordagem e explorar novas dimensões da análise de ocupação e uso do solo com ferramentas de código aberto.

Uma prioridade para trabalhos futuros consiste na validação alargada dos modelos em diferentes contextos ecológicos, dentro e fora da Faia Brava. A aplicação da metodologia a outras áreas, especialmente com diferentes padrões de vegetação ou uso do solo, permitirá aferir a sua generalização e sensibilidade a novas paisagens. Esta replicação deverá incluir campanhas de campo dedicadas, com recolha de amostras georreferenciadas para aumentar a fiabilidade da verdade terrestre e melhorar o treino dos algoritmos.

Paralelamente, será importante avaliar o impacto da inclusão de dados altimétricos (como MDS) em cenários multitemporais, por exemplo, para detetar alterações sazonais ou estruturais na vegetação. A possibilidade de monitorizar alterações verticais, como regeneração de matos ou crescimento de árvores jovens, pode abrir novas aplicações para a metodologia, nomeadamente em ecologia da sucessão e avaliação de impacto de projetos de restauro.

No domínio computacional recomenda-se explorar mais profundamente o potencial do *plugin* OTB integrado no QGIS, nomeadamente através da aplicação do algoritmo SVM com *kernel* radial (RBF) e da abordagem OBIA (*Object-Based Image Analysis*). Estas opções já se encontram disponíveis na interface gráfica do OTB, sem necessidade de recorrer a programação externa, permitindo configurar classificações mais avançadas com base em objetos segmentados e fronteiras não lineares. A utilização do *kernel* RBF é especialmente relevante em contextos com elevada heterogeneidade espectral, como o da Faia Brava, enquanto a OBIA permite integrar padrões espaciais e contextuais na análise, superando as limitações das abordagens baseadas apenas em pixels. Tal como revisto por Cini et al. (2024), esta integração de dados multissensoriais através do GEOBIA revela-se promissora para projetos de monitorização ambiental em ecossistemas complexos. Assim, futuras investigações poderão beneficiar da combinação destas funcionalidades, mantendo-se integralmente no ecossistema *open source* e potenciando uma maior precisão e robustez nos processos de classificação.

Em suma, as propostas aqui enunciadas apontam para a consolidação e evolução da abordagem desenvolvida nesta dissertação, alicerçada em *software* livre e técnicas de classificação supervisionada. A replicação em novas áreas, o reforço da validação com dados de campo e a integração de métodos mais avançados, como a OBIA e o *kernel* RBF, constituem caminhos



sólidos para o aprofundamento científico e técnico da metodologia, com forte potencial de aplicação em contextos de conservação e gestão territorial.

5.5. Contribuições para os Objetivos de Desenvolvimento Sustentável (ODS)

Este trabalho contribui diretamente para os ODS, em especial o ODS 15 – Proteger a Vida Terrestre, ao aplicar métodos acessíveis de monitorização ambiental numa area protegida. Também se relaciona com o ODS 13 – Ação Climática, ao oferecer ferramentas para detetar alterações no uso do solo associadas a processos como a desertificação. A utilização de *software* livre e interface gráficas está alinhada com o ODS 9 – Industria, Inovação e Infraestruturas, ao tornar tecnologias avançadas aplicáveis em contextos com poucos recursos. Por fim, o caráter replicável da metodologia reforça o ODS 4 – Educação de Qualidade, ao facilitar o acesso a técnicas de análise especial por parte de profissionais não especializados em programação.

BIBLIOGRAFIA

Agência para a Modernização Administrativa (AMA). (2023). OrtoSat2023 30 cm - Portugal Continental - 2023. Dados.gov.pt. <https://dados.gov.pt/pt/datasets/OrtoSat2023-30-cm-portugal-continental-2023/>

Anderson, K., & Gaston, K. J. (2013). Lightweight unmanned aerial vehicles will revolutionize spatial ecology. *Frontiers in Ecology and the Environment*, 11(3), 138–146. <https://doi.org/10.1890/120150>

Belgiu, M., & Drăguț, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, 24–31. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>

Blaschke, T., Hay, G. J., Kelly, M., Lang, S., Hofmann, P., Addink, E., Feitosa, R. Q., van der Meer, F., van der Werff, H., van Coillie, F., & Tiede, D. (2014). Geographic Object-Based Image Analysis – Towards a new paradigm. *ISPRS Journal of Photogrammetry and Remote Sensing*, 87, 180–191. <https://doi.org/10.1016/j.isprsjprs.2013.09.014>

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>

Butt, B. (2018). Environmental indicators and governance. *Current Opinion in Environmental Sustainability*, 32, 84–89. <https://doi.org/10.1016/j.cosust.2018.05.006>

Cini, E., Marzioletti, F., Paterni, M., Berton, A., Acosta, A., & Ciccarelli, D. (2024). Integrating UAV imagery and machine learning via Geographic Object Based Image Analysis (GEOBIA) for enhanced monitoring of *Yucca gloriosa* in Mediterranean coastal dunes. *Ocean & Coastal Management*, 258, 107377. <https://doi.org/10.1016/j.ocecoaman.2024.107377>

Coetzee, S., Ivánová, I., Mitasova, H., & Brovelli, M. A. (2020). Open geospatial software and data: A review of the current state and a perspective into the future. *ISPRS International Journal of Geo-Information*, 9(2), 90. <https://doi.org/10.3390/ijgi9020090>

de Castro, A. I., Torres-Sánchez, J., Peña, J. M., Jiménez-Brenes, F. M., Csillik, O., & López-Granados, F. (2018). An automatic Random Forest-OBIA algorithm for early weed mapping between and within crop rows using UAV imagery. *Remote Sensing*, 10(2), 285. <https://doi.org/10.3390/rs10020285>

Delgado, R., & Tibau, X.-A. (2019). Why Cohen’s Kappa should be avoided as performance measure in classification. *PLOS ONE*, 14(9), e0222916. <https://doi.org/10.1371/journal.pone.0222916>

Dias, L. S. O., Assunção, A. F., & Jesus, F. A. N. de. (2024). Free geospatial software and applications. In *Revolutionizing Earth Observation – New Technologies and Insights*. IntechOpen. <https://doi.org/10.5772/intechopen.1005682>



Direção-Geral do Território. (2023). Documentação técnica do OrtoSat2023. <https://smos.dgterritorio.gov.pt/cartografia-de-base>

Ferreira, L. G., Yoshioka, H., Huete, A., & Sano, E. E. (2003). Seasonal landscape and spectral vegetation index dynamics in the Brazilian Cerrado: An analysis within the Large-Scale Biosphere–Atmosphere Experiment in Amazonia (LBA). *Remote Sensing of Environment*, 87, 534–550. <https://doi.org/10.1016/j.rse.2002.09.003>

Figueiredo, G. C., & Vieira, C. A. O. (2007). Estudo do comportamento dos índices de exatidão global, Kappa e Tau, comumente usados para avaliar a classificação de imagens do sensoriamento remoto. Anais do XIII Simpósio Brasileiro de Sensoriamento Remoto, 5755–5762. Instituto Nacional de Pesquisas Espaciais (INPE).

Foody, G. M. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80(1), 185–201. [https://doi.org/10.1016/S0034-4257\(01\)00295-4](https://doi.org/10.1016/S0034-4257(01)00295-4)

Grandini, M., Bagli, E., & Visani, G. (2020). Metrics for multi-class classification: An overview (White paper). CRIF S.p.A. & University of Bologna. <https://arxiv.org/abs/2008.05756>

Grizonnet, M., Michel, J., Poughon, V., Inglada, J., Savinaud, M., & Cresson, R. (2017). Orfeo ToolBox: Open source processing of remote sensing images. *Open Geospatial Data, Software and Standards*, 2(15). <https://doi.org/10.1186/s40965-017-0031-6>

Haralick, R. M. (1986). Chapter 11: Statistical image texture analysis (pp. 247–272). In *Computer and Robot Vision* (Vol. 1). Ann Arbor, MI: Academic Press. https://haralick.org/book_chapters/Statistical_Image_Texture.pdf

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2^a ed). Springer.

Holloway, J., & Mengersen, K. (2018). Statistical machine learning methods and remote sensing for sustainable development goals: A review. *Remote Sensing*, 10(9), 1365. <https://doi.org/10.3390/rs10091365>

Hossin, M., & Sulaiman, M. N. (2015). A review on evaluation metrics for data classification evaluations. *International Journal of Data Mining & Knowledge Management Process*, 5(2), 1–11. <https://doi.org/10.5121/ijdkp.2015.5201>

Huang, S., Tang, L., Hupy, J. P., Wang, Y., & Shao, G. (2021). A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing. *Journal of Forestry Research*, 32(1), 1–6. <https://doi.org/10.1007/s11676-020-01155-1>

Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, 25(3), 295–309. [https://doi.org/10.1016/0034-4257\(88\)90106-X](https://doi.org/10.1016/0034-4257(88)90106-X)

IPMA – Instituto Português do Mar e da Atmosfera. (2025a). Classificação do clima de Portugal Continental. Instituto Português do Mar e da Atmosfera.

<https://www.ipma.pt/pt/educativa/tempo.clima/index.jsp?page=clima.pt.xml>.

IPMA - Instituto Português do Mar e da Atmosfera. (2025b). Classificação do clima de Portugal Continental. Instituto Português do Mar e da Atmosfera.

<https://www.ipma.pt/pt/oclima/observatorio.secas/pdsi/apresentacao/evolu.historica/>

Lourenço, P., Teodoro, A. C., Gonçalves, J. A., Honrado, J. P., Cunha, M., & Sillero, N. (2021). Assessing the performance of different OBIA software approaches for mapping invasive alien plants along roads with remote sensing data. *International Journal of Applied Earth Observation and Geoinformation*, 95, 102263. <https://doi.org/10.1016/j.jag.2020.102263>

Mahesh, B. (2020). Machine learning algorithms - A review. *International Journal of Science and Research (IJSR)*, 9(1), 381–386. <https://doi.org/10.21275/ART20203995>

Manomba-Mbadinga, N., Niculescu, S., Zaabar, N., Mombo, J.-B., & Xie, G. (2023). Grand Libreville (Gabon) coastline machine learning and convolutional neural network detection and automatic extraction of the methods. *SPIE Europe International Symposium, Remote Sensing, Earth Resources and Environmental Remote Sensing/GIS Applications XIV*, 12734. <https://doi.org/10.1117/12.2678897>

Maxwell, A. E., Warner, T. A., & Fang, F. (2018). Implementation of machine-learning classification in remote sensing: an applied review. *International Journal of Remote Sensing*, 39(9), 2784–2817. <https://doi.org/10.1080/01431161.2018.1433343>

Mountrakis, G., Im, J., & Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(3), 247–259. <https://doi.org/10.1016/j.isprsjprs.2010.11.001>

Müllerová, J., Brůna, J., Dvořák, P., Bartaloš, T., & Vítková, M. (2016). Does the data resolution/origin matter? Satellite, airborne and UAV imagery to tackle plant invasions. *ISPRS Archives, XLI-B7*, 903–908. <https://doi.org/10.5194/isprs-archives-XLI-B7-903-2016>

Landis, J. R., Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, v. 33, n. 1, p. 159.

Li, Y. (2018). Deep reinforcement learning: An overview. arXiv preprint arXiv:1701.07274. <https://arxiv.org/abs/1701.07274>

Orfeo Toolbox (OTB). (2024). Haralick Texture Extraction Application. Orfeo Toolbox CookBook. https://www.orfeo-toolbox.org/CookBook/Applications/app_HaralickTextureExtraction.html

OTB Development Team. (2023). *Orfeo Toolbox software documentation*. <https://www.orfeo-toolbox.org/CookBook/>

- Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, 26(1), 217–222. <https://doi.org/10.1080/01431160412331269698>
- Pelletier, C., Valero, S., Inglada, J., Champion, N., & Dedieu, G. (2020). A Random Forest-Based Approach for Land Cover Classification Using Sentinel-2 Data. *Remote Sensing*, 12(7), 1135. <https://doi.org/10.3390/rs12071135>
- Pettorelli, N., Vik, J. O., Mysterud, A., Gaillard, J.-M., Tucker, C. J., & Stenseth, N. C. (2005). Using the satellite-derived NDVI to assess ecological responses to environmental change. *Trends in Ecology & Evolution*, 20(9), 503–510. <https://doi.org/10.1016/j.tree.2005.05.011>
- Phan Thanh Noi, & Kappas, M. (2018). Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. *Sensors*, 18(1), 18. <https://doi.org/10.3390/s18010018>
- Radočaj, D., Šiljeg, A., Marinović, R., & Jurišić, M. (2023). State of major vegetation indices in precision agriculture studies indexed in Web of Science: A review. *Agriculture*, 13(3), 707. <https://doi.org/10.3390/agriculture13030707>
- Rosas-Chavoya, M., Gallardo-Salazar, J. L., López-Serrano, P. M., Alcántara-Concepción, P. C., & León-Miranda, A. K. (2022). QGIS: A constantly growing free and open-source geospatial software contributing to scientific development. *Cuadernos de Investigación Geográfica*, 48(1), 197–213. <https://doi.org/10.18172/cig.5143>
- Rouse, J. W., Haas, R. H., Schell, J. A. & Deering, D. W. (1973). Monitoring vegetation systems in the Great Plains with ERTS. 3rd ERTS Symposium, NASA SP-351, Vol. I, pp. 309–317. <https://ntrs.nasa.gov/citations/19740022614>
- Srivastava, D. K., & Bhambhu, L. (2009). Data classification using support vector machine. *Journal of Theoretical and Applied Information Technology*, 12(1), 1–7. https://jatit.org/volumes/twelfth_volume_1_2010.php
- Wang, F., Yi, Q., Hu, J., Xie, L., Yao, X., Xu, T., & Zheng, J. (2021). Combining spectral and textural information in UAV hyperspectral images to estimate rice grain yield. *International Journal of Applied Earth Observation and Geoinformation*, 102, 102397. <https://doi.org/10.1016/j.jag.2021.102397>
- Xue, J., & Su, B. (2017). Significant remote sensing vegetation indices: A review of developments and applications. *Journal of Sensors*, 2017, Article ID 1353691. <https://doi.org/10.1155/2017/1353691>

ANEXOS

Anexo 1- Desempenho das composições testadas com o algoritmo Random Forest (RF) aplicado ao OrtoFotoUAV. Apresentam-se os valores de precisão, revocação, F1-score por classe e o coeficiente Kappa total. A composição RGB+GLI+B2 obteve o melhor desempenho global (assinalado a verde), enquanto RGB+GLI+HS apresentou o pior resultado (assinalado a vermelho).

ORTOFOTOMOSAICO - RF													
Composição	Precisão				Revocação				F1-score				Kappa Total
	Classe1	Classe2	Classe3	Classe4	Classe1	Classe2	Classe3	Classe4	Classe1	Classe2	Classe3	Classe4	
RGB	0,987	0,639	0,940	0,745	0,991	0,897	0,944	0,438	0,989	0,746	0,942	0,551	0,757
RGB+GLI	0,996	0,642	0,940	0,727	0,996	0,884	0,948	0,446	0,996	0,744	0,944	0,553	0,758
RGB+GLI+B1	0,996	0,665	0,926	0,849	0,996	0,979	0,914	0,459	0,996	0,792	0,920	0,596	0,783
RGB+GLI+B2	0,996	0,723	0,941	0,869	0,991	0,987	0,897	0,597	0,994	0,835	0,919	0,707	0,824
RGB+GLI+B3	0,991	0,645	0,893	0,686	0,996	0,906	0,897	0,403	0,994	0,754	0,895	0,508	0,734
RGB+GLI+B4	0,996	0,646	0,921	0,873	1,000	0,987	0,948	0,382	0,998	0,781	0,934	0,531	0,773
RGB+GLI+B5	0,996	0,687	0,863	0,833	1,000	0,953	0,923	0,451	0,998	0,799	0,892	0,585	0,775
RGB+GLI+B6	0,991	0,653	0,929	0,811	0,996	0,944	0,953	0,425	0,994	0,772	0,941	0,558	0,773
RGB+GLI+B7	0,996	0,729	0,911	0,856	1,000	0,948	0,837	0,665	0,998	0,825	0,872	0,749	0,817
RGB+GLI+B8	0,991	0,677	0,914	0,806	0,996	0,979	0,910	0,446	0,994	0,800	0,912	0,575	0,777
RGB+GLI+HS	0,996	0,593	0,829	0,670	0,996	0,845	0,918	0,313	0,996	0,697	0,872	0,427	0,691

Anexo 2 - Desempenho das composições testadas com o algoritmo Support Vector Machine (SVM) aplicado ao OrtoFotoUAV. Apresentam-se os valores de precisão, revocação, F1-score por classe e o coeficiente Kappa total. A composição RGB+GLI+B2 obteve o melhor desempenho global (assinalado a verde), enquanto RGB+GLI+HS apresentou o pior resultado (assinalado a vermelho).

ORTOFOTOMOSAICO - SVM													
Composição	Precisão				Revocação				F1-score				Kappa Total
	Classe1	Classe2	Classe3	Classe4	Classe1	Classe2	Classe3	Classe4	Classe1	Classe2	Classe3	Classe4	
RGB	0,979	0,755	0,978	0,828	1,000	0,914	0,948	0,661	0,989	0,827	0,963	0,735	0,841
RGB+GLI	1,000	0,771	0,953	0,840	1,000	0,923	0,948	0,678	1,000	0,840	0,951	0,751	0,850
RGB+GLI+B1	0,959	0,742	0,927	0,976	1,000	1,000	0,987	0,532	0,979	0,852	0,956	0,689	0,840
RGB+GLI+B2	0,983	0,812	0,941	0,947	1,000	1,000	0,961	0,691	0,991	0,896	0,951	0,799	0,884
RGB+GLI+B3	0,983	0,819	0,881	0,872	1,000	0,953	0,953	0,644	0,991	0,881	0,915	0,741	0,850
RGB+GLI+B4	0,967	0,766	0,962	0,954	1,000	1,000	0,970	0,622	0,983	0,868	0,966	0,753	0,864
RGB+GLI+B5	0,979	0,833	0,931	0,897	1,000	0,987	0,931	0,712	0,989	0,904	0,931	0,794	0,877
RGB+GLI+B6	0,991	0,777	0,894	0,835	1,000	0,940	0,940	0,609	0,996	0,850	0,916	0,705	0,830
RGB+GLI+B7	0,987	0,773	0,982	0,853	1,000	0,918	0,961	0,700	0,994	0,839	0,972	0,769	0,860
RGB+GLI+B8	1,000	0,802	0,956	0,895	0,983	0,974	0,940	0,734	0,991	0,880	0,948	0,807	0,877
RGB+GLI+HS	0,983	0,644	0,875	0,809	1,000	0,931	0,991	0,326	0,991	0,761	0,930	0,465	0,750

Anexo 3 - Desempenho das composições testadas com o algoritmo RF aplicado às imagens OrtoSat2023 (cor verdadeira). Apresentam-se os valores de precisão, revocação, F1-score por classe e o coeficiente Kappa total. A composição RGB obteve o melhor desempenho global (assinalado a verde), enquanto RGB+GLI+HS apresentou o pior resultado (assinalado a vermelho).

ORTOSAT2023 COR VERDADEIRA - RF													
Composição	Precisão				Revocação				F1-score				Kappa
	Classe1	Classe2	Classe3	Classe4	Classe1	Classe2	Classe3	Classe4	Classe1	Classe2	Classe3	Classe4	
RGB	0,978	0,682	0,969	0,806	0,944	0,845	0,944	0,661	0,961	0,755	0,957	0,726	0,798
RGB+GLI	0,978	0,684	0,956	0,809	0,944	0,837	0,936	0,674	0,961	0,753	0,946	0,735	0,797
RGB+GLI+B1	0,991	0,470	0,948	0,548	0,944	0,670	0,854	0,395	0,967	0,552	0,898	0,459	0,621
RGB+GLI+B2	0,982	0,520	0,961	0,690	0,944	0,833	0,850	0,382	0,963	0,640	0,902	0,492	0,670
RGB+GLI+B3	0,974	0,555	0,943	0,811	0,948	0,914	0,850	0,386	0,961	0,690	0,894	0,523	0,700
RGB+GLI+B4	1,000	0,525	0,950	0,750	0,940	0,871	0,824	0,399	0,969	0,655	0,883	0,521	0,678
RGB+GLI+B5	0,956	0,551	0,938	0,830	0,936	0,923	0,837	0,378	0,946	0,690	0,884	0,519	0,691
RGB+GLI+B6	0,982	0,672	0,956	0,835	0,918	0,897	0,931	0,631	0,949	0,768	0,943	0,719	0,793
RGB+GLI+B7	0,982	0,609	0,964	0,749	0,931	0,837	0,927	0,536	0,956	0,705	0,945	0,625	0,744
RGB+GLI+B8	0,973	0,596	0,919	0,856	0,918	0,936	0,927	0,408	0,945	0,728	0,923	0,552	0,730
RGB+GLI+HS	0,986	0,479	0,781	0,651	0,901	0,781	0,627	0,425	0,942	0,594	0,695	0,514	0,578

Anexo 4 - Desempenho das composições testadas com o algoritmo SVM aplicado às imagens OrtoSat2023 (cor verdadeira). Apresentam-se os valores de precisão, revocação, F1-score por classe e o coeficiente Kappa total. A composição RGB+GLI+HS obteve o melhor desempenho global (assinalado a verde), enquanto RGB+GLI+B3 apresentou o pior resultado (assinalado a vermelho).

ORTOSAT2023 COR VERDADEIRA - SVM													
Composição	Precisão				Revocação				F1-score				Kappa
	Classe1	Classe2	Classe3	Classe4	Classe1	Classe2	Classe3	Classe4	Classe1	Classe2	Classe3	Classe4	
RGB	0,894	0,643	0,911	0,686	0,944	0,828	0,961	0,412	0,919	0,724	0,935	0,515	0,715
RGB+GLI	0,894	0,645	0,902	0,669	0,944	0,794	0,948	0,442	0,919	0,712	0,925	0,532	0,710
RGB+GLI+B1	0,911	0,523	0,874	0,397	0,966	0,691	0,923	0,223	0,938	0,595	0,898	0,286	0,601
RGB+GLI+B2	0,946	0,536	0,880	0,426	0,979	0,695	0,910	0,270	0,962	0,606	0,895	0,331	0,618
RGB+GLI+B3	0,938	0,495	0,864	0,414	0,979	0,639	0,901	0,258	0,958	0,558	0,882	0,317	0,592
RGB+GLI+B4	0,942	0,529	0,930	0,457	0,970	0,708	0,914	0,296	0,956	0,606	0,922	0,359	0,629
RGB+GLI+B5	0,894	0,542	0,906	0,455	0,974	0,725	0,910	0,258	0,932	0,620	0,908	0,329	0,622
RGB+GLI+B6	0,944	0,644	0,921	0,623	0,944	0,682	0,948	0,567	0,944	0,663	0,934	0,593	0,714
RGB+GLI+B7	0,958	0,548	0,941	0,597	0,974	0,777	0,953	0,330	0,966	0,643	0,947	0,425	0,678
RGB+GLI+B8	0,946	0,616	0,868	0,733	0,974	0,901	0,931	0,318	0,960	0,732	0,899	0,443	0,708
RGB+GLI+HS	0,966	0,642	0,969	0,848	0,974	0,910	0,953	0,502	0,970	0,753	0,961	0,631	0,780

Anexo 5 - Desempenho das composições testadas com o algoritmo RF aplicado às imagens OrtoSat2023 (falsa cor – NDVI). Apresentam-se os valores de precisão, revocação, F1-score por classe e o coeficiente Kappa total. A composição IRG+NDVI+B7 obteve o melhor desempenho global (assinalado a verde), enquanto IRG+NDVI+B4 apresentou o pior resultado (assinalado a vermelho).

ORTOSAT2023 FALSA COR - NDVI - RF													
Composição	Precisão				Revocação				F1-score				Kappa Total
	Classe1	Classe2	Classe3	Classe4	Classe1	Classe2	Classe3	Classe4	Classe1	Classe2	Classe3	Classe4	
IRG	1,000	0,703	0,932	0,766	0,966	0,721	1,000	0,717	0,983	0,712	0,965	0,741	0,801
IRG+NDVI	1,000	0,693	0,931	0,760	0,961	0,717	0,983	0,721	0,980	0,705	0,956	0,740	0,794
IRG+NDVI+B1	1,000	0,564	0,845	0,661	0,970	0,605	0,798	0,670	0,985	0,584	0,821	0,665	0,681
IRG+NDVI+B2	1,000	0,712	0,867	0,741	0,970	0,678	0,867	0,798	0,985	0,695	0,867	0,769	0,771
IRG+NDVI+B3	1,000	0,594	0,883	0,733	0,970	0,760	0,936	0,506	0,985	0,667	0,908	0,599	0,724
IRG+NDVI+B4	1,000	0,552	0,910	0,565	0,961	0,614	0,828	0,575	0,980	0,581	0,867	0,570	0,660
IRG+NDVI+B5	1,000	0,677	0,860	0,710	0,970	0,755	0,820	0,682	0,985	0,714	0,840	0,696	0,742
IRG+NDVI+B6	1,000	0,704	0,923	0,726	0,974	0,674	0,983	0,730	0,987	0,689	0,952	0,728	0,787
IRG+NDVI+B7	1,000	0,757	0,903	0,832	0,966	0,777	0,914	0,828	0,983	0,767	0,908	0,830	0,828
IRG+NDVI+B8	1,000	0,689	0,902	0,776	0,979	0,712	0,944	0,730	0,989	0,700	0,922	0,752	0,788
IRG+NDVI+HS	1,000	0,558	0,819	0,700	0,970	0,644	0,738	0,682	0,985	0,598	0,777	0,691	0,678

Anexo 6 - Desempenho das composições testadas com o algoritmo SVM aplicado às imagens OrtoSat2023 (falsa cor – NDVI). Apresentam-se os valores de precisão, revocação, F1-score por classe e o coeficiente Kappa total. A composição IRG+NDVI+HS obteve o melhor desempenho global (assinalado a verde), enquanto IRG+NDVI+B3 apresentou o pior resultado (assinalado a vermelho).

ORTOSAT2023 FALSA COR - NDVI - SVM													
Composição	Precisão				Revocação				F1-score				Kappa Total
	Classe1	Classe2	Classe3	Classe4	Classe1	Classe2	Classe3	Classe4	Classe1	Classe2	Classe3	Classe4	
IRG	1,000	0,832	0,861	0,630	0,974	0,575	0,931	0,790	0,987	0,680	0,895	0,701	0,757
IRG+NDVI	1,000	0,660	0,921	0,714	0,974	0,725	1,000	0,601	0,987	0,691	0,959	0,653	0,767
IRG+NDVI+B1	1,000	0,577	0,917	0,688	0,979	0,773	1,000	0,408	0,989	0,661	0,957	0,512	0,720
IRG+NDVI+B2	1,000	0,577	0,914	0,671	0,987	0,760	1,000	0,403	0,994	0,656	0,955	0,504	0,717
IRG+NDVI+B3	1,000	0,536	0,910	0,596	0,983	0,704	1,000	0,361	0,991	0,609	0,953	0,449	0,682
IRG+NDVI+B4	1,000	0,573	0,914	0,678	0,966	0,760	1,000	0,416	0,983	0,653	0,955	0,516	0,714
IRG+NDVI+B5	1,000	0,576	0,914	0,674	0,983	0,751	1,000	0,416	0,991	0,652	0,955	0,515	0,717
IRG+NDVI+B6	1,000	0,605	0,921	0,690	0,987	0,755	1,000	0,468	0,994	0,672	0,959	0,558	0,737
IRG+NDVI+B7	0,996	0,589	0,928	0,701	0,983	0,751	1,000	0,464	0,989	0,660	0,963	0,558	0,732
IRG+NDVI+B8	1,000	0,606	0,928	0,836	0,979	0,893	1,000	0,395	0,989	0,722	0,963	0,536	0,755
IRG+NDVI+HS	0,991	0,620	0,952	0,830	0,991	0,854	0,944	0,524	0,991	0,718	0,948	0,642	0,771

Anexo 7 - Desempenho das composições testadas com o algoritmo RF aplicado às imagens OrtoSat2023 (falsa cor – SAVI). Apresentam-se os valores de precisão, revocação, F1-score por classe e o coeficiente Kappa total. A composição IRG+SAVI+B7 apresentou o melhor desempenho global (assinalado a verde), enquanto IRG+SAVI+B4 registou o pior resultado (assinalado a vermelho).

ORTOSAT2023 FALSA COR - SAVI - RF													
Composição	Precisão				Revocação				F1-score				Kappa
	Classe1	Classe2	Classe3	Classe4	Classe1	Classe2	Classe3	Classe4	Classe1	Classe2	Classe3	Classe4	
IRG	1,000	0,703	0,932	0,766	0,966	0,721	1,000	0,717	0,983	0,712	0,965	0,741	0,801
IRG+SAVI	0,978	0,678	0,924	0,739	0,966	0,704	0,987	0,670	0,972	0,691	0,954	0,703	0,775
IRG+SAVI+B1	0,983	0,564	0,836	0,654	0,970	0,588	0,785	0,674	0,976	0,576	0,810	0,664	0,672
IRG+SAVI+B2	0,991	0,715	0,864	0,731	0,979	0,678	0,876	0,768	0,985	0,696	0,870	0,749	0,767
IRG+SAVI+B3	0,991	0,593	0,883	0,722	0,970	0,751	0,936	0,502	0,980	0,663	0,908	0,592	0,720
IRG+SAVI+B4	0,987	0,526	0,914	0,553	0,966	0,601	0,824	0,541	0,976	0,561	0,867	0,547	0,644
IRG+SAVI+B5	0,996	0,682	0,870	0,728	0,974	0,773	0,833	0,678	0,985	0,724	0,851	0,702	0,753
IRG+SAVI+B6	0,996	0,702	0,916	0,731	0,979	0,678	0,987	0,712	0,987	0,690	0,950	0,722	0,785
IRG+SAVI+B7	0,996	0,744	0,892	0,823	0,966	0,747	0,918	0,820	0,980	0,745	0,905	0,822	0,817
IRG+SAVI+B8	0,979	0,684	0,896	0,744	0,979	0,678	0,927	0,725	0,979	0,681	0,911	0,735	0,770
IRG+SAVI+HS	0,996	0,573	0,802	0,704	0,970	0,639	0,730	0,704	0,983	0,604	0,764	0,704	0,681

Anexo 8 - Desempenho das composições testadas com o algoritmo SVM aplicado às imagens OrtoSat2023 (falsa cor – SAVI). Apresentam-se os valores de precisão, revocação, F1-score por classe e o coeficiente Kappa total. A composição IRG+SAVI+HS apresentou o melhor desempenho global (assinalado a verde), enquanto IRG+SAVI+B3 registou o pior resultado (assinalado a vermelho).

ORTOSAT2023 FALSA COR - SAVI - SVM													
Composição	Precisão				Revocação				F1-score				Kappa
	Classe1	Classe2	Classe3	Classe4	Classe1	Classe2	Classe3	Classe4	Classe1	Classe2	Classe3	Classe4	
IRG	1,000	0,832	0,861	0,630	0,974	0,575	0,931	0,790	0,987	0,680	0,895	0,701	0,757
IRG+SAVI	1,000	0,675	0,910	0,714	0,974	0,712	1,000	0,622	0,987	0,693	0,953	0,665	0,770
IRG+SAVI+B1	1,000	0,562	0,914	0,632	0,970	0,721	1,000	0,412	0,985	0,632	0,955	0,499	0,701
IRG+SAVI+B2	1,000	0,568	0,910	0,655	0,983	0,751	1,000	0,391	0,991	0,647	0,953	0,489	0,708
IRG+SAVI+B3	1,000	0,528	0,900	0,576	0,987	0,678	1,000	0,356	0,994	0,594	0,947	0,440	0,674
IRG+SAVI+B4	1,000	0,572	0,914	0,651	0,970	0,734	1,000	0,425	0,985	0,643	0,955	0,514	0,710
IRG+SAVI+B5	1,000	0,576	0,907	0,674	0,974	0,751	1,000	0,416	0,987	0,652	0,951	0,515	0,714
IRG+SAVI+B6	1,000	0,624	0,900	0,702	0,987	0,755	1,000	0,485	0,994	0,683	0,947	0,574	0,742
IRG+SAVI+B7	1,000	0,584	0,928	0,686	0,983	0,742	1,000	0,459	0,991	0,654	0,963	0,550	0,728
IRG+SAVI+B8	1,000	0,605	0,947	0,760	0,983	0,863	1,000	0,408	0,991	0,712	0,973	0,531	0,751
IRG+SAVI+HS	0,987	0,618	0,968	0,826	1,000	0,863	0,923	0,528	0,994	0,720	0,945	0,644	0,771

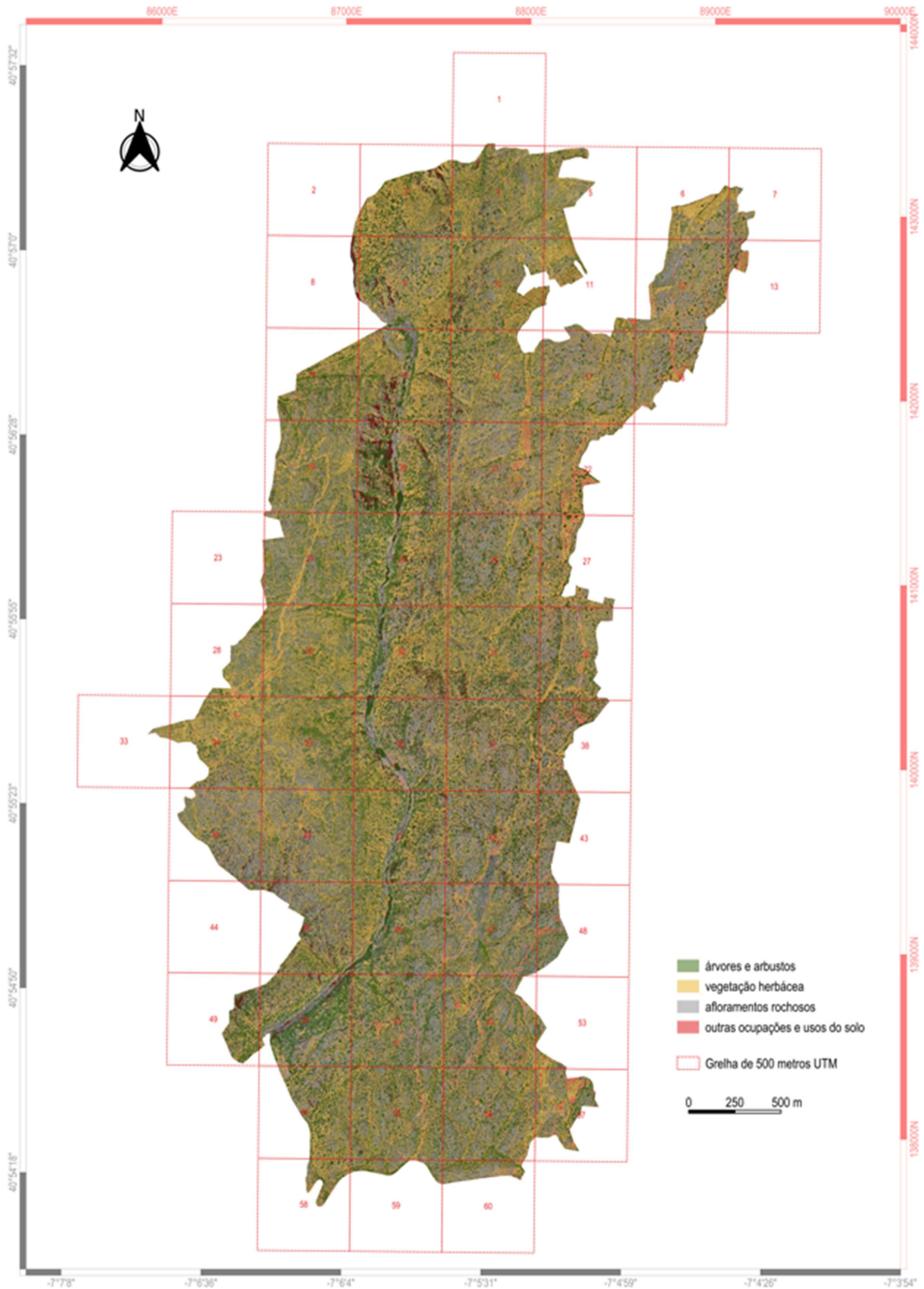
Anexo 9 - Desempenho das composições testadas com o algoritmo RF aplicado às imagens OrtoSat2023 (falsa cor – MSAVI). Apresentam-se os valores de precisão, revocação, F1-score por classe e o coeficiente Kappa total. A composição IRG+MSAVI+B7 apresentou o melhor desempenho global (assinalado a verde), enquanto IRG+MSAVI+B4 registou o pior resultado (assinalado a vermelho).

ORTOSAT2023 FALSA COR - MSAVI - RF													
Composição	Precisão				Revocação				F1-score				Kappa Total
	Classe1	Classe2	Classe3	Classe4	Classe1	Classe2	Classe3	Classe4	Classe1	Classe2	Classe3	Classe4	
IRG	1,000	0,703	0,932	0,766	0,966	0,721	1,000	0,717	0,983	0,712	0,965	0,741	0,801
IRG+MSAVI	0,978	0,674	0,927	0,736	0,966	0,700	0,987	0,670	0,972	0,686	0,956	0,701	0,774
IRG+MSAVI+B1	0,987	0,522	0,827	0,631	0,970	0,554	0,781	0,639	0,978	0,538	0,804	0,635	0,648
IRG+MSAVI+B2	0,987	0,668	0,849	0,730	0,979	0,665	0,820	0,764	0,983	0,667	0,834	0,746	0,742
IRG+MSAVI+B3	0,983	0,597	0,882	0,726	0,974	0,764	0,931	0,489	0,978	0,670	0,906	0,585	0,720
IRG+MSAVI+B4	0,978	0,532	0,903	0,569	0,961	0,605	0,798	0,567	0,970	0,566	0,847	0,568	0,644
IRG+MSAVI+B5	0,978	0,685	0,876	0,746	0,974	0,755	0,880	0,670	0,976	0,718	0,878	0,706	0,760
IRG+MSAVI+B6	0,983	0,691	0,923	0,723	0,974	0,682	0,979	0,695	0,978	0,687	0,950	0,709	0,777
IRG+MSAVI+B7	0,987	0,758	0,881	0,833	0,970	0,738	0,918	0,833	0,978	0,748	0,899	0,833	0,820
IRG+MSAVI+B8	0,979	0,682	0,901	0,746	0,979	0,682	0,936	0,717	0,979	0,682	0,918	0,731	0,771
IRG+MSAVI+HS	0,987	0,572	0,789	0,739	0,974	0,665	0,721	0,691	0,981	0,615	0,753	0,714	0,684

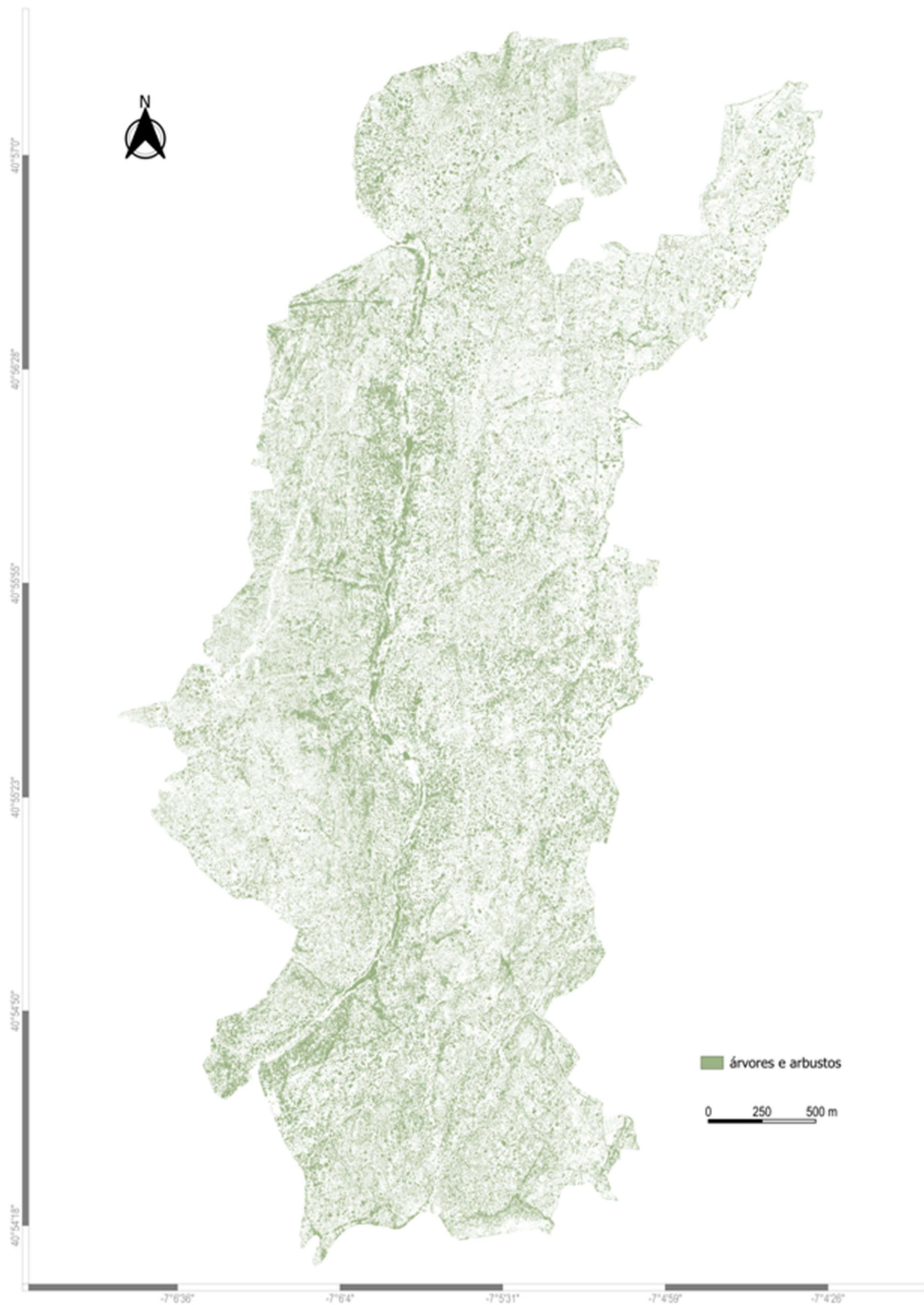
Anexo 10 - Desempenho das composições testadas com o algoritmo SVM aplicado às imagens OrtoSat2023 (falsa cor – MSAVI). Apresentam-se os valores de precisão, revocação, F1-score por classe e o coeficiente Kappa total. A composição IRG+MSAVI+HS apresentou o melhor desempenho global (assinalado a verde), enquanto IRG+MSAVI+B3 registou o pior resultado (assinalado a vermelho).

ORTOSAT2023 FALSA COR - MSAVI - SVM													
Composição	Precisão				Revocação				F1-score				Kappa Total
	Classe1	Classe2	Classe3	Classe4	Classe1	Classe2	Classe3	Classe4	Classe1	Classe2	Classe3	Classe4	
IRG	1,000	0,832	0,861	0,630	0,974	0,575	0,931	0,790	0,987	0,680	0,895	0,701	0,757
IRG+MSAVI	1,000	0,744	0,916	0,704	0,974	0,661	0,987	0,747	0,987	0,700	0,950	0,725	0,790
IRG+MSAVI+B1	1,000	0,670	0,921	0,649	0,970	0,652	0,948	0,665	0,985	0,661	0,934	0,657	0,745
IRG+MSAVI+B2	1,000	0,548	0,913	0,589	0,979	0,691	0,987	0,399	0,989	0,611	0,948	0,476	0,685
IRG+MSAVI+B3	1,000	0,520	0,900	0,559	0,987	0,665	1,000	0,348	0,994	0,584	0,947	0,429	0,667
IRG+MSAVI+B4	1,000	0,564	0,913	0,597	0,970	0,661	0,987	0,464	0,985	0,609	0,948	0,522	0,694
IRG+MSAVI+B5	1,000	0,575	0,913	0,658	0,974	0,742	0,987	0,429	0,987	0,648	0,948	0,519	0,711
IRG+MSAVI+B6	1,000	0,641	0,906	0,677	0,987	0,721	0,987	0,541	0,994	0,679	0,945	0,601	0,745
IRG+MSAVI+B7	1,000	0,588	0,928	0,683	0,983	0,734	1,000	0,472	0,991	0,653	0,963	0,558	0,730
IRG+MSAVI+B8	1,000	0,600	0,939	0,714	0,983	0,837	0,987	0,408	0,991	0,699	0,962	0,519	0,738
IRG+MSAVI+HS	0,983	0,617	0,968	0,824	1,000	0,863	0,918	0,524	0,991	0,719	0,943	0,640	0,768

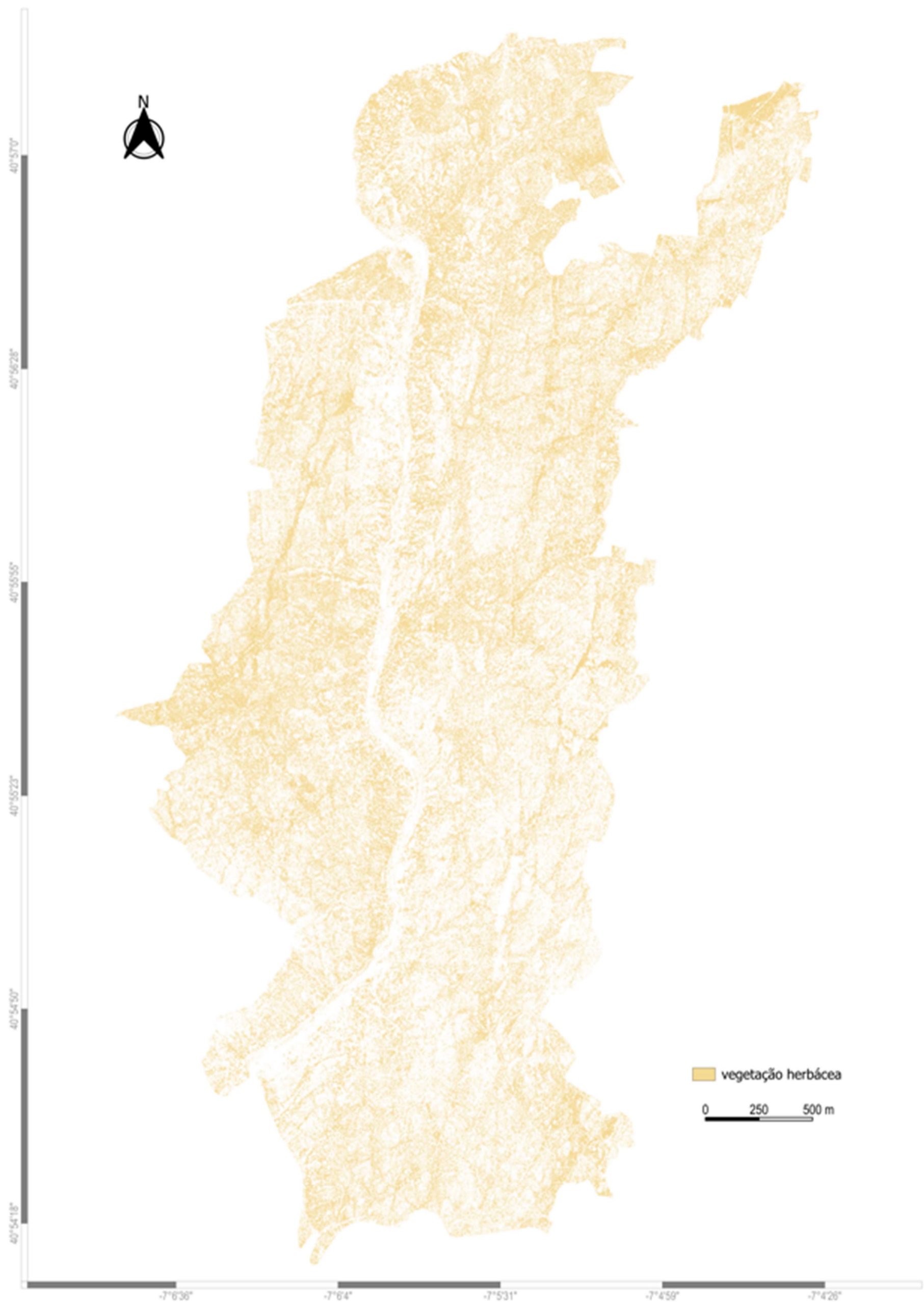
Anexo II -Mapa final de classificação da Reserva da Faia Brava. O mapa representa a distribuição espacial das quatro classes de ocupação do solo: árvores e arbustos, vegetação herbácea, afloramentos rochosos e Outras ocupações e usos do solo. A classificação foi realizada com o algoritmo Random Forest aplicado ao ortofotomosaico RGB, combinado com o índice GLI e a métrica de textura Entropia. A grelha UTM de 500 metros foi sobreposta para facilitar a referência espacial e análise por quadrícula.



Anexo 12 - Distribuição espacial da classe "árvores e arbustos" na Reserva da Faia Brava. Mapa obtido a partir do modelo final de classificação, destacando exclusivamente a mancha de vegetação arbórea e arbustiva identificada no território.



Anexo 13 - Distribuição espacial da classe "vegetação herbácea" na Reserva da Faia Brava. Mapa resultante da classificação final, evidenciando as áreas dominadas por vegetação herbácea no território.



Anexo 14 - Distribuição espacial da classe "afloramentos rochosos" na Reserva da Faia Brava. Representação isolada desta classe, conforme identificada no modelo final de classificação, evidenciando a sua concentração nas zonas de relevo mais exposto ou pedregoso.



Anexo 15 - Distribuição espacial da classe "Outras Ocupações e usos do Solo" na área de estudo. Esta categoria agrupa vários elementos, incluindo caminhos, solo nu, construções, corpos de água e zonas de sombra. A sua representação permite observar a localização e extensão destes elementos no território.

