

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE ENGENHARIA GEOGRÁFICA, GEOFÍSICA E ENERGIA



**Avaliação de metodologias de aprendizagem automática na
classificação de culturas agrícolas com base em imagens do
Sentinel-2**

Inês de Carvalho Fernandes Martins da Silva

Mestrado em Sistemas de Informação Geográfica – Tecnologias e Aplicações

Trabalho de projeto orientado por:
Prof. Doutor João Catalão Fernandes
Prof.^a Doutora Ana Navarro

Agradecimentos

Para a realização deste trabalho contei com o indispensável apoio, de forma mais ou menos direta, de pessoas e de entidades, que aproveito para mencionar aqui.

Em primeiro lugar, os meus agradecimentos ao Professor Doutor João Catalão Fernandes pela confiança, pelo entusiasmo e qualidade das discussões, pela disponibilidade e compreensão; à Professora Doutra Ana Navarro, pela partilha de conhecimento, pela precisão das orientações e críticas construtivas e pela qualidade do humor.

Ao Instituto de Financiamento da Agricultura e Pescas e à Fundação para a Ciência e Tecnologia pela oportunidade de me graduar através da participação num projeto real, com problemáticas concretas e implicações práticas.

À Professora Doutora Cristina Catita e Professora Doutora Paula Redweick pelo empenho, generosidade e boa disposição dentro e fora do contexto de aulas.

Aos meus companheiros de jornada Cláudia Belém, Christoph Bierer, João Duarte e Tomás Machado, aos companheiros de estágio de verão Bernardo Morais, Daniela Carrapiço, João Loureiro, Pedro Reis e Madalena Gil e companheiros de ‘gabinete’ Cláudia Silva, Carolina Rocha e Miguel Santos que fazem do trabalho um prazer e que me ajudaram todos os dias a fazer aquilo que mais gosto - aprender mais e ligar pontos.

Às amigas de sempre, Cátia Lopes e Maria João Falcão, pela presença e afeto.

À família que já me segue incondicionalmente há mais tempo, os Silvas, pais e irmã; à família mais recente, os Falcões, pelo apoio, entusiasmo, bom humor e bons conselhos; e ao companheiro Alexandre, que sabe sempre onde anda a generosidade, a paciência, a resiliência e o otimismo. Finalmente, um grande obrigada à pequena B por me ajudar a perspetivar o presente e o futuro de forma única, mesmo ainda sem estar cá fora.

Resumo

O elevado potencial das imagens de satélite abre novas possibilidades à recolha de informação para o desenvolvimento de produtos úteis ao mapeamento e monitorização agrícola que, a par do desenvolvimento de metodologias de análise de grandes volumes de dados, são instrumentos fundamentais na implementação de uma política agrícola comum (PAC) a nível europeu assente na automatização de processos de monitorização. O presente estudo tem como objetivo testar várias metodologias e algoritmos de aprendizagem automática para classificação de imagem com vista a construir modelos de classificação de culturas agrícolas e potenciar a transferência da aprendizagem que o classificador ganha em treino para permitir a sua aplicação em dados novos, atualizados anualmente. Usaram-se 84 imagens do satélite Sentinel-2 para seguimento do ciclo anual agrícola das culturas de verão em Portugal, entre Março e Setembro dos anos de 2018 e 2019, de três zonas geográficas distintas. Destas, extraíram-se as variáveis para a caracterização do perfil de cada parcela agrícola: os valores de refletância de cada banda e alguns marcadores fenológicos. Os dados foram processados e introduzidos em algoritmos de classificação supervisionada, para posterior comparação. De seguida, adotando o algoritmo que devolveu melhores resultados (*Random Forest*, RF) foram realizadas diversas classificações e análises de desempenho seguindo diferentes estratégias de amostragem. Os modelos de classificação com o RF mais bem sucedidos mostraram classificações com uma exatidão global (EG) entre os 78% e os 90%. Os modelos de classificação gerados com dados treino de 2018 permitiram classificar dados de qualquer zona do ano subsequente com uma EG superior a 77%. Houve uma maior degradação na EG da classificação na transferência do conhecimento entre regiões no mesmo ano ou em anos distintos do que na mesma região entre anos diferentes. Os resultados obtidos permitiram compreender os mecanismos de classificação de 37 classes de culturas agrícolas e delinear estratégias de amostragem baseadas em técnicas de transferência da aprendizagem para melhorar os modelos de classificação e possibilitar a sua aplicação entre anos e/ou zonas geográficas: a incorporação anual de dados de treino e a amostragem seletiva de parcelas corretamente classificadas (técnica de adaptação de domínio) podem ser estratégias a considerar futuramente para a melhoria dos modelos já desenvolvidos. O presente trabalho está integrado no projeto ‘Desenvolvimento de trabalho técnico-científico de apoio à implementação de um sistema de alerta para a monitorização da atividade agrícola com base em imagens de satélite e inteligência artificial (SAMAS-IA)’, do Instituto de Financiamento da Agricultura e Pescas, I.P. (IFAP) em parceria com a Faculdade de Ciências da Universidade de Lisboa.

Palavras-chave: aprendizagem automática, monitorização agrícola, mapeamento do tipo de cultura, Sentinel-2, classificação de imagem, transferência da aprendizagem, classificação supervisionada, *Random Forest*

Abstract

The potential of satellite imagery opens up new possibilities for acquiring information and product development useful for agricultural mapping and monitoring. Together with the development of methodologies for analyzing large volumes of data, these are important instruments in the implementation of a European Common Agricultural Policy (CAP) based on the automation of monitoring processes. The present study aims to test various methodologies and machine learning algorithms for image classification in order to develop crop classification models and use the transfer learning capabilities to its application in new data, updated annually. 84 images from the Sentinel-2 satellite were used to monitor the annual agricultural cycle of temporary summer crops in Portugal, between March and September 2018 and 2019, from three different regions. Each agricultural parcel profile was characterized using the extracted variables of: multispectral bands reflectance values and phenological markers. Data was processed and subjected to supervised classification algorithms for comparison, followed by classification and performance analysis using different sampling strategies with the method that returned the best results: *Random Forest, RF*. The most successful classification models showed classifications with overall accuracy (OA) values between 78% and 90%. The classification models built with training data from 2018 were able to classify data from any region in the following year with an OA greater than 77%. There was a greater decrease in the OA of the classification when transfer learning in between regions of the same year (or of different years) happened than in the same region between different years. The results allowed the understanding of classification mechanisms of 37 agricultural classes and outline sampling strategies based on transfer learning techniques to improve the models and enable their application between years and / or geographical regions. The incorporation of training data on a year basis and the selective sampling of correctly classified objects (domain adaptation technique) is a strategy to consider for the improvement of the models already created. This study is part of the project ‘Development of technical-scientific work to support the implementation of an alert system for monitoring agricultural activity based on satellite imagery and artificial intelligence (SAMAS-IA)’, from the Fisheries and Agriculture Financing Institute (IFAP) in partnership with the Faculty of Sciences of the University of Lisbon.

Keywords: machine learning, agriculture monitoring, crop type mapping, Sentinel-2, image classification, transfer learning, supervised classification, *Random Forest*

Índice

Agradecimentos.....	ii
Resumo.....	iii
Abstract	iv
Índice.....	v
Índice de Tabelas.....	vii
Índice de Figuras	ix
1 Introdução.....	1
1.1 Enquadramento.....	1
1.2 Motivação e Objetivos.....	4
1.3 Estrutura do trabalho	4
2 Estado da Arte	5
2.1 Estudos de referência desenvolvidos no âmbito do projeto Sen2Agri	5
2.2 Aprendizagem Automática e Inteligência Artificial.....	7
2.2.1 Tarefas de aprendizagem automática	8
2.2.2 Algoritmos de classificação.....	9
2.2.3 Transferência da aprendizagem em classificação.....	13
2.2.4 Transferência da aprendizagem em classificação de imagens de detecção remota - adaptação de domínio.....	14
2.2.5 Transferência da aprendizagem em classificação de imagens de detecção remota - uso de mapas temáticos de cobertura do solo	16
3 Dados e Métodos.....	17
3.1 Área de estudo.....	17
3.2 Dados.....	18
3.2.1 Dados <i>in situ</i> : Sistema de Identificação Parcelar (SIP).....	18
3.2.2 Dados de Detecção Remota: Imagens Sentinel-2.....	19
3.3 Metodologia.....	21
3.3.1 Preparação dos Dados	22
3.3.2 Pré-processamento dos dados.....	23
3.3.3 Dados Treino	25
3.3.4 Classificação: algoritmos e abordagem	27
3.3.5 Pós-Classificação.....	29
3.3.6 Melhoria dos modelos de classificação	31
3.3.7 Melhoria do conjunto de dados de treino: sobreamostragem e reamostragem seletiva.	32
3.4 Quadro resumo dos testes realizados.....	33
4 Resultados e Discussão	35

4.1	T01 – Classificação com diversos algoritmos e calibração do classificador sem reamostragem temporal	35
4.1.1	Análise das Matrizes de Confusão	36
4.2	T02 – Interpolação linear e reamostragem da série temporal.....	39
4.2.1	Análise das Matrizes de Confusão	40
4.2.2	Importância das variáveis para cada ano	43
4.3	T03 – Redução da dimensionalidade dos dados por PCA e Seleção de Variáveis (com interpolação).....	45
4.3.1	Análise de componentes principais (PCA).....	45
4.3.2	Seleção de Variáveis	46
4.4	T04 - Classificação em dados conjuntos (em diferentes proporções)	47
4.5	T05 – Transferência da aprendizagem no espaço com refletâncias de bandas.....	48
4.6	T06 – Transferência da aprendizagem no tempo com refletâncias de bandas	49
4.7	T07 – Transferência da aprendizagem no espaço e no tempo com refletâncias de bandas...	50
4.8	T08 – Transferência da aprendizagem no tempo: treino com dados de todas as zonas em diferentes proporções, validação em dados de 2019	50
4.9	T09 – Redução da dimensionalidade dos dados por marcadores fenológicos.....	51
4.9.1	Análise das Matrizes de Confusão	52
4.9.2	Importância de variáveis.....	53
4.10	T10 - Classificação por etapas do ciclo vegetativo	54
4.11	Mapas de confiança	55
4.11.1	Inspeção visual dos mapas de classificação	58
4.12	Análise global de resultados – melhores modelos e estratégias	61
5	Conclusões	64
	Referências Bibliográficas	66
	Anexos.....	70

Índice de Tabelas

Tabela 3.1- Número de parcelas agrícolas declaradas por região e por ano.....	18
Tabela 3.2 - Imagens Sentinel-2 no período do verão com uma percentagem de nuvens inferior a 10% nos dois anos em estudo (2018 e 2019) por região (NE, ND e NC).	20
Tabela 3.3 - Estrutura dos conjuntos de dados originais.	20
Tabela 3.4 - <i>Software</i> utilizado neste estudo.	22
Tabela 3.5 - Índices de vegetação selecionados e respetivas referências.	25
Tabela 3.6 - Valores dos principais hiperparâmetros de cada algoritmo de classificação usados.....	27
Tabela 3.7 – Valores dos hiperparâmetros do classificador RF usados e respetiva fonte.	31
Tabela 3.8 – Quadro resumo dos testes de classificação realizados.....	33
Tabela 4.1 – Métricas de desempenho e tempos de processamento por iteração dos modelos de classificação com diferentes algoritmos e conjuntos de dados: <i>Support Vector Machine</i> (SVM), <i>Multi-layer Perceptron</i> (MLP), <i>K-Nearest Neighbour</i> (kNN) e <i>Random Forest</i> (RF).	35
Tabela 4.2 – Métricas de desempenho dos modelos de classificação usando <i>Random Forest</i> (RF) com diferentes valores de parâmetros.	36
Tabela 4.3 – Matriz de confusão para a classificação realizada em dados de 2018, zona ND. Colunas representam a classificação e linhas representam os dados de referência. Os valores estão em absoluto, exceto quando indicado por %. A diagonal representa o número de parcelas classificadas corretamente pelo modelo.	37
Tabela 4.4 – Matriz de confusão para a classificação realizada em dados de 2019 da zona ND. Colunas representam a classificação e linhas representam os dados de referência. Os valores estão em absoluto, exceto quando indicado por %. A diagonal representa o número de parcelas classificadas corretamente pelo modelo.	37
Tabela 4.5 – Métricas de desempenho dos modelos de classificação para os dados originais e interpolados linearmente a 5 dias.	40
Tabela 4.6 – Matriz de confusão para a classificação de dados de 2018 de todas as zonas, com dados interpolados (a 5 dias). Colunas representam a classificação e linhas representam os dados de referência. Os valores estão em absoluto, exceto quando indicado por %. A diagonal representa o número de parcelas classificadas corretamente pelo modelo.	40
Tabela 4.7 – Matriz de confusão para a classificação de dados de 2019 de todas as zonas, com dados interpolados (a 5 dias). Colunas representam a classificação e linhas representam os dados de referência. Os valores estão em absoluto, exceto quando indicado por %. A diagonal representa o número de parcelas classificadas corretamente pelo modelo.	41
Tabela 4.8 – Resultados de EG dos modelos de classificação (com RF) desenvolvidos usando os dados originais interpolados a 5 dias e reduzidos por PCA.....	45
Tabela 4.9 – Resultados de EG dos modelos de classificação desenvolvidos usando os dados originais interpolados a 5 dias e reduzidos por Seleção de Variáveis.	46
Tabela 4.10 – Variáveis selecionadas para treino de modelos de classificação por Seleção de Variáveis para todos os conjuntos de dados interpolados a 5 dias.	47
Tabela 4.11 - Dimensão dos dados treino resultantes de combinação em proporções de 30%, 50% e 70% dos dados originais.....	47
Tabela 4.12 – Métricas de desempenho dos modelos de classificação para os dados interpolados com a totalidade das amostras para todas as zonas e para os dados das zonas concatenados em diferentes proporções.	48
Tabela 4.13 - Métricas de desempenho de classificadores treinados numa zona e testados noutra (para 2018 e para 2019).	48
Tabela 4.14 - Métricas de desempenho de classificadores treinados em 2018 e testados em 2019.	49
Tabela 4.15 - Métricas de desempenho de classificadores treinados em anos e zonas diferentes.....	50
Tabela 4.16 - Métricas de desempenho de classificadores treinados em 2018 (apenas parte da totalidade dos dados: 30%, 50% ou 70%) e testados em 2019 (totalidade dos dados).	51

Tabela 4.17 – Métricas de desempenho de classificadores desenvolvidos usando os dados originais interpolados a 5 dias (valores de refletância de bandas) e reduzidos por Marcadores Fenológicos.....	51
Tabela 4.18 - F1-score e EG obtida nos meses de início, meio e final da estação de crescimento das culturas temporárias de verão em todas as zonas para cada ano.	55
Tabela 4.19 – Percentagem do total de parcelas de 2019 com confianças de classificação superiores a 50% e 90% em modelos de classificação sem TA e com TA.	55

Índice de Figuras

Figura 3.1 – Localização da Área de Estudo mostrando a quadrícula UTM dos produtos Sentinel-2 usados neste estudo.	17
Figura 3.2 - Representatividade das classes nos conjuntos de dados.	19
Figura 3.3 - Fluxograma das diversas fases do trabalho de projeto.	21
Figura 3.4 – Diagrama de trabalho para pré-processamento, construção do modelo de classificação treinado em dados combinados das três zonas de um ano na proporção de 30% (cada zona) e teste em dados do mesmo ano (numa zona diferente) ou no ano seguinte.	26
Figura 3.5 – Procedimento de validação dos modelos de classificação – <i>Holdout Method</i> [57].	28
Figura 4.1 - Precisão vs revocação para as classes mais representadas em cada amostra teste de 2018 e 2019 (zona ND). A dimensão dos círculos ilustra a frequência de cada classe nos dados.	38
Figura 4.2 – Importância relativa das variáveis na classificação com RF para dados de 2018 (a azul) e de 2019 (a laranja), zona ND. As denominações das variáveis dizem respeito à banda, mês e dia de aquisição da imagem, respectivamente.	39
Figura 4.3 – Precisão vs revocação para as classes mais representadas em cada amostra teste de 2018 e 2019 (todas as zonas). A dimensão dos círculos ilustra a frequência de cada classe nos dados.	42
Figura 4.4 - Importância relativa das variáveis na classificação com RF para dados de 2018 e de 2019, todas as zonas.	43
Figura 4.5 - Importância relativa das variáveis na classificação com RF para dados de 2018 e de 2019, todas as zonas.	44
Figura 4.6 – Gráfico de refletâncias das bandas usadas para a identificação das culturas mais representativas em análise.	44
Figura 4.7 – Gráficos das refletâncias das bandas de maior importância na classificação para o conjunto total de dados, para as culturas agrícolas mais representativas.	45
Figura 4.8 – Precisão vs revocação para as classes mais representadas em cada amostra teste de 2018 (todas as zonas). A dimensão dos círculos ilustra a frequência de cada classe nos dados.	52
Figura 4.9 – Precisão vs revocação para as classes mais representadas em cada amostra teste de 2019 (todas as zonas). A dimensão dos círculos ilustra a frequência de cada classe nos dados.	53
Figura 4.10 – Importância das variáveis usadas na classificação dos dados para os dois anos (2018 e 2019).	53
Figura 4.11 - EG dos modelos treinados com dados ao longo da época de crescimento da vegetação para 2018 e para 2019 (todas as zonas).	54
Figura 4.12 - Mapa de culturas agrícolas para as parcelas da zona ND, ano de 2019 usando a metodologia proposta (T02) (EG = 90.9%).	56
Figura 4.13 – De cima para baixo: imagem das zonas NE, ND e NC de 2019. Da esquerda para a direita: cor verdadeira (RGB: 4.3.2), falsa cor (RGB: 8.4.3), mapa de confiança para modelo de classificação sem TA (T02) e mapa de confiança para modelo de classificação com TA no tempo (T06). Culturas não analisadas a cinzento.	57
Figura 4.14 – Distribuição de confianças de classificação para modelos de classificação sem TA (T02) e com TA no tempo (T06) para classificação em dados de 2019.	58
Figura 4.15 - Mapas de confiança (2019 zona ND, para modelo sem TA (T02). Os valores iguais a 1 são omitidos; valores inferiores a 0.5 estão representados a vermelho.	58
Figura 4.16 - Comparação de mapa de culturas obtidos usando um modelo treinado em dados de 2019 (zona ND - T02 – Interpolação linear e reamostragem da série temporal) com a informação do Sistema de Identificação de Parcelas (SIP). Os números indicam o tipo de cultura dos dados de referência das classes incorretamente classificadas e as cores correspondem ao resultado do classificador.	59

Figura 4.17 - Mapa de confiança (2019 zona ND) para as mesmas parcelas com modelo: sem TA (T02, em cima) e com TA no tempo (T06, em baixo) (apontador azul indica parcelas incorretamente classificadas)	60
Figura 4.18 - Mapa de confiança (2019 zona ND, para modelo sem TA (T02) e com TA (T06). Os valores iguais a 1 são omitidos; valores inferiores a 0.5 estão representados a vermelho.	61
Figura 4.19 - Resultados de exatidão global (EG) para os testes de classificação realizados neste estudo.	62
Figura 4.20 - EG média para modelos de classificação que operaram sem TA (T02) e com TA (T05, T06, T07 e T08).	63

1 Introdução

1.1 Enquadramento

‘Even though food production has doubled over the last 30 years, it will have to double yet again by 2025 if it is to feed an expected population of 8.5 billion. From recent statistics this seems unlikely, as in 2000 world agricultural production grew by only 1%, the lowest since the early 1990s.’

[1]

A produção alimentar é a motivação de muitas atividades humanas e, direta ou indiretamente, o motor no estabelecimento de prioridades políticas, no desenvolvimento de estratégias de ação e de agendas de investigação de inúmeras organizações. Atualmente, a atividade agrícola mundial depara-se com desafios que dizem respeito, não só, à produção direta de bens alimentares como também aos fatores de contexto que a influenciam, nomeadamente, o impacto das alterações climáticas e consequências ao nível dos ecossistemas e biodiversidade, da qualidade dos solos e da água.

A nível europeu, apesar da agricultura ter desenvolvido práticas, métodos e técnicas que contribuíram para a melhoria da produção alimentar face aos custos de produção, há a necessidade de aumentar essa produção nos anos mais próximos, sendo premente o desenvolvimento de modelos de produção cada vez mais responsivos, em que a produção agrícola seja independente da dimensão da exploração e que respeitem e trabalhem simultaneamente com o contexto ambiental, social e político [2].

Para os sucessos atingidos ao longo da evolução nas práticas agrícolas até à atualidade, muito contribuiu a tecnologia espacial: a agricultura de precisão, os sistemas de navegação por satélite e de observação da Terra são de grande utilidade em muitas explorações agrícolas da União Europeia (UE) e prevê-se a sua proliferação em todo o setor. Concretamente, no sentido da execução da política agrícola comum (PAC) focada na autossuficiência alimentar da UE, a informação obtida pelas técnicas de observação da Terra potencia a monitorização das políticas, por exemplo, na atribuição de subsídios e na agricultura inteligente e ambientalmente responsável. A proposta de regulamento do Parlamento Europeu e do Conselho que cria o programa espacial da União e a Agência da União Europeia para o Programa Espacial sugere a atribuição de recursos orçamentais no valor de 16 mil milhões de euros para ‘prosseguir e melhorar os sistemas Galileo, EGNOS [Serviço Europeu Complementar de Navegação Geoestacionário], Copernicus e SST [*Space Surveillance and Tracking*], bem como para lançar a iniciativa Govsatcom [*Governmental Satellite Communications*].’, sendo mais de dois mil milhões de Euros alocados ao programa Copernicus [2]. Pretende-se que estes recursos sejam reinvestidos no que a Comissão Europeia considerou prioritário: a realização de ‘observações com elevada resolução espaço-temporal’ e de ‘observações óticas com capacidades de formação de imagens hiperespectrais a fim de complementar e expandir as atuais medições através do Sentinel-2’ (*ibid.*). O elevado potencial das imagens de satélite na monitorização de práticas e gestão de recursos traduz-se, por exemplo, na distribuição aberta e gratuita das imagens dos satélites Sentinel, que indiretamente auxiliam à tomada de decisão sustentada de autoridades

políticas e públicas. É especificamente recomendado pela Comissão da Agricultura e do Desenvolvimento Rural (Parlamento Europeu) que se reforce o potencial do programa Copernicus na atividade agrícola, por exemplo, no que respeita ao uso e previsão de rendimento de colheitas das explorações agrícolas, gestão da irrigação, à ‘cartografia sazonal das superfícies cultivadas e aos mapas dos tipos de culturas, à ocupação dos solos’ [2] entre outros aspetos, fazendo-se uso dos desenvolvimentos ao nível da ‘análise dos grandes volumes de dados e das tecnologias de inteligência artificial’ (*ibid.*). Os dados dos projetos Sen2Agri (<http://www.esa-sen2agri.org/>) e Sen4CAP (<http://esa-sen4cap.org/>), os dois projetos mais recentes da Agência Espacial Europeia (*European Space Agency*, ESA) de apoio à implementação da PAC, desempenham um papel fundamental neste âmbito, nomeadamente na passagem ‘de uma abordagem de monitorização da PAC baseada em amostras para uma abordagem sistemática’ (*ibid.*) e uma consequente redução de recursos pela substituição dos controlos no terreno. O mapeamento do uso e cobertura do solo com recurso às tecnologias de deteção remota para monitorização dos tipos de ocupação do solo é particularmente importante no alcance deste objetivo. Assume-se que uma elevada exatidão em modelos de classificação de aprendizagem automática para uma dada cultura agrícola/cobertura do solo dão uma elevada probabilidade da sua deteção remotamente, no processo de monitorização [3].

No seguimento das missões Landsat e SPOT, a constelação Sentinel-2 e o seu Instrumento Multiespectral (*Multispectral Instrument*, MSI) medem a refletância solar espectral em 13 bandas, desde o visível e o infravermelho próximo (NIR) ao infravermelho de onda curta (SWIR). As bandas espectrais incluem três bandas para monitorização de nuvens e correção atmosférica a 60m de resolução espacial; seis bandas a 20m, quatro no domínio do limiar do vermelho (*Red Edge*) e duas bandas SWIR que fornecem informação indispensável sobre neve/gelo/nuvens e humidade da vegetação e as quatro bandas do azul, verde, vermelho e infravermelho próximo a 10m para aplicações terrestres [4]. Um período de revisita de 5 dias é assegurado por dois dos satélites da constelação (Sentinel 2A e 2B) desde 2017 e até, pelo menos, 2030 [5]. Para assegurar a continuidade no fornecimento de dados, prevê-se ainda o lançamento dos satélites óticos Sentinel-2C e 2D até 2020/2021 [6]. A elevada largura de varrimento (*swath width*) do Sentinel-2, o período de revisita de 5 dias, a elevada qualidade de imagem e as políticas de acesso livre e aberto aos produtos do Sentinel-2 abrem novas possibilidades à recolha de informação para o desenvolvimento de produtos úteis ao mapeamento e monitorização agrícola e ao delineamento de novos desafios no que respeita à monitorização da atividade agrícola a nível mundial. Os índices de vegetação e outras variáveis biofísicas obtidas por deteção remota são já usadas globalmente em sistemas de alerta nas mais diversas instituições, embora a monitorização ao nível do tipo de cultura agrícola ainda não se faça de forma sistemática [5].

Sen2Agri designa o sistema, baseado em algoritmos de aprendizagem automática, do Sentinel-2 desenhado para gerar automaticamente produtos necessários à monitorização da atividade agrícola a partir da ingestão sistemática de séries temporais dos satélites Sentinel-2 e Landsat-8 ao longo do ciclo agrícola: refletância de superfície sem nuvens, máscara com a dinâmica das culturas agrícolas, mapas do tipo de cultura agrícola e extensão da área para os principais grupos de culturas, fornecidos duas vezes ao longo do ciclo fenológico de cada cultura (no meio e no final do ciclo) e indicadores do estágio de desenvolvimento fenológico (métricas que informam sobre parâmetros-chave no ciclo de crescimento: data início, duração e data em que é atingido o crescimento máximo), nomeadamente o índice de vegetação da diferença normalizada

(*Normalized Difference Vegetation Index*, NDVI) e o índice de área foliar (*Leaf Area Index*, LAI), descrevendo o desenvolvimento vegetativo [7].

O projeto Sen4CAP (Sentinel para a Política Agrícola Comum) está ainda em fase de desenvolvimento e ambiciona fornecer aos agentes, Europeus e nacionais, algoritmos validados (e código para sua implementação) e promover a sua transferência para os institutos públicos nacionais, nomeadamente ao nível de produtos, metodologias e boas práticas para a monitorização agrícola que sejam relevantes para a gestão da PAC. O projeto focar-se-á na informação fornecida pelo Sentinel no apoio à modernização e simplificação das PAC pós-2020 [8].

Com a reforma da PAC, a implementar em 2021 (e a vigorar até 2027), pretende-se simplificar e modernizar processos, quer para a administração quer para o agricultor. Das mais relevantes fontes de informação, para conhecimento da atividade agrícola e sua validação no âmbito da informação geográfica destacam-se: a deteção remota – imagens de satélite que fornecem informação sobre a ocupação do solo para efeitos de monitorização; e o Sistema de Identificação do Parcelário – informação georreferenciada das parcelas agrícolas [9].

Por meio de deteção remota e de algoritmos de inteligência artificial (aprendizagem automática), pretende-se identificar as alterações no solo e validar a informação existente nos sistemas de identificação parcelar, e das operações agrícolas reportadas pelo agricultor, substituindo-se o atual sistema de controlo *in situ*. Conjugando a informação do limite das parcelas existentes no Sistema de Identificação do Parcelário, com os mapas de marcadores fenológicos das diversas culturas agrícolas podem-se obter estatisticamente valores que permitem caracterizar a parcela: identificar culturas temporárias/permanentes por padrões característicos e operações agrícolas (por exemplo, sementeira ou colheita) [9]. A realização destes controlos assegurará o pagamento das ajudas no mais curto prazo possível aos beneficiários que cumpram os requisitos estabelecidos pela regulamentação.

A nível nacional, é o Instituto de Financiamento da Agricultura e Pescas, I.P (IFAP) o principal responsável por ‘garantir o funcionamento dos sistemas de apoio e de ajudas diretas nacionais e comunitárias e a aplicação, a nível nacional, das regras comuns para os regimes de apoio direto no âmbito da PAC’ e ‘executar a política estratégica na área das tecnologias de informação e comunicação, para o setor da agricultura e pescas’ [10] entre outras atribuições.

Para isso, dispõe de ferramentas como o Sistema Integrado de Gestão e Controlo (SIGC) para monitorizar, administrativa e localmente, os pedidos de ajuda e ‘assegurar a verificação eficaz do respeito dos requisitos da concessão das ajudas’. O maior esforço a nível do controlo está relacionado com a identificação do tipo de cultura da parcela e da área elegível da parcela [11] para o que existe, entre outras aplicações, o Sistema de Identificação Parcelar (SIP) - sistema de informação geográfica que pretende caracterizar as explorações agrícolas. Tal como regulamentado pela PAC, o SIGC assenta i) no controlo físico no terreno e ii) no controlo por deteção remota, realizado por operadores que ‘visualizam as parcelas agrícolas sobre uma sequência multitemporal de imagens de satélite de alta (HR) e muito alta resolução (VHR) no sentido de validar o tipo de cultura e a área elegível’ (*idem*). Este é, portanto, um processo moroso, dispendioso e falível pelos recursos que emprega e pela subjetividade inerente à fotointerpretação. Para além disso, apenas 5% das parcelas declaradas são validadas por esta via. No sentido de tornar o processo de controlo mais eficiente do ponto de vista logístico mas também económico, geograficamente abrangente, robusto e automático, é intenção do IFAP

implementar estratégias informáticas que reforcem a sua capacidade de monitorizar de forma automática o que é declarado pelos agricultores e que permitam um melhor serviço público (com diminuição de penalizações por fiscalização incorreta das culturas, diminuição do tempo de operação e aumento do impacto ao nível da Comunidade Europeia no que respeita à prestação de serviços).

1.2 Motivação e Objetivos

O objetivo do presente trabalho é testar várias metodologias e algoritmos de aprendizagem automática para classificação de imagem com vista a construir modelos de classificação de culturas agrícolas e potenciar a transferência da aprendizagem (TA) que o classificador ganha em treino para permitir a sua aplicação em dados novos, atualizados anualmente. Tendo como dados de referência os dados das declarações anuais dos agricultores, realizadas no âmbito dos pedidos de apoio financeiro concedido no âmbito da PAC através do IFAP, pretende-se contribuir para a criação de um sistema capaz de promover, com robustez de desempenho, a TA que o classificador ganha em treino com dados já existentes para permitir a generalização e aplicação do modelo em dados distintos daqueles que foram usados para treino. Esta TA entre modelos de classificação desenvolvidos com recurso a dados limitados, a um período temporal e/ou a um espaço geográfico, e sujeitos a contingências relacionadas com a disponibilidade de imagens do Sentinel-2, será essencial à constante melhoria dos sistemas automáticos que se pretendem operacionais. É neste aspeto que se centra o presente estudo.

O desenvolvimento de sistemas de alerta para monitorização da atividade agrícola com base em inteligência artificial implica a criação de modelos de classificação capazes de identificar e classificar conjuntos de píxeis com a maior exatidão possível, de modo a minimizar as confirmações diretamente no terreno. Implica também, o desenvolvimento de técnicas de TA adquiridas por modelos de classificação construídos com conjuntos de dados situados num determinado tempo e espaço de modo a que o seu desempenho se mantenha satisfatório em dados novos, provenientes de aquisições mais recentes, assegurando a sua robustez o mais independentemente possível das necessidades sentidas.

O trabalho foi desenvolvido no âmbito do estudo – ‘Desenvolvimento de trabalho técnico-científico de apoio à implementação de um sistema de alerta para monitorização da atividade agrícola com base em imagem de satélite e inteligência artificial (SAMAS-IA)’, financiado pelo IFAP – pretende contribuir para o desenvolvimento de metodologias de monitorização automática, que se tornará obrigatória em 2020 com a nova PAC, em substituição do controlo local por amostragem descrito anteriormente.

1.3 Estrutura do trabalho

O presente trabalho-projeto está dividido em cinco capítulos e organiza-se da seguinte forma: no primeiro capítulo faz-se a contextualização do tema do trabalho e são enunciados os principais objetivos; no segundo capítulo, é apresentado o Estado da Arte dos temas abordados, com referência ao que atualmente se desenvolve no âmbito do tema, sustentando-se essa informação em literatura e metodologias atualmente em desenvolvimento; no terceiro capítulo, introduz-se a área de estudo e os dados usados no projeto e dá-se conta da metodologia utilizada; no quarto capítulo, são apresentados os principais resultados do estudo e sua discussão seguindo-se, no

quinto capítulo, as principais conclusões do trabalho e delineamento de caminhos de investigação futuros.

2 Estado da Arte

O presente capítulo tem como objetivo abranger o maior número possível dos tópicos relevantes ao estudo conduzido neste projeto por forma a ser elucidativo, sem ser exaustivo. Inclui as diversas áreas do conhecimento exploradas no trabalho de projeto contextualizando vários temas de investigação que lançam as bases da problemática deste estudo

A pesquisa de que se dá conta neste capítulo foi feita a partir de trabalhos realizados por diversos autores no âmbito dos projetos Sen2Agri e Sen4CAP e foi-se aprofundando à medida das necessidades específicas do contexto e dos problemas que foram surgindo. Portanto, procurou-se a base teórica para a contextualização deste estudo sobretudo em trabalhos de investigação que serviram de suporte à escolha da metodologia utilizada. Por questões de maior clareza, procurou fazer-se uma exposição em tópicos, uma vez que as áreas de conhecimento abrangidas são muito diversas e independentes entre si.

O Estado da Arte inclui uma revisão da literatura sobre:

- Utilização de imagens de satélite no desenvolvimento de sistemas de monitorização no âmbito das políticas agrícolas da União Europeia (ajudas comunitárias e ações de controlo);
- Processamento e classificação de imagem, com enfoque em imagens obtidas por sistemas de deteção remota; e identificação de classes de cobertura do solo, em particular, de culturas agrícolas;
- Metodologias de aprendizagem automática e de inteligência artificial na construção e melhoramento de modelos de classificação (supervisionada).

2.1 Estudos de referência desenvolvidos no âmbito do projeto Sen2Agri

Considerando os sistemas de monitorização desenvolvidos no âmbito das políticas agrícolas da União Europeia e os projetos desenvolvidos pela ESA, importa realçar alguns trabalhos de suporte e de análise das metodologias desenvolvidas para o sistema Sen2Agri. O sistema mais recente Sen4CAP, ainda em desenvolvimento, está disponível apenas em versão Beta lançada em Junho de 2019 e dedicada unicamente às agências nacionais. Uma nova versão gratuita e de acesso aberto estará disponível em Março de 2020, pelo que ainda carece de estudos aprofundados deste tipo.

No sentido de desenvolver métodos de substituição de controlos no terreno (*on-the-spot-checks*) para permitir uma aplicação mais económica e ajustada da PAC, o Centro Comum de Investigação da Comissão Europeia (*Joint Research Center, JRC*) selecionou a plataforma *Tensorflow* pela sua efetividade na resolução de um amplo leque de problemas de aprendizagem automática, chamando a atenção para o carácter ‘caixa-negra’ das redes neuronais profundas (algoritmo base dos processos de classificação). Preconiza-se que estes métodos de controlo tenham por base o conceito de marcador, isto é, um registo estruturado de uma observação

objetiva de comportamento do solo tal como retratada pelos sinais do Sentinel [3]. O marcador descreve não só a observação objetiva no sinal, mas também as condições de contexto que dão significado a essas observações, através de um cerne (*core*) – o comportamento detetável pelo sinal (uma continuidade, uma mudança, uma transição) - e de parâmetros - restrições aplicadas ao cerne do marcador para introdução de especificidades do contexto de modo a ser eficiente na realidade a que se refere. Aconselha-se também que a monitorização do uso do solo se faça com base em variáveis de interesse (*feature of interest*, FOI), ou seja, a unidade de terreno que expressa o comportamento descrito pelo marcador. Portanto, as avaliações do terreno devem ser conduzidas em conjuntos de píxeis, delimitadores de parcelas agrícolas. O agrupamento em conjuntos de píxeis permite alargar o tipo de sinal disponibilizado pelo satélite pela adição de valores estatísticos aos dados originais e, assim, conferir uma maior homogeneidade à cobertura do solo e melhoria da qualidade do sinal pela remoção de observações anómalas (*outliers*) e ruído [12].

Também recentemente, Defourny *et al.* [5] testaram a aplicabilidade dos métodos usados na plataforma Sen2Agri a sistemas agrícolas distintos no sentido de fornecer, em tempo útil, informação rigorosa a uma resolução espacial de 10m sobre áreas nacionais. Para isso, testaram o uso do sistema em oito experiências a nível local e nacional, abrangendo culturas e práticas agrícolas na Ucrânia, Mali e África do Sul e em áreas de 300 x 300 km em outros cinco países. Um dos produtos gerados pelo sistema é uma máscara com a dinâmica das culturas agrícolas recorrendo a uma série temporal. Para tal, o sistema preenche as lacunas temporais (advenientes, por exemplo, da cobertura nebulosa, usando uma máscara de nuvens) existentes no dados de entrada e utiliza um algoritmo supervisionado, *Random Forest* (RF), para realizar a classificação das parcelas. O modo de implementação do RF depende da existência ou não de dados *in situ*: no caso de existirem dados *in situ*, a implementação do RF baseia-se na distribuição de variáveis estatísticas, derivadas a partir da série temporal do NDVI e de variáveis temporais, extraídas do índice de diferença normalizada da água (*Normalized Difference Water Index*, NDWI), da intensidade e do índice do limiar do vermelho (*Brightness*) para cada tipo de cultura; no caso de não existirem dados *in situ*, os dados de calibração são selecionados de um mapa de referência e pré-processados antes de serem usados pelo algoritmo RF. Neste caso, a construção do modelo de classificação deve basear-se em variáveis temporais correspondentes a estágios específicos do ciclo de crescimento das plantas: data dos declives máximo e mínimo da curva de NDVI; data dos valores máximo e mínimo do NDVI e data do valor máximo da banda do vermelho. No processo de conceção de um outro produto do Sen2Agri – os mapas de tipo de cultura agrícola – a tarefa de classificação é baseada no algoritmo RF ou no *Support Vector Machine* (SVM), sendo que o RF é reportado como tendo melhores resultados por autores como Inglada (2015) pelo que é o algoritmo do sistema por defeito [5]. Técnicas de interpolação linear e reamostragem são passos no pré-processamento dos dados e as variáveis extraídas para a subsequente classificação são os valores de refletância nas bandas 3, 4, 5, 6, 7, 8, 11 e 12, o NDVI, o NDWI e a intensidade para cada data da série temporal (já interpolada e reamostrada). Os autores enunciam ainda estratégias de recolha de dados *in situ* e de medição da exatidão dos modelos de classificação. Neste último ponto, a exatidão global e o F1-score são métricas aconselhadas, embora nada seja mencionado sobre a dimensão das amostras e o impacto da representatividade das classes nos resultados da classificação.

Relativamente ao desempenho do sistema Sen2Agri na execução de mapas de cultura agrícola, os autores concluem que este está muito dependente da disponibilidade de dados *in situ* para calibração do mesmo. Os resultados mostram também que há culturas particularmente difíceis de

discriminar entre si, em que as diferenças ocorrem apenas ao nível do calendário das culturas, levando a uma diminuição na exatidão dos modelos. Por último, e medindo o desempenho dos modelos ao longo da estação de crescimento das plantas, os autores verificaram que os valores de exatidão global de produtos obtidos no final do ciclo de crescimento são superiores a 80% e o F1-score entre 0.6 e 0.98 com a exceção de algumas culturas, mais ‘difíceis’ de classificar [5]. Estes valores são mais elevados do que os obtidos para produtos da meia-estação e mais acentuados para determinadas classes agrícolas de acordo com o seu calendário específico. A dimensão dos campos de cultivo também interfere com a qualidade da classificação (por exemplo, campos de 20 x 20 m são pouco ‘compatíveis’ com a resolução espacial de 10 m do MSI do Sentinel-2). Resumidamente, os autores concluem que o sistema Sen2Agri tem sido usado por agentes nacionais e é, de facto, relevante no suporte à monitorização agrícola [5].

Previamente, já Valero *et al.* [13] se focaram num dos quatro produtos do Sen2Agri para uso operacional – uma máscara com a dinâmica das culturas agrícolas que separa, concretamente, áreas de cultivo de terrenos não cultivados – para explorar diferentes cadeias de processamento de forma a construir esse produto, usando três estratégias de classificação. Estas, baseiam-se quer no píxel, quer no objeto (nesta última, acrescentando-se uma tarefa de filtragem antes ou depois da classificação propriamente dita, que implica a atribuição de um valor médio a um conjunto de píxeis pertencentes a um mesmo objeto). A ideia dos autores foi perceber até que ponto a consideração de relações temporais ou espaciais entre os píxeis da imagem pelas estratégias testadas tem influência na exatidão da sua classificação. As estratégias usadas pelos autores assentam sobre dois tipos de técnicas comuns em reconhecimento de imagem: a extração de variáveis e a classificação supervisionada. A extração de variáveis fez-se usando métricas anuais como variáveis independentes para o mapeamento, tendo como base o NDVI, tradicionalmente usado para monitorização de culturas, e métricas calculadas com base em curvas de NDVI, representativas de cada cultura e do respetivo ciclo fenológico. A classificação foi feita usando os algoritmos de base do Sen2Agri e conduzida em 12 locais de teste (África do Sul, Argentina, Bélgica, Burkina Faso, China, França, Madagáscar, Maricopa, Marrocos, Paquistão, Rússia e Ucrânia), diversos em contextos agro ecológicos, padrões de paisagem, práticas agrícolas e condições de observação do satélite. Os resultados foram promissores, com exatidões globais em torno dos 90% no final do ciclo agrícola [13].

2.2 Aprendizagem Automática e Inteligência Artificial

A inteligência artificial (IA), campo de conhecimento que tem recebido muita atenção por diversos setores sociais e económicos em anos recentes, nasceu enquanto campo de conhecimento em 1956, com muitas definições e campos de ação. Genericamente, segundo Hu *et al.* [14], a IA foca-se no estudo e arquitetura de máquinas e métodos automáticos que desempenhem tarefas normalmente desenvolvidas por seres humanos, tais como, o reconhecimento de objetos a partir de imagens ou o entendimento do significado de frases em linguagem natural. Três dos fatores que mais contribuíram para que a IA ganhasse popularidade recentemente foram a proliferação e o acesso relativamente democratizado a grandes volumes de dados (*big data*), e o desenvolvimento de novos algoritmos e de poder computacional. A montante, contribuíram, entre outros, a emergência de sensores e a utilização generalizada da *web* pelos utilizadores com geração de grandes volumes de dados a grande velocidade. O desenvolvimento de novos

algoritmos e modelos capazes de relacionar diferentes campos de conhecimento e ainda a computação de elevado desempenho fornecem o poder de ligação entre a grande disponibilidade de dados e a existência de modelos mais complexos, permitindo o treino de modelos sofisticados em enormes conjuntos de dados em poucas horas ou dias.

Dentro do conceito mais alargado de IA inclui-se a aprendizagem automática. Este, assenta em métodos estatísticos ou técnicas de otimização numérica para derivar modelos de dados sem programar explicitamente todo e qualquer parâmetro ou passo de computação. Um tipo específico de aprendizagem automática é a aprendizagem profunda (*deep learning*), que se centra no desenvolvimento e utilização de redes neuronais profundas em tarefas de aprendizagem automática. Estas são tipos de redes neuronais artificiais que contêm múltiplas camadas escondidas (*hidden layers*) entre as camadas de processamento de entrada (*input*) e de saída (*output*) (vide secção 2.2.2). Dois dos tipos de redes neuronais mais populares na comunidade científica são as redes neuronais convolucionais (*convolutional neural networks*) e as redes neuronais recorrentes (*recurrent neural network*) [14].

2.2.1 Tarefas de aprendizagem automática

As tarefas de aprendizagem automática são genericamente agrupadas em três categorias: a aprendizagem supervisionada, em que ao computador é dado o *input*-exemplo e o *output*-pretendido (o objetivo é aprender a regra geral que atribui a cada *input* o *output* fornecido), a aprendizagem não-supervisionada, em que o *input*-exemplo são os próprios dados (o objetivo é encontrar uma estrutura nos dados) e a aprendizagem de reforço, em que o computador interage com um ambiente dinâmico tendo que atingir um dado objetivo de desempenho, sendo fornecido ao programa *feedback* em forma de recompensa/penalização no processo de resolução do problema [14].

As tarefas de aprendizagem automática podem ser usadas com diferentes fins: classificação (classificar um alvo numa categoria), *clustering* (encontrar grupos em dados) e regressão (prever valores desconhecidos através de modelos de regressão). Outras tarefas como a deteção de anomalias, a geração de dados, a visualização, a aprendizagem de características (*feature learning*), entre outras, são também parte do domínio da aprendizagem automática. No domínio deste estudo, a finalidade da IA que mais interessa é a classificação (de imagens obtidas por deteção remota) – em que se espera que o algoritmo de aprendizagem em presença dos *inputs* de duas ou mais classes produza um modelo que atribua a *inputs* não vistos em treino, uma, duas ou mais dessas classes com uma certa probabilidade associada [14]. A redução da dimensionalidade – em que os *inputs* são simplificados pela sua realocação a um espaço de menor dimensionalidade [14] – é também particularmente importante aqui, dado o elevado número de variáveis dos dados utilizados.

Em classificação de imagem, a tarefa do classificador é aprender a partir de imagens com etiquetas de classe. Assim, há dois problemas fundamentais a ter em conta na criação de um bom modelo de classificação: i) encontrar um espaço de variáveis (*feature space*) suficientemente discriminativo das classes entre si através da sua distribuição probabilística, seja por seleção de variáveis ou por mapeamento de amostras num novo espaço de variáveis (ex. de técnicas como a Análise de Componentes Principais, *Principal Component Analysis*, PCA) ; ii) construir um classificador robusto que mapeie amostras do espaço de variáveis para o espaço das etiquetas da

classe. O classificador pode ser um modelo generativo que aprende a distribuição de probabilidades conjunta $P(X, Y)$ ou um modelo discriminativo que aprende a distribuição de probabilidades condicionais $P(Y|X)$ [15].

De entre os problemas mais comuns encontrados na construção dos modelos de classificação, está o custo das amostras de treino para o classificador, que pode, por um lado, incluir muito ruído e, por outro, ser dispendioso pelo tempo e recursos que a etiquetagem manual de um grande número de amostras implica. Mais adiante, encontram-se referências a alguns estudos em detecção remota que abordam esta limitação e propõem soluções.

2.2.2 Algoritmos de classificação

A aprendizagem automática pode ser resumida como o ato de aprender uma função que atribui a variáveis de *input* (x) determinadas variáveis de *output* (Y).

$$Y = f(x) \quad (2.1)$$

Um algoritmo aprende o alvo gerando uma função a partir de dados de treino. Sendo a forma da função desconhecida, o trabalho em aprendizagem automática passa por avaliar diferentes algoritmos e aferir qual o que melhor se aproxima da função subjacente. Diferentes algoritmos fazem diferentes suposições ou viés sobre a forma desta função e como pode ser aprendida. Estas suposições facilitam, por um lado, o processo de aprendizagem e limitam, por outro, o que pode ser aprendido. Por exemplo, se um algoritmo assume que a função tem a forma de uma linha (tal como acontece em regressão linear) representada por:

$$f(x) = b_0 + b_1x_1 + b_2x_2 \quad (2.2)$$

em que b_0 , b_1 e b_2 são os coeficientes da linha que controlam a interceção e declive da reta e x_1 e x_2 as duas variáveis de *input*, o processo de aprendizagem de um modelo preditivo resume-se à estimação dos coeficientes da equação. Se a função subjacente a ser aprendida não tiver uma forma linear, o modelo terá um desempenho fraco e conduzirá a estimativas pouco rigorosas.

Genericamente, as técnicas de classificação distribuem-se por modelos paramétricos e não paramétricos. Um classificador que assume que os dados para cada classe individual estão distribuídos normalmente é um modelo paramétrico. Modelos não paramétricos não fazem suposições sobre a natureza estatística dos dados e usam os exemplos treino para prever o próximo exemplo – o número de parâmetros cresce com o número de exemplos disponíveis [16].

No domínio da classificação de imagem obtida por detecção remota, em particular, existem numerosas publicações sobre a aplicação de técnicas de aprendizagem automática que exploram os principais desafios à aplicação de ferramentas de classificação, as vantagens/desvantagens, níveis de desempenho e comportamento de diversos algoritmos usados na análise de dados provenientes de imagens de satélite. A elevada resolução espectral dos sensores usados na recolha de dados resulta em centenas de canais de observação. Estes, possibilitam desenvolver aplicações que requerem elevado poder discriminatório no domínio espectral mas trazem desafios metodológicos à análise automática dos dados devido a: i) elevada variabilidade espacial da assinatura espectral de cada classe de cobertura do solo; ii) efeitos atmosféricos e iii) a ‘maldição da dimensionalidade’ (*curse of dimensionality*). Em classificação supervisionada em particular, o principal problema relaciona-se com a proporção entre o número de amostras de treino e o

número de variáveis que impossibilita a obtenção de estimativas usadas em classificadores estatísticos. Com o aumento do número de variáveis dadas como *input* ao classificador dentro de um dado limite (dependente do número de amostras de treino e do tipo de classificador adotado), a exatidão da classificação diminui [17].

À data desta pesquisa sobre o estado da arte em classificação de imagens de detecção remota, os estudos encontrados propõem o desenvolvimento de metodologias que partam do enquadramento operacional de algoritmos SVM [23-26] e redes neuronais artificiais (RNA) [19]. Lary *et al.* [20] analisaram alguns desses algoritmos nas suas capacidades de modelação, concluindo que as redes neuronais artificiais e o SVM eram as duas técnicas de aprendizagem automática mais usadas, mas que, independentemente do seu bom desempenho, eram consideradas modelos ‘caixa-negra’, ou seja, não eram capazes de gerar equações preditivas de carácter prático.

No campo da aprendizagem profunda em aplicações de detecção remota, Ma *et al.* [21] revêem exaustivamente as técnicas e métodos mais usados em análise de imagem, nomeadamente, fusão e registo de imagem, classificação de cenas, detecção de objetos, classificação de cobertura e uso do solo, segmentação e análise baseada em objetos. Os autores relembram que, até ao desenvolvimento de técnicas de aprendizagem profunda, o foco da comunidade de detecção remota estava em modelos como o SVM e os classificadores *ensemble* (abordados de seguida), como por exemplo, o algoritmo RF para tarefas de classificação de imagem e de detecção de alterações.

Nas secções seguintes descrevem-se, brevemente, alguns dos algoritmos usados em classificação no campo da detecção remota, enumerando-se as suas principais vantagens e inconvenientes.

2.2.2.1 Support Vector Machine (SVM)

Na sua formulação inicial, a este método é apresentado um conjunto de amostras rotuladas e o algoritmo SVM em treino tem como objetivo encontrar o hiperplano ótimo (fronteira de decisão que minimiza classificações incorretas) capaz de separar o conjunto de dados num número discreto de classes de modo consistente com os exemplos de treino [22]. Um aspeto de generalização importante destes algoritmos é que, frequentemente, não são usados todos os exemplos treino na especificação deste hiperplano mas apenas um subconjunto de elementos que se localizam na margem desta superfície de separação - exemplos que formam as fronteiras das classes (vetores de suporte, *support vectors*). Isto é particularmente interessante quando o número de amostras treino é reduzido. Por ser um modelo não paramétrico, estes algoritmos procuram minimizar o erro de classificação em dados não vistos sem suposições feitas na sua distribuição de probabilidades [22]. Em classificação de imagem, o SVM tornou-se popular pela sua capacidade de lidar com dados de elevada dimensionalidade, bom desempenho com número limitado de amostras treino e esforço reduzido no seu design [17].

No entanto, os algoritmos SVM foram desenvolvidos para resolver problemas de classificação binária linear, que atribuem a uma dada amostra de teste uma de duas classes possíveis. Particularmente em detecção remota, os dados, adquiridos por sensores multiespectrais, representados pelos valores das suas variáveis, sobrepõem-se no espaço, o que dificulta a sua separabilidade linear com exatidão elevada. Outros desafios à sua aplicabilidade são: i) a escolha apropriada da função *kernel* – a função que atua na discriminação de dados por transformação da dimensionalidade do espaço (a aprendizagem é realizada neste espaço de dimensionalidade mais

elevada e os dados classificados posteriormente no espaço de *input* original [23]); ii) pode ser dispendioso do ponto de vista computacional [24]. Várias técnicas têm sido desenvolvidas, com sucesso, na adaptação deste algoritmo a tarefas de classificação de dados de imagens multiespectrais, nomeadamente no que respeita à não linearidade da separação de elementos no espaço de variáveis e na classificação em mais do que duas classes. O trabalho de Mountrakis [22] passa em revista alguns dos mais relevantes estudos feitos acerca da aplicação e melhoramento deste algoritmo em deteção remota.

2.2.2.2 Redes Neuronais Artificiais (RNA)

As RNA são algoritmos de aprendizagem automática, não-paramétricos inspirados nas redes neuronais biológicas. Estas são redes que consistem num grupo de nós (ou neurónios) conectados, que processam informação em paralelo e podem ser usadas para modelar relações complexas entre *inputs* e *outputs* e encontrar padrões em dados. Um nó recebe *input* de outros nós, ou de uma fonte externa. Cada *input* tem um peso associado (w) que determina a sua força e sinal e que pode ser modificado para mimetizar a aprendizagem sináptica. Cada unidade computa a função f , da soma ponderada dos seus *inputs* e o *output* subsequente servirá de *input* no nó seguinte (j).

$$y_i = f \left(\sum_j w_{ij} y_j \right) \quad (2.3)$$

w_{ij} refere-se ao peso da unidade j para a unidade i . A função f é denominada função de ativação e, genericamente, define o *output* de um nó, dado um, ou mais, *inputs* [25]. O modo como os nós se interconectam pode ocorrer apenas num sentido (*feed-forward*) ou em forma de ciclos (retropropagação). A rede neuronal é, na maioria dos casos, um sistema adaptativo que modifica a sua estrutura com base na informação que flui na rede durante a fase de treino – a aprendizagem faz-se pelo ajustamento dos pesos ao nível do nó de modo a minimizar a diferença entre a ativação do nó de saída e o resultado.

Devido à sua adaptabilidade e capacidade de produzir classificações com melhor exatidão do que outras geradas com métodos paramétricos, o uso de redes neuronais artificiais popularizou-se no campo da deteção remota. O tipo de rede neuronal mais usado é o perceptrão multicamada (*Multi-layer Perceptron*, MLP), um modelo do tipo *feed-forward*, que aprende uma função de ativação, linear ou não, e apresenta entre a camada de *input* e a de *output* uma ou mais camadas (ocultas) de nós de processamento [26] [27].

Alguns dos problemas de utilização das RNA relacionam-se com o seu design e implementação, nomeadamente i) a especificação do número e dimensão das camadas ocultas (não existem ferramentas universalmente aceites para estimar o número de camadas ocultas para um problema particular); ii) a escolha de valores de parâmetros da rede (o intervalo de pesos iniciais, a taxa de aprendizagem, a taxa de momento de treino, critério de paragem, o número de iterações) que é normalmente feita por tentativa e erro e que reforça o já demorado processo de treino, entre outras. Adicionalmente, sendo um método ‘caixa negra’ não é exatamente conhecida a forma como as RNA aprendem problemas particulares e aplicam as regras daí extraídas a casos não vistos ou como podem conclusões ser retiradas de redes treinadas [27].

2.2.2.3 *K-Nearest Neighbor (kNN)*

O algoritmo kNN é um método para classificar objetos com base nos exemplo de treino mais próximos no espaço de variáveis. Este é a mais simples e fundamental técnica de classificação quando há pouco ou nenhum conhecimento prévio sobre a distribuição dos dados. Assume que, a menor distância corresponde maior semelhança: píxeis próximos uns dos outros no espaço de variáveis pertencem, com elevada probabilidade, à mesma classe. Na sua forma mais simples, um píxel desconhecido é rotulado examinando os píxeis de treino disponíveis no espaço multiespectral e escolhendo a classe mais representada entre um número especificado (k) de vizinhos. Fundamentalmente, esta comparação requer o cálculo da distância do píxel desconhecido a cada um dos píxeis de treino. Por vezes, é mais adequado ponderar os vizinhos de modo que os vizinhos mais próximos contribuam mais para o ajuste do classificador, por exemplo.

O desempenho de um classificador kNN é determinado principalmente pela escolha de k , bem como pela métrica de distância aplicada. O raio da região local é determinado pela distância do elemento de ordem k ao elemento de classe desconhecida e diferentes valores de k conduzem a diferentes distribuições de probabilidade condicional de classe. A escolha ideal do valor ideal de k depende muito dos dados: em geral, com um baixo valor de k , a estimativa local tende a ser pobre devido à escassez de dados e pontos ruidosos, ambíguos ou incorretamente classificados; com um maior valor de k , há supressão dos efeitos do ruído, mas os limites da classificação entre os elementos tornam-se menos distintos [28].

Para cada píxel ser rotulado, é necessário avaliar tantas distâncias espetais quanto o número de píxeis de treino. Tal tarefa requer uma capacidade de processamento elevada, particularmente quando o número de bandas espectralis/ou o número de amostras treino é grande. O método não é considerado adequado, portanto, para conjuntos de dados hiperespetrais [29].

2.2.2.4 *Random Forest*

O algoritmo RF, é um método *ensemble* desenvolvido por Breiman [30] que agrega várias árvores de decisão, para modelar comportamentos complexos. Estes métodos são relativamente fáceis de usar para além de possibilitarem o posicionamento das variáveis de acordo com a sua capacidade preditiva. A sua ideia básica é a construção de uma coleção de modelos diferentes por treino em subconjuntos de dados, selecionados aleatoriamente do conjunto total de dados treino e combinar o seu *output*, calculando a média de *outputs* individuais ou retirando a resposta maioritária (voto por maioria) [31]. Para cada subconjunto de dados treino de cada árvore individual da RF, é incorporado um subconjunto de variáveis selecionado aleatoriamente e diferentes entre si.

Em deteção remota, este algoritmo tem-se tornado popular pela exatidão do seu desempenho e o sucesso no trabalho com dados de elevada dimensionalidade e multicolinearidade, rapidez no processamento e insensibilidade ao sobreajustamento (*overfitting*). Numerosos autores centraram a sua atenção neste classificador. Na aplicação a dados de deteção remota, Belgiu & Dragut [32] apresentam em revista o sucesso do RF em classificação destacando as suas características de parametrização e sensibilidade a mudanças nos métodos de amostragem, tamanho e representatividade das amostras de treino e ruído nos dados e, ainda, a utilização da medição da

importância das variáveis explicativas na classificação que o próprio algoritmo fornece na compreensão dos dados, na redução da sua dimensionalidade e na seleção das fontes mais relevantes dos mesmos e das alturas do ano para uma maior discriminação das classes na classificação.

Sage [33] destaca, ainda, outra grande vantagem deste algoritmo: a possibilidade de estimar a probabilidade de uma nova instância ser classificada numa das categorias possíveis. Estas probabilidades dão a confiança do classificador na previsão de classes para as novas amostras e indicam a fração do número total de árvores que vota nessa classe específica são computadas como a média das probabilidades estimadas para cada classe pelas árvores do RF.

2.2.3 Transferência da aprendizagem em classificação

Em aprendizagem automática, o conceito de transferência da aprendizagem (TA) refere-se ao uso do conhecimento adquirido por um modelo numa tarefa para resolver outras relacionadas com essa [34]. Pan & Yang [35] estabelecem os conceitos-chave para a sua compreensão: domínio, tarefa e probabilidade marginal.

O domínio \mathcal{D} consiste em dois componentes: um espaço de variáveis \mathcal{X} e uma distribuição de probabilidade marginal $P(X)$ no espaço das variáveis, onde $X = \{x_1, \dots, x_n\} \in \mathcal{X}$. Em dois domínios diferentes, os respetivos espaços de variáveis terão, provavelmente, distribuições de probabilidade marginal distintas.

O conceito de tarefa \mathcal{T} , tendo em conta um domínio $\mathcal{D} = \{\mathcal{X}, P(X)\}$, contém dois componentes: um espaço de etiquetas Y e uma função preditiva $f(\cdot)$ que não pode ser observada mas pode ser aprendida através dos dados treino. Esta consiste em pares $\{x_i, y_i\}$, em que $x_i \in \mathcal{X}$ e $y_i \in Y$. A função preditiva é usada para prever a etiqueta correspondente, $f(x)$ de uma nova amostra, x . Ou seja, $f(x)$ é a distribuição de probabilidade condicional $P(y|x)$. Dado um domínio-fonte \mathcal{D}_S (*source*, s) e uma tarefa de aprendizagem \mathcal{T}_S , um domínio-alvo \mathcal{D}_T (*target*, t) e uma tarefa de aprendizagem \mathcal{T}_T , a TA tem como objetivo melhorar a aprendizagem da função preditiva alvo $\hat{f}_T(\cdot)$ em \mathcal{D}_T , aprendendo a distribuição de probabilidades condicionais $P(Y_T|X_T)$ em \mathcal{D}_T com a informação ganha na fonte, em \mathcal{D}_S e \mathcal{T}_S , onde $\mathcal{D}_S \neq \mathcal{D}_T$ ou $\mathcal{T}_S \neq \mathcal{T}_T$ (os domínios são diferentes ou as tarefas são distintas).

Pan & Yang [35] destacam, ainda, as três principais questões de investigação que enformam a TA: i) o que transferir, ii) como transferir e iii) quando transferir. A primeira questão está relacionada com que porção de conhecimento pode ser transferida entre domínios ou tarefas, distinguindo entre conhecimento específico a domínios ou tarefas individuais e conhecimento comum a diferentes domínios tal que possa melhorar o desempenho do classificador no domínio ou tarefa alvos. Após perceber que conhecimento pode e deve ser transferido, importa saber como (questão ii)) e quando o fazer (e, do mesmo modo, quando o evitar – questão iii)).

No caso específico de o domínio-fonte e o domínio-alvo serem o mesmo, i.e. $\mathcal{D}_S = \mathcal{D}_T$, e as suas tarefas as mesmas, i.e. $\mathcal{T}_S = \mathcal{T}_T$, o problema de aprendizagem é um problema de aprendizagem automática tradicional.

Para o contexto específico da utilização de classificadores RF, Segev *et al.* [36] propõem dois algoritmos para a TA entre modelos RF que utilizam um modelo treinado num domínio-fonte e o adaptam a um domínio-alvo usando ajustamentos locais dos parâmetros da árvore e/ou da sua

arquitetura: um, que procura resolver o problema de modo *greedy* por modificações ótimas, locais à estrutura de cada árvore, e tenta expandir ou reduzir a árvore em torno dos nós individuais (SER – *structure expansion/reduction*); e outro, que não modifica a estrutura da árvore mas somente os parâmetros associados aos nós de decisão (STRUT – *structure transfer*).

2.2.4 Transferência da aprendizagem em classificação de imagens de detecção remota - adaptação de domínio

Em processos não estacionários no tempo ou no espaço, surgem os problemas de TA: este é o caso da análise em imagens de detecção remota, onde diferentes aquisições correspondem a diferentes condições atmosféricas (por exemplo, iluminação, ângulo de visada, humidade do solo e topografia) que podem afetar a assinatura espectral dos tipos de cobertura do solo e a distribuição das classes informativas no espaço das variáveis [37]. Surge, assim, a necessidade de desenvolver algoritmos que aprendam de forma eficiente a partir de um relativamente reduzido número de amostras rotuladas, usando o conhecimento de dados relacionados mas não etiquetados, ruidosos e/ou com uma distribuição distinta. Os algoritmos convencionais de aprendizagem automática estão desenvolvidos para resolver tarefas específicas e têm de ser reconstruídos assim que a distribuição probabilística do espaço de variáveis muda [34].

Portanto, no caso da construção de mapas de cobertura agrícola, os dados podem sofrer facilmente uma desatualização (por exemplo, com a mudança de culturas no ano seguinte pelo agricultor) e os dados obtidos para um dado período (ou região) podem não seguir a mesma distribuição probabilística mais tarde (ou noutra região). A TA torna-se especialmente importante porque um modelo treinado num período temporal (ou região, espacial) pode levar a que o desempenho da previsão desse modelo noutra período ou região seja reduzido [35].

Um outro obstáculo comum em classificação de imagem, e profusamente explorado em classificação de imagens adquiridas por detecção remota, é a insuficiência ou inexistência de amostras rotuladas a fornecer ao algoritmo de classificação para treino, devido aos recursos envolvidos em idas ao terreno, o que conduz à construção de classificadores sobreajustados e com pouca capacidade de generalização, impossibilitando a atualização de mapas de cobertura do solo por classificação supervisionada direta de cada imagem da série temporal. Para reduzir os custos de recalibração do classificador, será desejável usar métodos de TA no sentido de adaptar o modelo de classificação treinado com um conjunto de dados específico de um período temporal (domínio-fonte) para um novo período (o domínio-alvo) ou adaptar o modelo de classificação treinado numa dada zona geográfica (domínio-fonte) a outra zona geográfica (domínio-alvo).

A classificação de imagens obtidas por detecção remota, adquiridas sobre áreas geográficas vastas ou em intervalos de tempo curtos, depende grandemente da representatividade das amostras usadas para treino dos algoritmos de classificação, na definição do modelo. É provável que um modelo falhe se as amostras de treino são espectralmente diferentes de amostras usadas para a produção de um mapa de classificação. Estas alterações resultam das diferenças na aquisição e nas condições atmosféricas ou em mudanças na natureza do objeto observado atrás enunciadas. Tuia *et al.* [37] reconhecem esta problemática e centram-se no desenho de métodos de classificação para a ‘adaptação de domínio’ (*domain adaptation*, ou DA), previamente

introduzido por Pan & Yang [35] e propõem-se rever criticamente os avanços nestes métodos em deteção remota.

Recuperando os conceitos já descritos, supondo que, aos domínios fonte e alvo estão associadas as distribuições das probabilidades conjuntas $P_S(X, Y)$ e $P_T(X, Y)$, que definem os problemas de classificação num e noutra, respetivamente, onde X é o vetor-variável (por exemplo, as bandas espectrais da imagem-fonte e/ou outras variáveis adicionais usadas para caracterizar o contexto de um píxel – ex.: marcadores fenológicos) e Y é a variável-*output* associada ao conjunto de classes (informativa sobre o uso ou cobertura do solo), o objetivo do método DA é adaptar o classificador treinado no domínio-fonte para fazer estimativas no domínio-alvo, ou seja, adaptar modelos treinados para completar uma tarefa específica, a uma nova tarefa relacionada com a primeira, para a qual o conhecimento do modelo inicial é suficiente, embora imperfeito [37].

Adicionalmente, Tuia *et al.* [37] propõem uma taxonomia de métodos de DA¹: DA por 1) seleção de variáveis invariantes (*invariant feature selection*), 2) adaptação das distribuições dos dados (*adaptation of the data distributions*), 3) adaptação de classificadores (*adaptation of classifiers*) e por 4) adaptação do classificador por aprendizagem ativa (*adaptation of the classifier by active learning*). No primeiro método, um subconjunto do conjunto inicial de variáveis é selecionado, antes do treino, de modo a incluir as variáveis simultaneamente mais discriminativas das classes de cobertura do solo e robustas ou invariantes face a mudanças entre domínio-fonte e domínio-alvo. No segundo, tornam-se as distribuições dos dados de domínios fonte e alvo o mais semelhantes possível para manter o classificador inalterado, o que normalmente é conseguido com extração conjunta de variáveis. O objetivo é tornar as distribuições de dados semelhantes ao longo dos domínios e treinar um único modelo que possa classificar, então, o domínio-fonte e o domínio-alvo. Na terceira categoria de métodos – DA por adaptação de classificadores – o modelo de classificação é definido pelo treino no domínio-fonte e adaptado ao domínio-alvo considerando as amostras não rotuladas no domínio-alvo. A maioria destes métodos assume que o domínio-fonte e alvo partilham o mesmo conjunto de classes e de variáveis. Finalmente, no quarto grupo de métodos de DA – por adaptação do classificador por aprendizagem ativa – o classificador treinado no domínio-fonte adapta-se ao domínio-alvo pelo fornecimento de amostras rotuladas do próprio domínio-alvo, especialmente selecionadas pelo seu potencial informativo, que são incorporadas nos dados de treino. Este conjunto de métodos aborda uma questão central em classificação de imagem de deteção remota: a definição eficiente do conjunto de dados para treino de um classificador, de forma a melhorar o desempenho do modelo iterativamente através de amostragem [38]. O processo inicia-se com um classificador ótimo para o domínio-fonte e evolui gradualmente para a modelação da distribuição de dados no domínio-alvo [37].

Demir *et al.* [39] propõem a TA conduzida pela deteção de alterações para a atualização de mapas de cobertura do solo pela classificação de imagens de diferentes séries temporais adquiridas na mesma área geográfica. A TA por adaptação de domínio proposta centra-se na propagação de etiquetas do domínio-alvo para o domínio-fonte. O processo compõe-se de três fases: i) a definição do conjunto de treino-alvo por aplicação de algoritmo de deteção de alterações não supervisionado (não há extração de etiquetas do domínio-alvo); ii) o enriquecimento deste conjunto de treino-alvo por aprendizagem ativa segundo a qual em primeiro lugar, se etiquetam amostras cuja classe se alterou no tempo e só depois se etiquetam amostras não rotuladas (quer a classe se tenha alterado, quer não); iii) classificação da imagem alvo por classificação supervisionada. Os resultados deste método mostram que, para além de se obter um conjunto de

¹ em tradução livre

dados treino inicial fiável, a regra de prioridades estabelecida pelos autores resulta num aumento na convergência para a exatidão de desempenho na classificação desejada relativamente a outros métodos de aprendizagem ativa.

2.2.5 Transferência da aprendizagem em classificação de imagens de deteção remota - uso de mapas temáticos de cobertura do solo

As dificuldades na TA entre modelos de classificação de cobertura do solo agrícola estão em grande parte dependentes das soluções apresentadas para a adaptação de domínio, mas também, como chamam a atenção Paris *et al.* [40], pela compreensão e modelação do domínio do mapa de classificação da cobertura agrícola (domínio-alvo) no sentido de reduzir a discrepância entre este e o domínio da imagem multiespectral (domínio-fonte). Esta necessidade advém de dois problemas inerentes aos mapas de classificação: i) sendo estes, na maioria das vezes fornecidos em formato vetorial, as etiquetas atribuídas aos polígonos (parcelas agrícolas) não correspondem necessariamente a áreas espectralmente homogêneas e ii) há, com frequência, uma discrepância semântica entre a legenda do mapa e o conjunto de classes naturais discrimináveis nas imagens multiespectrais (por exemplo, uma classe ‘cultura’ pode, no terreno, incluir diferentes culturas agrícolas).

Os autores apresentam um método para reduzir esta discrepância sem aprofundarem estratégias de remoção de *outliers* estatísticos para classificação (i.e., *outliers* no domínio espectral - os píxeis menos fiáveis para classificação), pelo facto de estas já terem sido profusamente exploradas na literatura e por apresentarem desvantagens como o risco elevado de se removerem inadvertidamente amostras informativas, afetando-se assim a capacidade de generalização do classificador. Propõem então: i) a compreensão do domínio-fonte por decomposição semântica da legenda do mapa (quantas culturas naturais discrimináveis na imagem multiespectral cabem no mapa de classificação?), seguido da ii) modelação do domínio-fonte por extração, do mapa de classificação, de amostras que sejam, simultaneamente, confiáveis, informativas e representativas (estas são detetadas aplicando previamente *clustering* não supervisionado que gerará um conjunto de dados treino ‘confiável’) e iii) adaptação de domínio: um classificador é treinado na imagem nova com o conjunto de treino ‘confiável’ derivado de ii). Os autores usam um classificador ensemble SVM e os resultados são promissores, sendo a exatidão global do método proposto (com adaptação de domínio), de aproximadamente 80%, próxima da do método base (sem adaptação de domínio), cerca de 84% [40]. No caso particular da classificação de imagens e transferência de aprendizagem, estes autores salientam a necessidade de se explorar o potencial destas técnicas na análise de séries temporais, dadas as suas vantagens.

Adicionalmente, como referido, a proposta do JRC para a base operacional de monitorização da PAC assenta na plataforma *Tensorflow*. Esta, usa redes neuronais profundas no treino e teste de algoritmos [3] e foi desenvolvida em linguagem *Python* (em crescente popularidade e amplo desenvolvimento), permitindo a replicação das orientações dadas pelo JRC noutras bibliotecas de aprendizagem automática como a *theano* ou a *Scikit-learn* [41].

3 Dados e Métodos

3.1 Área de estudo

A área deste estudo abrange as regiões de Portugal continental ilustradas pelas molduras coloridas da figura 3.1. Usaram-se dados relativos às regiões da Beira Litoral e Interior (quadrado NE da quadrícula UTM) e da Estremadura e Ribatejo, e Alentejo (maioritariamente quadrados ND e NC, respetivamente) (figura 3.1). Estas três zonas situam-se entre as longitudes $9^{\circ}0'0''\text{W}$ e $7^{\circ}42'5''\text{W}$ e as latitudes $40^{\circ}38'37''\text{N}$ e $37^{\circ}51'18''\text{N}$ (datum WGS84). Na grande parte do território continental, onde se incluem estas zonas geográficas, o clima é temperado, com verão quente e seco nas regiões interiores do vale do Douro e nas regiões a sul do sistema montanhoso Montejunto-Estrela, exceto no litoral oeste do Alentejo e Algarve e quase todas as regiões a Norte do sistema montanhoso Montejunto-Estrela, onde o verão é seco e suave. No distrito de Beja, o clima é árido [42].

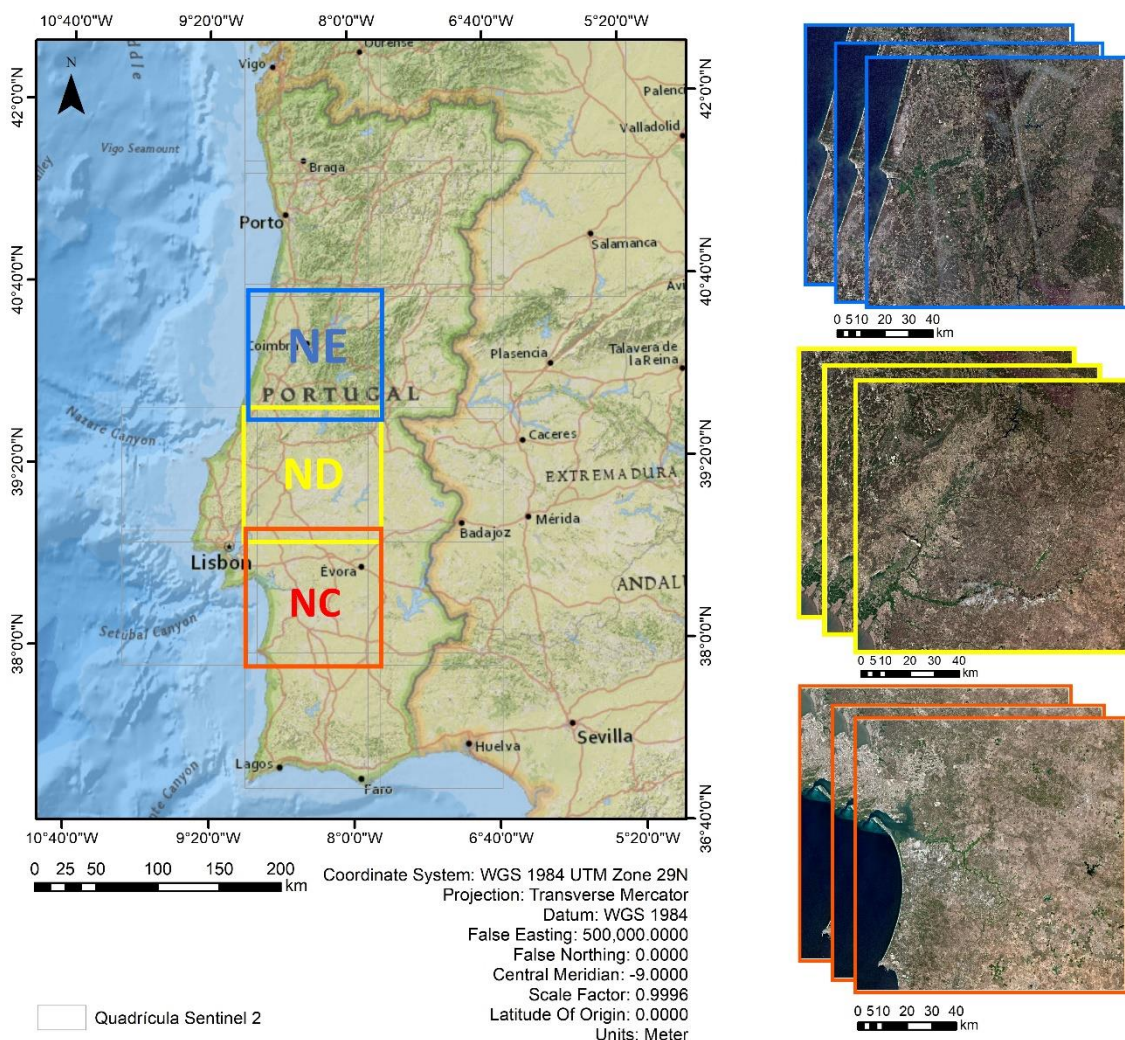


Figura 3.1 – Localização da Área de Estudo mostrando a quadrícula UTM dos produtos Sentinel-2 usados neste estudo.

A morfologia das áreas agrícolas em Portugal caracteriza-se pela disparidade ao nível da dimensão das parcelas agrícolas, sendo que na maior parte do território continental, se encontra parcelas minifundiárias de pequenas dimensões (entre 2.5ha, na Beira Litoral, e 4.3ha, na região de Entre Douro e Minho) e no Alentejo predominam os grandes latifúndios em campos abertos (parcelas com mais de 65ha, em média). Os sistemas de culturas variam entre monocultura e policultura, regadio e sequeiro, e intensivo/extensivo. A monocultura, associada à produção extensiva com pousio, é dominante nas explorações do Alentejo e Trás-os-Montes, enquanto que no Ribatejo e Oeste a agricultura é sobretudo intensiva; a policultura, normalmente regada e sem pousio (intensiva), é característica do litoral Norte e Centro e Algarve.

A superfície agrícola portuguesa caracteriza-se, ainda, pelas culturas temporárias nas regiões de Entre Douro e Minho, Beira Litoral, Ribatejo e Oeste e Madeira; pelas culturas permanentes em Trás-os-Montes e Alentejo; e pelas pastagens permanentes no Alentejo e Açores. As culturas temporárias mais relevantes são o milho, o arroz, a batata e as culturas hortícolas. De entre as culturas permanentes, destacam-se as árvores de frutos frescos e secos, a vinha e a oliveira [43].

3.2 Dados

3.2.1 Dados *in situ*: Sistema de Identificação Parcelar (SIP)

Usaram-se os dados de parcelas declaradas pelos agricultores em 2018 e 2019 no contexto da PAC, fornecidos pelo IFAP para desenvolver e treinar os modelos de classificação. Constam das declarações de cada agricultor, por exemplo, o tipo de cultura (temporária ou permanente) e, para cálculo da área, a delimitação geográfica da parcela agrícola. Para o projeto em que se insere este trabalho e para a área em estudo, o IFAP fornece a informação geográfica (limites de parcelas) e alfanumérica (OSA-ID e a declaração da ocupação do solo).

O sistema de coordenadas adotado no Sistema Integrado de Gestão e Controlo é o sistema ETRS89/Portugal TM06 (EPSG: 3763) e o sistema de referência das imagens do Sentinel-2 é o WGS84/UTM Fuso 29 N (EPSG: 32629), pelo que as coordenadas das parcelas foram transformadas entre os sistemas geodésicos ETRS89 e WGS84. A atualização da informação respeitante à ocupação e limite de cada parcela faz-se anualmente. Na tabela 3.1 enumera-se o número de parcelas declaradas por ano e região estudada.

Tabela 3.1- Número de parcelas agrícolas declaradas por região e por ano.

	2018NE	2018ND	2018NC	2019NE	2019ND	2019NC
Nr. total de parcelas	335794	357412	273798	335797	358245	275018
Parcelas sem cultura (%)	36.9	41.7	39.3	36.6	40.7	38.3
Área < 1000 m ² (%)	24.7	13.2	9.0	25.0	13.6	9.4
Parcelas a controlar (%)	31.1	37.2	41.5	31.3	37.9	42.4
Culturas temporárias (%)	20.7	13.0	9.7	15.1	7.2	7.0

As parcelas com menos de 1000 m² representam, aproximadamente, 9% a 25% do total das parcelas, aumentando de sul para norte. O número de parcelas com culturas temporárias

declaradas varia entre as 19123 na região NC (7% das parcelas da região, ano de 2019) e as 69659 parcelas na região NE (20.7% das parcelas da região, ano de 2018) Neste estudo, foram consideradas 181 classes agrícolas referentes a culturas temporárias de verão e, adicionalmente, as culturas permanentes de vinha e olival. Destas classes apenas uma parte se encontra representada nos dados. A proporção das diferentes culturas presentes nas parcelas em cada subconjunto de dados (correspondentes a cada série temporal em cada zona geográfica) apresenta-se na figura 3.2.

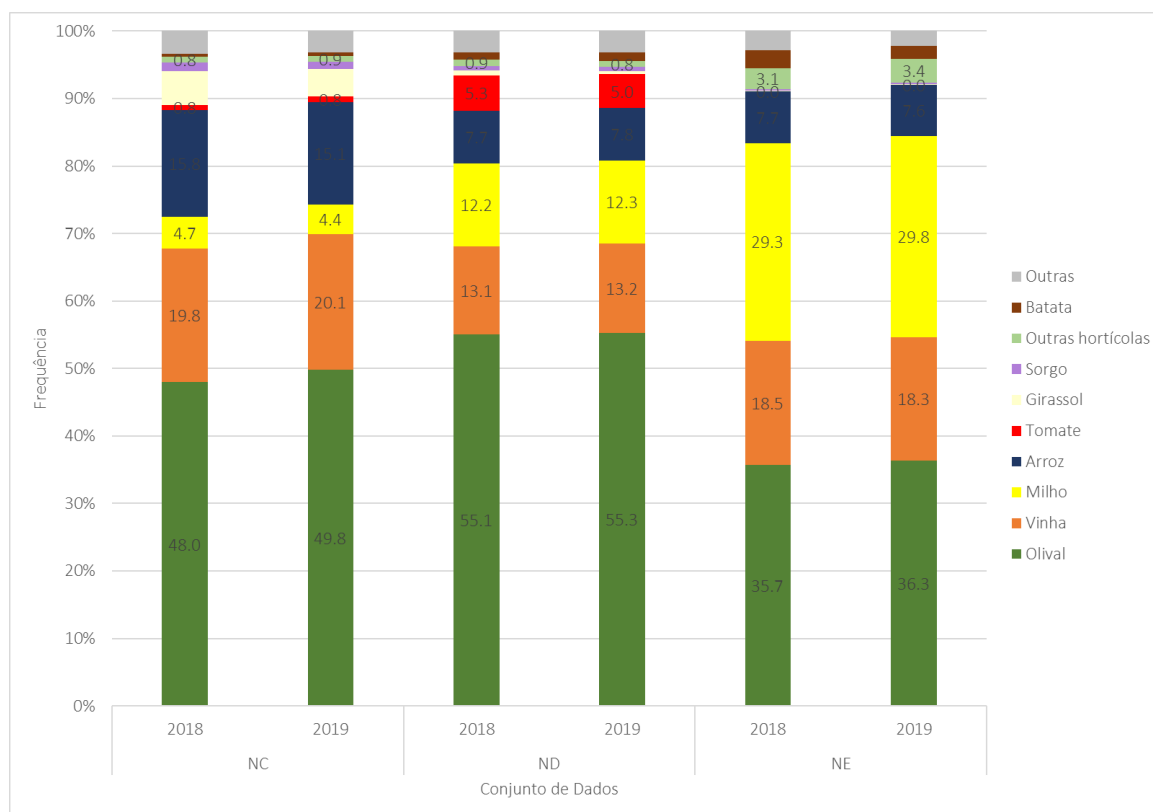


Figura 3.2 - Representatividade das classes nos conjuntos de dados.

3.2.2 Dados de Detecção Remota: Imagens Sentinel-2

As imagens do satélite Sentinel-2 foram descarregadas na plataforma online *Copernicus Open Access Hub* da ESA [6]. São produtos de nível 2A, adquiridos por ambos os satélites Sentinel-2A e 2B, durante o período ótimo para seguimento do ciclo anual agrícola das culturas de verão em Portugal, entre Março e Setembro dos anos de 2018 e 2019, nas quadrículas T29SNC, T29SND e T29TNE (abreviadas para NC, ND e NE, respetivamente) durante a órbita relativa 037 (figura 3.1). Com base neste período temporal, foi descarregado um número variável de imagens para cada zona geográfica de acordo com o representado na tabela 3.2, dependendo da presença de nuvens.

As imagens em cada quadrícula de 100 km x 100 km estão ortorretificadas, na projeção WGS84/UTM Fuso 29 N. O território nacional continental é coberto por 16 imagens (figura 3.1) adquiridas num trajeto único descendente. As imagens do satélite Sentinel-2 de nível 2A

Para a zona mais a norte – NE – há um maior número de parcelas identificadas, mas um conjunto menor de variáveis disponíveis para as caracterizar, isto é, um menor número de imagens na série temporal. Por outro lado, na zona mais a sul – NC – há um menor número de amostras (menor número de parcelas identificadas com as classes consideradas), mas um conjunto maior de variáveis para a caracterização de cada amostra. A zona ND, está entre as restantes quer em número de amostras, quer em número de variáveis disponíveis (exceto para o ano de 2018, em que se obtiveram apenas 90 variáveis, correspondentes a 10 datas). Em termos de número de amostras, a zona NE representa cerca de 47% do conjunto total de dados, seguida da zona ND (34%) e da zona NC (19%).

3.3 Metodologia

A metodologia deste trabalho foi delineada tendo em conta os diversos elementos presentes: i) os algoritmo(s) de classificação utilizado(s) e as particularidades dos seus parâmetros, ii) os conjuntos de dados disponíveis e o seu pré-processamento e iii) a conjugação de dados e algoritmo de classificação na construção do modelo de classificação.

Neste trabalho, seguiu-se o procedimento básico de um problema clássico de classificação supervisionada. Para tal, os dados originais provenientes das imagens foram preparados e pré-processados para serem, de seguida, introduzidos no algoritmo de classificação. Após a produção de uma decisão pelo classificador, foi feita a avaliação e delineadas estratégias para melhoria do modelo (figura 3.3).

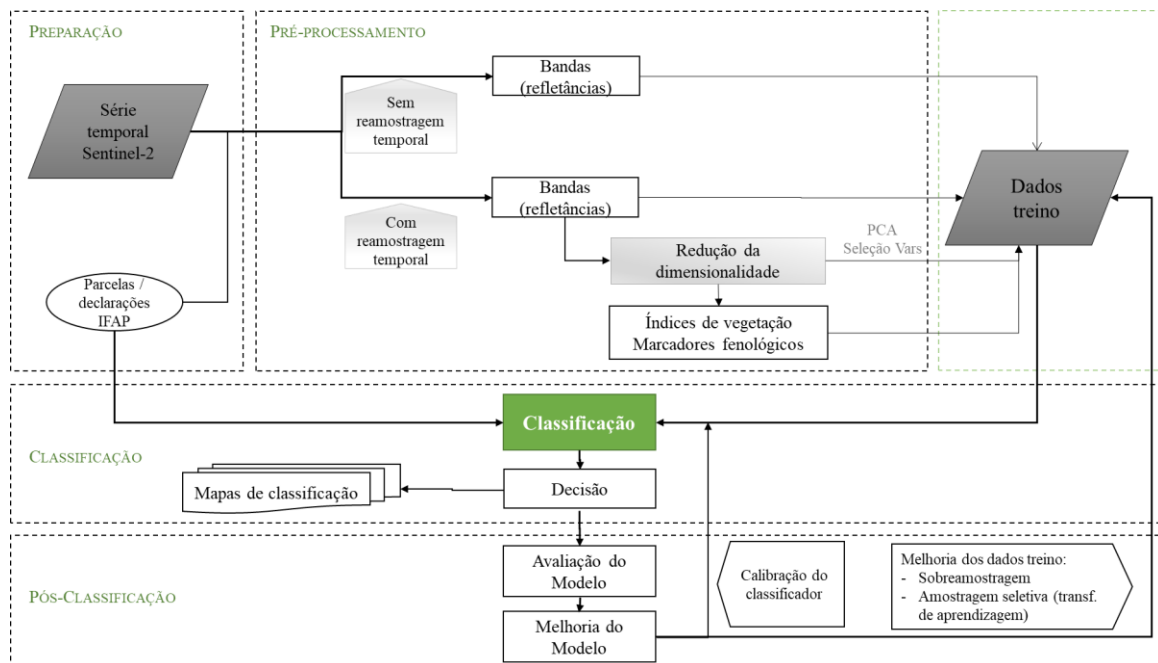


Figura 3.3 - Fluxograma das diversas fases do trabalho de projeto.

Os ficheiros de *input* usados foram:

1. Série temporal de imagens processadas a nível 2A (ortorretificadas, com valores de refletâncias para nove bandas) – este dará a matriz x de valores que caracterizam cada parcela (eq. (2.1)).
2. Dados de referência com amostras rotuladas (*shapefiles* usadas com localizações geográficas para as quais o tipo de cultura é conhecido) – que darão o vetor Y de etiquetas (classes) para cada parcela (eq. (2.1));

A implementação dos algoritmos de classificação foi realizada recorrendo a ferramentas disponibilizadas em linguagem Python 3.7 (nomeadamente a biblioteca Scikit-learn, v. 0.21.0 e Pandas 0.25.2). As principais aplicações informáticas usadas nas várias fases do estudo apresentam-se na tabela 3.4.

Tabela 3.4 - *Software* utilizado neste estudo.

Software	Tarefas
ESA SNAP	Preparação de dados: Processamento inicial das imagens Sentinel-2 (reamostragem a 10m de resolução)
Matlab R2017a	Preparação de dados: Tratamento de dados vetoriais (atribuição de valores e classes a cada parcela) Conversão de dados das imagens (refletâncias) em ficheiros .csv Extração de variáveis: índices de vegetação e marcadores fenológicos para cada parcela
Spyder v.3.3.3	Pré-processamento de dados em formato .csv Aplicação de ferramentas de classificação (biblioteca Scikit-learn v. 0.21.0 e Pandas 0.25.2) Classificação e validação dos resultados
Excel 365	Sistematização, visualização e análise de resultados
ArcGIS Desktop 10.6.1	Visualização, análise e discussão de resultados de classificação

Foram realizados diversos testes tendo como objetivo obter os modelos de classificação mais robustos, do ponto de vista da sua exatidão global (como referido, esta é apenas uma, entre muitas métricas disponíveis para avaliação da qualidade de um modelo e a que foi, nesta primeira fase, utilizada). Estes testes e respetivos procedimentos e objetivos, estão descritos na tabela 3.8 no final deste capítulo.

3.3.1 Preparação dos Dados

A fase de preparação dos dados consistiu na preparação dos objetos de classificação: as parcelas agrícolas. Cada parcela constitui um objeto de classificação e traduz-se pelo conjunto de píxeis existentes no interior dos limites georreferenciados - informação constante no Sistema de Identificação Parcelar, do IFAP. A cada parcela fez-se corresponder o valor da mediana do respetivo conjunto píxeis. Apenas um subconjunto dos dados iniciais foi considerado obedecendo a dois critérios: manutenção de todas as parcelas com dimensão igual ou superior a 1000 m² e remoção dos píxeis correspondentes a campos vazios (sem culturas agrícolas).

3.3.2 Pré-processamento dos dados

A fase inicial do pré-processamento dos dados incluiu:

- limpeza e transformação dos dados: teve como objetivo inspecionar e remover dados em falta ou nulos e remover parcelas únicas (i.e., cuja classe tem apenas um elemento), uniformizar os cabeçalhos entre conjuntos de dados quando necessário, atribuir a coluna de classificação de referência (de acordo com dados *in situ*);
- análise exploratória dos dados (dimensionalidade, visualização de distribuição de refletâncias por classe)
- preparação dos dados para processamento pelo algoritmo de classificação: geração de conjuntos de treino e de validação, extração e engenharia de variáveis (uniformização de número de variáveis).

3.3.2.1 Interpolação temporal e reamostragem

À semelhança da metodologia usada por Inglada [45], a existência de lacunas nas séries temporais devido a cobertura de nuvens, foi colmatada por reamostragem temporal. Neste caso, o processo de reamostragem usa funções de interpolação para encontrar os valores dos píxeis em épocas que estão ausentes na série temporal. O desempenho do processo de reamostragem depende da função de interpolação usada: neste estudo, optou-se pela interpolação linear. Este é um método determinístico, polinomial, que estima um novo valor de uma função entre dois valores adjacentes conhecidos, o último ponto antes da ausência de informação e o primeiro ponto depois da ausência de informação [46].

O método foi implementado através de uma grelha temporal com início na primeira data de aquisição comum às três séries temporais (correspondentes às três zonas em estudo), 30 de Abril, com um intervalo de amostragem de 5 dias, até ao final da época de cultivo das culturas consideradas – 27 de Setembro. Para cada píxel, a série temporal de cada banda espectral é linearmente interpolada, usando-se apenas os dados válidos, ou seja, píxeis sem nuvens, sem sombras e não saturados. Produz-se, assim, uma nova série temporal com todos os valores válidos originais e os interpolados na grelha temporal de 5 dias.

3.3.2.2 Redução da dimensionalidade dos dados por aplicação de PCA e por seleção de variáveis

Para diminuir e uniformizar o número de variáveis em cada conjunto de dados, foram utilizadas duas técnicas de redução da dimensionalidade:

- Análise de Componentes Principais (*Principal Component Analysis*, PCA): o objetivo foi selecionar os conjuntos de bandas e datas mais informativos de forma a caracterizar os perfis de cada cultura. A variabilidade numérica dos dados não se perde embora os valores que cada componente traduz se assemelhem a uma caixa negra, uma vez que, quer a variáveis originais, quer os seus valores, ficam sintetizados nos valores expressos pelas componentes

mais explicativas escolhidas [47]. Escolheu-se o mesmo valor ‘explicativo’, limite a partir do qual se decidiu o número de componentes a selecionar para cada conjunto de dados: 98%. Em todos os conjuntos de dados este valor foi satisfeito por trinta componentes. Ou seja, as trinta componentes escolhidas, que resultam de uma síntese das diversas variáveis, explicam 98% das variabilidade observada nos dados. Esta técnica permite selecionar variáveis que alinham os espaços de variáveis dos dados entre si que, por sua vez, possibilita aplicar o mesmo classificador a todos os domínios [37] realizando a TA desejada.

- Seleção de variáveis: usando o método *SelectKBest* (da biblioteca Scikit-learn, v. 0.21.0), que seleciona as variáveis mais importantes para cada conjunto de dados com base em testes estatísticos univariados [48]. Neste caso, usou-se o teste do qui-quadrado: o método aplica o teste por comparação de cada variável com a classe (etiqueta) para perceber se há uma relação estatisticamente significativa entre elas. As variáveis que apresentam valores mais baixos no teste são as menos dependentes da classe e são, portanto, excluídas por irrelevância no processo de classificação, até se atingir k variáveis a serem selecionadas, valor introduzido pelo utilizador. Neste estudo procurou-se reter apenas as variáveis comuns a todos os conjuntos de dados, um pouco na linha do enunciado por Tuia [37] (capítulo *Estado da Arte*) no método de adaptação de domínio por seleção de variáveis invariantes entre conjuntos de dados com origens diferentes de modo a facilitar a adaptação do domínio-fonte ao domínio-alvo nos processos de TA entre classificadores, no tempo e/ou espaço. Na prática, a seleção de variáveis permite uma maior flexibilidade na manipulação dos dados-treino. Os resultados desta opção são apresentados e discutidos na secção seguinte (capítulo *Resultados e Discussão*).

A redução da dimensionalidade permite concatenar os conjuntos de dados das diversas zonas num conjunto de dados único que foi utilizado para executar uma nova série de testes: treinar o classificador com dados provenientes das três zonas e testar o seu desempenho em dados de cada uma das zonas individualmente. Esta operação torna possível ainda manipular as proporções de dados provenientes das diferentes zonas a fornecer a uma base de dados-treino única. Esta flexibilidade na manipulação dos conjuntos de dados permite fornecer ao classificador dados-treino provenientes dos três conjuntos que sejam representativos desta disponibilidade de informação e é extremamente útil para perceber de que forma melhorar o desempenho de um classificador considerando a quantidade de dados existentes.

Os dados treino e os dados teste foram manipulados de modo a perceber que valor mínimo de amostras e qual a sua natureza (ano e zona geográfica) seria necessário fornecer ao classificador para se obter um nível de exatidão global de classificação interessante.

3.3.2.3 Redução da dimensionalidade dos dados por Marcadores Fenológicos

Como estratégia paralela de extração de variáveis, procedeu-se ao cálculo de marcadores fenológicos para cada parcela, a partir da série temporal com os valores de refletância interpolados, tal como descrito anteriormente. A variabilidade temporal dos índices vegetativos permite caracterizar uma parcela com atividade agrícola: espera-se que uma cultura temporária apresente uma elevada variabilidade do desvio padrão temporal ou do gradiente temporal do índice vegetativo. O uso dos perfis dados pelos marcadores fenológicos foi já extensamente

confirmado em identificação de culturas agrícolas, passando estes a substituir ou complementar os valores de refletância em bruto como variáveis para treino do classificador [5]. Adicionalmente, esta estratégia permite reduzir a dimensionalidade dos dados.

Os índices usados foram: NDVI (3.1), $CI_{RedEdge}$ (3.2), NDWI (3.3) e o BI (3.4), os quais foram calculados de acordo com as equações sumarizadas na tabela 3.5. As métricas calculadas, com base na série temporal, para cada um deles foram: o valor mínimo, o valor máximo, a média, o desvio padrão, o gradiente mínimo e o gradiente máximo. Esta operação resultou num subconjunto de 24 variáveis para cada ano e zona geográfica.

Tabela 3.5 - Índices de vegetação selecionados e respetivas referências.

Índice	Equação*	Autor / Obs.
NDVI (Normalized Difference Vegetation Index)	$\frac{NIR-Red}{NIR+Red} \quad (3.1)$	Rouse <i>et al.</i> [49]
$CI_{RedEdge}$ (RedEdge Chlorophyll Index)	$\frac{RedEdge_6}{RedEdge_5} \quad (3.2)$	Zarco-Tejada <i>et al.</i> [50]
NDWI (Normalized Difference Water Index)	$\frac{NIR-SWIR1}{NIR-SWIR2} \quad (3.3)$	Gao [51]
BI (Brightness Index)	$\sqrt{\sum_{i=1}^N b_i^2} \quad (3.4)$	b_i : B2, B3, B4, B5, B6, B7, B8, B11, B12

* NIR (banda 8 – infravermelho próximo); *Red* (banda 4 – vermelho); *RedEdge* (banda 5 e 6 – limiar do vermelho); SWIR1 (banda 11 – infravermelho de onda curta 1); e *Green* (banda 3 – verde) das imagens Sentinel-2.

3.3.3 Dados Treino

Os dados treino do algoritmo de classificação foram os descritos na tabela 3.3. Uma vez que se dispõe apenas de dados de 2018 e de 2019, interessa particularmente construir modelos para 2018 e aplicá-los a dados de 2019 para ter uma perceção da sua capacidade preditiva – TA no tempo. Do mesmo modo, interessa compreender a capacidade generalizadora de um modelo desenvolvido com dados de uma região específica. Para isso, testaram-se os modelos gerados em cada zona geográfica noutras, individualmente ou agrupadas - TA no espaço. Na prática, pretende-se avaliar a TA entre modelos de classificação: testaram-se modelos já treinados num dado domínio-fonte (referente a um ano e/ou a uma zona geográfica) num domínio-alvo diferente deste. Esta foi a operação base usada para aferir o desempenho de modelos de classificação em dados do ano seguinte e/ou de zonas geográficas distintas, ainda sem considerar a aplicação de metodologias mais complexas de seleção de amostras mais ou menos informativas dos dados treino (exploradas em numerosos trabalhos de que se dá conta no capítulo *Estado da Arte*).

Contando com a possibilidade de não existirem conjuntos de dados disponíveis para todas as zonas e todas as datas, foram ainda utilizadas estratégias de concatenação de dados provenientes dos diversos conjuntos de dados de forma a assegurar a ordem de proporções de cada conjunto de dados pertencente a cada zona geográfica na geração de um conjunto de dados uno, com amostras das várias regiões (NE, ND e NC).

A geração de dados de treino usando diferentes proporções dos dados permite responder a questões como:

1. Qual o número mínimo de amostras a reunir de modo a construir um modelo de classificação suficientemente robusto para amostras nunca vistas, por exemplo, de anos seguintes ou de zonas geográficas diferentes?
2. Com que exatidão é possível classificar dados de uma zona e/ou ano diferentes daqueles com que o modelo foi treinado?

Neste sentido, a cada um dos três conjuntos de dados iniciais (três zonas: NE, ND e NC) para o ano de 2018 foram selecionadas 30% das amostras, mantendo-se a proporção da representatividade de culturas/classes em cada zona. Estes subconjuntos foram reagrupados num conjunto de dados que serviu para treino do modelo de classificação. Cada um dos modelos foi subsequentemente testado com os restantes 70% dos dados, com ou sem TA no espaço/tempo (figura 3.4).

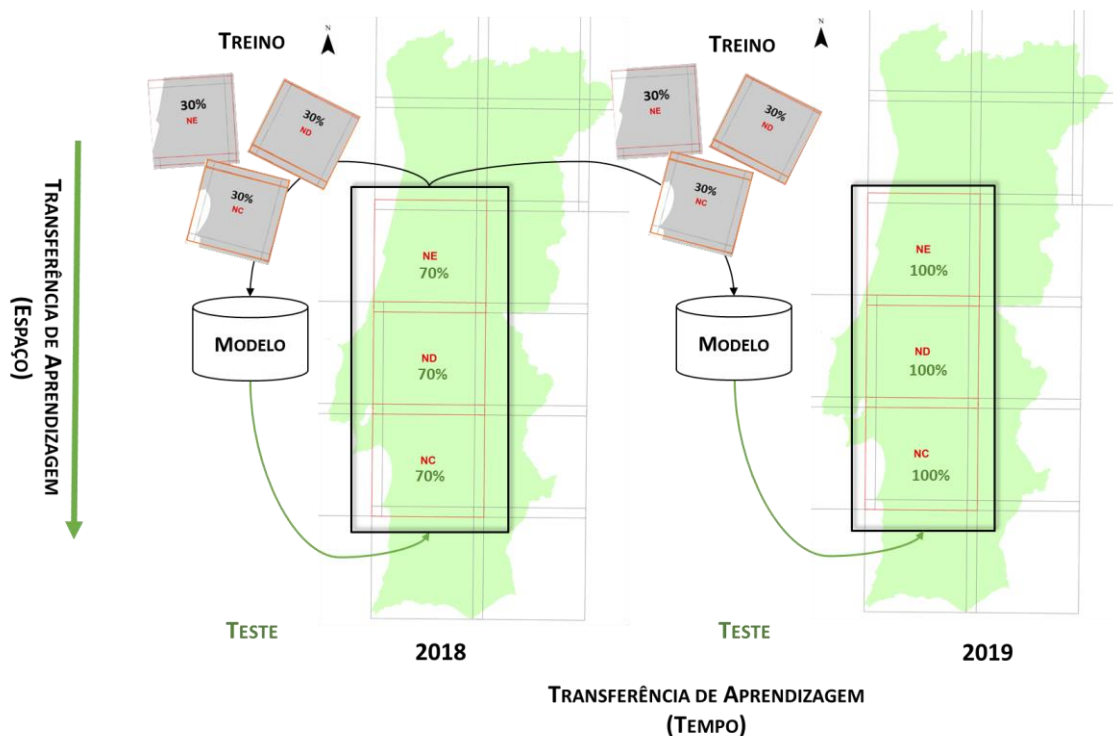


Figura 3.4 – Diagrama de trabalho para pré-processamento, construção do modelo de classificação treinado em dados combinados das três zonas de um ano na proporção de 30% (cada zona) e teste em dados do mesmo ano (numa zona diferente) ou no ano seguinte.

Do mesmo modo, criaram-se mais dois conjuntos de dados de treino do classificador fazendo variar a percentagem inicial de dados a retirar do conjunto de amostras de 2018: 50% e 70%. O processo de teste foi idêntico, mas para os restantes 50% e 30% de dados, respetivamente (tabela 3.8).

3.3.4 Classificação: algoritmos e abordagem

Com os dados treino preparados, foi feita a classificação das imagens por implementação inicial a título exploratório, dos algoritmos de classificação: *Support Vector Machines* (SVM), *Multi-layer Perceptron* (MLP), *K-Nearest Neighbor* (kNN) e *Random Forest* (RF), disponíveis em Scikit-learn v. 0.21.0. O RF foi aquele que apresentou resultados de exatidão global e o tempo de computação mais satisfatórios pelo que foi escolhido como base dos modelos de classificação desenvolvidos (*vide Resultados e Discussão*). Os parâmetros usados para a comparação de algoritmos foram os dados por defeito na biblioteca do Scikit-learn (tabela 3.6). Posteriormente, como se verá, estes parâmetros foram afinados para o algoritmo escolhido - RF.

Tabela 3.6 - Valores dos principais hiperparâmetros de cada algoritmo de classificação usados.

Algoritmo de Classificação	Valores dos principais hiperparâmetros (por defeito, Scikit-learn v. 0.21.0)	Fonte
SVM	$C=1.0$ $kernel='rbf'$ $gamma='scale'$	$class_weight=None$ $decision_function_shape='ovr'$ [52]
MLP	$hidden_layer_sizes=(100,)$ $activation='relu'$ $batch_size='auto'$ $learning_rate='constant'$	$tol=0.0001$ $momentum=0.9$ $early_stopping=False$ [53]
kNN	$n_neighbors=5$ $weights='uniform'$	$metric='minkowski'$ [54]
RF	$n_estimators=100$ $criterion='gini'$ $max_depth=None$ $min_samples_split=2$ $min_samples_leaf=1$	$max_features='auto'$ $bootstrap=True$ $class_weight=None$ $max_samples=None$ [55]

À semelhança do que foi desenvolvido para o projeto Sen2Agri (*vide Estado da Arte*), pretende-se que o sistema de classificação aqui desenvolvido assente num algoritmo capaz de lidar com a diversidade das assinaturas espectro-temporais de uma dada cultura devido às diversas datas de sementeira, cultivo e condições atmosféricas de cada zona agroclimática [5] usando dados *in situ* para mapeamento do tipo de cultura agrícola existente.

A estratégia de classificação utilizada baseia-se no objeto, como descrito na secção 3.3.1 *Preparação dos Dados*. Na abordagem usada, baseada na quadrícula do satélite (*Tile-Based*) [56], a classificação resulta da combinação, num único classificador, de vários classificadores RF treinados em vários subconjuntos do conjunto de dados treino. Estes subconjuntos correspondem a cada uma das zonas geográficas (NE, ND e NC) ou combinações entre elas [5]. Após o treino, a classificação dos dados de uma zona foi feita utilizando-se diversas estratégias e avaliando-se os modelos pelos resultados obtidos. Espera-se que usando classificadores distintos treinados em dados de zonas diferentes (exemplo do treino em ‘todas’ e teste numa das zonas – NC, ND ou NE), seguido de um voto de maioria resulte num bom compromisso entre complexidade e exatidão do modelo [56].

Para avaliar e escolher o modelo que melhor se ajusta aos dados (a melhor hipótese), é necessário definir a proporção de erro – o número de vezes que $f(x) \neq Y$ (eq. (2.1)) para uma amostra (x, y) . Contudo, uma baixa proporção de erro na classificação de um conjunto de dados de treino por um modelo não significa que este seja capaz de generalizar bem. Para obter uma avaliação correta do modelo é necessário testá-lo num conjunto de exemplos que ainda não viu.

Para a validação dos modelos, foi utilizado o método *train_test_split* (biblioteca Scikit-learn) com estratificação da amostragem: separação aleatória dos dados em dois conjuntos de treino e de teste na proporção 75/25%, com manutenção da proporção das classes no conjunto total, segundo o método *Holdout Validation* [56] representado na figura 3.5. Os parâmetros do algoritmo são especificados manualmente (o classificador não os aprende a partir dos dados treino durante o ajustamento do modelo) ou são adotados os existentes por defeito na biblioteca em uso. Os dados teste, ainda não vistos pelo modelo, são então utilizados para este prever as etiquetas das classes. Através da comparação das decisões do modelo com as etiquetas teste (dados de referência) estima-se a exatidão da generalização. Depois de obtida esta estimativa de como o modelo se comporta em dados não vistos em treino, usa-se a totalidade dos dados (treino + teste) para melhorar o seu desempenho na classificação.

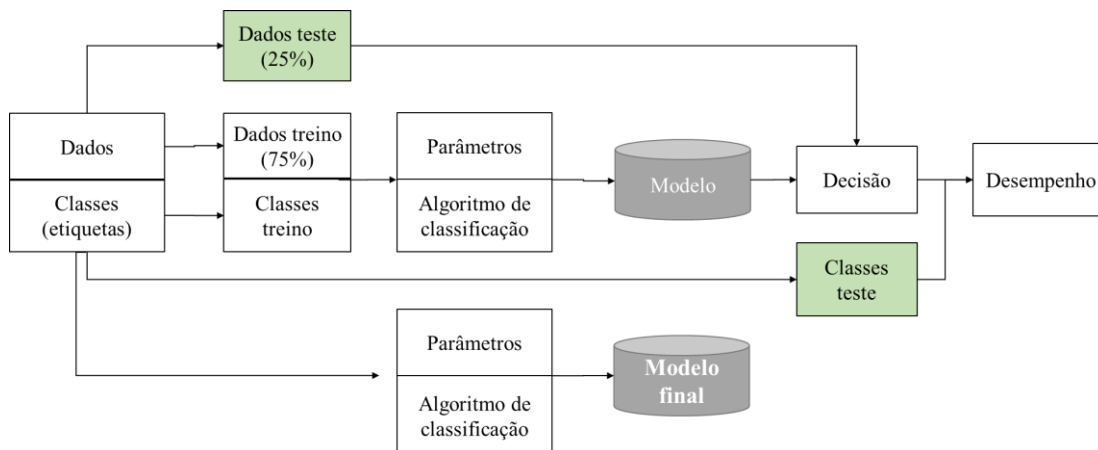


Figura 3.5 – Procedimento de validação dos modelos de classificação – *Holdout Method* [57].

Este método, tem a desvantagem de não garantir o uso de todos os dados disponíveis [58]: por um lado, se se treinar o classificador em apenas metade dos dados poder-se-á gerar uma hipótese muito fraca; por outro, se se reservar apenas 10% dos dados para teste (e os restantes 90% para treino) poder-se-á condicionar a estimativa da exatidão do modelo a um conjunto de dados teste pouco representativos do universo em estudo. Neste trabalho, procurou-se contornar esta desvantagem aumentando o número de vezes em que a separação em dados treino/teste se realizou: o classificador foi treinado com estes conjuntos de dados, em 10 iterações fazendo-se o ajustamento do modelo aos dados 10 vezes, sendo escolhido aquele que melhor desempenho apresentou nos dados teste de acordo com as métricas enunciadas na secção seguinte. O modelo guardado para teste com transferência de aprendizagem resultou da combinação destes dez modelos ajustados ao domínio de treino. Nestes casos, a validação foi feita num domínio diferente daquele em que o modelo foi treinado [39].

3.3.5 Pós-Classificação

3.3.5.1 Avaliação dos modelos de classificação

a. Métricas de avaliação do desempenho de classificadores e matrizes de confusão

A cada modelo treinado e testado no respectivo conjunto de dados foram calculadas métricas de avaliação de desempenho. Geraram-se matrizes de confusão - tabelas de contingência de dupla entrada onde as linhas se referem às classes no terreno (dados de referência) e as colunas são as classes previstas (pelo classificador) – para quantificar a concordância entre a classificação automática e a realidade observada no terreno. Cada célula representa o número de elementos (neste caso, parcelas) que, pertencendo de facto à classe-linha foram classificados pelo modelo na classe-coluna. (De notar que esta é a configuração de matriz de saída, por defeito, da biblioteca Scikit-learn. Outros softwares poderão apresentar as matrizes de confusão em formato transposto relativamente a este -).

Pelos valores das matrizes de confusão calculou-se a exatidão global (EG) de cada modelo, que reflete a frequência com que o modelo está correto - a capacidade deste identificar corretamente as instâncias usadas para teste.

$$EG = \left(\frac{\sum_{i=1}^r n_{ii}}{\sum_{i=1}^r \sum_{j=1}^r n_{ij}} \right) * 100 \quad (3.5)$$

Em que n é o número de parcelas e r é o número de classes.

Os erros de classificação por omissão em cada classe (representados pelos valores fora da diagonal, em cada linha) e os erros de comissão (os valores fora da diagonal em cada coluna) dão uma ideia da qualidade do modelo e da influência que cada classe tem no seu desempenho. Estes são inversamente proporcionais à exatidão do produtor e à exatidão do utilizador, respetivamente.

A exatidão do produtor (EP), ou revocação, refere-se à completude do modelo. É a proporção de parcelas corretamente classificadas em relação ao total das parcelas daquela classe no terreno. Para cada classe de parcelas no terreno, o número de parcelas corretamente classificadas é dividida pelo número total de parcelas daquela classe (total de linha).

$$EP_i = \sum_{j=1}^r \frac{n_{ii}}{n_{ij}} \quad (3.6)$$

Onde n é o número de parcelas, r é o número de classes.

A exatidão do utilizador (EU), ou precisão, é a capacidade do modelo de retornar instâncias relevantes, ou seja, a proporção de parcelas corretamente classificadas em relação à totalidade de parcelas (correta ou incorretamente) classificadas nessa classe (total de coluna).

$$EU_i = \sum_{j=1}^r \frac{n_{ii}}{n_{ji}} \quad (3.7)$$

Onde n é o número de parcelas, r é o número de classes.

Na maioria dos casos, a questão do compromisso possível entre precisão e revocação é uma questão a considerar. Um modelo genérico manterá um equilíbrio entre estas duas métricas.

O F1-score (entre 0 e 1) é média ponderada entre precisão e revocação, dando igual peso aos dois valores (em muitos casos deve, por não ter em conta o conhecimento do âmbito da classificação, ser olhado com cautela, porque os erros de predição têm implicação idêntica). O F1-score é obtido para cada classe (i) e calculada a média aritmética simples para obter um valor único (*macro-averaged* F1-score). Tendo em conta o número de amostras de cada classe, considera-se mais interessante olhar para o F1-score ponderado (*weighted* F1 score).

$$F1\ score = 2 * \frac{EU_i * EP_i}{EU_i + EP_i} \quad (3.8)$$

Por último, o coeficiente Kappa (k) é uma medida de concordância que dá ideia de quanto os valores observados se afastam dos esperados, fruto do acaso, ou seja, permite aferir o desempenho de um classificador que tem a conta a frequência de cada classe relativamente a outro que prevê aleatoriamente. Os seus valores, sempre iguais ou inferiores a 1, são interpretados de acordo com o esquema enunciado por Landis e Koch [59] em que valores < 0 indicam ausência de concordância, ou desempenho muito fraco do modelo; valores entre 0 e 0.20, indicam desempenho fraco; entre 0.21 e 0.40, desempenho razoável; entre 0.41 e 0.60, desempenho moderado; entre 0.61 e 0.80, bom desempenho e, entre 0.81 e 1, desempenho quase perfeito (ou concordância total entre estimadores).

$$k = \frac{N \sum_{i=1}^r n_{ii} - \sum_{i=1}^r (G_i C_i)}{N^2 - \sum_{i=1}^r (G_i C_i)} \quad (3.9)$$

Em que i é a classe; r é o número de classes; N é o número total de elementos classificados que estão a ser comparados à verdade no terreno n_{ii} é o número de elementos pertencentes à classe i no terreno que também foram classificados como tal pelo modelo; C_i é o número total de elementos classificados pelo modelo como pertencendo a i ; G_i é número total de elementos da classe i no terreno.

b. Mapas de Cobertura do Solo – Mapas de Confiança

Adicionalmente às matrizes de confusão, a representação visual dos resultados de classificação é extremamente útil para avaliar a qualidade de desempenho de um modelo de classificação e pode ser fornecida sob a forma de um mapa de cobertura do solo em que cada conjunto de parcelas apresenta um nível de confiança da decisão do classificador (*vide* secção 2.2.2.4 – *Random Forest*). Os valores de confiança da classificação para cada classe, para cada parcela, foram calculados recorrendo ao método *predict_proba* do classificador *Random Forest* (Scikit-learn [55]). Seguidamente, adicionou-se esta informação aos dados georreferenciados (formato vetorial) e desenvolveram-se mapas cujo aspeto visual resulta da decisão do classificador *versus* a conformidade, ou não, com a declaração feita pelo agricultor, usada como referência (*vide* figura 4.15 para exemplo).

3.3.6 Melhoria dos modelos de classificação

Após a avaliação do modelo, procuram-se formas de o melhorar. Sendo este um problema de classificação, suponha-se, à partida, que uma forma de melhorar o modelo seria por calibração dos seus parâmetros ou mesmo intervenção na sua estrutura [36] ou, por outro lado, por melhoria do conjunto de dados treino no sentido de tornar os dados mais equilibrados através da aplicação de técnicas de transferência de aprendizagem.

3.3.6.1 Calibração de parâmetros e reestruturação do classificador

Numa primeira fase, a implementação do algoritmo de classificação escolhido fez-se usando os parâmetros por defeito da biblioteca Scikit-learn v. 0.21.0 (tabela 3.7).

Tabela 3.7 – Valores dos hiperparâmetros do classificador RF usados e respetiva fonte.

Hiperparâmetro	Valor por defeito [55]	Fonte		
		<i>Randomized SearchCV</i> [60]	<i>GridSearch</i> [61]	Outros estudos: <i>Inglada</i> [56]
Nr. árvores de decisão (<i>n_estimators</i>)	10	1000	1000	100
Critério impureza (<i>criterion</i>)	‘gini’	‘gini’	‘gini’	‘gini’
Profundidade máxima de cada árvore (<i>max_depth</i>)	None	40	60	25
Nr. mínimo de amostras para separação (<i>min_samples_split</i>)	2	2	2	5
Nr. amostras por folha (<i>min_samples_leaf</i>)	1	1	3	1
Nr. máximo de variáveis por nó (<i>max_features</i>)	raiz quadrada do nr. total de variáveis	raiz quadrada do nr. total de variáveis	raiz quadrada do nr. total de variáveis	raiz quadrada do nr. total de variáveis

Para o RF, em particular usaram-se os parâmetros do construtor de classe para *sklearn.ensemble.RandomForestClassifier* [55]: 10 árvores de decisão, de profundidade tal que todas as folhas contivessem menos de 2 amostras, e um mínimo de 2 amostras por nó. O número máximo de variáveis selecionadas em cada árvore igualou a raiz quadrada do número total de variáveis tal como sugerido por Breiman [62]. Uma vez que, na prática, a escolha do critério na medição da impureza em cada nó não tem impacto significativo no desempenho do classificador [31] manteve-se o dado por defeito, i.e., ‘gini’.

Como estes parâmetros não são aprendidos pelo classificador através de treino, terão de ser definidos por testes sucessivos com diferentes valores para cada um deles até se encontrar aqueles que conduzem a melhores desempenhos. Por isso, procurou-se afinar os parâmetros do classificador por *RandomizedSearchCV* e por *GridSearch*, duas técnicas do Scikit-learn que testam o desempenho do classificador usando diversos valores dentro de intervalos fornecidos

pelo utilizador. Contudo, dado o elevado tempo de computação e os reduzidos ganhos em termos de desempenho (cerca de 2%), optou-se por adotar os valores propostos por Inglada [56] implementação de RF com 100 árvores de decisão, profundidade máxima de cada árvore de 25 níveis e número mínimo de 5 amostras por nó (*vide*, no capítulo 4 - *Resultados e Discussão*, secção 4.1, tabela 4.2, para informação mais detalhada). O número máximo de variáveis selecionadas em cada nó manteve-se igual à raiz quadrada do número total de variáveis (o valor por defeito e recomendado por Breiman [62]). Para além disso, e seguindo a metodologia de Inglada [56], fez-se o ajustamento do classificador 10 vezes com geração aleatória de conjuntos de treino e de teste a cada iteração para validação e combinação dos resultados de cada RF individual num modelo final.

Os diferentes métodos incluídos no classificador em Scikit-learn permitem ainda obter i) as probabilidades de pertença a uma classe ou ii) o valor da classe mais numerosa estimada para cada amostra dos dados-treino (determinados em cada nó final das árvores de decisão). Os resultados são discutidos no capítulo *Resultados e Discussão*.

3.3.7 Melhoria do conjunto de dados de treino: sobreamostragem e reamostragem seletiva

Em aprendizagem automática, conjuntos de dados desequilibrados resultam em problemas de classificação. Para todos os conjuntos de dados usados neste estudo, o número total de classes representadas varia entre as 26 e as 37 sendo que, em todos eles, mais de 80% das parcelas estão distribuídas por, 3 ou 4 dessas classes (figura 3.2). Neste caso, tem-se uma fração de desequilíbrio em que, em média, apenas 1/5 do total de classes está representado na quase totalidade das amostras usadas para treino do classificador, o que deverá ser tido em conta na validação destes modelos de classificação e no desenvolvimento de estratégias de melhoria dos classificadores considerados.

Classes desigualmente representadas podem causar problemas porque a amostragem aleatória (característica do RF) pode resultar em largas diferenças na proporção das classes sub-representadas nos dados originais relativamente à dos dados treino [31]. Esta situação refletir-se-á nas métricas de avaliação do classificador, resultando em valores de exatidão enviesados pela precisão das classes maioritárias (o chamado ‘paradoxo da exatidão’ [63]). A exatidão global é uma medida útil, mas terá de ser usada com reserva porque, tendo bom desempenho quando há igual número de amostras pertencendo a cada classe (conjunto de dados de treino equilibrado), o seu valor pode dar uma falsa ideia de elevada exatidão. As classes olival, vinha e arroz são as três classes mais representadas e destacam-se claramente das demais. A exatidão global do modelo vai basear-se sobretudo na precisão na classificação destas três classes. No capítulo *Resultados e Discussão*, far-se-á uma análise detalhada desta situação.

Neste estudo, a mitigação do efeito das classes sub-representadas na qualidade do modelo de classificação é uma questão em aberto, para discussão futura. Estratégias possíveis a implementar e testar passam por: i) acumular amostras sub-representadas de ano para ano por seleção ativa e incorporar estas amostras no novo conjunto de dados de treino (à semelhança da metodologia adotada por Demir *et al.* [18] para aprendizagem ativa de classificadores); ii) sobre-amostrar culturas sub-representadas por geração sintética de amostras – por exemplo, pela aplicação da técnica SMOTE (*Synthetic Minority Oversampling Technique*) [64] que opera, criando amostras

sintéticas das classes minoritárias (em vez de replicações simples das mesmas que poderiam gerar sobre-ajustamento do classificador) usando o algoritmo do vizinho mais próximo.

3.4 Quadro resumo dos testes realizados

Após uma bateria inicial de testes exploratórios sobre os dados com diversos algoritmos de classificação usando os valores de refletância disponíveis em cada série para as nove bandas consideradas (T01), dada a origem geográfica e temporal distinta dos dados, verificou-se que seria necessário incluir técnicas para garantir harmonização de variáveis. À semelhança de procedimentos seguidos noutros estudos do mesmo tipo [56], antes da implementação da classificação propriamente dita, fez-se interpolação linear das refletâncias para regularizar a série temporal (T02).

Paralelamente, para prevenir redundância de informação e sobre-ajustamento do classificador aos dados de treino, possibilitar a diminuição no número de dados e do esforço computacional usados na classificação procedeu-se a redução da dimensionalidade dos dados. Numa primeira fase, usaram-se técnicas de PCA e de seleção de variáveis (T03) e, numa segunda fase, os marcadores fenológicos calculados para cada cultura (T09). A criação de conjuntos de treino para os diversos testes, fez-se com dados com as amostras rotuladas usando várias estratégias para combinação entre os anos e zonas geográficas para procurar perceber o comportamento de cada modelo de classificação tendo em conta diversos cenários de disponibilidade dos dados (T04 e T08) e a transferência da aprendizagem entre anos e/ou zonas (T05, T06 e T07). Finalmente, fizeram-se testes para procurar perceber a variação na exatidão de classificação ao longo das várias etapas do ciclo de crescimento da vegetação (T10).

Tabela 3.8 – Quadro resumo dos testes de classificação realizados.

Código do teste	Procedimento	Características	Objetivo
T01	Classificação com diversos algoritmos e calibração do classificador sem reamostragem temporal	Variáveis: dados originais das refletâncias de bandas; algoritmos: SVM, MLP, kNN, RF.	Avaliar o desempenho de diferentes algoritmos populares em classificação de imagem; ponto de partida para classificação dos conjuntos de dados em estudo.
T02	Interpolação Linear e reamostragem da série temporal	Variáveis: refletâncias das bandas (com interpolação temporal a 5 dias); algoritmo: RF.	Perceber impacto da interpolação temporal dos dados na classificação; forma de uniformizar dados para proceder a TA.
T03	Redução da dimensionalidade dos dados por PCA e Seleção de Variáveis	Variáveis: componentes PCA ou variáveis selecionadas; algoritmo: RF.	Encontrar formas eficientes de redução de custos de processamento (tempo de computação) e de manutenção da interpretabilidade dos dados.
T04	Classificação em dados conjuntos (em diferentes proporções)	Variáveis: refletâncias das bandas (com interpolação temporal a 5 dias); junção de dados de cada ano, das três zonas, em proporções diferentes: 30%, 50% e 70%; algoritmo: RF.	No caso de redução na disponibilidade de dados e/ou de necessidade de redução de custos de processamento, perceber que quantidade de amostras é necessário extrair de cada zona geográfica (usadas em treino) para gerar modelos de classificação com desempenho interessante.

Código do teste	Procedimento	Características	Objetivo
T05	TA no espaço com refletâncias de bandas	Variáveis: refletâncias das bandas (com interpolação temporal a 5 dias); Sem redução da dimensionalidade; algoritmo: RF.	Testar modelos em dados de zonas não vistas em treino. Perceber a capacidade de reutilização de conhecimento no espaço sem aplicar técnicas mais específicas de transferência de aprendizagem.
T06	TA no tempo com refletâncias de bandas	Variáveis: refletâncias das bandas (com interpolação temporal a 5 dias); Sem redução da dimensionalidade; algoritmo: RF.	Testar modelos em dados de anos não vistos em treino. Perceber a capacidade de reutilização de conhecimento no tempo sem aplicar técnicas mais específicas de transferência de aprendizagem.
T07	TA no espaço e no tempo com refletâncias de bandas	Variáveis: refletâncias das bandas (com interpolação temporal a 5 dias); Sem redução da dimensionalidade; algoritmo: RF.	Testar modelos em dados de anos e zonas diferentes, não vistos em treino. Perceber a capacidade de reutilização de conhecimento simultaneamente no espaço e no tempo, sem aplicar técnicas mais específicas de transferência de aprendizagem.
T08	TA no tempo com dados interpolados: treino com dados de todas as zonas em diferentes proporções, teste em dados de 2019 em cada zona e em todas as zonas	Variáveis: refletâncias das bandas (com interpolação temporal a 5 dias); Sem redução da dimensionalidade; algoritmo: RF.	Testar modelos treinados em dados limitados em dados de anos não vistos em treino. No caso de indisponibilidade de imagens ou de limitações de poder computacional, avaliar o impacto que a quantidade de dados treino tem no desempenho do classificador, com TA no tempo.
T09	Redução da dimensionalidade dos dados por marcadores fenológicos	Variáveis: marcadores fenológicos; algoritmo: RF.	Testar outros métodos para redução da dimensionalidade dos dados, mais inteligíveis do que os anteriormente usados.
T10	Classificação por etapas do ciclo vegetativo	Variáveis: refletâncias das bandas (com interpolação temporal a 5 dias); algoritmo: RF.	Perceber o desempenho na classificação de modelos gerados com dados treino limitados às diversas fases do ciclo vegetativo.

4 Resultados e Discussão

Nesta secção analisam-se os desempenhos dos modelos de classificação gerados nos testes descritos no capítulo precedente. Os resultados apresentam-se pela ordem do enunciado na tabela 3.8. No final do capítulo, apresentam-se mapas de classificação e de confiança dos modelos com melhor desempenho para inspeção visual das classificações produzidas em termos da confiança atribuída a cada classe e das valências que o algoritmo de classificação usado neste estudo proporciona.

4.1 T01 – Classificação com diversos algoritmos e calibração do classificador sem reamostragem temporal

Inicialmente, foram utilizados os algoritmos de classificação: *Support Vector Machines* (SVM), *Multi-layer Perceptron* (MLP), *K-Nearest Neighbor* (kNN) e *Random Forest* (RF). Verificou-se que os resultados em termos das métricas usadas dos modelos são semelhantes para os algoritmos RF e kNN (EG com valores entre aproximadamente os 78% e os 91% para as diferentes zonas em ambas as datas), sendo registados valores mais baixos para o SVM (cerca de 76% a 85%) e valores mais elevados para o MLP (cerca de 84% a 93%). O que distingue claramente o desempenho dos classificadores são os tempos de processamento por iteração: o algoritmo mais rápido é o RF, seguido do kNN, SVM e MLP, que é consideravelmente superior aos restantes (tabela 4.1).

Tabela 4.1 – Métricas de desempenho e tempos de processamento por iteração dos modelos de classificação com diferentes algoritmos e conjuntos de dados: *Support Vector Machine* (SVM), *Multi-layer Perceptron* (MLP), *K-Nearest Neighbour* (kNN) e *Random Forest* (RF).

Conj. Dados Treino	Algoritmo de Classificação	2018				2019			
		EG (%)	Coef. Kappa (%)	F1-Score (%)	Tempo de processamento (por ref. ao melhor valor, =1)	EG (%)	Coef. Kappa (%)	F1-Score (%)	Tempo de processamento (por ref. ao melhor valor, =1)
NE	SVM	75.5	65.7	72	176	77.4	68.2	74	163
	MLP	83.9	77.9	81.9	421	84.5	78.5	82.4	667
	kNN	78.2	69.9	76.2	31	79.7	71.8	77.8	31
	RF	78.6	70.3	76.1	2	80.4	72.8	78.3	3
ND	SVM	83.7	73.4	80.4	42	85.2	75.6	82.5	71
	MLP	90.9	86.1	90.5	247	92.4	88.2	92	328
	kNN	89.2	82.9	88.2	8	90.5	84.8	89.6	17
	RF	89	82.7	88.1	2	90.3	84.6	89.4	2
NC	SVM	77.3	64.8	73.4	21	76.2	61	71	27
	MLP	91.9	88.3	91.3	87	92.7	89.1	92.1	101
	kNN	90.3	85.9	89.6	4	90.6	85.8	89.9	5
	RF	89.1	84.1	88.4	1	89.5	84.1	88.7	1

Embora os algoritmos kNN e RF apresentem resultados muito semelhantes para as três zonas em estudo e o algoritmo MLP mostre os melhores resultados de classificação, considerou-se que o RF é o algoritmo que melhor conjuga a qualidade de desempenho e o tempo de computação nos diversos conjuntos de dados (de notar que o tempo de processamento medido corresponde sobretudo a tempo de treino para todos os classificadores, com exceção do kNN, que apresenta um tempo de predição muito superior ao de treino). O RF apresenta também (como abordado na secção 2.2.2) valências que são úteis para a compreensão dos dados, como a possibilidade de calcular confianças de classificação e importância de variáveis explicativas na discriminação dos dados e, portanto, foi aquele pelo qual se optou para a construção dos modelos de classificação.

A mudança nos parâmetros de RF de acordo com o descrito na secção *Metodologia (Pré-processamento dos dados)* levou a uma melhoria destes resultados na ordem dos 1% a 3% na EG, resultados muito próximos dos valores obtidos por aplicação do algoritmo com melhor desempenho testado em 4.1 – T01 (MLP), pelo que se passou à implementação do RF com os novos valores e à geração de modelos combinados resultantes de ciclos de 10 iterações sobre os dados treino (tabela 4.2).

Tabela 4.2 – Métricas de desempenho dos modelos de classificação usando *Random Forest* (RF) com diferentes valores de parâmetros.

Algoritmo RF (Parâmetros)	Conjunto de Dados Treino	2018			2019		
		EG (%)	Coef. Kappa (%)	F1-Score (%)	EG (%)	Coef. Kappa (%)	F1-Score (%)
Estimadores = 10	NE	78.6	70.3	76.1	80.4	72.8	78.3
Min_samples_split = 2	ND	89.0	82.7	88.1	90.3	84.6	89.4
Max_depth = 'None'	NC	89.1	84.1	88.4	89.5	84.1	88.7
Estimadores = 100	NE	81.4	74.1	78.3	82.5	75.6	80.0
Min_samples_split = 5	ND	89.9	84.1	88.8	90.9	85.7	89.9
Max_depth = 25	NC	90.7	86.5	89.9	90.7	86.1	89.9

4.1.1 Análise das Matrizes de Confusão

As tabelas 4.3 e 4.4 mostram parte das matrizes de confusão para as classificações com os melhores resultados da EG realizadas em 2018 e 2019, na classificação da zona ND.

Por impossibilidade de representação da totalidade das classes e para auxílio na discussão de resultados, representam-se gráficos (figura 4.1) para as métricas de precisão e de revocação das classes mais representadas em cada amostra teste (para as matrizes com as classes representadas em mais do que 40 parcelas *vide* Anexo A).

A figura 4.1 e respetivas matrizes de confusão permitem realçar que os resultados da EG para 2018 e 2019 estão próximos dos 90% e que os coef. Kappa e F1-score são muito semelhantes; as classes com valor mais elevado de F1-score (maior precisão e revocação, simultaneamente), para ambos os modelos, são: o arroz, o milho, o olival e o tomate - entre 92% e 99%; a vinha apresenta valores na ordem dos 70% a 80% , respetivamente, para a precisão e a revocação devido à confusão com olival. O mesmo acontece em sentido inverso, mas em menor escala, tal como refletem os valores destas duas métricas para a classe olival – na ordem dos 90%.

Em classes com representatividade próximas de 1%, tem-se o pimento, com percentagens de revocação e precisão relativamente baixas (na ordem dos 65%). Algumas classes apresentam uma precisão significativamente superior à revocação: tais como, o girassol, a melancia, o melão, o sorgo e as outras hortícolas. Destacam-se o sorgo, com revocação próxima dos 30% e precisão próxima dos 90%, e as outras hortícolas, em que a revocação está abaixo dos 5% em contraste com uma precisão próxima dos 100% para 2018 e 60% para 2019. Isto deve-se a alguma confusão com as classes de olival e vinha – parcelas que são, na realidade, de sorgo e de outras hortícolas são frequentemente omitidas, por classificação em olival e vinha.

Para o ano de 2018, os valores de revocação mais baixos verificam-se para as classes outras hortícolas, melancia, melão e sorgo, mantendo-se sempre abaixo de 34%. O sorgo é confundido com olival na mesma proporção em que é classificado como o próprio e as restantes classes são demasiado escassas nos dados para se poderem tirar ilações sobre os seus valores.

Tabela 4.3 – Matriz de confusão para a classificação realizada em dados de 2018, zona ND. Colunas representam a classificação e linhas representam os dados de referência. Os valores estão em absoluto, exceto quando indicado por %. A diagonal representa o número de parcelas classificadas corretamente pelo modelo.

	ARROZ	BAT	MILHO	OLIV	HORT	PIM	TOM	VINHA	Total	Freq (%)	Revocação (%)	F1-score (%)
ARROZ	852	0	9	6	0	0	0	0	867	7.74	98.27	99
BAT	0	82	10	11	0	0	8	12	124	1.11	66.13	74
MILHO	2	2	1295	32	0	1	11	26	1372	12.25	94.39	94
OLIV	1	1	4	6016	0	0	0	144	6167	55.07	97.55	94
HORT	0	1	7	31	5	4	6	37	100	0.89	5	10
PIM	0	0	4	1	0	52	8	12	78	0.7	66.67	66
TOM	1	0	4	3	0	1	584	5	598	5.34	97.66	92
VINHA	0	0	8	438	0	2	0	1016	1464	13.07	69.4	73
Total	857	97	1374	6603	5	79	676	1312	11199			
Precisão (%)	99.42	84.54	94.25	91.11	100	65.82	86.39	77.44	100			

EG = 89.9%, Coef. Kappa = 84.1%, F1-score = 88.8%

Tabela 4.4 – Matriz de confusão para a classificação realizada em dados de 2019 da zona ND. Colunas representam a classificação e linhas representam os dados de referência. Os valores estão em absoluto, exceto quando indicado por %. A diagonal representa o número de parcelas classificadas corretamente pelo modelo.

	ARROZ	BAT	MILHO	OLIV	HORT	PIM	TOM	VINHA	Total	Freq (%)	Revocação (%)	F1-score (%)
ARROZ	852	0	5	2	0	0	1	4	864	7.76	98.61	99
BAT	0	110	12	9	1	0	5	5	144	1.29	76.39	77
MILHO	2	4	1282	48	0	1	8	25	1375	12.34	93.24	94
OLIV	0	1	4	6043	0	0	0	113	6161	55.3	98.08	95
HORT	0	4	11	26	4	4	2	32	87	0.78	4.6	6
PIM	0	0	6	2	0	54	2	4	70	0.63	77.14	72
TOM	0	0	6	1	0	2	544	3	556	4.99	97.84	93
VINHA	0	0	5	410	0	0	4	1052	1471	13.2	71.52	76
Total	855	129	1361	6601	7	70	617	1292	11141			
Precisão (%)	99.65	85.27	94.2	91.55	57.14	77.14	88.17	81.42	100			

EG = 90.9%, Coef. Kappa = 85.7%, F1-score = 89.9%

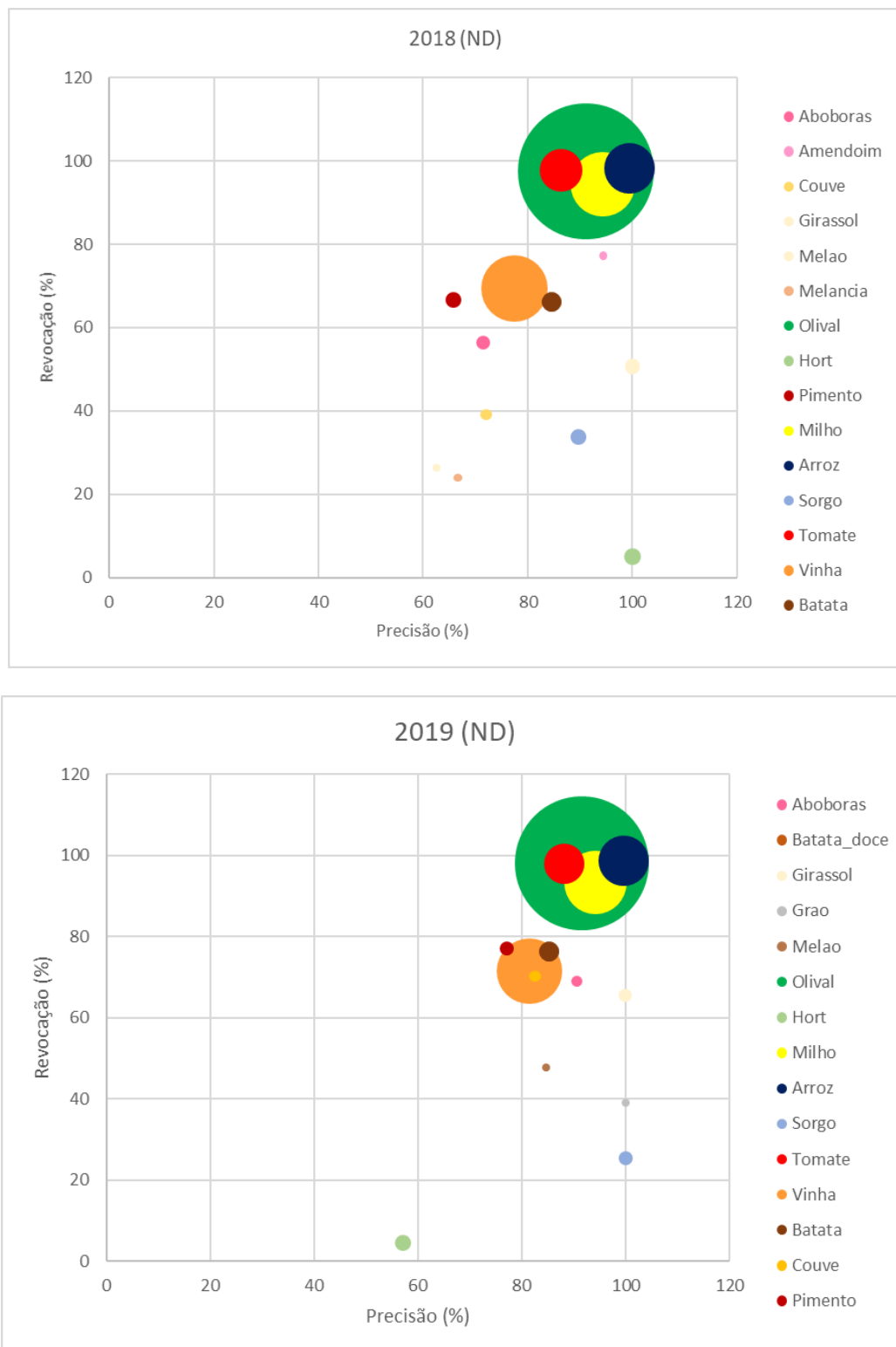


Figura 4.1 - Precisão vs revocação para as classes mais representadas em cada amostra teste de 2018 e 2019 (zona ND). A dimensão dos círculos ilustra a frequência de cada classe nos dados.

Importância de variáveis

Para os dados de 2018 e de 2019 da região ND, selecionaram-se as variáveis cujos valores de importância relativa são iguais ou superiores a 0.02, tal como mostra o gráfico da figura 4.2.

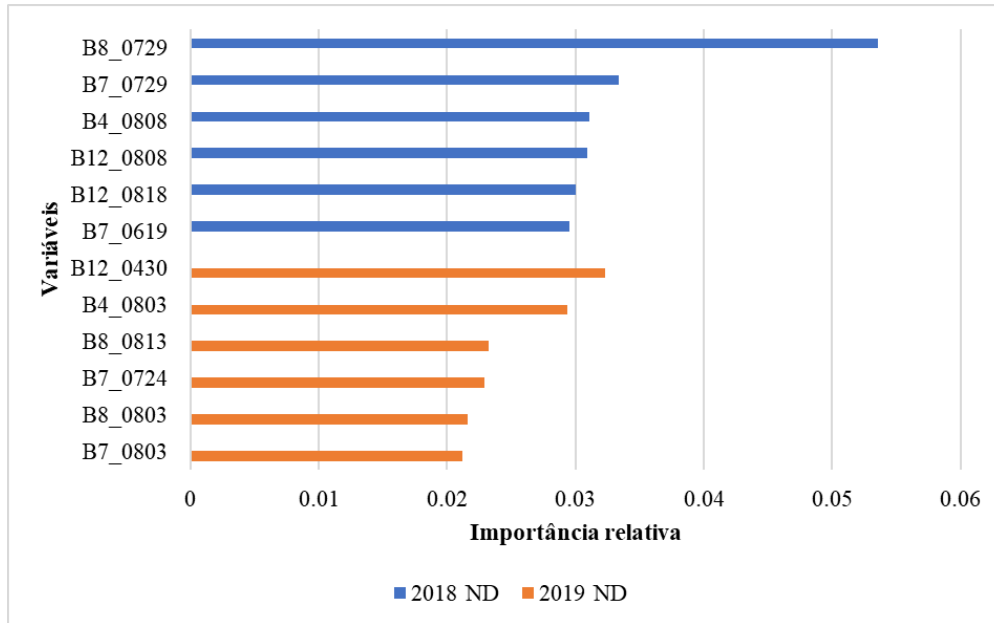


Figura 4.2 – Importância relativa das variáveis na classificação com RF para dados de 2018 (a azul) e de 2019 (a laranja), zona ND. As denominações das variáveis dizem respeito à banda, mês e dia de aquisição da imagem, respetivamente.

De entre as seis variáveis mais importantes para cada um dos conjuntos de dados, não existe uma comum aos dois, embora a maioria corresponda a valores de refletâncias de imagens adquiridas nos meses de Julho e Agosto (final da estação de crescimento).

As variáveis mais importantes considerando ambos os modelos (treinados nos dois conjuntos de dados distintos), ou seja, aquelas que mais contribuem para a separação de culturas diferentes são os valores de refletância correspondentes às bandas 7 e 8 (B7, limiar do vermelho, B8, infravermelho próximo) do dia 29 de Julho de 2018, seguidas de valores de refletância para a banda 12 (infravermelho de onda curta) de 30 de Abril de 2019 e para a banda 4 (vermelho) do dia 08 de Agosto de 2018.

4.2 T02 – Interpolação linear e reamostragem da série temporal

A interpolação linear simples, em intervalos regulares de 5 dias, resultou no aumento do número de variáveis, em média, 2.5 vezes em relação ao do conjunto inicial de variáveis, uma vez que se ficou com o equivalente a 31 imagens para cada banda, passando a 279 variáveis para todos os conjuntos de dados. Os modelos de classificação apresentam os valores de EG descritos na tabela 4.5.

Tabela 4.5 – Métricas de desempenho dos modelos de classificação para os dados originais e interpolados linearmente a 5 dias.

Conjunto de Dados Treino		Variáveis originais	Interpolação temporal (5 dias)		
Ano	Zona	EG (%)	EG (%)	Coef. Kappa (%)	F1-score (%)
2018	NE	81.4	81.0	73.6	78.0
	ND	89.9	89.8	84.0	88.9
	NC	90.7	90.3	85.9	89.4
	todas	NA*	85.3	79.2	83.5
2019	NE	82.5	82.2	75.1	79.6
	ND	90.9	90.9	85.6	89.9
	NC	90.7	90.6	85.9	89.9
	todas	NA*	86.0	80.0	84.4

A interpolação temporal modificou as observações, podendo introduzir erros que resultam numa diminuição da exatidão das classificações. No entanto, a EG dos modelos treinados em dados interpolados, quando diminuiu, foi de forma pouco significativa (máximo de 0.4%), pelo que esta operação permitiu, de modo robusto, uniformizar o número de variáveis entre os diferentes conjuntos de dados mantendo a sua integridade e foi utilizada para treino dos modelos subsequentes.

4.2.1 Análise das Matrizes de Confusão

As tabelas 4.6 e 4.7 mostram parte das matrizes de confusão para as classificações com melhores resultados de EG realizadas em 2018 e 2019, na classificação em todas as zonas.

Por impossibilidade de representação da totalidade das classes e para auxílio na discussão de resultados, representam-se gráficos (figura 4.3) para as métricas de precisão e de revocação das classes mais representadas em cada amostra teste (para as matrizes com as classes representadas em mais do que 40 parcelas *vide* Anexo A).

Tabela 4.6 – Matriz de confusão para a classificação de dados de 2018 de todas as zonas, com dados interpolados (a 5 dias). Colunas representam a classificação e linhas representam os dados de referência. Os valores estão em absoluto, exceto quando indicado por %. A diagonal representa o número de parcelas classificadas corretamente pelo modelo.

	ARROZ	BAT	GIRASS	MILHO	OLIV	TOM	VINHA	Total	Freq (%)	Revocação (%)	F1-score (%)
ARROZ	2983	0	0	44	23	0	5	3055	9.24	97.6	98.0
BAT	0	153	1	228	73	6	98	569	1.72	26.9	38.0
GIRASS	1	2	289	27	40	17	20	397	1.2	72.8	82.0
MILHO	31	32	2	5734	221	10	178	6230	18.8	92	87.0
OLIV	3	5	6	100	14068	1	553	14742	44.6	95.4	91.0
TOM	0	2	0	28	8	608	3	651	1.97	93.4	90.0
VINHA	2	3	0	103	1410	0	4064	5585	16.9	72.8	75.0
Total	3030	240	310	6900	16174	703	5255	33066			
Precisão (%)	98.5	63.8	93.2	83.1	86.98	86.49	77.34				

EG = 85.3%, Coef. Kappa = 79.2%, F1-score = 83.5%

Tabela 4.7 – Matriz de confusão para a classificação de dados de 2019 de todas as zonas, com dados interpolados (a 5 dias). Colunas representam a classificação e linhas representam os dados de referência. Os valores estão em absoluto, exceto quando indicado por %. A diagonal representa o número de parcelas classificadas corretamente pelo modelo.

	ARROZ	GIRASS	MILHO	OLIV	TOM	VINHA	Total	Freq (%)	Revocação (%)	F1-score (%)
ARROZ	2929	0	39	15	0	1	2986	9.07	98.09	99.0
GIRASS	2	236	9	41	14	13	316	0.96	74.68	83.0
MILHO	12	1	5706	294	10	210	6274	19.06	90.95	89.0
OLIV	0	1	83	14235	0	592	14917	45.31	95.43	91.0
TOM	1	1	11	4	585	9	614	1.87	95.28	90.0
VINHA	1	0	85	1407	1	4078	5577	16.94	73.12	75.0
Total	2955	250	6609	16401	679	5290	32920			
Precisão (%)	99.1	94.4	86.34	86.79	86.2	77.09	100			

EG = 86%, Coef. Kappa = 80%, F1-score = 84.4%

Considerando apenas as classes representadas em mais do que 40 parcelas, por questões de visualização, nos dados de ambos os anos, as percentagens de revocação e de precisão por classe mantêm-se semelhantes: as classes com valor mais elevado de F1-score para 2018, 2019, são: o arroz (98%, 99%), o olival (91%), tomate (90%) e milho (87%, 89%). Estas classes (à exceção do tomate) são, também, as mais representadas nas amostras de teste.

Para os dois anos, as classes que apresentam maior número de erros de comissão (complementares dos valores de precisão) são outras hortícolas (com uma percentagem de erros próxima dos 60%), a batata e o pimento (~30%) e a vinha (com cerca de 23% de erros de comissão).

Quanto aos erros de omissão (complementares dos valores de revocação), os valores mais elevados correspondem às parcelas com outras hortícolas, feijão, abóboras e sorgo. Estas classes representam, conjuntamente, apenas 3% dos dados. Relativamente às classes mais representativas, destaca-se a vinha, com cerca de 28% de erros de omissão maioritariamente atribuídos a olival. Todas as outras classes, com representatividade próxima e superior a 10% apresentam percentagens de revocação acima dos 90%.

Dada a sua baixa representatividade, a classe das outras hortícolas destaca-se como sendo a que maior confusão apresenta: cerca de 60% das parcelas classificadas como tal pertencem, no terreno, a outras classes (nomeadamente olival e vinha) e cerca de 97% das realmente existentes são atribuídas a outras classes. Portanto, mais uma vez, a confusão entre outras hortícolas e olival e vinha é marcada. De facto, olhando para o gráfico da figura 4.6 constata-se que esta é a classe que mais facilmente se mistura com as restantes, dificultando o processo de classificação.

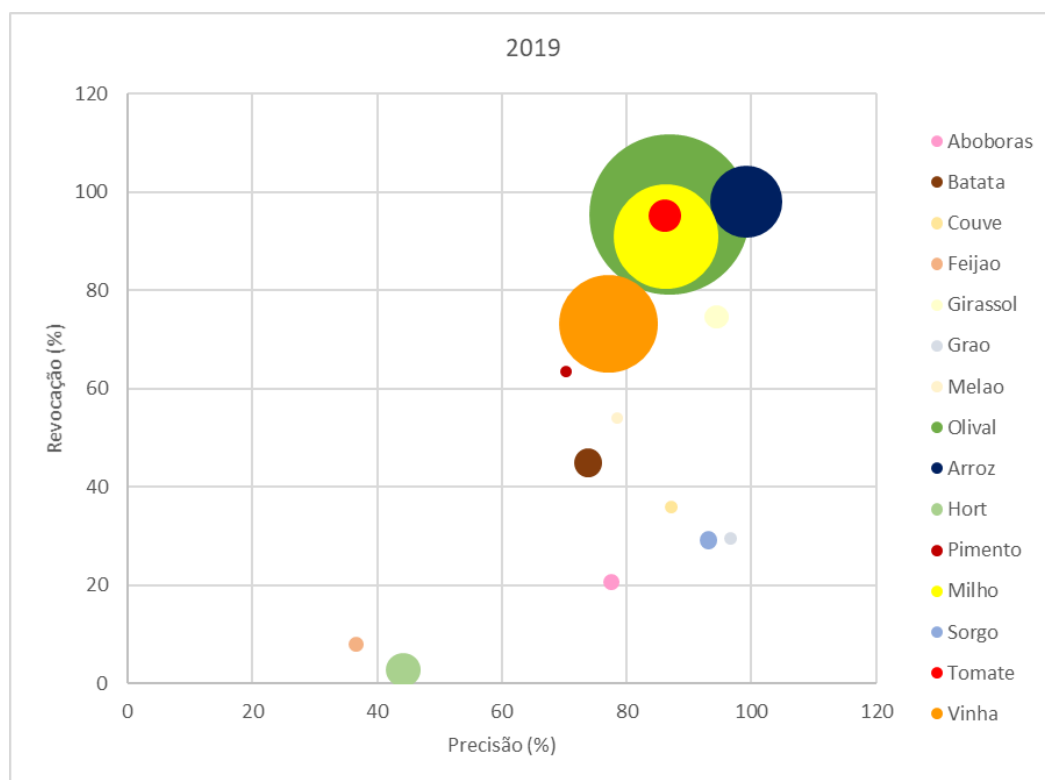
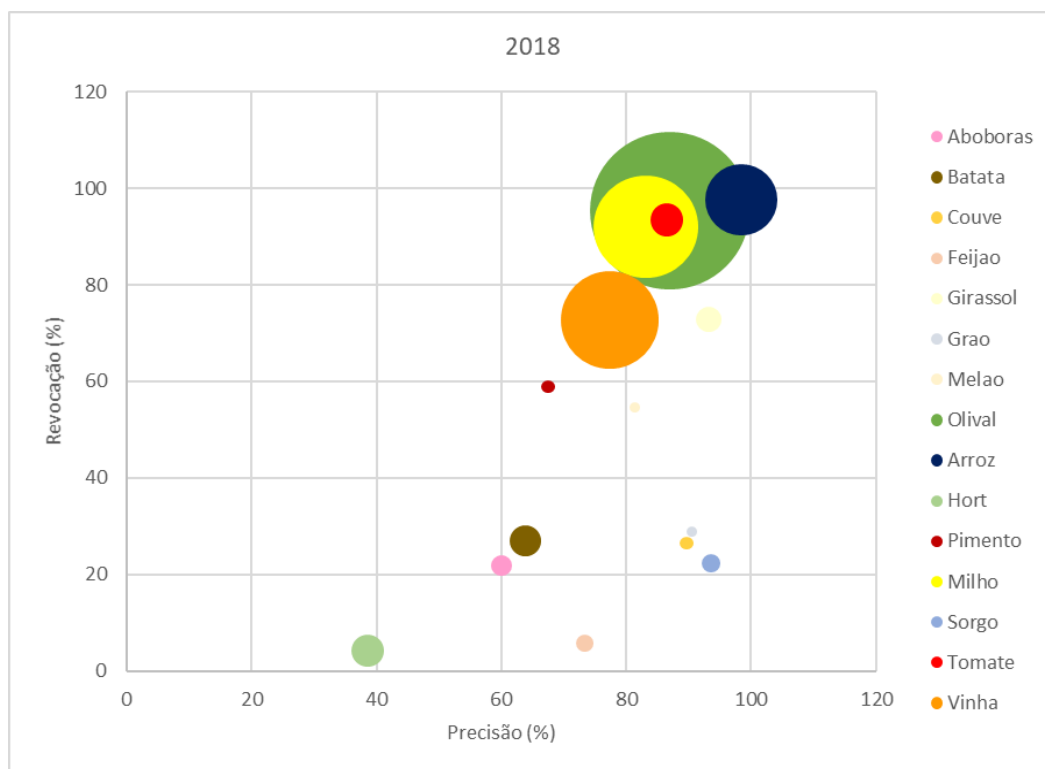


Figura 4.3 – Precisão vs revocação para as classes mais representadas em cada amostra teste de 2018 e 2019 (todas as zonas). A dimensão dos círculos ilustra a frequência de cada classe nos dados.

4.2.2 Importância das variáveis para cada ano

Na figura 4.4, representa-se a importância relativa das seis variáveis mais importantes na classificação dos dados para cada um dos anos estudados, para todas as zonas.

As variáveis mais importantes para a discriminação das culturas agrícolas em 2018 e 2019 não são totalmente concordantes, apesar de as seis variáveis mais importantes para os dois conjuntos de dados corresponderem aos valores das bandas do infravermelho próximo (B8), do vermelho (B4) e do limiar do vermelho (B7) em diferentes dias dos meses de Julho e Agosto (fase final do ciclo de cultura de verão), tal como se tinha verificado para a classificação em dados não interpolados.

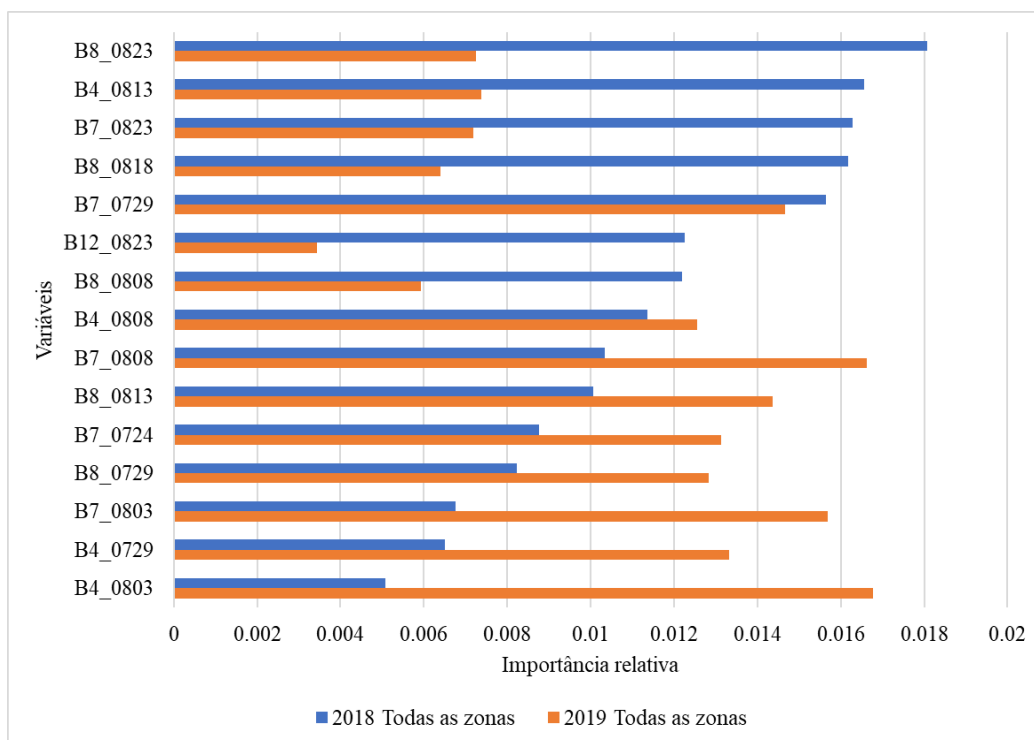


Figura 4.4 - Importância relativa das variáveis na classificação com RF para dados de 2018 e de 2019, todas as zonas.

Há duas variáveis que reúnem consenso na escala de importância relativa: B7_0729 (29 de Julho) e B4_0808 (08 de agosto). O gráfico mostra que a variação na importância de cada variável é gradual, em que as seis primeiras de cada conjunto de dados apresentam valores próximos entre si.

Na figura 4.5, representa-se a importância relativa das seis variáveis mais importantes na classificação dos dados para cada um dos anos estudados, para todas as zonas. Para o conjunto de dados correspondente à junção entre os dois anos e as três zonas em estudo, as doze variáveis mais importantes na discriminação das culturas agrícolas correspondem, na sua maioria, aos valores de refletância das imagens do mês de Agosto, para as bandas B4, B7 e B8. Com valor intermédio de importância, aparece a imagem adquirida em 30 de Abril para a banda B11 e, com o valor mínimo deste grupo, a imagem de 13 de Agosto da banda B12 (figura 4.5).

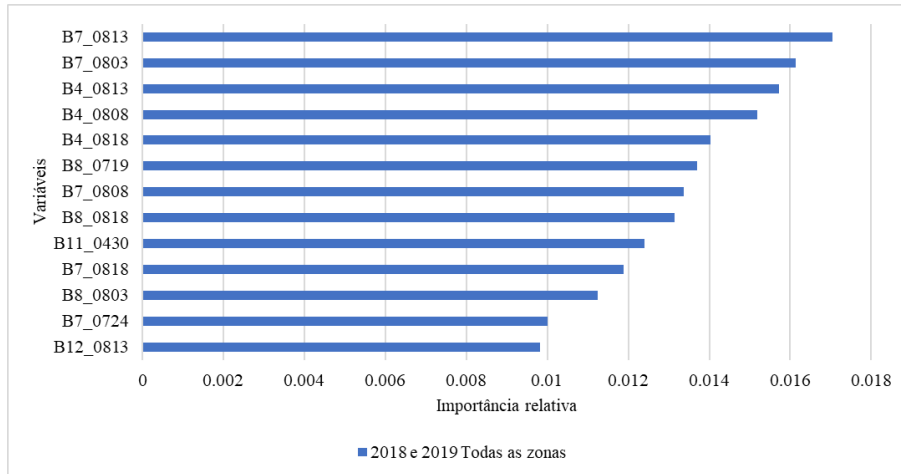


Figura 4.5 - Importância relativa das variáveis na classificação com RF para dados de 2018 e de 2019, todas as zonas.

Estas variáveis dão pistas importantes para perceber a confusão do classificador ao tentar discriminar culturas entre si. Gerar gráficos da série temporal para cada uma destas bandas, para cada cultura considerada, permite perceber a distância a que estas se encontram entre si, dentro das variáveis (refletâncias das bandas) que melhor as diferenciam. Veja-se, na figura 4.6, o perfil de refletância para cada cultura, obtido pela média dos valores de refletância para os dois anos considerados, para todas as zonas geográficas e para cada uma das bandas na série temporal:

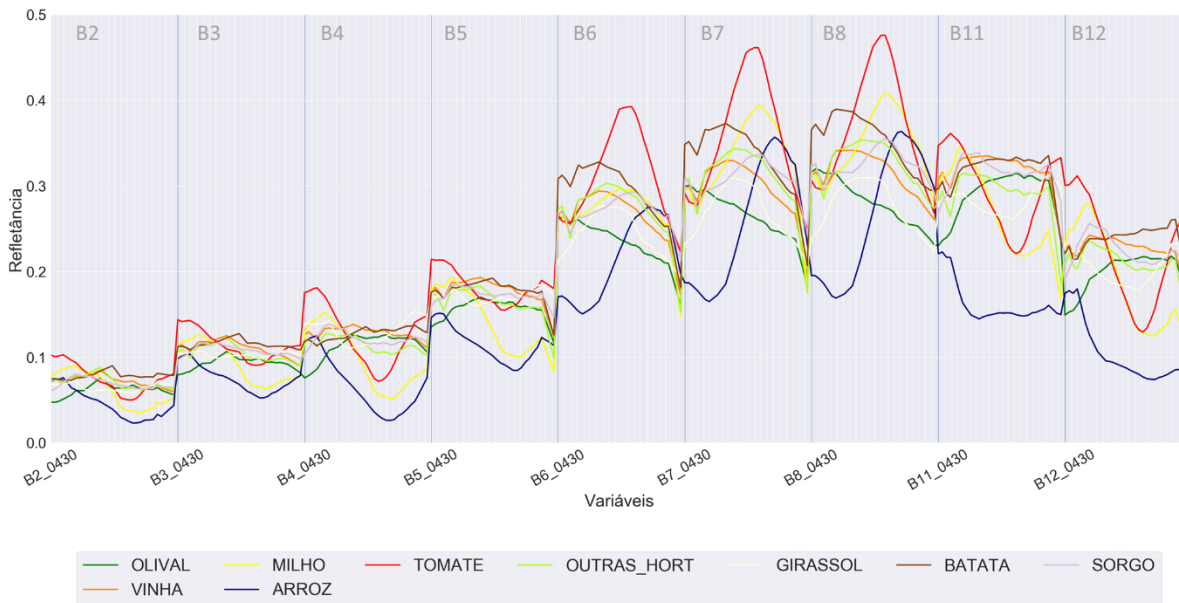


Figura 4.6 – Gráfico de refletâncias das bandas usadas para a identificação das culturas mais representativas em análise.

Aqui, pode ver-se um padrão regular para todas as culturas, em ciclos regulares ao longo do tempo que correspondem às 9 bandas analisadas na série temporal e aos seus valores de refletância nas datas estudadas (entre 30 de Abril e 27 de Setembro). Destacam-se, imediatamente, das restantes, as curvas a azul (arroz), a vermelho (tomate) e a amarelo (milho) nas bandas 6, 7 e 8; as curvas a laranja (vinha), a verde claro (outras hortícolas) e a verde escuro (olival) apresentam um comportamento semelhante, sendo mais difíceis de distinguir

visualmente. Esta dificuldade reflete-se na elevada confusão na classificação destas parcelas em particular, o que se refletirá nos valores de precisão e revocação para cada classe.

Olhando para cada uma das bandas em particular (figura 4.7) confirma-se que a distância entre as diversas culturas em termos de valores de refletância é maior para o final de Julho e mês de Agosto, ou seja, próximo do final do ciclo de crescimento.

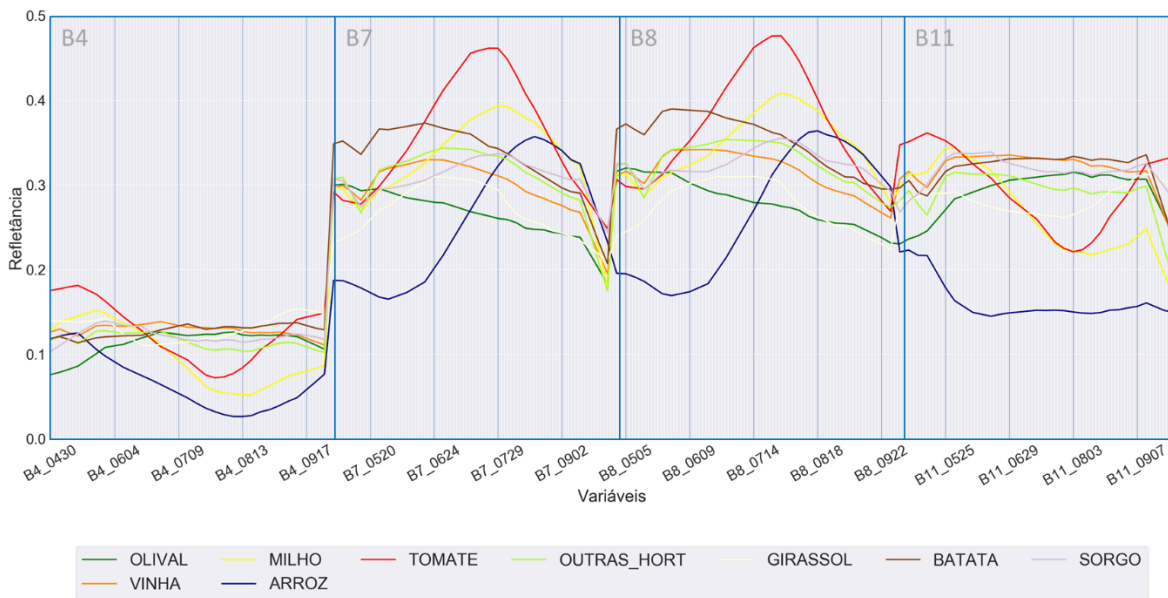


Figura 4.7 – Gráficos das refletâncias das bandas de maior importância na classificação para o conjunto total de dados, para as culturas agrícolas mais representativas.

4.3 T03 – Redução da dimensionalidade dos dados por PCA e Seleção de Variáveis (com interpolação)

4.3.1 Análise de componentes principais (PCA)

O procedimento de PCA aplicado aos dados interpolados resultou numa diminuição de 279 para 30 variáveis (componentes principais) e conduziu aos resultados de EG de classificação enumerados na tabela 4.8.

Tabela 4.8 – Resultados de EG dos modelos de classificação (com RF) desenvolvidos usando os dados originais interpolados a 5 dias e reduzidos por PCA.

Conjunto de Dados Treino-Teste (dados interpolados a 5 dias)		EG (%)	
Ano	Zona	Sem redução de dimensionalidade	Com PCA
2018	NE	81.0	80.8
	ND	89.8	89.7
	NC	90.3	89.5
	todas	85.3	83.7

Conjunto de Dados Treino-Teste (dados interpolados a 5 dias)		EG (%)	
Ano	Zona	Sem redução de dimensionalidade	Com PCA
2019	NE	82.2	82.1
	ND	90.9	90.2
	NC	90.6	91.4
	todas	86.0	85.0

Para a maioria dos modelos, os resultados de EG decresceram com a redução da dimensionalidade, como era esperado, embora de forma pouco significativa. Apesar de os dados serem bastante interessantes (sempre acima dos 80%), o facto de a técnica de PCA não possibilitar uma análise mais aprofundada das variáveis no processo de classificação conduziu à experimentação com técnicas de redução de dimensionalidade alternativas.

4.3.2 Seleção de Variáveis

O método de seleção de variáveis nos dados interpolados conduziu a resultados 5 a 7% inferiores aos resultados obtidos por PCA (tabela 4.9).

Tabela 4.9 – Resultados de EG dos modelos de classificação desenvolvidos usando os dados originais interpolados a 5 dias e reduzidos por Seleção de Variáveis.

Dados Treino-Teste (interpolados a 5 dias)		EG (%)		
Ano	Zona	Sem redução de dimensionalidade	Com PCA	Com Seleção de Variáveis
2018	NE	81.0	80.8	74.0
	ND	89.8	89.7	84.7
	NC	90.3	89.5	83.7
	todas	85.3	83.7	78.1
2019	NE	82.2	82.1	74.6
	ND	90.9	90.2	85.5
	NC	90.6	91.4	83.4
	todas	86.0	85.0	78.6

A opção pela escolha das variáveis comuns ao total dos dados (anos e zonas geográficas), conduziu a dados de treino com 24 variáveis do total de 279 resultantes da interpolação temporal, que foram usadas na totalidade para treino destes modelos. Estas variáveis estão enumeradas na tabela 4.10, e apontam, genericamente, para os valores de refletância das bandas 7, 8, 11 e 12, nas datas de aquisição das imagens em meados e finais de Agosto na sua grande maioria.

Tabela 4.10 – Variáveis selecionadas para treino de modelos de classificação por Seleção de Variáveis para todos os conjuntos de dados interpolados a 5 dias.

Variáveis selecionadas (dados interpolados)	Mês de aquisição
B7_0609 B8_0609	Junho
B11_0709 B11_0719	Julho
B4_0813 B7_0813 B7_0818 B7_0823 B7_0828 B8_0818 B8_0823 B8_0828 B11_0803 B11_0808 B11_0813 B11_0818 B11_0823 B11_0828 B12_0808 B12_0813 B12_0818 B12_0823	Agosto
B7_0902 B11_0902	Setembro

Estes dados confirmam o anteriormente observado relativamente à importância das variáveis na classificação dos dados interpolados: o grupo de variáveis mais importante, antes da classificação está nos valores de refletância das bandas 4, 7, 8, 11 e 12, particularmente nas imagens adquiridas no mês de Agosto.

4.4 T04 - Classificação em dados conjuntos (em diferentes proporções)

A junção dos dados das diferentes zonas, em diferentes proporções, de acordo com a figura 3.4 resultou em vários conjuntos de dados cuja dimensão se apresenta na tabela 4.11. Por exemplo, a primeira linha ('30%') indica que 39675 é o número de amostras treino resultante da junção de apenas 30% do número total de amostras de todas as zonas de 2018 (coluna '2018 todas'); 39500 é o número de amostras treino resultante da junção de apenas 30% do número total de amostras de todas as zonas de 2019 (coluna '2019 todas'); 79175, 30% do número total de amostras de ambos os anos de todas as zonas (coluna '2018 2019 todas') e assim sucessivamente para diferentes proporções. A representatividade das classes mantém-se nestes conjuntos de dados parciais. Como descrito na secção 3.3.3 e representado na figura 3.4, estes dados treino serviram para a aprendizagem de modelos que foram testados nas amostras teste em proporção 70%, 50% e 30% (as não vistas em treino).

Tabela 4.11 - Dimensão dos dados treino resultantes de combinação em proporções de 30%, 50% e 70% dos dados originais.

Dimensão dos dados treino	2018 todas	2019 todas	2018 2019 todas
70%	92576	92168	184744
50%	66126	65835	131961
30%	39675	39500	79175

Os resultados de desempenho para estes modelos (Tabela 4.1.12), mostram que a redução nos valores das principais métricas é mínima - apenas 1% - quando se aumenta a dimensão dos dados treino do classificador de 30% para 70% de amostras. O desempenho das classificações é robusto (sempre superior a 80% em EG) independentemente do número de amostras usadas para treino dos modelos, o que pode ser vantajoso em caso de escassez ou ausência de dados de treino. Por

referência aos valores obtidos aos modelos treinados com a totalidade dos dados disponíveis (explorados na secção 4.2 – T02), os resultados melhoram em apenas dois pontos percentuais, mas o tempo de computação aumenta significativamente.

Tabela 4.12 – Métricas de desempenho dos modelos de classificação para os dados interpolados com a totalidade das amostras para todas as zonas e para os dados das zonas concatenados em diferentes proporções.

Dados Treino (proporção)	2018			2019			2018 e 2019		
	EG (%)	Coef. Kappa (%)	F1- score (%)	EG (%)	Coef. Kappa (%)	F1- score (%)	EG (%)	Coef. Kappa (%)	F1-score (%)
70%	84.4	77.9	82.6	85.5	79.3	83.8	84.7	78.3	82.9
50%	82.4	74.8	80.4	85.1	78.7	83.3	84.3	77.6	82.4
30%	83.5	76.6	81.4	84.5	77.8	82.6	83.8	76.9	81.8

Na secção 4.8 – T08, apresentam-se os resultados relativos à utilização destes modelos com TA no tempo (modelos treinados com dados de 2018 e testados em dados de 2019).

4.5 T05 – Transferência da aprendizagem no espaço com refletâncias de bandas

Quando se procura fazer a transferência da aprendizagem (TA) do modelo no espaço, ou seja, usar o modelo para classificar dados de uma zona geográfica distinta daquela em que foi treinado, para o mesmo ano, os resultados da EG diminuem, em maior valor para as zonas geográficas NC e NE, as mais distantes geograficamente (tabela 4.13). No entanto, a EG para todas os modelos é superior aos 65.5% o que, não sendo um valor elevado, é interessante, considerando que não foi aplicada nenhuma técnica específica de transferência de aprendizagem, quer ao nível da estrutura do RF (*vide* secção 2.2.3 e [36]), quer ao nível da reamostragem dos dados treino (*vide* secções 2.2.4 e 2.2.5 e [34], [35], [38] e [41]).

O modelo que produz os melhores resultados é, para ambos os anos, o treinado em dados de NC e testado em dados de ND, o que indica que estas duas zonas apresentarão maior semelhança entre si do que quaisquer das restantes.

Tabela 4.13 - Métricas de desempenho de classificadores treinados numa zona e testados noutra (para 2018 e para 2019).

Domínio- fonte	Domínio-alvo	Com TA no espaço			
		EG (%)	Coef. Kappa (%)	F1-score (%)	
2018					
2018	NE	ND	80.0	68.8	76.6
		NC	75.9	63.4	71.6
	ND	NE	73.5	63.7	71.6
		NC	78.4	66.8	74.3
	NC	NE	65.5	54.7	65.7
		ND	80.0	70.4	80.2

Domínio-fonte	Domínio-alvo	EG (%)	Coef. Kappa (%)	F1-score (%)	
2019					
2019	NE	ND	79.6	68.5	77.1
		NC	74.3	61.0	71.2
	ND	NE	73.1	62.9	70.9
		NC	80.0	68.8	76.6
	NC	NE	69.9	59.3	69.0
		ND	83.8	75.0	83.4

O valor de EG mais baixo registado (65.5%) é o do modelo treinado em dados de 2018 NC e testado em 2018 NE. De modo geral, os resultados são tanto melhores quanto mais próximas se situam as zonas correspondentes aos dados do domínio-fonte e domínio-alvo.

Modelos treinados unicamente com dados da zona NC ou ND e testados em ND ou NC, respetivamente, classificam com uma EG superior a 78%, ou seja, o classificador necessita de dados de apenas umas das zonas para classificar a outra com uma exatidão relativamente robusta, o que é um resultado interessante. Este resultado também é indicativo de que NC e ND, sendo zonas geograficamente mais próximas apresentam maior homogeneidade espectral entre si (menor variabilidade intraclasse). A transferência da aprendizagem do classificador entre estas apresenta valores mais elevados de EG do que entre qualquer uma delas e NE.

Estes valores são baixos para serem considerados na escolha do modelo mais adequado aos dados de que se dispõe, mas permitem compreender os dados e delinear formas de reamostragem para melhoria dos modelos de classificação. As amostras corretamente classificadas poderão ser, por exemplo, posteriormente incorporados nos dados treino no desenvolvimento de um modelo ‘compósito’ deste.

4.6 T06 – Transferência da aprendizagem no tempo com refletâncias de bandas

Com transferência da aprendizagem do modelo no tempo, ou seja, usando-o para classificar dados do ano posterior ao daquele em que foi treinado, para a mesma zona geográfica, a EG mantém-se acima dos 80% exceto para a zona NE, 81.9% para a zona NC, 86.7% para a zona ND e 77.6% para a zona NE (tabela 4.14).

Tabela 4.14 - Métricas de desempenho de classificadores treinados em 2018 e testados em 2019.

Domínio-fonte	Domínio-alvo	EG (%)	Com TA no tempo		
			Coef. Kappa (%)	F1-score (%)	
2019					
2018	NE	NE	77.6	68.5	74.8
	ND	ND	86.7	78.8	85.4
	NC	NC	81.9	72.8	80.3
	todas	todas	81.2	72.9	79.1

Comparativamente com os resultados obtidos para transferência no espaço, verifica-se que, de modo geral, o desempenho da classificação sofre um maior impacto com a localização geográfica do que com o período temporal, ou seja, há maior variação das características das culturas com o espaço do que com o tempo (pelo menos, entre estes dois anos) o que permite uma melhor reutilização do conhecimento já adquirido por um modelo no domínio-fonte para classificação no domínio-alvo.

4.7 T07 – Transferência da aprendizagem no espaço e no tempo com refletâncias de bandas

Numa transferência da aprendizagem simultânea no tempo e no espaço, ou seja, para modelos testados em dados de, simultaneamente, anos e zonas geográficas distintas daqueles em que foram treinados, o valor da EG está sempre acima dos 77% exceto, mais uma vez, em modelos treinados em NC e testados em NE, acima dos 65% (tabela 4.15).

Tabela 4.15 - Métricas de desempenho de classificadores treinados em anos e zonas diferentes

Domínio-fonte	Domínio-alvo	Com TA no espaço e no tempo		
		EG (%)	Coef. Kappa (%)	F1-score (%)
2019				
NE	ND	80.0	68.3	76.6
	NC	72.9	58.1	69.1
2018	ND	NE	73.3	62.6
		NC	77.6	64.6
	NC	NE	65.9	54.8
		ND	78.5	68.0

4.8 T08 – Transferência da aprendizagem no tempo: treino com dados de todas as zonas em diferentes proporções, validação em dados de 2019

Nesta secção, apresentam-se os resultados relativos à exatidão global de modelos treinados com um número limitado de dados de 2018 (a 30%, 50% ou 70% de cada zona) e testados em todos os dados disponíveis de 2019 – TA no tempo. Pretendeu-se, com este procedimento, perceber o impacto da quantidade de dados treino tem no desempenho do classificador (*vide* secção 3.4 - *Quadro resumo dos testes realizados*) e responder à questão: qual a exatidão global gerada por um modelo treinado com dados de todas as zonas de um ano ao gerar classificações dos dados do ano seguinte?

A tabela 4.16 dá os valores de EG para modelos treinados em dados de 2018, em conjuntos treino formados por amostras desse ano agregadas em diferentes proporções (30%, 50% ou 70% de cada zona) e validados em dados do ano seguinte, de uma zona específica ou não.

Tabela 4.16 - Métricas de desempenho de classificadores treinados em 2018 (apenas parte da totalidade dos dados: 30%, 50% ou 70%) e testados em 2019 (totalidade dos dados).

Domínio-fonte (ano/zona/proporção de dados utilizada)		Domínio- alvo	Com TA no tempo		
			EG (%)	Coef. Kappa (%)	F1-score (%)
2019					
2018 todas	70%	NE	77.5	68.7	75
	50%		77.1	68	74.6
	30%		77	67.9	74.5
	70%	ND	86.5	78.7	85.3
	50%		86	77.9	84.7
	30%		86	77.8	84.6
	70%	NC	80.5	69.7	77.5
	50%		80.7	70	77.6
	30%		80.4	69.5	77.4
	70%	todas	81.1	72.8	79
	50%		80.8	72.2	78.6
	30%		80.7	72.1	78.5

Os resultados estão acima dos 77% de EG para modelos treinados com apenas 30% dos dados de todas as zonas do ano anterior. Acumulando estes dados com os de 2019, espera-se que seja possível classificar dados de 2020 com uma EG próxima destes valores também e que, ao longo dos anos, à medida que mais dados treino vão sendo adicionados a classificação de anos subsequentes seja cada vez mais exata.

4.9 T09 – Redução da dimensionalidade dos dados por marcadores fenológicos

Os resultados da EG dos modelos de classificação desenvolvidos usando os dados interpolados a 5 dias cujas variáveis são os valores de refletância para bandas e aqueles desenvolvidos em dados reduzidos por Marcadores Fenológicos estão representados na tabela 4.17. Usando estes últimos dados, verifica-se um decréscimo dos valores até cerca de 3 pontos percentuais, o que não parece significativo, dadas as vantagens desta estratégia de amostragem.

Tabela 4.17 – Métricas de desempenho de classificadores desenvolvidos usando os dados originais interpolados a 5 dias (valores de refletância de bandas) e reduzidos por Marcadores Fenológicos.

Conjunto de Dados Treino		Refletâncias de bandas		Marcadores fenológicos	
Ano	Zona	EG (%)	EG (%)	Coef. Kappa (%)	F1-score (%)
2018	NE	81.0	79.1	70.9	75.5
	ND	89.8	87.8	80.7	86.2
	NC	90.3	88.6	83.4	87.4
	todas	85.3	82.7	75.5	80.3

Conjunto de Dados Treino		Refletâncias de bandas		Marcadores fenológicos	
Ano	Zona	EG (%)	EG (%)	Coef. Kappa (%)	F1-score (%)
2019	NE	82.2	80.6	72.9	77.7
	ND	90.9	88.8	82.3	87.4
	NC	90.6	90.0	85.2	89.2
	todas	86.0	84.8	78.3	82.9

À semelhança do que se verificou para as técnicas de redução de dimensionalidade anteriores, os resultados da EG decresceram em relação aos resultados de modelos desenvolvidos sobre os dados interpolados. Os modelos desenvolvidos sobre os dados cujas variáveis são os marcadores fenológicos são, ainda assim, mais próximos dos encontrados para modelos treinados nos dados com PCA do que com métodos de seleção de variáveis, com a vantagem de permitirem a análise das variáveis mais importantes no processo de classificação.

Para além disso, o tempo de computação no teste destes modelos em dados de anos ou zonas distintos (que se pretende para a transferência da aprendizagem descrita no capítulo 3.3 - *Metodologia*) é mais reduzido do que o correspondente para o teste nos dados sem redução de dimensionalidade.

4.9.1 Análise das Matrizes de Confusão

Pela análise das matrizes de confusão (Anexo A) verifica-se a tendência apresentada nos resultados anteriores, ou seja, confusão entre vinha e olival; taxas de revocação muito baixas para a classe outras hortícolas; e maiores erros de comissão para as classes outras hortícolas, batata, grão de bico e sorgo. As classes com mais erros de omissão são as classes outras hortícolas e, em menor proporção, abóboras, 'couve' e sorgo (figuras 4.8 e 4.9).

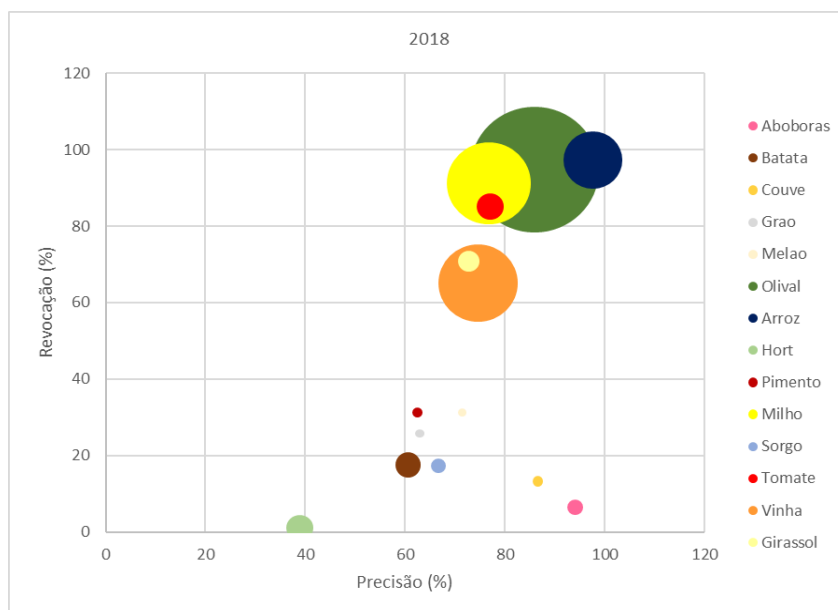


Figura 4.8 – Precisão vs revocação para as classes mais representadas em cada amostra teste de 2018 (todas as zonas). A dimensão dos círculos ilustra a frequência de cada classe nos dados.

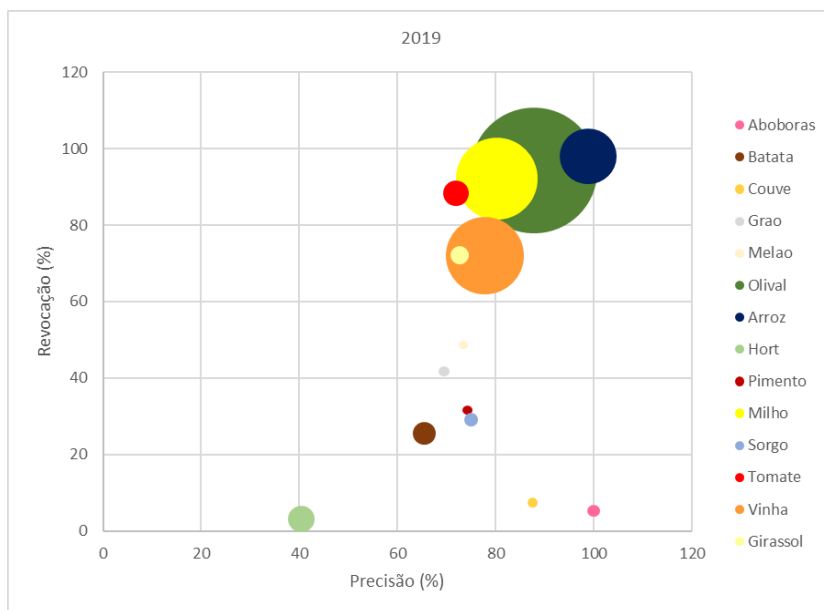


Figura 4.9 – Precisão vs revocação para as classes mais representadas em cada amostra teste de 2019 (todas as zonas). A dimensão dos círculos ilustra a frequência de cada classe nos dados.

4.9.2 Importância de variáveis

Das 24 variáveis geradas para cada subconjunto, foram selecionadas as seis mais importantes na discriminação das diversas culturas para os dois anos em análise, em conjunto (figura 4.10).

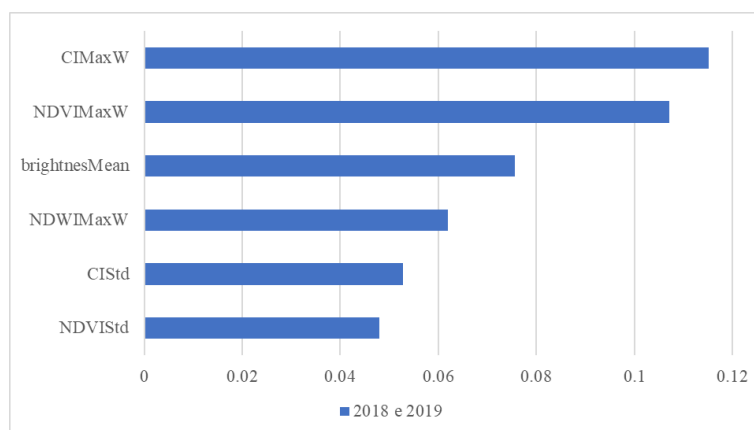


Figura 4.10 – Importância das variáveis usadas na classificação dos dados para os dois anos (2018 e 2019).

Os resultados mostram que as variáveis com maior poder discriminatório das várias culturas são o CIMaxW – o máximo do gradiente temporal do índice de clorofila (no limiar do vermelho) – seguido de perto pelo NDVIMaxW - o máximo do gradiente temporal do índice de vegetação da diferença normalizada. O valor médio do índice de intensidade (BrightnessMean) aparece de seguida, com valor de importância relativa mais de duas décimas abaixo dos primeiros.

4.10 T10 - Classificação por etapas do ciclo vegetativo

Para cada etapa do ciclo vegetativo de crescimento das culturas temporárias de verão, a EG dos modelos de classificação varia de acordo com o representado nos gráficos da figura 4.11.

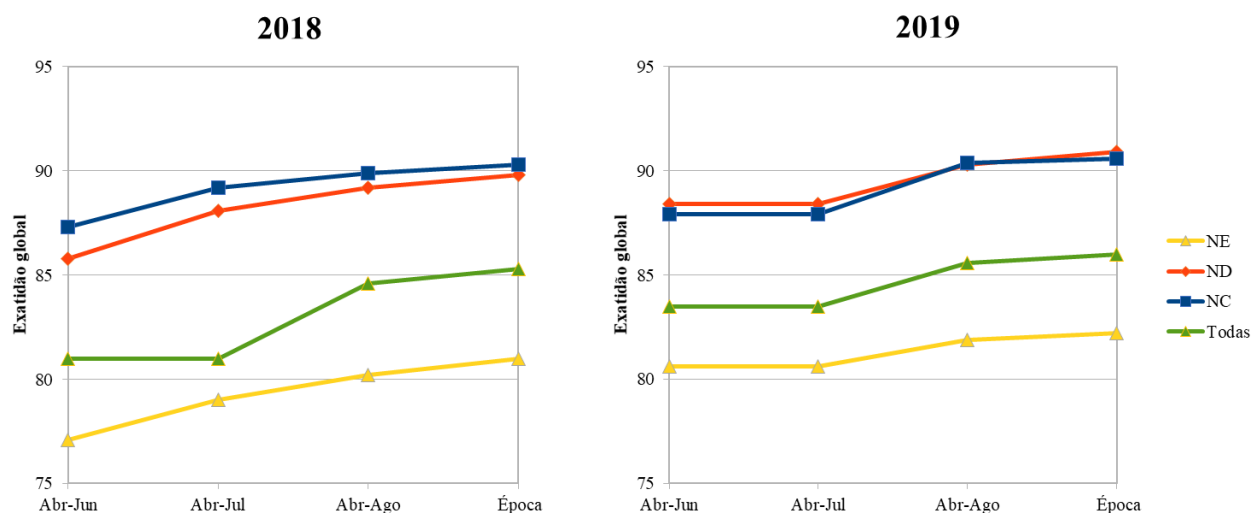


Figura 4.11 - EG dos modelos treinados com dados ao longo da época de crescimento da vegetação para 2018 e para 2019 (todas as zonas).

O desempenho dos modelos treinados com dados de 2019 foi sempre superior à dos treinados com dados de 2018, para o que contribui a maior disponibilidade de imagens ao longo da época de cultivo de 2019, como visto em resultados anteriores. Para os dois anos, os melhores modelos parciais são aqueles treinados com as imagens da série temporal que vai de Abril a Agosto. Estes resultados indicam que a disponibilidade de dados até Agosto é suficiente para gerar modelos de classificação com desempenho superior a 80% para qualquer das zonas, individual ou conjuntamente.

A tabela 4.18 apresenta os valores de F1-score e da EG para cada tipo de cultura. Distinguem-se as classificações feitas usando dados do início (entre Abril e Junho), meio (entre Abril e Agosto) e final do ciclo vegetativo (entre Abril e Setembro). Os valores da EG são superiores a 80% e o F1-score para as diferentes culturas varia entre 35% e 99%, com exceção da classe outras hortícolas (que nunca ultrapassa os 8%) e do sorgo (que passa de 8% no início da estação e atinge os 41% no final da época de crescimento). A principal dificuldade em classificar a classe outras hortícolas resulta, por um lado, da sua baixa representatividade nas amostras (em torno dos 2%) e, por outro, dos elevados erros de omissão devido a confusão com as culturas de milho, olival e vinha já detetados, ao longo da época, resultando em apenas 3 a 4% de parcelas de outras hortícolas corretamente classificadas.

Tabela 4.18 - F1-score e EG obtida nos meses de início, meio e final da estação de crescimento das culturas temporárias de verão em todas as zonas para cada ano.

Dados treino	Etapa	Cultura / F1-score								EG (%)
		Arroz	Girassol	Milho	Olival	Outras hortícolas	Sorgo	Tomate	Vinha	
2018	Abr-Jun	96.0	71.0	79.0	89.0	2.0	8.0	72.0	71.0	81.0
	todas	97.0	75.0	84.0	90.0	4.0	29.0	81.0	72.0	83.3
	as	98.0	81.0	87.0	91.0	7.0	32.0	87.0	73.0	84.6
	zonas	98.0	82.0	87.0	91.0	8.0	33.0	90.0	74.0	85.0
2019	Abr-Jun	98.0	67.0	84.0	90.0	2.0	17.0	80.0	72.0	83.5
	todas	98.0	76.0	88.0	90.0	3.0	37.0	86.0	73.0	85.0
	as	98.0	79.0	88.0	91.0	6.0	32.0	89.0	74.0	85.6
	zonas	99.0	83.0	88.0	91.0	6.0	41.0	90.0	76.0	86.2

Como esperado, os modelos de classificação parciais de início de estação dão menores valores nas métricas de desempenho, embora apresentem, ainda assim, um nível elevado. A classe que é discriminada mais cedo é o sorgo, que passa de 8% a 29%, em 2018, e de 17% a 37% em 2019, entre Junho e Julho. A maior parte das restantes culturas é mais facilmente identificável entre Junho e Agosto, exceto para as culturas de arroz, olival e milho, que mantêm um nível de F1-score relativamente constante ao longo da estação, o que poderá estar relacionado com o seu calendário de crescimento.

4.11 Mapas de confiança

Em todas as zonas em estudo, de um total de 436593 parcelas agrícolas declaradas em 2019, 110487 estiveram ocupadas por culturas temporárias de verão. Destas, na classificação sem TA (testes T02), cerca de 91% foram classificadas com uma confiança acima de 50%; aproximadamente 56% foram classificadas com uma confiança acima de 90%. Na classificação com TA, cerca de 78% das parcelas foram classificadas com uma confiança acima de 50% e, aproximadamente, 19% foram classificadas com uma confiança acima de 90% (tabela 4.19).

Tabela 4.19 – Percentagem do total de parcelas de 2019 com confianças de classificação superiores a 50% e 90% em modelos de classificação sem TA e com TA.

Confiança de classificação superior a:	Sem TA		Com TA	
	Nr. de parcelas	%	Nr. de parcelas	%
50%	100808	91.2	85928	77.8
90%	62289	56.4	20665	18.7

A figura 4.12 mostra o mapa de culturas agrícolas resultante da classificação pelo modelo treinado em dados de 2019, zona ND, a região mais densamente coberta por parcelas agrícolas (zona da Lezíria do Tejo, Ribatejo) e com grande variabilidade de culturas. Os dados usados para

classificação foram as refletâncias das bandas, interpolados temporalmente (resultados T03, vistos previamente).

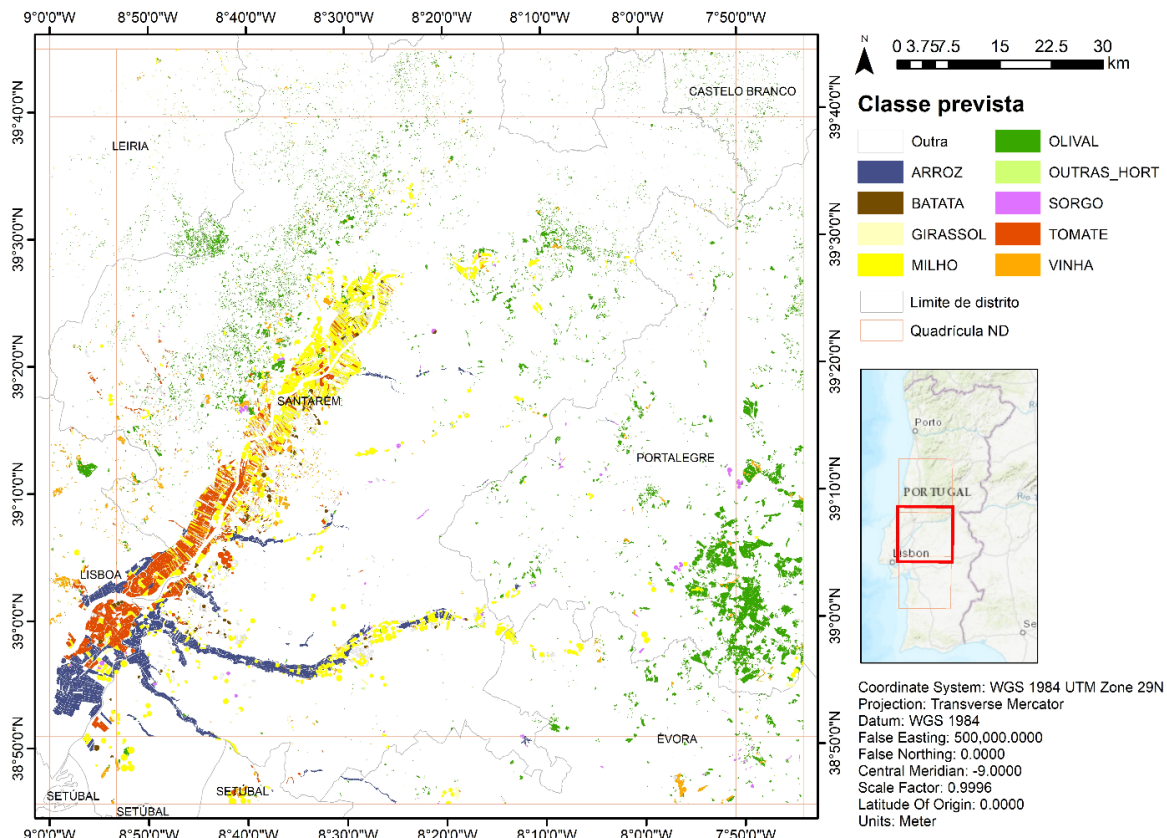


Figura 4.12 - Mapa de culturas agrícolas para as parcelas da zona ND, ano de 2019 usando a metodologia proposta (T02) (EG = 90.9%).

As parcelas classificadas têm dimensão variável: as maiores e, também, as mais concentradas no espaço são as de arroz, tomate e milho, enquanto que as parcelas de olival e vinha se encontram bastante mais dispersas. Visualizam-se ainda algumas parcelas de sorgo e batata dispersas pela região.

A figura 4.13 mostra os mapas das três zonas em estudo (em verdadeira e falsa cor, para visualização da distribuição de culturas agrícolas na área de estudo) e os mapas de confiança das respetivas classificações usando os modelos discutidos na secção 4.2 -T02). Nas duas colunas da direita, estão representados os valores das probabilidades de predição de classes usando os melhores modelos de classificação, sem TA (T02) e com TA no tempo (T06), respetivamente. Os tons mais escuros indicam uma maior confiança na predição da classe, ou seja, um maior número relativo de árvores de RF que votaram para a classe maioritária.

Em geral, onde não ocorreu TA, os modelos comportaram-se melhor e obteve-se uma maior confiança na classificação (figura 4.13, tons mais escuros). Estes resultados eram esperados dado que não se implementaram técnicas específicas para lidar com as diferenças entre domínio-fonte e domínio-alvo – (referidas nas secções 2.2.3, 2.2.4 e 2.2.5 e [34], [36], [35], [38] e [41]). Ainda assim, os valores das métricas analisadas são acima dos 80% (tabela 4.14), com exceção da classificação na zona NE, que apresenta menor disponibilidade de imagens na série temporal (que ficou nos 67.7%).

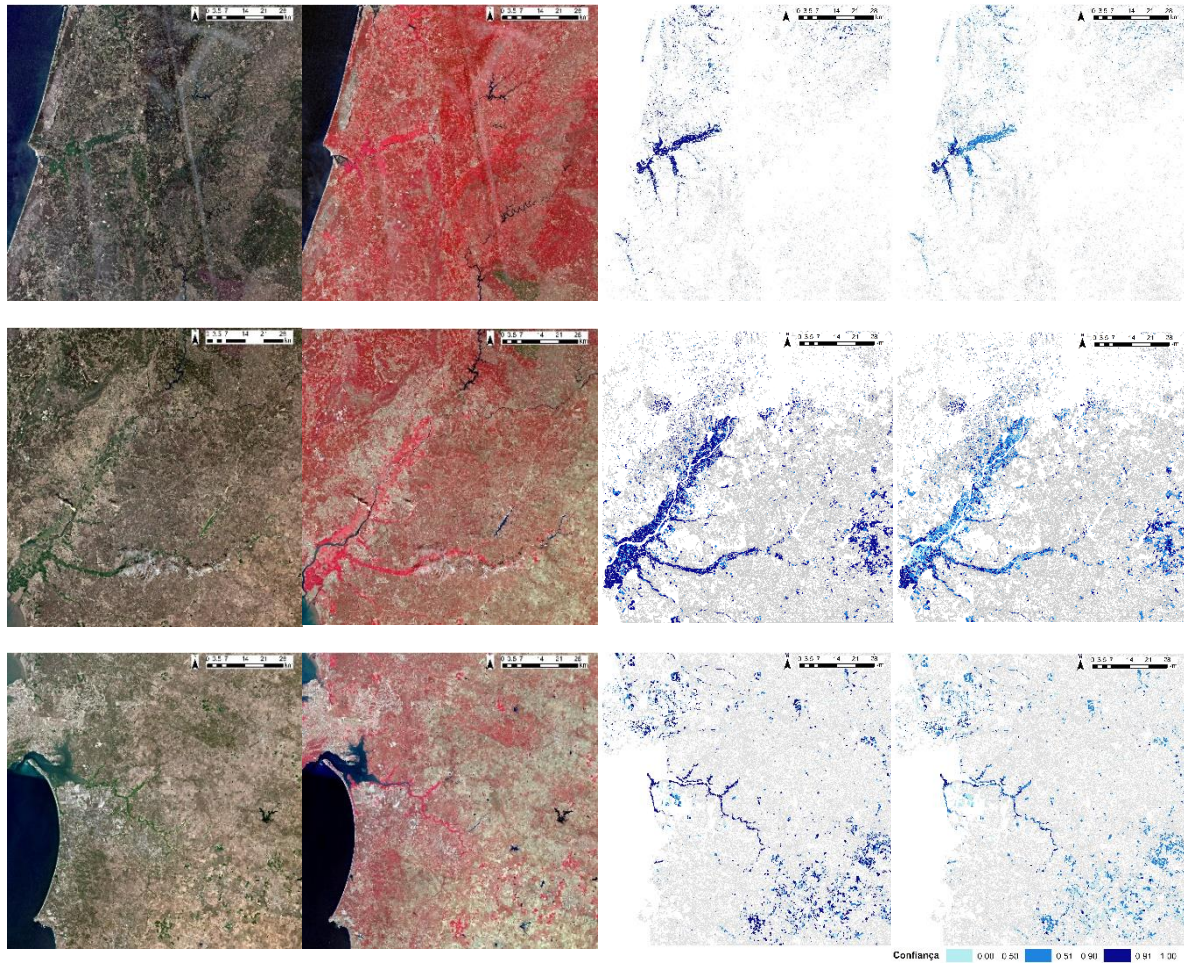


Figura 4.13 – De cima para baixo: imagem das zonas NE, ND e NC de 2019. Da esquerda para a direita: cor verdadeira (RGB: 4.3.2), falsa cor (RGB: 8.4.3), mapa de confiança para modelo de classificação sem TA (T02) e mapa de confiança para modelo de classificação com TA no tempo (T06). Culturas não analisadas a cinzento.

Com TA no tempo, o desempenho da classificação traduz-se numa maior incerteza na atribuição da classe a cada amostra, ou seja, maior dispersão das árvores de decisão do algoritmo pelas diversas etiquetas de classificação.

Olhando para a distribuição de confianças de classificação de ambos os resultados de classificação, sem e com TA (figura 4.14), verifica-se que a maior parte das decisões está concentrada em torno de valores de probabilidade mais elevados (tonalidade mais escura). Estatisticamente, há menor dispersão quando não ocorre TA, sendo a média de confiança de 84% e mediana nos 94%, com valores de 100% para algumas parcelas estimadas e mínimo de 38% (valores abaixo deste, até 12% são já considerados anómalos). O modelo em que ocorre TA entre anos (T06), produz decisões com menores valores de probabilidade associados: média de aproximadamente, 68.8%, mediana de 71% e valores mais dispersos, entre 11% e 100%. Pode-se dizer que, neste modelo, o desempenho da classificação se traduz numa maior incerteza na atribuição da classe a cada amostra, ou seja, maior dispersão das árvores de decisão do algoritmo pelas diversas etiquetas de classificação.

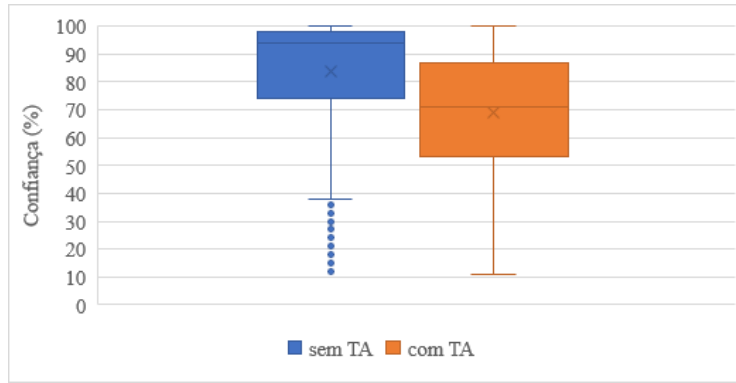


Figura 4.14 – Distribuição de confianças de classificação para modelos de classificação sem TA (T02) e com TA no tempo (T06) para classificação em dados de 2019.

4.11.1 Inspeção visual dos mapas de classificação

Na figura 4.15, podem ver-se, em maior detalhe, os resultados da classificação feita pelo modelo de 2019 em dados de 2019 (sem TA) – testes T02. A inspeção visual proporcionada por estes mapas permite evidenciar erros de classificação e diferenças entre confianças na classificação entre os diversos modelos (por questões de visualização, foram omitidos os valores iguais a 1).

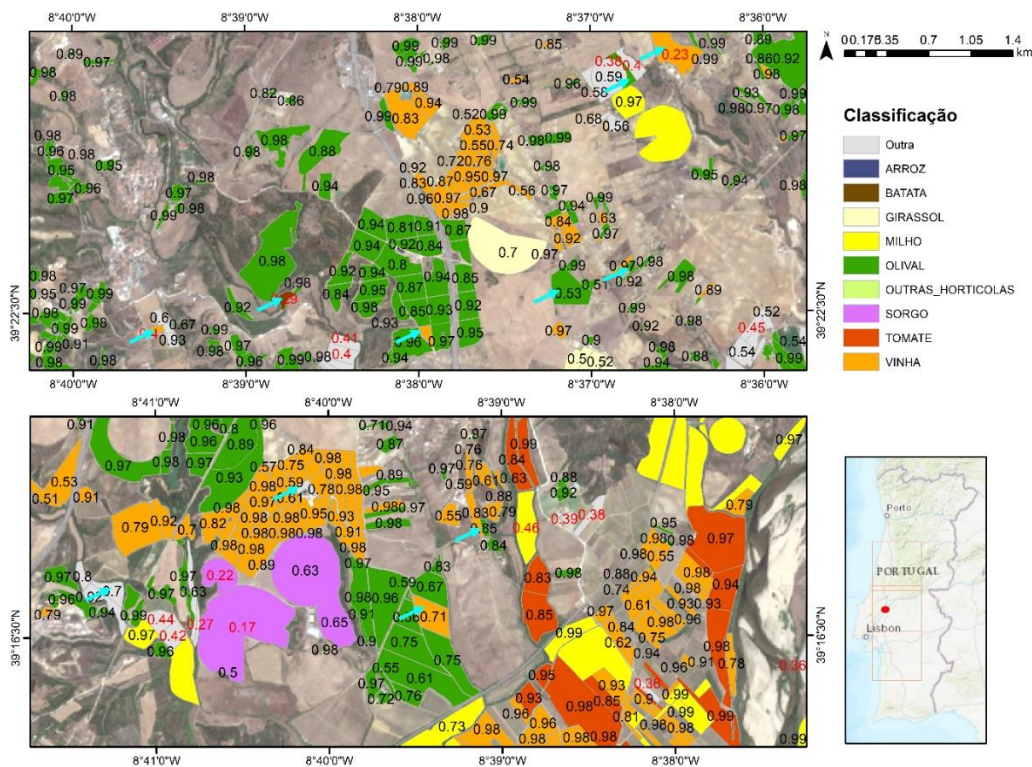


Figura 4.15 - Mapas de confiança (2019 zona ND, para modelo sem TA (T02). Os valores iguais a 1 são omitidos; valores inferiores a 0.5 estão representados a vermelho.

Para o modelo que operou sem TA, representam-se na figura 4.16 algumas das parcelas cujas classes geraram maior confusão e dificuldade na classificação, como mostraram os resultados dos testes T02. As etiquetas numéricas sobre os mapas indicam a cultura, nos dados de referência

(‘Ref.’), de parcelas que foram incorretamente classificadas pelo modelo, ou seja, todos os algarismos de ‘1’ a ‘9’ dispersos no mapa indicam parcelas que foram erradamente classificadas.

Na figura 4.16 (em cima), é evidente a elevada precisão na classificação de parcelas de milho, e algumas omissões de parcelas outras hortícolas atribuídas a milho (parcelas a amarelo que deveriam ser verde claro, sinalizadas com o algarismo ‘6’), a confusão entre olival e vinha (parcelas a laranja que deveriam ser verde escuro, sinalizadas com os algarismos ‘5’ e ‘9’); na figura 4.16 (em baixo): a confusão entre olival e vinha (novamente, parcelas a laranja que deveriam ser verde escuro, sinalizadas com os algarismos ‘5’ e ‘9’ dispersas pela área do mapa) (Nota: alguns ‘9’ indicam parcelas que, no terreno, são de vinha, mas que estão representadas a verde – portanto, mal classificadas como olival - que não são visíveis à dimensão do mapa que é possível representar aqui).

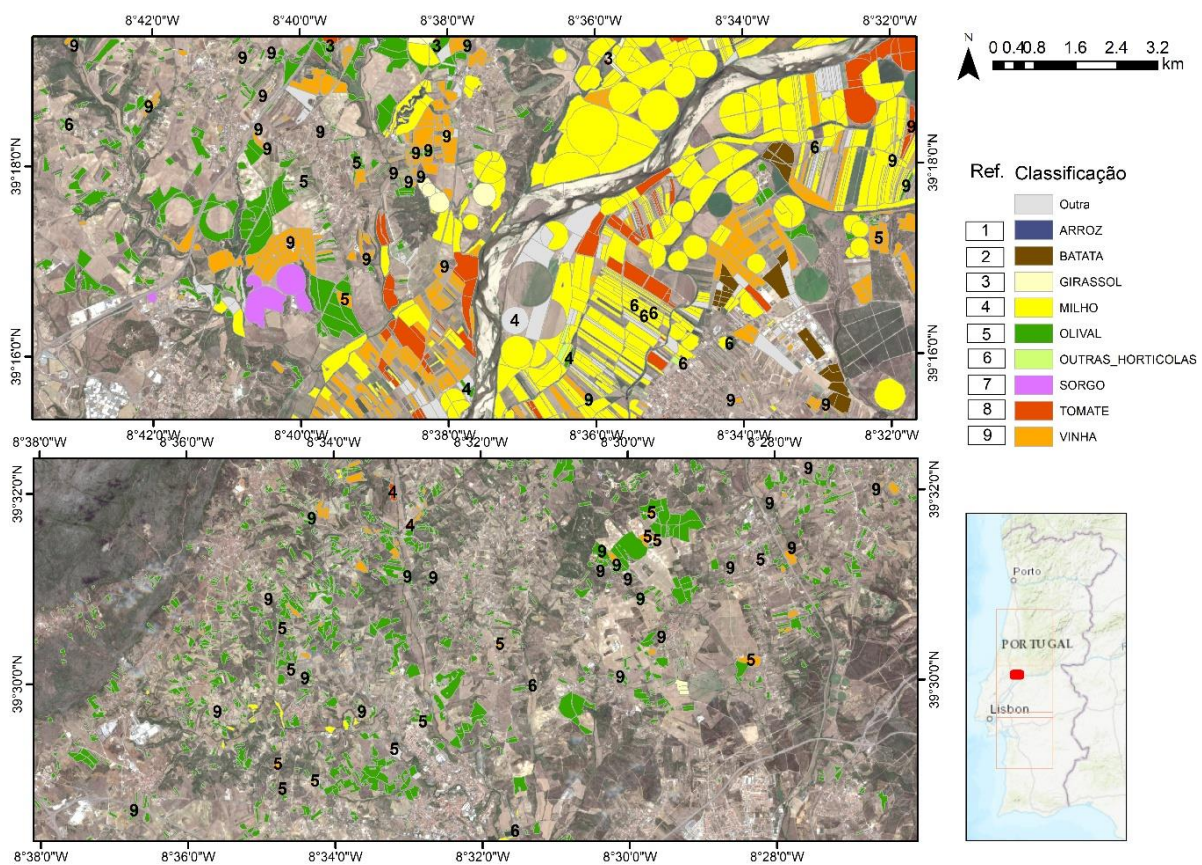


Figura 4.16 - Comparação de mapa de culturas obtidos usando um modelo treinado em dados de 2019 (zona ND - T02 – Interpolação linear e reamostragem da série temporal) com a informação do Sistema de Identificação de Parcelas (SIP). Os números indicam o tipo de cultura dos dados de referência das classes incorretamente classificadas e as cores correspondem ao resultado do classificador.

A figura 4.17 compara as confianças na classificação de cada um dos modelos analisados: em cima, sem TA (testes T02); em baixo, com TA (testes T06). As etiquetas numéricas indicam a probabilidade com que a cultura foi atribuída à parcela correspondente (por questões de visualização, foram omitidos os valores iguais a 1). Além da baixa confiança na predição pelo modelo treinado em dados de 2018 (muitos valores na ordem dos 60-75%), verifica-se que muitas classes são incorretamente atribuídas às parcelas em estudo (apontadores azuis) relativamente ao

modelo treinado com dados de 2019: apenas sete parcelas foram erradamente classificadas pelo modelo de 2019 (T02), enquanto que, no modelo de 2018 (T06), este número aumenta para vinte parcelas incorretamente classificadas só na área visível neste mapa. Para o modelo com TA, a confusão deu-se sobretudo para as classes tomate, milho, batata e olival. Este, classifica com muito menor confiança e fá-lo de forma incorreta com alguma frequência nas classes olival e vinha.

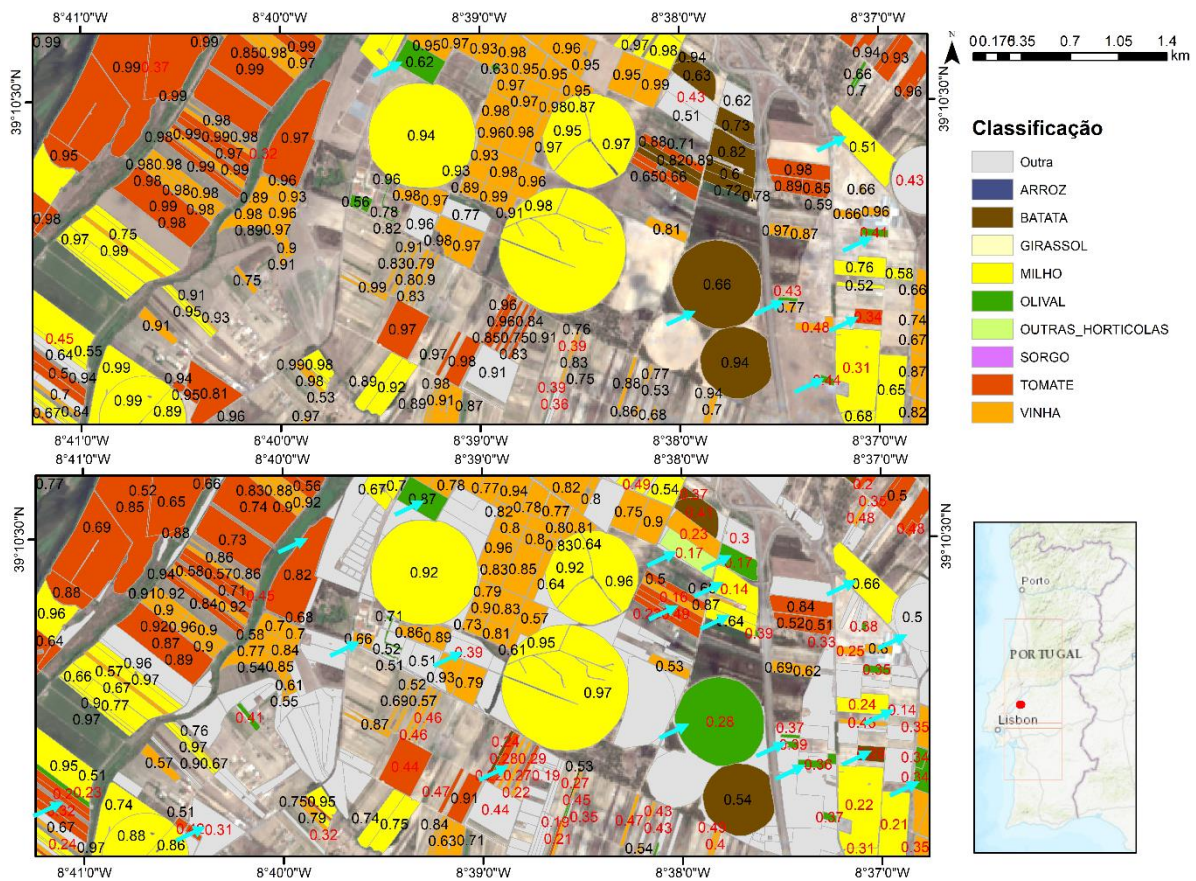


Figura 4.17 - Mapa de confiança (2019 zona ND) para as mesmas parcelas com modelo: sem TA (T02, em cima) e com TA no tempo (T06, em baixo) (apontador azul indica parcelas incorretamente classificadas)

Por exemplo, para as parcelas arroz, ambos os modelos classificam com elevada exatidão mas, como seria de esperar, com confianças distintas (Figura 4.18). A imagem de cima ilustra a classificação do modelo de 2019 com uma elevada precisão para as parcelas de arroz, e apenas 4 parcelas a que foram atribuídas culturas incorretas. A imagem de baixo ilustra a incerteza subjacente ao modelo de 2018 na classificação das parcelas de 2019, com mais parcelas incorretamente classificadas e confusão entre parcelas de arroz e de milho (confundidas com olival, ainda que com uma baixa confiança).

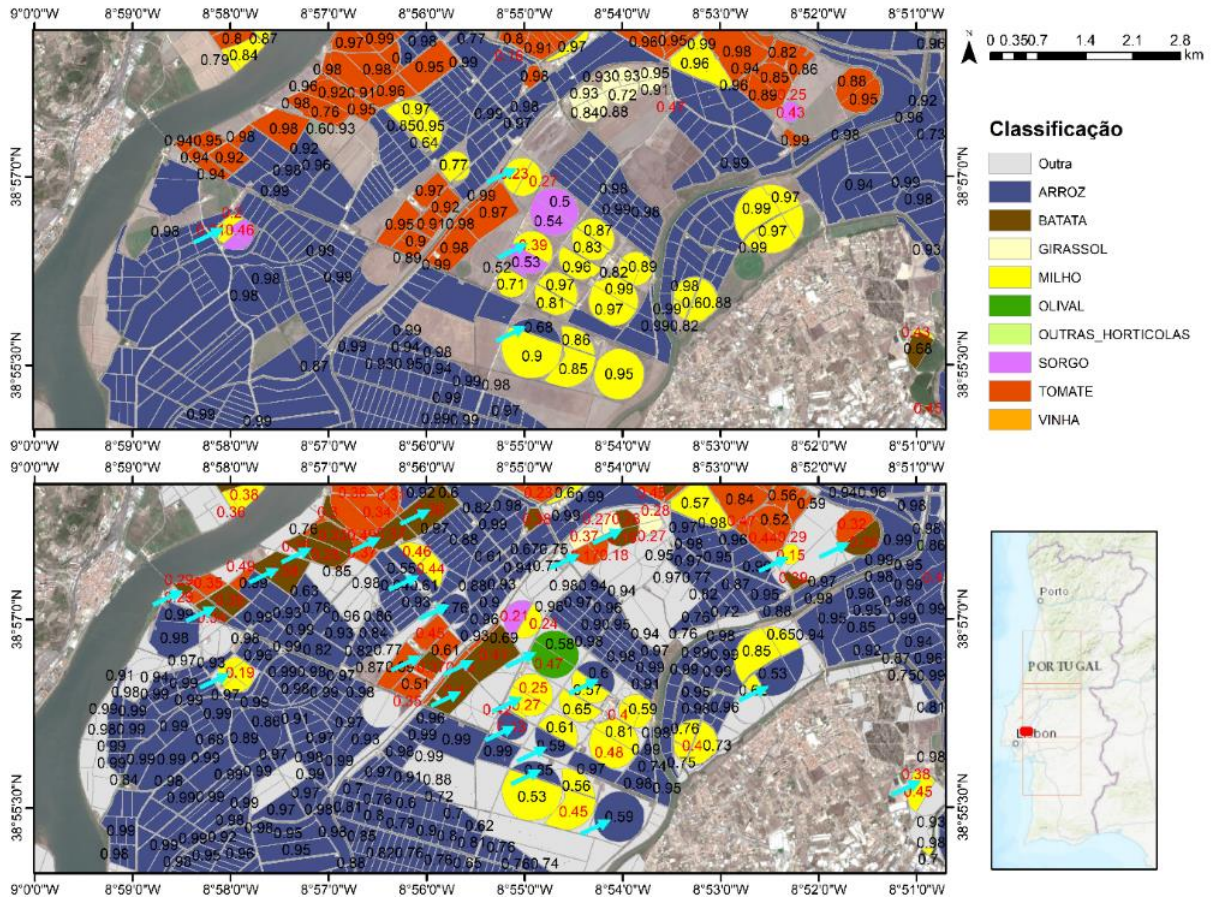


Figura 4.18 - Mapa de confiança (2019 zona ND, para modelo sem TA (T02) e com TA (T06). Os valores iguais a 1 são omitidos; valores inferiores a 0.5 estão representados a vermelho.

4.12 Análise global de resultados – melhores modelos e estratégias

Como esperado, os modelos de classificação gerados com domínios de treino com menor cobertura de imagens na série temporal mostraram os resultados de classificação mais fracos – caso da zona NE, mais a norte (figura 4.19). Em média, os modelos treinados com dados de 2019 (o ano com maior número de imagens disponíveis na série temporal) apresentaram melhores resultados de EG do que os modelos treinados com dados de 2018, na ordem de um a dois pontos percentuais: entre 80.4% e 89.4% para os primeiros e entre 79% e 88.6% para os últimos.

O número de amostras não teve, aparentemente, influência nestes resultados: a zona ND gerou modelos com EG muito próxima de modelos treinados em dados de NC, apesar de apresentar cerca do dobro de amostras desta zona.

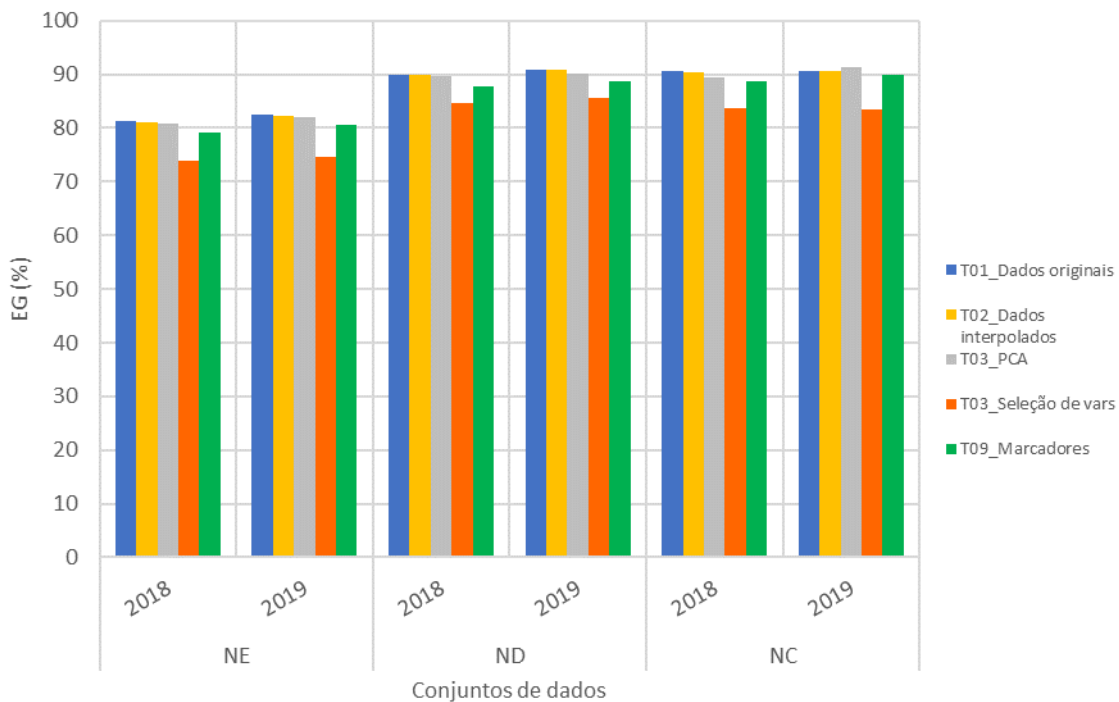


Figura 4.19 - Resultados de exatidão global (EG) para os testes de classificação realizados neste estudo.

Os modelos de classificação que ofereceram melhor desempenho foram aqueles treinados com os dados originais das refletâncias de bandas (coluna a azul) (resultados dos testes T01), com valores de EG entre os 81.4% e os 90.9%, seguidos dos modelos treinados com dados interpolados linearmente a 5 dias (T02) – EG entre os 81% e os 90.9%. Usando marcadores fenológicos (T09), os modelos apresentaram desempenho inferior (EG entre 79.1% e 90%), mas a dimensionalidade dos dados viu-se significativamente reduzida, o que é interessante do ponto de vista dos custos de processamento à medida que os conjuntos de dados vão aumentando em número de amostras. A estas duas estratégias de extração de variáveis – interpolação linear e cálculo de marcadores fenológicos a partir de índices – acresce a vantagem da flexibilidade na manipulação dos dados provenientes de domínios diferentes.

Recorrendo à técnica de PCA (T03) para reduzir a dimensão dos dados, preservou-se a EG dos modelos (de 82.1% a 91.4%, coluna a cinzento na figura 4.19), mas perdeu-se a interpretabilidade dos dados no sentido de perceber que variáveis concorrem para os resultados observados. Com seleção de variáveis (T03), os desempenhos na classificação foram mais fracos: para o modelo treinado com dados interpolados, EG entre 74% e 85.5%. O processo de seleção permitiu, ainda assim, perceber que variáveis comuns aos vários conjuntos de dados são mais importantes na caracterização dos dados em estudo: os valores de refletâncias das bandas 4, 7, 8, 11 e 12, em imagens adquiridas a meio e final do ciclo vegetativo das culturas temporárias de verão.

Para modelos que operaram com TA, os melhores resultados foram obtidos com TA no tempo (T06) - modelos de 2018 validados em 2019), em que a média da EG se manteve acima dos 81% (figura 4.19). Com TA no espaço (T05), os resultados da EG ficaram entre os 75 % e os 77%. Finalmente, em presença de dados treino limitados a 30% dos dados utilizados nos modelos gerados, o valor médio de EG é de 81% com TA no tempo (T08). Estes são resultados bastante interessantes já que, como referido, nenhuma estratégia mais complexa de TA, nomeadamente de

adaptação de domínio, foi utilizada para otimizar a classificação. Este será um bom ponto de partida para a complexificação das técnicas aplicadas.

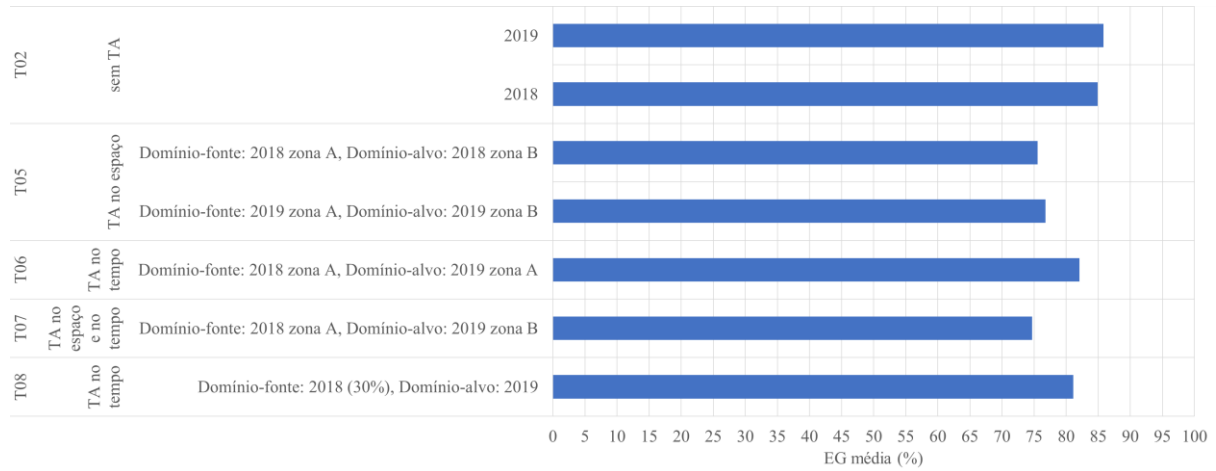


Figura 4.20 - EG média para modelos de classificação que operaram sem TA (T02) e com TA (T05, T06, T07 e T08).

5 Conclusões

Este trabalho testou diversas metodologias de classificação de parcelas agrícolas recorrendo a diferentes algoritmos e estratégias de manipulação de dados. A abordagem de classificação estudada baseou-se em três tipos de extração de variáveis que permitiram caracterizar os perfis das culturas: valores de refletâncias de nove bandas espectrais e subconjuntos destas obtidos por seleção de variáveis, componentes PCA e marcadores fenológicos (calculados a partir dos índices vegetativos).

Posteriormente, a avaliação de impactos na qualidade das classificações consoante as ferramentas usadas serviu para delinear caminhos a seguir consoante a disponibilidade e origem dos dados, para melhorar os modelos e ainda para compreender os mecanismos de transferência da aprendizagem subjacentes à classificação de dados em diferentes anos e zonas geográficas.

A análise quantitativa dos resultados mostra que o classificador RF operando sobre uma série temporal de refletâncias das bandas interpolada linearmente, oferece o melhor equilíbrio entre o desempenho da classificação e os custos de processamento para os conjuntos de dados em estudo. Este algoritmo possibilita ainda o cálculo de confianças de classificação e importância de variáveis explicativas na discriminação dos dados, mostrando-se muito útil na compreensão dos mesmos. Futuramente, maior esforço poderá ser dedicado à exploração e aplicação de outros algoritmos como o SVM e/ou as RNA, na senda de trabalhos desenvolvidos por outros autores e do recomendado pelo JRC (secção 2.2.2). Estes métodos mostraram-se mais exigentes do ponto de vista computacional mas, tendo em conta que mais dados treino podem ser adicionados anualmente, poderão conduzir à construção de modelos mais complexos, robustos e com melhor desempenho de classificação dos dados em estudo.

No caso de necessidade de redução do esforço de processamento, verificou-se que a diminuição do número de amostras de treino para 70%, 50% ou mesmo 30% de cada zona geográfica não prejudica significativamente a exatidão global da classificação (EG sempre superior a 82%), podendo ser uma boa estratégia de amostragem acoplada a outras técnicas para geração de conjuntos de dados treino que otimizem a construção de modelos de classificação com maior qualidade.

Os modelos de classificação gerados com dados treino de 2018 permitem classificar dados de qualquer zona do ano subsequente com uma EG superior a 80%, sem aplicar técnicas específicas de TA, quer ao nível da estrutura do algoritmo, quer ao nível da reamostragem dos dados treino (técnicas de adaptação de domínio, por exemplo). A possibilidade de TA entre anos e entre regiões foi analisada, tendo-se concluído que há uma maior degradação na exatidão da classificação na TA entre regiões no mesmo ano ou em anos distintos do que na mesma região entre anos diferentes, ou seja, o desempenho da classificação sofre um maior impacto com a localização geográfica do que com o período temporal. A incorporação anual de dados atualizados apresenta-se como forma adicional de enriquecer o modelo de classificação e introduzir a variabilidade necessária para aumentar o seu desempenho na classificação até ao alcance de níveis de robustez que permitam a classificação automática de dados de anos seguintes com exatidão global crescente. Adicionalmente, a amostragem seletiva de parcelas corretamente classificadas (técnica de adaptação de domínio) pode ser uma estratégia a considerar para melhoria dos modelos já criados.

O impacto de imagens em falta devido sobretudo a cobertura nebulosa foi claramente um fator limitante na qualidade da classificação: os conjuntos de dados teste com menor número de imagens apresentaram os piores resultados. Os períodos mais longos sem imagens úteis ocorrem nos meses de junho e julho, na fase de maior crescimento das culturas temporárias de verão. A distribuição irregular das imagens no tempo constitui uma limitação à implementação de processos de classificação baseados na variabilidade temporal das refletâncias, à perspectiva de transferibilidade da aprendizagem e ao conhecimento no espaço e no tempo. As metodologias de preenchimento de lacunas das imagens de satélite devido a cobertura nebulosa são potenciais pontos de discussão e a exploração de técnicas alternativas de interpolação como, por exemplo, o uso de técnicas de interpolação espaço-temporal poderão ser consideradas.

Adicionalmente, a utilização de um método de classificação supervisionada mostrou que o número de dados de referência e a sua qualidade (por exemplo, a representatividade de classes *in situ*) deverá ser tido em conta na geração de dados para treinar melhores modelos. Com este trabalho atingiram-se bons resultados de exatidão global dos modelos de classificação (na ordem dos 80% a 89%) mas, como acautelado inicialmente, estes resultados são enviesados pelas classes agrícolas mais representadas. Na análise efetuada ao nível da cultura, as culturas com melhor exatidão do produtor foram: o arroz, o olival e o milho com precisões superiores a 80% e o tomate, a vinha e o girassol com precisões superiores a 70%. Estas culturas representam 93% do total das culturas controladas. É importante melhorar a representatividade dos dados de treino para construção de modelos de classificação mais robustos, capazes de manter elevados níveis de precisão e revocação para as classes minoritárias nas amostras. Não ficou claro, através deste estudo, até que ponto a baixa confiança na classificação de grande parte das parcelas pertencentes a estas culturas pode advir da quantidade de amostras disponíveis para treino ou da sua qualidade - a variabilidade que as caracteriza, isto é, o seu perfil espectral - ou ambas.

Uma das questões que se colocam, então, é: como criar amostras de treino mais representativas a partir dos dados considerados, à partida, como sendo de outros domínios, isto é, de outras zonas geográficas/anos? A resposta poderá passar por utilizar metodologias de aprendizagem ativa para adaptação de domínio já aplicadas por alguns autores em classificação de imagens de deteção remota. Fundamentalmente, enriquecer as amostras de treino com as parcelas correspondentes às culturas menos representadas, introduzindo mais variabilidade nas amostras selecionadas (um pouco à semelhança dos estudos conduzidos por Demir *et al.* [23, 25] e outros autores).

Referências Bibliográficas

- [1] European Space Agency, «ESA - Agriculture». [Em linha]. Disponível em: https://www.esa.int/Applications/Observing_the_Earth/Agriculture. [Acedido: 19-Dez-2019].
- [2] H. Vautmans, «Parecer da Comissão da Agricultura e do Desenvolvimento Rural», 2018.
- [3] Devos W., L. G., M. P., e F. D., *JRC Technical Reports - Technical guidance on the decision to go for substitution of OTSC by monitoring*. 2018.
- [4] F. Gascon, «Sentinel-2 for agricultural monitoring», *Int. Geosci. Remote Sens. Symp.*, vol. 2018-July, pp. 8166–8168, 2018.
- [5] P. Defourny *et al.*, «Near real-time agriculture monitoring at national scale at parcel resolution: Performance assessment of the Sen2-Agri automated system in various cropping systems around the world», *Remote Sens. Environ.*, vol. 221, n. March 2018, pp. 551–568, 2019.
- [6] E. S. Agency, «Copernicus: Sentinel-2 - Satellite Missions - eoPortal Directory», 2019. [Em linha]. Disponível em: <https://directory.eoportal.org/web/eoportal/satellite-missions/c-missions/copernicus-sentinel-2>. [Acedido: 31-Dez-2019].
- [7] E. S. Agency, «System Description – Sentinel-2 for Agriculture», 2018. [Em linha]. Disponível em: <http://www.esa-sen2agri.org/operational-system/system-description/>. [Acedido: 19-Dez-2019].
- [8] E. S. Agency, «Sen4Cap - Sentinels for Common Agriculture Policy», 2017. [Em linha]. Disponível em: <http://esa-sen4cap.org/>. [Acedido: 19-Dez-2019].
- [9] N. Moreira, «Os desafios da PAC 21-27 aos sistemas de informação da Administração Pública», *Cad. Análise e Prospetiva Cultiv.*, vol. 16, pp. 89–95, 2019.
- [10] I. de F. da A. e Pescas, «O IFAP - Apresentação», 2016. [Em linha]. Disponível em: <https://www.ifap.pt/web/guest/o-ifap2>. [Acedido: 19-Dez-2019].
- [11] I. de F. da A. e Pescas, «Regras e Informações Básicas - IFAP», 2016. [Em linha]. Disponível em: <https://www.ifap.pt/sigc-regras>. [Acedido: 13-Jan-2020].
- [12] Devos W., L. G., M. P., e F. D., *Discussion document on the introduction of monitoring to substitute OTSC*. 2017.
- [13] S. Valero *et al.*, «Production of a dynamic cropland mask by processing remote sensing image series at high temporal and spatial resolutions», *Remote Sens.*, vol. 8, n. 1, pp. 1–21, 2016.
- [14] Y. Hu *et al.*, «Artificial Intelligence Approaches», *Geogr. Inf. Sci. Technol. Body Knowl.*, vol. 2019, n. Q3, pp. 1–12, 2019.
- [15] Y. Lu, «Transfer Learning for Image Classification», 2019.
- [16] P. K. Srivastava, D. Han, M. A. Rico-Ramirez, M. Bray, e T. Islam, «Selection of classification techniques for land use/land cover change investigation», *Adv. Sp. Res.*, vol. 50, n. 9, pp. 1250–1265, 2012.
- [17] F. Melgani e L. Bruzzone, «Classification of Hyperspectral Remote Sensing», vol. 42, n. August, pp. 1778–1790, 2004.
- [18] B. Demir, C. Persello, e L. Bruzzone, «Batch-mode active-learning methods for the

- interactive classification of remote sensing images», *IEEE Trans. Geosci. Remote Sens.*, vol. 49, n. 3, pp. 1014–1031, 2011.
- [19] R. P. de Lima e K. Marfurt, «Convolutional neural network based heterogeneous transfer learning for remote-sensing scene classification», *Int. J. Remote Sens.*, vol. 40, n. 22, pp. 8506–8527, 2019.
- [20] D. J. Lary, A. H. Alavi, A. H. Gandomi, e A. L. Walker, «Machine learning in geosciences and remote sensing», *Geosci. Front.*, vol. 7, n. 1, pp. 3–10, 2016.
- [21] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, e B. A. Johnson, «Deep learning in remote sensing applications: A meta-analysis and review», *ISPRS J. Photogramm. Remote Sens.*, vol. 152, n. November 2018, pp. 166–177, 2019.
- [22] G. Mountrakis, J. Im, e C. Ogole, «Support vector machines in remote sensing: A review», *ISPRS Journal of Photogrammetry and Remote Sensing*. 2011.
- [23] J. Holloway e K. Mengersen, «Statistical machine learning methods and remote sensing for sustainable development goals: A review», *Remote Sens.*, vol. 10, n. 9, 2018.
- [24] A. E. Mohamed, «Comparative Study of Four Supervised Machine Learning Techniques for Classification», *Int. J. Appl. Sci. Technol.*, vol. 7, n. 2, pp. 5–18, 2017.
- [25] D. J. Lary, «Artificial Intelligence in Geoscience and Remote Sensing», *Intech*, vol. i, n. tourism, p. 13, 2010.
- [26] «Neural network models (supervised) — scikit-learn 0.22.1 documentation», 2019. [Em linha]. Disponível em: https://scikit-learn.org/stable/modules/neural_networks_supervised.html. [Acedido: 22-Fev-2020].
- [27] T. Kavzoglu e P. M. Mather, «The use of backpropagating artificial neural networks in land cover classification», *Int. J. Remote Sens.*, vol. 24, n. 23, pp. 4907–4938, 2003.
- [28] S. B. Imandoust e M. Bolandraftar, «Application of K-Nearest Neighbor (KNN) Approach for Predicting Economic Events : Theoretical Background», *Int. J. Eng. Res. Appl.*, vol. 3, n. 5, pp. 605–610, 2013.
- [29] J. A. Richards e J. Xiuping, *Remote sensing digital image analysis: An introduction*, vol. 9783642300. 2013.
- [30] L. Breiman, «Bagging predictors», *Mach. Learn.*, vol. 24, n. 2, pp. 123–140, 1996.
- [31] M. Bowles, *Machine Learning in Python*. Wiley, 2015.
- [32] M. Belgiu e L. Drăgu, «Random forest in remote sensing: A review of applications and future directions», *ISPRS J. Photogramm. Remote Sens.*, vol. 114, pp. 24–31, 2016.
- [33] A. J. Sage, «Random Forest Robustness, Variable Importance, and Tree Aggregation», *ProQuest Diss. Theses*, p. 120, 2018.
- [34] D. Sarkar, «A Comprehensive Hands-on Guide to Transfer Learning with Real-World Applications in Deep Learning», 2018. [Em linha]. Disponível em: <https://towardsdatascience.com/a-comprehensive-hands-on-guide-to-transfer-learning-with-real-world-applications-in-deep-learning-212bf3b2f27a>. [Acedido: 22-Dez-2019].
- [35] S. J. Pan e Q. Yang, «WEMOTE -Word Embedding based Minority Oversampling Technique for Imbalanced Emotion and Sentiment Classification», *PLoS One*, vol. 8, n. 3, pp. 1–16, 2009.
- [36] N. Segev, M. Harel, S. Mannor, K. Crammer, e R. El-Yaniv, «Learn on Source, Refine on Target: A Model Transfer Learning Framework with Random Forests», *IEEE Trans.*

- Pattern Anal. Mach. Intell.*, vol. 39, n. 9, pp. 1811–1824, 2017.
- [37] D. Tuia, C. Persello, e L. Bruzzone, «Domain adaptation for the classification of remote sensing data: An overview of recent advances», *IEEE Geosci. Remote Sens. Mag.*, vol. 4, n. 2, pp. 41–57, 2016.
- [38] D. Tuia, M. Volpi, L. Copa, M. Kanevski, e J. Muñoz-Marí, «A survey of active learning algorithms for supervised remote sensing image classification», *IEEE J. Sel. Top. Signal Process.*, vol. 5, n. 3, pp. 606–617, 2011.
- [39] B. Demir, F. Bovolo, e L. Bruzzone, «Updating land-cover maps by classification of image time series: A novel change-detection-driven transfer learning approach», *IEEE Trans. Geosci. Remote Sens.*, vol. 51, n. 1, pp. 300–312, 2013.
- [40] C. Paris, L. Bruzzone, e D. Fernández-Prieto, «A novel method based on source domain understanding and modeling to transfer labels from land-cover vector maps to classifiers for multispectral images», *Int. Geosci. Remote Sens. Symp.*, vol. 2018-July, pp. 3619–3622, 2018.
- [41] W. Devos *et al.*, *Second discussion document on the introduction of monitoring to substitute OTSC : rules for processing applications in 2018-2019*. 2018.
- [42] I. P. do M. e da A. IPMA, «IPMA - Clima Normais», 2020. [Em linha]. Disponível em: <https://www.ipma.pt/pt/oclima/normais.clima/>. [Acedido: 16-Jan-2020].
- [43] J. A. dos S. Varela, «Estruturas da Propriedade e Culturas Regionais», 2008.
- [44] F. Warmerdam e E. Rouault, «SENTINEL2 – Sentinel-2 Products — GDAL documentation», 2019. [Em linha]. Disponível em: <https://gdal.org/drivers/raster/sentinel2.html>. [Acedido: 31-Dez-2019].
- [45] J. Inglada *et al.*, «Assessment of an operational system for crop type map production using high temporal and spatial resolution satellite optical imagery», *Remote Sens.*, vol. 7, n. 9, pp. 12356–12379, 2015.
- [46] M. Lepot, J. B. Aubin, e F. H. L. R. Clemens, «Interpolation in time series: An introductory overview of existing methods, their performance criteria and uncertainty assessment», *Water (Switzerland)*, vol. 9, n. 10, 2017.
- [47] «Decomposing signals in components (matrix factorization problems) — scikit-learn 0.22.1 documentation», 2019. [Em linha]. Disponível em: <https://scikit-learn.org/stable/modules/decomposition.html#pca>. [Acedido: 12-Out-2020].
- [48] «sklearn.feature_selection.SelectKBest — scikit-learn 0.22.1 documentation», 2019. [Em linha]. Disponível em: https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.SelectKBest.html. [Acedido: 12-Out-2020].
- [49] J. W. Rouse, R. H. Hass, J. A. Schell, e D. W. Deering, «Monitoring vegetation systems in the great plains with ERTS», *Third Earth Resour. Technol. Satell. Symp.*, vol. 1, pp. 309–317, 1973.
- [50] P. J. Zarco-Tejada, J. R. Miller, T. L. Noland, G. H. Mohammed, e P. H. Sampson, «Scaling-up and model inversion methods with narrowband optical indices for chlorophyll content estimation in closed forest canopies with hyperspectral data», *IEEE Trans. Geosci. Remote Sens.*, vol. 39, n. 7, pp. 1491–1507, 2001.
- [51] B.-C. Gao, «Naval Research Laboratory, 4555 Overlook Ave», *Remote Sens. Env.*, vol. 7212, n. April 1995, pp. 257–266, 1996.

- [52] «sklearn.svm.SVC — scikit-learn 0.22.1 documentation». [Em linha]. Disponível em: <https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html>. [Acedido: 10-Dez-2019].
- [53] «sklearn.neural_network.MLPClassifier — scikit-learn 0.22.1 documentation». [Em linha]. Disponível em: https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html. [Acedido: 10-Dez-2019].
- [54] «sklearn.neighbors.KNeighborsClassifier — scikit-learn 0.22.1 documentation». [Em linha]. Disponível em: <https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html>. [Acedido: 10-Dez-2019].
- [55] «sklearn.ensemble.RandomForestClassifier — scikit-learn 0.22.1 documentation», 2019. [Em linha]. Disponível em: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>. [Acedido: 16-Jan-2020].
- [56] J. Inglada, A. Vincent, M. Arias, B. Tardy, D. Morin, e I. Rodes, «Operational High Resolution Land Cover Map Production at the Country Scale Using Satellite Image Time Series», *Remote Sens.*, vol. 9, n. 1, p. 95, 2017.
- [57] S. Raschka, «Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning», 2018.
- [58] S. J. Russell e P. Norvig, *Artificial Intelligence A Modern Approach*; Pearson Education, 2003.
- [59] J. R. Landis e G. G. Koch, «The Measurement of Observer Agreement for Categorical Data», *Biometrics*, vol. 33, n. 1, p. 159, 1977.
- [60] «sklearn.model_selection.RandomizedSearchCV — scikit-learn 0.22.1 documentation». [Em linha]. Disponível em: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.RandomizedSearchCV.html. [Acedido: 25-Fev-2020].
- [61] «sklearn.model_selection.GridSearchCV — scikit-learn 0.22.1 documentation». [Em linha]. Disponível em: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html. [Acedido: 25-Fev-2020].
- [62] L. Breiman, «ST4_Method_Random_Forest», *Mach. Learn.*, vol. 45, n. 1, pp. 5–32, 2001.
- [63] F. J. Valverde-Albacete e C. Peláez-Moreno, «100% classification accuracy considered harmful: The normalized information transfer factor explains the accuracy paradox», *PLoS One*, vol. 9, n. 1, 2014.
- [64] N. V. Chawla, «Data mining for imbalanced datasets: An overview», em *Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy with severe factor XII deficiency*, vol. 57, n. 5, 2005, pp. 854–867.

Anexos

A. Matrizes de Confusão

- i. T01 – Matriz de confusão para a classificação realizada em dados de 2018 da zona ND. Colunas representam a classificação e linhas representam os dados de referência. Os valores são dados em valor absoluto, exceto quando indicado %. A diagonal representa o número de parcelas classificadas corretamente pelo modelo.

	ABO	AMEND OIM	ARR OZ	BAT	COUV E	GIRASS	MELAN CIA	MELAO	MILH O	OLIV	HORT	PIM	SORG O	TOM	VINHA	Total	Freq (%)	Revocação (%)	F1- score (%)
ABO	35	0	0	0	0	0	0	0	3	7	0	2	0	5	9	62	0.55	56.45	64
AMEND OIM	0	17	0	0	0	0	0	0	2	0	0	0	0	3	0	22	0.2	77.27	85
ARROZ	0	0	852	0	0	0	0	0	9	6	0	0	0	0	0	867	7.74	98.27	99
BAT	0	0	0	82	1	0	0	0	10	11	0	0	0	8	12	124	1.11	66.13	74
COUVE	0	0	0	3	18	0	0	0	5	6	0	1	0	3	10	46	0.41	39.13	51
GIRASS	0	0	0	1	0	39	0	0	10	2	0	0	0	19	6	77	0.69	50.65	67
MELAN CIA	1	0	0	1	0	0	6	0	0	3	0	0	0	11	2	25	0.22	24	35
MELAO	1	0	0	0	0	0	1	5	2	2	0	1	0	5	2	19	0.17	26.32	37
MILHO	0	1	2	2	0	0	1	1	1295	32	0	1	0	11	26	1372	12.25	94.39	94
OLIV	0	0	1	1	0	0	0	0	4	6016	0	0	1	0	144	6167	55.07	97.55	94
HORT	5	0	0	1	4	0	0	0	7	31	5	4	0	6	37	100	0.89	5	10
PIM	0	0	0	0	0	0	0	0	4	1	0	52	1	8	12	78	0.7	66.67	66
SORGO	2	0	1	0	1	0	0	0	7	27	0	0	26	2	11	77	0.69	33.77	49
TOM	0	0	1	0	0	0	0	0	4	3	0	1	0	584	5	598	5.34	97.66	92
VINHA	0	0	0	0	0	0	0	0	8	438	0	2	0	0	1016	1464	13.07	69.4	73
Total	49	18	857	97	25	39	9	8	1374	6603	5	79	29	676	1312	11199	NA	NA	NA
Precisão (%)	71.43	94.44	99.42	84.54	72	100	66.67	62.5	94.25	91.11	100	65.82	89.66	86.39	77.44	100	NA	NA	NA

EG =89.9%, Coef. Kappa = 84.1%, F1-score = 88.8%

ii. T01 – Matriz de confusão para a classificação realizada em dados de 2019 da zona ND. Colunas representam a classificação e linhas representam os dados de referência. Os valores são dados em valor absoluto, exceto quando indicado %. A diagonal representa o número de parcelas classificadas corretamente pelo modelo.

	ABO	ARROZ	BAT	BAT_DOCE	COUVE	GIRASS	GRAO	MELAO	MILHO	OLIV	HORT	PIM	SORGO	TOM	VINHA	Total	Freq (%)	Revocação (%)	F1-score (%)
ABO	29	0	0	0	0	0	0	1	2	5	0	0	0	1	4	42	0.38	69.05	75
ARROZ	0	852	0	0	0	0	0	0	5	2	0	0	0	1	4	864	7.76	98.61	99
BAT	0	0	110	1	1	0	0	0	12	9	1	0	0	5	5	144	1.29	76.39	77
BAT_DOCE	0	0	0	19	0	0	0	0	0	1	0	0	0	1	0	22	0.2	86.36	88
COUVE	0	0	0	0	33	0	0	0	4	1	0	1	0	3	5	47	0.42	70.21	67
GIRASS	0	0	1	0	0	38	0	0	2	1	0	0	0	11	5	58	0.52	65.52	71
GRAO	0	0	2	0	2	0	9	0	0	9	0	0	0	0	1	23	0.21	39.13	45
MELAO	0	0	0	0	0	0	0	11	1	1	1	0	0	5	4	23	0.21	47.83	55
MILHO	1	2	4	0	1	0	0	0	1282	48	0	1	0	8	25	1375	12.34	93.24	94
OLIV	0	0	1	0	0	0	0	0	4	6043	0	0	0	0	113	6161	55.3	98.08	95
HORT	1	0	4	1	2	0	0	0	11	26	4	4	0	2	32	87	0.78	4.6	6
PIM	1	0	0	0	0	0	0	0	6	2	0	54	0	2	4	70	0.63	77.14	72
SORGO	0	1	1	0	0	0	0	0	10	27	0	0	18	2	12	71	0.64	25.35	47
TOM	0	0	0	0	0	0	0	0	6	1	0	2	0	544	3	556	4.99	97.84	93
VINHA	0	0	0	0	0	0	0	0	5	410	0	0	0	4	1052	1471	13.2	71.52	76
Total	32	855	129	22	40	38	9	13	1361	6601	7	70	18	617	1292	11141	NA	NA	NA
Precisão (%)	90.62	99.65	85.27	86.36	82.5	100	100	84.62	94.2	91.55	57.14	77.14	100	88.17	81.42	100	NA	NA	NA

EG = 90.9%, Coef. Kappa = 85.7%, F1-score = 89.9%

iii. T02 – Matriz de confusão para a classificação realizada em dados de 2018 de todas as zonas, com dados interpolados (a 5 dias). Colunas representam a classificação e linhas representam os dados de referência. Os valores são dados em valor absoluto, exceto quando indicado %. A diagonal representa o número de parcelas classificadas corretamente pelo modelo.

	ABO	ARROZ	BAT	COUVE	FEIJAO	GIRASS	GRAO	MELAO	MILHO	OLIV	HORT	PIM	SORGO	TOM	VINHA	Total	Freq (%)	Revocação (%)	F1-score (%)
ABO	54	1	4	0	1	1	0	2	115	22	5	2	0	2	36	247	0.75	21.9	32.0
ARROZ	0	2983	0	0	0	0	0	0	44	23	0	0	0	0	5	3055	9.24	97.6	98.0
BAT	4	0	153	0	0	1	0	0	228	73	4	1	0	6	98	569	1.72	26.9	38.0
COUVE	2	0	4	26	1	0	0	0	34	8	3	0	0	8	12	98	0.3	26.5	41.0
FEIJAO	0	1	5	0	11	0	0	0	125	18	3	0	0	4	23	190	0.57	5.79	11.0
GIRASS	0	1	2	0	0	289	1	0	27	40	0	0	0	17	20	397	1.2	72.8	82.0
GRAO	0	0	0	0	0	1	19	0	3	35	0	0	1	0	6	66	0.2	28.8	44.0
MELAO	3	0	0	0	0	0	0	35	2	4	6	0	0	7	7	64	0.19	54.7	65.0
MILHO	6	31	32	1	1	2	0	1	5734	221	8	5	0	10	178	6230	18.8	92	87.0
OLIV	0	3	5	0	0	6	1	0	100	14068	3	0	2	1	553	14742	44.6	95.4	91.0
HORT	8	1	7	2	1	1	0	2	223	164	27	7	0	17	176	636	1.92	4.25	8.0
PIM	2	1	0	0	0	0	0	0	10	3	1	56	0	6	15	95	0.29	59	63.0
SORGO	2	4	1	0	0	1	0	0	74	38	1	2	44	1	29	197	0.6	22.3	36.0
TOM	0	0	2	0	0	0	0	0	28	8	0	1	0	608	3	651	1.97	93.4	90.0
VINHA	0	2	3	0	0	0	0	0	103	1410	2	1	0	0	4064	5585	16.9	72.8	75.0
Total	90	3030	240	29	15	310	21	43	6900	16174	70	83	47	703	5255	33066	NA	NA	NA
Precisão (%)	60	98.5	63.8	89.7	73.3	93.2	90.5	81.4	83.1	86.98	38.6	67.5	93.6	86.49	77.34	NA	NA	NA	NA

EG = 85.3%, Coef. Kappa = 79.2%, F1-score = 83.5%

iv. T02 – Matriz de confusão para a classificação realizada em dados de 2019 de todas as zonas, com dados interpolados (a 5 dias). Colunas representam a classificação e linhas representam os dados de referência. Os valores são dados em valor absoluto, exceto quando indicado %. A diagonal representa o número de parcelas classificadas corretamente pelo modelo.

	ABO	ARROZ	BAT	COUVE	FEIJAO	GIRASS	GRAO	MELAO	MILHO	OLIV	HORT	PIM	SORGO	TOM	VINHA	Total	Freq (%)	Revocação (%)	F1-score (%)
ABO	31	0	3	0	1	0	0	2	64	10	2	0	0	2	34	151	0.46	20.53	32.0
ARROZ	0	2929	0	0	0	0	0	0	39	15	0	0	2	0	1	2986	9.07	98.09	99.0
BAT	2	0	217	0	4	0	0	0	123	78	3	1	0	2	50	483	1.47	44.93	56.0
COUVE	0	0	6	34	3	1	0	1	26	6	3	0	0	5	10	95	0.29	35.79	51.0
FEIJAO	0	1	1	0	11	0	0	0	89	10	2	0	1	3	21	139	0.42	7.91	13.0
GIRASS	0	2	1	0	0	236	0	0	9	41	0	0	0	14	13	316	0.96	74.68	83.0
GRAO	0	1	7	0	0	1	29	0	3	48	0	0	0	0	9	98	0.3	29.59	45.0
MELAO	0	0	1	0	0	2	0	40	2	1	1	1	0	18	8	74	0.22	54.05	64.0
MILHO	2	12	17	0	6	1	0	0	5706	294	8	6	0	10	210	6274	19.06	90.95	89.0
OLIV	0	0	1	0	0	1	0	1	83	14235	3	0	1	0	592	14917	45.31	95.43	91.0
HORT	3	0	19	4	4	0	0	2	253	173	19	4	0	9	189	679	2.06	2.8	5.0
PIM	0	1	2	0	0	0	0	0	14	3	1	52	0	5	3	82	0.25	63.41	67.0
SORGO	0	3	1	0	0	3	0	0	64	32	0	0	54	1	27	185	0.56	29.19	44.0
TOM	0	1	2	0	0	1	0	0	11	4	0	1	0	585	9	614	1.87	95.28	90.0
VINHA	1	1	1	0	0	0	0	0	85	1407	1	0	0	1	4078	5577	16.94	73.12	75.0
Total	40	2955	294	39	30	250	30	51	6609	16401	43	74	58	679	5290	32920			
Precisão (%)	77.5	99.1	73.8	87.2	36.7	94.4	96.7	78.43	86.34	86.79	44.19	70.3	93.1	86.2	77.09	100	NA	NA	NA

EG = 86%, Coef. Kappa = 80%, F1-score = 84.4%

- v. T09 - Matriz de confusão para a classificação realizada em dados de 2018 de todas as zonas, com marcadores fenológicos. Colunas representam a classificação e linhas representam os dados de referência. Os valores são dados em valor absoluto, exceto quando indicado %. A diagonal representa o número de parcelas classificadas corretamente pelo modelo.

	ABO	ARROZ	BAT	COUVE	FEIJAO	GIRASS	GRAO	MELAO	MILHO	OLIV	HORT	PIM	SORGO	TOM	VINHA	Total	Freq (%)	Revocação (%)	F1-score (%)
ABO	16	3	5	0	0	2	0	0	159	18	2	2	3	5	31	6.48	12.0	6.48	12.0
ARROZ	0	2971	0	0	0	0	0	0	65	9	0	0	1	0	9	97.25	97.0	97.25	97.0
BAT	0	1	100	0	0	8	0	0	295	42	0	1	1	17	104	17.57	27.0	17.57	27.0
COUVE	0	0	3	13	0	10	1	0	44	3	1	0	0	18	5	13.27	23.0	13.27	23.0
FEIJAO	0	0	2	0	2	2	0	0	128	19	0	0	4	5	28	1.05	2.0	1.05	2.0
GIRASS	0	8	2	0	1	281	2	0	52	10	1	0	2	26	12	70.78	72.0	70.78	72.0
GRAO	0	1	3	0	0	14	17	0	14	12	0	0	0	1	4	25.76	37.0	25.76	37.0
MELAO	0	3	0	0	0	1	0	20	21	4	1	0	0	9	5	31.25	43.0	31.25	43.0
MILHO	0	34	9	1	0	10	1	1	5679	229	1	2	2	20	240	91.16	83.0	91.16	83.0
OLIV	0	3	3	0	0	13	1	0	166	13977	2	0	3	0	574	94.81	90.0	94.81	90.0
HORT	0	1	9	1	0	7	0	5	273	131	7	5	0	16	181	1.1	2.0	1.1	2.0
PIM	1	3	0	0	0	3	0	0	41	3	0	30	1	9	4	31.25	42.0	31.25	42.0
SORGO	0	3	2	0	1	10	0	0	110	21	0	1	34	4	11	17.26	27.0	17.26	27.0
TOM	0	7	6	0	0	4	1	0	71	5	0	0	0	554	3	85.1	81.0	85.1	81.0
VINHA	0	0	2	0	0	3	1	0	174	1761	2	0	0	1	3641	65.19	70.0	65.19	70.0
Total	17	3040	165	15	4	386	27	28	7396	16257	18	48	51	718	4875	33064	NA	NA	NA
Precisão (%)	94.12	97.73	60.61	86.67	50	72.8	62.96	71.43	76.78	85.98	38.89	62.5	66.67	77.16	74.69	100	NA	NA	NA

EG = 82.7%, Coef. Kappa = 75.5%, F1-score = 80.3%

vi. T09 – Matriz de confusão para a classificação realizada em dados de 2019 de todas as zonas, com marcadores fenológicos. Colunas representam a classificação e linhas representam os dados de referência. Os valores são dados em valor absoluto, exceto quando indicado %. A diagonal representa o número de parcelas classificadas corretamente pelo modelo.

	ABO	ARROZ	BAT	COUVE	FEIJAO	GIRASS	GRAO	MELAO	MILHO	OLIV	HORT	PIM	SORGO	TOM	VINHA	Total	Freq (%)	Revocação (%)	F1-score (%)
ABO	8	0	0	0	0	3	1	3	94	13	1	1	3	9	15	151	0.46	5.3	10.0
ARROZ	0	2926	0	0	0	0	0	0	37	19	0	0	0	0	4	2986	9.07	97.99	98.0
BAT	0	0	123	0	0	4	1	0	230	40	4	1	0	29	48	483	1.47	25.47	37.0
COUVE	0	0	2	7	0	4	1	1	47	3	0	0	1	25	4	95	0.29	7.37	14.0
FEIJAO	0	0	3	0	0	4	0	0	97	8	2	0	1	1	23	139	0.42	0	0.0
GIRASS	0	1	0	0	0	228	5	0	18	11	0	0	4	40	9	316	0.96	72.15	72.0
GRAO	0	0	1	0	0	30	41	0	10	7	0	0	0	6	2	98	0.3	41.84	52.0
MELAO	0	2	1	0	0	2	3	36	12	4	0	1	1	9	3	74	0.22	48.65	59.0
MILHO	0	13	14	0	0	4	1	0	5782	233	11	2	3	23	187	6274	19.06	92.16	86.0
OLIV	0	0	2	0	0	11	1	0	185	14066	9	0	2	5	636	14917	45.32	94.3	91.0
HORT	0	2	7	0	0	3	2	4	311	142	21	2	1	10	174	679	2.06	3.09	6.0
PIM	0	1	1	0	0	2	0	0	34	1	1	26	0	7	8	82	0.25	31.71	44.0
SORGO	0	8	0	0	0	3	0	0	84	20	0	0	54	9	7	185	0.56	29.19	42.0
TOM	0	1	9	0	0	2	1	2	48	4	0	0	0	542	5	614	1.87	88.27	79.0
VINHA	0	1	0	0	0	2	0	0	137	1422	1	0	0	0	4014	5577	16.94	71.97	75.0
Total	8	2958	188	8	0	314	59	49	7204	16020	52	35	72	753	5157	32918	NA	NA	NA
Precisão (%)	100	98.92	65.43	87.5	0	72.61	69.49	73.47	80.26	87.8	40.38	74.29	75	71.98	77.84	100	NA	NA	NA

EG = 84.8%, Coef. Kappa = 78.3%, F1-score = 82.9%