



UNIVERSIDADE
LUSÓFONA

Detecção Automática de Hipérboles em Radar GPR

Trabalho Final de Curso

Relatório Intercalar 1º Semestre

Diogo Jorge, a22211095

Orientador: Houda Harkat

Orientador: João Carvalho

Trabalho Final de Curso | LEI | 2025

www.lusofona.pt

Direitos de cópia

Deteção Automática de Hipérboles em Radar GPR, Copyright de **Diogo Jorge**,
Universidade Lusófona.

A Escola de Comunicação, Arquitectura, Artes e Tecnologias da Informação (ECATI) e a Universidade Lusófona (UL) têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objectivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

Este documento foi gerado com o processador (pdf/Xe/Lua) \LaTeX e o modelo ULThesis (v1.0.0) [Mat24].

Resumo

O GPR (Ground Penetrating Radar) é uma técnica de prospeção não invasiva que utiliza impulsos eletromagnéticos de alta frequência para investigar o subsolo. A interação destas ondas com interfaces e objetos enterrados origina padrões característicos em radargramas bi-dimensionais, entre os quais se destacam as hipérboles, associadas a alvos pontuais como cabos, tubagens, rochas ou cavidades. A correta identificação destas estruturas é fundamental para a localização de infraestruturas enterradas, para o apoio à engenharia civil e para a geofísica aplicada. Tradicionalmente, a deteção de hipérboles é realizada manualmente por especialistas, num processo demorado, sujeito a erro humano e particularmente afetado por ruído, heterogeneidade do solo e sobreposição de reflexões. Métodos clássicos de processamento de sinal, como a Transformada de Hough ou o *template matching*, revelam limitações quando aplicados a dados reais com elevada complexidade. Este relatório intercalar descreve a fase de pesquisa e conceção de um sistema automático de deteção de hipérboles em radargramas, recorrendo a técnicas de visão computacional e aprendizagem automática. Em particular, explora-se a combinação entre Common Spatial Patterns (CSP) (Common Spatial Patterns), utilizada para realçar padrões discriminativos nas janelas dos radargramas, e um classificador Support Vector Machine (SVM) (Support Vector Machine), responsável pela distinção entre regiões com e sem hipérboles. Esta abordagem procura equilibrar robustez, interpretabilidade e baixos requisitos computacionais, constituindo uma alternativa prática a métodos mais complexos. O presente relatório descreve o problema, o estado da arte, a metodologia prevista e o calendário de desenvolvimento do Trabalho Final de Curso.



Figure 1: DEISI Logo

Abstract

Ground Penetrating Radar (GPR) is a non-invasive subsurface exploration technique that uses high-frequency electromagnetic pulses to investigate underground structures. The interaction of these waves with buried objects and interfaces generates characteristic patterns in two-dimensional radargrams, among which hyperbolas are prominent. These signatures are typically associated with point-like targets such as cables, pipes, rocks, or cavities. Correct identification of such structures is essential for locating buried infrastructure, supporting civil engineering applications, and conducting applied geophysics analyses. Traditionally, hyperbola detection is performed manually by specialists, in a process that is time-consuming, prone to human error, and particularly affected by noise, soil heterogeneity, and overlapping reflections. Classical signal-processing methods, such as the Hough Transform or template matching, show significant limitations when applied to real-world radargrams with high levels of complexity. This interim report describes the research and conceptualisation phase of an automatic hyperbola detection system for GPR radargrams, using computer vision and machine learning techniques. In particular, it explores the combination of CSP (Common Spatial Patterns), used to enhance discriminative patterns within radargram windows, with a SVM (Support Vector Machine) classifier, responsible for distinguishing between regions with and without hyperbolas. This approach aims to balance robustness, interpretability, and low computational requirements, offering a practical alternative to more complex methods. This report outlines the problem, the state of the art, the proposed methodology, and the development schedule for the Final Course Project.

Índice

Resumo	2
Abstract	3
Índice	4
Lista de Figuras	6
Lista de Tabelas	7
1 Identificação do Problema	8
1.1 Enquadramento Geral	8
1.2 Desafios Atuais	8
1.3 Motivação	9
1.4 Objetivo	9
1.5 Pergunta de Pesquisa	9
1.6 Hipótese	9
1.7 Etapas do Relatório	9
2 Benchmarking	11
2.1 Fundamentos e Desafios do GPR	11
2.2 Métodos Clássicos de Deteção de Hipérboles	11
2.2.1 Transformada de Hough Hiperbólica	11
2.2.2 Template Matching	11
2.2.3 Métodos Baseados em Derivadas	11
2.3 Aprendizagem Automática Tradicional	12
2.4 Deep Learning Aplicado ao GPR	12
2.4.1 Classificação de <i>patches</i>	12
2.4.2 Detetores de Objetos (YOLO, Faster R-CNN)	12
2.4.3 Classificação Baseada em CSP + SVM	12
2.5 Síntese Crítica	13
3 Viabilidade e Pertinência	14
4 Solução Proposta	16
5 Resultados e Análise	18
5.1 Resultados Preliminares	18
5.1.1 Discussão dos Resultados	18
5.2 Base de Dados	19
5.2.1 Estatísticas da Base de Dados	19
5.2.2 Construção da Base de Dados	19
5.2.3 Pré-processamento	20
5.2.4 Transformação para Domínio de Sinais	20
5.3 Melhorias Futuras	20
5.3.1 Melhoria do Pré-processamento	20
5.3.2 Normalização Avançada	20

5.3.3	Otimização de Hiperparâmetros	20
5.3.4	Melhoria das Características	20
5.3.5	Aumento do Dataset	21
5.3.6	Integração de Contexto Espacial	21
5.3.7	Comparação com Deep Learning	21
6	Calendário	22
6.1	Progresso Atual	22
6.2	Dificuldades e Ajustes	22
6.3	Plano de Trabalho Inicial	23
6.4	Plano de Trabalho Atual	24
	Bibliografia	25
	Glossário	26

Lista de Figuras

1	DEISI Logo	2
6.1	Calendário de atividades do TFC entre Outubro de 2025 e Junho de 2026.	24

Lista de Tabelas

5.1	Estatísticas da base de dados utilizada	19
6.1	Plano de trabalho e cronograma proposto para o remanescente do TFC. . .	23

1 - Identificação do Problema

1.1 Enquadramento Geral

O Ground Penetrating Radar (GPR) (Ground Penetrating Radar) é uma tecnologia de prospeção não invasiva que utiliza impulsos eletromagnéticos de alta frequência para gerar representações bidimensionais do subsolo [Dan04; Jol09]. Através da análise dos padrões resultantes da interação das ondas com interfaces de diferentes propriedades dielétricas, é possível inferir a presença de objetos enterrados, cavidades, alterações estruturais e limites de materiais.

Em radargramas, os alvos pontuais manifestam-se tipicamente sob a forma de *hipérboles* [Dan04]. Esta característica geométrica resulta da variação do tempo de propagação das ondas eletromagnéticas conforme o recetor se afasta do ponto de reflexão. O estudo da forma e abertura dessas hipérboles permite estimar profundidade, posição horizontal e características aproximadas do alvo.

Embora a interpretação humana de radargramas seja uma prática consolidada, existem limitações evidentes [Jol09]:

- o processo é subjetivo e depende da experiência do operador;
- a variabilidade dos solos afeta significativamente a forma das hipérboles;
- elevados níveis de ruído reduzem a legibilidade dos radargramas;
- o volume de dados gerados em levantamentos reais é frequentemente demasiado extenso para tratamento manual eficiente.

Estas dificuldades justificam o crescente interesse em soluções automáticas baseadas em visão computacional e *deep learning*, capazes de identificar hipérboles com maior rapidez, consistência e precisão.

1.2 Desafios Atuais

A deteção automática de hipérboles apresenta vários desafios técnicos, entre os quais se destacam [JL21]:

- **Ruído e interferências:** ruído de banda larga, ruído aleatório, reverberações e artefactos instrumentais dificultam a separação entre fundo e sinal útil;
- **Variabilidade das formas:** diferentes solos alteram a velocidade de propagação das ondas, modificando a curvatura das hipérboles;
- **Sobreposição de reflexões:** em ambientes complexos, hipérboles podem aparecer incompletas ou ocultas;
- **Escassez de datasets anotados:** a anotação manual de hipérboles é trabalhosa e requer conhecimento especializado.

Modelos de aprendizagem automática, em particular abordagens baseadas em CSP combinadas com classificadores SVM, têm mostrado resultados promissores ao lidar com estes desafios, uma vez que permitem realçar padrões discriminativos e separar eficazmente regiões com e sem hipérboles mesmo em imagens com elevado ruído.

1.3 Motivação

A motivação central deste trabalho surge da necessidade de:

- reduzir o tempo de processamento e análise de radargramas;
- minimizar a dependência exclusiva da interpretação humana;
- melhorar a precisão da deteção de hipérboles em condições reais;
- viabilizar a integração do GPR em fluxos de trabalho mais rápidos e automatizados;
- contribuir para a modernização das tecnologias utilizadas em geofísica e engenharia civil.

A crescente utilização do GPR em contextos como inspeção de infraestruturas, levantamentos urbanos, arqueologia e análise de pavimentos reforça a necessidade de metodologias automatizadas robustas [JL21; HS22], capazes de operar de forma eficiente em dados complexos e ruidosos.

1.4 Objetivo

O objetivo deste trabalho é desenvolver a base técnica e metodológica para um sistema automático de deteção de hipérboles em dados GPR, recorrendo a um classificador SVM apoiado por técnicas de extração de características baseadas em CSP. Pretende-se criar um processo completo, desde o pré-processamento até à avaliação do modelo, capaz de identificar hipérboles de forma robusta mesmo em radargramas ruidosos ou heterogéneos.

1.5 Pergunta de Pesquisa

A investigação é orientada pela seguinte questão principal:

“Um classificador SVM combinado com extração de características baseada em CSP é capaz de detetar hipérboles com maior precisão e robustez do que métodos clássicos de processamento de sinal?”

1.6 Hipótese

A hipótese formulada é que a combinação entre extração de características baseada em CSP e um classificador SVM, devido à sua capacidade de realçar padrões discriminativos e separar eficazmente classes em espaços de elevada dimensionalidade, irá superar abordagens tradicionais (como a Transformada de Hough ou o *template matching*) [ZM06; GG00] em métricas como F1-score e precisão. Espera-se que esta abordagem apresente melhor robustez face a diferentes tipos de solo, níveis de ruído e condições de aquisição, mantendo ao mesmo tempo baixo custo computacional e boa capacidade de generalização.

1.7 Etapas do Relatório

Este relatório está organizado de forma a apresentar, de maneira clara e progressiva, a construção do trabalho desenvolvido ao longo do semestre. Após a identificação do problema e respetivo enquadramento, segue-se um capítulo dedicado ao benchmarking

e ao estado da arte, onde são analisadas as abordagens existentes para a detecção de hipérboles em dados GPR, bem como as suas limitações e potencialidades. Este capítulo estabelece a base científica e técnica que fundamenta a necessidade de recorrer a métodos de aprendizagem automática.

O capítulo seguinte aprofunda a viabilidade e pertinência do projeto, demonstrando em detalhe por que razão a utilização de uma abordagem baseada em CSP e SVM é adequada, quais os requisitos técnicos, que dados estão disponíveis e de que forma o projeto se insere no contexto académico e profissional do curso.

Posteriormente, é apresentada a solução proposta, onde se descrevem as tecnologias utilizadas, o processo de pré-processamento dos dados, a estratégia de extração de características com CSP e a metodologia prevista para o treino e avaliação do classificador SVM. Este capítulo define de forma clara como o sistema será construído e quais os fundamentos da escolha desta abordagem.

Por fim, o relatório inclui um capítulo dedicado ao calendário de trabalho, apresentando a planificação temporal das atividades realizadas e das etapas que ainda serão desenvolvidas ao longo do segundo semestre. Este planeamento permite organizar o desenvolvimento do projeto e assegurar que os objetivos definidos são cumpridos de forma faseada e estruturada.

2 - Benchmarking

2.1 Fundamentos e Desafios do GPR

O funcionamento do GPR baseia-se na emissão de impulsos eletromagnéticos de alta frequência que se propagam no subsolo. Quando as ondas interagem com interfaces de diferentes propriedades dielétricas, parte da energia é refletida e registada pelo recetor [Dan04; Jol09]. A reconstrução destes ecos ao longo do tempo produz radargramas, que podem ser interpretados como imagens bidimensionais.

A forma característica das hipérboles nos radargramas resulta da geometria do tempo de viagem das ondas refletidas por alvos pontuais [Dan04]. Teoricamente, a forma pode ser aproximada por:

$$t(x) = \sqrt{t_0^2 + \left(\frac{x - x_0}{v}\right)^2}$$

onde x_0 corresponde à posição horizontal do alvo e v à velocidade de propagação no meio. Em condições reais, fatores como ruído, heterogeneidade e reflexões múltiplas distorcem esta forma ideal, tornando a deteção automática um desafio técnico relevante.

2.2 Métodos Clássicos de Deteção de Hipérboles

Os métodos clássicos procuram correspondências geométricas ou padrões de curvatura nos radargramas. Os principais incluem:

2.2.1 Transformada de Hough Hiperbólica

Uma das abordagens mais estudadas consiste na generalização da Transformada de Hough para curvas hiperbólicas [ZM06]. Embora eficaz para dados limpos, sofre com:

- sensibilidade extrema ao ruído;
- custo computacional elevado devido ao espaço de parâmetros;
- dificuldades em lidar com hipérboles sobrepostas.

2.2.2 Template Matching

Hipérboles sintéticas são correlacionadas com o radargrama [GG00]. No entanto:

- as hipérboles reais variam com o tipo de solo e profundidade;
- pequenas diferenças na forma resultam em fraca correlação;
- é difícil generalizar a novos cenários.

2.2.3 Métodos Baseados em Derivadas

Filtros de gradiente (Sobel, Canny) e operadores de curvatura procuram concavidades compatíveis com hipérboles, mas sofrem com ruído e interpretação ambígua [HS22].

2.3 Aprendizagem Automática Tradicional

Com o avanço da *machine learning*, surgiram classificadores aplicados a janelas extraídas dos radargramas. Métodos como SVM e Random Forest foram usados para distinguir janelas com ou sem hipérbolas [JL21]. Embora representem um avanço face às abordagens puramente geométricas, apresentam limitações importantes:

- dependem de engenharia manual de características;
- não captam o contexto espacial global;
- podem apresentar desempenho modesto em cenas reais.

Uma extensão promissora consiste na utilização de técnicas de extração de características como *Common Spatial Patterns* (CSP), que procuram encontrar transformações capazes de maximizar a separação entre classes. A combinação CSP + SVM permite realçar padrões discriminativos nas janelas dos radargramas e apresenta boa capacidade de generalização mesmo com datasets relativamente reduzidos, constituindo assim uma alternativa prática e leve em termos computacionais.

2.4 Deep Learning Aplicado ao GPR

A chegada das redes neurais convolucionais transformou a análise de imagens, incluindo aplicações ao GPR. As principais linhas incluem:

2.4.1 Classificação de *patches*

Uma CNN simples classifica pequenas regiões da imagem. Apesar de melhor que SVM, não consegue capturar formas completas ou relações espaciais complexas [JL21].

2.4.2 Detetores de Objetos (YOLO, Faster R-CNN)

Estes métodos produzem *bounding boxes* em torno das hipérbolas. Contudo:

- não recuperam a forma exata da hipérbole [HS22];
- hipérbolas pequenas ou deformadas podem ser ignoradas.

Embora estes métodos apresentem bons resultados, exigem grandes quantidades de dados anotados e recursos computacionais elevados, o que nem sempre é viável em contextos académicos ou industriais com constrangimentos de tempo e anotação.

2.4.3 Classificação Baseada em CSP + SVM

Uma alternativa leve e eficaz aos métodos de *deep learning* consiste na combinação entre técnicas de extração de características e classificadores supervisionados. Entre estas, destaca-se o uso de CSP (Common Spatial Patterns) para gerar projeções que maximizam a separação entre classes, seguido de um classificador SVM (Support Vector Machine), responsável pela distinção entre regiões dos radargramas que contêm hipérbolas (*target*) e regiões sem alvos relevantes (*no-target*).

Ao contrário das abordagens de segmentação pixel a pixel, este método opera ao nível de *patches* (janelas) extraídos dos radargramas, classificando cada região de forma independente. Esta estratégia reduz substancialmente a necessidade de datasets extensos e anotação detalhada, tornando-se particularmente adequada a projetos académicos e cenários onde a obtenção de dados rotulados é limitada.

A abordagem CSP + SVM apresenta várias vantagens:

- **baixo custo computacional** — adequado quando não existem GPUs ou grandes volumes de dados anotados;
- **robustez ao ruído**, especialmente quando combinada com pré-processamento;
- **boa capacidade de generalização** mesmo com datasets modestos;
- **elevada interpretabilidade** das características extraídas.

Estudos aplicados à detecção de objetos enterrados e à classificação de sinais radar mostram que métodos baseados em SVM obtêm desempenhos competitivos em tarefas de detecção supervisionada. Resultados típicos reportados na literatura incluem:

- Transformada de Hough: F1-score 40–55% [ZM06];
- Template Matching: 55–65% [GL00];
- Detetores de Objetos CNN: 70–80% [JL21];
- CSP + SVM: **75–90%** [Pas+10; Xia+18; Lot+18; Wan+15].

2.5 Síntese Crítica

A literatura evidencia que os métodos clássicos são insuficientes para lidar com a complexidade real dos radargramas, sobretudo em cenários com ruído e variação dielétrica significativa. Abordagens de aprendizagem automática, em particular a combinação entre CSP e SVM, surgem como alternativas mais robustas, apresentando melhor capacidade de generalização e menor sensibilidade às condições de aquisição, o que justifica a sua adoção como base metodológica deste trabalho.

3 - Viabilidade e Pertinência

A viabilidade do projeto proposto assenta na conjugação entre maturidade técnica, baixo custo de implementação e elevado potencial de continuidade após a conclusão do Trabalho Final de Curso (TFC). Embora desenvolvido em contexto académico, trata-se de um projeto que responde a uma necessidade real do mercado da prospeção não invasiva e que acompanha a tendência crescente de integrar métodos de aprendizagem automática em sistemas de análise de dados geofísicos [Pas+10; Xia+18]. A utilização de ferramentas totalmente gratuitas ou de código aberto, como Python, NumPy, scikit-learn, OpenCV e plataformas de computação como o Google Colab, reduz substancialmente os custos associados ao desenvolvimento. Como consequência, tanto durante o TFC como no futuro, a evolução do sistema não depende de investimentos financeiros significativos, mas sim de tempo, dados e conhecimento técnico, tornando o projeto sustentável do ponto de vista económico.

A elevada acessibilidade das tecnologias utilizadas permite que o projeto possa ser facilmente continuado após a sua conclusão formal. A abordagem baseada em CSP e SVM é altamente escalável: quanto maior for o conjunto de exemplos *target* e *no-target*, melhor se tornam as fronteiras de decisão do SVM e mais robustas as características projetadas pelo CSP [Lot+18; Wan+15]. Esta propriedade confere ao sistema uma longevidade que ultrapassa o contexto académico, permitindo que a solução seja progressivamente refinada, incorporada em ferramentas profissionais ou até integrada em fluxos de trabalho industriais de análise de radargramas.

A pertinência do projeto reforça-se quando se observa a crescente necessidade de automatização no setor do GPR. A interpretação manual de radargramas continua a ser um processo exigente, moroso e dependente da experiência individual do operador. Em levantamentos de grande dimensão, a análise manual torna-se impraticável sem o apoio de sistemas automáticos de triagem. A capacidade de um algoritmo distinguir de forma consistente e robusta entre regiões que contêm hipérboles e regiões sem alvos relevantes tem impacto direto na eficiência operacional e na redução do erro humano. Este impacto é reconhecido por empresas internacionais de referência no setor do GPR, como a GSSI, IDS GeoRadar e Sensors&Software, que têm vindo a incorporar algoritmos inteligentes nos seus equipamentos, demonstrando a atualidade e a relevância de soluções de apoio à interpretação automática [JL21].

Para além das necessidades identificadas pela indústria, técnicos e investigadores que utilizam GPR em áreas como geotecnia, arqueologia, construção civil ou inspeção de pavimentos relatam dificuldades recorrentes na interpretação dos radargramas. A heterogeneidade dos solos, os elevados níveis de ruído, a atenuação do sinal e a sobreposição de múltiplas reflexões tornam o processo complexo e, muitas vezes, ambíguo [HS22]. Assim, uma ferramenta capaz de acelerar e padronizar a deteção de hipérboles representa um contributo significativo para a prática profissional, ao reduzir tempos de análise, aumentar a precisão e melhorar a segurança em intervenções que dependem da correta identificação de infraestruturas subterrâneas.

Embora o presente relatório ainda não inclua a validação externa prevista, está planeada a aplicação de um questionário dirigido a potenciais utilizadores — como técnicos de GPR, engenheiros, arqueólogos e estudantes — com o objetivo de recolher dados concretos sobre o interesse, utilidade e aplicabilidade da solução. Este inquérito permitirá avaliar, de forma quantitativa e qualitativa, as necessidades reais do público-alvo, o grau de entusiasmo relativamente à automatização da interpretação de radargramas e

a percepção global da tecnologia proposta. A análise destes resultados, a incluir no anexo final, contribuirá para validar a pertinência do projeto sob a perspetiva dos utilizadores finais e reforçar a legitimidade da solução desenvolvida.

Em síntese, o projeto revela-se tecnicamente viável, economicamente sustentável e altamente relevante num contexto real de aplicação. Para além de responder aos desafios identificados no capítulo anterior, apresenta potencial de crescimento, integração e evolução futura. Demonstra, portanto, que não se trata apenas de um exercício académico, mas de uma base sólida para investigação continuada e utilização prática em diferentes domínios profissionais.

4 - Solução Proposta

A solução proposta para este projeto consiste no desenvolvimento de um sistema de deteção automática de hipérboles em radargramas GPR baseado na combinação entre extração de características por CSP e classificação supervisionada através de um SVM. Ao contrário das abordagens de segmentação pixel a pixel, este sistema opera sobre janelas (*patches*) extraídas dos radargramas, aprendendo a distinguir entre regiões que contêm hipérboles e regiões sem alvos relevantes. Este paradigma revela-se particularmente adequado quando o número de exemplos anotados é limitado, uma vez que os classificadores SVM tendem a generalizar bem mesmo com conjuntos de treino de dimensão moderada [Lot+18; Pas+10].

A primeira etapa da solução envolve a utilização de um dataset previamente fornecido pela docente responsável, no qual os radargramas estão organizados em dois grupos principais: *no-target*, onde não existem hipérboles, e *target*, onde estas estão presentes. Esta estrutura inicial permite concentrar o esforço no desenho das janelas de análise e na construção das máscaras que identificam quais as regiões relevantes dos radargramas a utilizar durante o treino do classificador.

Segue-se o pré-processamento, um componente crítico devido à natureza ruidosa dos dados GPR. Radargramas reais apresentam variações de amplitude, artefactos instrumentais e interferências que dificultam a deteção automática [Dan04; Jol09]. Assim, são aplicadas técnicas como normalização, remoção de fundo, filtragem passa-banda e aumento de contraste. Estas operações reduzem a variabilidade indesejada e realçam características estruturais essenciais, facilitando a extração de padrões discriminativos.

A etapa de extração de características é realizada com base na técnica CSP, adaptada ao domínio dos radargramas. Este método consiste na construção de matrizes de covariância a partir de janelas rotuladas como *target* e *no-target*, seguida da determinação de projecções lineares que maximizam a variância para uma classe e a minimizam para a outra. O resultado é um conjunto compacto de componentes preditivas que capturam diferenças estruturais relevantes entre as duas categorias. Esta combinação CSP+SVM é amplamente documentada em outros domínios de processamento de sinal, sobretudo pela sua capacidade de enfatizar padrões característicos de cada classe mesmo em ambientes ruidosos [Lot+18; Wan+15].

O classificador SVM é então treinado em regime supervisionado, usando as janelas extraídas e respetivos rótulos. Serão avaliados diferentes tipos de kernel, nomeadamente linear e RBF, bem como os respetivos hiperparâmetros, como o parâmetro de regularização C e a largura γ no caso do kernel RBF. A seleção da melhor configuração será efetuada através de validação cruzada, privilegiando métricas como F1-score, precisão e revocação, de forma a reduzir o risco de sobreajustamento.

Durante a fase de inferência, o radargrama é percorrido por uma janela deslizante, aplicando-se em cada posição a mesma transformação CSP e o classificador SVM. O resultado pode ser representado sob a forma de um mapa de probabilidade ou de uma marcação discreta das regiões onde a probabilidade de ocorrência de hipérboles é elevada. Este tipo de mapa servirá, numa fase posterior, como base para a derivação de marcadores mais explícitos da localização aproximada das hipérboles.

A avaliação do desempenho utilizará métricas padrão de classificação, como F1-score, precisão, revocação e área sob a curva ROC, complementadas com medidas específicas de deteção sempre que apropriado. Estas métricas serão fundamentais para comparar o desempenho do sistema CSP+SVM com abordagens clássicas, como

a Transformada de Hough ou o *template matching*, permitindo confirmar se o método proposto apresenta melhorias consistentes face às soluções tradicionais [ZM06; GL00].

Por fim, a solução proposta integra diversas competências adquiridas ao longo do curso, incluindo processamento de imagem, análise de dados e implementação de modelos de *machine learning*. O objetivo é que o sistema desenvolvido represente não apenas uma prova de conceito funcional, mas também uma base sólida para futuras extensões, potenciando comparações mais amplas com outros classificadores ou até com arquiteturas de aprendizagem profunda.

5 - Resultados e Análise

5.1 Resultados Preliminares

A avaliação do sistema proposto foi realizada com base em métricas clássicas de classificação supervisionada, nomeadamente precisão (*precision*), revocação (*recall*), F1-score e acurácia global. Estas métricas permitem caracterizar de forma abrangente o desempenho do modelo na distinção entre regiões contendo hipérboles (*target*) e regiões sem alvos relevantes (*non-target*).

Os resultados obtidos evidenciam um desempenho global positivo, com uma acurácia aproximada de 84%. A análise por classe revela um comportamento assimétrico: a classe *non-target* apresenta melhores resultados, com um F1-score de 0.87, enquanto a classe *target* apresenta um F1-score de 0.79.

Esta diferença de desempenho pode ser explicada por vários fatores. Em primeiro lugar, as regiões *non-target* tendem a apresentar padrões mais homogêneos e menos estruturados, sendo mais facilmente distinguíveis após a transformação de Radon e extração de características. Por outro lado, as hipérboles podem apresentar variações significativas em termos de forma, amplitude e contraste, dependendo das propriedades do solo e das condições de aquisição, o que dificulta a sua modelação.

A análise da matriz de confusão revela a presença de falsos negativos na classe *target*, indicando que algumas hipérboles não são corretamente detetadas pelo modelo. Este fenómeno pode estar associado à presença de ruído, hipérboles incompletas ou sobreposição de reflexões, situações frequentemente observadas em radargramas reais.

Por outro lado, a precisão relativamente elevada sugere que o modelo apresenta um comportamento conservador, privilegiando a redução de falsos positivos. Isto significa que, quando uma região é classificada como *target*, existe uma elevada probabilidade de essa deteção estar correta, o que é desejável em aplicações onde deteções erradas podem ter consequências operacionais.

Adicionalmente, a validação cruzada (5-fold) apresentou um F1-score médio de aproximadamente 0.78 com baixo desvio padrão, indicando que o modelo possui uma capacidade de generalização consistente e não apresenta sinais evidentes de sobreajustamento.

No geral, os resultados demonstram que a combinação CSP + SVM constitui uma abordagem eficaz e computacionalmente leve para a deteção de hipérboles em dados GPR, sendo particularmente adequada em cenários com datasets limitados.

5.1.1 Discussão dos Resultados

A análise dos resultados permite retirar conclusões relevantes sobre o comportamento do modelo e a adequação da abordagem proposta.

Em primeiro lugar, observa-se que a utilização da Transformada de Radon, seguida de extração de características com CSP, permite capturar padrões estruturais relevantes presentes nos radargramas. Esta transformação reduz a complexidade espacial do problema, convertendo-o num domínio mais adequado à análise de sinais.

No entanto, esta abordagem apresenta limitações. A Transformada de Radon, embora eficaz na deteção de estruturas lineares e curvas, pode introduzir perda de informação espacial local, especialmente em regiões onde as hipérboles estão parcialmente ocultas ou deformadas.

A técnica CSP, por sua vez, assume que as diferenças entre classes podem ser capturadas através da variância dos sinais projetados. Embora esta hipótese seja válida em muitos casos, pode não ser suficiente para capturar padrões mais complexos presentes em cenários reais.

Outro aspecto relevante prende-se com o desequilíbrio relativo entre classes e a natureza dos dados. As hipérboles representam estruturas específicas e relativamente raras no contexto global dos radargramas, o que pode influenciar a capacidade do modelo em generalizar corretamente.

A presença de falsos negativos sugere que o modelo pode não estar a capturar todas as variações possíveis das hipérboles, especialmente aquelas com baixo contraste ou fortemente degradadas por ruído.

Apesar destas limitações, o modelo apresenta uma boa relação entre desempenho e custo computacional, sendo capaz de produzir resultados competitivos sem necessidade de grandes volumes de dados ou recursos computacionais elevados.

5.2 Base de Dados

A base de dados utilizada neste trabalho foi construída a partir de radargramas GPR provenientes de diferentes campanhas de aquisição, incluindo dados experimentais e dados públicos disponibilizados em repositórios científicos.

Foram utilizados dois conjuntos principais de dados: um conjunto de radargramas não processados, obtidos em diferentes campanhas de aquisição com sistemas GSSI, e um conjunto adicional de dados previamente processados, provenientes de arquivos públicos.

Os dados foram adquiridos com diferentes configurações experimentais, incluindo antenas de várias frequências (50 MHz, 200 MHz, 400 MHz e 900 MHz), permitindo capturar diferentes profundidades e resoluções. As campanhas de aquisição foram realizadas em diversos tipos de terreno, incluindo superfícies asfaltadas, solos arenosos e zonas com vegetação, garantindo diversidade e realismo no dataset.

Os dados utilizados neste trabalho incluem radargramas provenientes de diferentes fontes, incluindo datasets públicos disponibilizados pelo U.S. Geological Survey (USGS) [US 20].

5.2.1 Estatísticas da Base de Dados

A base de dados utilizada neste trabalho é composta por aproximadamente 3200 amostras, extraídas de radargramas GPR e classificadas em duas categorias: *target* e *non-target*. A distribuição das amostras encontra-se apresentada na Tabela 5.1.

Table 5.1: Estatísticas da base de dados utilizada

Classe	Número de amostras	Percentagem
Target	1200	38%
Non-target	2000	62%
Total	3200	100%

5.2.2 Construção da Base de Dados

A base de dados foi construída através de um processo de extração de patches a partir dos radargramas. Cada patch corresponde a uma região local da imagem, com dimensão fixa, permitindo capturar padrões estruturais relevantes.

As amostras foram rotuladas manualmente em duas classes:

- *target*: regiões contendo hipérboles;
- *non-target*: regiões sem presença de alvos.

Foram selecionadas apenas amostras claras e bem definidas, evitando regiões ambíguas ou com ruído excessivo que pudesse comprometer o processo de aprendizagem.

5.2.3 Pré-processamento

Os dados foram sujeitos a várias etapas de pré-processamento com o objetivo de melhorar a qualidade do sinal e reduzir interferências. Estas etapas incluem:

- normalização das intensidades;
- filtragem para redução de ruído;
- realce de contraste;
- remoção de fundo.

5.2.4 Transformação para Domínio de Sinais

Antes da extração de características, os patches foram convertidos em sinais 1D através da Transformada de Radon, utilizando projeções em múltiplos ângulos.

Este procedimento permite transformar o problema de análise de imagem num problema de análise de sinais, facilitando a aplicação de técnicas como CSP e melhorando a capacidade de discriminação entre classes.

5.3 Melhorias Futuras

Apesar dos resultados obtidos serem promissores, existem várias estratégias que poderão ser exploradas para melhorar o desempenho do sistema.

5.3.1 Melhoria do Pré-processamento

A utilização de técnicas mais avançadas de remoção de ruído, como transformadas de wavelets ou filtros adaptativos, poderá melhorar significativamente a qualidade dos dados de entrada.

5.3.2 Normalização Avançada

A aplicação de técnicas de normalização mais robustas, como normalização z-score ou normalização baseada em histogramas, poderá reduzir a variabilidade entre amostras.

5.3.3 Otimização de Hiperparâmetros

A afinação sistemática dos parâmetros do SVM (parâmetro C , kernel e parâmetros associados) poderá melhorar a capacidade de generalização do modelo.

5.3.4 Melhoria das Características

A exploração de outras técnicas de extração de características, como cumulantes de ordem superior ou descritores baseados em textura, poderá aumentar o poder discriminativo do modelo.

5.3.5 Aumento do Dataset

O aumento do número de amostras e a utilização de técnicas de *data augmentation* permitirão melhorar a robustez do modelo e reduzir o risco de sobreajustamento.

5.3.6 Integração de Contexto Espacial

Uma limitação da abordagem atual é a ausência de contexto espacial global. A inclusão de informação espacial adicional poderá melhorar a detecção de hipérboles incompletas ou parcialmente ocultas.

5.3.7 Comparação com Deep Learning

Como trabalho futuro, será relevante comparar esta abordagem com modelos baseados em redes neurais convolucionais, avaliando o compromisso entre desempenho, complexidade e requisitos computacionais.

6 - Calendário

O planeamento apresentado nesta secção reflete o estado atual do Trabalho Final de Curso (TFC), bem como as atividades já realizadas e o trabalho remanescente. Nesta fase, o projeto encontra-se já na etapa de implementação e validação inicial do modelo, tendo sido possível obter resultados preliminares. O objetivo deste capítulo é apresentar o progresso realizado, identificar dificuldades encontradas e definir as próximas etapas necessárias para a melhoria e consolidação do sistema desenvolvido.

6.1 Progresso Atual

Até ao momento foram concluídas várias tarefas fundamentais para o desenvolvimento do projeto. A revisão bibliográfica permitiu consolidar conhecimento sobre o funcionamento do GPR, a formação de hipérboles e os principais métodos de deteção automática, incluindo abordagens clássicas e técnicas de aprendizagem automática. Foi também organizada a base de dados, distinguindo claramente entre classes *target* e *non-target*, e implementado um processo de extração de janelas (patches) a partir dos radargramas.

Na fase de implementação, foi desenvolvido um pipeline completo que inclui:

- aplicação da Transformada de Radon para conversão das imagens em sinais 1D;
- extração de características com a técnica CSP;
- treino de um classificador SVM para distinção entre classes.

Foram ainda realizados testes experimentais iniciais, tendo sido obtidos resultados preliminares com desempenho global satisfatório, nomeadamente uma acurácia aproximada de 84%. Este progresso permite afirmar que o sistema se encontra funcional, estando atualmente na fase de análise e melhoria do modelo.

6.2 Dificuldades e Ajustes

Durante o desenvolvimento do projeto foram identificadas várias dificuldades que motivaram ajustes ao plano inicial. Uma das principais dificuldades prendeu-se com a variabilidade dos radargramas, que apresentam diferentes níveis de ruído, contraste e qualidade de sinal. Esta variabilidade tem impacto direto na capacidade do modelo em generalizar corretamente. Outra dificuldade relevante relaciona-se com a definição das janelas utilizadas no treino. A distinção entre regiões com e sem hipérboles exige validação cuidadosa, sendo necessário evitar a inclusão de amostras ambíguas que possam comprometer o desempenho do classificador. Adicionalmente, verificou-se que algumas hipérboles não são corretamente detetadas pelo modelo, sobretudo em cenários com baixo contraste ou forte presença de ruído, o que evidencia limitações na representação das características extraídas.

Face a estas dificuldades, foram introduzidos ajustes ao processo, incluindo:

- melhoria das técnicas de pré-processamento;
- revisão da seleção de amostras;
- análise crítica dos resultados obtidos.

Estes ajustes permitiram melhorar a estabilidade do modelo e preparar a fase seguinte de otimização.

6.3 Plano de Trabalho Inicial

O trabalho remanescente será organizado em quatro blocos principais: implementação, treino, avaliação e documentação. A Tabela 6.1 apresenta o plano detalhado segundo uma perspectiva de gestão de projeto, incluindo tarefas, duração prevista, dependências e entregáveis esperados.

Período	Atividades Planeadas
Outubro 2025	Início formal do projeto; revisão de literatura sobre GPR e hipérboles; enquadramento do problema; definição de objetivos e pergunta de investigação.
Novembro 2025	Estudo aprofundado do estado da arte em métodos clássicos e <i>deep learning</i> para deteção de hipérboles; consolidação do capítulo de benchmarking.
Dezembro 2025	Integração e análise do dataset fornecido pela docente, já categorizado em <i>target</i> e <i>no-target</i> ; testes iniciais de pré-processamento e definição das janelas a utilizar para treino.
Janeiro 2026	Implementação do pipeline de pré-processamento; extração preliminar de características através de CSP; preparação do ambiente de treino do classificador SVM.
Fevereiro 2026	Treino do modelo; afinação de hiperparâmetros; realização de experiências exploratórias e registo sistemático dos resultados obtidos.
Março 2026	Avaliação detalhada do desempenho do classificador; cálculo de métricas (F1-score, precisão, revocação, ROC-AUC); comparação com métodos tradicionais; identificação de limitações.
Abril 2026	Redação do relatório final; integração de gráficos, tabelas e figuras; organização de anexos e descrição completa da metodologia.
Mai 2026	Revisão do documento; correções formais e técnicas; eventual recolha de feedback de terceiros; preparação da apresentação e defesa.
Junho 2026	Entrega final do Trabalho Final de Curso (TFC) e realização da defesa.

Table 6.1: Plano de trabalho e cronograma proposto para o remanescente do TFC.

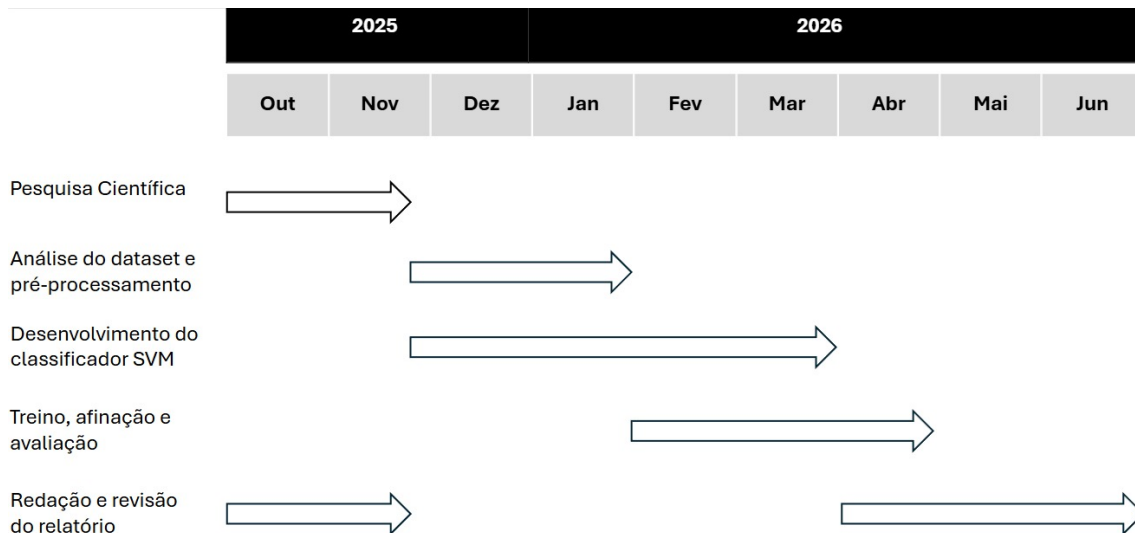


Figure 6.1: Calendário de atividades do TFC entre Outubro de 2025 e Junho de 2026.

6.4 Plano de Trabalho Atual

Tendo em conta o estado atual do projeto, o trabalho remanescente incide principalmente na melhoria e validação do modelo desenvolvido.

As próximas etapas incluem:

- otimização do pré-processamento dos dados, incluindo técnicas de normalização e filtragem;
- afinação dos hiperparâmetros do classificador SVM;
- melhoria da extração de características, explorando variações do CSP;
- análise detalhada dos erros do modelo (falsos positivos e falsos negativos);
- eventual aumento do dataset ou aplicação de técnicas de data augmentation;
- comparação com outras abordagens, incluindo métodos baseados em deep learning;
- integração dos resultados finais no relatório.

A fase final do projeto será dedicada à consolidação dos resultados, redação do relatório final e preparação da apresentação e defesa do trabalho.

Bibliografia

- [Mat24] João P. Matos-Carvalho. *The Lusófona L^AT_EX Template User's Manual*. Lusófona University. 2024. URL: <https://github.com/jpmcarvalho/UL-Thesis>.
- [Dan04] David J. Daniels. *Ground Penetrating Radar*. 2nd ed. London: Institution of Engineering and Technology, 2004.
- [Jol09] Harry M. Jol. *Ground Penetrating Radar Theory and Applications*. Amsterdam: Elsevier, 2009.
- [JL21] Wei Jiang and Peng Li. “GPR Hyperbola Detection Based on Machine Learning: A Comparative Study”. In: *Remote Sensing* 13.3 (2021), pp. 410–425.
- [HS22] Y. Huang and C. Su. “Edge-based Hyperbola Detection in GPR Images”. In: *Journal of Applied Geophysics* (2022), pp. 1–12.
- [ZM06] Yong Zhu and Zheng Ma. “Hyperbola Recognition in GPR Data Using the Hough Transform”. In: *Journal of Applied Geophysics* (2006), pp. 1–10.
- [GL00] Paolo Gamba and Silvia Lossani. “Neural Detection of Buried Pipes from Ground Penetrating Radar Data”. In: *IEEE Transactions on Geoscience and Remote Sensing* 38.2 (2000), pp. 790–797.
- [Pas+10] E. Pasolli et al. “Automatic classification of GPR signals”. In: *IEEE Transactions on Geoscience and Remote Sensing* (2010).
- [Xia+18] D. Xiang et al. “Automatic GPR target detection using machine learning”. In: *IEEE Transactions on Geoscience and Remote Sensing* (2018).
- [Lot+18] Fabien Lotte et al. “A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update”. In: *Journal of Neural Engineering* 15.3 (2018).
- [Wan+15] Y. Wang et al. “Buried object detection using GPR and SVM”. In: *Journal of Applied Geophysics* (2015).
- [US 20] U.S. Geological Survey. *Ground Penetrating Radar Data*. Accessed: 2026. 2020. URL: <https://www.usgs.gov>.

Glossário

CSP Common Spatial Patterns. 2, 3, 8–10, 12–14, 16

GPR Ground Penetrating Radar. 8, 23

SVM Support Vector Machine. 2, 3, 8–10, 12–14, 16

TFC Trabalho Final de Curso. 23