

# Classificação de pragas e doenças pela sintomatologia de folhas em Tomateiros

## *Image classification problem*

### *Problem Statement*

Um dos problemas agronómicos que pode ser solucionado com recurso a machine learning passa pela identificação e classificação de pragas e doenças usando modelos de deep learning aplicados à classificação de imagem. Esta solução proporciona por exemplo um auxílio no acompanhamento técnico providenciado a agricultores facilitando o acompanhamento e apoio técnico.

A proposta deste trabalho consiste na especificação de uma cultura, tomateiro, e na tentativa de implementação de um sistema que através de imagens de sintomas identifique a praga/doença do tomateiro num modelo preditivo de deep learning com recurso à classificação de imagens. Numa segunda fase, a ideia será que a predição dessa praga ou doença direcione para a gama de tratamentos possíveis de realizar na cultura. Integrando assim a identificação do problema com as soluções existentes de combate/cura.

### *Challenges*

Um dos desafios associados a este tipo de problemas passa pela disponibilidade de dados/fotos associadas à sintomatologia em diferentes culturas. Por essa mesma razão, é necessário especificar o problema a uma cultura para o qual seja possível obter esses dados/imagens. Essas imagens devem ainda ser de qualidade de forma a permitir o desenvolvimento de um modelo forte e capaz de identificar com certeza a praga ou doença presente na cultura. Neste caso iremos focar numa das culturas com maior importância em Portugal, o tomate.

Importa ainda referir que a criação de modelos com base em análise de imagens é algo que requer bastante tempo para correr o código factor que pode de certa forma condicionar a execução deste trabalho.

Outros desafios identificados passam pela garantia de qualidade do data set e das labels pois só com esta qualidade é possível garantir o correcto funcionamento do modelo. Poderá ainda haver uma distribuição desigual nas classes que poderá levar a tendências (biased model predictions) no modelo. A eventual ligação entre o modelo de predição e a recomendação de aplicação de produto fitofarmaco/tratamento poderá ser de mais difícil execução em relação ao que foi inicialmente previsto.

### *Dataset*

A forma mais exequível de realizar este projecto academicamente passa pela utilização de fotografias em repositórios online. Naturalmente, a escalabilidade do projecto e a adaptação a outras culturas passa pela obtenção de mais dados fotográficos que sairá do âmbito deste trabalho.

Foi possível identificar um DataSet no Kaggle que agrupa diferentes pragas e doenças em tomateiros que irá ser o data set utilizado neste projecto.

### *Method or Algorithm*

Iremos utilizar a CNN – Convolution Neural Network baseado na arquitectura ResNet50 pre-trained em ImageNet para desenvolver este modelo de machine learning de classificação de imagens.

### *Evaluation*

A análise dos resultados será baseada em:

Metricas de performance tais como accuracy, precision, recall, e F1-score.

Confusion Matrix

Testes Estatísticos

Class Weights

Uma vez que o dataset já se encontra dividido em test e train será difícil executar uma cross validation.