

Inteligencia Computacional - 2023/2024

Fase 1 - Desenvolvimento sustentável

Identificação da saúde de produtos vegetais

Introdução

À medida que o planeta enfrenta desafios crescentes relacionados à segurança alimentar, e problemas com o uso responsável dos recursos naturais e mitigação das mudanças climáticas, a identificação precoce e eficaz de doenças em produtos vegetais como plantas e frutas emerge como uma solução vital no âmbito do desenvolvimento sustentável. Esta capacidade de detetar e responder a doenças nas culturas agrícolas não apenas promove práticas agrícolas mais eficientes, mas também desempenha um papel fundamental na preservação da biodiversidade, na redução do uso de produtos químicos e na promoção de sistemas alimentares sustentáveis, sendo este trabalho, o objetivo para a resolução destes problemas.

Datasets

No decorrer da pesquisa para este trabalho, encontrei um conjunto de dados que se alinha perfeitamente com o tema do projeto. Trata-se do 'New Plant Diseases Dataset,' um dataset que contém 87000 de imagens que podem ser utilizadas e categorizadas no uso deste projeto.

Este dataset tem 38 classes, onde irei escolher 10 para a realização deste trabalho, onde poderá ser explorado um leque de possibilidades.

Segue-se o link deste dataset em baixo:

- <https://www.kaggle.com/datasets/pavansanagapati/images-dataset>
- <https://www.kaggle.com/datasets/vipooooool/new-plant-diseases-dataset/data>

Testes e Eficácia da Rede

Treinar uma rede neural com MLPs (Multi-Layer Perceptrons) para identificar doenças em plantas envolve várias etapas, incluindo a preparação dos dados, o design da rede, o treino e a validação do desempenho da rede. Este processo irá ser realizado fazendo o carregamento dos conjuntos de dados, pré processamento das imagens, agrupamento das imagens por classes, será feita a escolha do número de camadas, o número de neurônios em cada camada e a função de ativação adequada.

Relativamente ao treino da rede, pode ser utilizado um algoritmo de otimização, como o Gradiente Descendente Estocástico (SGD) ou o Adam, e definida a taxa de aprendizagem.

O treino ocorrerá usando o conjunto de treino, e irei ajustar os pesos da rede para minimizar a função de perda. Este processo será repetido por várias épocas até que o desempenho pare de melhorar ou pare.

O conjunto de validação será utilizado para monitorar o desempenho da rede, o que permitirá avaliar se a rede está a aprender bem e detectar possíveis problemas de overfitting.

Por último, será feita uma avaliação final do modelo usando o conjunto de testes. Essa avaliação, irá conter dados que a rede não viu durante o treino, garantindo uma medida imparcial do desempenho na identificação de doenças em plantas, e para quantificar o desempenho do modelo, irei utilizar métricas de avaliação adequadas, como a matriz de confusão.

Conclusão

Uma das principais vantagens deste trabalho é a capacidade de monitorar e diagnosticar doenças de forma precoce, permitindo aos agricultores agir de maneira mais eficaz, reduzindo o uso desnecessário de produtos químicos e minimizando o impacto ambiental. Além disso, contribui para a preservação da biodiversidade, promovendo práticas agrícolas mais sustentáveis.