

# Report of Potato Pest Classifier

## Introdução

Como identificar as pragas das batatas? A batata é a quinta cultura mais produzida no mundo, o que faz com que tenha uma grande importância na segurança alimentar global. De acordo com o relatório do Ministério da Agricultura de Bangladesh, existem 19 pragas na cultura da batata. No entanto, devido às limitações de imagens no Google Imagens, portais online, jornais e amostras partilhadas publicamente, focamos-nos em oito tipos mais comumente prejudiciais de pragas da batata. Esses oito tipos de pragas são mais nocivos em comparação com os outros, o que também justifica a escolha destes. Portanto iremos desenvolver um modelo de classificação de imagens que ajudem os agricultores a identificar a praga que esteja a ameaçar a cultura.

## Dados

Para treinar o modelo fomos buscar um dataset do kaggle que continha 495 fotografias de 8 espécies de pragas: *Agrotis ipsilon* (Hufnagel), *Amrasca devastans* (Distant), *Aphis gossypii* Glover, *Bemisia tabaci* (Gennadius), *Brachytrypes portentosus* Lichtenstein, *Epilachna vigintioctopunctata* (Fabricius), *Myzus persicae* (Sulzer), *Phthorimaea operculella* (Zeller). Fizemos algumas transformações como saturação, normalização e redimensionamento. em baixo fica um exemplo das fotografias antes e depois de serem processadas.

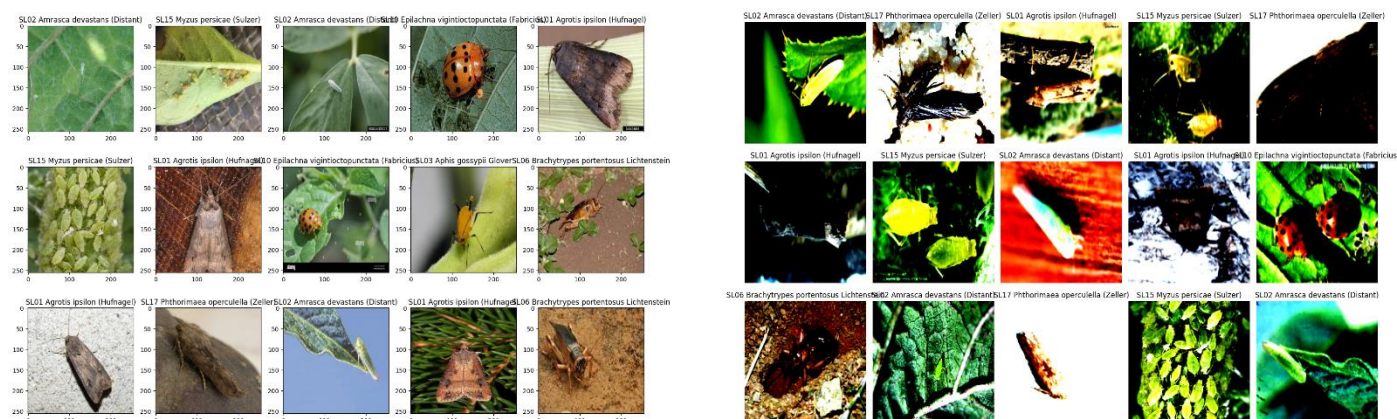


Fig1: antes de transformação e depois

## Organização de dados

Das 495 fotografias selecionamos 70% para o conjunto de treino, 20% para o conjunto de validação e 10% para o conjunto de teste.

Espécie	Conjunto de Treino	Conjunto de Validação	Conjunto de Teste
<i>Agrotis ipsilon</i> (Hufnagel)	97	27	15
<i>Amrasca devastans</i> (Distant)	43	12	7
<i>Aphis gossypii</i> Glover	25	7	5
<i>Bemisia tabaci</i> (Gennadius)	24	7	4
<i>Brachytrypes portentosus</i> Lichtenstein	24	7	4
<i>Epilachna vigintioctopunctata</i> (Fabricius)	49	14	7
<i>Myzus persicae</i> (Sulzer)	52	15	8
<i>Phthorimaea operculella</i> (Zeller)	29	8	5
TOTAL	343 (70%)	97 (20%)	55 (10%)

Fig2: Tabela de organização de dados

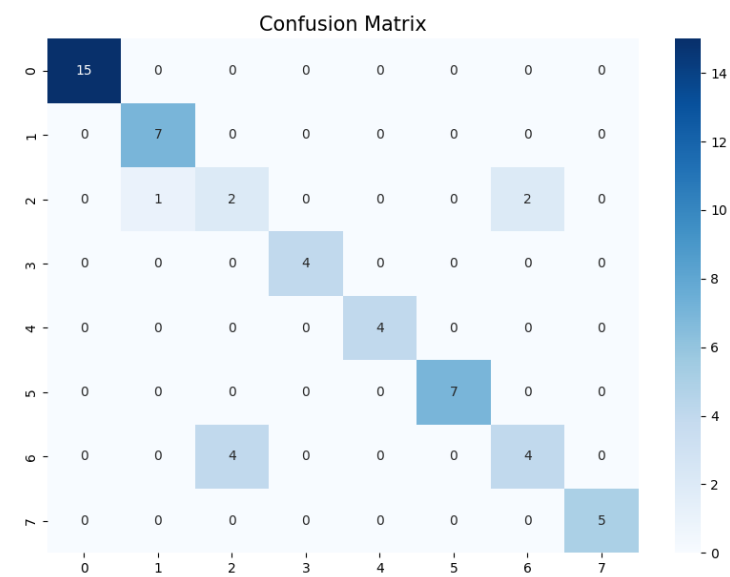
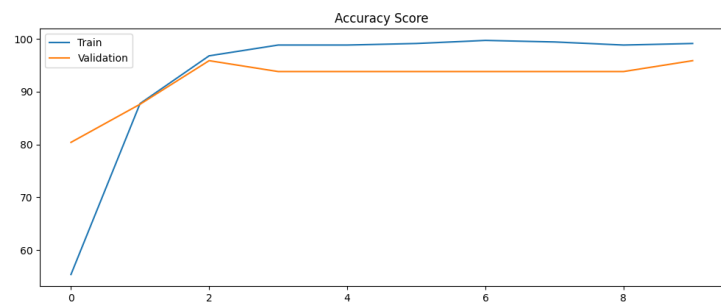
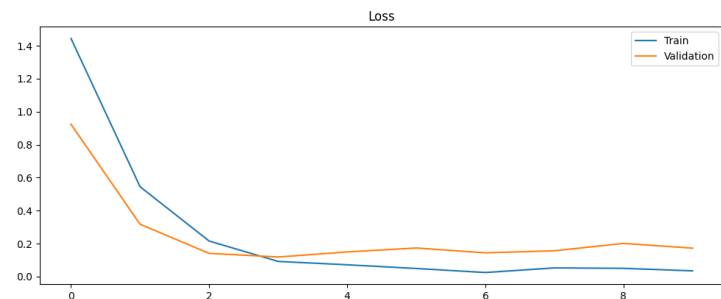
## Métodos

Usamos o modelo *ResNet 50* e como hiperparametros usamos learning rate de 0.0001, epoch 10 e batch size 32, para o modelo alcançar uma precisão maior e apenas durar aproximadamente 4 a 5 minutos de treino.

## Resultados

Test accuracy: 87,27%

Species	Precision	Recall	F1-score	Support
0	1,00	1,00	1,00	15
1	0,88	1,00	0,93	7
2	0,33	0,40	0,36	5
3	1,00	1,00	1,00	4
4	1,00	1,00	1,00	4
5	1,00	1,00	1,00	7
6	0,67	0,5	0,57	8
7	1,00	1,00	1,00	5
Accuracy			0,87	55
Macro avg	0,85	0,84	0,84	55
weighted avg	0,88	0,87	0,87	55



## Análise

Podemos ver que o nosso modelo converge a partir da terceira epoch podendo ter finalizado o seu treino no terço do tempo que demorou e de forma a reduzir o risco de overfitting, se bem que o validation set prevêem de certa forma este problema. Experimentamos diferentes learning rate e epochs para atingir a maior precisão no menor tempo possível.

A precision, Recall e a F1-score foram apenas baixas na *Aphis gossypii* Glover **2** e na *Myzus persicae* (Sulzer) **6** possivelmente devido à particularidade de ter cores diferentes ou pelo baixo número de imagens usadas para o seu treino e validação, o que dificulta a sua identificação.

Já nas restantes espécies acontece o oposto nestas métricas por ter mais elementos que possam discriminar os insetos como no caso das pintinhas na *Epilachna vigintioctopunctata* (Fabricius) **5** ou da forma característica da *Agrotis ipsilon* (Hufnagel) **0** e por estas ocorrerem mais vezes no conjunto de treino e validação como aparece na métrica support e na confusion matrix.

A média macro tira a média aritmética das métricas individuais de cada classe, tratando todas de forma igual e a weighted average é a ponderada que neste caso são relativamente boas.

## Referências

Md. Simul Hasan Talukder, Rejwan Bin Sulaiman, Mohammad Raziuddin Chowdhury, Musarrat Saber Nipun, and Taminul Islam. (2023). Potato Pests dataset [Data set]. Kaggle. <https://doi.org/10.34740/KAGGLE/DSV/6246312>

**Contribuições:** Todos os membros do grupo contribuíram para o trabalho de igual forma e acompanharam todas as etapas.

Trabalho elaborado por:  
Diogo Simão, 25585;  
Miguel Ferreira, 23382;  
Rubén Torrado, 26042