

# An End-to-End Deep Learning System for Hop Classification

Pedro Castro, Gladston Moreira, *Member, IEEE*, and Eduardo Luz

**Abstract**—Automatic classification of plant species is a very challenging and widely studied problem in the literature. Distinguishing different varieties within the same species is an even more challenging task although less explored. Nevertheless, for some species distinguishing the varieties within the species can be of paramount importance.

Hops, a plant widely used in beer production, has over 250 cataloged varieties. Although the varieties have similar appearances, their chemical components, which influence the aroma and flavor of the drink, are quite heterogeneous. Therefore, it is important for producers to distinguish which variety the plant belongs to in a simple manner.

In this work, an end-to-end deep learning system is proposed to automate the task of hop classification. Five architectures are proposed and evaluated with an uncontrolled environment dataset that includes 12 varieties of hops on 1592 images, from three different cell phone cameras. The best architecture automatically detects the hop leaves on the image and performs the classification using the information of up to 10 leaves. The method achieved an accuracy of 95.69% with an inference time of 672ms. To reach such figures, state-of-the-art convolutional blocks were explored along with data augmentation techniques. Our results show that the system is robust and has a low computational cost.

**Index Terms**—Hop, Convolutional neural network, Leaf recognition, Data augmentation.

## I. INTRODUÇÃO

**O** *Humulus lupulus* L., também conhecido como lúpulo, é uma trepadeira cujas flores são um dos principais componentes utilizados na fabricação de cerveja em todo o mundo. Ele entrega sabor, amargor e aroma à bebida [1] e auxilia na estabilização da espuma [2]. Além disso, tem um papel importante na conservação da mesma devido às suas propriedades antimicrobianas e antioxidantes, que, recentemente, têm sido estudadas para a preservação de alimentos [3]. Em menor escala, possui aplicações farmacológicas [4], [5] na produção de medicamentos e cosméticos. Dentre os usos medicinais estão seus efeitos anti-inflamatórios, antitérmicos, ansiolíticos e sedativos [6]. Ainda, é possível encontrar o consumo alimentício da planta [7], [8].

Existem mais de 250 variedades catalogadas [9] desta planta e, dentre os principais atributos que diferem uma da outra, estão os alfa-ácidos, beta-ácidos e óleos essenciais [10]. Esses componentes conferem à cerveja uma combinação única de características, podendo, inclusive, influenciar na categoria da mesma.

No Brasil, a produção de lúpulo está em estágio inicial, com um dos primeiros experimentos bem-sucedidos de cultivo tendo ocorrido em 2016 [11], [12]. Apesar de essa estar

em crescimento, a maior parte do insumo consumido pelas cervejarias do país ainda é importado [13]. Além disso, uma fatia considerável do cultivo nacional ainda é incipiente. Com isso, poucos viveiros possuem certificado de origem do lúpulo e, portanto, podem gerar um produto com características diferentes do esperado. Dessa maneira, a distinção de a qual variedade a planta pertence é importante para o produtor.

Existem muitas formas documentadas de se identificar a variedade do lúpulo a partir dos seus ácidos e óleos essenciais utilizando métodos como cromatografia, espectrometria de massa, eletroforese capilar e ressonância magnética nuclear [14]–[18]. No entanto, esses utilizam equipamentos e processos físicos e químicos custosos, complexos, inacessíveis ou indisponíveis aos produtores.

Atualmente, muitos modelos de redes neurais convolucionais (CNN) têm sido aplicados à classificação de espécies de plantas, flores e frutas a partir de imagens [19]. Uma vez que podem ser utilizadas fotos sem qualquer tipo de processamento, esse se torna um processo simples, barato e acessível a qualquer pessoa com um dispositivo móvel com câmera. Esses métodos têm se mostrado eficientes para a identificação de espécies. Porém, para o lúpulo, o interesse reside na classificação da variedade, que, em taxonomia, é um nível inferior à espécie. De acordo com [20], nas variedades intraespecíficas, os indivíduos possuem características de genótipo ou fenótipo em comum. No caso do lúpulo, a morfologia visual observável é bem similar, como destacado na Fig. 1, que contém exemplos de folhas jovens de duas variedades distintas.



Fig. 1. Exemplo de duas variedades distintas de Lúpulo com morfologia visual similar.

Na literatura, pouco tem-se explorado a diferenciação entre variedades pertencentes à mesma espécie. Este estudo objetiva estabelecer uma *baseline* para a classificação de 12 variedades de lúpulo e desenvolver uma arquitetura para identificá-las, com uma boa acurácia, a partir de fotos de folhas obtidas em ambiente não controlado. Propomos aqui uma sistema ponta-a-ponta para realização da tarefa, capaz de segmentar automaticamente as folhas e classificar a espécie. O sistema proposto é baseado em técnicas de aprendizagem profunda e obteve, na base de dados construída para este projeto, acurácia de 95,69%.

As contribuições deste trabalho podem ser sumarizadas como:

- Um sistema ponta-a-ponta para realização da tarefa, baseado em técnicas estado-da-arte;
- Uma abordagem combinando múltiplas visões/folhas para validar a variedade;
- Comparação detalhada de 5 arquiteturas ponta-a-ponta (detecção, extração de características e classificação);
- Construção de uma base de dados com 1592 imagens de 12 variedades Disponível em <http://www.decom.ufop.br/csilab/databases/> <https://github.com/pedrohnc/ufop-hvd-l-oader>;
- Um protocolo de avaliação visando reprodutibilidade do trabalho;
- Um estudo detalhado do impacto de técnicas de aumento de dados para permitir o treinamento de redes profundas;

Este trabalho está dividido obedecendo à seguinte ordem: Seção II, trazendo uma revisão dos trabalhos relacionados; Seção III, detalhando a metodologia utilizada, as arquiteturas propostas e os aumentos de dados aplicados; Seção IV, com os experimentos das arquiteturas, as contribuições individuais dos aumentos de dados e discussão sobre os resultados; e Seção V, encerrando com as conclusões e sugestões de trabalhos futuros.

## II. TRABALHOS RELACIONADOS

Até onde sabemos, este trabalho é o primeiro a abordar classificação automática de variedades de lúpulo. Assim, a revisão de literatura teve foco em trabalhos que realizam identificação automática de variedades de plantas utilizando visão computacional e de espécies de plantas a partir de redes de convolução.

O trabalho proposto em [21] foi um dos primeiros a se preocupar com o reconhecimento automatizado de variedades de plantas. Não há muitas informações sobre o *dataset*, mas é citado que foram testadas variedades de cenouras, couve de bruxelas e salsão. O sistema proposto, chamado Visor, baseia-se em características extraídas da imagem como histogramas (ex.: intensidade média de cada cor e proporção de pixels verdes), medidas morfológicas locais (ex.: número de pontas em cada folha) e globais (ex.: silhueta mais ou menos alongada). O sistema compara as características de uma nova imagem com as informações contidas em uma base de dados, gerando uma pontuação de similaridade. Então, associa a variedade àquela mais similar. Na época da publicação, o sistema ainda estava em construção e não foram disponibilizadas informações claras sobre desempenho e acurácia.

Em [22], os autores desenvolveram um método para a identificação automática de 5 variedades de *Ficus deltoidea* a partir da imagem da folha. Para a construção da base de dados, as folhas foram posicionadas sobre um papel branco e digitalizadas com um escâner. Em seguida, é aplicado um processo de remoção de ruídos. O *dataset* conta com 345 imagens divididas igualmente entre as classes. Para a classificação, foi utilizada extração de características de formato, cor e textura, todas obtidas a partir de funções do Matlab. Então, as características são passadas por uma Análise de Componente Principal e entregues a uma KNN. A melhor acurácia obtida foi de 99,32%.

Algumas das primeiras pesquisas a explorar redes de convolução para a classificação de plantas foram os trabalhos de [23] e [24]. Em ambos, uma CNN foi utilizada para extração de características e, para a classificação, foram testados os seguintes modelos de aprendizagem de máquina: *Extremely Randomized Tree*, *Multilayer Perceptron* e *Support Vector Machine* (SVM). O primeiro estudo foi avaliado com o *dataset* LifeCLEF 2014 e reconheceu próximo de 35% das imagens. Essa base de dados é considerada difícil, pois possui 47.000 imagens com 500 espécies diferentes de ervas, árvores e samambaias, obtidas ao ar livre. O segundo alcançou uma acurácia de 99,6% no *dataset* MalayaKew (MK), que contém imagens de folhas em ambiente controlado de 44 espécies.

Posteriormente, em [25] e [26], utilizou-se uma CNN completa para extração de características e também classificação de plantas das bases de dados PlantCLEF 2015, PlantCLEF 2016 e TARBIL. Para validar os modelos, o primeiro estudo comparou os resultados com uma rede composta por uma Scale-Invariant Feature Transform (SIFT) para extração de características e uma SVM para classificação. O segundo implementou uma rede que utiliza Local Binary Patterns (LBP) e descritores GIST como extratores e também SVM para classificar. Esses trabalhos demonstraram que as redes neurais convolucionais completas tiveram desempenho superior.

A abordagem vista em [27] emprega uma rede de convolução para a classificação do *dataset* LifeClef 2016. Como esse possui imagens de centenas de espécies de plantas com diferentes partes das mesmas, o estudo se destaca por construir uma rede que determina o órgão (ex.: galho, flor, fruta, folha, folha escaneada, tronco e planta inteira) que está sendo avaliado na foto e outra para classificar a planta. Então, os resultados são combinados em uma etapa chamada fusão de características para que a mesma apresente a classificação final. As duas redes compartilham as primeiras camadas de extração de características em baixo nível. O modelo proposto obteve uma acurácia de 68,9%.

Os autores em [28] recorreram a uma ResNet com 26 camadas de profundidade e 8 blocos de construção residuais para classificar 100 espécies de plantas ornamentais em um *dataset* com 10.000 imagens, alcançando uma acurácia de 91,78%. As fotos foram obtidas em ambiente aberto com um *smartphone* no campus da Beijing Forestry University.

O trabalho de [29] utilizou duas versões da GoogleNet (uma original e outra modificada) para classificar folhas de 33 espécies de plantas do *dataset* Flavia, que contém fotos das folhas segmentadas, e algumas extras sem segmentação, totalizando 3.867 imagens. A rede modificada adiciona dois módulos Inception a mais na GoogleNet. A acurácia alcançada foi de 99,6% para a rede original e 99,8% para a modificada.

A proposta em [30] consiste em uma rede denominada Multiscale Fusion Convolutional Neural Network (MSF-CNN), que recebe como entrada a imagem original e várias versões em escalas menores, com objetivo de extrair características diferentes em cada tamanho de imagem. As informações obtidas são fundidas ao longo da rede e, ao final, todas são concatenadas para gerar a predição. A rede é avaliada nos *datasets* MalayaKew e LeafSnap Plant Leaf, atingindo a acurácia de 99,82% e 85,28%, respectivamente. O primeiro

conjunto de dados contém apenas imagens de folhas destacadas em ambiente controlado, enquanto o segundo possui fotos das plantas ao ar livre.

No estudo de [31], os autores sugerem a utilização de uma FasterRCNN padrão e uma modificada, com objetivo de realizar a detecção de folhas em uma imagem junto com a classificação de cada uma. A FasterRCNN modificada possui uma InceptionV2 com *Batch Normalization* no lugar das camadas de convolução que fornecem informação para a Region Proposal Network e para a Region of Interest Pooling com o intuito de oferecer características extraídas em várias escalas da imagem. A rede modificada melhorou a capacidade de detecção de folhas e a acurácia do reconhecimento de suas classes. Contudo, os valores não foram informados.

A abordagem em [32] utiliza uma rede siamesa, onde duas redes convolucionais paralelas, com os mesmos pesos, recebem duas imagens de folhas e aprendem os pesos para ordená-las espacialmente, mantendo próximas as folhas da mesma classe e distantes as folhas de classes distintas. Após a separação, uma KNN é utilizada para a classificação das folhas no espaço aprendido. Os *datasets* para avaliar o modelo foram o Flavia, o Swedish e o Leafsnap e as acurácias alcançadas em cada um deles foram, respectivamente, de 95,32%, 91,37% e 91,75%.

Um modelo de atenção em duas vias foi a escolha do trabalho proposto em [33]. A primeira via é responsável pela detecção da família da planta para identificar a maior região de interesse da imagem através de um mapa de calor. Então, a imagem é recortada nessa região e passa por uma segunda rede que irá classificar a espécie da planta. Nos *datasets* MalayaKew, ICL, FLOWERS102 e CFH Plant, o modelo obteve acurácia de 99,8%, 99,9%, 97,2% e 79,5%, respectivamente.

Em [34], foca-se na classificação de variedades de batata doce. Ele utiliza uma rede baseada em uma ResNet com 34 camadas pré-treinada com a ImageNet. O estudo capturou 3000 imagens de 5 variedades de batata doce. As imagens foram obtidas sob ambiente e iluminação controlados. Elas ainda passaram por um pré-processamento de melhoria de imagem e remoção de áreas irrelevantes. Além disso, os atributos mais importantes das imagens foram destacados. A acurácia obtida foi de 96,33%.

A pesquisa apresentada em [35] trabalha com um *dataset* próprio de mais de 15.000 imagens de folhas contendo 15 espécies de plantas. As imagens foram obtidas partir de uma câmera digital e um aparelho celular. Foi utilizada uma Mask R-CNN para segmentar a folha e uma VGG16 para realizar a classificação, atingindo uma acurácia de 91,5% na base de teste.

O estudo realizado em [36] aborda a classificação de 9 variedades de sementes de milho com 1000 amostras de cada. As imagens foram obtidas em ambiente controlado e passaram por processamento e segmentação. Foi utilizada uma VGG16 para extração de características. Para a classificação foram testadas Redes Neurais, SVM cúbica e quadrática, K-Nearest-Neighbor, Boosted Tree, Bagged Tree e Análise Discriminante Linear. A melhor acurácia foi com a rede neural, atingindo 98,1% de acurácia.

Os autores em [37] propuseram algumas variações de CNN com até 3 camadas convolucionais e 2 densas para classificar

15 variedades de grãos de trigo. Foram geradas 15000 imagens (1000 de cada variedade), a partir de uma câmera de celular em ambiente controlado, com os grãos sobre um fundo branco. O melhor desempenho alcançado foi de de 97,53% acurácia.

Por fim, em [38] foi elaborada uma rede convolucional baseada na VGG16 para classificar 12 variedades diferentes de feijão, distribuídas em 3 espécies. Para a tarefa foram produzidas 2200 imagens de folhas de feijão, em frente e verso, a partir de uma câmera digital. As folhas foram cortadas da planta e fotografadas dentro de uma caixa com iluminação artificial e foco controlado. A acurácia para a diferenciação das variedades foi de 91,37%.

O presente trabalho se diferencia por apresentar uma abordagem ponta-a-ponta, que permite combinar a inferência de várias folhas detectadas na mesma imagem para gerar uma única predição final. Para essa abordagem, considera-se que cada imagem possui apenas uma classe, o que é esperado em fotos de plantações de lúpulo. Além disso, a inferência a partir das folhas é tão importante quanto a da imagem completa e, portanto, ambas são utilizadas. Outras contribuições desse estudo são a comparação de redes para a extração de característica aplicadas ao lúpulo e a investigação das participações individuais de cada aumento de dados na otimização do modelo.

### III. METODOLOGIA

Nesta seção são descritos os protocolos utilizados para obtenção das imagens e construção do *dataset*. Além disso, a metodologia empregada para a tarefa de classificação ponta-a-ponta é detalhada.

#### A. Base de Dados e Protocolo de Avaliação

Foram coletadas 1592 fotos de folhas de lúpulo em 3 viveiros por diferentes pessoas e dispositivos móveis (celular-phone cameras) distintos com o objetivo de aumentar a representatividade da base de dados. As imagens foram obtidas em ambiente natural, com variação de clima, luminosidade, foco, oclusão, resolução, distância e ângulo. As amostras contêm folhas adultas e jovens. Não houve remoção de *outliers* nem de imagens com baixa qualidade. Tampouco foi adotado pré-processamento. A Fig. 2 contém exemplos de cada uma das 12 variedades utilizadas neste trabalho. Para viabilizar reprodutibilidade e futura comparação com a literatura, disponibilizamos publicamente o conjunto de dados registrado sob o DOI [39], denominado UFOP Hop Varieties Dataset (UFOP-HVD). Propomos um protocolo para avaliação dos métodos, em que o *dataset* deve ser dividido em 70% para treino, 15% para validação e 15% para teste. A Tabela I contém o número de imagens por classe de cada conjunto. Os aparelhos utilizados na captura das fotos foram Motorola Moto G7, Samsung Galaxy A11 e Apple iPhone 11. Na Tabela II pode ser conferido um sumário de todas as resoluções de imagens obtidas pelos aparelhos divididas por variedade. A base de dados está disponível em <http://www.decom.ufop.br/csilab/databases/>. Para permitir reprodutibilidade, as imagens já estão particionadas em treino, teste e validação. Os arquivos de anotação também estão disponíveis.

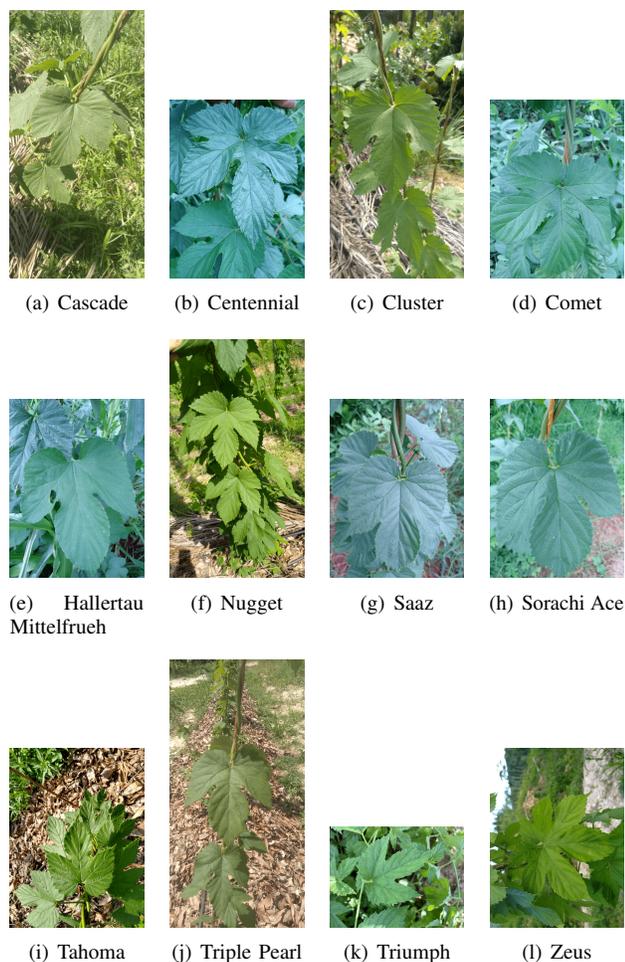


Fig. 2. Exemplos das 12 variedades de Lúpulo utilizadas neste trabalho.

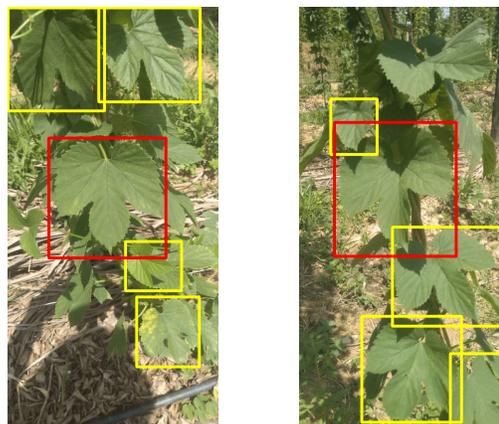
TABELA I

DIVISÃO POR CLASSE/VARIEDADE DO TAMANHO DO CONJUNTO DE DADOS DE TREINO, VALIDAÇÃO E TESTE.

Variedade	Treino	Validação	Teste	Total
Cascade	78	16	16	<b>110</b>
Centennial	89	18	18	<b>125</b>
Cluster	73	15	15	<b>103</b>
Comet	121	25	25	<b>171</b>
Hallertau Mittelfrueh	99	21	21	<b>141</b>
Nugget	82	17	17	<b>116</b>
Saaz	122	25	25	<b>172</b>
Sorachi Ace	137	28	28	<b>193</b>
Tahoma	85	17	17	<b>119</b>
Triple Pearl	84	17	17	<b>118</b>
Triumph	71	15	15	<b>101</b>
Zeus	87	18	18	<b>123</b>
<b>Total</b>	<b>1128</b>	<b>232</b>	<b>232</b>	<b>1592</b>

Cada imagem pode conter uma ou mais folhas de um lúpulo da mesma variedade, as quais foram demarcadas com *bounding boxes* todas que, subjetivamente, estavam mais completas, nítidas e com menor oclusão, conforme os exemplos da Fig. 3. A folha com maior área foi rotulada como principal (*bounding box* vermelha), enquanto as demais, como folhas extras (em amarelo). As anotações das caixas delimitadoras estão disponibilizadas com o *dataset* no formato Pascal VOC

[40].

Fig. 3. Metodologia de rotulagem: a folha com maior área foi rotulada como principal (*boundingbox* vermelha), enquanto as demais, como folhas extras (amarelo).

A partir do *dataset* original, foi gerado um novo chamado *Cropped Dataset* contendo todas as folhas recortadas pelas *bounding boxes*. O número de folhas por classe nessa base de dados pode ser conferido na tabela III, enquanto a média de folhas por imagem está representada na tabela IV.

### B. Blocos para Extração de Características/Modelos Base

Antes de introduzir as arquiteturas para classificação ponta-a-ponta, é preciso detalhar os modelos base que as compõem, presentes na Fig. 4. Ambos recebem como entrada uma imagem redimensionada para 300x300 enviada a um bloco de extração de características [41]–[43], que é um conjunto de camadas de convolução capaz de produzir informações que servem como base para as camadas de classificação da rede. Neste trabalho, os blocos de extração de características utilizados foram as partes convolucionais dos modelos InceptionV3 [44], ResNet50 [45] e EfficientNetB3 [46], obtidas removendo-se suas camadas de classificação. A saída do bloco passa por uma *Global Average Pooling* [47] e está ligada a uma camada densa com 1024 neurônios, seguida por uma com 512. Por fim, uma camada de classificação com 12 neurônios e função *softmax* classifica a variedade do lúpulo contido na imagem. As camadas densas, com exceção da última, utilizam a função de ativação *ReLU*. Os modelos chamados Full Base (Fig. 4(a)) são treinados com a imagem/foto inteira do lúpulo, sem cortes, enquanto os modelos chamados de *Cropped Base* (Fig. 4(b)) trabalham com a folha recortada (*Cropped Dataset*). O intuito do último é ignorar o restante do cenário, como galhos e solo, e se concentrar apenas na folha, reduzindo interferência do ambiente. Além disso, as camadas totalmente conectadas foram incrementadas com operação de *dropouts* com *drop rates* de 0,2, 0,5 e 0,8. O objetivo dessa combinação é encontrar um bom modelo como ponto de partida para ser otimizado.

### C. Arquitetura Ponta-a-Ponta

Foram propostas cinco arquiteturas: Full (Fig. 5(a)), Main-Cropped (camada interna da Fig. 5(b)), Main-Cropped-Full (camada externa da Fig. 5(b)), Multi-Cropped (camada interna

TABELA II

TOTAL DE IMAGENS DE CADA VARIEDADE SEPARADO POR RESOLUÇÃO ESPECIFICADA EM ALTURA POR LARGURA.

Variedade	Resolução (Altura x Largura)								
	1040x520	1032x581	3391x2345	4096x2304	3683x3024	3024x4032	4032x3024	3072x4096	4096x3072
Cascade	51	9	0	0	1	3	46	0	0
Centennial	0	0	0	0	0	0	0	83	42
Cluster	30	0	0	23	0	0	50	0	0
Comet	55	10	0	0	0	3	47	12	44
Hallertau Mittelfrueh	0	0	0	0	0	0	0	67	74
Nugget	0	8	0	58	0	2	48	0	0
Saaz	0	7	1	0	0	1	48	39	76
Sorachi Ace	0	8	0	0	0	0	50	76	59
Tahoma	0	8	0	0	0	3	47	31	30
Triple Pearl	60	8	0	0	0	1	49	0	0
Triumph	0	9	0	42	0	5	45	0	0
Zeus	0	8	0	0	0	8	42	63	2
<b>Total</b>	<b>196</b>	<b>75</b>	<b>1</b>	<b>123</b>	<b>1</b>	<b>26</b>	<b>472</b>	<b>371</b>	<b>327</b>

TABELA III

NÚMERO DE FOLHAS DA BASE CROPPED DATASET.

Variedade	Treino	Validação	Teste	Total
Cascade	318	39	57	<b>414</b>
Centennial	158	23	21	<b>202</b>
Cluster	316	58	75	<b>449</b>
Comet	212	52	57	<b>321</b>
Hallertau Mittelfrueh	165	30	37	<b>232</b>
Nugget	301	74	67	<b>442</b>
Saaz	265	40	48	<b>353</b>
SorachiAce	286	43	61	<b>390</b>
Tahoma	203	32	39	<b>274</b>
TriplePearl	299	56	62	<b>417</b>
Triumph	260	56	65	<b>381</b>
Zeus	196	48	40	<b>284</b>
<b>Total</b>	<b>2979</b>	<b>551</b>	<b>629</b>	<b>4159</b>

TABELA IV

MÉDIA DE FOLHAS EXTRAÍDAS POR IMAGEM.

Variedade	Treino	Validação	Teste	Total
Cascade	4,08	2,44	3,56	<b>3,76</b>
Centennial	1,78	1,28	1,17	<b>1,62</b>
Cluster	4,33	3,87	5,00	<b>4,36</b>
Comet	1,75	2,08	2,28	<b>1,88</b>
Hallertau Mittelfrueh	1,67	1,43	1,76	<b>1,65</b>
Nugget	3,67	4,35	3,94	<b>3,81</b>
Saaz	2,17	1,60	1,92	<b>2,05</b>
SorachiAce	2,09	1,54	2,18	<b>2,02</b>
Tahoma	2,39	1,88	2,29	<b>2,30</b>
TriplePearl	3,56	3,29	3,65	<b>3,53</b>
Triumph	3,66	3,73	4,33	<b>3,77</b>
Zeus	2,25	2,67	2,22	<b>2,31</b>
<b>Total</b>	<b>2,64</b>	<b>2,38</b>	<b>2,71</b>	<b>2,61</b>

da Fig. 5(c)) e Multi-Cropped-Full (camada externa da Fig. 5(c)). Elas utilizam uma FasterRCNN [48] composta por uma ResNet50 como extrator de características para a detecção de folhas. O tamanho da imagem de entrada da rede de detecção é 640x640, enquanto das de classificação, é 300x300. Para isso, todas as imagens foram redimensionadas utilizando a interpolação INTER\_AREA da biblioteca OpenCV [49].

1) *Full*: Realiza a inferência a partir da imagem completa utilizando o modelo Full Base.

2) *Main-Cropped*: Recebe como entrada a folha principal recortada por uma rede treinada para a detecção de folhas e utiliza a Cropped Base para a classificação. Como a detecção

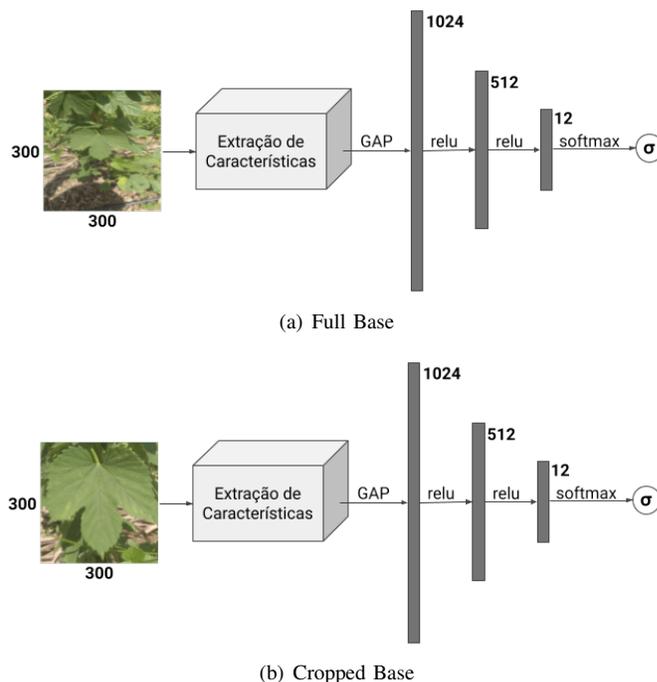


Fig. 4. Arquiteturas para os modelos base.

pode retornar várias folhas, o critério adotado para escolha da folha principal é a que possui a maior área dentre elas.

3) *Main-Cropped-Full*: Soma o resultado da Main Cropped a uma inferência realizada na imagem completa.

4) *Multi-Cropped*: Recebe como entrada todas as folhas detectadas e encaminha cada uma delas à Cropped Base. Ao final, soma-se as predições.

5) *Multi-Cropped-Full*: Combina a saída da Multi Cropped com o resultado da Full Base aplicada à imagem completa.

#### D. Aumento de Dados para o Treinamento

Com o propósito de melhorar o desempenho dos modelos base, foram aplicados aumentos de dados, tanto nas imagens inteiras, quanto nas folhas principais recortadas. Entre eles, estão o ajuste de brilho, onde as imagens foram clareadas (Fig. 6(b)) ou escurecidas (Fig. 6(c)) em 20% e aplicação da escala de

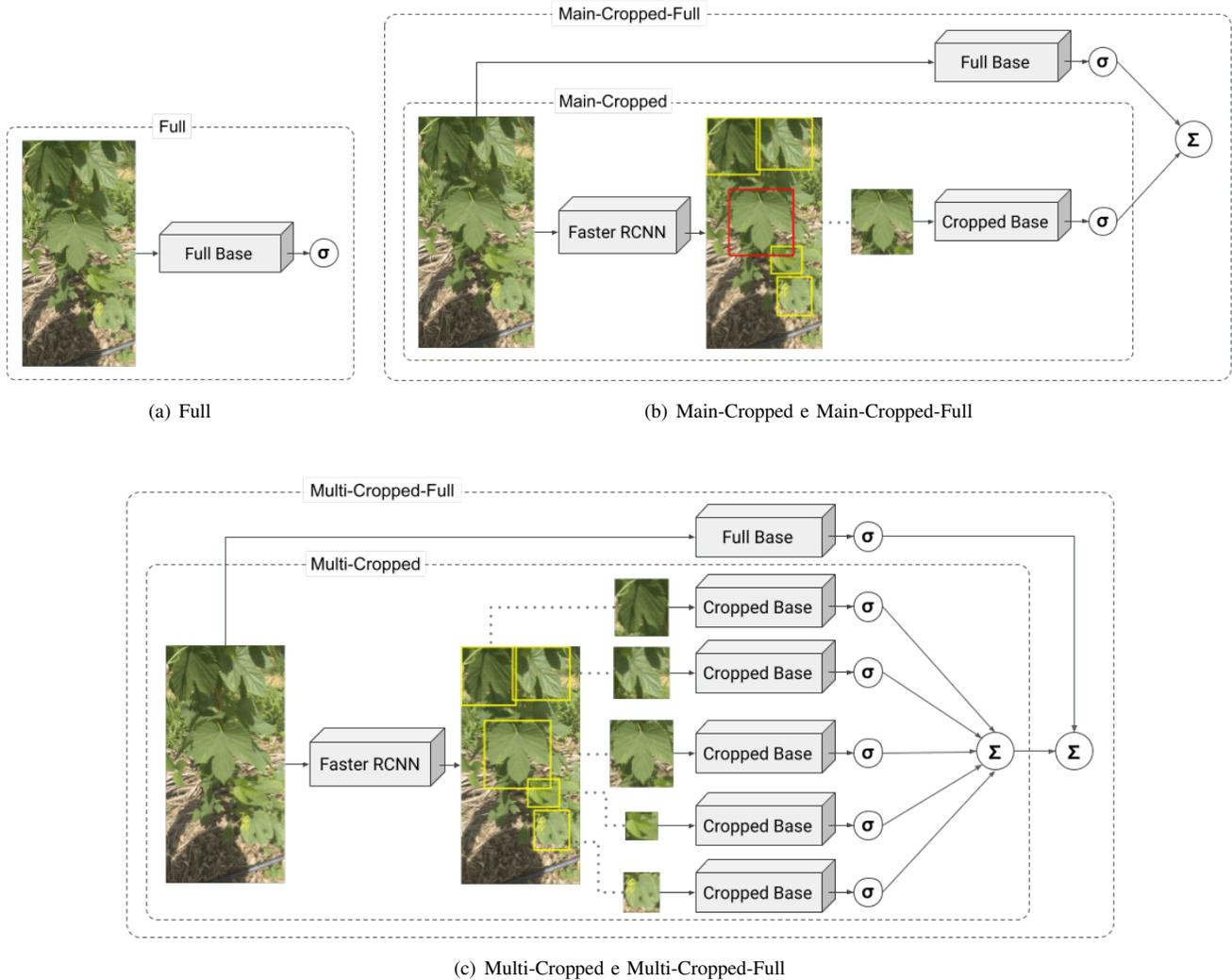


Fig. 5. Arquiteturas Ponta-a-Ponta Propostas.

cinza (Fig. 6(d)), ambos com objetivo de tornar a aprendizagem invariante em relação ao clima e cor das folhas. Também foi utilizado desfoque (Fig. 6(e)), com tamanhos do *kernel* em 5x5, 7x7, 9x9 e 11x11, para emular câmeras fotográficas com baixa resolução de imagem ou ausência de foco. Outra geração de dados foi a partir da oclusão, com o posicionamento aleatório de retângulos na imagem obstruindo parte da mesma. Nessa técnica, inspirada na regularização *cutout* [50], as áreas retangulares tiveram três preenchimentos: opaco (sem transparência) na Fig. 6(f), escuro (60 a 65% de transparência) na Fig. 6(g) e claro (75 a 85% de transparência) na Fig. 6(h). Além da oclusão, para diminuir área de visualização disponível da imagem, também foi utilizado o recorte (Fig. 6(i)) de 20 e 50% na altura e largura da imagem. Foram empregadas rotações com valores múltiplos de 22,5° nas imagens (Fig. 6(j)) e também rotações 3D (Fig. 6(k)), com 25 e 40° nos outros eixos. Nas rotações, os espaços gerados foram preenchidos com 0 em todos os *pixels*. Por fim, as folhas extras (não-principais) foram adotadas como aumento de dados para treinamento da *Cropped Base*. No entanto, como essas em sua maioria já possuem uma resolução inferior em relação à principal, não

foram aplicados aumentos que deteriorassem mais ainda a sua qualidade. As únicas transformações executadas nas folhas extras (de menor resolução) foram as rotações no eixo principal.

### E. Treinamento

A função de custo utilizada para otimizar os modelos de classificação foi a *categorical cross entropy* [51], [52] por ser indicada para problemas de reconhecimento multi-classes. A equação da mesma apresenta-se na Eq. (1), onde  $k$  é o número de classes,  $y$  o vetor com o rótulo verdadeiro e  $\hat{y}$  a saída do modelo com as previsões. A métrica para avaliar a performance foi a acurácia.

$$L(\theta) = - \sum_{i=1}^k y_i \log(\hat{y}_i) \quad (1)$$

Para o treinamento da rede de detecção, utilizou-se a função de custo da *FasterRCNN*, contida na Eq. (2). As variáveis  $t_i^*$  e  $p_i^*$  referem-se às coordenadas e rótulos verdadeiros da âncora  $i$ , respectivamente, enquanto  $p_i$  e  $t_i$  são os valores preditos.  $N_{cls}$  e  $N_{reg}$  correspondem aos termos de normalização e  $\lambda$  é

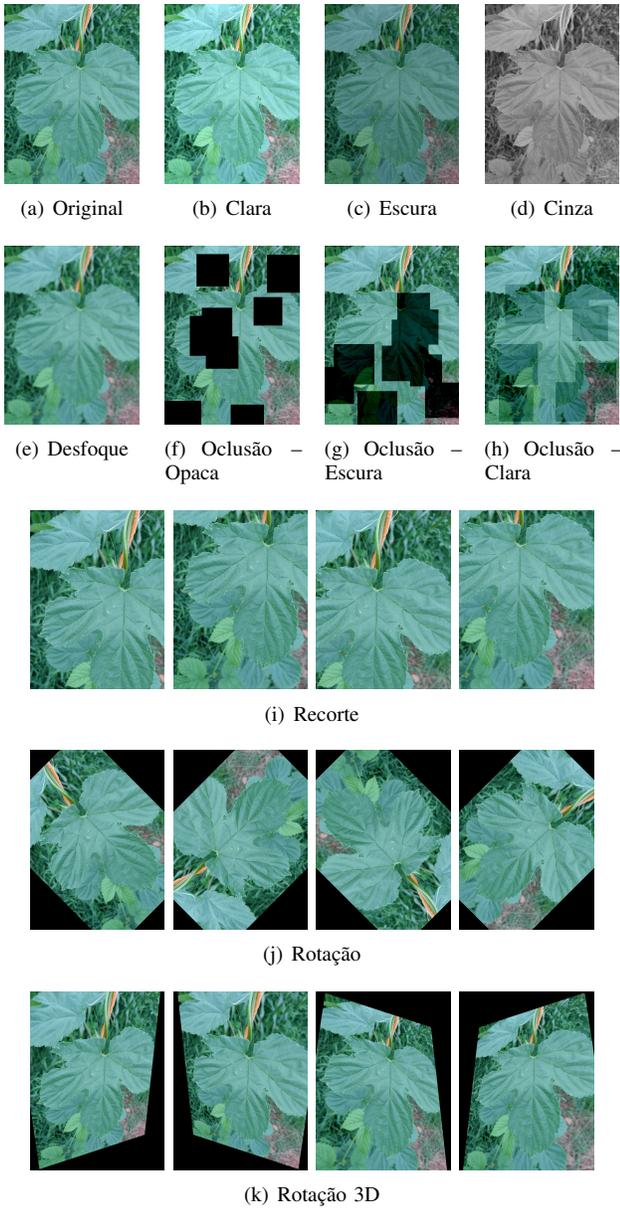


Fig. 6. Exemplos das técnicas de aumento de dados empregadas.

um parâmetro de balanceamento para os termos. Mais detalhes sobre a função podem ser encontrados no artigo original em [48].

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*) \quad (2)$$

#### IV. EXPERIMENTOS

Esta seção contém informações sobre os experimentos realizados para a escolha dos modelos base, os testes com as arquiteturas propostas e as contribuições individuais de cada aumento de dado aplicado. Além disso, são exibidos e discutidos os resultados.

Todos os modelos neste trabalho foram implementados utilizando a biblioteca TensorFlow e treinados em uma NVIDIA GeForce GTX TITAN X com 12GB de memória. Optou-se pelo otimizador AdaGrad [53] com um *learning rate* de  $1 \times 10^{-3}$ . Esta escolha é justificada pelo estudo apresentado em [54], que mostrou que esse algoritmo, quando aliado a essa taxa de aprendizagem e aplicado às redes InceptionV3 e ResNet50, apresenta bom desempenho na classificação de imagens em uma classe específica de problemas.

Os modelos base tiveram seus pesos inicializados com a técnica *Glorot Uniform* [55] nas camadas densas, enquanto os das camadas de extração de características tiveram seus pesos pré-treinados com a ImageNet. Não houve congelamento de nenhuma camada. As tabelas V e VI exibem a acurácia dos modelos Full Base e Cropped Base, respectivamente, combinando o modelo de extração de características com a taxa de *dropout*. O par que obteve o melhor desempenho, para ambas as bases, foi a ResNet50 com uma taxa de dropout de 0,5. Logo, estes foram os modelos selecionados para serem otimizados e compor as arquiteturas propostas. Para que as inferências destes modelos não se tornassem enviesadas ao conjunto de teste, os valores obtidos correspondem à predição realizada na base de validação do conjunto de treinamento. O tamanho do mini-lote (*mini-batch*) foi 32 para InceptionV3 e ResNet50 e 16 para EfficientNet B3.

TABELA V  
ACURÁCIA DO MODELO FULL BASE NOS DADOS DE VALIDAÇÃO.

Modelo	Taxa de Dropout	Acurácia	Épocas
InceptionV3	0,2	0,5776	152
InceptionV3	0,5	0,6293	92
InceptionV3	0,8	0,6595	170
ResNet50	0,2	0,7457	89
<b>ResNet50</b>	<b>0,5</b>	<b>0,8017</b>	<b>59</b>
ResNet50	0,8	0,7586	115
EfficientNet B3	0,2	0,7931	168
EfficientNet B3	0,5	0,7931	237
EfficientNet B3	0,8	0,7845	211

TABELA VI  
ACURÁCIA DO MODELO CROPPED BASE NOS DADOS DE VALIDAÇÃO.

Modelo	Taxa de Dropout	Acurácia	Épocas
InceptionV3	0,2	0,5733	40
InceptionV3	0,5	0,5733	21
InceptionV3	0,8	0,6897	143
ResNet50	0,2	0,7672	43
<b>ResNet50</b>	<b>0,5</b>	<b>0,7802</b>	<b>63</b>
ResNet50	0,8	0,7457	251
EfficientNet B3	0,2	0,7716	101
EfficientNet B3	0,5	0,7759	228
EfficientNet B3	0,8	0,7759	175

Após a determinação da rede para extração de característica (modelos base), aplicou-se os aumentos de dados descritos na seção III. Como critério de escolha de quais técnicas de aumento seriam selecionadas para o treinamento de cada base, os modelos foram treinados com cada uma individualmente, com o objetivo de verificar as que mais contribuem para a

classificação. Na base de validação, as mais apropriados para o Full Base foram Ajuste de Brilho, Oclusão – Opaca, Desfoque (kernel: 9x9), Recorte (50%), Rotação 3D e Rotação (Múltipla de 22,5°). Para o Cropped Base foram aplicadas Ajuste de Brilho, Oclusão – Clara, Desfoque (kernel: 9x9), Recorte (20%), Rotação 3D, Rotação (Múltipla de 22,5°) e Folhas Extras Rotacionadas (Múltipla de 22,5°). A tabela VII mostra os ganhos individuais dos aumentos de dados em cada um dos modelos base, já considerando a base de teste. Nota-se que os maiores ganhos absolutos estão relacionados à rotação das imagens, que aumentaram a acurácia em mais de 0,12 pontos percentuais tanto nas imagens inteiras quanto nas folhas recortadas. Outro que teve um bom desempenho foi a Rotação 3D. Por outro lado, o Ajuste de Brilho e a Escala de Cinza pioraram o desempenho dos modelos Cropped Base. Ainda na tabela VII, constam o ganho por imagem, que representa o ganho absoluto dividido pelo número de amostras geradas para aquele aumento. Pode-se observar que o Desfoque com *kernel* 11x11 teve uma boa contribuição para ambos os modelos, atentando-se ao número pequeno de novas imagens que foram produzidas. O bom desempenho pode estar relacionado ao fato de a base de dados conter muitas fotos desfocadas, o que ajudou os modelos as reconhecerem melhor nessas condições.

Para completar a construção das arquiteturas, foi necessário treinar a rede de detecção FasterRCNN. Com a finalidade de simplificar o treinamento, a rede apenas detecta folhas, sem classificá-las. O modelo foi inicializado com os pesos treinados no *dataset* COCO [56]. O limiar de Intersection Over Union (IoU) [57], [58] foi de 0,7 no primeiro estágio e 0,6 no segundo. A IoU é uma métrica utilizada para definir se duas predições de *bounding boxes* pertencem ao mesmo objeto e é fornecida a partir da divisão da interseção das áreas pela sua união. Na FasterRCNN, ela foi utilizada para a etapa de Non-Maximum Suppression (NMS) [59], [60], ou seja, remoção de todas as predições pertencentes ao mesmo objeto que não possuem o maior *score*. Embora não tenha sido utilizado um valor mínimo de *score* para o treinamento, durante a inferência são removidas todas as predições com pontuação abaixo de 0,5. Além disso, o número de folhas detectadas por imagem é limitada a 10, tomando as com melhor pontuação. Visando aprimorar o desempenho, foram utilizados os seguintes aumentos de dados combinados: ajuste de brilho, rotações múltiplas de 90° e *flip* horizontal e vertical. Quando nenhuma folha é detectada, as arquiteturas que utilizam a folha principal (Main-Cropped e Main-Cropped-Full) são instruídas a utilizar a imagem inteira como se fosse uma detecção. O tamanho do *batch* foi de 2. A avaliação do modelo por classe na base de teste pode ser conferida na tabela VIII, onde os verdadeiro-positivos (VP) correspondem às folhas rotuladas que ele detectou, os falso-negativos (FN) se referem às não detectadas e os falso-positivos (FP) são as detecções incorretas. Para determinar se uma caixa delimitadora detectada corresponde a um rótulo foi utilizada uma IoU de 0,6. Pode-se observar que a Triumph foi a classe que obteve maior precisão. No entanto, ela e a Comet tiveram as menores revocações. Essa baixa sensibilidade pode ser explicada pela pouca quantidade de folhas por imagem destas variedades na base de treino em relação à de teste.

A tabela IX contém a acurácia final das arquiteturas propostas com e sem aumento de dados. Em ambas, a que obteve o melhor desempenho foi a Multi-Cropped-Full, alcançando 95,69% na base de teste. Os resultados mostram que tanto a imagem inteira quanto a recortada contribuem para a classificação final, uma vez que os modelos que somam as saídas de ambas foram superiores às predições individuais de cada uma. Além disso, a utilização de folhas extras, mesmo que não tenham a qualidade da folha principal, mostrou-se eficaz. Os ganhos com a utilização da imagem inteira e várias folhas recortadas podem estar relacionados ao fato de fotos mais próximas da planta produzirem imagens com melhor resolução e nitidez, porém em número menor, enquanto fotos mais distantes, possuem qualidade inferior, mas são disponíveis em maior número. Essa tabela também disponibiliza o tempo de inferência de cada arquitetura em segundos. Pode-se observar que a etapa de detecção é a que gasta mais tempo durante a predição. Ainda, como esperado, os modelos que utilizam várias folhas consomem ligeiramente mais tempo em relação aos que inferem apenas com a folha principal.

As tabelas X, XI, XII, XIII e XIV detalham as métricas de precisão, revocação e F1-score por classe das 5 arquiteturas. Nas arquiteturas mais simples, Full e Main-Cropped, a precisão da classificação da variedade Nugget mostrou-se baixa, tanto sem e com aumento de dados. Já com as arquiteturas Main-Cropped-Full e Multi-Cropped-Full, que mesclam várias predições, foi possível chegar a uma boa precisão com o aumento de dados, o que reforça a robustez dessa combinação. A variedade Cluster, que teve um desempenho ruim na métrica de sensibilidade em todos os modelos, teve um ganho considerável quando aplicado o aumento de dados, o que demonstra a importância desta técnica no problema abordado.

Nas Figs. 7 e 8, podem ser conferidas as matrizes de confusão da Multi-Cropped-Full sem e com aumento de dados, respectivamente. Pode-se observar que a rede teve mais dificuldades em reconhecer as classes Cluster, Triumph e Nugget, quando não aplicado o aumento de dados. Já para a Cascade, Centennial e Zeus, que já tinham um bom resultado sem aumento de dados, não houve melhora considerável. O modelo conseguiu aprender bem as características dessas variedades sem precisar de mais amostras. Com o aumento de dados, os resultados foram satisfatórios para todas as classes. No entanto, a classificação da Triumph ainda foi inferior às demais. Pode ser que o aumento da sensibilidade na detecção das folhas gere mais amostras e melhore o desempenho nessa variedade.

TABELA VII  
GANHO NA ACURÁCIA POR AUMENTO DE DADOS APLICADO AOS MODELOS FULL BASE E CROPPED BASE.

Com Aumento	Full Base			Cropped Base		
	Acurácia	Ganho	Ganho / Imagem	Acurácia	Ganho	Ganho / Imagem
Base (Sem Aumento)	0,7802	-	-	0,7802	-	-
Ajuste de Brilho	0,8147	0,0345	0,0173	0,7716	-0,0086	-0,0043
Desfoque (kernel: 5x5)	0,7888	0,0086	0,0086	0,8060	0,0258	0,0258
Desfoque (kernel: 7x7)	0,8147	0,0345	0,0345	0,7974	0,0172	0,0172
Desfoque (kernel: 9x9)	0,8147	0,0345	0,0345	0,8060	0,0258	0,0258
Desfoque (kernel: 11x11)	0,8233	0,0431	<b>0,0431</b>	0,8190	0,0388	<b>0,0388</b>
Escala de Cinza	0,8103	0,0301	0,0301	0,7543	-0,0259	-0,0259
Oclusão – Clara	0,8103	0,0301	0,0301	0,7931	0,0129	0,0129
Oclusão – Escura	0,8103	0,0301	0,0301	0,7974	0,0172	0,0172
Oclusão – Opaca	0,8060	0,0258	0,0258	0,7845	0,0043	0,0043
Recorte (20%)	0,8534	0,0732	0,0183	0,7888	0,0086	0,0021
Recorte (50%)	0,8578	0,0776	0,0194	0,8060	0,0258	0,0065
Rotação (Múltipla de 22,5°)	<b>0,9095</b>	<b>0,1293</b>	0,0086	0,8922	0,1120	0,0075
Rotação (Múltipla de 45°)	0,8793	0,0991	0,0142	0,8707	0,0905	0,0129
Rotação 3D	0,8836	0,1034	0,0129	0,8319	0,0517	0,0065
Folhas Extras	-	-	-	0,8147	0,0345	0,0210
Folhas Extras Rotacionadas (22,5°)	-	-	-	<b>0,9224</b>	<b>0,1422</b>	0,0036
Folhas Extras Rotacionadas (45°)	-	-	-	0,9095	0,1293	0,0070

TABELA VIII  
AVALIAÇÃO DO MODELO FASTERRCNN.

Variedade	VP	FN	FP	Precisão	Revocação	F1-Score
Cascade	47	10	7	0,8704	0,8246	0,8468
Centennial	19	2	3	0,8636	0,9048	0,8837
Cluster	61	14	8	0,8841	0,8133	0,8472
Comet	44	13	8	0,8462	0,7719	0,8073
Hallertau Mittelfrueh	29	8	9	0,7632	0,7838	0,7733
Nugget	57	10	7	0,8906	0,8507	0,8702
Saaz	39	9	6	0,8667	0,8125	0,8387
Sorachi Ace	55	6	6	0,9016	0,9016	0,9016
Tahoma	31	8	8	0,7949	0,7949	0,7949
Triple Pearl	48	14	8	0,8571	0,7742	0,8136
Triumph	45	20	2	0,9574	0,6923	0,8036
Zeus	38	2	3	0,9268	0,9500	0,9383
<b>Total</b>	<b>513</b>	<b>116</b>	<b>75</b>	<b>0,8724</b>	<b>0,8156</b>	<b>0,8431</b>

TABELA IX  
ACURÁCIA E TEMPO DE INFERÊNCIA DAS ARQUITETURAS PROPOSTAS.

Arquitetura	Sem Aumento de Dados			Com Aumento de Dados		
	Acurácia	Tempo de inferência sem detecção (s)	Tempo de inferência com detecção (s)	Acurácia	Tempo de inferência sem detecção (s)	Tempo de inferência com detecção (s)
Full	0,7802	0,0214	0,0214	0,9052	0,0218	0,0218
Main-Cropped	0,7802	0,0107	0,5707	0,9181	0,0108	0,5708
Main-Cropped-Full	0,8017	0,0322	0,5922	0,9440	0,0326	0,5926
Multi-Cropped	0,7672	0,0891	0,6491	0,9267	0,0900	0,6500
<b>Multi-Cropped-Full</b>	<b>0,8190</b>	<b>0,1117</b>	<b>0,6717</b>	<b>0,9569</b>	<b>0,1127</b>	<b>0,6727</b>

## V. CONCLUSÃO

Apesar de a classificação de variedades de lúpulo não ter se mostrado um problema de fácil resolução, foi possível elaborar uma arquitetura com uma boa precisão a partir da combinação da classificação da imagem inteira com a de suas folhas recortadas. Assim, torna-se viável a construção de uma ferramenta para identificar a variedade de maneira fácil. Sugere-se aplicar este protocolo a outras classificações de plantas que possuem muitas folhas para se investigar a robustez.

Ainda, pôde-se avaliar os impactos de cada aumento de dados neste tipo de classificação, onde alguns foram essenciais, como as rotações, e outros até prejudicaram o desempenho,

como o ajuste de brilho e a aplicação da escala de cinza. As contribuições também puderam ser avaliadas em relação ao total de imagens geradas, o que pode ser relevante em um cenário com limitação de *hardware* para o treinamento.

Como trabalhos futuros, seria interessante experimentar rotações com variações de ângulos menores que 22,5° e similar combinações de vários aumentos de dados (ex.: Rotação + Desfoque). Adicionalmente, pode ser relevante utilizar bases de extração de características diferentes para a Full Base e Cropped Base e testar camadas densas com outras configurações, com o objetivo de aprenderem características diferentes. Outra possibilidade é empregar folhas segmentadas em vez de

TABELA X  
 AVALIAÇÃO POR CLASSE DA ARQUITETURA FULL.

Variedade	Sem Aumento de Dados			Com Aumento de Dados		
	Precisão	Revocação	F1-Score	Precisão	Revocação	F1-Score
Cascade	0,9333	0,8750	0,9032	0,9375	0,9375	0,9375
Centennial	0,7619	0,8889	0,8205	1,0000	0,9444	0,9714
Cluster	0,7143	0,6667	0,6897	0,8235	0,9333	0,8750
Comet	0,6452	0,8000	0,7143	0,9200	0,9200	0,9200
Hallertau Mittelfrueh	0,9500	0,9048	0,9268	0,9545	1,0000	0,9767
Nugget	0,7143	0,5882	0,6452	0,6667	0,5882	0,6250
Saaz	0,9500	0,7600	0,8444	0,9231	0,9600	0,9412
Sorachi Ace	0,8065	0,8929	0,8475	0,9655	1,0000	0,9825
Tahoma	0,6471	0,6471	0,6471	0,9286	0,7647	0,8387
Triple Pearl	0,7647	0,7647	0,7647	0,9412	0,9412	0,9412
Triumph	0,5833	0,4667	0,5185	0,8571	0,8000	0,8276
Zeus	0,8500	0,9444	0,8947	0,8500	0,9444	0,8947
<b>Média</b>	<b>0,7767</b>	<b>0,7666</b>	<b>0,7680</b>	<b>0,8973</b>	<b>0,8945</b>	<b>0,8943</b>
<b>Média Ponderada</b>	<b>0,7843</b>	<b>0,7802</b>	<b>0,7783</b>	<b>0,9042</b>	<b>0,9052</b>	<b>0,9033</b>

TABELA XI  
 AVALIAÇÃO POR CLASSE DA ARQUITETURA MAIN-CROPPED.

Variedade	Sem Aumento de Dados			Com Aumento de Dados		
	Precisão	Revocação	F1-Score	Precisão	Revocação	F1-Score
Cascade	0,8235	0,8750	0,8485	1,0000	1,0000	1,0000
Centennial	1,0000	0,9444	0,9714	0,9412	0,8889	0,9143
Cluster	0,5455	0,4000	0,4615	0,9286	0,8667	0,8966
Comet	0,7692	0,8000	0,7843	0,9259	1,0000	0,9615
Hallertau Mittelfrueh	0,9091	0,9524	0,9302	0,9524	0,9524	0,9524
Nugget	0,7778	0,4118	0,5385	0,7333	0,6471	0,6875
Saaz	0,8148	0,8800	0,8462	0,9231	0,9600	0,9412
Sorachi Ace	0,8214	0,8214	0,8214	0,9655	1,0000	0,9825
Tahoma	0,6000	0,7059	0,6486	0,8235	0,8235	0,8235
Triple Pearl	0,7000	0,8235	0,7568	0,9444	1,0000	0,9714
Triumph	0,6111	0,7333	0,6667	0,8571	0,8000	0,8276
Zeus	0,8824	0,8333	0,8571	0,9444	0,9444	0,9444
<b>Média</b>	<b>0,7712</b>	<b>0,7651</b>	<b>0,7609</b>	<b>0,9116</b>	<b>0,9069</b>	<b>0,9086</b>
<b>Média Ponderada</b>	<b>0,7820</b>	<b>0,7802</b>	<b>0,7748</b>	<b>0,9160</b>	<b>0,9181</b>	<b>0,9164</b>

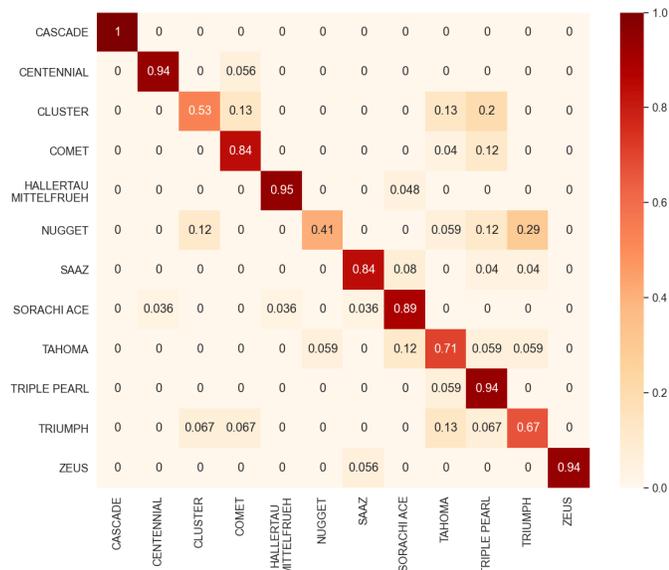


Fig. 7. Matriz de Confusão da arquitetura Multi-Cropped-Full sem aumento de dados.

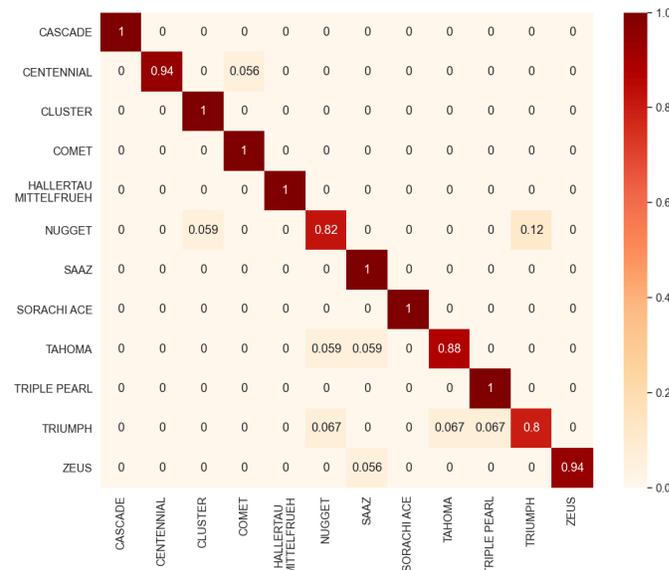


Fig. 8. Matriz de Confusão da arquitetura Multi-Cropped-Full com aumento de dados.

recortadas para que a rede consiga focar ainda mais na folha e ignorar o restante do cenário.

Por fim, ampliar a base de dados para mais classes pode ser

desafiador, uma vez que aumenta-se o número de variedades que têm características visuais em comum.

TABELA XII  
 AVALIAÇÃO POR CLASSE DA ARQUITETURA MAIN-CROPPED-FULL.

Variedade	Sem Aumento de Dados			Com Aumento de Dados		
	Precisão	Revocação	F1-Score	Precisão	Revocação	F1-Score
Cascade	0,8824	0,9375	0,9091	0,9412	1,0000	0,9697
Centennial	0,8947	0,9444	0,9189	1,0000	0,9444	0,9714
Cluster	0,7000	0,4667	0,5600	0,9375	1,0000	0,9677
Comet	0,8077	0,8400	0,8235	0,9600	0,9600	0,9600
Hallertau Mittelfrueh	0,9091	0,9524	0,9302	1,0000	1,0000	1,0000
Nugget	0,5833	0,4118	0,4828	0,9231	0,7059	0,8000
Saaz	0,9130	0,8400	0,8750	0,9259	1,0000	0,9615
Sorachi Ace	0,8276	0,8571	0,8421	1,0000	1,0000	1,0000
Tahoma	0,6316	0,7059	0,6667	0,8333	0,8824	0,8571
Triple Pearl	0,7500	0,8824	0,8108	0,9444	1,0000	0,9714
Triumph	0,5882	0,6667	0,6250	0,9231	0,8000	0,8571
Zeus	0,9444	0,9444	0,9444	0,8947	0,9444	0,9189
<b>Média</b>	<b>0,7860</b>	<b>0,7874</b>	<b>0,7824</b>	<b>0,9403</b>	<b>0,9364</b>	<b>0,9363</b>
<b>Média Ponderada</b>	<b>0,7984</b>	<b>0,8017</b>	<b>0,7964</b>	<b>0,9445</b>	<b>0,9440</b>	<b>0,9424</b>

TABELA XIII  
 AVALIAÇÃO POR CLASSE DA ARQUITETURA MULTI-CROPPED.

Variedade	Sem Aumento de Dados			Com Aumento de Dados		
	Precisão	Revocação	F1-Score	Precisão	Revocação	F1-Score
Cascade	0,9375	0,9375	0,9375	1,0000	1,0000	1,0000
Centennial	1,0000	0,8889	0,9412	1,0000	0,8333	0,9091
Cluster	0,5455	0,4000	0,4615	0,8667	0,8667	0,8667
Comet	0,8696	0,8000	0,8333	0,8929	1,0000	0,9434
Hallertau Mittelfrueh	0,9091	0,9524	0,9302	1,0000	1,0000	1,0000
Nugget	0,6250	0,2941	0,4000	0,7778	0,8235	0,8000
Saaz	0,8400	0,8400	0,8400	0,9600	0,9600	0,9600
Sorachi Ace	0,8000	0,8571	0,8276	0,9655	1,0000	0,9825
Tahoma	0,5263	0,5882	0,5556	0,9333	0,8235	0,8750
Triple Pearl	0,5333	0,9412	0,6809	0,8889	0,9412	0,9143
Triumph	0,5882	0,6667	0,6250	0,8571	0,8000	0,8276
Zeus	1,0000	0,8333	0,9091	0,9444	0,9444	0,9444
<b>Média</b>	<b>0,7645</b>	<b>0,7500</b>	<b>0,7452</b>	<b>0,9239</b>	<b>0,9161</b>	<b>0,9186</b>
<b>Média Ponderada</b>	<b>0,7796</b>	<b>0,7672</b>	<b>0,7628</b>	<b>0,9285</b>	<b>0,9267</b>	<b>0,9262</b>

TABELA XIV  
 AVALIAÇÃO POR CLASSE DA ARQUITETURA MULTI-CROPPED-FULL.

Variedade	Sem Aumento de Dados			Com Aumento de Dados		
	Precisão	Revocação	F1-Score	Precisão	Revocação	F1-Score
Cascade	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
Centennial	0,9444	0,9444	0,9444	1,0000	0,9444	0,9714
Cluster	0,7273	0,5333	0,6154	0,9375	1,0000	0,9677
Comet	0,8400	0,8400	0,8400	0,9615	1,0000	0,9804
Hallertau Mittelfrueh	0,9524	0,9524	0,9524	1,0000	1,0000	1,0000
Nugget	0,8750	0,4118	0,5600	0,8750	0,8235	0,8485
Saaz	0,9130	0,8400	0,8750	0,9259	1,0000	0,9615
Sorachi Ace	0,8333	0,8929	0,8621	1,0000	1,0000	1,0000
Tahoma	0,6316	0,7059	0,6667	0,9375	0,8824	0,9091
Triple Pearl	0,5926	0,9412	0,7273	0,9444	1,0000	0,9714
Triumph	0,5882	0,6667	0,6250	0,8571	0,8000	0,8276
Zeus	1,0000	0,9444	0,9714	1,0000	0,9444	0,9714
<b>Média</b>	<b>0,8248</b>	<b>0,8061</b>	<b>0,8033</b>	<b>0,9533</b>	<b>0,9496</b>	<b>0,9508</b>
<b>Média Ponderada</b>	<b>0,8344</b>	<b>0,8190</b>	<b>0,8160</b>	<b>0,9568</b>	<b>0,9569</b>	<b>0,9562</b>

#### AGRADECIMENTOS

Os autores agradecem à Atlântica Hops, Hops Brasil e Bra-zuca Lúpulos pela cessão das imagens, à Lúpulo Zona da Mata pela contribuição técnica e à UFOP pelo suporte financeiro. Os autores também agradecem à NVIDIA Corporation, pela doação da NVIDIA Titan X pascal usada nos experimentos deste trabalho.

#### REFERÊNCIAS

- [1] G. Astray, P. Gullón, B. Gullón, P. E. Mune-kata, and J. M. Lorenzo, "Humulus lupulus l. as a natural source of functional biomolecules," *Applied Sciences*, vol. 10, no. 15, p. 5074, 2020.
- [2] C. Almaguer, C. Schönberger, M. Gastl, E. K. Arendt, and T. Becker, "Humulus lupulus—a story that begs to be told. a review," *Journal of the Institute of Brewing*, vol. 120, no. 4, pp. 289–314, 2014.
- [3] B. Kramer, J. Thielmann, A. Hickisch, P. Muranyi, J. Wunderlich, and C. Hauser, "Antimicrobial activity of hop extracts against foodborne pathogens for meat applications," *Journal of applied microbiology*, vol. 118, no. 3, pp. 648–657, 2015.

- [4] T. Nuutinen, "Medicinal properties of terpenes found in cannabis sativa and humulus lupulus," *European journal of medicinal chemistry*, vol. 157, pp. 198–228, 2018.
- [5] L. Chadwick, G. Pauli, and N. Farnsworth, "The pharmacognosy of humulus lupulus l.(hops) with an emphasis on estrogenic properties," *Phytomedicine*, vol. 13, no. 1-2, pp. 119–131, 2006.
- [6] P. Zanoli and M. Zavatti, "Pharmacognostic and pharmacological profile of humulus lupulus l." *Journal of ethnopharmacology*, vol. 116, no. 3, pp. 383–396, 2008.
- [7] M. Biendl and C. Pinzl, "Hops and health," *MBAA TQ*, vol. 46, pp. 1–7, 2009.
- [8] M. Van Cleemput, K. Cattoor, K. De Bosscher, G. Haegeman, D. De Keuleleire, and A. Heyerick, "Hop (humulus lupulus)-derived bitter acids as multipotent bioactive compounds," *Journal of natural products*, vol. 72, no. 6, pp. 1220–1230, 2009.
- [9] J. Healey, *The Hops List: 265 Beer Hop Varieties From Around the World*. Julian Healey, 2016.
- [10] B. Steenackers, L. De Cooman, and D. De Vos, "Chemical transformations of characteristic hop secondary metabolites in relation to beer properties and the brewing process: a review," *Food Chemistry*, vol. 172, pp. 742–756, 2015.
- [11] A. da Rosa Almeida, M. V. d. O. B. Maciel, B. Cardoso Gasparini Gandolpho, M. H. Machado, G. L. Teixeira, F. C. Bertoldi, C. M. Noronha, L. Vitali, J. M. Block, and P. L. M. Barreto, "Brazilian grown cascade hop (humulus lupulus l.): Lc-esi-ms-ms and gc-ms analysis of chemical composition and antioxidant activity of extracts and essential oils," *Journal of the American Society of Brewing Chemists*, pp. 1–11, 2020.
- [12] A. d. R. Almeida, M. V. d. O. B. Maciel, M. H. Machado, G. C. Bazzo, R. D. de Armas, V. B. Vitorino, L. Vitali, J. M. Block, and P. L. M. Barreto, "Bioactive compounds and antioxidant activities of brazilian hop (humulus lupulus l) extracts," *International Journal of Food Science & Technology*, vol. 55, no. 1, pp. 340–347, 2020.
- [13] T. R. Arruda, P. F. Pinheiro, P. I. Silva, and P. C. Bernardes, "A new perspective of a well-recognized raw material: Phenolic content, antioxidant and antimicrobial activities and  $\alpha$ - and  $\beta$ -acids profile of brazilian hop (humulus lupulus l.) extracts," *LWT*, vol. 141, p. 110905, 2021.
- [14] L. De Cooman, E. Everaert, and D. De Keuleleire, "Quantitative analysis of hop acids, essential oils and flavonoids as a clue to the identification of hop varieties," *Phytochemical Analysis: An International Journal of Plant Chemical and Biochemical Techniques*, vol. 9, no. 3, pp. 145–150, 1998.
- [15] M. Kovačević and M. Kač, "Determination and verification of hop varieties by analysis of essential oils," *Food Chemistry*, vol. 77, no. 4, pp. 489–494, 2002.
- [16] R. A. Shellie, S. D. Poynter, J. Li, J. L. Gathercole, S. P. Whittock, and A. Koutoulis, "Varietal characterization of hop (humulus lupulus l.) by gc–ms analysis of hop cone extracts," *Journal of separation science*, vol. 32, no. 21, pp. 3720–3725, 2009.
- [17] M. A. Farag, E. A. Mahrous, T. Lübken, A. Porzel, and L. Wessjohann, "Classification of commercial cultivars of humulus lupulus l.(hop) by chemometric pixel analysis of two dimensional nuclear magnetic resonance spectra," *Metabolomics*, vol. 10, no. 1, pp. 21–32, 2014.
- [18] L. M. Duarte, L. H. C. Adriano, and M. A. L. de Oliveira, "Capillary electrophoresis in association with chemometrics approach for bitterness hop (humulus lupulus l.) classification," *Electrophoresis*, vol. 39, no. 11, pp. 1399–1409, 2018.
- [19] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Computers and electronics in agriculture*, vol. 147, pp. 70–90, 2018.
- [20] M. A. Jenks, "Plant nomenclature," *Purdue University-Department of Horticulture and Landscape Architecture. Disponivel*, 2011.
- [21] G. W. Horgan, M. Talbot, and J. Davey, "Towards automatic recognition of plant varieties," in *Challenge of Image Retrieval*, 1998, pp. 1–8.
- [22] A. F. A. Nasir, M. N. A. Rahman, N. Mat, and A. R. Mamat, "Automatic identification of ficus deltoidea jack (moraceae) varieties based on leaf," *Modern Applied Science*, vol. 8, no. 5, p. 121, 2014.
- [23] N. Sünderhauf, C. McCool, B. Upcroft, and T. Perez, "Fine-grained plant classification using convolutional neural networks for feature extraction." in *CLEF (Working Notes)*, 2014, pp. 756–762.
- [24] S. H. Lee, C. S. Chan, P. Wilkin, and P. Remagnino, "Deep-plant: Plant identification with convolutional neural networks," in *2015 IEEE international conference on image processing (ICIP)*. IEEE, 2015, pp. 452–456.
- [25] B. P. Tóth, M. J. Tóth, D. Papp, and G. Szücs, "Deep learning and svm classification for plant recognition in content-based large scale image retrieval." in *CLEF (Working Notes)*, 2016, pp. 569–578.
- [26] H. Yalcin and S. Razavi, "Plant classification using convolutional neural networks," in *2016 Fifth International Conference on Agro-Geoinformatics (Agro-Geoinformatics)*. IEEE, 2016, pp. 1–5.
- [27] S. H. Lee, Y. L. Chang, C. S. Chan, and P. Remagnino, "Plant identification system based on a convolutional neural network for the lifecycle 2016 plant classification task." *CLEF (Working Notes)*, vol. 1, pp. 502–510, 2016.
- [28] Y. Sun, Y. Liu, G. Wang, and H. Zhang, "Deep learning for plant identification in natural environment," *Computational intelligence and neuroscience*, vol. 2017, 2017.
- [29] W.-S. Jeon and S.-Y. Rhee, "Plant leaf recognition using a convolution neural network," *International Journal of Fuzzy Logic and Intelligent Systems*, vol. 17, no. 1, pp. 26–34, 2017.
- [30] J. Hu, Z. Chen, M. Yang, R. Zhang, and Y. Cui, "A multiscale fusion convolutional neural network for plant leaf recognition," *IEEE Signal Processing Letters*, vol. 25, no. 6, pp. 853–857, 2018.
- [31] X. Zhu, M. Zhu, and H. Ren, "Method of plant leaf recognition based on improved deep convolutional neural network," *Cognitive Systems Research*, vol. 52, pp. 223–233, 2018.
- [32] B. Wang and D. Wang, "Plant leaves classification: A few-shot learning method based on siamese network," *IEEE Access*, vol. 7, pp. 151 754–151 763, 2019.
- [33] Y. Zhu, W. Sun, X. Cao, C. Wang, D. Wu, Y. Yang, and N. Ye, "Ta-cnn: Two-way attention models in deep convolutional neural network for plant recognition," *Neurocomputing*, vol. 365, pp. 191–200, 2019.
- [34] D. I. Mercurio and A. A. Hernandez, "Classification of sweet potato variety using convolutional neural network," in *2019 IEEE 9th International Conference on System Engineering and Technology (ICSET)*. IEEE, 2019, pp. 120–125.
- [35] K. Yang, W. Zhong, and F. Li, "Leaf segmentation and classification with a complicated background using deep learning," *Agronomy*, vol. 10, no. 11, p. 1721, 2020.
- [36] S. Javanmardi, S.-H. M. Ashtiani, F. J. Verbeek, and A. Martynenko, "Computer-vision classification of corn seed varieties using deep convolutional neural network," *Journal of Stored Products Research*, vol. 92, p. 101800, 2021.
- [37] S. Lingwal, K. K. Bhatia, and M. S. Tomer, "Image-based wheat grain classification using convolutional neural network," *Multimedia Tools and Applications*, pp. 1–25, 2021.
- [38] H. Tavakoli, P. Alirezazadeh, A. Hedayatipour, A. B. Nasib, and N. Landwehr, "Leaf image-based classification of some common bean cultivars using discriminative convolutional neural networks," *Computers and Electronics in Agriculture*, vol. 181, p. 105935, 2021.
- [39] P. Castro, "Ufop hop varieties dataset (ufop-hvd)," 2021. [Online]. Available: <https://doi.org/10.6084/m9.figshare.14933178>
- [40] M. Everingham and J. Winn, "The pascal visual object classes challenge 2012 (voc2012) development kit," *Pattern Analysis, Statistical Modelling and Computational Learning, Tech. Rep.*, vol. 8, p. 5, 2011.
- [41] T. Wiatowski and H. Bölskei, "A mathematical theory of deep convolutional neural networks for feature extraction," *IEEE Transactions on Information Theory*, vol. 64, no. 3, pp. 1845–1866, 2017.
- [42] Y. H. Liu, "Feature extraction and image recognition with convolutional neural networks," in *Journal of Physics: Conference Series*, vol. 1087, no. 6. IOP Publishing, 2018, p. 062032.
- [43] I. Namatēvs, "Deep convolutional neural networks: Structure, feature extraction and training," *Information Technology & Management Science (Sciendo)*, vol. 20, no. 1, 2017.
- [44] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.
- [45] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [46] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International Conference on Machine Learning*. PMLR, 2019, pp. 6105–6114.
- [47] M. Lin, Q. Chen, and S. Yan, "Network in network," *arXiv preprint arXiv:1312.4400*, 2013.
- [48] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: towards real-time object detection with region proposal networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 6, pp. 1137–1149, 2016.
- [49] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.
- [50] T. DeVries and G. W. Taylor, "Improved regularization of convolutional neural networks with cutout," *arXiv preprint arXiv:1708.04552*, 2017.

- [51] S. Bhatnagar, D. Ghosal, and M. H. Kolekar, "Classification of fashion article images using convolutional neural networks," in *2017 Fourth International Conference on Image Information Processing (ICIIP)*. IEEE, 2017, pp. 1–6.
- [52] Y. Ho and S. Wookey, "The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling," *IEEE Access*, vol. 8, pp. 4806–4813, 2019.
- [53] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," *Journal of machine learning research*, vol. 12, no. 7, 2011.
- [54] I. Kandel, M. Castelli, and A. Popovič, "Comparative study of first order optimizers for image classification using convolutional neural networks on histopathology images," *Journal of Imaging*, vol. 6, no. 9, p. 92, 2020.
- [55] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the thirteenth international conference on artificial intelligence and statistics. JMLR Workshop and Conference Proceedings*, 2010, pp. 249–256.
- [56] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *European conference on computer vision*. Springer, 2014, pp. 740–755.
- [57] M. A. Rahman and Y. Wang, "Optimizing intersection-over-union in deep neural networks for image segmentation," in *International symposium on visual computing*. Springer, 2016, pp. 234–244.
- [58] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 658–666.
- [59] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, "Distance-iou loss: Faster and better learning for bounding box regression," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 07, 2020, pp. 12 993–13 000.
- [60] J. Hosang, R. Benenson, and B. Schiele, "Learning non-maximum suppression," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4507–4515.



**Pedro Castro** é bacharel e mestre em Ciência da Computação pela Universidade Federal de Ouro Preto (UFOP), em 2008 e 2012, respectivamente. Atualmente, é Técnico Administrativo no Núcleo de Tecnologia da Informação (NTI) da UFOP. Seus interesses de pesquisa incluem meta-heurística, sistemas embarcados, aprendizagem de máquinas e aprendizagem profunda.



**Gladston Moreira** é mestre em matemática pela Universidade Federal de Minas Gerais em 2003, e doutor em engenharia elétrica em 2011. Atualmente é Professor Associado do Departamento de Computação da Universidade Federal de Ouro Preto, e professor permanente do Programa de Pós-Graduação em Ciência da Computação. Seus interesses de pesquisa incluem otimização multiobjetivo, reconhecimento de padrões e estatística espacial.



**Eduardo Luz** é bacharel em Engenharia Elétrica pela Universidade Federal de Minas Gerais (2005), e doutor em Ciência da Computação pela Universidade Federal de Ouro Preto (2019). Atua como Professor Adjunto no Departamento de Computação (DECOM) da Universidade Federal de Ouro Preto e nos cursos de pós-graduação em Ciência da Computação e Mestrado Profissional em Instrumentação, Controle e Automação. Seus interesses de pesquisa incluem processamento de sinais biomédicos, sistemas embarcados, reconhecimento de padrões e aprendizagem

de máquina.