

JULIANO DE PAULA GONÇALVES

**APRENDIZADO PROFUNDO APLICADO À QUANTIFICAÇÃO DA SEVERIDADE
DE SINTOMAS DE DOENÇAS E PRAGAS FOLIARES**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Engenharia Agrícola, para obtenção do título de *Magister Scientiae*.

Orientador: Francisco de Assis de Carvalho Pinto

**VIÇOSA - MINASGERAIS
2020**

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade
Federal de Viçosa - Campus Viçosa**

T

G635a
2020
Gonçalves, Juliano de Paula, 1992-
Aprendizado profundo aplicado à quantificação da
severidade de sintomas de doenças e pragas foliares / Juliano de
Paula Gonçalves. – Viçosa, MG, 2020.
70 f. : il. (algumas color.) ; 29 cm.

Orientador: Francisco de Assis de Carvalho Pinto.
Dissertação (mestrado) - Universidade Federal de Viçosa.
Inclui bibliografia.

1. Café - Doenças e pragas. 2. Soja - Doenças e pragas.
3. Trigo - Doenças e pragas. 4. Análise foliar. 5. Inteligência
artificial. I. Universidade Federal de Viçosa. Departamento de
Engenharia Agrícola. Programa de Pós-Graduação em
Engenharia Agrícola. II. Título.

CDD 22. ed. 633.73978

JULIANO DE PAULA GONÇALVES

**APRENDIZADO PROFUNDO APLICADO À QUANTIFICAÇÃO DA SEVERIDADE
DE SINTOMAS DE DOENÇAS E PRAGAS FOLIARES**


Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Engenharia Agrícola, para obtenção do título de *Magister Scientiae*.

APROVADA: 24 de julho de 2020.

Assentimento:



Juliano de Paula Gonçalves
Autor



Francisco de Assis de Carvalho Pinto
Orientador

**A Deus e aos meus pais Vicente de Paula Gonçalves e Cleusa Aparecida Pinto
Gonçalves, pelo amor, carinho, incentivo e confiança...**

Dedico.

AGRADECIMENTOS

A Deus por estar ao meu lado em todos os momentos que vivi ao entrar para Universidade Federal de Viçosa, principalmente em momentos de fraqueza onde pensei que não iria conseguir chegar até aqui. Obrigado pela saúde e o dom do conhecimento para que pudesse construir um mundo melhor através desta pesquisa.

Aos meus pais Vicente de Paula Gonçalves e Cleusa Aparecida Pinto Gonçalves e minha irmã Bruna de Paula Gonçalves por todo esforço, incentivo, carinho e compreensão desde o período escolar até a universidade. Todas as conquistas que obtive na vida foram graças a vocês. Obrigado por tudo e serei eternamente grato.

Agradeço a todas as pessoas que contribuíram para o desenvolvimento desta pesquisa, especialmente:

Ao amigo e professor Dr. Francisco de Assis de Carvalho Pinto, que esteve ao meu lado como orientador desde 2014. Todas as oportunidades no âmbito acadêmico foram conquistadas graças a sua competência e dedicação em ensinar e incentivar ir além nos conhecimentos.

Aos Coorientadores, Dr. Daniel Marçal de Queiroz e a Dr^a. Flora Maria de Melo Villar, pelos conhecimentos compartilhados e principalmente por acreditarem e motivarem a desenvolver um trabalho de qualidade.

Ao coorientador Dr. Emerson Medeiros Del Ponte, no qual tive a oportunidade de convidar neste último ano para fazer parte desta pesquisa. Obrigado pela amizade e por todo esforço e paciência em compartilhar todos os conhecimentos em uma área nova na minha carreira como pesquisador, a fitopatologia.

Ao pesquisador da Embrapa Informática Agropecuária, Jayme Garcia Anal Barbedo por toda solicitude e recomendações na condução desta pesquisa. Para mim foi motivo de honra ter trabalhado e acrescentado suas sugestões e ensinamentos no desenvolvimento desta pesquisa.

A minha namorada Camylla Engelender Oliveira Cruz, pela ajuda na coleta de dados e pré-processamentos dos dados. Além disso, por sempre dizer: “Você vai conseguir meu amor”. Minha eterna gratidão.

Aos colegas do Laboratório de Mecanização Agrícola (LMA) Jorge, Rodrigo, Marcelo, Fernando, Lucas e André pela amizade e troca de conhecimentos que me fizeram crescer profissionalmente e pessoalmente.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

Enfim, agradeço a todos os demais familiares e amigos que não foram citados aqui, mas que contribuíram para que eu pudesse chegar até aqui.

RESUMO

GONÇALVES, Juliano de Paula Gonçalves, M.Sc., Universidade Federal de Viçosa, julho de 2020. **Aprendizado profundo aplicado à quantificação da severidade de sintomas de doenças e pragas foliares.** Orientador: Francisco de Assis de Carvalho Pinto.

Os métodos propostos atualmente para a segmentação de imagens digitais superaram a exatidão e precisão do método visual na estimativa da severidade de doenças foliares. Entretanto, a combinação de fatores como exatidão, precisão, aplicabilidade em uma gama de condições e rapidez, faz com que as estimativas visuais continuem sendo o método mais usado, tanto em ambiente controlado, como em campo. Os objetivos deste estudo foram: (1) avaliar o aprendizado profundo, mais especificamente, as redes neurais convolucionais (CNNs) na segmentação semântica de imagens digitais para estimativa de severidade de doenças e pragas foliares; e (2) investigar os fatores que afetam os desempenhos dos modelos de CNNs na segmentação semântica de imagens digitais. O conjunto de dados de imagem deste estudo reuniu 766 imagens compostas por uma única folha com sintomas cloróticos e necróticos causados por duas doenças fúngicas (ferrugem da soja e mancha bronzeada do trigo) e um inseto-praga (bicho-mineiro do café). As imagens foliares com sintomas de ferrugem da soja e mancha bronzeada do trigo foram capturadas em laboratório, com a iluminação homogênea e o plano de fundo padronizado. Diferentemente, as imagens foliares com sintomas do inseto-praga bicho-mineiro do café foram capturadas em condição de campo, com iluminação irregular e fundo complexo. Três classes semântica foram anotadas manualmente em cada imagem: fundo da imagem, área foliar saudável e área foliar lesionada. Os fatores investigados que poderiam afetar o desempenho dos modelos de CNN foram: plano de fundo da imagem, aumento de dados, resolução da imagem, *backbone*, número de dobras K na validação cruzada *k-fold*, algoritmo de otimização e aprendizado por transferência. O modelo de Rede de Pirâmides Características (FPN) tendeu a apresentar o melhor desempenho na segmentação do conjunto de imagens de teste, atingindo uma Interseção-sobre-União (IoU) de 98,9% para o fundo da imagem, 94,1% para área foliar saudável e 79,4% para a área lesionada. Quando previsões em nível de pixel foram usadas para calcular a severidade das folhas, o modelo FPN superou o *software* Assess, referência para os fitopatologistas, em todas as doenças e pragas analisadas, com coeficientes de concordância de 0,97, 0,98 e 0,98 para ferrugem da soja, mancha bronzeada do trigo e bicho-mineiro do café, respectivamente. Apresentando um método totalmente automático de segmentação, o tempo de inferência e estimativa da severidade dos

modelos de CNNs não ultrapassou os 6s, tempo aproximadamente 20 vezes menor que o obtido pelo usuário no modo manual de segmentação do Assess. Os experimentos computacionais mostraram que os fatores analisados tenderam a impactar o desempenho do modelo FPN na segmentação semântica de imagens foliares. Os resultados se mostraram promissores, principalmente pela capacidade dos modelos de segmentar imagens em condições desafiadoras de iluminação e fundo complexo de forma precisa e automática.

Palavras-chave: Redes neurais convolucionais. Segmentação semântica. Processamento de imagens.

ABSTRACT

GONÇALVES, Juliano de Paula Gonçalves, M.Sc., Universidade Federal de Viçosa, July, 2020.
Deep learning applied to quantify the severity of symptoms of leaf diseases and pests.
Adviser: Francisco de Assis de Carvalho Pinto.

The methods currently proposed for digital image segmentation have surpassed the accuracy and precision of the visual method in estimating the severity of leaf diseases. However, the combination of factors such as accuracy, precision, applicability in a range of conditions and speed of analysis, makes that visual estimates continue to be the most used method, both in a controlled environment and in the field. The objectives of this study were: (1) to evaluate the deep learning, more specifically, the performance of convolutional neural networks (CNNs) in the semantic segmentation of digital images to estimate the severity of diseases and leaf pests; and (2) to investigate the factors affecting the performance of CNN models in the semantic segmentation of digital images. The image dataset of this study gathered 766 images composed of a single leaf with chlorotic and necrotic symptoms caused by two fungal diseases (soybean rust and wheat tan spot) and a pest insect (coffee leaf miner). Leaf images with symptoms of soybean rust and wheat tan spot were obtained in the laboratory, with homogeneous lighting and standardized background. Conversely, the leaf images with symptoms of the coffee leaf miner were captured under field conditions, with irregular lighting and complex background. Three semantic classes were manually annotated on each image: background of the image, healthy and damaged leaf area. The factors investigated that could affect the performance of the CNN models were image background, data increase, image resolution, backbone, number of folds (k) in the k-fold cross-validation, optimization algorithm and transfer learning. The Feature Pyramid Network (FPN) model tended to present the best performance in the segmentation of the test image set, presenting an Intersection-Over-Union (IoU) of 98.9% for the image background, 94.1% for healthy leaf area, and 79.4% for the damaged area. When pixel-level predictions were used to calculate leaf severity, the FPN model outperformed the Assess software, the reference used for plant pathologists, for all diseases and pests analyzed, with concordance coefficients of 0.97, 0.98, and 0.98 for soybean rust, wheat tan spot, and coffee leaf miner, respectively. Presenting a fully automatic segmentation method, the inference time and estimate of the severity of the CNN models did not exceed 6s, a time approximately 20 times shorter than that obtained by the user in the manual mode of segmentation of the Assess. The computational experiments showed that the factors analyzed tended to impact the

performance of the FPN model in the semantic segmentation of leaf images. The results were promising, mainly because of the models' ability to segment images under challenging lighting and complex background conditions in an accurate and automatic way.

Keywords: Convolutional neural networks. Semantic segmentation. Image processing.

SUMÁRIO

1 INTRODUÇÃO GERAL	12
1.1 REFERÊNCIAS	15
2 ARTIGO I - ESTIMATIVA AUTOMÁTICA DA SEVERIDADE DE DOENÇAS E PRAGAS DE PLANTAS: UMA ABORDAGEM DE SEGMENTAÇÃO SEMÂNTICA DE IMAGENS DIGITAIS.....	18
RESUMO.....	18
ABSTRACT	19
2.1 INTRODUÇÃO.....	20
2.2 MATERIAL E MÉTODOS.....	23
2.2.1 Conjunto de dados de imagem.....	23
2.2.2 Anotação das imagens	24
2.2.3 Modelos de segmentação semântica de última geração	26
2.2.4 Implementação dos modelos	27
2.2.5 Inicialização.....	28
2.2.6 Pré-processamento e aumento de dados	28
2.2.7 Validação cruzada.....	29
2.2.8 Métricas de avaliação dos modelos	29
2.2.9 Comparação com o método referência na estimativa da severidade	30
2.3 RESULTADOS	31
2.3.1 Comparação de desempenho de modelos de CNN.....	31
2.3.2 Severidade estimada pelos modelos de CNN.....	33
2.3.3 Severidade estimada pelo <i>software</i> Assess.....	36
2.3.4 Tempo de inferência	38
2.4 DISCUSSÃO	39
2.5 CONCLUSÃO.....	42

2.6 REFERÊNCIAS	43
3 ARTIGO II - FATORES QUE INFLUENCIAM O DESEMPENHO DAS REDES NEURAIAS CONVOLUCIONAIS NA SEGMENTAÇÃO SEMÂNTICA DE PRAGAS DE PLANTAS	49
RESUMO.....	49
ABSTRACT	50
3.1 INTRODUÇÃO.....	51
3.2 MATERIAL E MÉTODOS.....	53
3.2.1 Conjunto de dados de imagem.....	53
3.2.2 Configuração experimental.....	54
3.2.3 Métrica de avaliação.....	56
3.3 RESULTADOS E DISCUSSÃO.....	56
3.3.1 Plano de fundo da imagem	57
3.3.2 Aumento de dados	59
3.3.3 Resolução da imagem.....	60
3.3.4 Backbone	61
3.3.5 Número de dobras k.....	62
3.3.6 Algoritmo de otimização	62
3.3.7 Aprendizado por transferência.....	64
3.4 CONCLUSÃO.....	65
3.5 REFERÊNCIAS	66
4. CONCLUSÃO GERAL	70

1 INTRODUÇÃO GERAL

A severidade de doenças de plantas em unidades foliares é definida como a proporção ou percentagem da área afetada pela doença (Preston, 2017). O monitoramento da severidade é fundamental para avaliar e prever epidemias (Moreno-Velázquez et al., 2020), avaliar a eficácia de fungicidas (Peng et al., 2020), cultivares resistentes a doenças (Jacobs et al., 2020) e estimar a perda de rendimento das culturas (Bock e Chiang, 2019). Portanto, é imprescindível o desenvolvimento de métodos fáceis de usar, altamente exatos e precisos e aplicáveis a uma gama de situações (Nutter et al., 1993).

A estimativa da severidade pode ser feita visualmente ou por análise de imagens digitais. No método visual, se popularizou o uso dos “Diagramas de Área Padrão” (*Standard Area Diagrams - SAD*). O SAD é um conjunto de imagens, coloridas ou não, com severidade conhecida que auxilia o avaliador na estimativa da severidade. Diversos estudos mostraram que o uso dos SADs aumentam a acurácia e precisão do avaliador, principalmente com pouca experiência (Dolinski et al., 2017; Fantin et al., 2018; Santos and Spósito, 2018; Sterling et al., 2020). Entretanto, o método visual é comprovadamente subjetivo, pois, o desempenho do avaliador depende de fatores como a característica da lesão, experiência e fadiga do avaliador (Bock et al., 2008; Moya et al., 2005).

Com desenvolvimento de tecnologias de visão computacional aplicadas na segmentação de imagens digitais, o método de análise de imagem digital superou a exatidão e precisão do método visual, mesmo com auxílio dos SADs, na estimativa da severidade (Bock et al., 2009). Os *softwares* mais populares entre os fitopatologistas são o QUANT (Vale et al., 2003) e o *Assess* (Lamari, 2002). O *software* QUANT oferece cinco maneiras para segmentar a imagem e, conseqüentemente, estimar a severidade: limiar, seleção de cores por paleta, intervalo de coordenadas de cores nos espaços de cores RGB (vermelho-verde-azul) ou HSV (matiz-saturação-valor), amostragem de cores e funções discriminantes. O *software* *Assess* versão 2.0, em seu modo manual, permite que o usuário trabalhe com imagens no espaço de cores RGB, HSV ou L^*a^*b (Luminosidade e coordenadas cromáticas a^* e b^*) e selecione o canal e o limiar de segmentação que melhor se ajuste para segmentar a folha e os sintomas da doença.

Pensando na portabilidade e no aproveitando dos recursos que um *smartphone* pode oferecer, como câmera e capacidade de processamento e análise, Pethybridge e Nelson (2015) desenvolveram o Leaf Doctor. O Leaf Doctor é um aplicativo para *smartphone* que permite que o usuário toque na tela e selecione até oito cores diferentes que representam tecidos saudáveis.

O usuário move um controle deslizante para seleção do limiar até que apenas os tecidos sintomáticos sejam transformados em um tom azul. A imagem pixelizada resultante é então analisada para calcular a porcentagem da doença. Alternativamente, Barbedo (2016a) propôs um algoritmo semiautomático que manipula o histograma dos canais H (do espaço de cor HSV) e a^* (do espaço de cor L^*a^*b) e define o valor do limiar de forma automática para segmentar os tecidos saudáveis e doentes. Ao fim do processo, o usuário decide qual canal (H ou a^*) fornece a melhor diferenciação.

Todos os *softwares* e algoritmos citados até aqui foram desenvolvidos com base no método de segmentação limiar. A principal limitação deste método é que eles são precisos somente se existir um contraste óbvio entre duas regiões a ser segmentada. Entretanto, o contraste entre a folha e fundo da imagem, por exemplo, pode não ser tão óbvio assim, principalmente se houver outros elementos de cor verde. Além disso, em condições de campo, a incidência solar sobre a superfície foliar pode causar reflexões especulares, reduzindo o contraste entre o tecido saudável e lesionado.

Alguns estudos adotaram diferentes estratégias na aquisição de imagens digitais foliares: introdução de painéis na cor azul ou preta ao plano de fundo da imagem, digitalização, alteração do ângulo de captura das imagens para redução do reflexo na superfície foliar e remoção manual do fundo da imagem (Barbedo, 2016b). Essas medidas, de fato, criam o contraste necessário para segmentar a imagem de forma precisa, mas elevam o tempo necessário para realização das estimativas. Portanto, algoritmos e *softwares* usados para estimar a severidade tiveram sua aplicação direcionada, principalmente, a ambientes controlados, aplicados no suporte para o desenvolvimento dos SADs (Costa Lage et al., 2015; Freire Fagundes-Nacarath et al., 2018; Melo et al., 2020). De fato, Del Ponte et al. (2017) mostraram em sua revisão que de 103 estudos realizados para proposição dos SADs, 82 usaram a segmentação de imagens digitais como a estimativa real da severidade.

A fim de abordar condições mais desafiadoras para os métodos de segmentação, como fundo complexo e iluminação irregular, alguns trabalhos aplicaram o método do crescimento de região para segmentação de sintomas de doenças (Jothiaruna et al., 2019; Ma et al., 2017). Este método consiste em agrupar pixels vizinhos semelhantes, por exemplo, em cor e textura, ou sub-regiões de agrupamento em regiões homogêneas maiores, de acordo com um critério de crescimento (Oliveira et al., 2016). Apesar de apresentarem bom desempenho no processo de segmentação da mancha da doença, a seleção das sementes no mapa crescimento é de maneira interativa, o que significa que os métodos não são automáticos. Além disso, não foi testada a capacidade dos algoritmos em segmentar a área saudável da folha, requisito necessário na

estimativa da severidade de doenças foliares em percentagem.

Recentemente, alguns estudos testaram o desempenho das redes neurais convolucionais (*Convolutional Neural Networks* - CNNs) na estimativa de doenças foliares e obtiveram resultados promissores (Esgario et al., 2020; Liang et al., 2019; Wang et al., 2017). A classificação a nível de imagem feita pelos modelos de CNNs é automática, o que elimina a necessidade da intervenção humana no processo e permite avaliar um número grande de imagens em um período relativamente curto. Entretanto, permite apenas estimar a severidade em categorias, apresentando pouca aplicabilidade para fins de pesquisas, especialmente em avaliações de cultivares resistentes e efeito de tratamento de fungicidas foliares. Para essas situações, a porcentagem de severidade é mais informativa (Bock et al., 2016).

Além dos avanços científicos na classificação de imagens, as CNNs apresentaram desempenhos encorajadores na segmentação semântica de imagens digitais. Com um método totalmente automático, as CNNs atribuem uma classe semântica a cada pixel na imagem de entrada (Krestenitis et al., 2019). A aplicação de modelos de CNNs na segmentação semântica de doenças e pragas foliares é incipiente. O primeiro trabalho foi conduzido por Lin et al. (2019), que obtiveram 50 imagens de folhas de pepino afetadas pelo oídio em uma plataforma de fenotipagem estacionária, onde um fundo de difusão na parte inferior fornece iluminação uniforme e uma fonte de luz artificial periférica na parte superior para minimizar a sombra. O modelo de CNN Unet superou outros métodos tradicionais de aprendizado de máquina em uma amostra de 20 imagens: a acurácia foi de 96,08% e a interseção sobre a união de 72,11%. Apesar de resultados promissores, as imagens foram capturadas em condições controladas, com fundo simples e luminosidade controlada, condição na qual grande parte dos métodos propostos na estimativa da severidade são validados.

Diante do exposto, o objetivo deste trabalho foi avaliar o desempenho dos modelos de aprendizado profundo, mais especificamente as CNN, na segmentação semântica e estimativa da severidade de doenças e pragas foliares. Primeiramente, uma coleção de imagens com plantas doentes ou necrosadas com três sintomas distintos foi coletada em condições de laboratório e de campo, usando câmeras e scanners de mesa. O primeiro artigo deste estudo teve como objetivo treinar e comparar cinco modelos de CNN de última geração na segmentação semântica e estimativa da severidade de doenças foliares. O segundo artigo teve o objetivo analisar os fatores que influenciam o desempenho das CNN na segmentação semântica.

1.1 REFERÊNCIAS

- Barbedo, J.G.A., 2016a. A novel algorithm for semi-automatic segmentation of plant leaf disease symptoms using digital image processing. *Trop. Plant Pathol.* 41, 210–224. <https://doi.org/10.1007/s40858-016-0090-8>
- Barbedo, J.G.A., 2016b. A review on the main challenges in automatic plant disease identification based on visible range images. *Biosyst. Eng.* 144, 52–60. <https://doi.org/10.1016/j.biosystemseng.2016.01.017>
- Bock, C.H., Chiang, K.-S., Del Ponte, E.M., 2016. Accuracy of plant specimen disease severity estimates: concepts, history, methods, ramifications and challenges for the future. *CAB Rev. Perspect. Agric. Vet. Sci. Nutr. Nat. Resour.* 11. <https://doi.org/10.1079/PAVSNR201611032>
- Bock, C.H., Chiang, K.S., 2019. Disease incidence-severity relationships on leaflets, leaves, and fruit in the pecan-venturia effusa pathosystem. *Plant Dis.* 103, 2865–2876. <https://doi.org/10.1094/PDIS-11-18-1950-RE>
- Bock, C.H., Cook, A.Z., Parker, P.E., Gottwald, T.R., 2009. Automated image analysis of the severity of foliar citrus canker symptoms. *Plant Dis.* 93, 660–665. <https://doi.org/10.1094/PDIS-93-6-0660>
- Bock, C.H., Parker, P.E., Cook, A.Z., Gottwald, T.R., 2008. Characteristics of the Perception of Different Severity Measures of Citrus Canker and the Relationships Between the Various Symptom Types. *Plant Dis.* 92, 927–939. <https://doi.org/10.1094/PDIS-92-6-0927>
- Costa Lage, D.A., Marouelli, W.A., da S. S. Duarte, H., Café-Filho, A.C., 2015. Standard area diagrams for assessment of powdery mildew severity on tomato leaves and leaflets. *Crop Prot.* 67, 26–34. <https://doi.org/10.1016/j.cropro.2014.09.014>
- Del Ponte, E.M., Pethybridge, S.J., Bock, C.H., Michereff, S.J., Machado, F.J., Spolti, P., 2017. Standard Area Diagrams for Aiding Severity Estimation: Scientometrics, Pathosystems, and Methodological Trends in the Last 25 Years. *Phytopathology* 107, PHYTO-02-17-006. <https://doi.org/10.1094/PHYTO-02-17-0069-FI>
- Dolinski, M.A., Duarte, H. da S.S., da Silva, J.B., May De Mio, L.L., 2017. Development and validation of a standard area diagram set for assessment of peach rust. *Eur. J. Plant Pathol.* 148, 817–824. <https://doi.org/10.1007/s10658-016-1138-9>
- Esgario, J.G.M., Krohling, R.A., Ventura, J.A., 2020. Deep learning for classification and severity estimation of coffee leaf biotic stress. *Comput. Electron. Agric.* 169, 105162. <https://doi.org/10.1016/j.compag.2019.105162>
- Fantin, L.H., Braga, K., Canteri, M.G., Dias, A.R., Borges, E.P., 2018. Development and validation of diagrammatic scale to assess target spot severity in cotton. *Australas. Plant Pathol.* 47, 491–497. <https://doi.org/10.1007/s13313-018-0576-6>
- Freire Fagundes-Nacarath, I.R., Debona, D., Costa Novaes, M.I., Rodrigues, F.Á., 2018. A set of standard area diagrams to assess white mold severity on the leaflets of common beans. *Crop Prot.* 114, 60–67. <https://doi.org/10.1016/j.cropro.2018.08.007>

- Jacobs, R.L., Adhikari, T.B., Pattison, J., Yencho, G.C., Fernandez, G.E., Louws, F.J., 2020. Assessing Rate-Reducing Foliar Resistance to Anthracnose Crown Rot and Fruit Rot in Strawberry. *Plant Dis.* 104, 398–407. <https://doi.org/10.1094/PDIS-04-19-0687-RE>
- Jothiaruna, N., Joseph Abraham Sundar, K., Karthikeyan, B., 2019. A segmentation method for disease spot images incorporating chrominance in Comprehensive Color Feature and Region Growing. *Comput. Electron. Agric.* 165, 104934. <https://doi.org/10.1016/J.COMPAG.2019.104934>
- Krestenitis, M., Orfanidis, G., Ioannidis, K., Avgerinakis, K., Vrochidis, S., Kompatsiaris, I., 2019. Oil Spill Identification from Satellite Images Using Deep Neural Networks. *Remote Sens.* 11, 1762. <https://doi.org/10.3390/rs11151762>
- Lamari, L., 2002. ASSESS: Image Analysis Software for Plant Disease Quantification. American Phytopathological Society, St. Paul, Minnesota: APS Press.
- Liang, Q., Xiang, S., Hu, Y., Coppola, G., Zhang, D., Sun, W., 2019. PD2SE-Net: Computer-assisted plant disease diagnosis and severity estimation network. *Comput. Electron. Agric.* 157, 518–529. <https://doi.org/10.1016/j.compag.2019.01.034>
- Lin, K., Gong, L., Huang, Y., Liu, C., Pan, J., 2019. Deep Learning-Based Segmentation and Quantification of Cucumber Powdery Mildew Using Convolutional Neural Network. *Front. Plant Sci.* 10. <https://doi.org/10.3389/fpls.2019.00155>
- Ma, J., Du, K., Zhang, L., Zheng, F., Chu, J., Sun, Z., 2017. A segmentation method for greenhouse vegetable foliar disease spots images using color information and region growing. *Comput. Electron. Agric.* 142, 110–117. <https://doi.org/10.1016/j.compag.2017.08.023>
- Melo, V.P. de, Mendonça, A.C. da Si., Souza, H.S. de, Gabriel, L.C., Bock, C., Eaton, M.J., Schwan-Estrada, K.R.F., Nunes, W.M. de C., 2020. Reproducibility of Development and Validation Process of Standard Area Diagram by Two Laboratories: an Example Using the *Botrytis cinerea*/*Gerbera jamesonii* Pathosystem. *Plant Dis.* PDIS-08-19-1708-RE. <https://doi.org/10.1094/PDIS-08-19-1708-RE>
- Moreno-Velázquez, M., Hernández-Ramos, L., Alvarado-Rosales, D., Hernández-Pablo, S., Quezada-Salinas, A., 2020. First Report of Powdery Mildew Caused by *Erysiphe heraclei* on Fennel (*Foeniculum vulgare*) in Mexico. *Plant Dis.* PDIS-08-19-1604. <https://doi.org/10.1094/PDIS-08-19-1604-PDN>
- Moya, E.A., Barrales, L.R., Apablaza, G.E., 2005. Assessment of the disease severity of squash powdery mildew through visual analysis, digital image analysis and validation of these methodologies. *Crop Prot.* 24, 785–789. <https://doi.org/10.1016/j.cropro.2005.01.003>
- Nutter, F.W., Gleason, M.L., Jenco, J.H., Christians, N.L., 1993. Assessing the Accuracy, Intra-rater Repeatability, and Inter-rater Reliability of Disease Assessment Systems. *Phytopathology* 83, 806–812. <https://doi.org/10.1094/Phyto-83-806>
- Oliveira, R.B., Filho, M.E., Ma, Z., Papa, J.P., Pereira, A.S., Tavares, J.M.R.S., 2016. Computational methods for the image segmentation of pigmented skin lesions: A review. *Comput. Methods Programs Biomed.* <https://doi.org/10.1016/j.cmpb.2016.03.032>

- Peng, F., Miu, S., Yang, Z., Fu, Y., Yang, Y., Yu, Y., Bi, C., 2020. Rapid Quantification of Fungicide Effectiveness on Inhibiting Wheat Stripe Rust Pathogen (*Puccinia striiformis* f. sp. *tritici*). *Plant Dis.* PDIS-09-19-1836-RE. <https://doi.org/10.1094/PDIS-09-19-1836-RE>
- Pethybridge, S.J., Nelson, S.C., 2015. Leaf Doctor: A New Portable Application for Quantifying Plant Disease Severity. *Plant Dis.* 99, 1310–1316. <https://doi.org/10.1094/PDIS-03-15-0319-RE>
- Preston, R., 2017. CHAPTER 2: Measuring Plant Diseases, in: Madden, L. V., Hughes, G., van den Bosch, F. (Eds.), *The Study of Plant Disease Epidemics*. The American Phytopathological Society, pp. 11–31. <https://doi.org/https://doi.org/10.1094/9780890545058.002>
- Santos, R.F., Spósito, M.B., 2018. Improving assessments of anthracnose severity on grapevine leaves through the development of a standard area diagram set. *Australas. Plant Pathol.* 47, 357–364. <https://doi.org/10.1007/s13313-018-0566-8>
- Sterling, A., Gómez-Torres, A.K., Suárez-Córdoba, Y.D., Loaiza-Molina, L.C., Sierra-Hayer, J.F., 2020. Standard area diagrams to assess black crust (*Phyllachora huberi*) severity on rubber tree leaflets. *Eur. J. Plant Pathol.* 156, 827–837. <https://doi.org/10.1007/s10658-020-01933-x>
- Vale, F.X.R., Fernandes Filho, E.I., Liberato, J.R., 2003. QUANT. A software for plant disease severity assessment., in: *In: 8th International Congress of Plant Pathology*. Christchurch New Zealand, p. 105.
- Wang, G., Sun, Y., Wang, J., 2017. Automatic Image-Based Plant Disease Severity Estimation Using Deep Learning. *Comput. Intell. Neurosci.* 2017, 1–8. <https://doi.org/10.1155/2017/2917536>

2 ARTIGO I - ESTIMATIVA AUTOMÁTICA DA SEVERIDADE DE DOENÇAS E PRAGAS DE PLANTAS: UMA ABORDAGEM DE SEGMENTAÇÃO SEMÂNTICA DE IMAGENS DIGITAIS

RESUMO

A estimativa da severidade de sintomas causados por doenças e pragas é uma das tarefas mais importantes na fitopatologia. As estimativas visuais ainda prevalecem, mas confiabilidade é prejudicada por inconsistências devido a fenômenos ópticos e cognitivos e à subjetividade intrínseca da tarefa. Os métodos de segmentação de imagens digitais aparecem como uma solução promissora, mas grande parte adotou o método de segmentação limiar, embora preciso, requer imagens capturadas em condições controladas de fundo de imagem e luminosidade. Neste estudo, avaliamos cinco modelos de redes neurais convolucionais (CNNs) na segmentação semântica e estimativa da severidade do bicho-mineiro do café, ferrugem da soja, e mancha bronzeada do trigo. Os resultados experimentais mostraram que os modelos de CNN apresentaram bom desempenho na segmentação e estimativa da severidade. O modelo de Redes de Pirâmides Características (FPN) tendeu a apresentar melhor desempenho na segmentação do conjunto de imagens de teste, atingindo uma Interseção-sobre-União (IoU) média de 90,8% sob condições desafiadoras de iluminação e de plano de fundo de imagens. Quando previsões em nível de pixel foram usadas para calcular a severidade das folhas, o modelo FPN teve o melhor desempenho para todos os três conjuntos de dados de imagem, com coeficientes de concordância de 0,97, 0,98 e 0,98 para ferrugem da soja, mancha bronzeada do trigo e bicho-mineiro do café, respectivamente.

ABSTRACT

Estimation of severity symptoms caused by diseases and pests is one of the most important tasks in plant pathology. Visual assessments still prevail, but their reliability is hindered due to inconsistencies of optical and cognitive phenomena, and the intrinsic subjectivity of the task. Digital image segmentation methods appear as a promising solution, however, most of them adopted the threshold method, which although accurate, it requires images captured under controlled conditions of background and brightness. In this study, five models of convolutional neural networks (CNNs) were evaluated for semantic segmentation and estimation of severity symptoms of coffee leaf miner, soybean rust and wheat tan spot. Experimental results showed that the CNNs models performed well in segmentation and severity estimation. The Feature Pyramid Network (FPN) model achieved the best performance in the test dataset segmentation, in which the Intersection-over-Union (IoU) averaged 90.8% under challenging conditions of lighting and background. When pixel-level predictions were used to calculate leaf severity, the FPN model performed the best for all three image datasets, with concordance coefficients 0.98, 0.97 and 0.98 for coffee leaf miner, soybean rust and wheat tan spot datasets, respectively.

2.1 INTRODUÇÃO

A produção agrícola, florestal e ornamental e pastoril é constantemente ameaçada por pragas e patógenos, como fungos, vírus, bactérias e nematoides (Oerke, 2006; Savary et al., 2017). A capacidade de detectar e quantificar com precisão esses sintomas é essencial para a pesquisa e decisões de gerenciamento (Bock et al., 2020). A estimativa da severidade em porcentagem da área afetada é fundamental para determinar os efeitos de tratamento e o rendimento das culturas (Madden et al., 2007). No entanto, as estimativas de severidade são inerentemente complexas e, dependendo do método de avaliação, vieses não são incomuns (Bock et al., 2020). Estimativas tendenciosas podem ser muito prejudiciais (Chiang et al., 2014), portanto, é necessário se dispor de métodos mais confiáveis e eficazes para estimativa de gravidade (Franceschi et al., 2020).

A estimativa visual da severidade percebida pelo olho humano pode ser realizada em diferentes escalas (nominal, ordinal ou razão) (Bock et al., 2020). Dada a variabilidade dos sintomas da doença (posição, tamanho, forma e cor) e a natureza subjetiva das estimativas de severidade visual, independentemente da escala, a falta de precisão entre os avaliadores têm sido a norma e não a exceção (Bock et al., 2010; Nita et al., 2003; Nutter et al., 1993). Para melhorar a precisão das estimativas visuais foram desenvolvidas ferramentas denominadas "Diagrama de Área Padrão" (*Standard Area Diagrams - SAD*). O SAD é um conjunto de imagens de sintomas em preto e branco ou coloridos em um número limitado (5 a 10 imagens) com a severidade conhecida para ajudar os avaliadores a segmentar mentalmente a área afetada durante a estimativa visual. Entretanto, independentemente de todos os avanços no desenvolvimento de novas escalas e SADs, a qualidade das avaliações visuais depende da capacidade e experiência do avaliador - e essas habilidades não podem ser transportadas de um avaliador para outro.

Para superar este problema, o uso de análise de imagem digitais como substituto das estimativas visuais vem sendo proposto nas últimas décadas (Clément et al., 2015; Lindow e Webb, 1983; Škaloudová et al., 2006). Os primeiros *softwares* comerciais específicos para doenças de plantas foram lançados em meados dos anos 2000 (Lamari, 2002; Vale et al., 2003), executados em plataformas estacionárias e desenvolvidos com base na seleção manual do limiar de segmentação. Entretanto, todo o processo pode ser lento devido à necessidade de diferentes instrumentos para aquisição, preparação e processamento de imagens. Recentemente, o aplicativo para *smartphone* Leaf Doctor (Pethybridge e Nelson, 2015), permitiu reduzir consideravelmente o tempo na tarefa, uma vez que a aquisição de imagens (câmera do

smartphone ou *tablet*) e o limiar de segmentação são conduzidos no aplicativo usando a tecnologia de tela sensível ao toque. Além disso, métodos de limiares semiautomáticos e automáticos foram propostos para acelerar o processo de segmentação das imagens (Barbedo, 2014, 2016b; Lamari, 2002).

A principal limitação de métodos do limiar de segmentação, independente se manuais, semiautomáticos ou automáticos, é a perda de precisão se o contraste entre áreas saudáveis e doentes na folha não for nítido. O fundo complexo, reflexões especulares e presença de simultânea luz e sombra tem sido um problema, especialmente quando a cor verde está presente ao fundo da imagem (Barbedo, 2016a). Adicionar um fundo padrão, ajustar condições de luz ou levar o material para o laboratório são soluções propostas para minimizar esses problemas da aquisição de imagens no campo, mas com o custo de velocidade e escalabilidade do processo (Barbedo, 2016a; Karisto et al., 2018; Stewart et al., 2016). Portanto, não é inesperado que os diversos software e algoritmos propostos tenham encontrado uso em pesquisas de pequena escala, como no desenvolvimento do SAD (Del Ponte et al., 2017).

As condições mais complexas, ou desafiadoras, da segmentação de sintomas de doenças de plantas em imagens vêm sendo abordada, como fundo complexo e iluminação irregular, para as quais as informações de cores não são suficientes. O método chamado crescimento de região, que tem como base o agrupamento de pixels vizinhos ou sub-regiões de áreas maiores por similaridade, como cor e textura, foi proposto com relativo sucesso, porém, depende da seleção manual das sementes no mapa de crescimento (Jothiaruna et al., 2019; Ma et al., 2017). Além disso, não foi investigado a segmentação da área foliar saudável, segmentação essencial para estimativa da severidade.

Métodos de aprendizado profundo, especialmente métodos de classificação supervisionados com base em redes neurais convolucionais (*Convolutional Neural Networks - CNNs*), constituem uma das técnicas mais poderosas no reconhecimento de padrões em imagens (Ferentinos, 2018). A aplicação das CNNs aos problemas fitopatológicos começou a integrar vários trabalhos após 2015, principalmente visando a classificação da imagem para diagnosticar o problema (Barbedo, 2018). Apesar da grande aplicação na identificação de doenças foliares, poucos estudos abordaram seu desempenho na estimativa da severidade. Alguns estudos usaram as CNNs para estimar valores categorizados de severidade de doenças de plantas com resultados promissores de precisão, variando de 86,51 a 91% (Esgario et al., 2020; Liang et al., 2019; Wang et al., 2017). Embora atraente, a classificação da imagem pode ser insuficiente para situações em que uma estimativa direta da severidade, e não em classes, se faz necessária. Além disso, todas as imagens usadas nestes trabalhos foram capturadas em

condições controladas, com fundo padronizado de imagem.

Alternativamente, as CNNs trouxeram recentes avanços significativos na segmentação semântica de imagens e agora uma grande variedade de arquiteturas está disponível para a comunidade científica. Na segmentação semântica, as CNNs atribuem uma classe semântica a cada pixel na imagem de entrada. Um dos primeiros trabalhos, desenvolvido por Long et al. (2015), propôs uma rede totalmente convolucional, onde modelos de CNNs voltados para classificação, como AlexNet (Krizhevsky et al., 2012), VGG net (Simonyan e Zisserman, 2014) e GoogLeNet (Szegedy et al., 2015), foram adaptados para tarefa de classificação a nível de pixel através da substituição da camada totalmente conectada por uma camada de convolução transposta, ou deconvolução. Posteriormente, grande parte dos modelos de CNNs aplicados na segmentação semântica adotou uma estrutura codificador-decodificador (Krestenitis et al., 2019). Na seção de codificação, camadas de convoluções e *max pooling* extraem informações de contexto. Na seção de decodificação, camadas de deconvolução e *upsampling* recuperam a informação espacial.

As CNNs têm sido extremamente bem-sucedidas na segmentação semântica, incluindo condução autônoma (R. Palafox et al., 2019; Wang et al., 2020), análise de imagens médicas (Borne et al., 2020; Y. Zhang et al., 2020) e imagens de sensoriamento remoto (Shichen Guo et al., 2019; Krestenitis et al., 2019). Na agricultura, as CNN superaram os métodos tradicionais de aprendizado de máquina na segmentação, tanto em precisão quanto em velocidade, para aplicações em mapeamento do uso da terra (Du et al., 2019; Yang et al., 2019) e segmentação de plantas (Xu et al., 2020). O uso das CNNs na segmentação semântica de imagens a fim resolver problemas de doenças de plantas é incipiente e foi testado em condições com fundo padronizado e luminosidade controlada (Lin et al., 2019), condição na qual grande parte dos métodos propostos na estimativa severidade até o momento foram validados.

Diante do exposto, com o presente trabalho é proposta uma implementação e teste de um novo fluxo de trabalho para a estimativa automática da severidade de doenças de plantas com base na segmentação semântica. Primeiramente, uma coleção de imagens com plantas doentes ou necrosadas com três sintomas distintos foi coletada em condições de laboratório e de campo, usando uma variedade de câmeras ou scanners de mesa. Essas imagens foram usadas para treinar e comparar cinco modelos de CNN para segmentação semântica de última geração. Um método padrão usado na pesquisa de doenças de plantas foi incluído para comparação.

2.2 MATERIAL E MÉTODOS

2.2.1 Conjunto de dados de imagem

O conjunto de dados de imagens deste estudo reuniu 766 imagens com sintomas de doença ou praga foliar (Tabela 1). As imagens de folhas de café lesionadas pelo inseto-praga bicho-mineiro (*Leucoptera coffeella*) foram obtidas durante a realização deste estudo em plantações com ocorrência natural da praga. As folhas foram destacadas e fotografadas em campo usando um *smartphone* Motorola One XT1941-3 (Motorola Inc., Chicago, IL, EUA) na resolução de 13 MPixels. O *smartphone* foi posicionado horizontalmente a uma distância de aproximadamente 20 cm de uma única folha exposta a uma ampla faixa de condições de iluminação e de fundo.

As imagens de folhas de soja foram obtidas de um repositório público e foram usadas em estudos anteriores para o desenvolvimento de um SAD para auxiliar as estimativas visuais de severidade (Franceschi et al. 2020) (Tabela 1). As folhas foram coletadas de vasos de plantas cultivadas em casa de vegetação e obtidas com um *scanner* Hewlett Packard (Modelo 2130) com resolução de 300 dpi, sobre um fundo azul. As folhas de trigo que apresentam sintomas típicos da mancha bronzeada foram obtidas de plantas de várias cultivares de trigo cultivadas em casa de vegetação e inoculadas para avaliação dos genótipos quanto à resistência à mancha bronzeada (Tabela 1). As folhas foram colocadas sobre um fundo preto e as seções das folhas foram fotografadas usando uma câmera Nikon D-3000 com uma resolução de 5,6 MP (Kuhnem et al. 2020).

Tabela 1. Informações dos conjuntos de dados de imagens usadas para avaliar o desempenho dos modelos de CNN na segmentação semântica de sintomas cloróticos e necróticos foliares causados por duas doenças fúngicas e um inseto-praga.

Espécie hospedeira	Doença/Praga	Número de imagens	Localização, Brasileiro	Estado	Fonte
<i>Coffea arabica</i>	Bicho mineiro	406	Viçosa, Minas Gerais		Este estudo
<i>Glycine max</i>	Ferrugem	208	Londrina, Paraná		(Franceschi et al., 2020)
<i>Triticum aestivum</i>	Mancha bronzeada	152	Passo Fundo, Rio Grande do Sul		Kuhnem et al. (2020)

2.2.2 Anotação das imagens

Para que os modelos de CNN sejam treinados a classificar corretamente cada pixel em uma imagem, uma anotação de referência é necessária, indicando um rótulo de classe para cada pixel. Para auxiliar na tarefa, foi desenvolvido um algoritmo para segmentar a imagem e gerar sua anotação de referência. O algoritmo foi desenvolvido no Python 3.6 (<http://www.python.org>) com base nas bibliotecas Numpy (<https://numpy.org>) e OpenCV (Open Source Computer Vision Library), uma biblioteca de *software* de visão computacional e aprendizado de máquina de código aberto (<http://opencv.org>). Na Figura 1 é apresentado o fluxograma do algoritmo proposto.

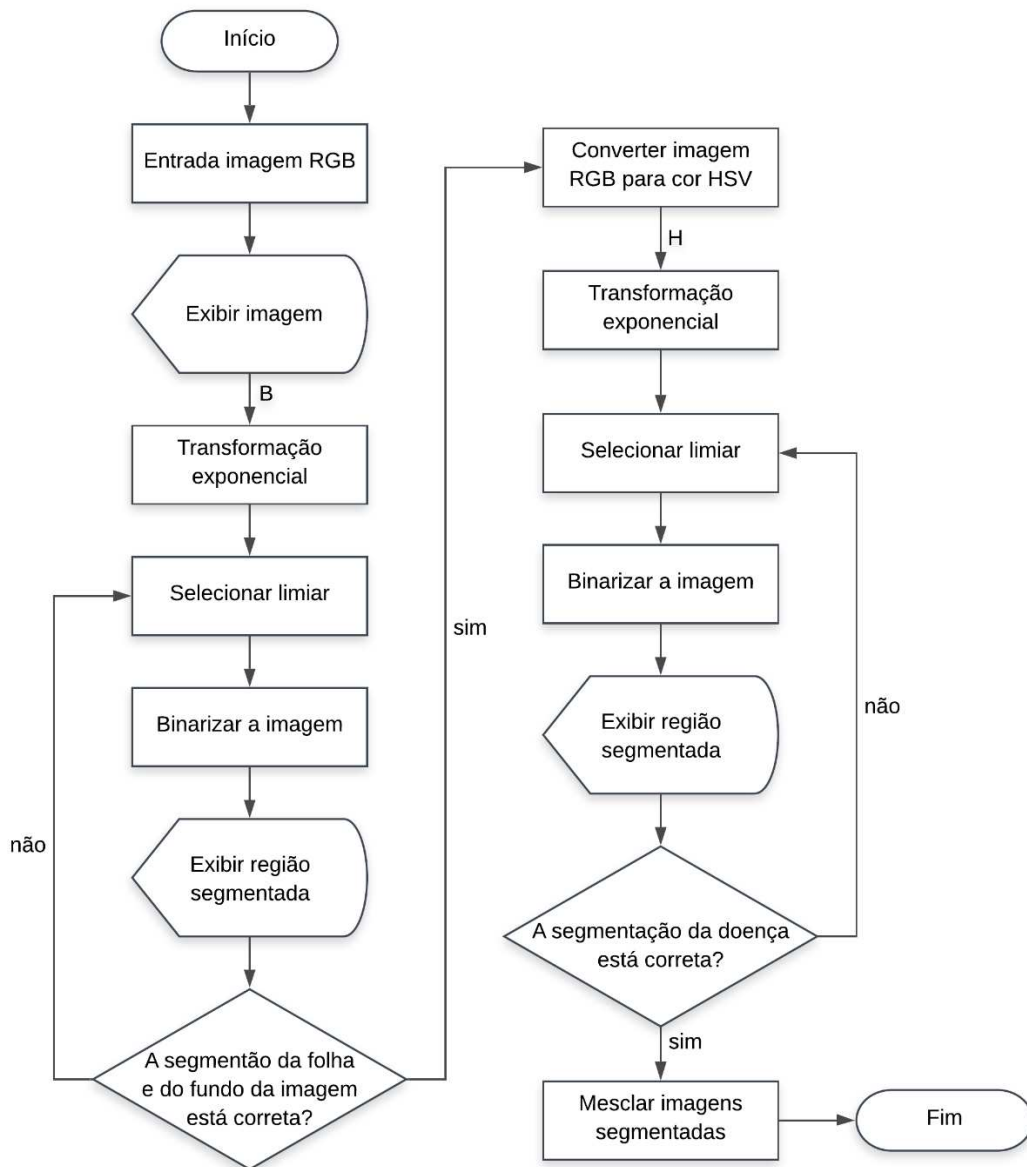


Figura 1. Fluxograma do algoritmo de seleção do limiar desenvolvido para segmentar o fundo da imagem, tecido saudável e sintomas de doenças de uma folha de planta.

O método de segmentação aplicado ao algoritmo é o limiar. Este método não é capaz de segmentar o fundo das imagens em condição de campo de forma satisfatória, principalmente numa condição em que plantas, folhas, solo, ou ambos apareçam. Portanto, os pixels de fundo das imagens em condição de campo foram anotados manualmente.

O algoritmo de segmentação proposto foi dividido basicamente em duas etapas. Na primeira, o objetivo é segmentar o fundo da imagem da folha. Para isso, aplica-se uma transformação exponencial sobre o canal B (Equação 1), em escala de cinza, para melhorar o contraste entre o plano de fundo e a folha.

$$I_s = c[b^{I_e} - 1], \quad (1)$$

em que,

I_s - pixel de saída;

I_e é o pixel de entrada

c - escala para o maior pixel da imagem;

b - coeficiente de ajuste exponencial, neste caso 1,001.

Após a melhoria de contraste, seleciona-se o limiar do pixel e ajusta-se seu valor para obter a melhor segmentação do fundo da imagem e da folha.

Na segunda etapa, o objetivo foi segmentar o tecido da folha com sintomas da doença. Primeiro, transforma-se a imagem RGB no espaço de cor HSV. Para melhorar o contraste do sintoma da doença aplica-se a Equação (1) no canal H. Em seguida, seleciona-se o limiar do pixel e ajusta-se seu valor para obter a melhor segmentação do sintoma da doença. Quando necessário, foram aplicadas correções manuais sobre as máscaras onde o algoritmo de segmentação não obteve êxito. Finalmente, as duas imagens segmentadas são mescladas, gerando uma imagem colorida classificada em fundo, região da folha saudável e região sintomática de folha. Como resultado, uma máscara de três classes foi gerada para cada imagem e foram utilizadas no banco de dados como anotação de referência de cada pixel (Figura 2). A proporção da área foliar afetada, conhecida como severidade (Bock et al., 2009), foi estimada em cada imagem pela Equação 2.

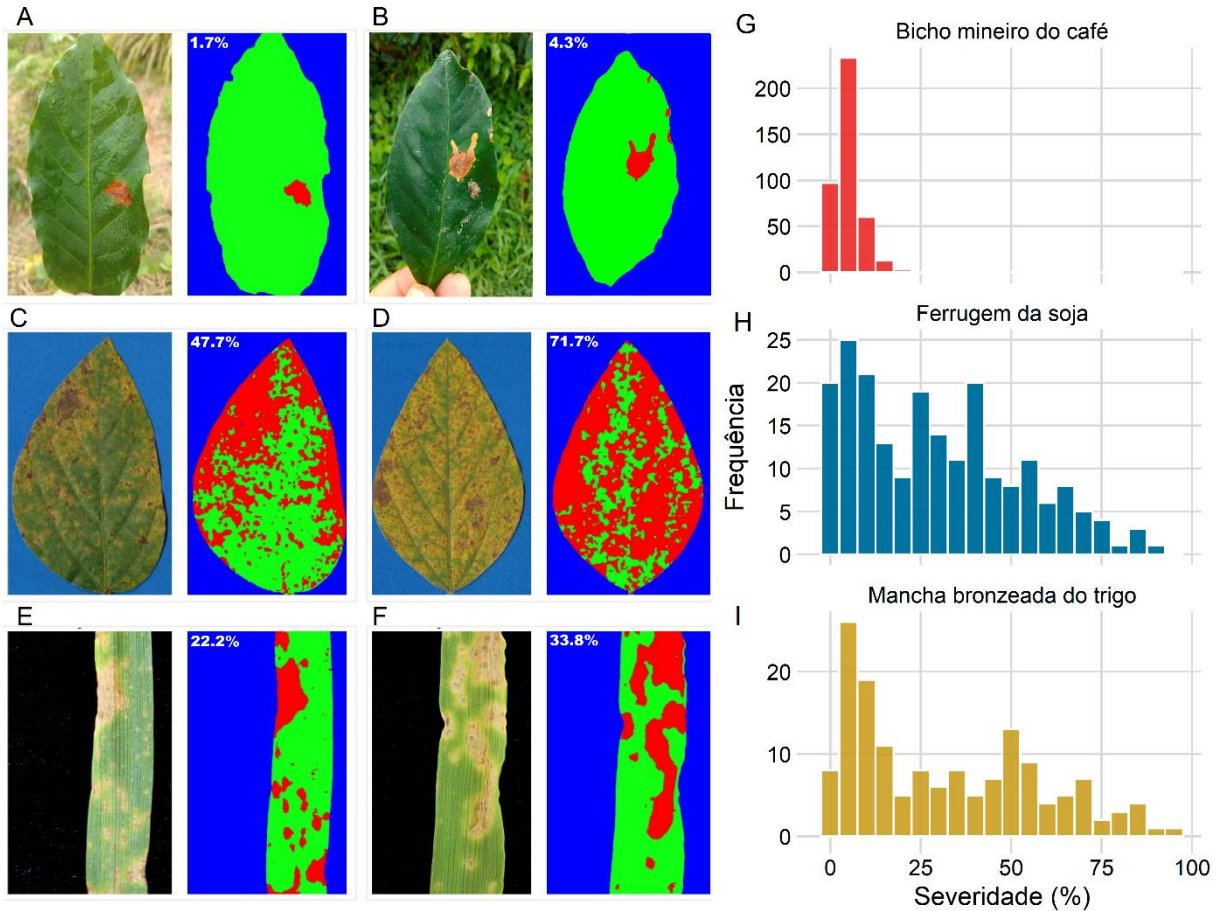


Figura 2. A. Imagens de exemplo de sintomas foliares causados pelo bicho-mineiro do café (A, B), ferrugem da soja (C, D) e mancha bronzeada do trigo (E, F) que foram anotadas individualmente para três classes de pixels em cada imagem: fundo (azul), folha saudável (verde) e folha sintomática necrótico (vermelho). Distribuição da severidade de referência para o conjunto de dados de café, (G, $n = 406$), soja (H, $n = 208$) e trigo (I, $n = 152$).

$$Severidade = \frac{L}{L + S} \times 100 \quad (2)$$

em que,

L - número de pixels rotulados como lesão;

S - número de pixels rotulados como saudáveis.

2.2.3 Modelos de segmentação semântica de última geração

Para avaliação do desempenho dos modelos de CNN na segmentação semântica e na estimativa da severidade de doenças e pragas foliares, foram selecionados cinco modelos de CNNs de última geração seguindo critérios como relevância e disponibilidade. Os modelos selecionados foram o Unet (Ronneberger et al., 2015), Segnet (Badrinarayanan et al., 2017), PSPNet (Zhao et al., 2017), FPN (Lin et al., 2017) e DeepLabv3+ (Chen et al., 2018) com duas variantes.

2.2.4 Implementação dos modelos

Todos os treinamentos foram realizados pela plataforma *Google Colaboratory*, ou “Colab”. O Colab é um serviço de *notebook* Jupyter em nuvem, gratuito, com objetivo de disseminar o ensino e pesquisa em aprendizado de máquina. Os modelos foram implementados usando o Python 3.6 e, com exceção do Deeplabv3+, foram treinados usando o Keras (<https://keras.io/>), biblioteca construída sobre o Tensorflow para treinamento de modelos de aprendizado profundo. O modelo Deeplabv3+ foi treinado usando o Tensorflow (<https://www.tensorflow.org/>), biblioteca de código aberto usada por pesquisadores para treinar modelos de aprendizado profundo (Kamilaris e Prenafeta-Boldú, 2018). A máquina usada para era equipada com uma CPU (unidade central de processamento) Intel Core i3-4150 (3.50 GHz) com 4 GB RAM.

Para estimar os desvios entre o valor previsto e o valor de referência para cada pixel, adotamos a função de perda entropia cruzada, algoritmo para resolução de problemas de otimização combinatória (de Boer et al., 2005). A função de perda original era uma soma de entropia cruzada entre o valor previsto e o valor de referência para cada pixel. Testes preliminares mostraram que os modelos de segmentação acabavam aprendendo os padrões das classes fundo e saudável, mas não aprendia os padrões da classe lesão. Portanto, foi introduzido um fator de ponderação γ na função de perda, a fim de forçar o modelo a ser sensível aos pixels pertencentes aos sintomas da doença. A função de perda modificada foi definida conforme Equação 2.

$$L(g_{ij}, \gamma_j, p_{ij}) = -\frac{1}{n} \sum_{i=0}^n \sum_{j=0}^m \gamma_j g_{ij} \log p_{ij}, \quad (3)$$

em que,

$L(g_{ij}, \gamma_j, p_{ij})$ – distância entre duas distribuições de probabilidade g (valor de referência) e p (valor de predição), adimensional;

g_{ij} - denota o valor de referência na i -ésima instância (pixel) pertencente à classe j ;

γ_j - fator de ponderação pertencente à classe j ;

p_{ij} - probabilidade prevista da i -ésima instância (pixel) pertencente à classe j , obtida com a função *softmax*;

n - número de pixels para cada tamanho de lote adotado;

m – número de classes.

O fator γ_j adotado para a classe lesão foi igual a 3 e para as classes fundo e saudável foi igual a 1. Para atualizar os pesos da rede foi utilizado o método de descida de gradiente estocástico (*Stochastic Gradient Descent* - SGD), algoritmo de otimização comum em trabalhos de aprendizagem profunda (Equação 4).

$$w_{t+1} = w_t - \alpha \nabla L(w_t), \quad (4)$$

em que,

w_t - pesos na iteração t ;

α - taxa de aprendizagem.

2.2.5 Inicialização

Se os parâmetros de inicialização da rede não forem adequados, a rede pode convergir para regiões de mínimo local, ou seja, apresentar soluções locais. Treinar redes com inicialização aleatória dos pesos exigem maior quantidades de dados e tempo computacionais extensivos. Alguns pesquisadores utilizam os pesos de redes pré-treinadas para outra tarefa ou de mesmo domínio para inicializar o processo de treinamento, trazendo ganhos significativos de desempenho aos modelos (Mohanty et al., 2016; Wang et al., 2017). Este processo é definido como aprendizado por transferência.

Neste estudo, foi adotado o aprendizado por transferência em todos os modelos, com exceção do SegNet. Os modelos de CNN de classificação a nível de pixel Unet, PSPNet, FPN e Deeplabv3+ adaptam modelos de CNNs de classificação a nível de imagem à sua arquitetura, que, compostos de camadas de convoluções e *max pooling*, são responsáveis por extrair informações de contexto e espaciais, embora informações espaciais em um nível menor. Neste caso, o modelo de CNN de classificação a nível de imagem recebe a denominação de “*backbone*”. Os modelos Unet, PSPNet e FPN foram treinados usando como *backbone* o modelo Inception-V3 (Szegedy et al., 2016), que iniciou o processo de treinamento com os pesos ajustados no conjunto de dados ImageNet (Russakovsky et al., 2015). O modelo DeepLabv3+ foi treinado testando duas variantes de *backbone*, denominadas Xception (Chollet, 2017) e MobilenetV2 (Sandler et al., 2018), que inicializaram os treinamentos com os pesos ajustados no conjunto de dados Pascal VOC 2012 (Everingham et al., 2014).

2.2.6 Pré-processamento e aumento de dados

Devido às limitações de memória do sistema utilizado e o tempo de conexão dos

notebooks do Colab (vida útil máxima de até 12 horas), os treinamentos dos modelos de CNN podem ser inviabilizados pela dimensão de cada imagem no banco de dados. Portanto, cada imagem e sua anotação de referência foram redimensionados para $768 \times 480 \times 3$ pixels (altura \times largura \times profundidade) antes do treinamento, a fim de reduzir a exigência computacional e diminuir o tempo de treinamento. Esta resolução ($768 \times 480 \times 3$ pixels) atendeu o padrão de entrada de todos os modelos.

Para aumentar a capacidade de generalização e impedir o super ajuste (*overfitting*) dos modelos, as imagens foram aumentadas com rotações ($90, 180$ e 270°), espelhamento (vertical e horizontal) e diminuição ou aumento aleatório do brilho (Liu et al., 2017). Como resultado, o total de imagens para treinamento era aumentado seis vezes. Além disso, durante o processo de treinamento, a função *ImageDataGenerator* do Keras foi utilizada para introduzir aleatoriamente rotações (entre 0 e 38°), deslocamento (entre 0 e $0,2$ unidades de altura e largura das imagens) e *zoom* (entre 0 e $0,2$).

2.2.7 Validação cruzada

O conjunto de dados deste estudo foi dividido em dois subconjuntos, com 80% das imagens usadas para treinamento (612 imagens) e 20% (154 imagens) para teste, mantendo-se esta proporção em cada doença/hospedeiro analisado (Tabela 1). O conjunto de dados de treinamento foi utilizado para pesquisa dos hiperparâmetros, enquanto o conjunto de dados de teste foi utilizado para estimar o desempenho dos modelos e comparar com o método referência na quantificação de doenças foliares. A estratégia adotada para o treinamento e busca dos parâmetros ideais foi a validação cruzada *k-fold*, com $k = 5$. Os resultados obtidos sobre o conjunto de teste foram apresentados como média.

Os hiperparâmetros ajustados para o treinamento dos modelos foram os seguintes: taxa de aprendizado fixada de $0,045$, momento de $0,9$ e o tamanho de lote de 2 , limitação computacional imposta pela arquitetura mais profunda, neste estudo, o Deeplabv3+ com o *backbone* Xception. Todos os modelos de CNN concluíram o processo de convergência em 10 épocas, com alguns modelos convergindo com aproximadamente 5 épocas.

2.2.8 Métricas de avaliação dos modelos

Para avaliação e comparação do desempenho dos modelos, foram empregados três métricas: Precisão (P), Recordação (R) e Interseção-sobre-União (IoU), também conhecida como índice de Jaccard. As métricas adotadas foram calculadas conforme Equações 5, 6 e 7, respectivamente.

$$P = \frac{TP}{TP + FP} * 100 \quad (5)$$

$$R = \frac{TP}{TP + FN} * 100 \quad (6)$$

$$IoU = \frac{TP}{TP + FP + FN} * 100 \quad (7)$$

em que,

TP - número de verdadeiros positivos;

FP - número de falsos positivos;

FN - número de falsos negativos.

As métricas P, R e IoU variam de 0 a 1, sendo que 0 significa inexistência de sobreposição das classes semânticas entre a máscara de referência e a máscara predita e 1 significa sobreposição perfeita. As métricas foram calculadas para cada classe e a média geral calculada a partir do valor de cada classe.

2.2.9 Comparação com o método referência na estimativa da severidade

As predições a nível de pixels pelos modelos de CNN foram transformadas em máscaras de segmentação para estimar a severidade. Para fins de comparação, foi usado o modo manual do *software* Assess versão 2.0 para estimar a severidade das 155 imagens do conjunto de teste. Os resultados da estimativa da severidade dos modelos de CNNs e do Assess foram comparados com a severidade obtida sobre as anotações de referências (Seção 2.2.2), consideradas o “padrão ouro”. Devido à baixa precisão do Assess em segmentar imagens foliares com o fundo complexo, este foi removido nas imagens de campo.

A qualidade das estimativas de severidade obtidas pelos modelos de CNN e pelo Assess foi avaliada com base na análise do coeficiente de correlação de concordância de Lin (*Lin's Concordance Correlation Coefficient* – LCCC) (Lin, 1989). Este é um dos mais populares coeficiente para avaliar a acurácia geral entre estimativas de severidade pelo método visual e por meio de *softwares* em estudos fitopatométricos (Bock et al., 2016; Madden et al., 2007). O LCCC permite avaliar o grau em que as estimativas caíram na linha de concordância (45 °, coeficiente angular = 1 e interceptação = 0). A acuraria geral (ρ_c) é um produto de precisão (r de *Pearson*) e viés geral (C_b), calculado a partir do viés sistemático (v) e do viés constante (μ). O desvio de ρ_c , r , C_b e v de 1 e μ de zero indica perda de precisão e perda de concordância. As análises das estatísticas do LCCC foram realizadas no ambiente estatístico R usando a função

epi.ccc do pacote irr (Gamer et al., 2012). Além disso, a raiz do erro médio quadrático (*Root Mean Square Error* - RMSE) também foi calculada (Wallach e Goffinet, 1989).

Para avaliar a rapidez das estimativas, foram contabilizados os tempos para obtenção do valor de severidade por imagem. O tempo necessário para estimar a severidade pelos modelos de CNN englobou o tempo de leitura das imagens, redimensionamento, inferência do modelo e transformação das etiquetas semânticas para gerar a máscara de segmentação. O tempo de inferência foi obtido utilizando a o Colab e a CPU. Para CPU foram realizadas dez inferências para calcular o tempo médio de estimativa por imagem. O tempo necessário para estimar severidade pelo Assess foi obtido pela soma dos tempos de leitura das imagens e seleção dos parâmetros de segmentação.

2.3 RESULTADOS

2.3.1 Comparação de desempenho de modelos de CNN

Conforme descrito na Seção 2.2.4, foi introduzido o hiperparâmetro λ devido ao desequilíbrio do número de pixels entre as classes. A classe lesão possui menos exemplos que as demais e, portanto, existe uma tendência natural dos modelos em classificar os pixels em função das classes dominantes. Nestes casos, o desequilíbrio favorece os falsos negativos em classes com poucos exemplos positivos. Portanto, sem o fator de aumento de sensibilidade λ para classe lesão, foi notado, nos testes preliminares, alta precisão (P) e baixa recordação (R) na classe lesão. Entretanto, podemos observar que o hiperparâmetro λ adotado conseguiu equilibrar os valores de P e R. Para quatro modelos a diferença entre P e R não ultrapassou 4%. O modelo com maior desequilíbrio foi o PSPNet, com P e R de 79,4 e 87,3%, respectivamente (Figura 3).

De forma geral, os modelos de CNN tiveram um bom desempenho na tarefa, alcançando uma IoU nas classes semânticas fundo (F), saudável (S) e lesão (L) de 94,4, 87,5 e 64,7%, respectivamente. O modelo FPN tendeu a apresentar o melhor desempenho, com IoU nas classes F, S e L de 98,9, 94,1 e 79,4%, respectivamente. Ele tendeu a superar o modelo Unet nas classes S e L em 0,2 e 1,5%, respectivamente. Embora o Deeplabv3+ com o *backbone* Xception seja o melhor algoritmo testado no conjunto de dados PASCAL VOC 2012, este tendeu a ser superado pela FPN em todas classes semânticas. Os modelos que tenderam a apresentar pior desempenho, no geral, foram o PSPNet, SegNet e o Deeplabv3+ com o *backbone* Mobilenetv2 (Figura 3).

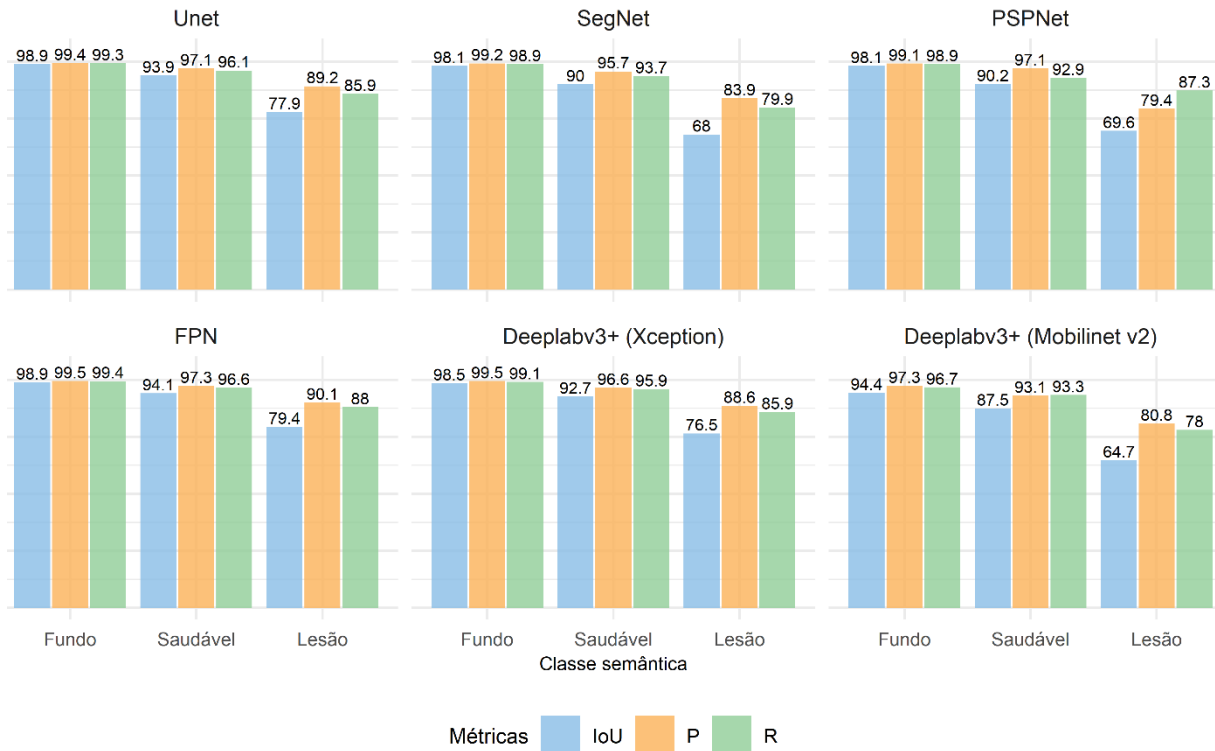


Figura 3. Desempenho dos modelos de CNN na segmentação semântica das imagens do conjunto de teste em três classes semânticas: plano de fundo da imagem, tecido foliar saudável e tecido foliar lesionado. Precisão (P); recordação (R); e interseção sobre união (IoU).

Na Figura 4 são apresentados os resultados da segmentação semântica realizada pelos modelos de CNNs no conjunto de dados do teste. Observa-se que as máscaras resultantes do processo de segmentação dos modelos Unet, FPN e Deeplabv3+ com o backbone Xception são altamente semelhantes as máscaras de referência. No entanto, em algumas imagens, os modelos PSPNet, SegNet e Deeplabv3+ com o backbone Mobilenetv2 apresentaram maior dificuldade em segmentar a folha. Nota-se na terceira e quarta linha da imagem que tais modelos não conseguiram segmentar uma parte da folha, considerada saudável, anotando tais pixels como fundo da imagem.

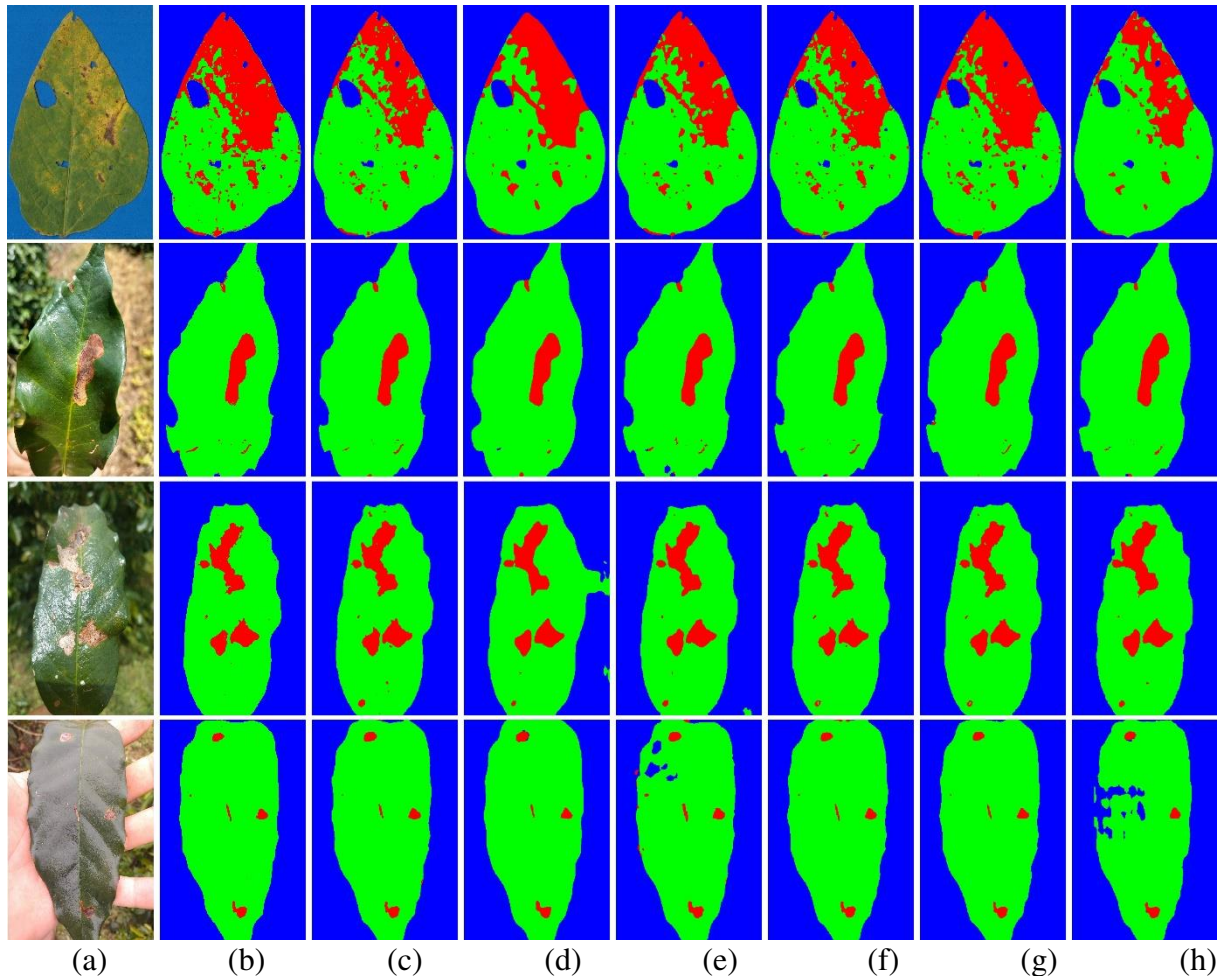


Figura 4. Comparação visual dos resultados da segmentação semântica em imagens de teste. (a) Imagem original; (b) anotação de referência; (c) Unet; (d) SegNet; (e) PSPNet; (f) FPN; (g) Deeplabv3+ com o backbone Xception; e (h) Deeplabv3+ com o backbone Mobilenetv2.

2.3.2 Severidade estimada pelos modelos de CNN

As máscaras obtidas do processo de predição em nível de pixel pelos modelos de CNN foram usadas para estimar a severidade (Equação 2). A análise dos resultados obtidos foi feita separadamente para cada uma das doenças/praga presente no conjunto de imagens de teste, o que permitiu entender melhor a fonte dos erros e comparar entre os modelos da CNN. Como esperado, considerando que os resultados da métrica IoU tendeu a se aproximar de 1, todos os modelos foram capazes de fornecer boas estimativas de severidade para a maioria das combinações de modelo de CNN *versus* imagem. Entretanto, diferenças no desempenho entre os modelos de CNN e doenças/pragas foram observadas e, portanto, foram explicados a seguir separadamente para cada doença/praga.

Os coeficientes de concordância de Lin (ρ_c) para o bicho-mineiro do café variou de 0,82 (PSPNet) a 0,98 (Unet e FPN) (Figura 5). Em todos os modelos, a precisão (r) foi menor que o viés generalizado (C_b), o que levou a uma redução de seu produto, ρ_c . Esse foi o caso das

previsões do modelo Segnet, PSPNet e DeepLabv3+ (MobileNetV2), em que a presença de erros relativamente maiores causou uma diminuição na precisão e exatidão.

Os modelos PSPNet e o DeepLabv3+ (MobileNetV2) apresentaram tendência a superestimar a severidade na maioria dos erros, principalmente no modelo PSPNet que teve o pior desempenho (Figura 5). A superestimação da severidade foi sugerida por valores de mudança de escala (v) superiores a 1,0: 1,27 para PSPNet e 1,08 para DeepLabv3+ (MobileNetV2). Esses modelos tenderam a classificar erroneamente os tecidos saudáveis como sintomáticos e mostraram os maiores valores de RMSE.

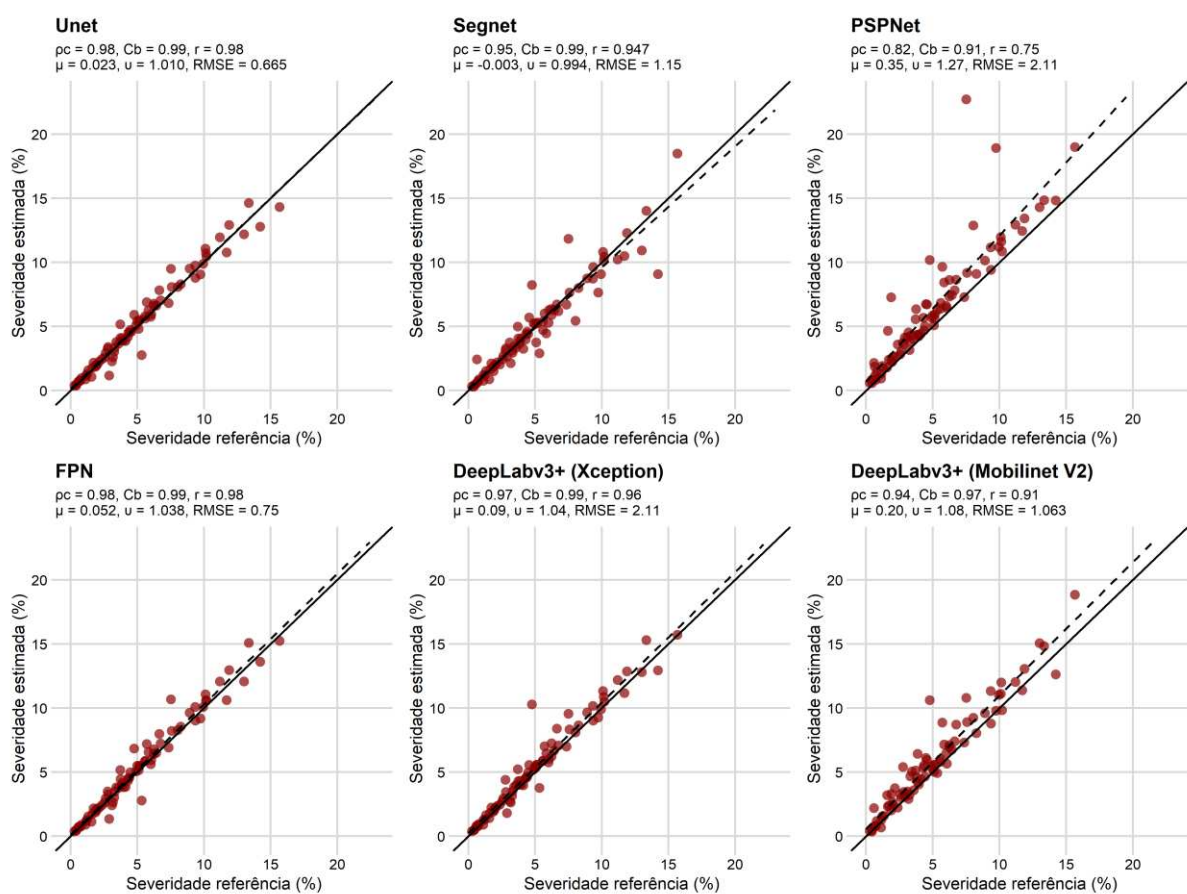


Figura 5. Gráfico de dispersão da linha de regressão de primeira ordem mais bem ajustada (linha tracejada) para a relação entre a severidade estimada (máscara resultante da segmentação dos modelos de CNN) e a severidade referência (obtidas sobre as anotações de referência). A linha sólida representa a linha de concordância perfeita. As estatísticas representam o coeficiente de concordância de Lin (pc) e os respectivos intervalos de confiança de 95%, dados pelo produto da precisão (r) e do coeficiente de viés (Cb). O último é calculado a partir das medidas de mudança de localização (μ) e mudança de escala (v). RMSE é o erro quadrático médio entre as duas séries.

Os valores de concordância de Lin (pc) para ferrugem da soja foram geralmente altos e relativamente próximos, variando de 0,93 a 0,97 (Figura 6). Novamente, os modelos Unet e FPN mostraram previsões em maior concordância com os valores de referência. Da mesma

forma, para as imagens do café, os valores de precisão (r) foram geralmente inferiores ao viés generalizado (C_b) para todos os modelos. Os valores de mudança de localização (μ) foram geralmente mais próximos de zero, sugerindo que não há viés constante para todos, exceto PSPNet e Deeplabv3+ (MobilinetV2).

Os modelos PSPNet, Deeplabv3+ (MobilinetV2) e SegNet mostraram a presença de vieses sistemáticos devido aos maiores valores de mudanças de escala ($v > 1$), o que indica que a severidade é cada vez mais superestimada à medida que a severidade aumenta nas imagens (Figura 6). De fato, isso foi observado na maioria dos modelos para os quais o conjunto de 10 imagens com severidade maior que 50% teve sua severidade prevista consistentemente acima do valor de referência, com poucas exceções. Esses modelos apresentaram os menores valores de precisão ($r < 0,91$) e maiores de RMSE (> 8).

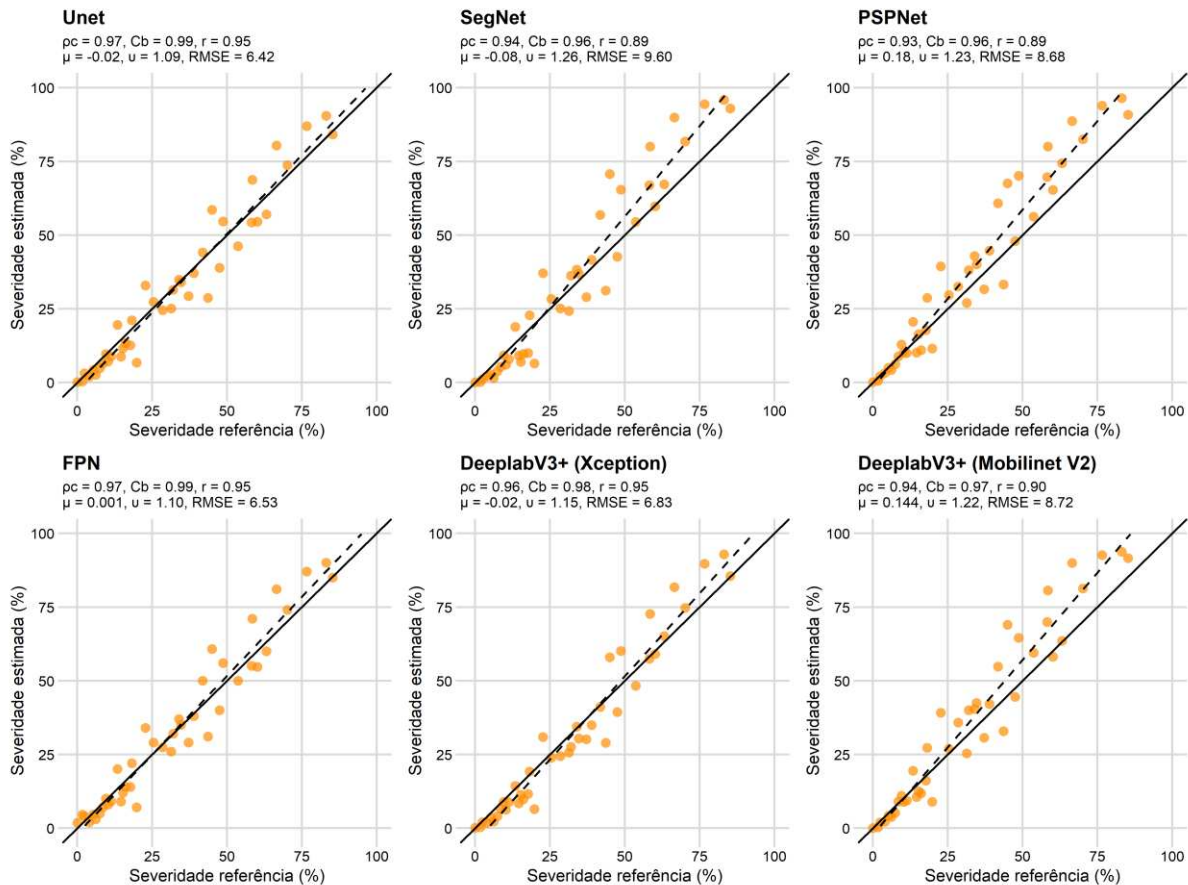


Figura 6. Gráfico de dispersão da linha de regressão de primeira ordem mais bem ajustada (linha tracejada) para a relação entre a severidade estimada (máscara resultante da segmentação dos modelos de CNN) e a severidade referência (obtidas sobre as anotações de referência). A linha sólida representa a linha de concordância perfeita. As estatísticas representam o coeficiente de correlação de concordância de Lin (ρ_c) e os respectivos intervalos de confiança de 95%, dados pelo produto da precisão (r) e do coeficiente de viés (C_b). O último é calculado a partir das medidas de mudança de localização (μ) e mudança de escala (v). RMSE é o erro quadrático médio entre as duas séries.

Os valores de concordância de Lin (ρ_c) para mancha bronzeada do trigo foram, em média, relativamente superiores aos calculados para as imagens de teste de café e soja (Figura 7). Para apenas um modelo (DeepLabV3+ MobilineV2), o coeficiente de concordância foi menor que 0,98. Ao contrário das outras duas imagens de corte, o modelo com pior desempenho tendeu a subestimar a gravidade para a grande maioria das imagens, o que foi evidenciado pelos valores negativos de deslocamento de localização e de escala superiores a 1,0, sugerindo vieses constantes e sistemáticos. Para os outros modelos, houve uma leve tendência a subestimar com mais frequência a doença, mas com erros geralmente pequenos e boa precisão.

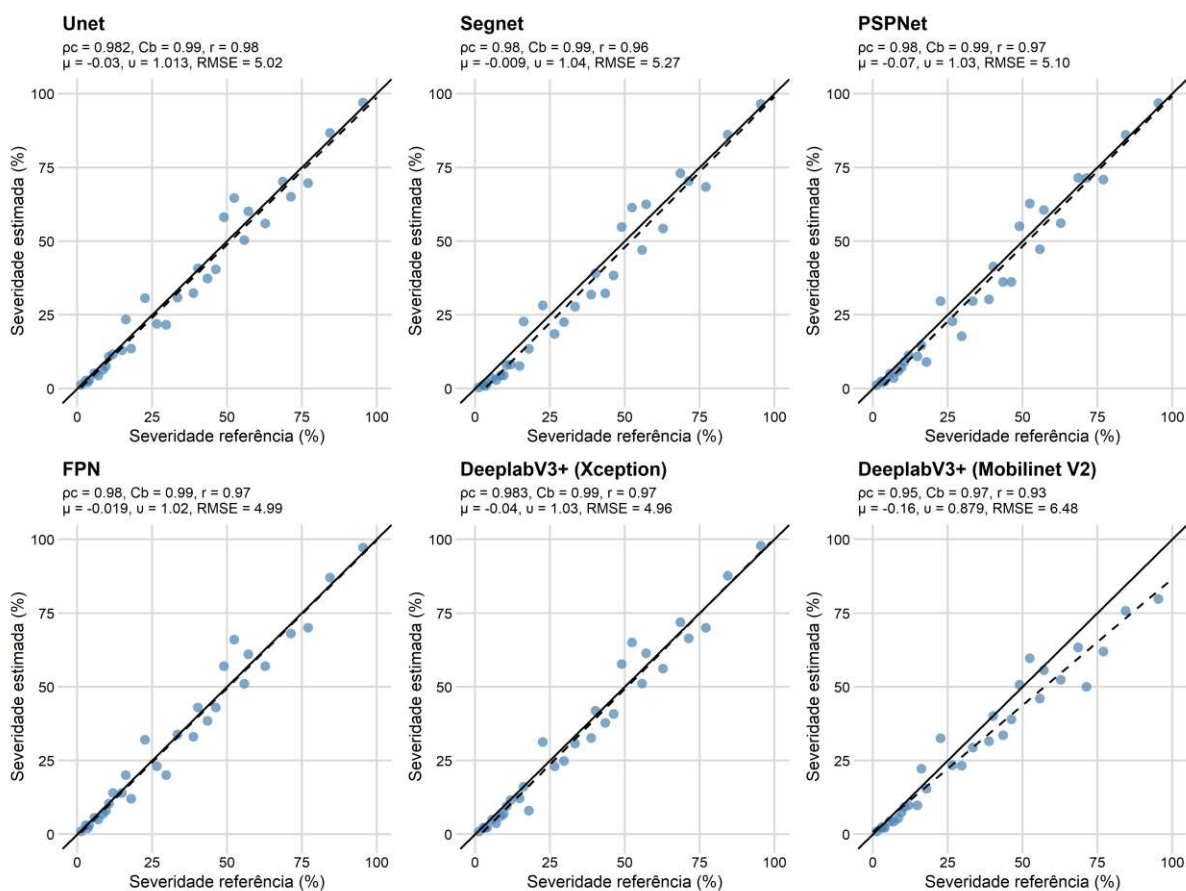


Figura 7. Gráfico de dispersão da linha de regressão de primeira ordem mais bem ajustada (linha tracejada) para a relação entre a severidade estimada (máscara resultante da segmentação dos modelos de CNN) e a severidade referência (obtidas sobre as anotações de referência). A linha sólida representa a linha de concordância perfeita. As estatísticas representam o coeficiente de correlação de concordância de Lin (ρ_c) e os respectivos intervalos de confiança de 95%, dados pelo produto da precisão (r) e do coeficiente de viés (C_b). O último é calculado a partir das medidas de mudança de localização (μ) e mudança de escala (ν). RMSE é o erro quadrático médio entre as duas séries.

2.3.3 Severidade estimada pelo *software* Assess

Os valores de concordância de Lin (ρ_c) obtidos usando o método de limiar de segmentação ajustado manualmente no *software* Assess foi maior ou igual a 0,94 para os três

conjuntos de imagens de culturas (Figura 8A-C). Semelhante à severidade prevista pelos modelos de CNNs, a precisão foi geralmente menor do que o viés generalizado. Apenas quatro folhas de café apresentaram severidade superestimada com erros maiores que 4%. O valor da mudança de local maior que zero sugeria um viés constante devido à superestimação mais frequente em toda a faixa de severidade (Figura 8A). Resultados semelhantes foram encontrados para a ferrugem da soja, cuja concordância relativamente menor foi devido à superestimação devido ao viés constante (Figura 8B). Semelhante a uma parte da gravidade prevista pelo modelo, a mancha marrom de trigo foi subestimada de forma mais consistente com menor precisão. O viés mais baixo, no entanto, não afetou a concordância bastante alta ($\rho_c = 0,97$) (Figura 8C).

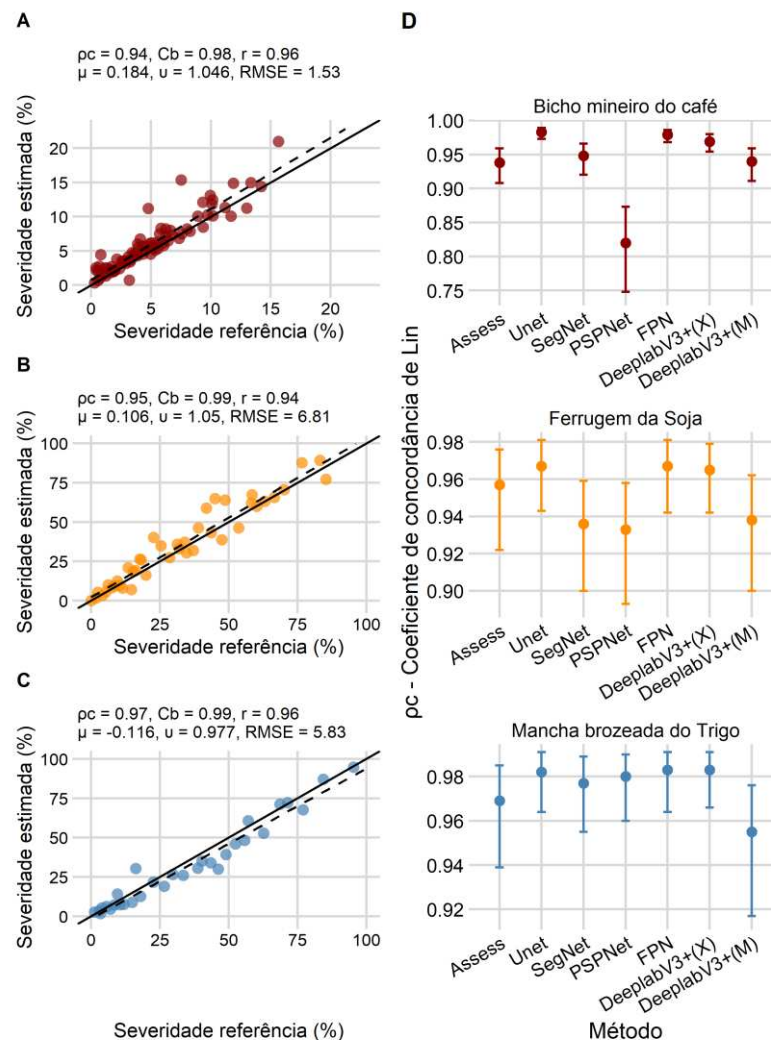


Figura 8. Gráfico de dispersão da linha de regressão de primeira ordem mais bem ajustada (linha tracejada) para a relação entre a severidade estimada (*software* Assess) e a severidade referência (obtidas sobre as anotações de referência) para o bicho-mineiro do café (A), ferrugem da soja (B) e mancha bronzeada do trigo (C). As estatísticas representam o coeficiente de correlação de concordância de Lin (ρ_c) e os respectivos intervalos de confiança de 95%, dados pelo produto da precisão (r) e do coeficiente

de viés (Cb). O último é calculado a partir das medidas de mudança de localização (μ) e mudança de escala (ν). RMSE é o erro quadrático médio entre as duas séries.

Os intervalos de confiança de 95% para os coeficientes de concordância sugerem que todos os seis modelos de CNNs estimaram valores de severidade semelhantes às obtidas pelo Assess. No entanto, em alguns casos, como no bicho-mineiro, os modelos FPN, Unet e DeepLabv3+ (Xception) foram superiores à avaliação do Assess. Os modelos Deeplabv3+ (MobilenetV2) e SegNet apresentaram desempenho semelhante ao Assess para o bicho-mineiro, e o modelo PSPNet apresentou o pior desempenho.

Os valores médios relativamente mais altos de concordância e sobreposição de intervalos de confiança sugerem que todos modelos podem ser usados para prever a severidade da ferrugem da soja e da mancha do trigo. No entanto, estimativas mais precisas, ou mais próximas aos valores de referência, foram obtidas ao usar os mesmos três modelos que melhor previram a severidade do bicho-mineiro: FPN, Unet e DeepLabv3 + (Xception) (Figura 8D).

2.3.4 Tempo de inferência

Na Tabela 2 é apresentado o tempo médio para estimativa da severidade por imagem, calculado sobre as imagens do conjunto de teste. Todos os modelos de CNN foram treinados nas mesmas condições, ou seja, usando a mesma dimensão de imagem, de otimizador, de taxa de aprendizado e de tamanho de lote. Portanto, o fator que pode explicar a variabilidade no tempo médio gasto para estimar a severidade entre os modelos de CNN é o número de parâmetros dos modelos. Os modelos com menor número de parâmetros, como o PSPNet, SegNet e o Deeplabv3+ com o *backbone* Mobilenetv2, tenderam a gastar menos tempo por imagem para estimar a severidade. O modelo com pior desempenho foi o Deeplabv3+ (Xception), entretanto foi aproximadamente 20x inferior ao operador do *software* Assess.

Tabela 2. Número de parâmetros dos modelos de CNN e o tempo médio para estimar a porcentagem de severidade por imagem do conjunto de imagens de teste.

Método	Número de parâmetros (Milhões)	Tempo de inferência Colab (s)	Tempo de inferência CPU (s)
Unet	29,6	1,94	4,79
Segnet	7,8	1,75	3,37
PSPNet	2,8	1,87	3,14
FPN	25,0	1,94	4,94
Deeplabv3+ (Xception)	41,3	1,94	5,14
Deeplabv3+ (Mobilenetv2)	2,1	1,75	3,03
Assess	-	-	105,09

2.4 DISCUSSÃO

Até o momento, o método de segmentação de imagens digitais mais utilizado em *softwares* e algoritmos propostos para estimativa da severidade ainda é o limiar de segmentação. Este método utiliza somente a informação de cor para realização da segmentação e, portanto, apresenta bom desempenho somente se existir um contraste óbvio entre duas regiões a ser segmentada. Barbedo (2016), em sua revisão, apresenta as soluções adotadas por diversos estudos para promover o contraste em imagens foliares: captura das imagens das folhas sobre um fundo padronizado (azul ou preto), utilização de um digitalizador de mesa, remoção manual do fundo da imagem e alteração do ângulo de captura das imagens para redução do reflexo na superfície foliar. Esses tipos de soluções podem aumentar a exatidão e a precisão, mas também podem desacelerar significativamente os processos de aquisição de imagens e estimativa de severidade.

Recentemente, o aprendizado profundo e, mais especificamente a segmentação semântica, vem sendo aplicado com objetivo de desenvolver sistemas rápidos, automáticos e precisos na segmentação de imagens. Em sistemas automáticos não é necessário nenhum ajuste prévio de parâmetro no momento da segmentação, deixando o processo rápido e independente da experiência do usuário. A limitação atual do emprego dos modelos de CNN é a falta de dados anotados adequadamente para treinamento dos modelos. Embora exista uma grande quantidade de imagens capturadas em uma gama de condições disponíveis na literatura, o processo de anotação das imagens pode ser extremamente demorado e difícil, mesmo com o desenvolvimento de algoritmos para auxiliar na tarefa. Até onde se sabe, este é o primeiro conjunto de dados disponibilizados para a comunidade científica para treinamento de modelos de CNN para segmentação semântica aplicados na estimativa da severidade de doenças foliares, disponível em <https://osf.io/hx2um/>.

As imagens deste estudo foram capturadas estrategicamente para apresentar uma variedade de condições e ambientes em que agricultores ou especialistas avaliam a severidade. Em condição controlada, as folhas foram coletadas e fotografadas ou escaneadas em ambiente com iluminação homogênea e o fundo padronizado na cor azul e preto, condição em que grande parte dos métodos são validados. Em condição de campo, as folhas foram coletadas e fotografadas estrategicamente para expor as imagens às condições mais desafiadoras para os métodos de segmentação. Barbedo (2016b), em sua revisão, discute três fatores extrínsecos que afetam o desempenho dos algoritmos de segmentação: fundo complexo, luz especular (reflexo de alta intensidade que ocorre em certos ângulos de visão) e a presença simultânea de sombra

e iluminação direta. No total, 766 imagens foram capturadas em condições de laboratório e campo e gerada a anotação de referência. Apesar de apresentar um número reduzido de imagens, culturas e doenças, se comparado a bancos de dados públicos como o *Digipathos* (<https://www.digipathos-rep.cnptia.embrapa.br/>) e *PlantVillage* (Hughes e Salathe, 2015), se mostrou suficientemente robusto para inferir o potencial de segmentação dos modelos de CNN na segmentação nas mais variadas condições.

O modelo FPN alcançou o melhor desempenho na pontuação IoU entre os modelos analisados. O modelo Unet, segundo colocado, utiliza um modelo de rede totalmente convolucional, composta basicamente de duas etapas: a contração e a expansão. Na seção de contração, a saída do mapa de características pela camada convolucional se torna menor à medida que a rede se aprofunda, acarretando perdas de informações de pequenos alvos. Entretanto, a FPN propõe um descendente e uma estrutura conectada horizontalmente que combina baixa resolução, recursos semânticos fortes com alta resolução e recursos semânticos fracos para detecção e segmentação de pequenos alvos. Portanto, a variabilidade do tamanho dos objetos de classe no conjunto de dados de teste, principalmente a lesão foliar, pode explicar a tendência do melhor desempenho da FPN frente a Unet.

Embora o Deeplabv3+ com o backbone Xception seja o melhor algoritmo testado no conjunto de dados PASCAL VOC 2012 (Everingham et al., 2014), este tendeu a ser superado pela FPN na pontuação IoU. Um dos problemas do Deeplabv3+ com o backbone Xception são as convoluções dilatadas. Elas aumentam o campo receptivo (informações de contexto), mas diminuem as informações de bordas, reduzindo seu desempenho em pequenos objetos de classe (Song Guo et al., 2019; S. Zhang et al., 2020). Além disso, o Deeplabv3+ com o backbone Xception é o algoritmo com maior número de parâmetros (Tabela 2). Como explicado por Lobo Torres et al. (2020), quanto maior o número de parâmetros no modelo maior a exigência na quantidade de amostras de treinamento. Portanto, o número de amostras de treinamento deste trabalho pode não ter sido suficiente para que o DeepLabv3+ com o backbone Xception expressasse todo o seu potencial. De fato, a quantidade de imagens utilizadas no treinamento do DeepLabv3+ foi de 10.582 (Chen et al., 2018), cerca de 20 vezes maior que as imagens de treinamento deste estudo.

Os experimentais computacionais mostraram que os modelos PSPNet, SegNet e o Deeplabv3+ com o *backbone* MobilenetV2 tenderam a apresentar pior desempenho. Estes modelos apresentam um número reduzido de parâmetros se comparado aos demais (Tabela 2). O conjunto de dados de imagem neste estudo é desafiador devido a grandes variações de textura, forma, cor e escala, exigindo assim uma maior quantidade de parâmetros nos modelos

para segmentar as imagens. O Deeplabv3+ com o *backbone* Mobilenetv2, por exemplo, em termos de parâmetros, é cerca de 20 vezes menor que o Deeplabv3+ com o *backbone* Xception. Portanto, o número reduzido de parâmetros destes modelos pode ter comprometido a sua capacidade de segmentação. Além disso, o modelo SegNet não utiliza os pesos pré-treinados do VGG 16 para inicializar a rede. Apesar deste fator não ter sido investigado neste estudo, alguns trabalhos mostraram que a inicialização da rede com os pesos pré-treinados em grandes conjuntos de imagens melhorou o desempenho dos modelos CNN a nível de classificação de imagens (Mohanty et al., 2016; Wang et al., 2017).

As estimativas de severidade devem fornecer resultados reproduzíveis (precisos) e exatos da estimativa da severidade. Alguns trabalhos mostraram que métodos manuais, como o Assess, tendem a apresentar melhor exatidão que algoritmos semiautomáticos e automáticos (Barbedo, 2016b), pois permite ao usuário ajustar os parâmetros de segmentação manualmente para obter o melhor resultado possível. Entretanto, os modelos de CNN são capazes de extrair e aprender recursos de contexto através das sucessivas convoluções e, portanto, segmentar imagens em condições variadas sem a necessidade de ajuste de parâmetros por parte do usuário. Na Seção 2.3.3 foi mostrado que, para cada doença, pelo menos um modelo de CNN superou o Assess. Para fins de comparação, o fundo das imagens foi removido para facilitar a análise do Assess e, mesmo assim, quatro modelos de CNN superaram o Assess em exatidão.

O tempo de análise de imagens digitais para estimativa da severidade pode se tornar um limitante em situações onde se exige um número elevado de medições. Em métodos manuais como o QUANT (Vale et al., 2003), Assess (Lamari, 2002) e o *Leaf Doctor* (Pethybridge e Nelson, 2015), o usuário pode ultrapassar facilmente 50s no processo de segmentação da imagem digital (Barbedo, 2016b; Pethybridge e Nelson, 2015). Entretanto, os resultados mostraram que os tempos da estimativa da severidade por imagem dos modelos de CNN foram bem inferiores aos tempos de *softwares* e algoritmos manuais de segmentação e, inclusive, aos métodos visuais. Além disso, os modelos de CNN aplicados na segmentação semântica são automáticos e, portanto, não necessitam de nenhuma experiência prévia do avaliador para segmentação das imagens.

Um aspecto importante deste trabalho que deve ser considerado é que as imagens de campo foram capturadas após o destacamento da folha, ou seja, o método aplicado foi destrutivo. Portanto, apesar de sofrerem com iluminação irregular e possuir o fundo complexo, não foram capturadas em condições reais de campo. Numa pesquisa futura, para segmentação de imagens para estimar a severidade em condições reais de campo deve ser introduzido o YOLO (*You Only Look Once*), método desenvolvido por Redmon et al. (2016), nos quais usam

"*bounding boxes*" para localizar os objetos de interesse em cena. Além disso, outro problema comum na fitopatologia, mas que não foi abordado neste estudo, é a segmentação de doenças de ocorrência simultânea na superfície foliar em classes semânticas distintas. De fato, essa é uma limitação que, até onde se sabe, não foi abordada em nenhum estudo de segmentação de imagens digitais.

2.5 CONCLUSÃO

Neste artigo, estudamos o problema estimativa da severidade de doenças e pragas foliares. Os modelos de CNN apresentaram grande potencial para segmentação semântica de imagens digitais de doenças foliares e seus hospedeiros para estimar a severidade de doenças. O modelo de Rede de Pirâmides Características (FPN) alcançou o melhor desempenho entre os modelos testados e superou o *software* Assess versão 2.0 em precisão e exatidão na estimativa da severidade. O tempo para estimar a severidade por imagem pelos modelos de CNNs aplicados na segmentação semântica foi, aproximadamente, 20 vezes inferior ao do *software* Assess.

2.6 REFERÊNCIAS

- Badrinarayanan, V., Kendall, A., Cipolla, R., 2017. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 2481–2495. <https://doi.org/10.1109/TPAMI.2016.2644615>
- Barbedo, J.G.A., 2018. Factors influencing the use of deep learning for plant disease recognition. *Biosyst. Eng.* 172, 84–91. <https://doi.org/10.1016/j.biosystemseng.2018.05.013>
- Barbedo, J.G.A., 2016a. A review on the main challenges in automatic plant disease identification based on visible range images. *Biosyst. Eng.* 144, 52–60. <https://doi.org/10.1016/j.biosystemseng.2016.01.017>
- Barbedo, J.G.A., 2016b. A novel algorithm for semi-automatic segmentation of plant leaf disease symptoms using digital image processing. *Trop. Plant Pathol.* 41, 210–224. <https://doi.org/10.1007/s40858-016-0090-8>
- Barbedo, J.G.A., 2014. An Automatic Method to Detect and Measure Leaf Disease Symptoms Using Digital Image Processing. *Plant Dis.* 98, 1709–1716. <https://doi.org/10.1094/PDIS-03-14-0290-RE>
- Bock, C.H., Barbedo, J.G.A., Del Ponte, E.M., Bohnenkamp, D., Mahlein, A.-K., 2020. From visual estimates to fully automated sensor-based measurements of plant disease severity: status and challenges for improving accuracy. *Phytopathol. Res.* 2, 9. <https://doi.org/10.1186/s42483-020-00049-8>
- Bock, C.H., Chiang, K.-S., Del Ponte, E.M., 2016. Accuracy of plant specimen disease severity estimates: concepts, history, methods, ramifications and challenges for the future. *CAB Rev. Perspect. Agric. Vet. Sci. Nutr. Nat. Resour.* 11. <https://doi.org/10.1079/PAVSNR201611032>
- Bock, C.H., Cook, A.Z., Parker, P.E., Gottwald, T.R., 2009. Automated Image Analysis of the Severity of Foliar Citrus Canker Symptoms. *Plant Dis.* 93, 660–665. <https://doi.org/10.1094/PDIS-93-6-0660>
- Bock, C.H., Poole, G.H., Parker, P.E., Gottwald, T.R., 2010. Plant Disease Severity Estimated Visually, by Digital Photography and Image Analysis, and by Hyperspectral Imaging. *CRC. Crit. Rev. Plant Sci.* 29, 59–107. <https://doi.org/10.1080/07352681003617285>
- Borne, L., Rivière, D., Mancip, M., Mangin, J.F., 2020. Automatic labeling of cortical sulci using patch- or CNN-based segmentation techniques combined with bottom-up geometric constraints. *Med. Image Anal.* 62, 101651. <https://doi.org/10.1016/j.media.2020.101651>
- Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation, in: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Springer Verlag, pp. 833–851. https://doi.org/10.1007/978-3-030-01234-2_49
- Chiang, K.-S., Liu, S.-C., Bock, C.H., Gottwald, T.R., 2014. What Interval Characteristics Make a Good Categorical Disease Assessment Scale? *Phytopathology* 104, 575–585. <https://doi.org/10.1094/PHYTO-10-13-0279-R>

- Chollet, F., 2017. Xception: Deep Learning with Depthwise Separable Convolutions, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp. 1800–1807. <https://doi.org/10.1109/CVPR.2017.195>
- Clément, A., Verfaille, T., Lormel, C., Jaloux, B., 2015. A new colour vision system to quantify automatically foliar discolouration caused by insect pests feeding on leaf cells. *Biosyst. Eng.* 133, 128–140. <https://doi.org/10.1016/j.biosystemseng.2015.03.007>
- Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., Schiele, B., R&d, D.A., Darmstadt, T.U., 2016. The Cityscapes Dataset for Semantic Urban Scene Understanding.
- de Boer, P.-T., Kroese, D.P., Mannor, S., Rubinstein, R.Y., 2005. A Tutorial on the Cross-Entropy Method. *Ann. Oper. Res.* 134, 19–67. <https://doi.org/10.1007/s10479-005-5724-z>
- Del Ponte, E.M., Pethybridge, S.J., Bock, C.H., Michereff, S.J., Machado, F.J., Spolti, P., 2017. Standard Area Diagrams for Aiding Severity Estimation: Scientometrics, Pathosystems, and Methodological Trends in the Last 25 Years. *Phytopathology* 107, PHYTO-02-17-006. <https://doi.org/10.1094/PHYTO-02-17-0069-FI>
- Du, Z., Yang, J., Ou, C., Zhang, T., 2019. Smallholder Crop Area Mapped with a Semantic Segmentation Deep Learning Method. *Remote Sens.* 11, 888. <https://doi.org/10.3390/rs11070888>
- Esgario, J.G.M., Krohling, R.A., Ventura, J.A., 2020. Deep learning for classification and severity estimation of coffee leaf biotic stress. *Comput. Electron. Agric.* 169, 105162. <https://doi.org/10.1016/j.compag.2019.105162>
- Everingham, M., Eslami, S.M.A., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A., 2014. The Pascal Visual Object Classes Challenge: A Retrospective. *Int. J. Comput. Vis.* 111, 98–136. <https://doi.org/10.1007/s11263-014-0733-5>
- Ferentinos, K.P., 2018. Deep learning models for plant disease detection and diagnosis. *Comput. Electron. Agric.* 145, 311–318. <https://doi.org/10.1016/j.compag.2018.01.009>
- Franceschi, V.T., Alves, K.S., Mazaro, S.M., Godoy, C. V., Duarte, H.S.S., Del Ponte, E.M., 2020. A new standard area diagram set for assessment of severity of soybean rust improves accuracy of estimates and optimizes resource use. *Plant Pathol.* 69, 495–505. <https://doi.org/10.1111/ppa.13148>
- Gamer, M., Lemon, J., Fellows, I., Singh, P., 2012. Various coefficients of interrater reliability and agreement. R package version 0.84. <https://cran.r-project.org/package=irr>.
- Guo, Shichen, Jin, Q., Wang, H., Wang, X., Wang, Y., Xiang, S., 2019. Learnable Gated Convolutional Neural Network for Semantic Segmentation in Remote-Sensing Images. *Remote Sens.* 11, 1922. <https://doi.org/10.3390/rs11161922>
- Guo, Song, Li, T., Kang, H., Li, N., Zhang, Y., Wang, K., 2019. L-Seg: An end-to-end unified framework for multi-lesion segmentation of fundus images. *Neurocomputing* 349, 52–63. <https://doi.org/10.1016/J.NEUCOM.2019.04.019>
- Hughes, D.P., Salathe, M., 2015. An open access repository of images on plant health to enable

the development of mobile disease diagnostics. arXiv, 1511.08060.

Jothiaruna, N., Joseph Abraham Sundar, K., Karthikeyan, B., 2019. A segmentation method for disease spot images incorporating chrominance in Comprehensive Color Feature and Region Growing. *Comput. Electron. Agric.* 165, 104934. <https://doi.org/10.1016/j.compag.2019.104934>

Kamilaris, A., Prenafeta-Boldú, F.X., 2018. Deep learning in agriculture: A survey. *Comput. Electron. Agric.* 147, 70–90. <https://doi.org/10.1016/j.compag.2018.02.016>

Karisto, P., Hund, A., Yu, K., Anderegg, J., Walter, A., Mascher, F., McDonald, B.A., Mikaberidze, A., 2018. Ranking Quantitative Resistance to *Septoria tritici* Blotch in Elite Wheat Cultivars Using Automated Image Analysis. *Phytopathology* 108, 568–581. <https://doi.org/10.1094/PHYTO-04-17-0163-R>

Krestenitis, M., Orfanidis, G., Ioannidis, K., Avgerinakis, K., Vrochidis, S., Kompatsiaris, I., 2019. Oil Spill Identification from Satellite Images Using Deep Neural Networks. *Remote Sens.* 11, 1762. <https://doi.org/10.3390/rs11151762>

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet Classification with Deep Convolutional Neural Networks, in: *Advances in Neural Information Processing Systems*. pp. 1097–1105.

Lamari, L., 2002. *ASSESS: Image Analysis Software for Plant Disease Quantification*. American Phytopathological Society, St. Paul, Minnesota: APS Press.

Liang, Q., Xiang, S., Hu, Y., Coppola, G., Zhang, D., Sun, W., 2019. PD2SE-Net: Computer-assisted plant disease diagnosis and severity estimation network. *Comput. Electron. Agric.* 157, 518–529. <https://doi.org/10.1016/j.compag.2019.01.034>

Lin, K., Gong, L., Huang, Y., Liu, C., Pan, J., 2019. Deep Learning-Based Segmentation and Quantification of Cucumber Powdery Mildew Using Convolutional Neural Network. *Front. Plant Sci.* 10. <https://doi.org/10.3389/fpls.2019.00155>

Lin, L.I.-K., 1989. A Concordance Correlation Coefficient to Evaluate Reproducibility. *Biometrics* 45, 255. <https://doi.org/10.2307/2532051>

Lindow, S.E., Webb, R.R., 1983. Quantification of foliar plant disease symptoms by microcomputer-digitized video image analysis. *Phytopathology* 73, 520–524.

Liu, B., Zhang, Y., He, D., Li, Y., 2017. Identification of Apple Leaf Diseases Based on Deep Convolutional Neural Networks. *Symmetry (Basel)*. 10, 11. <https://doi.org/10.3390/sym10010011>

Lobo Torres, D., Queiroz Feitosa, R., Nigri Happ, P., Elena Cué La Rosa, L., Marcato Junior, J., Martins, J., Olã Bressan, P., Gonçalves, W.N., Liesenberg, V., 2020. Applying Fully Convolutional Architectures for Semantic Segmentation of a Single Tree Species in Urban Environment on High Resolution UAV Optical Imagery. *Sensors* 20, 563. <https://doi.org/10.3390/s20020563>

Long, J., Shelhamer, E., Darrell, T., 2015. Fully Convolutional Networks for Semantic Segmentation.

- Ma, J., Du, K., Zhang, L., Zheng, F., Chu, J., Sun, Z., 2017. A segmentation method for greenhouse vegetable foliar disease spots images using color information and region growing. *Comput. Electron. Agric.* 142, 110–117. <https://doi.org/10.1016/J.COMPAG.2017.08.023>
- Madden, L., Hughes, G., Van Den Bosch, F., 2007. *The study of plant disease epidemics*. APS Press, St Paul.
- Mohanty, S.P., Hughes, D.P., Salathé, M., 2016. Using Deep Learning for Image-Based Plant Disease Detection. *Front. Plant Sci.* 7, 1–10. <https://doi.org/10.3389/fpls.2016.01419>
- Nita, M., Ellis, M.A., Madden, L. V., 2003. Reliability and Accuracy of Visual Estimation of Phomopsis Leaf Blight of Strawberry. *Phytopathology* 93, 995–1005. <https://doi.org/10.1094/PHYTO.2003.93.8.995>
- Nutter, F.W., Gleason, M.L., Jenco, J.H., Christians, N.L., 1993. Assessing the Accuracy, Intra-rater Repeatability, and Inter-rater Reliability of Disease Assessment Systems. *Phytopathology* 83, 806–812. <https://doi.org/10.1094/Phyto-83-806>
- Oerke, E.-C., 2006. Crop losses to pests. *J. Agric. Sci.* 144, 31–43. <https://doi.org/10.1017/S0021859605005708>
- Pethybridge, S.J., Nelson, S.C., 2015. Leaf Doctor: A New Portable Application for Quantifying Plant Disease Severity. *Plant Dis.* 99, 1310–1316. <https://doi.org/10.1094/PDIS-03-15-0319-RE>
- R. Palafox, P., Betz, J., Nobis, F., Riedl, K., Lienkamp, M., 2019. SemanticDepth: Fusing Semantic Segmentation and Monocular Depth Estimation for Enabling Autonomous Driving in Roads without Lane Lines. *Sensors* 19, 3224. <https://doi.org/10.3390/s19143224>
- Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You Only Look Once: Unified, Real-Time Object Detection, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp. 779–788. <https://doi.org/10.1109/CVPR.2016.91>
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation, in: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Springer Verlag, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L., 2015. ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis.* 115, 211–252. <https://doi.org/10.1007/s11263-015-0816-y>
- Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C., 2018. MobileNetV2: Inverted Residuals and Linear Bottlenecks, in: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society, pp. 4510–4520. <https://doi.org/10.1109/CVPR.2018.00474>
- Savary, S., Bregaglio, S., Willocquet, L., Gustafson, D., Mason D’Croz, D., Sparks, A., Castilla, N., Djurle, A., Allinne, C., Sharma, M., Rossi, V., Amorim, L., Bergamin, A., Yuen, J., Esker, P., McRoberts, N., Avelino, J., Duveiller, E., Koo, J., Garrett, K., 2017. Crop health

and its global impacts on the components of food security. *Food Secur.* 9, 311–327. <https://doi.org/10.1007/s12571-017-0659-1>

Simonyan, K., Zisserman, A., 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition.

Škaloudová, B., Křivan, V., Zemek, R., 2006. Computer-assisted estimation of leaf damage caused by spider mites. *Comput. Electron. Agric.* 53, 81–91. <https://doi.org/10.1016/j.compag.2006.04.002>

Stewart, E.L., Hagerty, C.H., Mikaberidze, A., Mundt, C.C., Zhong, Z., McDonald, B.A., 2016. An Improved Method for Measuring Quantitative Resistance to the Wheat Pathogen *Zymoseptoria tritici* Using High-Throughput Automated Image Analysis. *Phytopathology* 106, 782–788. <https://doi.org/10.1094/PHYTO-01-16-0018-R>

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going Deeper with Convolutions.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the Inception Architecture for Computer Vision, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp. 2818–2826. <https://doi.org/10.1109/CVPR.2016.308>

Vale, F.X.R., Fernandes Filho, E.I., Liberato, J.R., 2003. QUANT. A software for plant disease severity assessment., in: In: 8th International Congress of Plant Pathology. Christchurch New Zealand, p. 105.

Wallach, D., Goffinet, B., 1989. Mean squared error of prediction as a criterion for evaluating and comparing system models. *Ecol. Modell.* 44, 299–306. [https://doi.org/10.1016/0304-3800\(89\)90035-5](https://doi.org/10.1016/0304-3800(89)90035-5)

Wang, G., Sun, Y., Wang, J., 2017. Automatic Image-Based Plant Disease Severity Estimation Using Deep Learning. *Comput. Intell. Neurosci.* 2017, 1–8. <https://doi.org/10.1155/2017/2917536>

Wang, X., Ma, H., You, S., 2020. Deep clustering for weakly-supervised semantic segmentation in autonomous driving scenes. *Neurocomputing* 381, 20–28. <https://doi.org/10.1016/j.neucom.2019.11.019>

Xu, L., Li, Y., Xu, J., Guo, L., 2020. Two-level attention and score consistency network for plant segmentation. *Comput. Electron. Agric.* 170, 105281. <https://doi.org/10.1016/j.compag.2020.105281>

Yang, Q., Liu, M., Zhang, Z., Yang, S., Ning, J., Han, W., 2019. Mapping Plastic Mulched Farmland for High Resolution Images of Unmanned Aerial Vehicle Using Deep Semantic Segmentation. *Remote Sens.* 11, 2008. <https://doi.org/10.3390/rs11172008>

Zhang, S., Ma, Z., Zhang, G., Lei, T., Zhang, R., Cui, Y., 2020. Semantic Image Segmentation with Deep Convolutional Neural Networks and Quick Shift. *Symmetry (Basel)*. 12, 427. <https://doi.org/10.3390/sym12030427>

Zhang, Y., Miao, S., Mansi, T., Liao, R., 2020. Unsupervised X-ray image segmentation with task driven generative adversarial networks. *Med. Image Anal.* 62, 101664. <https://doi.org/10.1016/j.media.2020.101664>

Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J., 2017. Pyramid Scene Parsing Network, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp. 6230–6239. <https://doi.org/10.1109/CVPR.2017.660>

3 ARTIGO II - FATORES QUE INFLUENCIAM O DESEMPENHO DAS REDES NEURAI CONVOLUCIONAIS NA SEGMENTAÇÃO SEMÂNTICA DE INJÚRIAS CAUSADAS PELO BICHO MINEIRO EM FOLHAS DE CAFÉ

RESUMO

Pragas e patógenos comprometem a produtividade das culturas e produzem perdas econômicas importantes em todo o mundo. A segmentação dos sintomas causados por estes organismos a partir de imagens digitais é uma tarefa fundamental, porém desafiadora, principalmente em situações com plano de fundo complexo e iluminação heterogênea. O aprendizado profundo, mais especificamente, as redes neurais convolucionais, trouxeram recentes avanços científicos na segmentação semântica de imagens digitais, porém, sua aplicação a fim de solucionar problemas fitopatológicos é incipiente. Este estudo fez uma investigação do efeito do plano fundo da imagem, aumento de dados, tamanho da imagem, *backbone*, número de dobras (k) na validação cruzada *k-fold*, algoritmo de otimização e aprendizado por transferência no desempenho do modelo de Rede de Pirâmides Características (FPN) para segmentação semântica. O conjunto de dados de imagem deste estudo reuniu 405 imagens de folhas de café danificadas pelo bicho-mineiro de folhas (*Leucoptera coffeella*) obtidas em plantações de café com ocorrência natural da praga. Três classes semântica foram anotadas manualmente em cada imagem: fundo da imagem, área foliar saudável e lesionada. A melhor performance na segmentação obteve uma Interseção-sobre-União (IoU) de 98,70% para o fundo da imagem, 97,89% para área foliar saudável e 84,52% para a área lesionada. Os experimentos computacionais mostraram que os fatores analisados impactaram o desempenho do modelo FPN na segmentação semântica de imagens foliares.

Palavras-chave: Aprendizado profundo, redes neurais convolucionais, processamento de imagens, segmentação de imagens

ABSTRACT

Pests and pathogens compromise crop yields and produce major economic losses worldwide. The segmentation of symptoms caused by these organisms from digital images is a fundamental, but challenging task, especially in situations with complex backgrounds and heterogeneous lighting. Deep learning, more specifically, convolutional neural networks, have brought recent scientific advances in the semantic segmentation of digital images, however, its application to solve plant pathology problems is incipient. This study investigated the effects of the image background, data increase, image size, backbone, number of folds (k) in the k-fold cross-validation, optimization algorithm and transfer learning in the performance of the Feature Pyramid Network (FPN) model for semantic segmentation. The image dataset of this study was composed of 405 images of coffee leaves damaged by the leaf miner (*Leucoptera coffeella*) obtained from coffee plantations with natural occurrence of the pest. Three semantic classes were manually annotated on each image: background of the image, healthy and damaged leaf area. The best performance in the segmentation obtained an Intersection-Over-Union (IoU) of 98.70% for the image background, 97.89% for healthy leaf area and 84.52% for the damaged area. The computational experiments showed that the factors analyzed impacted the performance of the FPN model in the semantic segmentation of leaf images.

Keywords: Deep learning, Convolutional neural networks, Image processing, image segmentation

3.1 INTRODUÇÃO

Pragas e patógenos de plantas afetam a produtividade das culturas e colocam em risco a segurança alimentar (Oerke, 2006). A capacidade de detectar e quantificar com precisão os sintomas causados por estes organismos é essencial para compreender os efeitos de tratamentos, perda de rendimento e classificar corretamente os fenótipos de germoplasma (Bock et al., 2020). Embora avanços científicos em genética, química e robótica tenham contribuído significativamente para a evolução da tecnologia agrícola, a produção alimentar precisa ser aumentada devido ao rápido crescimento populacional (Radoglou-Grammatikis et al., 2020). Portanto, a implementação de métodos de monitoramento de culturas eficientes e acessíveis para detectar e quantificar pragas e doenças de plantas poderia facilitar o manejo de plantas e contribuir para o aumento da produção de alimentos.

A aplicação do aprendizado profundo, mais especificamente, as redes neurais convolucionais (*Convolutional Neural Networks* - CNNs), aos problemas fitopatológicos começou a integrar vários trabalhos após 2015 e grandes progressos foram feitos, principalmente, na detecção e identificação automática de pragas e doenças (Barbedo, 2019; Lu et al., 2017; Mohanty et al., 2016). As CNNs constituem uma das técnicas mais poderosas no reconhecimento de padrões em imagem, dotadas de um método totalmente automático, alcançaram elevados níveis de precisão na classificação de imagens tanto em laboratório quanto em campo.

Várias abordagens de caráter investigativo foram propostas para comparar arquiteturas populares de CNN, como AlexNet (Krizhevsky et al., 2012), VGG net (Simonyan e Zisserman, 2014), GoogLeNet (Szegedy et al., 2015), ResNet (He et al., 2016) e DenseNet (Huang et al., 2017) na classificação de pragas e doenças de plantas (Ferentinos, 2018; Lu et al., 2017; Mohanty et al., 2016; S. Zhang et al., 2019). Algumas abordagens buscaram investigar os fatores que afetam o desempenho de modelos, como aprendizado por transferência (Chen et al., 2020; Kaya et al., 2019; Mohanty et al., 2016), aumento de dados (Barbedo, 2019; Liu et al., 2017; Mehdipour Ghazi et al., 2017), algoritmos de otimização (Liu et al., 2017), fundo da imagem (Barbedo, 2018), qualidade da imagem (Barbedo and Castro, 2019), entre outros. Todos esses fatores mencionados se mostraram relevantes e impactaram o desempenho dos modelos de CNN na classificação de imagens.

Para controle eficiente de pragas e doenças de plantas, é importante não apenas identificar os sintomas, mas também estimar a severidade, uma vez que a identificação e a quantificação dos sintomas são duas variáveis igualmente importantes nas áreas de entomologia

e fitopatologia. Neste sentido, alguns estudos usaram as CNNs para classificar imagens foliares em intervalos categorizados de severidade com resultados promissores de precisão (Esgario et al., 2020; Liang et al., 2019; Wang et al., 2017). Entretanto, estimar a severidade em poucas (2 a 4) categorias pode não ser útil para alguns objetivos da pesquisa, especialmente em avaliações de cultivares resistentes e efeito de tratamento de fungicidas foliares. Para essas situações, o percentual de área afetada é mais informativa e geralmente é fornecida por avaliações visuais (Bock et al., 2016). Nestes casos, se faz necessário aplicar métodos capazes de rotular cada pixel da imagem em classes semânticas.

As CNNs trouxeram recentes avanços científicos na segmentação semântica de imagens digitais. Na segmentação semântica, também conhecida como classificação a nível de pixel, as CNN atribuem uma classe semântica a cada pixel na imagem de entrada. Um dos primeiros trabalhos, desenvolvido por Long et al. (2015), propôs uma rede totalmente convolucional, em que modelos de CNNs voltados para classificação, como AlexNet, VGG net e GoogLeNet, foram adaptados para tarefa de classificação a nível de pixel por meio da substituição da camada totalmente conectada por uma camada de convolução transposta, ou deconvolução. Posteriormente, grande parte dos modelos de CNNs aplicados na segmentação semântica adotou uma estrutura codificador-decodificador (Krestenitis et al., 2019). Na seção de codificação, camadas de convoluções e *max pooling* extraem informações de contexto. Na seção de decodificação, camadas de deconvolução e *upsampling* recuperam a informação espacial. O modelo UNet (Ronneberger et al., 2015), por exemplo, com estrutura codificador-decodificador, tem mostrado um potencial impressionante na segmentação de imagens médicas, mesmo em situações em que o número de imagens de treinamento rotuladas é pequeno.

A aplicação de modelos de CNNs para segmentação semântica na agricultura têm sido bem-sucedidas, incluindo mapeamento de áreas de cultivos (Du et al., 2019), sistemas de irrigação por pivô central (Saraiva et al., 2020), detecções de plantações (Clark e McKechnie, 2020) e segmentação de plantas (Majeed et al., 2020; Xu et al., 2020). Em estudos de fitopatologia, o uso das CNNs para segmentação semântica é incipiente, aplicado em condições com fundo padronizado e luminosidade controlada (Lin et al., 2019). Embora o desempenho dos modelos de CNN na segmentação seja encorajador, os diversos fatores que o afetam ainda não são bem compreendidos e podemos variar em diferentes conjuntos de dados.

Este estudo apresenta uma análise detalhada dos principais fatores que afetam o desempenho de modelos de CNNs na segmentação semântica de imagens. Imagens de folhas de café danificadas pelo inseto-praga bicho-mineiro do café (*Leucoptera coffeella*) foram estrategicamente fotografadas em condição de campo para expor as folhas às condições mais

desafiadoras para os métodos de segmentação: fundo complexo e iluminação irregular. O modelo de Rede de Pirâmides Características (FPN) foi selecionado para investigação dos seguintes fatores que podem afetar seu desempenho na segmentação semântica: fundo da imagem, aumento de dados, tamanho da imagem, *backbone*, número de dobras na validação cruzada, algoritmo de otimização e aprendizado por transferência. Esta é a principal contribuição deste estudo.

3.2 MATERIAL E MÉTODOS

3.2.1 Conjunto de dados de imagem

Imagens de folhas de café arábica danificadas pelo bicho-mineiro (*Leucoptera coffeella*) foram obtidas em plantações de café com ocorrência natural da praga. No total, 405 imagens de folhas que exibiam vários níveis de severidade foram destacadas da planta e fotografadas em campo usando o smartphone Motorola One XT1941-3 (Motorola Inc., Chicago, IL, EUA) na resolução de 13 MPixels. O smartphone foi posicionado horizontalmente a uma distância de aproximadamente 20 cm de uma única folha exposta a uma ampla faixa de condições de iluminação e fundo.

Para que os modelos da CNN aprendam a classificar cada pixel em uma imagem, é necessária uma anotação de referência. Para esta tarefa foi desenvolvido um algoritmo com base em limiar de segmentação, que permite ao usuário selecionar os limiares mais adequados para discriminar o fundo, tecidos saudáveis e lesões. Devido à baixa precisão de segmentação de algoritmos desenvolvidos com base na seleção do limiar em segmentar plano de fundo complexo, este foi anotado manualmente. Como resultado, uma máscara de três classes foi gerada. Na Figura 1 são apresentados exemplos de imagens acompanhados de sua anotação de referência e a frequência da severidade real do conjunto de imagens, expressa em porcentagem de área afetada, determinada digitalmente pela ferramenta. Este algoritmo e o conjunto de imagens deste trabalho está disponível em <https://osf.io/hx2um/>.

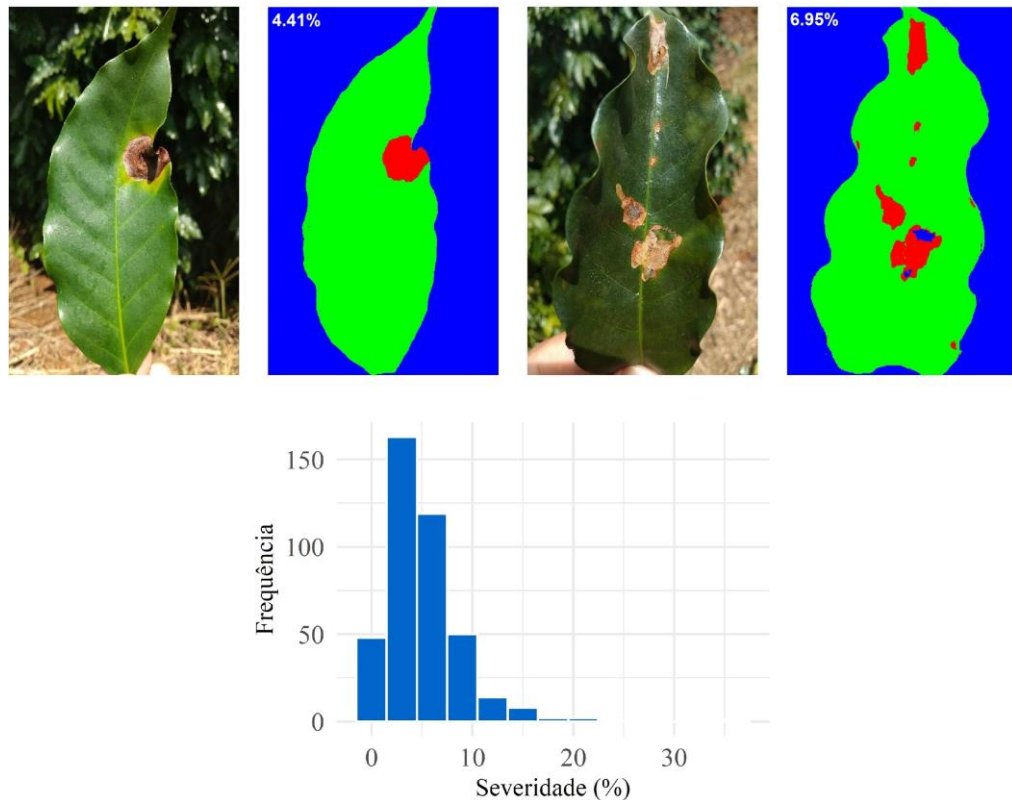


Figure 1. Frequência da severidade do bicho mineiro do café.

3.2.2 Configuração experimental

Para investigação dos fatores que influenciam o desempenho das CNN na segmentação semântica de pragas de plantas foi selecionado o modelo de Rede de Pirâmides Características (FPN) proposto por Lin et al. (2017). Para implementação do FPN foi utilizada a linguagem de programação Python 3.6 e o ambiente de desenvolvimento integrado Anaconda 3. Os códigos foram escritos com base na biblioteca Keras (<https://keras.io/>) com o Tensorflow (<https://www.tensorflow.org/>) como *back-end*. Os treinamentos foram realizados em um computador equipado com uma CPU (unidade central de processamento) Intel Core i7-8700 (3,70 GHz) e 16 GB de RAM. A GPU (unidade de processamento gráfico) usada foi a NVIDIA GeForce RTX 2080 com 8 GB de memória física.

Os fatores investigados neste estudo que podem afetar o desempenho das CNNs na segmentação semântica de pragas de plantas foram o plano fundo da imagem, aumento de dados, tamanho da imagem, *backbone*, número de dobras (k) na validação cruzada *k-fold*, algoritmo de otimização e aprendizado por transferência. Os planos de fundo da imagem investigados foram fundo complexo, presente nas imagens originais e fundo simples, resultantes de uma etapa de pré-processamento onde os pixels de fundo das imagens originais foram convertidos manualmente para cor branca.

O conjunto de dados foi dividido em um conjunto de treinamento (80% das amostras) e um conjunto de teste (20%). A primeira variante de treinamento proposta e usada como referência na comparação adotou a seguinte configuração experimental: imagens originais com fundo complexo, aumento de dados de treinamento (Figura 2), imagem com dimensão de entrada de $512 \times 512 \times 3$ pixels, *backbone* Inception-v3 (Szegedy et al., 2016) pré-treinado no conjunto de dados ImageNet (Russakovsky et al., 2015), algoritmo de otimização descida de gradiente estocástico (SGD) e validação cruzada *k-fold*, com $k = 5$. As demais variantes propostas são descritas na Tabela 1.



Figura 2. Resultado do processo de aumento de dados.

Tabela 1. Variantes propostas para investigação dos fatores que afetam o desempenho das CNNs na segmentação semântica de pragas de plantas.

Fatores	Fundo da imagem	Aumento de dados	Resolução da imagem	Backbone	N. de dobras K	Otimizador	Aprendizado por transferência
Fundo da imagem	Simples	Presente	512 x 512	Inception-v3	k = 5	SGD	ImageNet
Aumento de dados	Complexo	Ausente	512 x 512	Inception-v3	k = 5	SGD	ImageNet
Resolução da imagem	Complexo	Presente	128 x 128	Inception-v3	k = 5	SGD	ImageNet
			256 x 256				
Backbone	Complexo	Presente	512 x 512	VGG-19	k = 5	SGD	ImageNet
N. de dobras K	Complexo	Presente	512 x 512	Inception-v3	k = 3	SGD	ImageNet
					k = 10		
Otimizador	Complexo	Presente	512 x 512	Inception-v3	k = 5	RMSprop	ImageNet
Aprendizado por transferência	Complexo	Presente	512 x 512	Inception-v3	k = 5	SGD	Ausente

Os hiperparâmetros foram padronizados em todos os treinamentos: a entropia cruzada categórica (de Boer et al., 2005) como função de perda, o tamanho de lote igual a 2, a taxa de aprendizado foi definida de 0,001 e decaimento igual a $1e-5$. Para o algoritmo de otimização, descida de gradiente estocástico (SGD), foi adotado um momento *Nesterov* de 0,9. Cada um dos experimentos foi executado em 30 épocas. A razão pela qual se definiu o número de épocas

igual a 30 se deve ao fato de que após 30 épocas, nenhuma melhoria adicional foi notada em testes preliminares.

3.2.3 Métrica de avaliação

As lesões causadas pelo bicho-mineiro de folhas (*Leucoptera coffeella*) representa apenas um pequeno segmento da imagem. Neste caso, métricas como acurácia, precisão e recordação não são recomendadas, pois podem levar a um falso senso de superioridade, elevado pelo desempenho na classe de fundo da imagem. Portanto, a interseção sobre união (IoU), também conhecida como índice de Jaccard, tem sido amplamente usada em problemas de segmentação semântica para avaliar quantitativamente a semelhança entre a máscara de referência e a máscara predita (Rahman e Wang, 2016). A IoU é definida como a razão da área de interseção entre a máscara de referência e a máscara predita sobre a área de união entre a máscara de referência e a máscara predita (Equação 1). A IoU varia de 0 a 1, sendo que 0 significa inexistência de sobreposição das classes semânticas entre a máscara de referência e a máscara predita e 1 significa sobreposição perfeita.

$$IoU = \frac{ground\ truth \cap prediction}{ground\ truth \cup prediction} = \frac{TP}{TP + FP + FN} \quad (1)$$

em que,

TP - número de verdadeiros positivos;

FP - número de falsos positivos;

FN - número de falsos negativos.

3.3 RESULTADOS E DISCUSSÃO

Na Figura 3 são apresentados resultados da segmentação semântica realizada pelo modelo FPN no conjunto de dados de teste usando a configuração experimental de referência, especificada na Seção 2.2. Observa-se que o modelo apresentou bom desempenho e as máscaras resultantes do processo de segmentação foram altamente semelhantes as máscaras de referência.

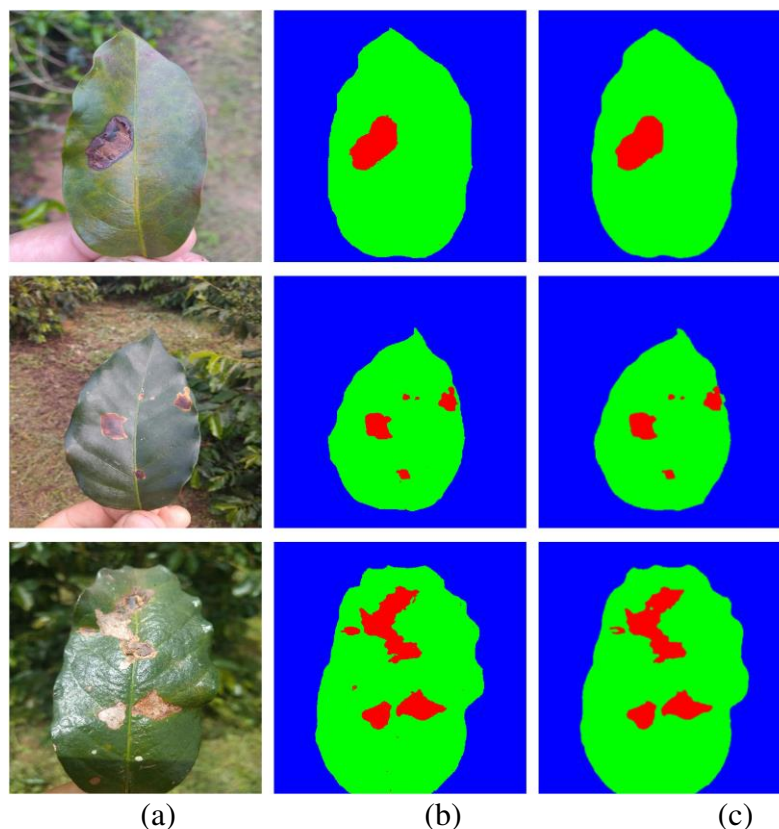


Figura 3. Comparação visual dos resultados da segmentação semântica em imagens de teste. (a) imagem original; (b) anotação de referência; (c) resultado da segmentação do modelo FPN.

Nas próximas seções são apresentados e discutidos os resultados do efeito de cada fator estudado no desempenho da CNN. A IoU foi calculada sobre as imagens do conjunto teste para cada classe e a média geral calculada a partir da média de cada classe.

3.3.1 Plano de fundo da imagem

Os resultados experimentais mostraram que o pré-processamento aplicado ao plano de fundo das imagens para criação de um fundo simples tendeu a melhorar o desempenho do modelo FPN na predição das classes semânticas saudável e lesão, obtendo uma IoU de 97,96 e 81,61%, respectivamente (Figura 4). O treinamento com o fundo complexo afetou principalmente a classe lesão, com IoU de 80,18%.

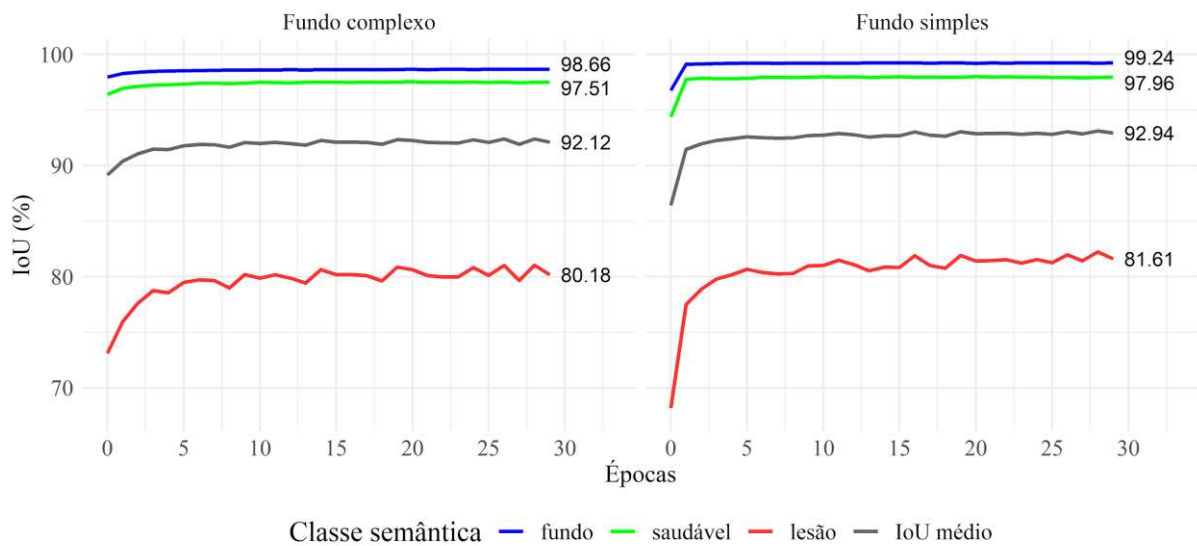


Figura 4. Variação da IoU com o número de épocas em função do plano de fundo da imagem.

Diversos algoritmos validados até o momento para segmentação de pragas e doenças foliares adotaram uma etapa de pré-processamento, em que os pixels de fundo da imagem são alterados manualmente para cor azul, preta, ou branca (Barbedo, 2016a; Pethybridge e Nelson, 2015). Além disso, alguns trabalhos optaram por introduzir um fundo padronizado nas cores azul ou preto no momento da captura da imagem foliar (Barbedo, 2016b). Isto porque, imagens capturadas em condições de campo, com fundo complexo, afetam o desempenho de algoritmos de segmentação, principalmente em fundos com presença de elementos de cor verde (Barbedo, 2016b).

Embora a remoção do fundo tenha melhorado a IoU do modelo FPN na segmentação das imagens (0,45% na classe saudável e 1,43% na classe lesão), os resultados foram semelhantes e mostram que o modelo FPN foi capaz de segmentar o fundo da imagem com precisão. Isto implica que o modelo é capaz de extrair informações não apenas de cor (a nível de pixel), mas também de contexto, como por exemplo, a textura. De fato, o desempenho na classe de fundo com fundo simples e complexo foram semelhantes, com IoU de 99,24 e 98,66%, respectivamente. Portanto, a segmentação de imagens foliares com o fundo complexo para estimar a severidade pelo modelo FPN se apresenta como uma ótima alternativa, uma vez que a introdução de fundo padronizado ou etapa de pré-processamento para remover o fundo complexo reduz a capacidade de aquisição e processamento de grandes quantidades de imagens.

3.3.2 Aumento de dados

A comparação entre os experimentos computacionais com e sem aumento de dados no conjunto de dados de treinamento mostram que a inclusão de dados de imagens (Figura 2) tendeu a melhorar os valores de IoU na segmentação das imagens (Figura 5). A IoU média sem e com aumento de dados foi de 90,53 e 92,12%, respectivamente. O treinamento sem aumento de dados afetou principalmente a classe lesão, com IoU de 76,19%.

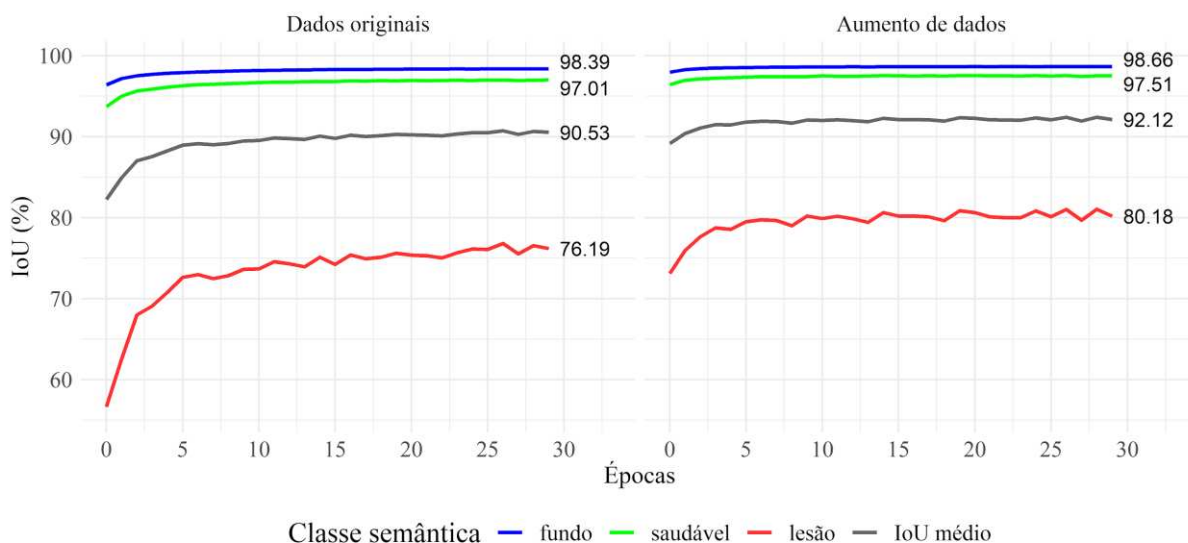


Figura 5. Variação da IoU com o número de épocas com e sem aumento de dados.

Técnicas de aumento de dados, que incluem rotação, *flip* horizontal e vertical e aumento e diminuição do zoom, são usadas para aumentar artificialmente o número de dados de treinamento e diminuir as chances de *overfitting* (Krizhevsky et al., 2012). Durante o processo de treinamento, o *overfitting* é caracterizado por um ajuste excessivo no conjunto de treinamento acompanhado de uma degradação no desempenho sobre o conjunto de teste (Srivastava et al., 2014). A definição do número de épocas igual a 30 permitiu acompanhar consistentemente o aumento do IoU nas classes semânticas até o modelo convergir sem manifestações de deterioração do desempenho no conjunto teste mesmo sem o aumento de dados (Figura 5).

Diversos estudos que utilizaram CNNs na classificação de pragas e doenças de plantas aplicaram e/ou investigaram diretamente o impacto de técnicas de aumentos de dados na acurácia e obtiveram ganhos significativos (Barbedo, 2019, 2018; Liu et al., 2017). Em segmentação semântica, Huang et al. (2020) compararam o desempenho do modelo de CNN com e sem aumento de dados e mostraram que a inclusão de aumento de dados pode melhorar os valores da IoU para mapear áreas de degelo.

Embora o mais lógico para melhorar o desempenho na segmentação pelos modelos de CNNs seja a aquisição de novos dados de imagens, este processo pode consumir tempo, recursos humanos e financeiros. Além disso, o processo de anotação de referência das imagens, em alguns casos, como neste estudo, pode ser extremamente demorado e difícil, mesmo com o auxílio de algoritmos específicos na tarefa (GitHub¹). Técnicas como rotação, flip horizontal e vertical podem ser aplicadas as anotações de referência de cada imagem e, portanto, gerar uma anotação correspondente de forma simples e rápida. Portanto, o aumento de dados, principalmente no treinamento de modelos de CNN para segmentação semântica, se apresenta como uma excelente alternativa para melhorar o desempenho dos modelos na segmentação de pragas foliares.

3.3.3 Resolução da imagem

A resolução das imagens afetou o processo de segmentação (Figura 6). Os resultados mostraram que o processo de treinamento e teste com imagens de resolução 512x512x3 tendeu a apresentar melhores resultados, com a IoU média de 92,11%. O redimensionamento das imagens para 128x128x3, fizeram com que a IoU média caísse para 88,02%. A classe lesão foi a mais afetada, com a IoU de 70,67%.

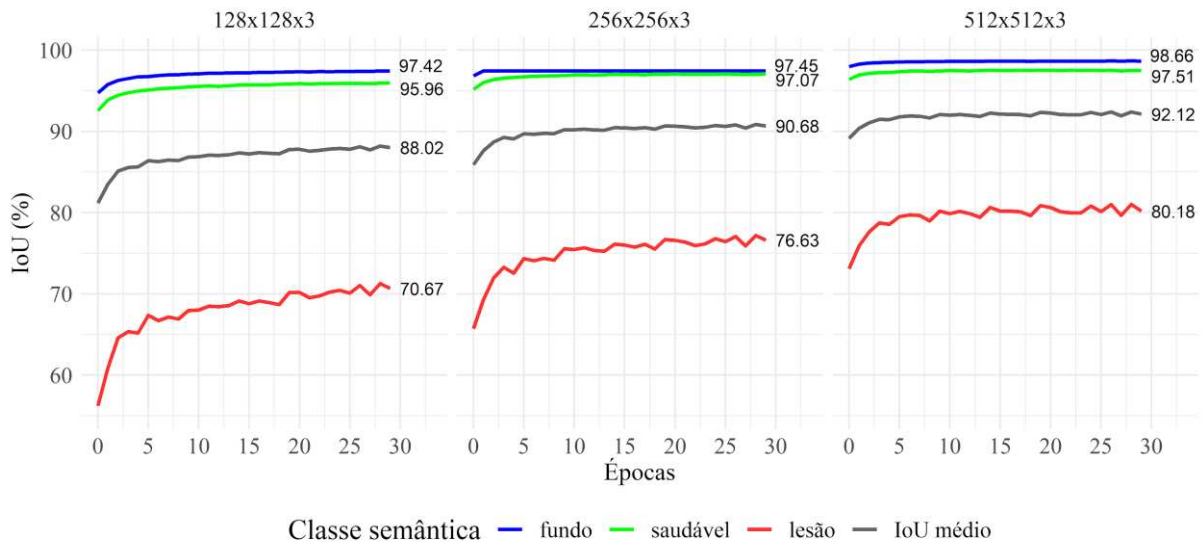


Figura 6. Variação da IoU com o número de épocas de acordo com o tamanho da imagem.

O redimensionamento de imagens em treinamento de modelos de CNN é comum e busca adaptar o tamanho da imagem de entrada exigido pelos modelos e/ou reduzir os custos computacionais. Muitos métodos de redimensionamento foram propostos, entretanto, quando uma imagem é redimensionada, geralmente resulta em perda de informação e deformação das estruturas (T. Zhang et al., 2019). As lesões causadas pelo bicho-mineiro de folhas representam

pequenos segmentos na imagem e, portanto, durante o processo de redimensionamento é a classe que mais é afetada pela perda de informação e deformação.

3.3.4 Backbone

Os resultados mostraram que o *backbone* VGG-19 foi capaz de extrair maior quantidade de informações semânticas e obteve uma IoU média de 92,76% (Figura 7). O melhor desempenho do VGG-19 frente ao Inception-v3 foi na classe lesão, com 81,83 e 80,18%, respectivamente.

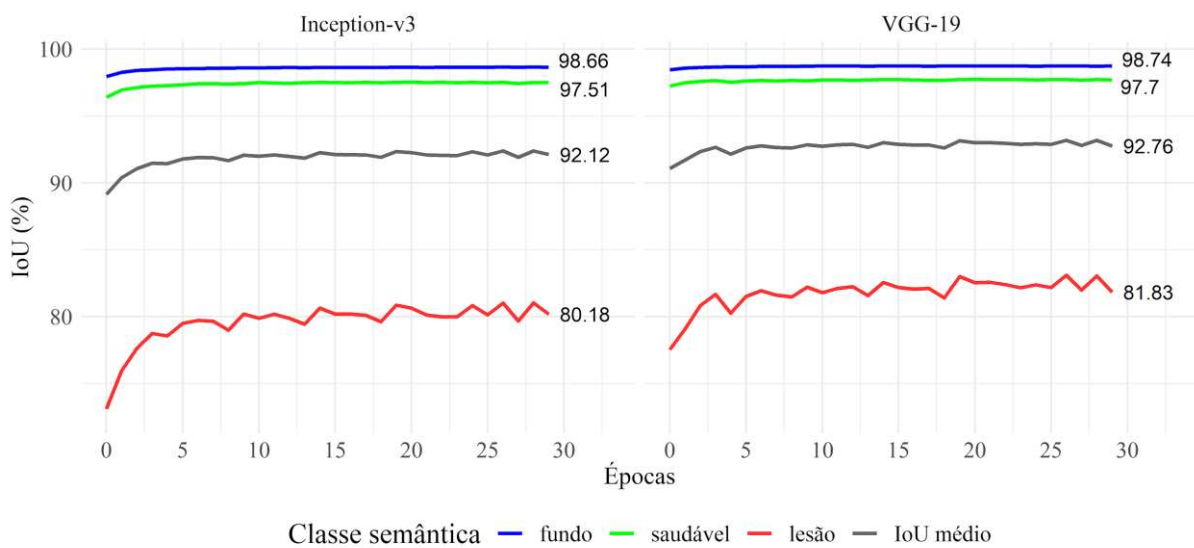


Figura 7. Variação da IoU com o número de épocas de acordo com o backbone.

Em geral, os modelos de segmentação semântica adaptam os *backbones* de modelos de CNN de classificação a nível de imagem, como os investigados neste estudo (Fan et al., 2018). Os *backbones*, compostos de camadas de convoluções e *max pooling*, são responsáveis por extrair informações de contexto e espaciais, embora informações espaciais em um nível menor. O modelo VGG-19, composto por 19 camadas, consiste em camadas de convolução composta apenas por filtros 3×3 , camadas de *max-pooling* 2×2 seguidas por duas camadas totalmente conectadas e camada *Softmax* como saída. Filtros menores extraem maior quantidade de informações distribuídas localmente e, portanto, favorecem a preservação de maior quantidade de informação espacial, principalmente de bordas e podem justificar o melhor desempenho do modelo VGG como backbone.

3.3.5 Número de dobras k

Na validação cruzada *k-fold*, o conjunto de dados de treinamento D é dividido aleatoriamente em subconjuntos k mutuamente exclusivos de tamanhos iguais ($D_1, D_2, D_3, \dots, D_k$). Após a divisão, um único subconjunto k é retido para validação e o restante dos subconjuntos ($D_1, D_2, D_3, \dots, D_{k-1}$) é utilizado para treinamento. Portanto, a definição do número de subconjuntos k influencia no número de imagens de treinamento e no tempo total do processo para avaliar o desempenho do modelo em um conjunto de dados de teste. Neste estudo, foi testada a validação cruzada *k-fold* ($k = 3, 5$ e 10), configurações mais usadas em aprendizado de máquinas (Huang et al., 2020).

Os resultados obtidos mostram que a IoU média aumentou à medida em que k aumenta, o que indica que mais dados de treinamento (90% em $k = 10$, 80% em $k = 5$ e 67% em $k = 3$) podem levar a melhores resultados (Figura 8). Entretanto, a diferença na IoU média entre os treinamentos com menor ($k = 3$) e maior ($k = 10$) número de imagens foi de apenas 0,4%, apesar da redução de aproximadamente 25% da quantidade de imagens, mostrando assim a capacidade de generalização do modelo FPN.

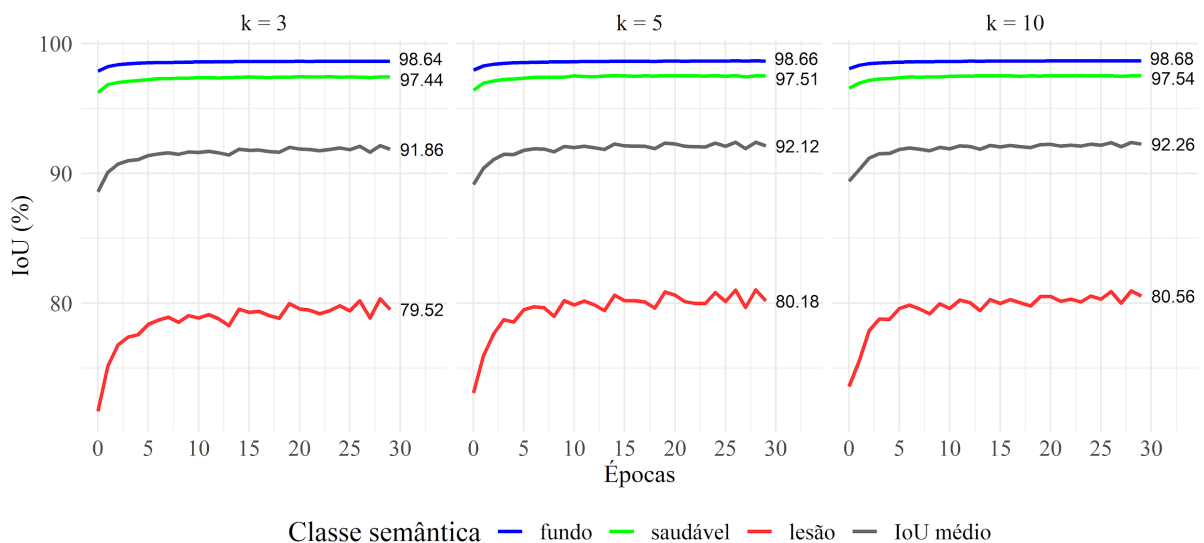


Figura 8. Variação da IoU com o número de épocas em função do número de subconjuntos k na validação cruzada.

3.3.6 Algoritmo de otimização

O algoritmo de otimização dos parâmetros de rede impactou significativamente o processo de segmentação das imagens (Figura 9). Os resultados mostraram que o algoritmo RMSprop (Tieleman e Hinton, 2012) atingiu uma IoU média de 93,70%, 1,58% superior à

obtida pelo otimizador SGD. A classe lesão apresentou o maior acréscimo no desempenho com o otimizador RMSprop, com a IoU de 84,52%, 4,34% superior à obtida pelo otimizador SGD.

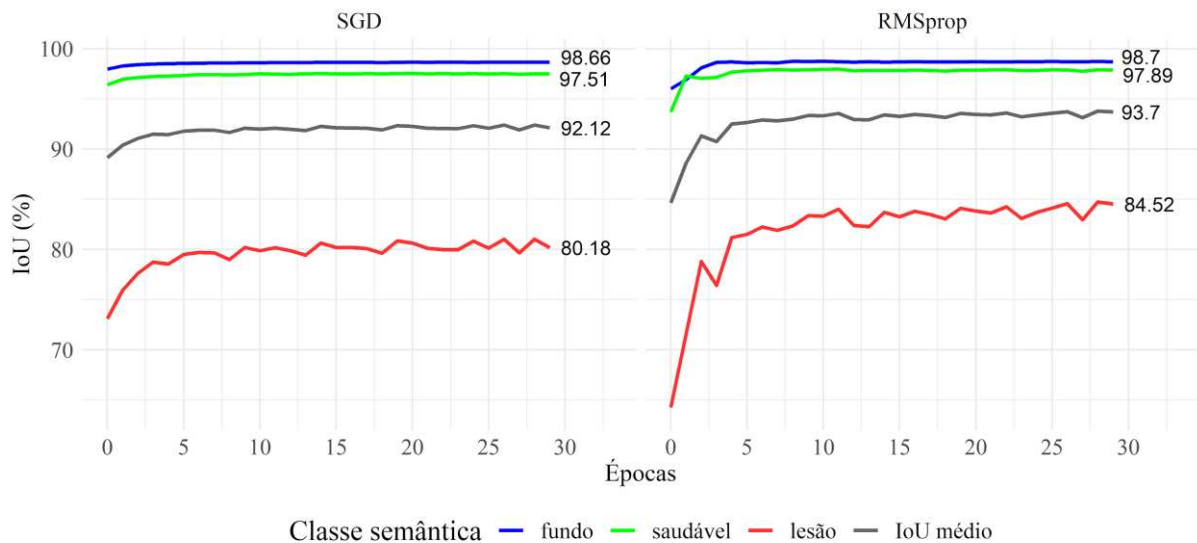


Figura 9. Variação da IoU com o número de épocas em função do tipo de algoritmo otimizador.

A descida do gradiente estocástico é o mais popular e amplamente usado entre todos os otimizadores no treinamento de CNN (Tang et al., 2019). Entretanto, o desempenho da descida do gradiente estocástico (SGD) depende criticamente de como as taxas de aprendizado são ajustadas e diminuídas ao longo do tempo (Schaul et al., 2013). Grande parte das funções de perda em modelos de CNN não são convexas e, portanto, dependendo do ajuste da taxa de aprendizado, há uma chance de ficar preso em um mínimo local e a perda nunca convergir para o valor mínimo global. O bom ajuste da taxa de aprendizado e de seu decaimento, por exemplo, no tempo, em etapas ou exponencialmente, permite que o SGD quase certamente convirja para um mínimo global (Vasudevan, 2020). No entanto, em alguns casos, como neste estudo, é difícil ajustar uma taxa inicial de aprendizado ideal (Tang et al., 2019).

Alternativamente, alguns algoritmos como o RMSProp são adaptativos e foram propostos para ajustarem a taxa de aprendizado automaticamente e usados para treinar redes neurais profundas (Vasudevan, 2020). O RMSprop utiliza médias móveis exponenciais dos gradientes quadrados passados para adaptar a taxa de aprendizado (Tang et al., 2019). Sua aplicação em treinamento de modelos de CNN para segmentação semântica é incipiente, entretanto, aplicado com frequência no treinamento de proposição e desenvolvimento de novos modelos de CNN para segmentação (Chaurasia e Culurciello, 2017; Chollet, 2017; Sandler et al., 2018; Tan e Le, 2019).

3.3.7 Aprendizado por transferência

Duas abordagens de aprendizagem foram exploradas: inicializando com os pesos ajustados no treinamento do modelo Inception-v3 no conjunto de dados *ImageNet*; e a partir do zero, em que os pesos são inicializados aleatoriamente. Os experimentos computacionais realizados mostraram que o aprendizado por transferência tendeu ao melhor desempenho, com IoU média de 92,12% (Figura 10). A classe lesão foi a mais afetada quando o treinamento se iniciou com os pesos aleatoriamente, com a IoU de 77,20%.

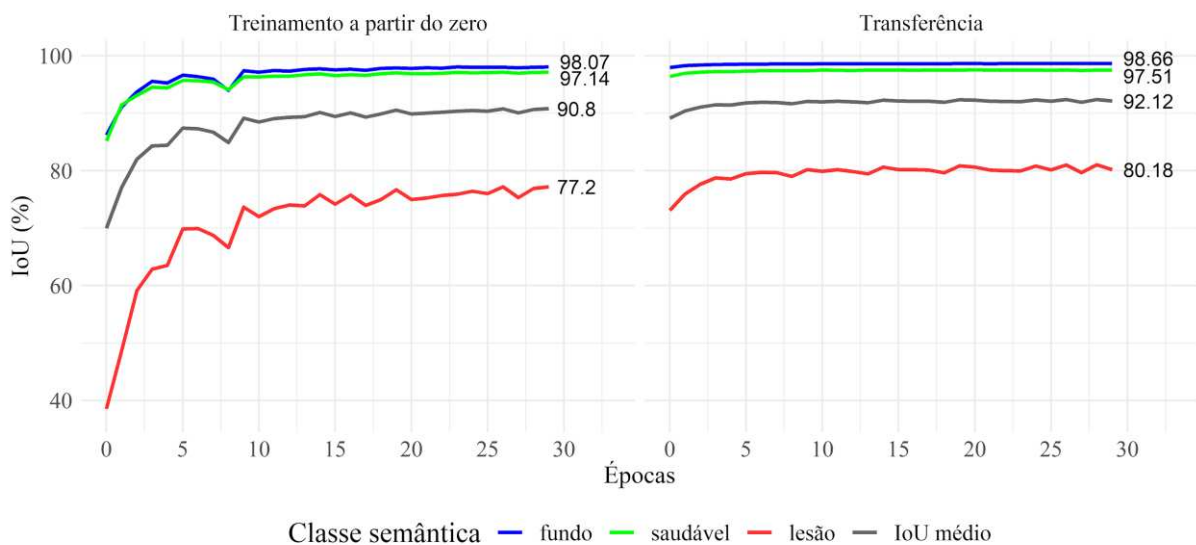


Figura 10. Variação da IoU com o número de épocas de acordo com o método de inicialização dos pesos da rede.

Wang et al. (2019) propuseram um método totalmente automatizado com base em CNN, denominado SegRoot, adaptado para segmentar raízes de solos com fundo complexo e obtiveram o melhor desempenho de segmentação usando modelos com pesos pré-treinados no ImageNet. Em fitopatologia, Mohanty et al. (2016) comparam essas duas abordagens de aprendizagem (transferência versus treinamento a partir do zero) para identificar 14 espécies de culturas e 26 doenças (ou ausência delas) e mostraram que o aprendizado de transferência produziu os melhores resultados.

O processo de transferência de aprendizagem acelerou a convergência do modelo FPN, que ocorreu em aproximadamente 5 épocas (Figura 10). A convergência do modelo é um fator importante em treinamentos de modelos e os resultados mostraram que o processo de aprendizagem por transferência pode reduzir significativamente o tempo de treinamento. Além disso, como relatado anteriormente, o processo de aquisição e anotação de referência das imagens é dispendioso e usar o aprendizado por transferência pode trazer ganhos de performance sem a necessidade de novas aquisições e anotações.

3.4 CONCLUSÃO

Neste artigo, foi investigado diversos fatores que influenciam o desempenho das redes neurais convolucionais na segmentação semântica de pragas de plantas. Embora ligeiramente superior, a performance na segmentação das imagens com fundo simples e complexo foram semelhantes mostrando a robustez do modelo FPN em trabalhar em condições desafiadoras de plano de fundo da imagem. A ausência do aumento de dados e redução do tamanho das imagens treinamento, reduziram a performance do modelo na segmentação das imagens. O *backbone* com maior capacidade de extrair informações de contexto e espaciais foi o VGG-19. O aumento do número de imagens de treinamento na validação cruzada (90% em $k = 10$, 80% em $k = 5$ e 67% em $k = 3$) melhorou o desempenho do modelo na segmentação, entretanto o crescimento de desempenho foi modesto mesmo com o aumento de aproximadamente 25% na quantidade de imagens. O algoritmo com melhor desempenho na otimização dos parâmetros de rede foi o RMSprop. O aprendizado por transferência realizado pelo backbone inception-v3 treinado no conjunto de dados do *Imagenet* melhorou a performance na segmentação das imagens e levou um menor número de épocas para convergir, se comparado ao treinamento a partir do zero.

3.5 REFERÊNCIAS

- Barbedo, J.G.A., 2019. Plant disease identification from individual lesions and spots using deep learning. *Biosyst. Eng.* 180, 96–107. <https://doi.org/10.1016/j.biosystemseng.2019.02.002>
- Barbedo, J.G.A., 2018. Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification. *Comput. Electron. Agric.* 153, 46–53. <https://doi.org/10.1016/j.compag.2018.08.013>
- Barbedo, J.G.A., 2016a. A novel algorithm for semi-automatic segmentation of plant leaf disease symptoms using digital image processing. *Trop. Plant Pathol.* 41, 210–224. <https://doi.org/10.1007/s40858-016-0090-8>
- Barbedo, J.G.A., 2016b. A review on the main challenges in automatic plant disease identification based on visible range images. *Biosyst. Eng.* 144, 52–60. <https://doi.org/10.1016/j.biosystemseng.2016.01.017>
- Barbedo, J.G.A., Castro, G.B., 2019. Influence of image quality on the identification of psyllids using convolutional neural networks. *Biosyst. Eng.* 182, 151–158. <https://doi.org/10.1016/J.BIOSYSTEMSENG.2019.04.007>
- Bock, C.H., Barbedo, J.G.A., Del Ponte, E.M., Bohnenkamp, D., Mahlein, A.-K., 2020. From visual estimates to fully automated sensor-based measurements of plant disease severity: status and challenges for improving accuracy. *Phytopathol. Res.* 2, 9. <https://doi.org/10.1186/s42483-020-00049-8>
- Bock, C.H., Chiang, K.-S., Del Ponte, E.M., 2016. Accuracy of plant specimen disease severity estimates: concepts, history, methods, ramifications and challenges for the future. *CAB Rev. Perspect. Agric. Vet. Sci. Nutr. Nat. Resour.* 11. <https://doi.org/10.1079/PAVSNR201611032>
- Chaurasia, A., Culurciello, E., 2017. LinkNet: Exploiting Encoder Representations for Efficient Semantic Segmentation. 2017 IEEE Vis. Commun. Image Process. 1–4. <https://doi.org/10.1109/VCIP.2017.8305148>
- Chen, Junde, Chen, Jinxiu, Zhang, D., Sun, Y., Nanekaran, Y.A., 2020. Using deep transfer learning for image-based plant disease identification. *Comput. Electron. Agric.* 173, 105393. <https://doi.org/10.1016/j.compag.2020.105393>
- Chollet, F., 2017. Xception: Deep Learning with Depthwise Separable Convolutions, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp. 1800–1807. <https://doi.org/10.1109/CVPR.2017.195>
- Clark, A., McKechnie, J., 2020. Detecting Banana Plantations in the Wet Tropics, Australia, Using Aerial Photography and U-Net. *Appl. Sci.* 10, 2017. <https://doi.org/10.3390/app10062017>
- de Boer, P.-T., Kroese, D.P., Mannor, S., Rubinstein, R.Y., 2005. A Tutorial on the Cross-Entropy Method. *Ann. Oper. Res.* 134, 19–67. <https://doi.org/10.1007/s10479-005-5724-z>
- Du, Z., Yang, J., Ou, C., Zhang, T., 2019. Smallholder Crop Area Mapped with a Semantic Segmentation Deep Learning Method. *Remote Sens.* 11, 888.

<https://doi.org/10.3390/rs11070888>

Esgario, J.G.M., Krohling, R.A., Ventura, J.A., 2020. Deep learning for classification and severity estimation of coffee leaf biotic stress. *Comput. Electron. Agric.* 169, 105162. <https://doi.org/10.1016/j.compag.2019.105162>

Fan, L., Wang, W.-C., Zha, F., Yan, J., 2018. Exploring New Backbone and Attention Module for Semantic Segmentation in Street Scenes. *IEEE Access* 6, 71566–71580. <https://doi.org/10.1109/ACCESS.2018.2880877>

Ferentinos, K.P., 2018. Deep learning models for plant disease detection and diagnosis. *Comput. Electron. Agric.* 145, 311–318. <https://doi.org/10.1016/j.compag.2018.01.009>

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep Residual Learning for Image Recognition.

Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely Connected Convolutional Networks, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp. 2261–2269. <https://doi.org/10.1109/CVPR.2017.243>

Huang, L., Luo, J., Lin, Z., Niu, F., Liu, L., 2020. Using deep learning to map retrogressive thaw slumps in the Beiluhe region (Tibetan Plateau) from CubeSat images. *Remote Sens. Environ.* 237. <https://doi.org/10.1016/j.rse.2019.111534>

Kaya, A., Keceli, A.S., Catal, C., Yalic, H.Y., Temucin, H., Tekinerdogan, B., 2019. Analysis of transfer learning for deep neural network based plant classification models. *Comput. Electron. Agric.* 158, 20–29. <https://doi.org/10.1016/J.COMPAG.2019.01.041>

Krestenitis, M., Orfanidis, G., Ioannidis, K., Avgerinakis, K., Vrochidis, S., Kompatsiaris, I., 2019. Oil Spill Identification from Satellite Images Using Deep Neural Networks. *Remote Sens.* 11, 1762. <https://doi.org/10.3390/rs11151762>

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet Classification with Deep Convolutional Neural Networks, in: *Advances in Neural Information Processing Systems*. pp. 1097–1105.

Liang, Q., Xiang, S., Hu, Y., Coppola, G., Zhang, D., Sun, W., 2019. PD2SE-Net: Computer-assisted plant disease diagnosis and severity estimation network. *Comput. Electron. Agric.* 157, 518–529. <https://doi.org/10.1016/j.compag.2019.01.034>

Lin, K., Gong, L., Huang, Y., Liu, C., Pan, J., 2019. Deep Learning-Based Segmentation and Quantification of Cucumber Powdery Mildew Using Convolutional Neural Network. *Front. Plant Sci.* 10. <https://doi.org/10.3389/fpls.2019.00155>

Lin, T.-Y., Dollar, P., Girshick, R., He, K., Hariharan, B., Belongie, S., 2017. Feature Pyramid Networks for Object Detection, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp. 936–944. <https://doi.org/10.1109/CVPR.2017.106>

Liu, B., Zhang, Y., He, D., Li, Y., 2017. Identification of Apple Leaf Diseases Based on Deep Convolutional Neural Networks. *Symmetry (Basel)*. 10, 11. <https://doi.org/10.3390/sym10010011>

Long, J., Shelhamer, E., Darrell, T., 2015. Fully Convolutional Networks for Semantic

Segmentation.

Lu, Y., Yi, S., Zeng, N., Liu, Y., Zhang, Y., 2017. Identification of rice diseases using deep convolutional neural networks. *Neurocomputing* 267, 378–384. <https://doi.org/10.1016/j.neucom.2017.06.023>

Majeed, Y., Karkee, M., Zhang, Q., Fu, L., Whiting, M.D., 2020. Determining grapevine cordon shape for automated green shoot thinning using semantic segmentation-based deep learning networks. *Comput. Electron. Agric.* 171, 105308. <https://doi.org/10.1016/j.compag.2020.105308>

Mehdipour Ghazi, M., Yanikoglu, B., Aptoula, E., 2017. Plant identification using deep neural networks via optimization of transfer learning parameters. *Neurocomputing* 235, 228–235. <https://doi.org/10.1016/j.neucom.2017.01.018>

Mohanty, S.P., Hughes, D.P., Salathé, M., 2016. Using Deep Learning for Image-Based Plant Disease Detection. *Front. Plant Sci.* 7, 1–10. <https://doi.org/10.3389/fpls.2016.01419>

Oerke, E.-C., 2006. Crop losses to pests. *J. Agric. Sci.* 144, 31–43. <https://doi.org/10.1017/S0021859605005708>

Pethybridge, S.J., Nelson, S.C., 2015. Leaf Doctor: A New Portable Application for Quantifying Plant Disease Severity. *Plant Dis.* 99, 1310–1316. <https://doi.org/10.1094/PDIS-03-15-0319-RE>

Radoglou-Grammatikis, P., Sarigiannidis, P., Lagkas, T., Moscholios, I., 2020. A compilation of UAV applications for precision agriculture. *Comput. Networks* 172, 107148. <https://doi.org/10.1016/j.comnet.2020.107148>

Rahman, M.A., Wang, Y., 2016. Optimizing intersection-over-union in deep neural networks for image segmentation, in: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Springer Verlag, pp. 234–244. https://doi.org/10.1007/978-3-319-50835-1_22

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation, in: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Springer Verlag, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L., 2015. ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis.* 115, 211–252. <https://doi.org/10.1007/s11263-015-0816-y>

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C., 2018. MobileNetV2: Inverted Residuals and Linear Bottlenecks, in: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society, pp. 4510–4520. <https://doi.org/10.1109/CVPR.2018.00474>

Saraiva, M., Protas, É., Salgado, M., Souza, C., 2020. Automatic Mapping of Center Pivot Irrigation Systems from Satellite Images Using Deep Learning. *Remote Sens.* 12, 558. <https://doi.org/10.3390/rs12030558>

- Schaul, T., Zhang, S., Lecun, Y., 2013. No More Pesky Learning Rates.
- Simonyan, K., Zisserman, A., 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition.
- Srivastava, N., Hinton, G., Krizhevsky, A., Salakhutdinov, R., 2014. Dropout: A Simple Way to Prevent Neural Networks from Overfitting, *Journal of Machine Learning Research*.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going Deeper with Convolutions.
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the Inception Architecture for Computer Vision, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp. 2818–2826. <https://doi.org/10.1109/CVPR.2016.308>
- Tan, M., Le, Q. V., 2019. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks.
- Tang, M., Huang, Z., Yuan, Y., Wang, C., Peng, Y., 2019. A Bounded Scheduling Method for Adaptive Gradient Methods. *Appl. Sci.* 9, 3569. <https://doi.org/10.3390/app9173569>
- Tieleman, T., Hinton, G., 2012. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. COURSERA Neural networks Mach. Learn. 4, 26–31.
- Vasudevan, S., 2020. Mutual Information Based Learning Rate Decay for Stochastic Gradient Descent Training of Deep Neural Networks. *Entropy* 22, 560. <https://doi.org/10.3390/e22050560>
- Wang, G., Sun, Y., Wang, J., 2017. Automatic Image-Based Plant Disease Severity Estimation Using Deep Learning. *Comput. Intell. Neurosci.* 2017, 1–8. <https://doi.org/10.1155/2017/2917536>
- Wang, T., Rostamza, M., Song, Z., Wang, L., McNickle, G., Iyer-Pascuzzi, A.S., Qiu, Z., Jin, J., 2019. SegRoot: A high throughput segmentation method for root image analysis. *Comput. Electron. Agric.* 162, 845–854. <https://doi.org/10.1016/J.COMPAG.2019.05.017>
- Xu, L., Li, Y., Xu, J., Guo, L., 2020. Two-level attention and score consistency network for plant segmentation. *Comput. Electron. Agric.* 170, 105281. <https://doi.org/10.1016/j.compag.2020.105281>
- Zhang, S., Zhang, Subing, Zhang, C., Wang, X., Shi, Y., 2019. Cucumber leaf disease identification with global pooling dilated convolutional neural network. *Comput. Electron. Agric.* 162, 422–430. <https://doi.org/10.1016/j.compag.2019.03.012>
- Zhang, T., Yu, M., Guo, Y., Liu, Y., 2019. Content-Aware Retargeted Image Quality Assessment. *Information* 10, 111. <https://doi.org/10.3390/info10030111>

4. CONCLUSÃO GERAL

Com resultados obtidos neste trabalho, pôde-se concluir que:

1. O modelo de Rede de Pirâmides Características (FPN) alcançou o melhor desempenho na segmentação do conjunto de imagens de teste, sob condições desafiadoras de fundo de imagem e iluminação.
2. O modelo FPN superou o software Assess, referência para os fitopatologistas, em todas as doenças e pragas analisadas, com coeficientes de concordância de 0,97, 0,98 e 0,98 para ferrugem da soja, mancha bronzeada do trigo e bicho-mineiro do café, respectivamente.
3. Apresentando um método totalmente automático de segmentação, o tempo de inferência e estimativa da severidade dos modelos de CNNs não ultrapassou os 6s, tempo aproximadamente 20 vezes maior que o obtido pelo modo manual de segmentação do software Assess.
4. Todos os fatores analisados tenderam a afetar o desempenho do modelo FPN na segmentação semântica de imagens foliares.