

Classification of Weeds in Agricultural Crops with Ensembles of Convolutional Neural Networks

Vitor M. de Souza Carvalho¹, Elloá B. Guedes¹, Marcos Filipe A. Salame²

¹Grupo de Pesquisa em Sistemas Inteligentes
Universidade do Estado do Amazonas
Av. Darcy Vargas, 1200 – Manaus – AM

²Embrapa Amazônia Ocidental
Caixa Postal 319 – 69.010-970 – Manaus – AM

{vmdsc.eng, ebgcosta}@uea.edu.br,marcos.salame@embrapa.br

Abstract. *The present work aimed at proposing and evaluation ensembles of convolutional neural networks to address the classification of agricultural crops versus weeds from images of seedlings in their early stages. To do so, there was a training data preparation phase, adoption of five convolutional neural network architectures and the proposition of three ensemble with different voting strategies. Upon considering the performance of individual networks, an accuracy of 95.77% was reported for MobileNet, meanwhile the ensemble with Support Vector Machine smart voting had accuracy of 97.08%. The results obtained consider a Computational Vision task that enhances the development of digital agriculture, favoring a better yield of agricultural production.*

Resumo. *Este trabalho teve por objetivo propor e avaliar o uso de comitês de redes neurais convolucionais aplicados à classificação de culturas agrícolas e de ervas daninhas a partir de imagens de mudas. Para tanto, foi considerada uma preparação dos dados de treinamento, a utilização de cinco arquiteturas de redes neurais convolucionais e a elaboração de três comitês com diferentes agregações. Para as redes individuais, uma acurácia de 95,77% foi obtida com o uso da rede MobileNet, e o comitê mediado por uma Máquina de Vetores de Suporte obteve acurácia de 97,08%. Os resultados obtidos colaboram em uma tarefa de Visão Computacional que potencializa o desenvolvimento da agricultura digital, favorecendo um melhor rendimento da produção agrícola.*

1. Introdução

De acordo com o Ministério da Agricultura, Pecuária e Abastecimento, o Brasil consolidou sua posição como um dos maiores produtores e fornecedores de alimentos para o mundo. Além dos fatores naturais, como extensão territorial agricultável e clima propício, menciona-se o investimento em tecnologia e pesquisa como componentes deste bom resultado [MAPA 2019]. Tendo em mente o progresso desse setor, a agricultura digital têm se consolidado de maneira expressiva, sendo caracterizada pelo uso e desenvolvimento de soluções tecnológicas relacionadas à área da Informática com o objetivo de melhorar a produção agrícola, obtendo mais qualidade e produtividade [Tang et al. 2002]. No Brasil, em especial, as técnicas de Processamento de Imagem e de Visão Computacional já figuram explicitamente como componentes integrantes dos eixos estratégicos a serem desenvolvidos e adotados nas próximas duas décadas [Embrapa 2014].

As técnicas previamente mencionadas já colaboram, por exemplo, na detecção de doenças por análise foliar em diversas espécies, na classificação de diferentes tipos de pragas, na análise da qualidade de alimentos e na inspeção automática de sementes [Vibhute and Bodhe 2012]. Mais recentemente, considerando as contribuições de *Deep Learning* nas tarefas de Visão Computacional, muitos resultados também têm emergido na agricultura onde, segundo um recente *survey*, identificou-se que em diferentes cenários, as técnicas de *Deep Learning* ultrapassam a performance das estratégias anteriores, que ainda existem desafios no tocante à disponibilidade e variabilidade das bases de dados e que há muito potencial de uso de tais técnicas em tarefas agrícolas de classificação e previsão [Kamilaris and Prenafeta-Boldú 2018].

No contexto agrícola, entre os fatores bióticos capazes de proporcionar redução no rendimento das culturas, encontram-se as plantas daninhas, que podem afetar a produção agrícola e econômica, causando um aumento significativo nos custos de produção [Christoffoleti and Passini 1999]. Elas consistem em espécies vegetais que se desenvolvem onde não são desejadas e competem com as culturas pelos recursos naturais, como nutrientes do solo, água, luz e espaço [da Silva et al. 1999]. Também são responsáveis por causar mais danos que benefícios às culturas de interesse do produtor e, além de resultar em perdas na lavoura e propiciar a sobrevivência de pragas e doenças, também podem prejudicar a colheita, provocando desde ferimentos nas mãos dos trabalhadores até a inviabilização de processos mecanizados [Melhorança and Karam 2019].

As práticas utilizadas para o controle de plantas daninhas estão agrupadas em cinco métodos, a citar: controle preventivo, cultural, físico/mecânico, biológico e químico. Estes métodos visam evitar perdas devido à competição por recursos, beneficiar as condições de colheita, evitar o aumento da infestação e proteger o ambiente [de Oliveira Jr. et al. 2011]. Ainda assim, é muito comum que a tomada de decisão de controle de plantas daninhas baseie-se em inspeções visuais, e essa subjetividade pode impedir uma recomendação mais precisa da estratégia de controle [Voll et al. 2003].

Com a agricultura digital, tem-se a motivação para o desenvolvimento de soluções tecnológicas mais precisas e seguras, cujo benefício colabora para o aumento da produtividade e rentabilidade por meio da otimização de processos e custos. Nesta perspectiva, o presente trabalho tem por objetivo propor e avaliar os resultados obtidos com comitês de redes neurais convolucionais, aplicados ao problema de classificação de ervas daninhas e de culturas agrícolas de interesse comercial, a partir de imagens de mudas das respectivas espécies. A adoção de classificadores baseados em redes neurais convolucionais profundas organizados segundo comitês, deu-se pelo potencial de melhor desempenho destes em relação aos classificadores individuais, o que mostra-se efetivo em decorrência da investigação de diferentes estratégias de agregação dos resultados.

Para apresentar o que se propõe, este trabalho está organizado como segue. Uma breve fundamentação teórica sobre aprendizado por comitê encontra-se descrita na Seção 2. As pesquisas afins disponíveis na literatura encontram-se detalhadas na Seção 3. Em seguida, os materiais e métodos utilizados para condução das atividades deste trabalho encontram-se descritos na Seção 4. Os resultados obtidos são mostrados e discutidos na Seção 5. Por fim, as considerações finais são listadas na Seção 6.

2. Aprendizado por Comitê

Nas tarefas de classificação mediante aprendizado supervisionado, tem-se por objetivo construir modelos classificadores com boas capacidades indutivas a partir de exemplos e

rótulos fornecidos com base em um conjunto de dados de treinamento. Ao invés de se restringir a um único modelo para tal tarefa, o Aprendizado por Comitê (do inglês, *Ensemble Learning*) visa agregar diferentes classificadores e combiná-los com o intuito de obter um classificador cujo desempenho supera a dos modelos individuais para o mesmo fim [Rokach 2010]. Diversos relatos na literatura mostram o bom desempenho desta técnica de aprendizado em problemas de detecção e rastreamento de objetos, segmentação e análise de cenas, reconhecimento de imagens, recuperação da informação, Bioinformática, dentre outros [Zhang and Ma 2012].

A proposição de comitês para a resolução de problemas baseia-se em três pilares: (1) a seleção/amostragem dos dados; (2) o treinamento dos classificadores integrantes do comitê; e (3) as estratégias de combinação/agregação dos classificadores [Polikar 2012]. De maneira geral, a pedra angular da proposição de bons comitês reside na diversidade, de modo que os modelos componentes introduzam vieses distintos, os quais podem ser estrategicamente combinados com vistas a diminuir o erro geral [Polikar 2006, Rokach 2010].

Algumas razões que justificam o uso de comitês de modelos são: (i) a capacidade destes em reduzir a variância em decorrência da agregação, ainda que os modelos individuais mantenham um viés relativamente similar, o que é especialmente útil para classificação de exemplos ainda não vistos; (ii) a possibilidade de lidar com um volume de dados maior do que pode ser efetivamente manipulado por um único classificador; (iii) a possibilidade de usar estratégias de reamostragem de dados para treinamento de múltiplos classificadores, nos casos de bases de dados muito pequenas; e (iv) os benefícios da divisão-e-conquista, em que cada classificador pode se aprimorar em uma parte de um problema mais amplo, contribuindo de forma mais especializada para uma melhor solução geral [Polikar 2006].

3. Trabalhos Relacionados

A tarefa de classificação de mudas de plantas em imagens com o uso de técnicas de *Machine Learning* e *Deep Learning* já foi abordada por outros autores da literatura, considerando diferentes modelos e eventualmente fazendo uso de técnicas de segmentação e extração de características em imagens. Alguns destes trabalhos encontram-se descritos a seguir, os quais foram selecionados por levarem em conta a mesma base de dados utilizada neste trabalho, conforme será descrito posteriormente.

O trabalho de Dyrmann et al., por exemplo, considera a tarefa de classificação multi-classe com vistas a distinguir 8 espécies, das quais 2 são culturas de interesse e as demais são plantas daninhas [Dyrmann et al. 2018]. Nesta abordagem, tem-se a classificação das espécies de plantas como um todo, e também de suas folhas individualmente. Na metodologia adotada pelos autores, foi feita a segmentação das informações relativas ao solo, a separação individual das folhas e a utilização de classificadores baseados em Máquinas de Vetores de Suporte. Apoiados por um método de correção dos potenciais erros, a aferição dos resultados da classificação para as plantas resultou em acurácia de até 96,3%, e para as folhas de 96,7%.

O trabalho de Alimboyong et al., considera classificação de 12 espécies de mudas em imagens, as quais compreendem 3 culturas de interesse e 9 plantas daninhas [Alimboyong et al. 2018]. Inicialmente, os autores realizaram um pré-processamento nas imagens disponíveis, visando a conversão do espaço de cores, efetuaram um redimensionamento das imagens e realizaram procedimentos de *data augmentation*, partindo de 4.234 exemplos para 118.750. Os dados resultantes foram então divididos em partições

de treinamento (70%), validação (20%) e teste (10%). Após esta etapa, seguiu-se o treino da rede neural convolucional AlexNet, a qual demonstrou acurácia superior a 99,69% no conjunto de teste. Nota-se, porém, que os resultados podem ter sofrido um viés positivo decorrente dos dados artificialmente aumentados nos conjuntos de teste e validação.

Evoluindo o trabalho anterior, o mesmo grupo de autores considerou a não adoção de técnicas de *data augmentation* e a proposição de uma arquitetura baseada na rede AlexNet, mas com apenas uma camada completamente conectada anterior à camada de saída [Alimboyong et al. 2019]. Mantendo a mesma proporção entre dados de treinamento, validação e teste, os resultados da avaliação deste modelo no conjunto de teste denotaram acurácia de 90.15% a qual, segundo os autores, foi alcançada com menos parâmetros em relação ao trabalho anterior, mostrando um bom potencial de uso prático.

De maneira geral, embora as métricas de desempenho mostrem-se muito positivas, percebe-se ainda a exploração de poucas arquiteturas canônicas de redes neurais convolucionais, ponderações acerca do uso de *data augmentation* e ainda uma lacuna quanto ao uso de comitês de modelos para este problema. Algumas destas questões serão endereçadas na solução proposta, cuja metodologia de obtenção encontra-se detalhada a seguir.

4. Metodologia

A tarefa de aprendizado supervisionado considerada no escopo deste trabalho é a de classificação binária, em que consideram-se classes disjuntas de plantas de culturas de interesse e ervas daninhas. A técnica de validação cruzada do tipo *holdout* foi adotada para treinar e testar os modelos propostos, considerando a proporção de 70% dos dados disponíveis para treinamento, 10% para validação e os 20% restante para testes. Para esta tarefa, considerou-se propor modelos, avaliá-los individualmente, mas também combiná-los segundo comitês sujeitos à diferentes estratégias de agregação.

Em relação ao treino, que foi conduzido exclusivamente com a partição de dados previamente especificada para este fim, adotou-se uma estratégia baseada em validação cruzada *k-fold*, em que k modelos foram propostos e cada um deles foi treinado com $k - 1$ *folds*. Adotou-se $k = 5$ em virtude dos custos computacionais do treino dos modelos, de tal modo que cada treinamento a ser conduzido utilizou apenas uma parte dos dados disponíveis, observando-se a existência de algumas interseções entre os exemplos vistos pelos diferentes modelos. Após os treinamentos individuais, houve a combinação dos k modelos em um comitê. A estratégia de partição dos dados adotada teve em vista promover diversidade no viés induzido por cada modelo [Polikar 2012].

Quanto às métricas de desempenho para aferir os resultados obtidos, foram escolhidas a acurácia e o *F-Score*. Estas métricas foram coletadas nos modelos individuais e também nos diferentes comitês propostos, considerando exclusivamente o desempenho no tocante aos dados disponíveis no conjunto de testes.

4.1. Dados Experimentais

Para endereçar o problema em questão considerou-se os exemplos disponíveis na base de dados *Plant Seedlings Dataset*. Esta base foi construída por meio de uma colaboração entre as universidades dinamarquesas de *Aarhus Universitet Flakkebjerg* e *Syddansk Universitet* e contempla a captura de imagens de mudas de plantas ao longo de um período de 20 dias, com coletas de imagens repetidas em intervalos de 2 a 3 dias a fim de ilustrar diferentes estágios de crescimento [Giselsson et al. 2017]. Esta base contém 960 imagens de 12 espécies distintas de plantas, com diferentes tamanhos e fases de crescimento conforme ilustrado na Figura 1.

Figura 1: Exemplos oriundos do *Plant Seedlings Dataset*. Fonte: [Giselsson et al. 2017]



Na ocasião da criação desta base de dados, a ênfase principal dos autores consistia em fomentar a proposição de algoritmos capazes de distinguir espécies de interesse das plantas daninhas, com vistas a colaborar no desenvolvimento de técnicas automatizadas de eliminação de tais interferências. Neste sentido, a base contém exemplos de 3 culturas comumente encontradas nos campos aráveis dinamarqueses e de amplo interesse comercial (milho, trigo e beterraba), além de 9 tipos de ervas daninhas. A Tabela 1 sumariza os nomes comuns, nomes científicos e quantidade de exemplos destas plantas na base de dados.

Tabela 1: Espécies encontradas na base de dados [Giselsson et al. 2017].

Nome Comum	Nome Científico	Exemplos
Amor de Hortelão	<i>Galium aparine</i>	335
Beterraba	<i>Beta vulgaris</i>	463
Bolsa de Pastor	<i>Capsella bursa-pastoris</i>	274
Milho	<i>Zea mays</i>	257
Trigo	<i>Tricicum aestivum</i>	253
Apera Sedosa	<i>Apera spica-venti</i>	762
Ansarina Branca	<i>Chenopodium album</i>	538
Granza dos Campos	<i>Sinapis arvensis</i>	452
Cauda de Raposa	<i>Alopecurus myosuroides</i>	309
Camomila Selvagem	<i>Matricaria perforata</i>	607
Erva de São Roberto	<i>Geranium pusillum</i>	576
Erva Estrela	<i>Stellaria media</i>	713

Após esta etapa inicial, os proponentes desta base aplicaram algoritmos de Visão Computacional, especialmente de segmentação, visando manter apenas um exemplo de planta em cada imagem. Desta forma, a base de dados resultante, em sua segunda versão, contém então 5.539 exemplos, em escala de cores RGB com alturas e larguras variadas, ocupando em torno de 1.7 GB em disco. Esta segunda versão foi utilizada no escopo deste trabalho.

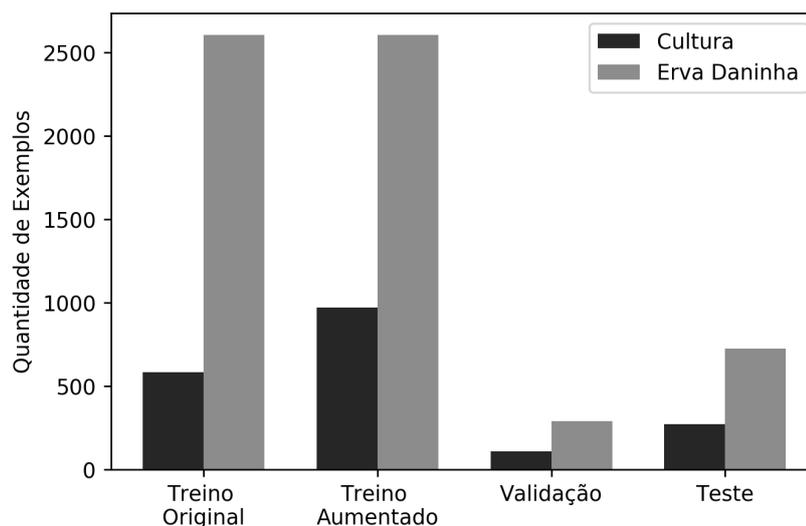
4.2. Preparação da Base de Dados

Para a preparação da base de dados, foram utilizadas técnicas de redimensionamento, *data augmentation* e normalização. No tocante ao redimensionamento, todas as imagens foram ajustadas para 256×256 pixels. Em seguida, os exemplos de milho, trigo e beterraba

foram agrupados na classe de culturas de interesse (classe positiva) e os exemplos remanescentes foram agrupados na classe relativa às ervas daninhas (classe negativa). Nota-se que as classes encontram-se desbalanceadas. A separação nas partições de treino, validação e teste foi feita de maneira aleatória, respeitando-se as proporções previamente estabelecidas.

Tendo em vista o desbalanceamento e, em decorrência, a menor quantidade de exemplos disponíveis para a classe positiva na etapa de treinamento, optou-se por adotar estratégias de *data augmentation*, baseadas principalmente em rotações e espelhamentos, e no estabelecimento de um limiar de 450 exemplos por espécie. Nota-se que os exemplos de teste e de validação não foram sujeitos ao aumento artificial de dados, visando manter verossimilhança com o cenário real no âmbito da avaliação a ser realizada posteriormente. A Figura 2 ilustra o quantitativo de exemplos em cada uma das partições a serem utilizadas para treinamento, teste e validação.

Figura 2: Histograma da distribuição de exemplos nas partições de treino, teste e validação.



Após o aumento dos dados de treinamento, ressalta-se que houve uma seleção aleatória e subdivisão destes dados em *5-folds* disjuntos, com vistas a viabilizar a estratégia concebida para treinamento dos modelos. Por fim, quanto à normalização dos dados para fase de treinamento, escalonou-se os valores dos *pixels* do intervalo $[0,255]$ para o intervalo $[0,1]$. Esta normalização é tipicamente sugerida pela literatura com vistas a promover uma atualização dos pesos no gradiente descendente com vistas à convergência [Chollet 2018].

4.3. Modelo, Arquiteturas e Comitês

Considerando o papel protagonista das redes neurais convolucionais em diversas tarefas de Visão Computacional do estado da arte [Khan et al. 2018], este foi o modelo escolhido para endereçar a tarefa em questão, todavia, levando em conta as diferentes arquiteturas possíveis para as mesmas, especialmente aquelas já consolidadas na literatura e com bom desempenho em tarefas de Visão Computacional, *tradeoffs* entre o número de camadas e

parâmetros treináveis, tal como os custos computacionais necessários para o treinamento devem ser ponderados.

Neste contexto, para o escopo do presente trabalho, optou-se por utilizar redes neurais convolucionais com menos parâmetros quando comparadas à soluções canônicas da literatura. Assim, as arquiteturas escolhidas foram:

1. **LeNet**. Proposta inicialmente para a tarefa de reconhecimento de dígitos manuscritos, esta rede é composta por 2 camadas convolucionais seguidas por camadas de *max pooling* com vistas a extração de características. Por fim, uma última camada convolucional é seguida de duas camadas completamente conectadas para classificação da saída [LeCun et al. 1998];
2. **AlexNet**. Almejando o uso de uma arquitetura de rede neural convolucional com bom desempenho reportado nos trabalhos relacionados, esta rede é composta por 5 camadas convolucionais iniciais e 3 camadas completamente conectadas ao final para produzir a classificação. Possui também camadas intermediárias de *dropout* e *max pooling* [Krizhevsky et al. 2012];
3. **MobileNet**. Visando a utilização em dispositivos móveis e embarcados, esta rede neural convolucional baseia-se em operações de convolução separáveis em profundidade, o que diminui o ônus das operações a serem realizadas nas primeiras camadas [Howard et al. 2017];
4. **ShuffleNet**. Baseia-se em duas operações introduzidas pelos autores, as chamadas convoluções em grupo, que são múltiplas convoluções em que cada uma abrange uma porção dos canais de entrada, e os embaralhamentos de canais, que aleatoriamente misturam os canais de saída das convoluções em grupo. De acordo com seus proponentes, esta arquitetura possui baixo custo computacional ao passo que mantém boa acurácia [Zhang et al. 2017];
5. **EffNet**. Assemelha-se às redes MobileNet e ShuffleNet no tocante ao uso de operações de convolução separáveis em profundidade, mas introduz um novo bloco convolucional que reduz o ônus computacional ao passo que ultrapassa o desempenho do estado da arte para algumas bases de dados amplamente conhecidas [Freeman et al. 2018].

Considerando as arquiteturas previamente mencionadas, tem-se o número de parâmetros treináveis, não-treináveis e totais conforme apresentado na Tabela 2. Nota-se que o somatório do número de parâmetros totais é inferior ao das arquiteturas VGG16 e VGG19 [Simonyan and Zisserman 2015], o que enseja que a combinação das mesmas, segundo um comitê, pode vir a ser menos onerosa que arquiteturas bem estabelecidas na literatura, mediante os termos considerados. As arquiteturas consideradas foram treinadas conforme a metodologia previamente descrita e avaliadas individualmente perante o desempenho no conjunto de teste.

Tabela 2: Parâmetros das redes neurais convolucionais consideradas.

Arquitetura	Parâmetros Treináveis	Parâmetros Não-Treináveis	Parâmetros Totais
LeNet	25.869.426	0	25.869.426
AlexNet	34.068.794	0	34.068.794
MobileNet	3.209.026	21.888	3.230.914
ShuffleNet	1.900.754	39.120	1.939.874
EffNet	634.466	2.240	636.706
Total	65.682.466	63.248	65.745.714

Além do uso e da avaliação das arquiteturas individuais, três técnicas de agregação foram utilizadas na composição de comitês das mesmas, a citar: (1) votação por maioria; (2) votação mediada por uma Máquina de Vetores de Suporte (SVM, do inglês *Support Vector Machine*); e (3) votação mediada por uma Rede Neural Artificial (RNA) *multilayer perceptron*. No caso das estratégias 2 e 3, o treinamento dos modelos mediadores da votação foi feito com a integralidade dos dados disponíveis no conjunto de treinamento, seguido da avaliação no conjunto de testes. Um *kernel* linear foi utilizado na SVM e apenas uma camada oculta com 2 neurônios, função de ativação ReLU e otimizador Adam foram utilizados na RNA. Não houve testes ou busca em *grid* com demais parâmetros e hiperparâmetros.

A implementação da preparação da base de dados, do treino e teste das arquiteturas das redes neurais convolucionais e dos modelos mediadores da votação foi conduzida utilizando a linguagem Python, os *frameworks* Keras¹ e Sci-Kit Learn² e o ambiente de desenvolvimento Jupyter Notebooks³. A plataforma Google Cloud deu suporte ao processamento necessário, na qual foi alocada uma máquina virtual com processador de 16 núcleos virtuais, 48 GB de memória principal e 8 GB de GPU para aceleração das operações de obtenção dos parâmetros ajustáveis.

5. Resultados e Discussão

Após a execução da metodologia proposta, aferiu-se o desempenho das arquiteturas individualmente, segundo as métricas de acurácia e F-Score. Estes resultados encontram-se sintetizados na Tabela 3, ordenados de maneira decrescente pela acurácia.

Tabela 3: Desempenho individual dos modelos de redes neurais convolucionais. As siglas VP, VN, FP e FN denotam, respectivamente, valores verdadeiros positivos, verdadeiros negativos, falsos positivos e falsos negativos.

Modelo	VP	VN	FP	FN	Acurácia	F-Score
<i>MobileNet</i>	237	715	33	9	0,9577	0,9186
<i>AlexNet</i>	224	706	46	18	0,9356	0,8750
<i>EffNet</i>	148	700	122	24	0,8531	0,6697
<i>LeNet</i>	88	650	182	74	0,7424	0,4074
<i>ShuffleNet</i>	173	386	97	338	0,5623	0,4430

É interessante notar que embora nenhuma técnica de pré-processamento tenha sido aplicada às imagens com o intuito de remover características do solo ou acentuar os elementos relativos às folhas, a exemplo de cores ou aspectos da morfologia, observa-se que três das melhores redes obtiveram acurácia superior a 85%, com a *MobileNet*, em especial, alcançando o maior valor dentre as redes observadas. No caso das redes *LeNet* e *ShuffleNet*, em particular, se consideradas as proporções entre as classes e os aspectos de precisão e revocação, tem-se um baixo desempenho aferido pelo F-Score.

Ao analisar o desempenho obtido com a rede *MobileNet* em comparação com os experimentos apresentados em [Alimboyong et al. 2019], nota-se a utilização de cerca de um décimo do número de parâmetros da *AlexNet* em sua versão original. Apesar disso,

¹<http://keras.io>

²<http://scikit-learn.org>

³<https://jupyter.org/>

há que se considerar que houve o uso de *data augmentation* no treino da MobileNet, que esta teve acesso a apenas 4/5 dos dados de treino e que as tarefas de classificação são distintas, sendo a aqui considerada, de classificação binária, e a do trabalho relacionado multi-rótulo, o que impede comparações mais diretas entre as acurácias observadas em ambos os casos. Todavia, o resultado aqui alcançado com uma rede compacta, especialmente projetada para dispositivos embarcados, amplia a possibilidade de utilização de tal resultado em um contexto prático apoiado por *smartphones*.

Após esta etapa, partiu-se então para a elaboração dos comitês, treinamento dos mediadores de votação, quando foi o caso, e teste. O desempenho dos comitês de modelos é reportado na Tabela 4. Ressalta-se que todos os comitês utilizaram todas as redes obtidas no passo anterior.

Tabela 4: Desempenho dos comitês de redes neurais convolucionais. As siglas utilizadas são as mesmas da Tabela 3.

Comitê	VP	VN	FP	FN	Acurácia	F-Score
Votação por Maioria	201	715	69	9	0,9215	0,8375
Votação Mediada por SVM	252	713	18	11	0,9708	0,9421
Votação Mediada por RNA	243	716	25	10	0,9647	0,9328

Os resultados da agregação dos modelos individuais pelos comitês possuem todos acurácia superior a 92% na fase de testes, mostrando-se superior à acurácia média dos modelos individuais (81,02%). O desempenho do comitê com votação por maioria talvez tenha sido viesado negativamente em decorrência da baixa precisão da rede LeNet e da baixa revocação da rede ShuffleNet. Quando considerada uma votação mediada por SVM e RNA, estes comitês foram capazes de minimizar tais vieses e produzir uma agregação melhor, o que culminou em uma acurácia maior do que nos modelos individuais. O melhor resultado observado contemplou a utilização de SVM como mediador de votação com acurácia de 97%.

Os bons resultados obtidos pelos comitês foram reflexo da diversidade no treinamento em decorrência dos dados disponibilizados à cada modelo, do uso de arquiteturas distintas de redes neurais convolucionais e também do bom desempenho individual de algumas das redes na tarefa de classificação binária considerada. Entretanto, considerando-se as duas redes com baixo desempenho individual, melhorias mediante técnicas de *fine-tuning* ou teste de outras arquiteturas ainda mostram-se possíveis.

6. Considerações Finais

Este trabalho teve por objetivo investigar a utilização de comitês de redes neurais convolucionais sujeitos à diferentes estratégias de votação aplicados ao problema de identificação de ervas daninhas e culturas de interesse em imagens de mudas, obtidas nos 20 primeiros dias de plantio. Para endereçar esta tarefa, foi utilizada a base de dados *Plant Seedlings Dataset* e considerou-se uma tarefa de classificação binária. O treinamento foi conduzido de modo a induzir vieses diferentes em arquiteturas distintas de redes neurais convolucionais com vistas ao posterior cancelamento dos mesmos na fase de agregação.

Foram treinadas e testadas cinco arquiteturas de redes neurais convolucionais para o problema em questão, a citar: LeNet, ShuffleNet, AlexNet, EffNet e MobileNet, tendo esta última obtido melhor desempenho entre elas, com acurácia de 95,77%. Todas as redes

foram combinadas em comitês sujeitos à três estratégias de votação, por maioria, mediada por RNA e mediada por SVM, em que este último comitê alcançou 97,08% de acurácia na tarefa considerada. Além do bom desempenho para a tarefa proposta, observou-se que o total de parâmetros treináveis do melhor comitê foi inferior ao das arquiteturas canônicas VGG16 e VGG19, amplamente utilizadas em tarefas de Visão Computacional.

De maneira geral, os comitês propostos para as tarefas possuíram um bom desempenho na etapa de testes, o que enseja um bom potencial de uso em cenários práticos sob condições análogas. Os trabalhos relacionados na literatura para a mesma base de dados consideraram uma classificação multi-rótulo, o que dificulta análises comparativas de desempenho com os resultados aqui propostos. Apesar desta dificuldade, ressalta-se que abordar o problema como uma tarefa de classificação binária pode facilitar a utilização e adaptação dos comitês de modelos aqui propostos para outros contextos, diminuindo o ônus de especialistas humanos na anotação de exemplos, o que pode ser particularmente útil mediante outras culturas de interesse e tipos distintos de plantas daninhas que podem ali incidir.

Ressalta-se ainda que além da proposição e avaliação dos comitês de modelos, o resultado individualmente obtido pela MobileNet pode vir a ser embarcado em dispositivos móveis, conforme propósito original desta arquitetura. Isto pode ser particularmente útil no contexto prático da agricultura, em que muitas avaliações desta natureza precisam ser realizadas em campo.

Como trabalhos futuros, almeja-se a melhoria dos modelos individuais, que obtiveram baixo F-Score, mediante técnicas de *fine-tuning*; o treino, teste e avaliação de outras arquiteturas de redes neurais convolucionais com baixo número de parâmetros; e a análise da viabilidade e desempenho do uso de técnicas de segmentação, seguidas de extração de características e posterior classificação com modelos clássicos de Aprendizado de Máquina, como uma alternativa à classificação com redes neurais convolucionais.

Agradecimentos

Os autores agradecem o apoio financeiro e material provido pela Fundação de Amparo à Pesquisa do Estado do Amazonas (FAPEAM) e pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) por meio do Programa PPP Edital 04/2017.

Referências

- Alimboyong, C. R., Hernandez, A. A., and Medina, R. P. (2018). Classification of plant seedling images using deep learning. In *IEEE TENCON Region 10 Conference*, pages 1839–1844, Coréia do Sul. IEEE.
- Alimboyong, C. R., Hernandez, A. A., and Medina, R. P. (2019). An improved deep neural network for classification of plant seedling images. In *IEEE 15th International Colloquium on Signal Processing & its Applications*, pages 217–222, Malásia. IEEE.
- Chollet, F. (2018). *Deep Learning with Python*. Manning Publications, Nova Iorque, Estados Unidos, 1 edition.
- Christoffoleti, P. J. and Passini, T. (1999). Manejo integrado de plantas daninhas na cultura do feijão. FANCELLI, A. L.; DOURADO NETO, D. Feijão irrigado: estratégias básicas de manejo.
- da Silva, A. A., Silva, J. F., Ferreira, F. A., Ferreira, L. R., and Silva, R. R. (1999). *Controle de plantas daninhas*. Abeas, Brasília, DF, 1 edition.

- de Oliveira Jr., R. S., Constantin, J., and Inoue, M. H. (2011). *Biologia e manejo de plantas daninhas*. Omnipax, Brasil, 1 edition.
- Dyrmann, M., Christiansen, P., and Midtiby, H. S. (2018). Estimation of plant species by classifying plants and leaves in combination. *J Field Robotics.*, 35:202–212.
- Embrapa (2014). Visão 2014–2034 – o futuro do desenvolvimento tecnológico da agricultura brasileira (síntese).
- Freeman, I., Roese-Koerner, L., and Kummert, A. (2018). Effnet: An efficient structure for convolutional neural networks. <http://arxiv.org/abs/1801.06434>.
- Giselsson, T. M., Jørgensen, R. N., Jensen, P. K., Dyrmann, M., and Midtiby, H. S. (2017). A public image database for benchmark of plant seedling classification algorithms.
- Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., and Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. <http://arxiv.org/abs/1704.04861>.
- Kamilaris, A. and Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, 147:70–90.
- Khan, S., Rahmani, H., Shah, S. A. A., and Bennamoun, M. (2018). *A Guide to Convolutional Neural Networks for Computer Vision*. Morgan & Claypool, Austrália, 1 edition.
- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems*, pages 1097–1105, Estados Unidos. Association for Computing Machinery.
- LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324.
- MAPA (2019). Ministério da agricultura, pecuária e abastecimento, panorama do agronegócio brasileiro. Disponível em <http://www.agricultura.gov.br>. Acesso em 26 de jun. de 2019.
- Melhorança, A. L. and Karam, D. (2019). Tratos culturais. Disponível em <http://bit.ly/2xaSIk7>. Acesso em 25 de jun. de 2019.
- Polikar, R. (2006). Ensemble based systems in decision making. *IEEE Circuits and Systems Magazine*, 6(3):21–45.
- Polikar, R. (2012). *Ensemble Learning*, pages 1–34. Springer Science+Business Media, Estados Unidos. Capítulo do livro: “*Ensemble Machine Learning: Methods and Applications*”, (C. Zhang & Y. Ma, eds.).
- Rokach, L. (2010). Ensemble-based classifiers. *Artif Intell Rev*, 33:1–39.
- Simonyan, K. and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. pages 1–14, Estados Unidos.
- Tang, S., Zhu, Q., Zhou, X., Liu, S., and Wu, M. (2002). A conception of digital agriculture. In *IEEE International Geoscience and Remote Sensing Symposium*, pages 3026–3028, Canada. IEEE.
- Vibhute, A. and Bodhe, S. K. (2012). Applications of image processing in agriculture: A survey. *International Journal of Computer Applications*, 52(2):34–40.

- Voll, E., Adegas, F. S., Gazziero, D. L. P., Brighenti, A. M., and de Oliveira, M. C. N. (2003). Amostragem do banco de sementes da flora emergente de plantas daninhas. *Pesquisa Agropecuária Brasileira*, 38(2):211–218.
- Zhang, C. and Ma, Y. (2012). *Ensemble Machine Learning: Methods and Applications*. Springer Science+Business Media, Estados Unidos.
- Zhang, X., Zhou, X., Lin, M., and Sun, J. (2017). Shufflenet: An extremely efficient convolutional neural network for mobile devices. <http://arxiv.org/abs/1707.01083>.