

**Anais da XV Jornada
de Iniciação Científica da
Embrapa Amazônia Ocidental**

*Empresa Brasileira de Pesquisa Agropecuária
Embrapa Amazônia Ocidental
Ministério da Agricultura, Pecuária e Abastecimento*

Anais da XV Jornada de Iniciação Científica da Embrapa Amazônia Ocidental

*Everton Rabelo Cordeiro
Eduardo Ossamu Nagao
Inocencio Junior de Oliveira
Jony Koji Dairiki
Maria Geralda de Souza
Ronaldo Ribeiro de Moraes
Editores Técnicos*

Embrapa
Brasília, DF
2019

Exemplares desta publicação podem ser adquiridos na:

Embrapa Amazônia Ocidental

Rodovia AM-010, Km 29,
Estrada Manaus/Itacoatiara,
Manaus, AM
69010-970

Caixa Postal 319

Fone: (92) 3303-7800

Fax: (92) 3303-7820

www.embrapa.br

www.embrapa.br/fale-conosco/sac

**Unidade responsável pelo
conteúdo e edição**

Embrapa Amazônia Ocidental

Comitê de Publicações da Unidade

Presidente: *Cheila de Lima Boijink*

Secretária-executiva: *Gleise Maria*

Teles de Oliveira

Membros: *Maria Augusta Abtibol Brito
de Sousa, Maria Perpétua Beleza Pereira
e Marcos Vinícius Bastos Garcia*

Revisão de texto

Maria Perpétua Beleza Pereira

Normalização bibliográfica

Maria Augusta Abtibol Brito de Sousa
(CRB 11/420)

Capa, projeto gráfico e editoração
eletrônica

Gleise Maria Teles de Oliveira

1ª edição

Publicação digital (2019)

Todos os direitos reservados.

A reprodução não autorizada desta publicação, no todo ou em parte,
constitui violação dos direitos autorais (Lei nº 9.610).

Dados Internacionais de Catalogação na Publicação (CIP).

Embrapa Amazônia Ocidental.

Jornada de Iniciação Científica da Embrapa Amazônia Ocidental (14. : 2018: Manaus, AM).
Anais da XV Jornada de Iniciação Científica da Embrapa Amazônia Ocidental; editores,
Everton Rabelo Cordeiro... [et al.], editores técnicos. – Brasília, DF: Embrapa, 2019.

PDF (143 p.).

ISBN 978-85-7035-948-3

1. Iniciação científica. 2. Comunicação científica. 3. Pesquisa. I. Cordeiro, Everton Rabelo. II. Nagao, Eduardo Ossamu. III. Oliveira, Inocencio Junior de. IV. Dairiki, Jony Koji. V. Souza, Maria Geralda de. VI. Morais, Ronaldo Ribeiro de. VII. Título. VIII. Embrapa Amazônia Ocidental.

CDD 630.72

Técnicas de aprendizagem de máquina aplicadas à identificação de cultivares de guaranazeiro por meio da morfologia de venação de folíolos

Allex de Lima Sousa¹

Marcos Filipe Alves Salame²

Resumo – O guaranazeiro (*Paullinia cupana* var. *sorbilis*) é uma planta nativa amazônica com um fruto amplamente utilizado em indústrias de bebidas e energéticos. Entretanto, a proliferação de pragas e doenças locais tem reduzido a produtividade dessa espécie na Amazônia, o que motivou a Empresa Brasileira de Pesquisa Agropecuária (Embrapa) a lançar 19 cultivares resistentes aos principais fatores que afetam a guaranaicultura na região. Contudo, a distinção dessas cultivares ainda depende de métodos manuais e passíveis de falha humana, e este trabalho apresenta uma abordagem capaz de realizar a distinção dessas cultivares utilizando o padrão de nervura das folhas por meio de métodos de aprendizagem de máquinas. Como resultado, foi alcançada uma acurácia superior a 85%, indicando que a nervura central é um descritor válido quando se trata de classificação dessas cultivares desenvolvidas pela Embrapa.

Termos de indexação: aplicativo móvel, classificação por meio de folhas, aprendizado de máquina.

¹Bolsista de Iniciação Científica, Paic/Fapeam/Embrapa Amazônia Ocidental, Manaus, AM.

²Engenheiro de computação, M.Sc. em Ciência da Computação, analista da Embrapa Amazônia Ocidental, Manaus, AM.

Machine learning techniques applied to the identification of guarana plant cultivars using vein patterns

Abstract – The guarana plant (*Paullinia cupana* var. *sorbilis*) is a native Amazon plant with a fruit widely used in the refrigerants and energy drinks industries. However, the proliferation of local pests and diseases has reduced the productivity of this species in the Amazon, which motivated Embrapa to launch 19 resistant cultivars to the main factors that affect this biodiversity in the region, but the distinction of these cultivars still depends on manual methods besides being susceptible to human failure. Thus, this work presents an approach capable of distinguishing these cultivars using the leaf vein pattern, through machine learning methods. As a result, an accuracy greater than 85% was reached, indicating that the central vein pattern is a valid descriptor when it comes to the classification of these cultivars developed by Embrapa.

Index terms: mobile app, leaf classification, machine learning.

Introdução

O guaranazeiro (*Paullinia cupana* var. *sorbilis*) é uma planta nativa do bioma Amazônia e popular no Brasil devido as suas propriedades gastronômicas e energéticas. Entretanto, índices históricos da produção nacional do guaraná destacam o estado da Bahia como o maior produtor do fruto (IBGE, 2013; Silva et al., 2016).

Dentre as principais causas dessa baixa produtividade de guaranazeiros na Amazônia, há a ausência de uma diversidade genética suficientemente resistente à proliferação de pragas e doenças. Esse cenário não é novo e motivou, no final da década de 1970, a Embrapa a iniciar um programa de melhoramento genético do guaranazeiro, o qual objetivava a expansão de produção do guaraná na Amazônia por meio de pesquisas que desenvolveram e disponibilizaram 19 cultivares resistentes aos principais fatores que afetam a guaranaicultura na região Amazônica (Tricaud et al., 2016).

Contudo, o método utilizado para a distinção dessas cultivares desenvolvidas ainda é bastante manual e passível de falha humana. Há também um problema em relação à distribuição das mudas dessas cultivares, uma vez que, sem um especialista presente, elas podem ser facilmente confundidas entre si. Assim, surge a possibilidade de agregar métodos computacionais de reconhecimento de padrões para auxiliar nesse processo.

A utilização de técnicas de aprendizagem de máquinas para a classificação de plantas a partir de descritores morfológicos, como folhas, por exemplo, possui considerável abundância na literatura em consequência da importante contribuição que o reconhecimento de espécimes vegetais, por sistemas computacionais, oferece ao setor agrônomo e computacional. Nesse cenário, este trabalho objetivou explorar métodos de identificação das cultivares de guaranazeiro da Embrapa por meio de imagens da superfície abaxial de folíolos, utilizando técnicas de aprendizagem profunda e clássicas. Ambas são

categorias de técnicas de aprendizagem de máquinas, uma subárea da inteligência computacional.

Material e Métodos

Ainda é comum a ausência de imagens das folhas de cultivares de guaranazeiro da Embrapa em bases de imagem públicas, que disponibilizam seus dados para pesquisas na área de aprendizado de máquina e visão computacional. Assim, foi necessário realizar a coleta das folhas em campo e, em seguida, fazer a aquisição das imagens por fotografias e escaneamento de folíolos amostrados, conforme ilustra a Figura 1. Após essa etapa, o conjunto de dados foi organizado e validado pelos pesquisadores que atuam com as cultivares de guaranazeiro na Embrapa. Ao todo, foram coletadas amostras foliares das cultivares BRS-Maués, BRS-Luzéia e CG-612.

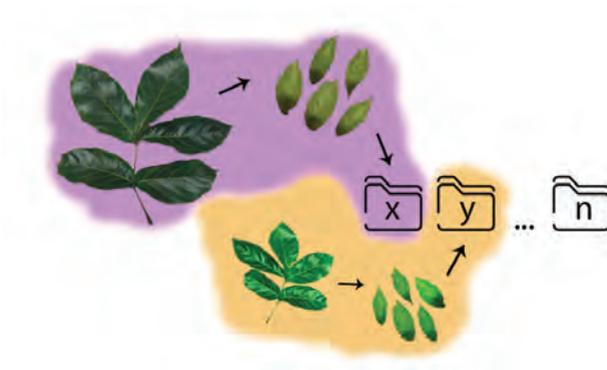


Ilustração: Marcos Filipe Alves Salame

Figura 1. Após a coleta das folhas, os folíolos são destacados da ráquis. Em seguida, a superfície abaxial de cada folíolo é fotografada e escaneada. Por fim, as imagens são organizadas para formar o conjunto de dados utilizado neste trabalho.

Essas variedades são resistentes aos principais fatores que afetam guaranazeiros oriundos de sementes e possuem alta produtividade de guaraná. Não obstante, a quantidade de cultivares selecio-

nada (três) visa experimentar uma primeira abordagem de classificação não binária com essa espécie, ou seja, explorar uma metodologia diferente de trabalhos anteriores.

Já com o conjunto de dados preparado, algumas técnicas de Processamento Digital de Imagens (PDI) foram utilizadas antes das imagens seguirem para os algoritmos de classificação. Baseando-se na metodologia proposta por Larese et al. (2014) e Grinblat et al. (2016), apenas a região central da superfície abaxial dos folíolos coletados foi mantida no conjunto de dados, de modo a preservar a morfologia da venação dos folíolos, conforme apresenta a Figura 2. Todas as imagens foram convertidas do espectro RGB para tons de cinza visando diminuir a influência de anomalias e cores, uma vez que se busca utilizar apenas o padrão da nervura central dos folíolos como alvo do processo de aprendizagem da máquina.

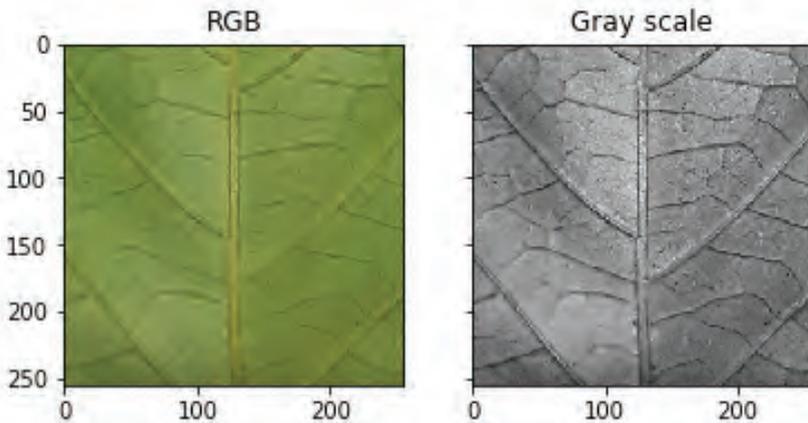


Ilustração: Marcos Filipe Alves Salame

Figura 2. Exemplo da nervura central da superfície abaxial de um folíolo. À direita, a mesma imagem da esquerda após a conversão para tons de cinza e que será utilizada na base de dados final.

A base final utilizada neste trabalho conta com 426 imagens de 256 x 256 *pixels*, fotografadas e escaneadas, organizadas com as seguintes classes: BRS-Maués (152), BRS-Luzéia (145) e CG-612 (129) conforme especificações no parágrafo anterior.

Máquinas de Vetores de Suporte (*Support Vector Machines – SVM*) foi a técnica de classificação explorada neste estudo, devido a sua ampla utilização na literatura. A implementação dessa técnica ocorreu por meio da linguagem de programação *Python*, mediante a biblioteca *Scikit-learn*. Adotou-se também o uso de validação cruzada *K-Fold*, com $k=10$ e uma estratégia de divisão de *folds* randômica e estratificada (i.e., todos os *folds* com a mesma quantidade de dados).

Os experimentos foram implementados por JupyterLab 0.31.12 Beta e executados em máquina com sistema operacional GNU/Linux Ubuntu 16.04 LTS, com 64 GiB de memória RAM, uma GPU Nvidia Quadro K1200 e uma CPU Intel Xeon E3-1225 v5, com 4 núcleos físicos de 3,3GHz, cada.

Resultados

A otimização dos hiperparâmetros, conduzida pela função *Grid-SearchCv* da biblioteca *Scikit-learn*, identificou os parâmetros $\gamma=1e-5$ e $C=100.0$ como a melhor combinação para o SVM, utilizando a *Radial Basis Function* (RBF) como *kernel*. A acurácia obtida, a partir da validação cruzada com 10 *folds*, foi de $85.17 \pm 2.99\%$. A performance de aprendizado do modelo pode ser observada na Figura 3.

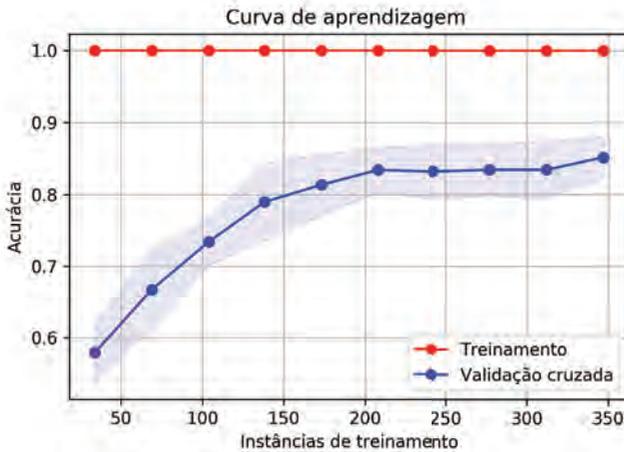


Figura 3. Performance média das técnicas de classificação utilizadas.

Discussão

Os resultados alcançados com o SVM indicam a presença de um padrão real na morfologia da nervura central dos folíolos. Dessa forma, uma nova possibilidade de identificação dessas cultivares surge, visto que esse tipo de abordagem não é apresentado no manual disponibilizado pelo Ministério da Agricultura, Pecuária e Abastecimento do Brasil, em 2010. Essa nova perspectiva para classificação de espécies vegetais está sendo vista cada vez mais na literatura, já que o padrão de venação dos folíolos é tido como uma espécie de “impressão digital” de unidades foliares.

Conclusões

Os recentes avanços em hardware e áreas como a aprendizagem de máquinas já viabilizam participação sólida em processos de identificação de espécies vegetais por meio de descritores morfológicos, como folhas. Dentre as vantagens oferecidas, esse tipo de

abordagem ajuda a diminuir o tempo do processo e erro humano, uma vez que muitas características necessitam de uma análise mais robusta, devido à alta similaridade das características das folhas entre cultivares. Em trabalhos futuros, espera-se ampliar o estudo com algoritmos de aprendizagem profunda de máquinas e outros métodos clássicos de classificação. Além disso, técnicas como *transfer learning* e *ensemble learning* podem ser úteis em uma abordagem com Redes Neurais Convolutivas.

Agradecimentos

À Fundação de Amparo à Pesquisa do Estado do Amazonas (Fapeam), pela bolsa de pesquisa, e aos pesquisadores Firmino José do Nascimento Filho, André Luiz Atroch e Adauto Maurício Tavares, pelas informações técnicas referentes ao guaraná.

Referências

GRINBLAT, G. L.; UZAL, L. C.; LARESE, M. G.; GRANITTO, P. M. Deep learning for plant identification using vein morphological patterns. **Computers and Electronics in Agriculture**, v. 127, p. 418-424, 2016.

IBGE. **Produção agrícola municipal**: culturas temporárias e permanentes. 2013. Disponível em: <<http://www.ibge.gov.br/home/estatistica/economia/pam/2013/default.shtm>>. Acesso em: 02 maio 2018.

LARESE, M. G.; NAMÍAS, R.; CRAVIOTTO, R. M.; ARANGO, M. R.; GALLO, C.; GRANITTO, P. M. Automatic classification of legumes using leaf vein image features. **Pattern Recognition**, v. 47, n. 1, p. 158-168, 2014.

SILVA, E. F. da; SOUSA, S. B. de; SILVA, G. F. da; SOUSA, N. R.; NASCIMENTO FILHO, F. J. do; HANADA, R. E. TRAP and SRAP markers to find genetic variability in complex polyploid *Paullinia cupana* var. *sorbilis*. **Plant Gene**, v. 6, p. 43-47, 2016.

TRICAUD, S.; PINTON, F.; PEREIRA, H. Saberes e práticas locais dos produtores de guaraná (*Paullinia cupana* Kunth var. *sorbilis*) do médio Amazonas: duas organizações locais frente à inovação. **Boletim do Museu Paraense Emílio Goeldi**. Ciências Humanas, v. 11, n. 1, p. 33-53, 2016.