

**Banco de imagens de lâminas foliares de espécies do Cerrado: técnicas de extração e classificação de imagens por visão computacional**

Deborah R. Bambil<sup>1,\*</sup>, Hemerson Pistori<sup>2</sup>, Francielli Bao<sup>3</sup>, Ieda M. Bortolotto<sup>1</sup>

<sup>1</sup> PPG Biologia Vegetal, Universidade Federal de Mato Grosso do Sul (UFMS), Campus de Campo Grande, 79070-900, Campo Grande, MS Brasil.

<sup>2</sup> INOVISÃO, Universidade Católica Dom Bosco (UCDB), 79117-900, Campo Grande, MS Brasil.

<sup>3</sup> Instituto de Biociências, Universidade Estadual Paulista (UNESP), Campus de Rio Claro, 13506-900, Rio Claro, SP, Brasil.

O artigo será enviado para a revista Biotropica.

## RESUMO

As características morfológicas das folhas de Angiospermas são importantes para identificar espécies usar a visão computacional para auxiliar o trabalho dos taxonomistas na extração de caracteres morfológicos pode ser uma solução para fazer a identificação vegetal. O objetivo deste estudo foi verificar a eficácia da visão computacional na classificação de lâminas foliares de espécies arbóreas e arbustivas do Cerrado através de caracteres (atributos) morfológicos. Foram coletadas 40 folhas de 30 espécies arbóreas e arbustivas, as imagens foram feitas de ambas as faces, sendo 80 imagens com câmera do celular e 80 com scanner, totalizando por espécie 160 imagens. Dessas imagens foram extraídos 226 atributos de cor, forma e textura pelo software Inovtaxon para fazer a classificação testando os algoritmos: *SVM*, *AdaBoost*, *Random Forest*. Os resultados mostraram que há eficácia da visão computacional na identificação de lâminas foliares, os aparelhos de celular e scanner obtiveram resultados similares. O algoritmo *SVM* e *Random Forest* apresentaram melhor desempenho comparado ao *Adaboost* que obteve um baixo desempenho na classificação mostrando que suas características de treinamento não foram a melhor solução para classificar as lâminas foliares. Concluímos que a diferença dos dispositivos de captura de imagens celular e scanner não foram significativos e o algoritmo *SVM* mostrou melhor desempenho para classificar o banco de imagens de lâminas foliares de espécies arbóreas e arbustivas do Cerrado.

Palavras-chave: Aprendizagem de máquina; Atributos; Cor; Forma; Textura;

Algoritmos; Identificação; Classificação; Morfologia; Taxonomia.

## INTRODUÇÃO

A identificação e classificação de plantas com base em caracteres morfológicos tem sido utilizadas desde Teofrasto, há cerca de 300 a. C. (Barroso 1986). Mesmo com o progresso contínuo da classificação biológica (taxonomia), que inclui caracteres moleculares, genéticos e etno-ecológicos, além da metagenoma e da metaboloma para simplificar e ordenar a imensa diversidade de vida em unidades coerentes chamadas taxa (Ruggiero *et al.* 2015), ainda hoje a morfologia vegetal se mantém como um atributo importante na identificação e classificação de plantas.

Para um especialista em determinado táxon, a identificação botânica faz parte da sua rotina em função do seu conhecimento científico, mas a nomenclatura botânica também se faz importante para outras áreas de estudo como para química, genética, medicina, microbiologia, farmacologia e farmácia (Bennett & Balick 2014). A identificação de plantas pelo público em geral que não faz uso científico, nem sempre é relevante ou interessante (Stagg & Donkin 2013), a solução pode ser feita tornando a identificação taxonômica mais atrativa (Hershey 2002), pois é relevante saber qual espécie dispõe para fazer uso medicinal, alimentício, bem como evitar espécies que possuem substâncias tóxicas que podem causar incidentes.

É possível facilitar e expandir a identificação botânica através da extração de caracteres morfológicos usando outra área de pesquisa: a visão computacional que desenvolve bases matemáticas e algorítmicas pelas quais informações do ambiente ou objeto são extraídas de imagens para análise, processamento digital e reconhecimento de padrões por métodos analíticos e estatísticos (Ponce & Forsyth 2012), que vêm empregando seu uso em distintas áreas. No contexto da biologia vegetal, está sendo utilizado na classificação de pólenes, através da extração de atributos de cor, textura e

forma (Gonçalves *et al.* 2016), em estudos fenológicos (Mariano *et al.* 2016), medição automática de predação na área foliar (Machado *et al.* 2016), dentre outros.

O uso da visão computacional permite o desenvolvimento de técnicas de aferição e extração de informações de forma automatizada de lâminas foliares (Lee & Chen 2006). O estudo foliar implica na coleta de inúmeros atributos como a largura e o comprimento, cor, textura e forma (Filho *et al.* 2014). Porém, a folha pode apresentar grande variação em seu padrão estrutural dentro de uma mesma espécie em função de fatores ambientais como solo e iluminação (Plotze *et al.* 2005), dificultando a identificação da espécie. A variação da quantidade de luz influencia diretamente no tamanho e na forma de folhas, o que gera inúmeras dificuldades em torno da identificação (Joppa *et al.* 2011). Respostas assimétricas conduzem à plasticidade fenotípica e não somente a luz pode causar este efeito, por conta desta plasticidade, até mesmo as folhas de um mesmo indivíduo podem diferir quando situados em determinados ambientes (Guerin *et al.* 2012).

O Cerrado é considerado um *hotspot* de biodiversidade (Myers *et al.* 2000) e apresenta uma variedade de fisionomias, como o Cerrado *sensu stricto*, Cerradão e Mata de Galeria (Simon *et al.* 2009), essas fisionomias são caracterizadas por formações campestres, savânicas e florestais (Durigan *et al.* 2007), compreendendo uma rica flora, com mais de 7000 espécies, o que requer maiores esforços para preservação e conservação deste bioma que já apresenta espécies em extinção (Klink & Machado 2005), desta forma, um enfoque na identificação e classificação de espécies por folhas pode facilitar e otimizar os esforços amostrais. Além disso, as diferentes fisionomias oferecem espécies que ora podem ser encontradas em locais abertos e ora em locais fechados e, assim recebem diferentes intensidades de luz, isso afeta a forma como as espécies usam esses ambientes (Ribeiro & Walter 2008), este cenário se torna um

desafio para os sistemas computacionais (Lee & Chen 2006), que podem contribuir para conservação e preservação da diversidade de espécies.

Os bancos de dados de folhas coletadas em áreas com grande diversidade biológica, bem organizados podem constituir um repositório de informações para estudos botânicos. A possibilidade de se formarem bancos de dados com informações de identificação de espécies através da captura de imagens requer um esforço para elaboração, onde primeiramente deve-se ter conhecimento de técnicas de coleta de imagem, extração de atributos e classificação (Joly *et al.* 2014).

Para isso comparar dispositivos de captura de imagens é importante devido a recorrente evolução tecnológica, possibilitando a escolha adequada de acordo com sua qualidade, resolução de imagem e adaptabilidade (Hasinoff & Kutulakos 2011). Provar algoritmos de aprendizagem possibilita verificar o melhor desempenho na classificação, analisar os resultados e dentre eles escolher o melhor método para usar em trabalhos decorrentes.

Diante desse preceito, o objetivo deste estudo foi verificar a eficácia da visão computacional para identificação de lâminas foliares de espécies arbóreas e arbustivas do Cerrado, propondo um método de classificação de lâminas de folhas utilizando atributos de caracteres morfológicos. Para atingir tal objetivo, foram elaboradas as seguintes questões: (i) Existe diferença nos resultados de imagens digitalizadas com scanner e fotos de câmera de celular para aplicação na identificação de espécies com base na classificação de lâminas foliares? (ii) Qual algoritmo de aprendizagem automática (*Support vector machine – SVM, Adaptive Boosting – AdaBoost, Random Forest*) apresenta melhor desempenho para classificar lâminas de folhas usando atributos de cor, forma e textura?

## MATERIAL E MÉTODOS

### *Área de estudo*

O estudo foi realizado com lâminas foliares coletadas na Reserva Particular de Patrimônio Natural (20°27'S, 54°37'W) da Universidade Federal de Mato Grosso do Sul - UFMS, localizada no município de Campo Grande, MS - Brasil. Essa reserva tem área de 40 hectares, e possui diferentes fisionomias: Cerrado, Cerradão e Mata de Galeria (Bueno *et al.* 2013). O clima é AW savanna tropical chuvoso (Koeppen 1948), com temperatura média anual de 23°C (Embrapa 1985).

### *Coleta das lâminas foliares em campo*

Em uma área delimitada de 0.5 ha foram selecionadas 30 espécies aleatórias de árvores e arbustos com diâmetro à altura do peito (DAP)  $\geq 4,77$  cm (Tabela 1; Fig.1). Foram realizadas coletas de 40 amostras referentes à menor unidade foliar observável (folha simples, folíolo ou foliólulo), evitando a necessidade de conhecimento prévio sobre o morfologia de cada espécie. As espécies foram previamente identificadas no estudo de Bueno *et al.* (2013) e exemplares estão depositados no herbário CGMS (Universidade Federal de Mato Grosso do Sul).

TABELA 1. Lista das 30 espécies de árvores e arbustos do Cerrado que foram coletadas amostras para o experimento, a nomenclatura segue o APGIII (2009).

Família	Nome científico	Voucher (CGMS)
	<i>Annona crassiflora</i> Mart.	24932
Annonaceae	<i>A. coriacea</i> Mart.	24931
	<i>Xylopia aromatica</i> (Lam.) Mart.	24900
Anacardiaceae	<i>Astronium fraxinifolium</i> Schott.	24898
Bignoniaceae	<i>Tabebuia aurea</i> (Silva Manso) Benth & Hook.f ex S.Moore.	24933
Caryocaraceae	<i>Caryocar brasiliense</i> Cambess.	24934
Chrysobalanaceae	<i>Licania humilis</i> Cham. & Schltdl.	24891
Combretaceae	<i>Terminalia argentea</i> Mart.	24879
Connaraceae	<i>Connarus suberosus</i> Planch.	24893
Dilleniaceae	<i>Curatella americana</i> L.	24926
	<i>Andira cuyabensis</i> Benth.	24902

	<i>Bowdichia virgilioides</i> Kunt.	24906
Fabaceae	<i>Copaifera langsdorffii</i> Desf.	24922
	<i>Dimorphandra mollis</i> Benth.	24923
	<i>Diptychandra aurantiaca</i> Tul.	24903
	<i>Stryphnodendron obovatum</i> Benth.	24904
Lythraceae	<i>Lafoensia pacari</i> A. St.-Hil.	24917
	<i>Byrsonima coccolobifolia</i> Kunth.	24883
Malpigiaceae	<i>B. verbascifolia</i> (L.) DC.	24882
Malvaceae	<i>Eriotheca pubescens</i> (Mart. & Zucc.) Schott & Endl.	24877
	<i>Pseudobombax</i> sp	24880
Melastomataceae	<i>Miconia albicans</i> (Sw.) Triana.	24916
Myrtaceae	<i>Eugenia</i> sp	24823
Opiliaceae	<i>Agonandra brasiliensis</i> Miers ex Benth. & Hook. f.	24895
Rhamnaceae	<i>Rhamnidium elaeocarpum</i> Reissek.	22982
Sapindaceae	<i>Matayba guianenses</i> Aubl.	24878
Sapotaceae	<i>Chrysophyllum marginatum</i> (Hook. & Arn.) Radlk.	24885
	<i>Qualea grandiflora</i> Mart.	24910
	<i>Qualea multiflora</i> Mart.	24913
Vochysiaceae	<i>Salvertia convallariodora</i> A. St.-Hil.	24875

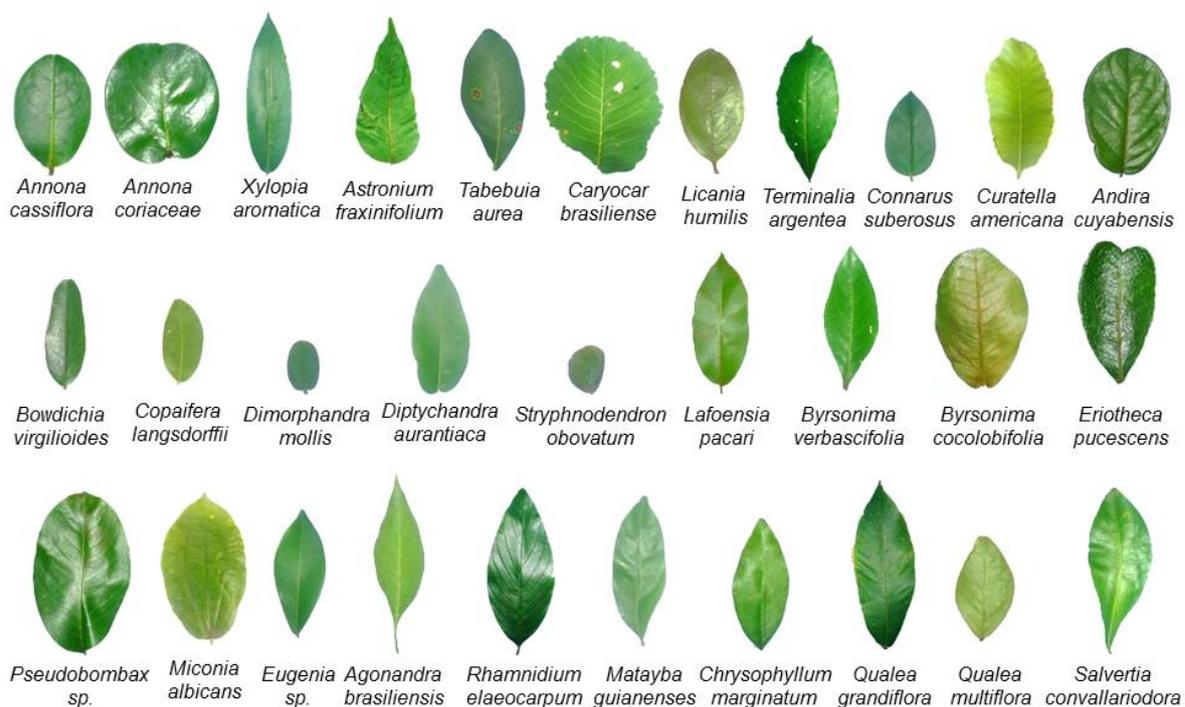


FIGURA 1. Imagens meramente ilustrativas invariantes a escala, das 30 espécies amostradas de árvores e arbustos do Cerrado.

*Captura e processamento das imagens*

Para avaliação do método de captura de imagens foi utilizado um aparelho de celular (Kumar *et al.* 2012) e um scanner (Plotze *et al.* 2005). As imagens foram feitas de ambas as faces adaxial e abaxial das lâminas foliares. O aparelho de celular modelo lumia 820, com câmera de 8 mega pixels, foi instalado em suporte com ângulo de 90° a 20 cm das folhas, com fundo branco, flash do aparelho, a resolução da imagem foi de 2000 x 3552 pixels (Fig.2A). Da mesma forma foi feito com o scanner modelo HP Scanjet G2710, ambas as faces das lâminas foliares foram digitalizadas, com fundo branco e resolução de imagem de 2550 x 3510 pixels (Fig.2B). Foram capturadas 160 imagens de cada espécie, sendo 80 com aparelho celular e 80 com scanner totalizando em um banco com 4.800 imagens. O material foi processado fresco para manter seus aspectos originais e após a captura de imagens foi realizada a extração e a classificação dos atributos.

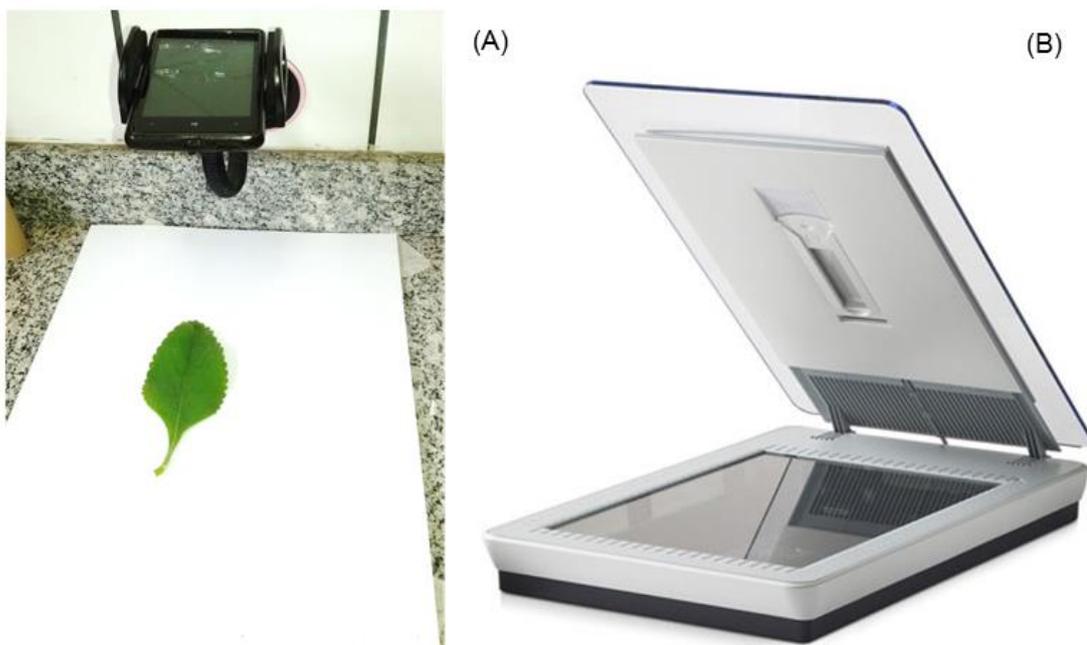


FIGURA 2. Aparelhos usados para captura de imagens. (A) aparelho de celular Lumia fixo no suporte e (B) Scanner HP Scanjet G2710.

#### *Extração de Atributos das imagens*

Foram extraídos 226 atributos de cada imagem (Tabela 2) usando o software Inovtaxon desenvolvido pelo grupo Inovisão que extrai características das imagens. Para classificar as lâminas foliares, foram extraídos os atributos de cor, forma e textura.

Dentro do espaço de cores foram extraídos os padrões RGB que define R como vermelho; G como verde e B como azul, utilizado para especificar as cores primárias (Kruse *et al.* 2014), o HSV especifica H como Matiz sendo a associação das principais cores, como azul, vermelho, amarelo, verde e roxo; S como saturação indica que quanto maior a saturação mais viva é a cor e V como valor classifica a cor quanto clara ou escura, assim quanto maior o valor, mais clara é a cor (Szeliski 2010, Filho *et al.* 2014). CIELab são padrões de cores que indicam L como luminosidade, a\* como as coordenadas cromáticas entre vermelho e verde e b\* como as coordenadas cromáticas entre amarelo e azul (Kruse *et al.* 2014).

Com relação aos atributos de forma, foram extraídos os 7 momentos de Hu que são características invariantes à escala, translação e rotação. A importância de analisar estes momentos invariantes se caracteriza por poder identificar a região da folha independente da escala, translação e rotação. Também foram extraídos histogramas de gradientes orientados – HOG que representa um descritor de conteúdo de imagem e vem sendo usado com a pretensão de encontrar objetos, por meio de uma busca por mudanças de tonalidade (Arafat *et al.* 2016).

Os atributos de textura extraídos foram *Gray level Co-occurrence* - GLCM que usa um método estatístico para avaliar valores em tons de cinza em determinado espaço, formulando uma matriz resultando em diferentes medidas caracterizando a textura da imagem (Filho *et al.* 2014). Padrão binário local – LBP é um descritor que usa um conceito de padrão uniforme (Zhou *et al.* 2012) onde o código de um pixel é obtido por comparação com a intensidade dos pixels vizinhos para classificar a textura em escala

de cinza (He *et al.* 1981), assim, possibilita lidar com a variação de iluminação. Este método tem sido muito utilizado para extração de elementos como folhas (Pahikkala *et al.* 2015).

TABELA 2. Descrição dos 226 atributos de Cor, Forma e Textura que foram extraídos das imagens das lâminas foliares.

COR	FORMA	TEXTURA
cor_rmin	hu_0	glcm_cont_1_0
cor_rmax	hu_1	glcm_cont_1_45
cor_rmediamedia	hu_2	glcm_cont_1_90
cor_rdesvio	hu_3	glcm_cont_2_0
cor_gmin	hu_4	glcm_cont_2_45
cor_gmax	hu_5	glcm_cont_2_90
cor_gmedia	hu_6	glcm_diss_1_0
cor_gdesvio	hog_0 À 127	glcm_diss_1_45
cor_bmin		glcm_diss_1_90
cor_bmax		glcm_diss_2_0
cor_bmedia		glcm_diss_2_45
cor_bdesvio		glcm_diss_2_90
cor_hmin		glcm_homo_1_0
cor_hmax		glcm_homo_1_45
cor_hmedia		glcm_homo_1_90
cor_hdesvio		glcm_homo_2_0
cor_smin		glcm_homo_2_45
cor_smax		glcm_homo_2_90
cor_smedia		glcm_asm_1_0
cor_sdesvio		glcm_asm_1_45
cor_vmin		glcm_asm_1_90
cor_vmax		glcm_asm_2_0
cor_vmedia		glcm_asm_2_45
cor_vdesvio		glcm_asm_2_90
cor_cielmin		glcm_ener_1_0
cor_cielmax		glcm_ener_1_45
cor_cielmedia		glcm_ener_1_90
cor_cieldesvio		glcm_ener_2_0
cor_cieamin		glcm_ener_2_45
cor_cieamax		glcm_ener_2_90
cor_cieamedia		glcm_corr_1_0
cor_cieadesvio		glcm_corr_1_45
cor_ciebmin		glcm_corr_1_90
cor_ciebmax		glcm_corr_2_0
cor_ciebmedia		glcm_corr_2_45
cor_ciebdesvio		glcm_corr_2_90

### *Classificação da imagem com algoritmos de aprendizagem de máquina*

Para classificação dos atributos foram testados três algoritmos de aprendizagem de máquina, com a configuração padrão do Software Weka 3.7.9, sendo eles: *SVM*, *AdaBoost* e *Random Forest* (Gonçalves *et al.* 2016, Arafat *et al.* 2016), para validar a aprendizagem de máquina usamos validação cruzada – VC 10 dobras (Gonçalves *et al.* 2016).

O *SVM* é um algoritmo que resolve problemas de classificação de imagens por regiões e otimiza o tempo para aprendizagem, desempenhando uma metodologia para resolver problemas de programação quadrática da máquina de vetores de suporte (Mattila *et al.* 2013). O *SVM*, em problemas de classificação binários, delimita um hiperplano separando as classes, a distância desse hiperplano é definida por uma margem estabelecida por vetores de suporte, separando o conjunto de dados (Kremic & Subasi 2016). O banco de imagens de folhas de Cerrado possui múltiplas classes. O método usado pelo *SVM* neste caso, foi a classificação emparelhada com 1 versus 1, estimando probabilidades de classe para cada par de classes, acoplando estimativas em conjunto (Hastie & Tibshirani 1998).

O *AdaBoost* é um algoritmo de reforço, baseado em meta-heurística, que classifica os dados particionando em instâncias, para tanto usa múltiplos classificadores (Kumar *et al.* 2008, Freund & Schapire 1996). Ao se deparar com instâncias classificadas negativamente, atribui pesos, para que no próximo ciclo possa fazer novas tentativas classificando corretamente, construindo um classificador forte (Malhi *et al.* 2013). Assim, se adapta positivamente a novas tentativas de acertar.

O *Random Forest* usa um método de aprendizagem que proporciona precisão, executando com eficiência quando trabalha com uma base de dados robusta (Manfredi *et al.* 2014, Breiman 2001). Esse classificador faz a aprendizagem por conjuntos de árvores que recebem treinamento para montar florestas aleatórias de decisão (Kremic & Subasi 2016), classificando o que for essencial para explorar os dados e mesmo quando ocorre formação de modelos fracos esses são reunidos para aprendizagem e formação de um modelo forte (Corcoran *et al.* 2013).

#### *Análise de dados*

Os resultados foram analisados com a métrica de Percentual de Classificação Correta – PCC que mostrou o resultado geral na classificação dos algoritmos. E também foi usada a métrica da medida – F para avaliar o desempenho dos algoritmos na classificação de cada classe que corresponde a uma espécie, sua composição é a média de outras duas métricas: precisão e revocação (López-García *et al.* 2010):

$$F = 2 * \frac{\text{precisão} * \text{revocação}}{\text{precisão} + \text{revocação}}$$

A formula da métrica de precisão é composta por porcentagem de verdadeiro positivo – VP que corresponde a amostra positiva, ou seja, a amostra pertence a classe e o classificador classificou a amostra positivamente. Falso positivo – FP é quando a amostra não pertence à classe, mas o classificador afirma que pertence classificando a amostra como positivo (Kruse *et al.* 2014):

$$\text{Precisão} = \frac{VP}{VP + FP}$$

A formula da métrica de revocação é composta por porcentagem de VP e de Falso negativo – FN é quando a amostra pertence a classe e o classificador classifica a amostra como negativo (Kruse *et al.* 2014):

$$\text{Revocação} = \frac{VP}{VP + FN}$$

Para avaliar os resultados da medida-F de cada classe foi construído um modelo de estatística ANOVA no ambiente R de programação estatística (R Development Core Team 2013). Este modelo foi construído para quantificar as variações que ocorreram no experimento e posteriormente o teste de Tukey para fazer análise de comparações das técnicas em pares (Gonçalves *et al* 2016).

Outra forma de visualizar o desempenho individual de classificação de cada espécie é por meio da matriz de confusão. Ela apresenta o número de amostras classificadas (Arafat *et al.* 2016) para cada classe (espécie). O número de acertos de cada classe é observado quando a linha da classe informada corresponde a coluna da classe predita. É possível encontrar claramente o número de acertos de cada classe, localizando a linha diagonal principal da matriz (Arafat *et al.* 2016).

## **RESULTADOS**

### *Métodos de captura de imagens: celular e scanner*

O resultado geral por captura de imagem e por algoritmo de aprendizagem de máquina da classificação de lâminas foliares de espécies arbóreas e arbustivas revela alto desempenho do Percentual de Classificação Correta (PCC), para cada dispositivo (Tabela 3). Tanto o scanner, quanto o celular, apresentaram PCC acima de 95% com algoritmos SVM (98% e 96%) e Random Forest (95% e 97%), respectivamente, com melhor desempenho. Já o AdaBoost em ambos foi baixo o desempenho 6%, não diferindo significativamente nos resultados entre imagens digitalizadas com scanner e fotos de celular (Tabela 3).

TABELA 3. Probabilidades associadas aos métodos de captura de imagens scanner e celular, avaliados por algoritmos de aprendizagem de máquina *SVM*, *Random Forest* e *AdaBoost*.

Comparação de dispositivo de captura de imagem com o mesmo algoritmo de aprendizagem de máquina	Df	P
Scanner ( <i>SVM</i> ) vs. Celular ( <i>SVM</i> )	0,017133333	0,2276976
Scanner ( <i>Random Forest</i> ) vs. Celular ( <i>Random Forest</i> )	0,020100000	0,0977253
Scanner ( <i>AdaBoost</i> ) vs. Celular ( <i>AdaBoost</i> )	0,006733333	0,9513343

*Desempenho dos algoritmos de aprendizagem: Adaboost, Random Forest e SVM para a classificação de imagens usando os de atributos de cor, forma e textura*

Para verificar o desempenho dos algoritmos de aprendizagem automática foi feito o teste de hipótese da medida-F que mostrou que houve uma diferença média significativa no desempenho de classificação entre os algoritmos, (ANOVA:  $F = 8310$ ,  $p < 0,001$ ). Apresentaram melhor desempenho para classificar as lâminas das folhas os algoritmos *SVM* e *Random Forest* (Fig.3). O teste de Tukey apresentou diferença média significativa para os pares *SVM vs. AdaBoost*, *Random Forest vs. AdaBoost*, e para os pares *SVM vs. Random Forest* não houve diferença média significativa em ambos métodos de captura de imagem (Fig.3).

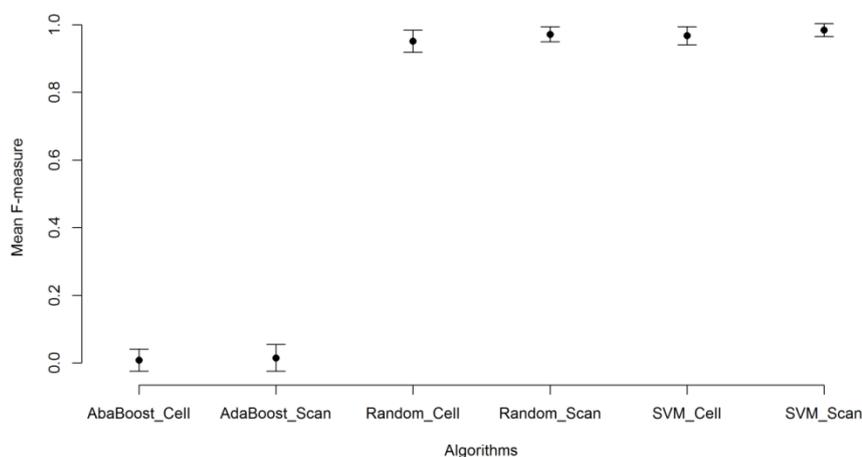


FIGURA 3. Diferença no desempenho dos algoritmos na classificação dos atributos, em ambos os experimentos de captura de imagem.

A classificação individual observada na matriz de confusão mostrou que o melhor desempenho foi com o método de captura de imagem scanner e classificação feita com o algoritmo SVM obtendo resultados de 74 a 80 amostras classificadas nas classes pertencentes (Fig.4), onde 80 é o número total de amostras por classe, ou seja, obtiveram todas as amostras classificadas as classes pertencentes. A matriz do algoritmo *AdaBoost* mostra o baixo desempenho individual de classificação também com o método de captura de imagem scanner, conseguindo classificar apenas quatro espécies em suas classes pertencente, ainda sim com baixo número de amostras (Fig.5).

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	aa	ab	ac	ad	Classes/Espécies		
80	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	a	<i>Pseudobombax</i> sp
0	79	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	b	<i>Bowdichia virgilioides</i>
0	0	80	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	c	<i>Annona coriacea</i>
0	1	0	74	0	0	0	0	0	0	1	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	d	<i>Diptychandra aurantiaca</i>
0	0	0	0	80	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	e	<i>Licania humilis</i>
0	0	0	0	0	77	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	f	<i>Miconia albicans</i>	
0	0	0	0	0	0	80	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	g	<i>Qualea grandiflora</i>
0	0	0	0	0	0	0	78	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	h	<i>Terminalia argentea</i>
0	0	0	0	0	0	0	0	80	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	i	<i>Chrysophyllum marginatum</i>
0	0	0	0	0	0	0	0	0	79	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	j	<i>Qualea multiflora</i>
0	0	0	1	0	0	0	0	0	0	77	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	k	<i>Copaifera langsdorffii</i>
0	0	0	0	0	0	0	0	0	0	0	80	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	l	<i>Byrsonima coccolobifolia</i>
0	0	0	0	0	0	0	0	0	0	0	0	79	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	m	<i>Stryphnodendron obovatum</i>
0	0	0	0	0	0	0	0	0	0	0	0	0	79	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	n	<i>Andira cuyabensis</i>	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	80	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	o	<i>Tabebuia aurea</i>	
0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	76	0	0	0	0	0	0	0	0	0	0	0	0	0	p	<i>Eugenia</i> sp	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	79	0	0	0	0	0	0	0	1	0	0	0	0	q	<i>Salvertia convallariodora</i>	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	80	0	0	0	0	0	0	0	0	0	0	0	r	<i>Rhamnidium elaeocarpum</i>	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	80	0	0	0	0	0	0	0	0	0	0	s	<i>Byrsonima verbascifolia</i>	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	77	0	0	0	0	0	0	0	0	0	t	<i>Annona cassiflora</i>	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	80	0	0	0	0	0	0	0	0	u	<i>Dimorphandra mollis</i>	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	80	0	0	0	0	0	0	0	v	<i>Connarus suberosus</i>	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	80	0	0	0	0	0	0	w	<i>Caryocar brasiliense</i>	
0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	79	0	0	0	0	0	x	<i>Matayba guianenses</i>	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	78	0	0	0	0	y	<i>Xylopia aromatica</i>	
0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	77	0	0	0	0	z	<i>Curatella americana</i>	
0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	77	0	0	0	aa	<i>Astronium fraxinifolium</i>	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	80	0	0	0	ab	<i>Eriotheca pucescens</i>	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	80	0	0	ac	<i>Lafoesia pacari</i>	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	78	0	ad	<i>Agonandra brasiliensis</i>	

FIGURA 4. Matriz de Confusão SVM (scanner). Valores em amarelo mostra valores das amostras classificadas nas suas classes pertencentes.

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	aa	ab	ac	ad	Classes/Espécies	
24	10	0	0	46	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	a	<i>Pseudobombax sp</i>
24	56	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	b	<i>Bowdichia virgilioides</i>
1	40	0	0	24	0	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	c	<i>Annona coriacea</i>	
24	56	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	d	<i>Diptychandra aurantiaca</i>	
2	0	0	0	62	0	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	e	<i>Licania humilis</i>	
23	0	0	0	57	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	f	<i>Miconia albicans</i>	
0	0	0	0	64	0	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	g	<i>Qualea grandiflora</i>	
0	0	0	0	64	0	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	h	<i>Terminalia argentea</i>	
24	56	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	i	<i>Chrysophyllum marginatum</i>	
0	0	0	0	64	0	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	j	<i>Qualea multiflora</i>	
24	52	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	k	<i>Copaifera langsdorffii</i>	
23	56	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	l	<i>Byrsonima coccolobifolia</i>	
24	54	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	m	<i>Stryphnodendron obovatum</i>	
0	0	0	0	64	0	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	n	<i>Andira cuyabensis</i>	
0	5	0	0	59	0	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	o	<i>Tabebuia aurea</i>	
24	56	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	p	<i>Eugenia sp</i>	
2	0	0	0	64	0	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	q	<i>Salvertia convallariodora</i>	
24	4	0	0	52	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	r	<i>Rhamnidium elaeocarpum</i>	
23	55	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	s	<i>Byrsonima verbascifolia</i>	
24	56	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	t	<i>Annona cassiflora</i>	
24	1	0	0	55	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	u	<i>Dimorphandra mollis</i>	
0	0	0	0	64	0	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	v	<i>Connarus suberosus</i>	
12	56	0	0	4	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	w	<i>Caryocar brasiliense</i>	
0	0	0	0	64	0	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	x	<i>Matayba guianenses</i>	
4	0	0	0	63	0	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	y	<i>Xylopia aromatica</i>	
20	0	0	0	56	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	z	<i>Curatella americana</i>	
1	0	0	0	63	0	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	aa	<i>Astronium fraxinifolium</i>	
0	3	0	0	61	0	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	ab	<i>Eriotheca pucescens</i>	
23	56	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	ac	<i>Lafaensia pacari</i>	
24	56	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	ad	<i>Agonandra brasiliensis</i>	

FIGURA 5. Matriz de Confusão *AdaBoost* (scanner). Valores em amarelo mostra valores das amostras classificadas nas suas classes pertencentes.

## Discussão

Nossos resultados evidenciam a eficácia da visão computacional na identificação de lâminas foliares de espécies arbóreas e arbustivas do Cerrado, através do método de extração de atributos de cor, forma e textura, além disso, observamos que o desempenho dos algoritmos *SVM* e *Random Forest* em ambos dispositivos de captura de imagens oferecem alto percentual de classificação para folhas.

Outros estudos que usaram diferentes métodos em visão computacional também comprovaram eficácia na classificação para identificação de espécies arbóreas e arbustivas, como o *LeafSnap*, que faz a identificação a partir da extração de atributos da curvatura da folha e compara com imagens existentes em um banco de imagens, em resposta uma ou mais espécies são sugeridas (Kumar *et al.*2012). Já o *Pl@ntnet* utiliza outro método de classificação, com algoritmo *k-means*, para identificar as espécies por aproximação, também com espécies existentes em um banco de imagens (Joly *et al.* 2014).

A extração de atributos foi de acordo com a avaliação de características morfológicas de espécies arbóreas do Cerrado que ofereceu maior precisão para classificação correta. Assim como, para classificar espécies florestais foram extraídos atributos de variação de forma, disposição, funções e estruturas das folhas, que mostraram ser capazes de lidar com as características de um banco de imagens (Filho et al. 2014), considerando que lâminas foliares não são facilmente reconhecidas com apenas um método, mas sim, pela combinação de métodos, o que possibilita o êxito na classificação correta (Mattila *et al.* 2013).

#### *Métodos de captura de imagens: celular e scanner*

As vantagens dos testes realizados com aparelho de celular giram em torno da perspectiva de posteriormente podermos alimentar um banco de dados que será utilizado na elaboração de um aplicativo de identificação de imagens, como o LeafSnap (Kumar *et al.* 2012), porém usamos maior seleção de atributos. O aparelho de celular por funcionar a bateria oferece a mobilidade que o scanner de mesa não tem, justamente pela necessidade de conexão com computador e rede de energia. Outro fator negativo para o scanner é que as folhas assimétricas, com ondas, característica presente em espécies do Cerrado, são prensadas no dispositivo e perdem essa informação, contudo nosso banco de dados por apresentar a captura de vários atributos chegou a um percentual de acerto similar entre scanner e celular.

Outra questão para ausência de diferenças entre scanner e celular seria, possivelmente, porque os dispositivos foram configurados para captura de imagens em níveis próximos de resolução para uma comparação justa. Ambos tiveram manipulação de luz e fundo branco o aparelho de scanner com a lâmpada interna (Mokhtarian & Abbasi 2004, Sicard *et al.* 2015), e o aparelho de celular com flash, pois a iluminação é um fator fundamental para manter um controle preciso (Sicard *et al.* 2015), o tratamento

de luz influencia diretamente nos resultados referentes aos atributos de cor (Hasinoff & Kutulakos 2011).

Com base nos resultados dos experimentos, para comparar o desempenho dos dispositivos de captura com celular e scanner, respaldamos a escolha do dispositivo a critério do objetivo e adaptabilidade do estudo, pois os equipamentos não apresentaram diferença significativa referente à qualidade de imagens aplicadas na classificação de lâminas foliares.

*Desempenho dos algoritmos de aprendizagem: Adaboost, Random Forest e SVM para a classificação de imagens usando os de atributos de cor, forma e textura*

Os resultados de desempenho dos algoritmos foram essenciais para mostrarmos a eficácia da classificação das imagens e verificar qual obteria melhor desempenho. Comparar as diferentes técnicas de classificação possibilita diminuir problemas que correspondem a escolha do melhor algoritmo para classificar os atributos (López-García *et al.* 2010). Os melhores resultados obtidos foram pelo *SVM* e o *Random Forest* na classificação, utilizando os atributos de cor, forma e textura, mostram a capacidade dos algoritmos para treinar uma alta quantidade de dados e várias entradas de variáveis (Kremic & Subasi 2016).

O melhor desempenho do *SVM* pode ser devido a cada imagem ser composta por uma lâmina foliar, isto facilita a classificação da imagem (Dua *et al.* 2012), livre de ruídos. Este fator possibilita agrupar os dados pertencentes à mesma classe (Mattila *et al.* 2013). Em experimento com espécies florestais Filho *et al.* (2014) também utilizou o *SVM* com imagens individuais das lâminas foliares e conseqüentemente obteve um bom desempenho para classificar as imagens.

O algoritmo *Random Forest* obteve resultados extremamente próximos ao do *SVM*, uma das suas diferenças é que o *Random Forest* tem o potencial de lidar com alta

quantidade de variáveis de entrada e selecionar as consideradas significativas, atribuindo importância a cada variável (Manfredi *et al.* 2014), sendo apto a resolver problemas de classificação, pois consegue equilibrar os erros nos conjuntos de dados (Kremic & Subasi 2016).

O algoritmo *Adaboost* é um algoritmo simples de ser implementado e flexível para ser usado em diversos problemas (Schapire & Singer 1999), não precisa de componentes de hardware pesados para processar as análises e faz uma análise simplificada, mas podendo não ser capaz de resolver problemas complexos ou com muitos parâmetros (Schapire 1999). Usa classificadores fracos para formar um classificador forte e realizar sua função (Kumar *et al.* 2008). Todavia por mais benefícios que suas características de treinamento possam oferecer, não foi a solução para a classificação do nosso banco de imagens, ao produzir classificadores fracos não conseguiu discriminar as classes sobrecarregando o treinamento (Wu & Nagahashi 2013). Diante do baixo resultado no desempenho de classificação, ou a meta heurística adotada pelo padrão do Weka, mostrou não ter sido um pressuposto de problemas interessantes para o algoritmo resolver.

O *SVM* obteve destaque na classificação individual das espécies com um alto número de acertos com todas as classes, e mesmo que as características das folhas possam ser semelhantes entre as espécies, dificultando até mesmo o reconhecimento por botânicos (Kazmi *et al.* 2015). Isso mostrou a capacidade de classificar do *SVM* lidando com variabilidade fenotípica, e diversas variáveis morfológicas (Kremic & Subasi 2016).

## CONCLUSÕES

Os resultados mostraram que existe eficácia da visão computacional na identificação de lâminas foliares, e que a interação com as informações botânicas podem agilizar simplificando o processo e aumentando a rapidez no trabalho em diversas áreas de estudo que necessitam de conhecimento taxonômico, e que hoje demandam de tempo de especialistas e recursos financeiros para chegar a um herbário fazer a comparação com material estéril, esse mesmo material pode ainda ser processado em imagens se tornando dados que venham ser classificados e construir um banco de identificação não só para comparação humana com imagens, mas uma comparação artificial que não exige conhecimentos específico. Não apenas uma solução para botânicos, mas para todos que queiram ter conhecimentos de identificação de espécies, pois sugerimos que o resultado deste experimento incentive iniciativas para compartilhar dessa metodologia em que a visão computacional facilitará a identificação de espécies vegetais para distintas finalidades e regiões, para ciência ou conhecimento popular, as máquinas não iram substituir o trabalho dos botânicos e sim auxilia-los.

## REFERÊNCIAS

- ARAFAT, S. Y., M. I. SAGHIR, M. ISHTIAQ, and U. BASHIR. 2016. Comparison of Techniques for Leaf Classification. 3–8.
- BARROSO, G. M. 1986. Sistemática de angiospermas do Brasil. *In* Viçosa: Universidade Federal de Viçosa. pp. 88–94.
- BENNETT, B. C., and M. J. BALICK. 2014. Does the name really matter? the importance of botanical nomenclature and plant taxonomy in biomedical research. *J. Ethnopharmacol.* 152: 387–392. Available at: <http://dx.doi.org/10.1016/j.jep.2013.11.042>.
- BREIMAN, L. 2001. Random forests. *Mach. Learn.* 45: 5–32.
- BUENO, M. L., D. R. M. NEVES, A. F. SOUZA, E. OLIVEIRA JUNIOR, G. A. DAMASCENO-JUNIOR, V. PONTARA, V. A. LAURA, and J. A. RATTER. 2013. Influence of edaphic factors on the floristic composition of an area of cerradão in the Brazilian central-west. *Acta Bot. Brasilica* 27: 445–455.
- CORCORAN, J. M., J. F. KNIGHT, and A. L. GALLANT. 2013. Influence of multi-source and multi-temporal remotely sensed and ancillary data on the accuracy of random forest classification of wetlands in northern Minnesota. *Remote Sens.* 5: 3212–3238.
- DUA, S., U. RAJENDRA ACHARYA, P. CHOWRIAPPA, and S. VINITHA SREE. 2012. Wavelet-based energy features for glaucomatous image classification. *IEEE Trans. Inf. Technol. Biomed.* 16: 80–87.
- DURIGAN, G., M. F. DE SIQUEIRA, and G. A. D. C. FRANCO. 2007. Threats to the Cerrado remnants of the state of São Paulo, Brazil. *Sci. Agric.* 64: 355–363. Available at: [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0103-90162007000400006&lng=en&nrm=iso&tlng=en](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-90162007000400006&lng=en&nrm=iso&tlng=en).

- FILHO, P. L. P., L. S. OLIVEIRA, S. NISGOSKI, and A. S. BRITTO. 2014. Forest species recognition using macroscopic images. *Mach. Vis. Appl.* 25: 1019–1031.
- FREUND, Y., and R. R. E. SCHAPIRE. 1996. Experiments with a New Boosting Algorithm. *Int. Conf. Mach. Learn.* 148–156. Available at: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.51.6252>.
- GONÇALVES, A. B., J. S. SOUZA, G. G. DA SILVA, M. P. CEREDA, A. POTT, M. H. NAKA, and H. PISTORI. 2016. Feature extraction and machine learning for the classification of Brazilian Savannah pollen grains. *PLoS One* 11: 1–20.
- GUERIN, G. R., H. WEN, and A. J. LOWE. 2012. Leaf morphology shift linked to climate change. *Biol. Lett.* 8: 882–886. Available at: <http://rsbl.royalsocietypublishing.org/content/roybiolett/early/2012/06/28/rsbl.2012.0458.full.pdf>.
- HASINOFF, S. W., and K. N. KUTULAKOS. 2011. Light-efficient photography. *IEEE Trans. Pattern Anal. Mach. Intell.* 33: 2203–2214.
- HASTIE, T., and R. TIBSHIRANI. 1998. Classification by pairwise coupling. *Ann. Stat.* 26: 451–471.
- HE, D., L. NANG, S. SHERBROOKE, C. CENTRE, R. SENSING, and M. ROAD. 1981. And texture. 2769–2772.
- HERSHEY, D. R. 2002. Plant Blindness: “We have Met the Enemy and He is Us.” *Plant Sci. Bull.* 48: 78–85. Available at: <http://www.botany.org/bsa/psb/2002/psb48-3.html>.
- JOLY, A., P. BONNET, S. SELMI, D. BARTH, A. JOLY, P. BONNET, S. SELMI, P. IDENTIFICATION, T. L. CAPPELLATO, N. FERRO, M. HALVEY, and W. K. CLEF. 2014. LifeCLEF Plant Identification Task 2014 e Go ” To cite this version : LifeCLEF Plant Identification Task 2014. 724–737.

- JOPPA, L. N., D. L. ROBERTS, and S. L. PIMM. 2011. How many species of flowering plants are there? *Proc. R. Soc. B* 278: 554–559.
- KAZMI, W., F. GARCIA-RUIZ, J. NIELSEN, J. RASMUSSEN, and H. J. ANDERSEN. 2015. Exploiting affine invariant regions and leaf edge shapes for weed detection. *Comput. Electron. Agric.* 118: 290–299. Available at: <http://dx.doi.org/10.1016/j.compag.2015.08.023>.
- KLINK, C. A., and R. B. MACHADO. 2005. Conservation of the Brazilian Cerrado. *Conserv. Biol.* 19: 707–713.
- KREMIC, E., and A. SUBASI. 2016. Performance of Random Forest and SVM in Face Recognition. 13: 287–293.
- KRUSE, O. M. O., J. M. PRATS-MONTALBÁN, U. G. INDAHL, K. KVAAL, A. FERRER, and C. M. FUTSAETHER. 2014. Pixel classification methods for identifying and quantifying leaf surface injury from digital images. *Comput. Electron. Agric.* 108: 155–165.
- KUMAR, N., P. N. BELHUMEUR, A. BISWAS, D. W. JACOBS, W. J. KRESS, I. C. LOPEZ, and J. V. B. SOARES. 2012. Leafsnap: A computer vision system for automatic plant species identification. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* 7573 LNCS: 502–516.
- KUMAR, N., P. BELHUMEUR, and S. NAYAR. 2008. FaceTracer: A search engine for large collections of images with faces. *In* *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. pp. 340–353.
- LEE, C. L., and S. Y. CHEN. 2006. Classification of leaf images. *Int. J. Imaging Syst. Technol.* 16: 15–23.
- LÓPEZ-GARCÍA, F., G. ANDREU-GARCÍA, J. BLASCO, N. ALEIXOS, and J. M. VALIENTE.

2010. Automatic detection of skin defects in citrus fruits using a multivariate image analysis approach. *Comput. Electron. Agric.* 71: 189–197.
- MACHADO, B. B., J. P. M. ORUE, M. S. ARRUDA, C. V SANTOS, D. S. SARATH, W. N. GONCALVES, G. G. SILVA, H. PISTORI, A. R. ROEL, and J. F. RODRIGUES-JR. 2016. BioLeaf: A professional mobile application to measure foliar damage caused by insect herbivory. *Comput. Electron. Agric.* 129: 44–55. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0168169916307670>.
- MALHI, Y. ET AL. 2013. Introductory Plant Biology. *Comput. Electron. Agric.* 88: 287–301. Available at: <http://www.pnas.org/content/106/48/20359.full.pdf?withds=yes%5Cn> <http://www.pnas.org/cgi/doi/10.1073/pnas.0903410106>.
- MANFREDI, M., C. GRANA, S. CALDERARA, and R. CUCCHIARA. 2014. A complete system for garment segmentation and color classification. *Mach. Vis. Appl.* 25: 955–969.
- MARIANO, G. C., L. P. C. MORELLATO, J. ALMEIDA, B. ALBERTON, M. G. G. DE CAMARGO, and R. DA S. TORRES. 2016. Modeling plant phenology database: Blending near-surface remote phenology with on-the-ground observations. *Ecol. Eng.* 91: 396–408.
- MATTILA, H., P. VALLI, T. PAHIKKALA, J. TEUHOLA, O. S. NEVALAINEN, and E. TYYSTJÄRVI. 2013. Comparison of chlorophyll fluorescence curves and texture analysis for automatic plant identification. *Precis. Agric.* 14: 621–636.
- MOKHTARIAN, F., and S. ABBASI. 2004. Matching shapes with self-intersections: Application to leaf classification. *IEEE Trans. Image Process.* 13: 653–661.
- MYERS, N., R. A. MITTERMEIER, C. G. MITTERMEIER, G. A. B. DA FONSECA, and J. KENT. 2000. Biodiversity hotspots for conservation priorities. *Nature* 403: 853–858. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10706275>.

- PAHIKKALA, T., K. KARI, H. MATTILA, A. LEPISTÖ, J. TEUHOLA, O. S. NEVALAINEN, and E. TYYSTJÄRVI. 2015. Classification of plant species from images of overlapping leaves. *Comput. Electron. Agric.* 118: 186–192. Available at: <http://dx.doi.org/10.1016/j.compag.2015.09.003>.
- PLOTZE, R. DE O., M. FALVO, J. G. PÁDUA, L. C. BERNACCI, M. L. C. VIEIRA, G. C. X. OLIVEIRA, and O. M. BRUNO. 2005. Leaf shape analysis using the multiscale Minkowski fractal dimension, a new morphometric method: a study with *Passiflora* (Passifloraceae). *Can. J. Bot.* 83: 287–301.
- PONCE, J., and D. FORSYTH. 2012. Computer vision: a modern approach Available at: <http://www.inria.fr/centre/paris-rocquencourt/actualites/computer-vision-a-modern-approach>.
- R DEVELOPMENT CORE TEAM. 2013. R Software. *R A Lang. Environ. Stat. Comput.*
- RIBEIRO, J. F., and B. M. T. WALTER. 2008. As principais fitofisionomias do bioma Cerrado. *In Cerrado: Ecologia e flora*. pp. 151–212.
- RUGGIERO, M. A., D. P. GORDON, T. M. ORRELL, N. BAILLY, T. BOURGOIN, R. C. BRUSCA, T. CAVALIER-SMITH, M. D. GUIRY, and P. M. KIRK. 2015. A higher level classification of all living organisms (Correction). *PLoS One* 10: 1–54.
- SCHAPIRE, R. E. 1999. A brief introduction to boosting. *IJCAI Int. Jt. Conf. Artif. Intell.* 2: 1401–1406.
- SCHAPIRE, R. E., and Y. SINGER. 1999. Improved boosting algorithms using confidence-rated predictions. *Mach. Learn.* 37: 297–336.
- SICARD, C., C. GLEN, B. AUBIE, D. WALLACE, S. JAHANSHAHI-ANBUHI, K. PENNING, G. T. DAIGGER, R. PELTON, J. D. BRENNAN, and C. D. M. FILIPE. 2015. Tools for water quality monitoring and mapping using paper-based sensors and cell phones. *Water Res.* 70: 360–369.

- SIMON, M. F., R. GREYER, L. P. DE QUEIROZ, C. SKEMA, R. T. PENNINGTON, and C. E. HUGHES. 2009. Recent assembly of the Cerrado, a neotropical plant diversity hotspot, by in situ evolution of adaptations to fire. *Proc. Natl. Acad. Sci.* 106: 20359–20364. Available at: <http://www.pnas.org/content/106/48/20359.full.pdf?with->  
[ds=yes%5Cnhttp://www.pnas.org/cgi/doi/10.1073/pnas.0903410106](http://www.pnas.org/cgi/doi/10.1073/pnas.0903410106).
- STAGG, B. C., and M. DONKIN. 2013. Teaching botanical identification to adults: experiences of the UK participatory science project “Open Air Laboratories.” *J. Biol. Educ.* 47: 104–110. Available at: <http://www.tandfonline.com/doi/abs/10.1080/00219266.2013.764341>.
- SZELISKI, R. 2010. *Computer Vision : Algorithms and Applications*. Computer (Long Beach, Calif). 5: 832. Available at: [http://research.microsoft.com/en-us/um/people/szeliski/book/drafts/szelski\\_20080330am\\_draft.pdf](http://research.microsoft.com/en-us/um/people/szeliski/book/drafts/szelski_20080330am_draft.pdf).
- WU, S., and H. NAGAHASHI. 2013. A new method for solving overfitting problem of gentle AdaBoost. *Fifth Int. Conf. Graph. Image Process. (ICGIP )* 9069: 6.
- ZHOU, S., Q. LIU, J. GUO, and Y. JIANG. 2012. ROI-HOG and LBP Based Human Detection via Shape Part-Templates Matching. 109–115.