

OBJETIVOS DE
DESENVOLVIMENTO
SUSTENTÁVEL



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

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

Anais da XVI Jornada de Iniciação Científica da Embrapa Amazônia Ocidental

*Cláudia Majolo
Inocencio Junior de Oliveira
Jony Koji Dairiki
Maria Geralda de Souza
Ronaldo Ribeiro de Moraes
Editores Técnicos*

Embrapa
Brasília, DF
2020

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
Everton Rabelo Cordeiro

Secretária-executiva
Gleise Maria Teles de Oliveira

Membros
José Olenilson Costa Pinheiro, Maria Augusta Abtibol Brito de Sousa e Maria Perpétua Beza Pereira

Revisão de texto
Maria Perpétua Beza Pereira

Normalização bibliográfica
Maria Augusta Abtibol Brito de Sousa

Projeto gráfico e editoração eletrônica
Gleise Maria Teles de Oliveira

1ª edição
Publicação digital (2020)

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 (16 : 2019 : Manaus).
Anais da XVI Jornada Científica da Embrapa Amazônia Ocidental / Claudia Majolo ... [et al.], editores técnicos. – Brasília, DF : Embrapa, 2020.
PDF (130 p.) : il. color.

ISBN 978-65-86056-10-5

1. Iniciação científica. 2. Comunicação científica. 3. Pesquisa. I. Majolo, Cláudia. II. Título. III. Embrapa Amazônia Ocidental.

CDD 501

Tecnologia da Informação

Uso de *Ensemble Learning* para identificação de ataques de trips em guaranazeiro

Richardson Allan Ferreira de Souza¹

Marcos Filipe Alves Salame²

Resumo – O ataque do trips (*Pseudophilotrips adisi*) tem afetado o sistema de agricultura familiar, reduzindo a produção e restringindo a exploração comercial do guaraná no Amazonas. Tecnologias computacionais podem auxiliar na identificação dos danos causados por indivíduos *P. adisi*, contribuindo para o seu controle e monitoramento. Este trabalho objetivou o desenvolvimento de um classificador para identificar danos em folíolos de guaranazeiro atacados pelo trips com base em imagens obtidas de smartphones. Técnicas de *Transfer Learning* foram empregadas para criar extratores de características, métodos de *Over-Sampling* foram utilizados para lidar com as amostras desbalanceadas no conjunto de dados e os classificadores *Support Vector Machine*, *Multilayer Perceptron* e *Random Forest* foram usados para prever cada amostra de teste num método de votação, trazendo como resultado uma classificação dos folíolos de guaranazeiro atacados, com acurácia acima de 90%.

Termos de indexação: trips, guaraná, comitês de máquina.

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

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

Using Ensemble Learning to identify attacks by thrips in guarana

Abstract – The attack by thrips (*Pseudophilotrips adisi*) has affected the family farming system, reducing production and restricting commercial exploitation of guarana in the Amazon. Computer technologies can assist in identifying damage caused by *P. adisi* individuals, contributing to their control and monitoring. This paper aimed the development of mechanisms to identify damage caused in guarana leaves by thrips based on images obtained from smartphones. We used Transfer Learning techniques to create feature pullers, Over-Sampling methods to deal with unbalanced samples in dataset and Support Vector Machine, Multilayer Perceptron and Random Forest to predict each test sample on a voting method. As result, we obtained a rating of guarana leaves attacked with accuracy above 90%.

Index terms: thrips, guarana, ensemble.

Introdução

Evidenciada pelo crescimento da demanda de sementes pelas indústrias de bebidas, a cultura do guaraná vem assumindo importância cada vez maior no cenário econômico, sendo também fonte de renda para inúmeros produtores rurais da região amazônica (Senra et al., 2008).

O ataque do tripses (*P. adisi*) afeta de forma direta a produtividade, uma vez que ele se alimenta do conteúdo celular das estruturas jovens das plantas de guaranazeiro. Como consequência do ataque, pequenas pontuações são formadas, por onde sugam a seiva, causando alterações morfológicas e de desenvolvimento, as quais afetam a produção de grãos e limitam a exploração comercial do guaraná na região. Alternativas de controle e monitoramento têm

sido alvo de pesquisas desenvolvidas pela Empresa Brasileira de Pesquisa Agropecuária (Embrapa) (Fontes; Tavares, 2018).

A utilização de procedimentos computacionais para reconhecer automaticamente o ataque do tripses a folíolos do guaranzeiro pode contribuir para o seu controle e monitoramento. Assim, o objetivo deste trabalho foi a elaboração de um classificador capaz de identificar danos em folíolos jovens do guaranzeiro atacados por indivíduos *P. adisi* a partir do desenvolvimento de um preditor, formado por três técnicas de aprendizado de máquina, *Support Vector Machine* (SVM), *Multilayer Perceptron* (MLP) e *Random Forest* (RF), com a aplicação de *Ensemble Learning* (EL) para obter um conjunto de votos majoritários com a finalidade de alcançar maior acurácia e mais precisão na identificação dos danos.

Material e Métodos

Para a construção de uma base de imagens, folíolos de até 1,7 centímetros da cultivar BRS-Maués foram isolados em placas de Petri com até dez indivíduos *P. adisi* adultos e inspecionados durante 27 horas. Nesse intervalo de 27 horas, os exemplares de tripses alimentaram-se dos folíolos e foram sendo formadas injúrias na parte superior (adaxial) destes. Para a captura das imagens foram utilizadas as câmeras digitais dos smartphones Motorola Moto G5 XT1672 e Xiaomi Mi 9 com as funções de HDR e Flash desativadas, tendo apenas a luz natural.

Para elaborar o conjunto de dados (*dataset*), 26 imagens foram manualmente selecionadas, dimensionadas para 512 x 512 pixels e subdivididas em 6.240 regiões de 32 x 32 pixels. Cada região foi observada e classificada manualmente. As regiões que apresentavam pequenas pontuações, evidenciando o ataque, formaram a base para a classe positiva, e as que não apresentaram nenhum dado aparente foram usadas para compor a base para a classe negativa, chegando

ao número de 520 amostras da classe positiva e 3.123 amostras da classe negativa.

A representação da imagem foi feita empregando um método conhecido como *transfer learning*, em que foi utilizada a arquitetura canônica de *Convolutional Neural Networks* (CNN) chamada VGG16 como extrator de características. A camada de entrada e a camada totalmente conectada da arquitetura foram removidas, assim como os três últimos blocos das camadas de convolução. Cada amostra da classe positiva e da classe negativa foi aferida ao extrator, que tinha como saída um vetor de característica com 2.048 atributos para representar cada amostra.

Para balancear o número de exemplos da classe positiva foram utilizados algoritmos *Synthetic Minority Oversampling TEchnique* (SMOTE) propostos por Chawla et al. (2002), os quais são capazes de gerar amostras por interpolação das instâncias da classe minoritária. Ao final do processo foi possível obter um dataset com as classes balanceadas contendo 3.120 amostras para cada classe.

Após o balanceamento das classes, o *dataset* foi aleatoriamente separado uma vez em um conjunto de treino com 70% e um conjunto de teste com 30% do total.

Tabela 1. Separação do *dataset* para treino e teste.

Conjunto	Classe Positiva	Classe Negativa	Total
Treino	2.185	2.183	4.368
Teste	935	937	1.872

Os classificadores SVM, MLP e RF foram então treinados com as amostras do conjunto de treino e em seguida cada classificador foi usado para classificar amostras do conjunto de teste. A predição feita por classificador é agregada e uma predição final é definida com a maioria absoluta (Saha; Ekbal, 2013). Essa é uma técnica de EL

básica mas poderosa, conhecida como *Hard Voting Classifier* (HVC) e está representada na Figura 1.

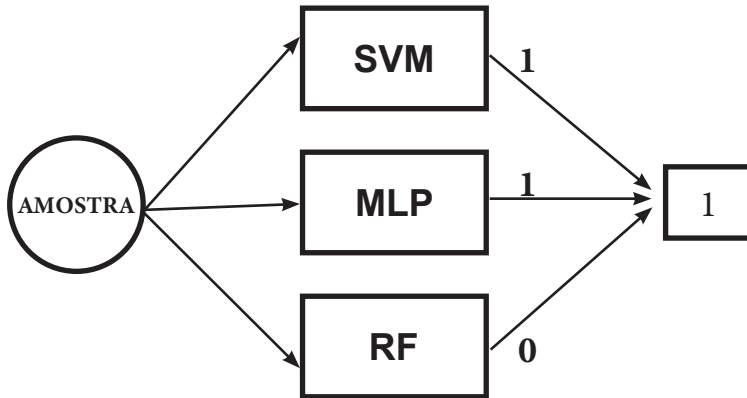


Figura 1. Fluxo de votação dos classificadores.

Resultados

Os resultados da acurácia e o índice Kappa alcançados durante a avaliação do conjunto de testes podem ser vistos na Figura 2, que apresenta o resultado individual e o resultado da combinação dos três modelos, empregando o método HVC. Na Figura 3 pode-se observar a matriz de confusão de cada modelo.

Discussão

O método SMOTE foi fundamental no balanceamento das amostras positivas, o que favoreceu não somente o alcance de uma boa acurácia, mas também um bom desempenho no índice Kappa, indicando que os classificadores conseguem identificar bem tanto as amostras da classe negativa quanto da classe positiva.

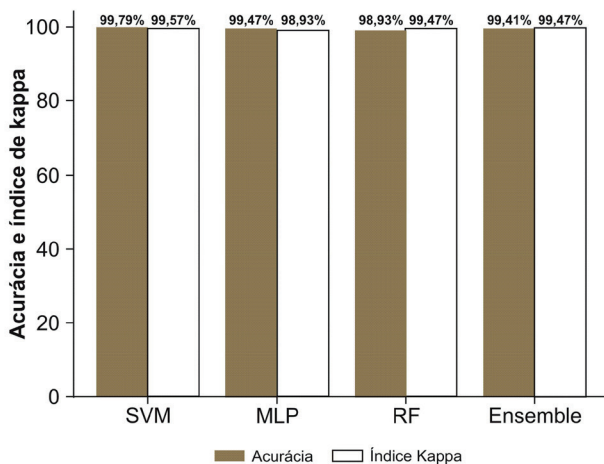


Figura 2. Porcentagem de acurácia e índice Kappa obtidos pelos classificadores SVM = *Support Vector Machine*; MLP = *Multilayer Perceptron*; RF = *Random Forest*; *Ensemble* = combinação dos três modelos.

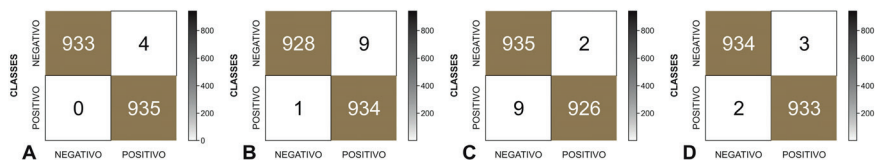


Figura 3. Matriz de confusão dos classificadores: SVM = *Support Vector Machine* (A); MLP = *Multilayer Perceptron* (B); RF = *Random Forest* (C); e *Ensemble* = combinação dos três modelos (D).

O uso de CNN como extrator de características também foi essencial para conseguir vetores de características representativos para o problema de forma automática, agilizando a tarefa de representar a imagem para que os classificadores pudessem encontrar padrões com mais facilidade.

Analisando a matriz de confusão apresentada na Figura 3, pode-se perceber que os classificadores SVM e MLP tendem a identificar mais amostras da classe positiva e o classificador RF a reconhecer mais amostras da classe negativa. Com isso, agregando os três classificadores para a tarefa de classificação, foi possível obter um novo classificador mais equilibrado na predição das amostras, conforme Figura 3D.

Conclusões

De modo geral, as técnicas utilizadas contribuíram para criação de modelos computacionais capazes de identificar automaticamente folíolos de guaranazeiro atacados por tripes. Todavia, apesar de ser possível obter amostras positivas com a técnica SMOTE, é imprescindível gerar variações de imagens com cenários reais do ataque para que se obtenha um ambiente mais próximo do real.

Agradecimentos

Os autores agradecem ao pesquisador Adauto Maurício Tavares pelas informações técnicas referentes ao tripes e guaraná; a Julliane da Silva Fontes, pelo auxílio nos experimentos; e à Fundação de Amparo à Pesquisa do Estado do Amazonas (Fapeam), pela disponibilização da bolsa de iniciação científica.

Referências

CHAWLA, N.; BOWYER, K. W.; HALL, L. O.; KEGELMEYER, W. P. SMOTE: synthetic minority over-sampling technique. **Journal of Artificial Intelligence Research**, v. 16, p. 321-357, 2002.

FONTES, J. da S.; TAVARES, A. M. Aspectos da biologia da linhagem sexuada do tripes-do-guaranazeiro (*Pseudophilothrips adisi*, zur Strassen) (Thysanoptera: Phlaeothripidae). In: JORNADA DE INICIAÇÃO CIENTÍFICA DA EMBRAPA AMAZÔNIA OCIDENTAL, 14., 2017, Manaus. **Anais...** Brasília, DF: Embrapa, 2018. p. 67-75. Disponível em: <http://ainfo.cnptia.embrapa.br/digital/bitstream/item/185856/1/XIV-Jornada-IC-67.pdf>. Acesso em: 23 jul. 2019.

SAHA, S.; EKBAL, A. Combining multiple classifiers using vote based classifier ensemble technique for named entity recognition. **Data & Knowledge Engineering**, v. 85, p. 15-39, 2013.

SENRA, T. V.; GARCIA, M. V. B.; TAVARES, A. M. Comportamento e controle de tripes do guaranazeiro, *Liothrips adisi* (Thysanoptera: Phlaeothripidae). In: SEMINÁRIO DE PÓS-GRADUAÇÃO NA EMBRAPA AMAZÔNIA OCIDENTAL, 2008, Manaus. **Integrando esforços para o desenvolvimento da Amazônia**. Manaus: Embrapa Amazônia Ocidental, 2008. p. 19-21. Disponível em: <http://ainfo.cnptia.embrapa.br/digital/bitstream/item/109053/1/Comportamento-Livro-Seminario-Pos-Grad.pdf>. Acesso em: 23 jul. 2019.