

MÉTODO COMPUTACIONAL UTILIZANDO SUPPORT VECTOR MACHINE PARA CLASSIFICAÇÃO DE FOLHAS DOENTES BASEADO EM TEXTURA

LUCAS R. TREVISAN¹, LEONARDO P. MAGALHÃES², ANA C. M. LIMA³,
TAMARA M. GOMES⁴, FABRICIO ROSSI⁵

¹ Doutorando - Dpt. Engenharia de Biosistemas, ESALQ/USP - Piracicaba-SP, lucas.renato.trevisan@usp.br/(14)98192-2925

² Doutorando - Dpt. Engenharia de Biosistemas, ESALQ/USP - Piracicaba-SP, leonardo.magalhaes@usp.br

³ Doutoranda - Dpt. Engenharia de Biosistemas, ESALQ/USP - Piracicaba-SP, ana.murad.lima@usp.br

⁴ Professor Doutor do Dpt. de Engenharia de Biosistemas, FZEA/USP, Pirassununga-SP, fabricio.rossi@usp.br

⁵ Professora Doutora do Dpt. de Engenharia de Biosistemas, FZEA/USP, Pirassununga-SP, tamara.gomes@usp.br

Apresentado no
XLIX Congresso Brasileiro de Engenharia Agrícola - CONBEA 2020
23 a 25 de novembro de 2020 - Congresso On-line

RESUMO: Folhas de plantas com patologias ou deficiências apresentam fatores como cor e textura alterados. Cor e textura extraídas das folhas das plantas podem ser utilizadas para identificação da presença de patologias. Métodos computacionais surgem para avaliações mais rápidas, não invasivas e mais precisas, possibilitando uma tomada de decisão precoce, sendo a segmentação e a classificação problemas de interesse. No problema de classificação, Support vector Machine (SVM) tem demonstrado grande acurácia. SVM é um classificador binário que separa um conjunto de dados em dois grupos de atributos. Este trabalho propõe extrair informações de cor (RGB) e textura (energia, contrastes, dissimilaridade, homogeneidade e correlação) e apresentar uma metodologia para segmentação e classificação de plantas doentes. O modelo SVM foi satisfatório realizando uma classificação com uma taxa de acerto de 86,25% (13,25% taxa de erro), sendo eficaz em indicar a existência de doenças nas folhas. Este trabalho demonstrou que cor e textura extraídas de imagens digitais obtidas por smartphone em RGB podem ser utilizados de uma forma eficaz na segmentação, bem com os atributos de textura energia e contraste na classificação de folhas de plantas para distinção de folhas saudáveis daquelas que apresentam alguma forma de patologia.

PALAVRAS-CHAVE: Visão Computacional; Classificação; Aprendizagem de Máquinas.

COMPUTATIONAL METHODOLOGY USING SUPPORT VECTOR MACHINE FOR CLASSIFYING LEAVES WITH DISEASE BASED ON TEXTURE

ABSTRACT: Plant leaves with pathologies or deficiencies present factors such as color and texture altered. Color and texture extracted from the leaves of plants can be used to identifying presence of pathologies. Computational methods emerge for faster, non-invasive and more accurate assessments, enabling early decision making, with segmentation and classification being problems of interest. In the classification problem, support vector machine (SVM), has shown great accuracy. SVM is a binary classifier that separates a data set into two groups of attributes. This work proposes to extract color (RGB) and texture (energy, contrasts, dissimilarity, homogeneity and correlation) information from plant leaves and present a methodology for segmentation and classification of plants that present signals of disease. The SVM model was satisfactory, performing a classification with an accuracy rate of 86.25% (13.78% error rate), proving to be effective in indicating the existence of leaf diseases. This work demonstrated that color and texture extracted from digital images obtained by smartphone in RGB can be used in an effective way in the segmentation, as well as energy and contrast textured features in the classification of plant leaves to distinct healthy leaves from those that present some sort of pathology.

KEYWORDS: Computer Vision; Classification; Machine Learning

INTRODUÇÃO: Folhas de plantas com patologias apresentam atributos de cor e textura alterados devido a modificações em suas estruturas ou atividades biológicas. Cor e textura podem ser utilizadas para identificar a presença de patologias, infestações e deficiência de nutrientes (ARIVAZHAGAN; SHEBIAH; ANANTHI, 2013). Métodos computacionais baseados nesses atributos surgem para avaliações mais rápidas e precisas (JENIFA et al., 2019). Em relação à cor, muitos trabalhos adotam imagem em RGB (red, blue, green) (ALI et al., 2017). Zheng et al., (2006) destacaram 21 parâmetros de textura sendo que quatro desses (energia, entropia, homogeneidade e contraste) apresentam bons resultados na detecção de plantas saudáveis. Grandes desafios dentro da visão computacional estão na segmentação e criação de modelos de classificação. A segmentação se refere à separação das informações de interesse na imagem (SINGH; MISRA, 2017). No problema de classificação, Support Vector Machine (SVM) tem demonstrado grande acurácia. SVM é um classificador binário que separa um conjunto de dados por um hiperplano (DUBEY et al., 2018), sendo amplamente utilizado, como na classificação de doenças em folhas de algodão (JENIFA et al., 2019); soja (ARAUJO; PEIXOTO, 2019); entre outros. Dessa forma, este trabalho busca apresentar uma metodologia para segmentação e classificação de folhas doentes baseados em cor e textura, contribuindo para um diagnóstico mais rápido e melhorando a eficácia nas tomadas de decisões.

MATERIAL E MÉTODOS: Neste estudo foi utilizado o pepino japonês (*Cucumis sativus* L.), cultivar Tikara (Agristar), sendo as folhas colhidas após o início da frutificação. As folhas foram individualmente dispostas em um fundo rosa para obtenção de imagens de sua face adaxial (Figuras 1a e 1b), com Smartphone (MOTOROLA G5S PLUS, câmera de 13 megapixels). Através de uma análise visual, as imagens foram manualmente classificadas como saudáveis (FIGURA 1a) ou doentes (FIGURA 1b), sendo as folhas doentes aquelas apresentando características como manchas pretas, maior senescência ou necrose. 70% das imagens foram separadas para a etapa de treinamento e 30% para a etapa de teste (classificação).

Um algoritmo em linguagem python foi desenvolvido para segmentação das folhas e cálculo dos valores de textura. O algoritmo foi estruturado para ler a imagem em RGB, separar os canais de cores, e obter o quociente entre os canais G e B (G/B). Como as intensidades de verde são muito maiores do que as intensidades de azul nos pixels pertencentes à folha, G/B deve ser maior nestas regiões e menor em regiões do *background*. Estudos iniciais mostraram então que o limiar $G/B = 1.2$ foi adequado para separar os pixels pertencentes ao *background*.

Também em linguagem python, um algoritmo foi desenvolvido para varrer a imagem segmentada, após convertê-la para escala de cinza, e determinar a Matriz de Co-ocorrência de níveis de cinza (GLCM, em inglês) utilizando uma vizinhança $d = 2$ pixels e ângulo de deslocamento $\theta = 0^\circ$. Baseados na GLCM foram então determinados os valores de textura para os índices: contraste, dissimilaridade, homogeneidade, energia de acordo com as equações apresentadas por Esefpour e Massah (2012). Estudos preliminares demonstraram que a avaliação do Contraste vs. Energia foram os melhores índices para classificação das folhas.

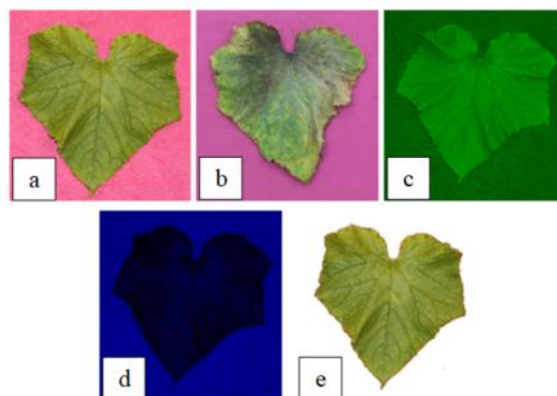


FIGURA 1. (a) Folha saudável; (b) folha doente; (c) canal verde da imagem original; (d) canal azul da imagem original; (e) imagem com *background* removido.

O modelo SVM foi obtido através da biblioteca *scikit-learn* em linguagem python, sendo inicialmente o modelo ajustado com os dados de treinamento. Após o treinamento, foi possível realizar a predição do restante dos dados, onde o algoritmo deveria obter as informações de textura da imagem e identificar se aquela folha era uma folha doente ou uma folha saudável. A acurácia do modelo foi avaliada pelo método da razão de erro (ER), ou seja, a porcentagem de vezes que o algoritmo classificou uma folha doente como sendo uma folha saudável. Todos os algoritmos utilizados neste trabalho podem ser encontrados no diretório do github (<https://github.com/lucastrevisan90/Python-GEBIO>). O fluxograma apresentado na FIGURA 2 ilustra todo o processamento realizado neste trabalho.

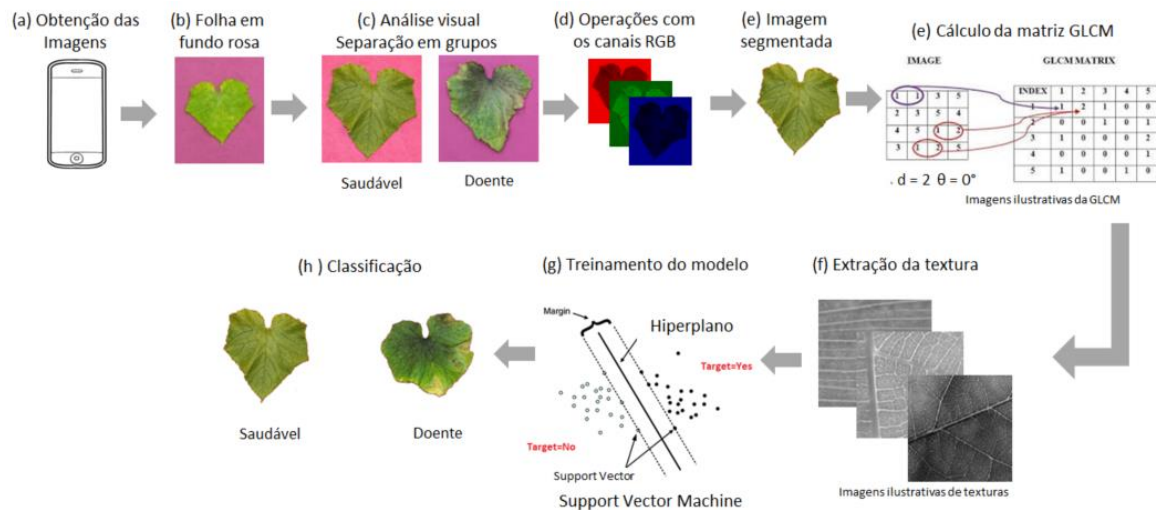


FIGURA 2. Fluxograma do projeto. (a) – (c) coleta de imagens e análise visual. (d) – (h) Metodologias computacionais desenvolvidas em linguagem python.

RESULTADOS E DISCUSSÃO: O uso dos canais azul e verde foi satisfatório para determinação de uma matriz de intensidades que possibilitasse a utilização de um *threshold* para segmentação das imagens (FIGURA 1e). Para cada imagem segmentada, a GLCM foi determinada e os valores de textura obtidos (TABELA 1).

A homogeneidade apresenta relação com a distribuição dos níveis de cinza na imagem e é inversamente proporcional ao contraste e à energia (TABELA 1) (ITO et al., 2009). O atributo de dissimilaridade é a medida da distância espacial entre dois objetos dentro de uma região de interesse (ABORISADE et al., 2014), os valores encontrados nesse trabalho foram maiores para plantas saudáveis. Por fim, correlação, está relacionada com a forma como um pixel se correlaciona com seus vizinhos, podendo apresentar correlação direta ou inversa (ABORISADE et al., 2014). Folhas saudáveis apresentaram correlações menores, porém, muito próximos, daqueles para folhas doentes.

O atributo energia está diretamente relacionado com o nível de brilho na escala de cinza (STORY et al., 2010), fator que se demonstrou menor para as folhas doentes. Já o contraste é um parâmetro de textura que mede as variações locais dentro de uma imagem, ou seja, quanto maior a complexidade de uma superfície, mais alto será o valor do contraste (STORY et al., 2010). Observou-se que as folhas saudáveis apresentaram valores maiores de contraste.

TABELA 1. Valores médios dos índices de textura calculados para as folhas saudáveis e aquelas apresentando algum tipo de doença.

	Contraste	Homogeneidade	Dissimilaridade	Energia	Correlação
Saudável	103,55 ± 34,05	2,17 ± 0,81	0,86 ± 0,045	0,84 ± 0,049	0,97 ± 0,006
Doente	91,23 ± 46,78	2,36 ± 1,22	0,78 ± 0,090	0,74 ± 0,100	0,98 ± 0,003

Valores de média ± desvio padrão.

Os resultados para energia e contraste foram plotados onde se observa uma divisão entre os grupos de valores, o que possibilitou a separação em dois grupos pelo hiperplano plotado na FIGURA

3. Estes valores foram então utilizados para o treinamento do modelo SVM. Em seguida, o modelo foi submetido a teste de acurácia utilizando os dados de classificação e apresentou ER de 0.1375.

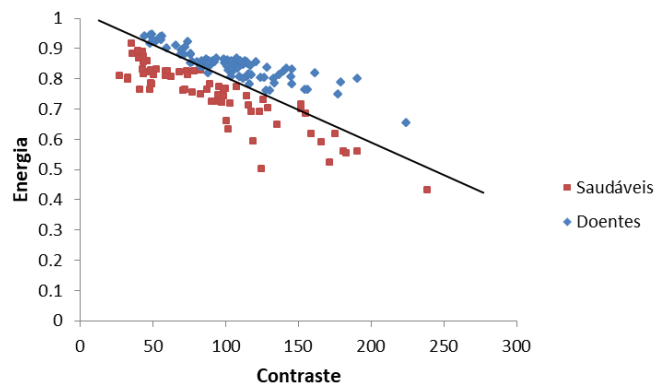


FIGURA 3. Valores dos índices de textura de Energia e Contraste.

CONCLUSÃO: Neste trabalho foram desenvolvidos algoritmos para pré-processamento, processamento, extração de informações de interesse e classificação de imagens. Foi apresentada uma metodologia capaz de segmentar o *background* das imagens e manter as informações na região de interesse utilizando apenas os canais R, G e B. Foi também desenvolvido um algoritmo para extração de atributos de textura das imagens, dentre os quais, Contraste e Energia foram utilizados para elaboração de um modelo SVM para classificação de folhas de plantas entre saudáveis e doentes. O modelo SVM foi satisfatório realizando uma classificação com 86,25% de acerto. Este trabalho demonstrou que as informações de cor e textura extraídas de imagens digitais obtidas por smartphone em RGB podem ser utilizados de uma forma eficaz na segmentação e classificação de folhas de plantas para identificação de folhas saudáveis ou folhas que apresentam alguma forma de patologia. Como trabalhos futuros, pretende-se avançar para modelos de classificação não binários.

AGRADECIMENTOS: À CAPES, à ESALQ/USP, e à sociedade em geral que mantém o financiamento da pesquisa científica.

REFERÊNCIAS

- ABORISADE, D. O. et al. Comparative analysis of textural features derived from GLCM for ultrasound liver image classification. *Energy*, v. 2, p. 10, 2014.
- ALI, Hussam et al. Symptom based automated detection of citrus diseases using color histogram and textural descriptors. *Computers and Electronics in Agriculture*, v. 138, p. 92-104, 2017.
- ARAUJO, Juliana Mariana Macedo; PEIXOTO, Zelia Myriam Assis. A new proposal for automatic identification of multiple soybean diseases. *Computers and Electronics in Agriculture*, v. 167, p. 105060, 2019.
- ARIVAZHAGAN, Sai et al. Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features. *Agricultural Engineering International: CIGR Journal*, v. 15, n. 1, p. 211-217, 2013.
- DUBEY, Yogita K.; MUSHRIF, Milind M.; TIPLE, Sonam. Superpixel based roughness measure for cotton leaf diseases detection and classification. In: **2018 4th International Conference on Recent Advances in Information Technology (RAIT)**. IEEE, 2018. p. 1-5.
- ITO, Rodrigo Hajime; KIM, Hae Yong; SALCEDO, Walter Jaimes. Classificação de texturas invariante a rotação usando matriz de co-ocorrência. In: **8th International Information and Telecommunication Technologies Symposium**. 2009. p. 1-6.
- JENIFA, A.; RAMALAKSHMI, R.; RAMACHANDRAN, V. Classification of Cotton Leaf Disease Using Multi-Support Vector Machine. In: **2019 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS)**. IEEE, 2019. p. 1-4.
- SINGH, Vijai; MISRA, Ak K. Detection of plant leaf diseases using image segmentation and soft computing techniques. *Information processing in Agriculture*, v. 4, n. 1, p. 41-49, 2017.
- STORY, David et al. Lettuce calcium deficiency detection with machine vision computed plant features in controlled environments. *Computers and electronics in agriculture*, v. 74, n. 2, p. 238-243, 2010.
- ZHENG, Chaoxin; SUN, Da-Wen; ZHENG, Liyun. Recent applications of image texture for evaluation of food qualities—a review. *Trends in Food Science & Technology*, v. 17, n. 3, p. 113-128, 2006.