

DETERMINAÇÃO DO ESTÁDIO DE MATURAÇÃO EM UVAS VINÍFERAS COM APRENDIZADO DE MÁQUINA – *RANDOM FOREST*

THAÍS P. DE AZEVEDO¹, DANIEL DOS S. COSTA², VICTOR GUSTAVO DA S. OLIVEIRA³, SABRINA R. NAZÁRIO⁴, GLEYCE KELLY D. A. FIGUEIREDO⁵, WOTSON LUCAS A. DA SILVA⁶

¹ Enga Agrícola e Ambiental, Prof. Assistente, Colegiado de Eng. Agrícola e Ambiental, Universidade Federal do Vale do São Francisco, UNIVASF, Juazeiro – BA, Fone: (0XX74) 2102.7621, thais.azevedo@univasf.edu.br.

² Engo Agrícola e Ambiental, Prof. Adjunto, Colegiado de Eng. Agrícola e Ambiental, UNIVASF, Juazeiro-BA.

³ Engo Computação, Graduando em Eng da Computação, Colegiado de Eng. da Computação, UNIVASF, Juazeiro-BA..

⁴ Enga Agrícola e Ambiental, Graduanda em Eng Agrícola e Ambiental, Colegiado de Eng. Agrícola e Ambiental, UNIVASF, Juazeiro-BA.

⁵ Tecnóloga da Construção Civil, Prof. Doutora, Faculdade de Engenharia Agrícola, UNICAMP, Campinas-SP.

⁶ Engo Agrícola e Ambiental, Graduando em Eng Agrícola e Ambiental, Colegiado de Eng. Agrícola e Ambiental, UNIVASF, Juazeiro-BA.

Apresentado no
XLVIII Congresso Brasileiro de Engenharia Agrícola - CONBEA 2019
17 a 19 de setembro de 2019 - Campinas - SP, Brasil

RESUMO: A determinação do ponto de colheita de uvas de mesa e vinho é realizada por meio de técnicas destrutivas e com uma amostragem que não é representativa de toda a área a ser colhida. Neste sentido, este trabalho teve como objetivo avaliar o desempenho do método de aprendizado de máquinas *Random Forest* a partir de dados de imagem de refletância e fluorescência, na classificação do estágio de maturação de uvas da variedade Cabernet Sauvignon. Para isso, amostras em diferentes estádios de maturação foram colhidas durante oito semanas e suas imagens obtidas com o uso de câmera fotográfica, após, as imagens foram processadas para se obter informações dos espaços de cor RGB, rgb, Lab e HSV. Os dados de imagem normalizados por vetor unitário e com normalização por área foram utilizados para implementar classificadores usando *Random Forest*, onde apresentaram uma acurácia de 98,44% e índice Kappa de 0,96. Assim, *Random Forest* pode permitir o desenvolvimento de um sistema relativamente simples e automatizado para classificar a uva quanto ao estágio de maturação.

PALAVRAS-CHAVE: aprendizado de máquinas, técnicas não-destrutivas, agricultura digital.

DETERMINATION OF THE RIPENING STAGE IN VINIFEROUS GRAPES WITH MACHINE LEARNING – *RANDOM FOREST*

ABSTRACT: The determination of the point of collection of table grapes and wine shall be carried out by means of destructive techniques and with sampling that is not representative of the entire area to be collected. In this sense, this study aimed to evaluate the performance of the Random Forest machine learning method based on image data of reflectance and fluorescence, in the classification of the ripening stage of Cabernet Sauvignon grapes. For this purpose, samples at different stages of ripeness were collected for eight weeks and their images obtained with the use of a photographic camera, after which the images were processed to obtain information from the RGB, rgb, Lab and HSV color spaces. Image data normalized by unit vector and normalized by area were used to implement classifiers using Random Forest, where they showed an accuracy of 98.44% and a Kappa index of 0.96. Thus,

Random Forest can allow the development of a relatively simple and automated system to classify the grapes according to their stage of ripeness.

KEYWORDS: machine learning, non-destructive techniques ,digital agriculture.

INTRODUÇÃO: A colheita da uva no estágio de maturação adequado é de fundamental importância para a qualidade dos vinhos. É neste período que a qualidade enológica da uva é determinada, pois é quando ocorre a síntese, degradação e/ou a translocação de compostos de interesse para as bagas, como: açúcares, terpenos, antocianinas, taninos, ácidos orgânicos, entre outros (CONDE et al., 2007). No período de maturação também podem ocorrer injúrias ocasionadas por pragas ou doenças que poderão interferir na composição físico-química das bagas (CHAVARRIA et al., 2007). A determinação do estágio de maturação também é importante por este estar relacionado à propriedades internas da fruta, como doçura e firmeza (MAGWZA & OPARA, 2015). Além disso, ajuda a determinar o tempo de armazenamento antes do consumo. A determinação da maturação é feita em um número limitado de amostras que não são representativos das condições físico-químicas típicas e variabilidade encontrada nos lotes. Assim, técnicas rápidas, inteligentes e não destrutivas são cada vez mais necessárias (WU & SUN, 2013). Sistemas de visão computacional têm sido utilizados para determinação de atributos de qualidade de alguns alimentos (JACKMAN et al., 2012). Diante do exposto, este trabalho teve como objetivo avaliar o desempenho do método de aprendizado de máquinas *Random Forest*, a partir de dados de imagem de refletância e fluorescência, na classificação do estágio de maturação de uvas da variedade Cabernet Sauvignon.

MATERIAL E MÉTODOS: As amostras de uva (*Vitis vinifera* L.), da variedade Cabernet Sauvignon, foram coletadas no município de Lagoa Grande – PE, localizado na região do Vale do Submédio São Francisco. A região é caracterizada por possuir clima do tipo BSh, segundo a classificação de Köppen (ALVARES et al., 2013), correspondendo a uma região de clima árido. As videiras onde foram realizadas as coletas estavam em uma área de 2,2 ha com irrigação localizada, com espaçamento entre linhas de 3 m e entre plantas de 1m.

Foram realizadas entre os meses de maio e agosto de 2017 oito coletas de 72 bagas, em diferentes datas, com a primeira ocorrendo 73 dias após a poda, abrangendo vários estágios de maturação, desde antes do véraison até o estágio de colheita, totalizando ao final do experimento 576 bagas. Para obter amostras representativas, foi adotado o método de amostragem estratificado. Uma fileira do meio do parreiral foi escolhida e, dela, selecionou-se seis plantas do centro.

O sistema de aquisição de imagens de refletância produzido no Laboratório de Energia na Agricultura (LENA) da Universidade Federal do Vale do São Francisco (UNIVASF) é constituído por uma câmera fotográfica Canon T5i, caixa de interior preto fosco, fonte de alimentação ajustável e uma caixa de controle de acendimento dos LEDs.

Foram obtidas cinco imagens de cada baga, sendo cada foto adquirida com iluminação de LEDs diferentes (vermelho, verde, azul, branco frio e branco quente). A câmera foi configurada para uma obtenção manual das imagens com ISO-100, tempo de exposição 1/2s, F/5.6, distância focal de 48mm. Ao final, foram adquiridas um total de 1200 imagens com resolução de 3456 x 5184 pixels para serem analisadas. Foi realizado o realce das imagens e após, os pixels foram convertidos nos espaços de cor RGB, rgb, Lab, e HSV.

Os valores médios dos pixels das imagens obtidos dos espaços de cor após o processamento das imagens das bagas foram organizados em uma tabela que possuía todos os espaços de cor obtidos para cada iluminante utilizado, o número da amostra, e o estágio de maturação em que a amostra se encontrava segundo a classificação de Coombe (1995) adaptada. Nesta etapa

foram realizados dois tipos de normalização com o intuito de diminuir a influência dos atributos que possuíam maior faixa de medição, e que com isso poderiam influenciar no classificador. Desta forma, foram aplicadas a normalização por área e a normalização por vetor unitário visto que os valores do espaço de cor Lab são negativos e estes dois métodos de normalização se aplicam a dados com esta característica.

Após a normalização dos dados estes foram analisados a partir da técnica de aprendizado de máquina supervisionado, onde o conjunto de dados foi dividido em treinamento e teste com 2/3 e 1/3 dos dados respectivamente, além disso, as saídas desejadas foram três estádios de maturação, sendo eles: green para as bagas totalmente verdes, véraison para as bagas com início de pigmentação e ripe para as bagas totalmente pigmentadas.

O algoritmo utilizado para a obtenção dos classificadores do estágio de maturação em uvas da variedade Cabernet Sauvignon foi *Random Forest*. O pacote instalado para o algoritmo foi o randomForest (LIAW & WIENER, 2002), onde foram utilizadas 600 árvores visto que com este número houve estabilização do erro de classificação. Além deste pacote, utilizou-se o pacote caret (KUHN, 2013), para a escolha dos melhores valores para alguns parâmetros. O modelo gerado teve sua eficiência avaliada por meio da matriz de confusão, do índice de Kappa e da acurácia.

RESULTADOS E DISCUSSÃO: Os dados com normalização por vetor unitário e com normalização por área foram classificados por meio do algoritmo *Random Forest*, não havendo diferença entre os índices Kappa, acurácia, sensibilidade e especificidade diante dos dois tipos de normalização. A Tabela 1 expõe os resultados obtidos com *Random Forest* e normalização dos dados por vetor unitário.

Os dados com normalização por vetor unitário apresentaram uma acurácia de 98,44% e índice Kappa de 0,96. Esses parâmetros mostram que o classificador foi eficiente em realizar a classificação das amostras quanto ao estágio de maturação. Pela matriz de confusão podemos observar que o maior número de amostras classificadas de forma incorreta ocorreu para a classe véraison, que é a classe onde as uvas começam a pigmentar, ou seja, estágio entre maduras e verdes.

TABELA 1. Matriz de confusão, Sensibilidade e Especificidade para cada classe, Acurácia e índice Kappa dos dados normalizados por vetor unitário classificados usando *Random Forest* para treinamento e teste.

	Green	Ripe	Veraison	Sensibilidade	Especificidade	Acurácia	Kappa
Treinamento							
Green	110	0	1	0,9821429	0,9963235	0,973958	0,944331
Ripe	0	253	5	0,9921569	0,9612403		
Veraison	2	2	11	0,64705882	0,98910082		
Teste							
Green	51	0	0	0,9807692	1	0,984375	0,963303
Ripe	0	135	1	0,9926471	0,9821429		
Veraison	1	1	3	0,75	0,9893617		

Esses resultados são corroborados por Pereira et al. (2017) que testaram modelos de *Random Forest* para previsão de estágio de maturação em mamão, obtendo desempenho médio de 94,3%, o que demonstra que dados de imagem associados a aprendizado de máquinas é eficiente na classificação de produtos agrícolas.

CONCLUSÕES: Através do processamento de imagens de refletância e fluorescência associado a aprendizado de máquina, resultados satisfatórios são relatados para classificação de uva da variedade Cabernet Sauvignon com base no seu estágio de maturação. *Random Forest* é uma técnica que pode permitir o desenvolvimento de um sistema relativamente simples e automatizado para classificação de uvas de acordo com diferentes níveis de maturidade. Mais investigações são necessárias para alcançar níveis robustos de previsão de maturação, principalmente no que se refere a diferenciar os vários níveis do estágio *véraison*.

REFERÊNCIAS

- ALVARES, C.A., J.L. STAPE, P.C. SENTELHAS, J.L.M. GONC, ALVES, 2013: Modeling monthly mean air temperature for Brazil. – **Theor. Appl. Climatol.** 113, 407–427.
- CHAVARRIA, G. et al. Incidência de doenças e necessidade de controle em cultivo protegido de videira. **Revista Brasileira de Fruticultura**, v.29, n.3, p.477- 482, 2007.
- CONDE, C.; SILVA, P.; FONTES, N.; DIAS, A.C.P.; TAVARES, R.M.; SOUSA, M.J.; AGASSE, A.; DELROT, S.; GERÓS, H. Biochemical changes throughout grape berry development and fruit and wine quality. Ed. **Global Science Book**, , v.1., n.1, p.1-22, 2007.
- COOMBE, B.G. Growth stages of the grapevine: adoption of a system for identifying grapevine growth stages. *Australian Journal Grape and Wine Research*, v.1, p.104-110, 1995.
- JACKMAN, P.; SUN, D.-W.; ELMASRY, G. Robust colour calibration of an imaging system using a colour space transform and advanced regression modelling. **Meat Sci.** 91, p. 402–407, 2012.
- KUHN, M. Caret: classification and regression training. R package version 5.1624. 2013.
- LIAW, A.; WIENER, M. Classification and regression by Random Forest. *R News*, v.2, n.3, p.18-22, 2002
- MAGWAZA, L.S.; OPARA, U.L. Analytical methods for determination of sugars and sweetness of horticultural products a review. **Scientia Horticulturae** 184, p.179–192, 2015.
- PEREIRA, L. F. S.; BARBON JR., S.; VALOUSB, N. A.; BARBIN, D. F. Predicting the ripening of papaya fruit with digital imaging and random forests. **Computers and Electronics in Agriculture**. p. 76-82, 2017.
- WU, D.; SUN, D.-W. Colour measurements by computer vision for food quality control – a review. **Trends Food Sci. Technol.** 29, p. 5–20, 2013.