

APLICAÇÃO DE AMOSTRAS SINTÉTICAS EM ESPECTROSCOPIA ATR-FTIR PARA A DETERMINAÇÃO DE PROPRIEDADES DE MEL

Ellisson Henrique de Paulo^{1,2}(PG), Evandro Messias²(PG), Emanuele Catarina da Silva Oliveira³(PQ), Márcia Helena Cassago Nascimento²(PQ), Pedro Henrique Pereira da Cunha²(PQ), Paulo Roberto Filgueiras²(PQ), Marco Flôres Ferrão¹(PQ).

¹Universidade Federal do Rio Grande do Sul, Avenida Bento Gonçalves, 9500, Agronomia – 90650-001 – Porto Alegre-RS, Brasil, ellisson.hp@gmail.com, mfferrao@gmail.com.

²Universidade Federal do Espírito Santo, Avenida Fernando Ferrari, 514, Goiabeiras - 29075-910 - Vitória-ES, Brasil, evandroqt@gmail.com, marcia.cassago@gmail.com, pedrohenrique@hotmail.com, paulo.filgueiras@ufes.br.

³Instituto Federal do Espírito Santo, Avenida Ministro Salgado Filho, 1000, Soteco - 29106-010 – Vila Velha-ES, Brasil, emanuele.oliveira@ifes.edu.br.

Resumo

O mel é valorizado por seu alto valor nutricional e benefícios à saúde, tornando o controle de sua qualidade crucial. A análise físico-química, embora essencial para garantir a pureza e qualidade do mel, pode ser complexa e demorada. Em contraste, a espectroscopia na região do infravermelho médio com reflectância total atenuada (ATR-FTIR) oferece uma abordagem rápida e não destrutiva. A combinação dessa técnica com as regressões por mínimos quadrados parciais (PLS) e por vetores de suporte (SVR) pode produzir modelos para estimar as propriedades do mel. Todavia, em situações em que o número de amostras é pouco, os modelos de regressão podem sofrer com a falta de representatividade. Frente a isso, o método de aumento de amostra com ruído gaussiano (SARGAN) foi usado para gerar amostras sintéticas. Os modelos SARGAN-PLS são a melhor estratégia para atividade diastásica, acidez livre, açúcar redutor, cinzas, hidroximetilfurfural, sacarose aparente e umidade, com R^2_p iguais a 0,870, 0,878, 0,852, 0,803, 0,736, 0,679 e 0,950, respectivamente. Os resultados destacam a importância das amostras sintéticas na estimativa das propriedades do mel.

Palavras-chave: Mel. ATR-FTIR. SARGAN. PLS. SVR.

Área do Conhecimento: Ciências Exatas e da Terra.

Introdução

Nos últimos anos, a produção de mel no Brasil aumentou significativamente devido à modernização do manejo, novas tecnologias e maior demanda (ABEMEL, 2024). O mel, sendo um alimento natural e de fácil digestão, é valorizado por seu alto valor nutricional, rápida absorção de carboidratos e benefícios no tratamento de feridas e saúde bucal (ZAMMIT YOUNG *et al.*, 2023). O crescimento do consumo reflete o interesse por produtos naturais. Assim, os órgãos de controle devem garantir a autenticidade e os padrões nutricionais do mel (BRASIL, 2000).

A análise físico-química do mel é crucial para garantir sua qualidade e autenticidade, fornecendo informações sobre sua composição, pureza e características como umidade, pH e açúcar (BRASIL, 2000). No entanto, essas análises podem ser limitadas por sua complexidade, necessidade de técnicas laboratoriais avançadas, o uso de reagentes e o tempo necessário para obter resultados (EL SOHAIMY, 2015). Além disso, algumas análises podem ser destrutivas, o que pode ser experimentalmente desvantajoso. Como alternativa, a espectroscopia na região do infravermelho médio com reflectância total atenuada (ATR-FTIR, do inglês *attenuated total reflectance Fourier transform infrared*) oferece uma abordagem rápida, não destrutiva e sem reagentes para obter informações espectrais sobre as bandas de ligações ou grupos funcionais na composição química do mel (COZZOLINO, 2011). Sendo uma técnica de detecção eficiente e não invasiva, a ATR-FTIR é amplamente utilizada para controle de qualidade, identificação e detecção de adulterações em alimentos, além de estudos sobre estruturas e mudanças conformacionais durante o processamento de alimentos (SKAFF, 2022). A combinação de métodos quimiométricos com espectroscopia aprimora ainda mais a análise. Os métodos

quimiométricos, como a regressão por mínimos quadrados parciais (PLS, do inglês *partial least squares*) e a regressão por vetores de suporte (SVR, do inglês *support vector regression*), podem usadas para quantificar diversas propriedades físico-químicas de mel (SE *et al.*, 2019; PAI *et al.*, 2009).

Todavia, em conjuntos de dados com poucas amostras, os modelos quimiométricos podem estar suscetíveis a sobreajuste, ter baixa representatividade, desempenho limitado e dificuldade na validação (HE *et al.*, 2008). Frente a isso, métodos de sobreamostragem têm sido propostos na literatura com o objetivo de mitigar esses problemas (YANG *et al.*, 2018). Porém, essas propostas se concentram principalmente em problemas qualitativos, nos quais, são gerados dados sintéticos somente para as variáveis independentes (HE *et al.*, 2008). Neste trabalho, a metodologia de aumento de amostra em regressão com ruído gaussiano (SARGAN, do inglês *sample augmentation in regression with Gaussian noise*) foi utilizada para gerar amostras sintéticas (SILVA *et al.*, 2011). A técnica SARGAN foi acoplada às regressões PLS e SVR para quantificar atividade de água (AA), atividade diastásica (AD), acidez livre (AL), açúcar redutor (AR), cinzas (C), hidroximetilfurfural (HMF), pH, sacarose aparente (SA) e umidade (U) do mel a partir de espectros ATR-FTIR.

Metodologia

Neste estudo, 36 amostras de mel produzidas no estado do Espírito Santo foram utilizadas. A análise físico-química foi realizada seguindo procedimentos padrão. O teor de umidade (%m/m), hidroximetilfurfural (HMF) (%mg/kg) e acidez livre (%meq/kg) foram determinados experimentalmente seguindo os métodos 962.19, 969.38 B e 980.23 da Associação Oficial de Químicos Analíticos (AOAC, 1997). A atividade diastásica (DN, *diastase number*), os açúcares redutores (%m/m), o teor de cinzas (%m/m) e a sacarose aparente (%m/m) foram quantificados pelos métodos disponibilizados pela (CAC, 1990). O pH das amostras foi determinado através de um pHmetro digital. A atividade de água (aw) foi determinada pelo analisador DECAGON (Aqualab, CX-2) (KUROISHI, 2012).

Os espectros ATR-FTIR foram adquiridos em espectrômetro da Agilent Technologies modelo Cary 630 FTIR. As análises foram realizadas em temperatura ambiente (25 °C) em triplicata. Os espectros foram coletados na região do infravermelho médio compreendida entre 4000 e 650 cm⁻¹, com 32 scans e resolução de 4 cm⁻¹. Os modelos foram desenvolvidos no *software* MATLAB® versão R2024a com scripts próprios. Os espectros ATR-FTIR foram organizados em uma matriz **X** que posteriormente foi normalizada. Os resultados das propriedades físico-químicas foram usados para compor um vetor **y** para cada propriedade. Os modelos SARGAN foram desenvolvidos para cada propriedade individualmente. O algoritmo SARGAN gerou uma matriz de dados sintéticos e um vetor sintético para cada propriedade baseados nos dados obtidos experimentalmente. Esses dados foram misturados com os dados experimentais originais. A partir disso, foram desenvolvidos dois modelos de regressão, o PLS e o SVR. Os modelos foram testados com as 36 amostras originais e com o conjunto de dados combinado (dados originais e os dados sintéticos) para comparação. Antes da construção dos modelos, as amostras foram separadas entre calibração (70%) e validação (30%) usando o algoritmo *Kennard-Stone* (KENNARD & STONE, 1969). Os modelos foram avaliados pela linearidade (coeficiente de determinação R²) e erros de calibração (RMSEC, do inglês *root mean squared error of calibration*) e validação (RMSEP, do inglês *root mean squared error of prediction*). As métricas de avaliação foram calculadas pelas Equações 1, 2 e 3, onde y_i é o valor de referência, $\hat{y}_i$ é o valor previsto pelo modelo, $\bar{y}_i$ é o valor médio de referência, $ncal$ é o número de amostras de calibração e $npred$ é o número de amostras de previsão (OLIVIERI, 2014).

$$RMSEC = \sqrt{\sum_{i=1}^{ncal} \frac{(y_i - \hat{y}_i)^2}{ncal - LV - 1}} \quad (1)$$

$$RMSEP = \sqrt{\sum_{i=1}^{npred} \frac{(y_i - \hat{y}_i)^2}{npred}} \quad (2)$$

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y}_i)^2} \quad (3)$$

Resultados

A Figura 1 traz os espectros originais (a) e os sintéticos (b) para o teor de umidade. Na qual, 84 amostras sintéticas foram produzidas pelo SARGAN. Essas sintéticas foram combinadas com as 36 amostras originais, totalizando-se um conjunto de 120 amostras. A quantidade de amostras utilizadas para a construção dos modelos de cada propriedade pode ser observada nas Tabelas 1 e 2.

Subtraindo-se os valores totais pela quantidade de amostras originais, tem-se que o SARGAN produziu 63 amostras para AA, 34 amostras para AD, AL e SA, 44 amostras para AR e pH, 54 amostras para C e 66 amostras para HMF.

Quanto aos resultados dos modelos de regressão, houve uma predominância dos modelos PLS sobre os de SVR. A Figura 2 mostra o gráfico dos melhores modelos com os valores de referência *versus* valores preditos para cada propriedade. Com exceção à AA e pH, todos os modelos SARGAN-PLS e SARGAN-SVR apresentaram resultados melhores que suas versões sem a geração de amostras sintéticas para todas as propriedades. A melhor estratégia para estimar AA foi o modelo PLS com R^2_p igual a 0,801. Os modelos SARGAN-PLS para AD, AL, AR, C, HMF, SA e U apresentaram R^2_p iguais a 0,870, 0,878, 0,852, 0,803, 0,736, 0,679 e 0,950, respectivamente. A propriedade pH foi a única que não apresentou valor de R^2_p satisfatório, com o modelo PLS sendo a estratégia que produziu o maior valor (0,433). Os melhores valores de RMSEP encontrados foram 0,021, 0,122, 0,511 meq/kg, 0,152%, 0,780%, 0,526%, 0,136, 0,708% e 0,163% para AA, AD, AL, AR, C, HMF, pH, SA e U, respectivamente.

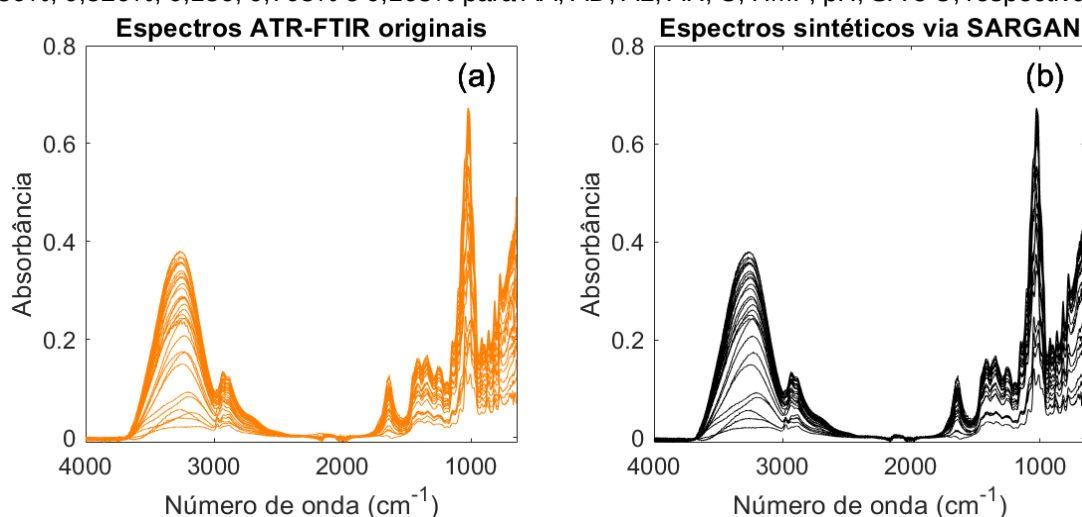


Figura 1. Espectros ATR-FTIR originais (a) e espectros sintéticos via SARGAN (b).

Fonte: Os autores.

Tabela 1. Resultados dos modelos de regressão para AA, AD, AL e AR.

Modelo	Prop	amostras	VL	RMSEC	RMSEP	R^2_c	R^2_p
SVR	AA	36	-	0,006	0,038	0,992	0,443
PLS	AA	36	4	0,030	0,021	0,815	0,801
SARGAN-SVR	AA	99	-	0,128	0,197	0,804	0,492
SARGAN-PLS	AA	99	10	0,367	0,331	0,792	0,790
SVR	AD	36	-	0,189	1146,000	0,981	0,448
PLS	AD	36	8	0,792	0,856	0,773	0,703
SARGAN-SVR	AD	70	-	0,212	0,636	0,973	0,779
SARGAN-PLS	AD	70	8	0,096	0,122	0,932	0,870
SVR	AL	36	-	56875,000	123503,000	0,300	0,740
PLS	AL	36	7	0,686	0,426	0,845	0,571
SARGAN-SVR	AL	70	-	0,115	2,047	1,000	0,952
SARGAN-PLS	AL	70	6	0,355	0,511	0,946	0,878
SVR	AR	36	-	32341,000	15796,000	0,748	0,467
PLS	AR	36	4	0,366	0,212	0,500	0,140
SARGAN-SVR	AR	80	-	0,081	0,784	0,999	0,941

SARGAN-PLS AR 80 9 0,115 0,152 0,936 0,852

Fonte: Os autores.

Tabela 2. Resultados dos modelos de regressão para C, HMF, pH, SA e U.

Modelo	Prop	amostras	VL	RMSEC	RMSEP	R ² _c	R ² _p
SVR	C	36	-	0,106	0,535	0,974	0,592
PLS	C	36	7	0,313	0,559	0,839	0,647
SARGAN-SVR	C	90	-	0,709	1,580	0,989	0,936
SARGAN-PLS	C	90	10	0,587	0,780	0,932	0,803
SVR	HMF	36	-	27569,000	4286,000	0,063	0,449
PLS	HMF	36	4	2,590	3,639	0,513	0,315
SARGAN-SVR	HMF	102	-	0,812	1,963	0,939	0,712
SARGAN-PLS	HMF	102	10	0,463	0,526	0,843	0,736
SVR	pH	36	-	0,285	0,198	0,043	0,300
PLS	pH	36	5	0,059	0,092	0,710	0,433
SARGAN-SVR	pH	80	-	0,351	0,332	0,337	0,226
SARGAN-PLS	pH	80	5	0,151	0,136	0,436	0,304
SVR	SA	36	-	20171,000	16496,000	0,563	0,726
PLS	SA	36	6	0,799	0,989	0,787	0,526
SARGAN-SVR	SA	70	-	0,221	1,072	0,993	0,804
SARGAN-PLS	SA	70	9	0,343	0,708	0,949	0,679
SVR	U	36	-	11851,000	31252,000	0,915	0,579
PLS	U	36	6	0,421	0,690	0,800	0,764
SARGAN-SVR	U	120	-	0,475	0,474	0,984	0,983
SARGAN-PLS	U	120	8	0,139	0,163	0,973	0,950

Fonte: Os autores.

Discussão

A maioria das propriedades apresentaram elevada correlação com os espectros ATR-FTIR, indicando que podem ser estimadas por eles. Com exceção dos modelos para atividade de água e pH, os modelos SARGAN-PLS e SARGAN-SVR apresentaram erros menores que as versões sem amostras sintéticas para todas as propriedades. Isso significa que o acréscimo de amostras aumenta a exatidão dos modelos de predição. Através da distribuição dos valores de cada propriedade, o SARGAN verifica a quantidade necessária de amostras a serem produzidas e isso gerou valores distintos em cada uma. O SARGAN cria valores sintéticos tanto para os dados espectrais quanto para as propriedades físico-químicas. Neste contexto, o objetivo é ajustar a distribuição dos dados para que ela se aproxime de uma distribuição normal (Gaussiana). Isso é benéfico para modelos lineares, pois a suposição de normalidade dos resíduos pode melhorar a eficácia e a confiabilidade das inferências feitas por esses modelos. Os modelos SARGAN-PLS mostraram altos valores de R² tanto na calibração quanto na previsão, sugerindo que a maioria das propriedades do mel segue uma distribuição linear. De acordo com os valores de R²_p, o modelo para umidade é capaz de explicar 95% da variabilidade nos valores de umidade durante a previsão. A umidade é detectada por volta de 3200 cm⁻¹ devido à absorção da hidroxila da água. A acidez livre, com 87,8% de variância explicada, é identificada por bandas carboxílicas em 1700 cm⁻¹ (PAULIUC *et al.*, 2021). A atividade diastásica, com 87,0% de variância, está associada a bandas de absorção típicas de estruturas enzimáticas. Os açúcares redutores, com 85,2% de variância, podem se correlacionar com a região de 1175 a 940 cm⁻¹.

correspondente ao alongamento C–O em carboidratos. As cinzas, com 80,3% de variância, estão relacionadas a minerais e compostos inorgânicos. O HMF, com 73,6% de variância, é detectado em torno de 1700 cm^{-1} devido a bandas carbonílicas (PAULIUC *et al.*, 2021). A sacarose aparente, com 67,9% de variância, é identificada por bandas associadas a glicose e frutose. Já atividade de água, a melhor estratégia foi o modelo PLS, indicando que não há necessidade de gerar amostras sintéticas, pois a propriedade tem boa relação com os espectros ATR-FTIR. O pH foi a única propriedade que não obteve um modelo satisfatório. Os valores de pH variaram entre 4 e 4,8, o que representa uma variação muito baixa entre as amostras. Isso pode ajudar a explicar os resultados insuficientes de R^2 para pH.

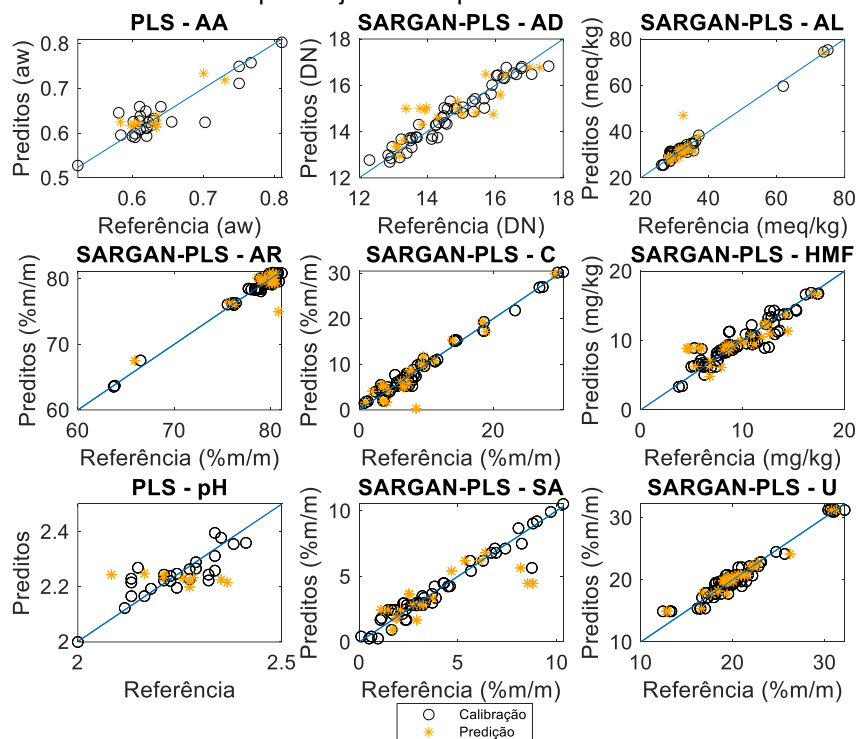


Figura 2. Melhores modelos de regressão para atividade de água (AA), atividade diastásica (AD), acidez livre (AL), açúcar redutor (AR), cinzas (C), hidroximetilfurfural (HMF), pH, sacarose aparente (SA) e umidade (U).

Fonte: Os autores.

Conclusão

Neste estudo, diversas propriedades físico-químicas do mel foram estimadas via ATR-FTIR e calibração multivariada com metodologia inovadora de geração de amostras sintéticas. Os modelos SARGAN-PLS predominaram como melhor estratégia, demonstrando que há uma enorme necessidade de gerar amostras sintéticas para se obter um conjunto mais representativo. Apenas atividade de água pode ser estimada diretamente com o PLS sem a necessidade de amostragem sintética e o pH possui baixa correlação com espectros ATR-FTIR. Este trabalho demonstra a aplicabilidade da combinação de quimiometria e espectroscopia ATR-FTIR no desenvolvimento de metodologias não invasivas e alternativas para a caracterização do mel.

Referências

ASSOCIAÇÃO BRASILEIRA DOS EXPORTADORES DE MEL (ABEMEL). **Dados estatísticos do mercado de mel**. Rio Claro - SP: ABEMEL, 2024. Disponível em: <https://www.brazilltsbee.com.br>. Acesso em: 08 ago. 2024.

ASSOCIATION OF OFFICIAL ANALYTICAL CHEMISTS. **Official methods of analysis of AOAC international**. 16. ed. Maryland: AOAC, 1997. 1141p.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. **Instrução Normativa nº 11, de 17 de novembro de 2000: Regulamento técnico de identidade e qualidade do mel.** Seção 001. Brasília - DF, 2000.

CODex ALIMENTARIUS COMMISSION (CAC). **Official methods of analysis.** Vol. 3, p.15-39, 1990.

COZZOLINO, D.; CORBELLA, E.; SMYTH, H. E. Quality control of honey using infrared spectroscopy: A review. **Applied Spectroscopy Reviews**, v. 46, n. 7, p. 523-538, 2011.

EL SOHAIMY, S. A.; MASRY, S. H. D.; SHEHATA, M. G. Physicochemical characteristics of honey from different origins. **Annals of Agricultural Sciences**, v. 60, n. 2, p. 279-287, 2015.

HE, H.; BAI, Y.; GARCIA, E. A.; LI, S. ADASYN: Adaptive Synthetic Sampling Approach for Imbalanced Learning. In: INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS (IJCNN), 2008, Hong Kong. **Anais...** Hong Kong: IEEE, 2008. p. 1322-1328.

KUROISHI, Alini Mari; QUEIROZ, Marise Bonifácio; ALMEIDA, Mareci Mendes de; QUAIST, Leda Battestin. Avaliação da cristalização de mel utilizando parâmetros de cor e atividade de água. **Braz. J. Food Technol.**, Campinas, v. 15, n. 1, p. 84-91, 2012.

KENNARD, RW; STONE, LA. Computer Aided Design of Experiments. **Technometric**. 11: p. 137-48, 1969.

OLIVIERI, AC. Analytical figures of merit: From univariate to multiway calibration, **Chem Rev.** 114: p. 5358-5378, 2014.

PAI, Ping-Feng; YANG, Shun-Ling; CHANG, Ping-Teng. Forecasting output of integrated circuit industry by support vector regression models with marriage honey-bees optimization algorithms. **Expert Systems with Applications**, v. 36, n. 7, p. 10746-10751, 2009.

PAULIUC, Daniela; CIURSA, Paula; ROPCIUC, Sorina; DRANCA, Florina; OROIAN, Mircea. Physicochemical parameters prediction and authentication of different monofloral honeys based on FTIR spectra. **Journal of Food Composition and Analysis**, v. 102, p. 104021, 2021.

SE, Kuan Wei; GHOSHAL, Sib Krishna; ABDUL WAHAB, Roswanira. Laser-induced breakdown spectroscopy unified partial least squares regression: An easy and speedy strategy for predicting Ca, Mg and Na content in honey. **Measurement**, v. 136, p. 1-10, 2019.

SILVA, I. B. V.; ADEODATO, P. J. L. PCA and Gaussian noise in MLP neural network training improve generalization in problems with small and unbalanced data sets. In: IEEE INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS (IJCNN), 2011, San Jose. **Anais...** San Jose: IEEE, 2011. p. 2664-2669.

SKAFF, W.; EL HAJJ, R.; HANNA-WAKIM, L.; ESTEAPHAN, N. Detection of adulteration in honey by infrared spectroscopy and chemometrics: Effect on human health. **Journal of Food Processing and Preservation**, v. 46, n. 10, e15438, 2022.

YANG, Y.; LI, Y.; YI, K.; WU, Z. Synthetic data approach for classification and regression. In: IEEE 29th INTERNATIONAL CONFERENCE ON APPLICATION-SPECIFIC SYSTEMS, ARCHITECTURES AND PROCESSORS (ASAP), 2018, Washington, DC. **Anais...** Washington, DC: IEEE, 2018. p. 1-8.

ZAMMIT YOUNG, Gerard-William; BLUNDELL, Renald. A review on the phytochemical composition and health applications of honey. **Heliyon**, v. 9, n. 2, e12507, 2023.

Agradecimentos

UFES, UFRGS, IFES, FEST, CAPES, FAPES, CNPq.