



MONITORAMENTO DA DINÂMICA DE CONDICIONAMENTO COM GLICEROL EM BLENDS DE TABACO POR IMAGEM HIPERESPECTRAL E APRENDIZADO DE MÁQUINA

João V. F. Gonçalves¹; Maurilio G. Nespeca¹; Douglas W. M. Flores¹; Denise White¹, Sri S. V. Aruchamy¹; Marcelo M. Caetano¹

¹British American Tobacco. Brazil-Labs, Avenida Frederico Augusto Ritter, 8000, Cachoeirinha, Rio Grande do Sul, Brasil.

joao_vitor_ferreira_goncalves@bat.com

Palavras-Chave: Quimiometria, espectroscopia NIR, controle de processo.

Introdução

A incorporação de agentes umectantes, como o glicerol, constitui uma etapa relevante no processamento de matrizes de tabaco, pois contribui para a retenção de umidade e para a manutenção das propriedades físicas do material durante o processamento e o armazenamento. O glicerol é amplamente empregado como umectante em produtos de tabaco, sendo sua aplicação associada ao aumento da capacidade de retenção de água e à redução da friabilidade do material (CARMINES; GAWORSKI, 2005). Entretanto, a resposta do tabaco ao condicionamento pode variar em função das características da matriz, tornando importante compreender a evolução temporal do material após a aplicação do umectante.

Métodos convencionais para caracterização de constituintes e umidade em matrizes complexas geralmente dependem de análises laboratoriais pontuais e podem exigir preparo de amostra. Nesse contexto, a imagem hiperespectral (*Hyperspectral Imaging*, HSI) combina informação espectral e espacial e permite avaliações rápidas, não destrutivas e sem contato, com potencial para monitoramento de atributos físicos e químicos em materiais agrícolas e alimentícios (MA et al., 2019; SAHA; MANICKAVASAGAN, 2021). Em tabaco, a HSI associada à Quimiometria já demonstrou viabilidade para classificação rápida de materiais segundo características de qualidade, cor e posição foliar, inclusive em aplicações industriais (MARCELO et al., 2019).

A elevada dimensionalidade e a colinearidade dos dados hiperespectrais tornam necessária a aplicação de métodos multivariados capazes de extrair padrões relevantes da assinatura espectral. A análise discriminante por mínimos quadrados parciais (PLS-DA) é amplamente utilizada para redução de dimensionalidade e discriminação de classes em dados espectroscópicos (BARKER; RAYENS, 2003), enquanto o Random Forest (RF) combina múltiplas árvores de decisão e apresenta robustez para classificação e estimação da importância relativa das variáveis (BREIMAN, 2001). A integração entre HSI e aprendizado de máquina tem ampliado o potencial dessas técnicas para classificação e monitoramento não destrutivo de produtos agrícolas e alimentícios (SAHA; MANICKAVASAGAN, 2021).



Embora aplicações de HSI em tabaco já tenham demonstrado desempenho em tarefas de classificação, seu uso para acompanhar temporalmente alterações espectrais associadas ao condicionamento com glicerol permanece pouco explorado. Além disso, uma resposta espectral observada após a aplicação do umectante não representa, necessariamente, uma medida direta da massa de glicerol absorvida, pois pode refletir simultaneamente alterações de água, distribuição do umectante e propriedades da própria matriz vegetal.

Dessa forma, o presente trabalho teve como objetivo avaliar o potencial da imagem hiperespectral SWIR associada a métodos quimiométricos e de aprendizado de máquina para monitorar a dinâmica espectral de blends de tabaco após o condicionamento com glicerol. Especificamente, buscou-se discriminar as condições inicial (0 min) e final de referência (1440 min), caracterizar a progressão temporal por meio de um índice espectral relativo, estimar o tempo de estabilização espectral dos blends e identificar as regiões do espectro mais relevantes para a discriminação temporal.

Material e Métodos

Foram avaliados quatro blends de tabaco, codificados como S1, S2, S3 e S4, após aplicação de glicerol durante o condicionamento. Para cada blend, foram acompanhadas seis unidades amostrais (T0–T5) em esquema temporal escalonado, com aquisições entre 0 e 360 min e uma condição final de referência após 24 h (1440 min). Em cada combinação unidade $\times$ tempo foram adquiridas cinco imagens hiperespectrais, tratadas como repetições técnicas. Das 1.075 aquisições previstas, 1.073 foram utilizadas após exclusão de duas imagens corrompidas.

As imagens foram adquiridas em refletância difusa com câmera Specim SWIR (Spectral Imaging Ltd., Oulu, Finlândia), detector MCT e objetiva OLES56 de 56 mm. O sistema opera nominalmente entre 1000 e 2500 nm; utilizaram-se 100 quadros s^{-1} , 7 ms de exposição e *binning* espectral e espacial igual a 1. A distância câmera–amostra foi de 70 cm. A refletância relativa foi calculada a partir das referências branca e escura. O fundo foi removido por segmentação espectral baseada em PLS-DA e, para a análise temporal, utilizou-se um espectro representativo por imagem. Foram mantidas 271 bandas entre 1004,41 e 2497,74 nm. Valores ausentes pontuais em bandas internas foram corrigidos por interpolação linear no eixo espectral. Os espectros foram pré-processados pela primeira derivada de Savitzky–Golay, com janela de 15 pontos e polinômio de segunda ordem.

Para cada blend, ajustou-se um Random Forest com 300 árvores para discriminar 0 e 1440 min. O desempenho foi estimado por validação *leave-one-unit-out*, mantendo todas as imagens de uma unidade T0–T5 exclusivamente no treinamento ou na validação em cada iteração. Foram calculadas acurácia balanceada, precisão, sensibilidade, especificidade, F1 e ROC-AUC por imagem e após agregação das aquisições técnicas por unidade $\times$ condição. Após a validação, o modelo foi reajustado com todos os espectros dos extremos e aplicado aos tempos intermediários. O escore contínuo de pertencimento à condição final foi



normalizado, dentro de cada unidade, pelos valores médios de 0 e 1440 min, originando o índice relativo de progressão espectral (IRPE), em escala de 0 a 100.

As cinco imagens de cada unidade $\times$ tempo foram resumidas por média e desvio-padrão do IRPE. Para cada uma das 24 trajetórias T0–T5 foram ajustados modelos exponencial assintótico e linear com platô, comparados pelo critério de informação de Akaike (AIC). O *breakpoint* do modelo linear com platô foi adotado como estimador principal do tempo de estabilização, mantendo-se o tempo para atingir IRPE = 95 no modelo exponencial como análise de sensibilidade. Os tempos foram resumidos por blend pela mediana das seis unidades, com intervalo de confiança *bootstrap* de 95% obtido por 10.000 reamostragens no nível da unidade.

Resultados e Discussão

Os modelos Random Forest discriminaram as condições inicial (0 min) e final de referência (1440 min) com elevado desempenho sob validação *leave-one-unit-out*. S1 e S2 apresentaram separação praticamente completa entre os extremos, enquanto S3 e S4 exibiram desempenho inferior, porém ainda consistente, com ROC-AUC superiores a 0,93. A manutenção do desempenho quando uma unidade T0–T5 inteira foi excluída da calibração indica que a discriminação não dependeu apenas de características específicas das imagens utilizadas no treinamento.

Tabela 1. Desempenho dos modelos Random Forest na discriminação entre as condições inicial (0 min) e final de referência (1440 min) sob validação *leave-one-unit-out*.

Blend	S1		S2		S3		S4	
Level	Img OOF	Unit OOF	Img OOF	Unit OOF	Img OOF	Unit OOF	Img OOF	Unit OOF
Balanced acc.	1.00	1.00	0.97	1.00	0.87	0.83	0.88	0.92
F1	1.00	1.00	0.97	1.00	0.86	0.83	0.88	0.91
ROC AUC	1.00	1.00	1.00	1.00	0.96	0.97	0.93	0.94

A validação agrupada é particularmente relevante em dados hiperespectrais, nos quais aquisições provenientes de uma mesma unidade tendem a compartilhar estrutura espectral. Assim, o desempenho observado fornece uma avaliação mais rigorosa da capacidade de generalização dos classificadores. A diferença de separabilidade entre os blends também indica que a magnitude da alteração espectral entre 0 e 24 h foi dependente da matriz. Estudos anteriores demonstraram que a HSI-NIR é sensível a variações de composição e qualidade em tabaco e que diferenças no teor de água afetam substancialmente sua assinatura NIR (MARCELO et al., 2019; LAI et al., 2025). Dessa forma, a classificação obtida deve ser interpretada como reconhecimento de uma mudança multivariada do estado espectral do material durante o condicionamento, e não como resposta exclusiva ao glicerol.

A aplicação dos classificadores aos tempos intermediários revelou dois padrões gerais de progressão. S1 e S2 apresentaram evolução mais lenta em direção à condição de 24 h, enquanto S3 e S4 atingiram valores elevados do índice relativo de progressão espectral

(IRPE) mais precocemente. A variabilidade entre T0–T5 foi relevante em alguns tempos intermediários, demonstrando que unidades de um mesmo blend não necessariamente seguiram trajetórias idênticas.

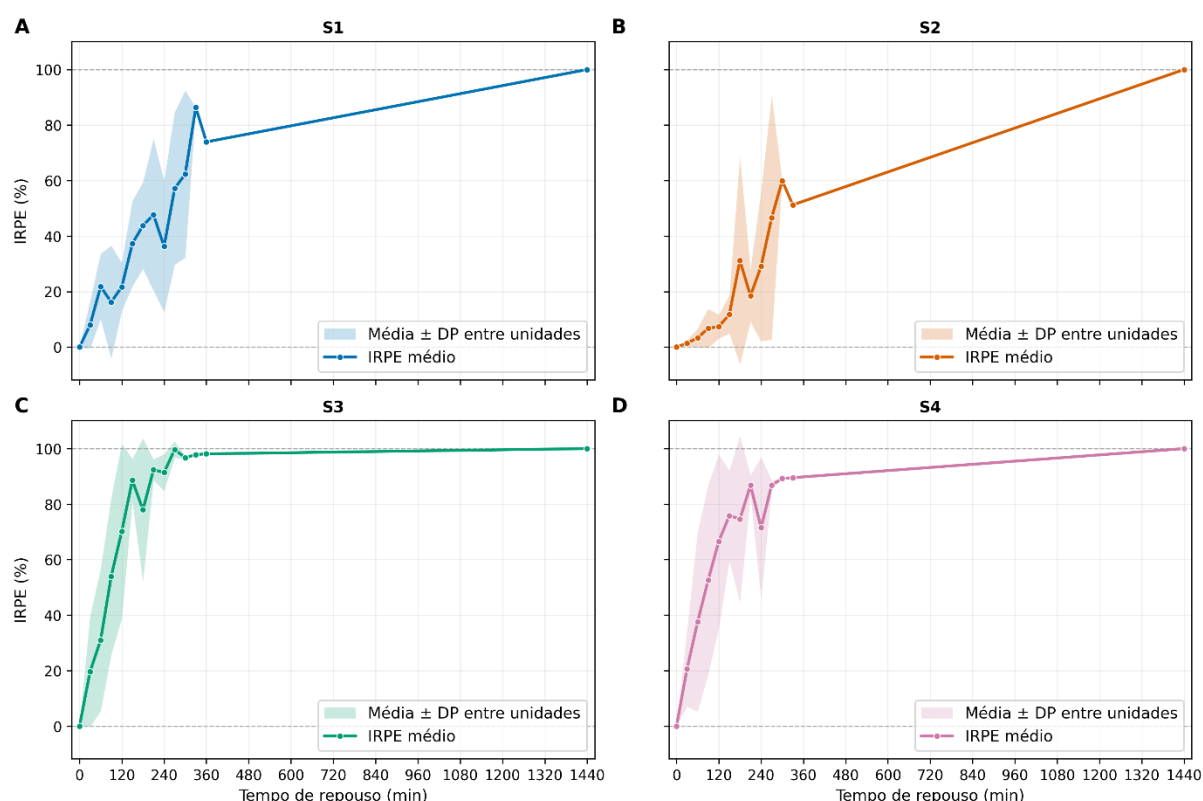


Figura 1. Evolução média do índice relativo de progressão espectral (IRPE) ao longo do tempo para os blends S1, S2, S3 e S4.

As diferenças observadas são compatíveis com uma resposta dependente da matriz, embora o delineamento não permita atribuí-las causalmente a um constituinte específico. O glicerol atua como umectante e apresenta elevada afinidade por água, enquanto estudos em matrizes de tabaco demonstram que sua presença pode aumentar a retenção e modificar o estado da água no material (CARMINES; GAWORSKI, 2005; TIAN et al., 2024). Esse aspecto é particularmente importante para a interpretação do IRPE, pois a resposta SWIR pode refletir simultaneamente redistribuição do glicerol, alterações no estado e teor de água e interações com os demais constituintes orgânicos da matriz. Consequentemente, o IRPE deve ser entendido como uma medida relativa da progressão espectral associada ao condicionamento, e não como percentual físico ou mássico de glicerol absorvido. A dispersão entre unidades reforça ainda a importância de preservar T0–T5 como unidades experimentais independentes, em vez de representar cada blend por uma única trajetória.

Entre os modelos avaliados, o linear com platô apresentou menor AIC em 19 das 24 trajetórias e foi adotado como modelo principal. As estimativas indicaram estabilização mais precoce para S3 e S4, com *breakpoints* medianos próximos de 160 min, e mais tardia para S1

e S2, com valores superiores a 400 min. O modelo exponencial assintótico foi mantido como análise de sensibilidade da dinâmica temporal.

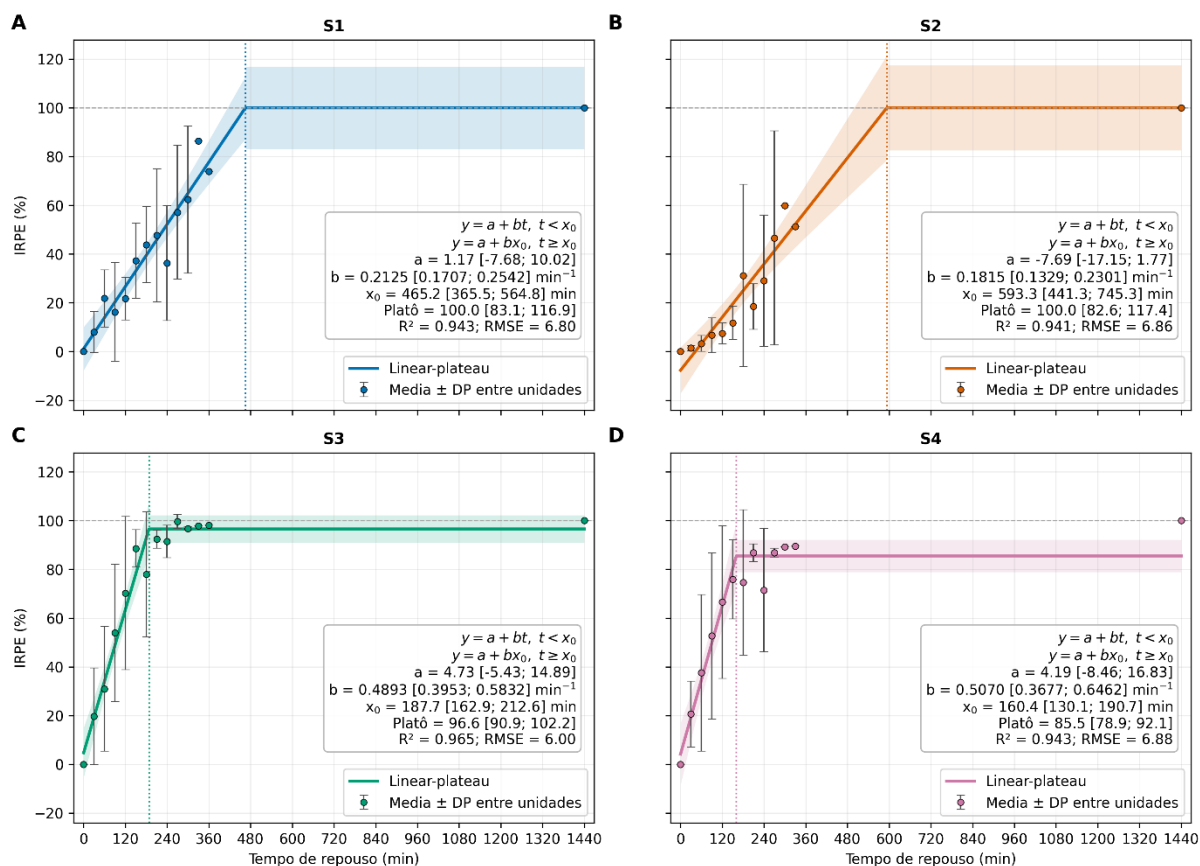


Figura 2. Tempos de estabilização espectral dos blends estimados pelos modelos linear com platô e exponencial assintótico, com intervalos de confiança obtidos por *bootstrap* das unidades amostrais.

A concordância entre o padrão temporal do IRPE e os breakpoints estimados reforça a existência de diferentes dinâmicas de condicionamento entre os materiais. A interpretação de S1 e S2, entretanto, requer cautela: em cinco das seis unidades de S1 e em todas as unidades de S2, o breakpoint ocorreu após o último tempo intermediário efetivamente amostrado, situando-se no intervalo entre aproximadamente 300–360 e 1440 min. Nesses casos, o tempo de estabilização representa uma estimativa do modelo e não a observação direta do momento em que o platô foi atingido. Para S3 e S4, a maior parte dos breakpoints ocorreu dentro da faixa temporal amostrada, conferindo maior suporte experimental à estabilização mais precoce. Além disso, a sobreposição dos intervalos de confiança não permite tratar os tempos estimados como diferenças estatísticas formais entre blends. Do ponto de vista de aplicação, os resultados demonstram principalmente a capacidade da metodologia de distinguir materiais com dinâmica espectral rápida e lenta, enquanto experimentos futuros deverão aumentar a resolução temporal após 300–360 min para refinar as estimativas de S1 e S2.

A importância das variáveis do Random Forest mostrou que a discriminação entre as condições inicial e final foi sustentada por múltiplas regiões do SWIR, com recorrência de contribuições aproximadamente em 1220–1260, 1680–1900 e 2250–2290 nm. A distribuição da importância, entretanto, variou entre os blends, reforçando o caráter multivariado e dependente da matriz da resposta espectral.

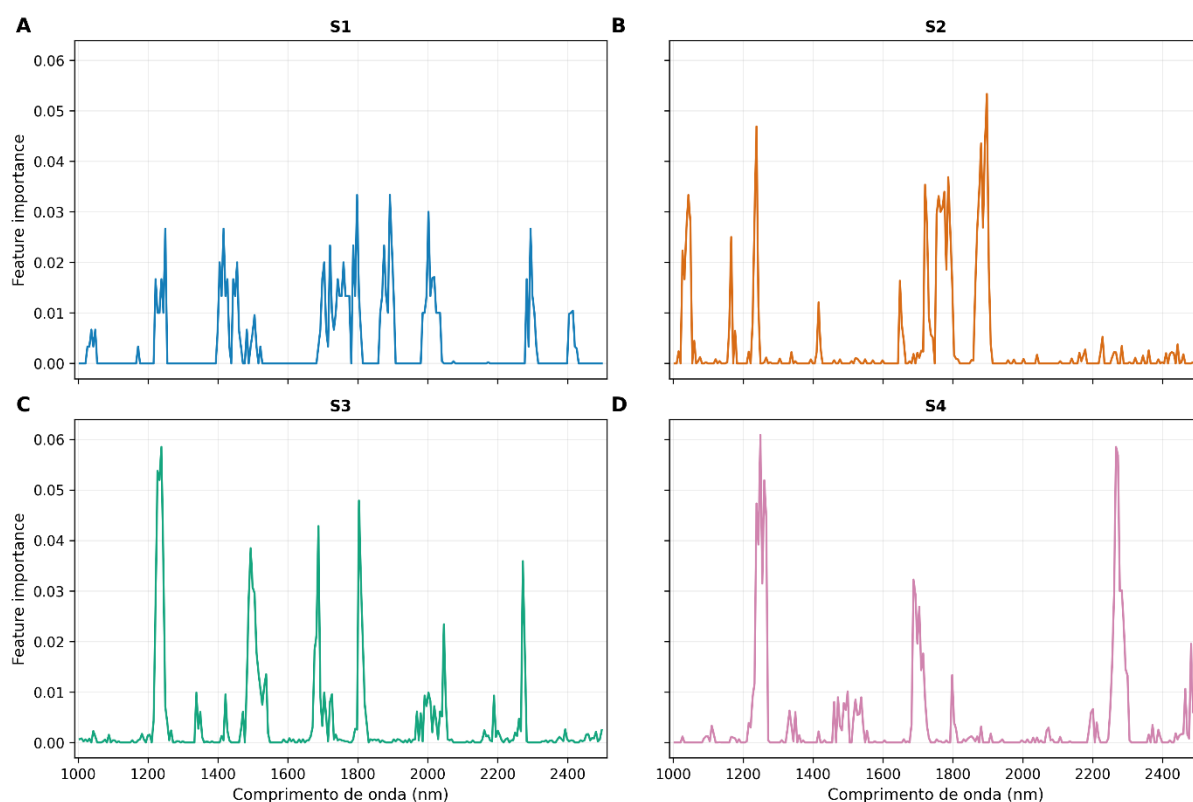


Figura 3. Importância relativa das bandas espectrais nos modelos Random Forest para discriminação entre as condições de 0 e 1440 min nos quatro blends de tabaco.

As regiões destacadas são quimicamente plausíveis para um sistema constituído por glicerol, água e matriz vegetal, mas não são específicas de um único componente. Em tabaco, absorções associadas à água são particularmente importantes próximo de 1450 e 1940 nm, enquanto sobretons e bandas de combinação relacionados a grupos C–H ocorrem aproximadamente entre 1670–1790 e 2220–2380 nm (LAI et al., 2025). Para o glicerol, foram reportadas bandas NIR próximas de 1500 e 2080 nm associadas a grupos contendo O–H e contribuições em aproximadamente 1710, 2270 e 2320 nm relacionadas a C–H e CH₂ (CHOWDHURY et al., 2023). A sobreposição dessas regiões com as variáveis relevantes dos modelos sustenta a sensibilidade espectral às transformações ocorridas durante o condicionamento, mas não permite atribuir a resposta exclusivamente ao glicerol. Adicionalmente, devido à elevada correlação entre bandas adjacentes no NIR, a importância por impureza do Random Forest deve ser interpretada como indicação de intervalos espectrais de interesse, e não como identificação causal de bandas químicas específicas.



Conclusões

A combinação entre imagem hiperespectral SWIR e aprendizado de máquina mostrou-se capaz de discriminar os estados inicial e final de referência e de acompanhar as alterações espectrais ocorridas durante o condicionamento dos diferentes blends. O índice relativo de progressão espectral (IRPE) permitiu representar a evolução temporal do processo, evidenciando dinâmicas distintas entre os materiais, com estabilização espectral mais precoce em S3 e S4 e mais tardia em S1 e S2. Para estes últimos, entretanto, os tempos estimados devem ser interpretados com cautela devido à menor resolução temporal entre as últimas avaliações intermediárias e 24 h. A contribuição de múltiplas regiões do SWIR reforça o caráter multivariado da resposta e sua provável associação conjunta às alterações de glicerol, água e matriz vegetal, não devendo ser interpretada como uma medida direta da quantidade de glicerol absorvida. Dessa forma, a metodologia apresenta potencial como ferramenta rápida e não destrutiva para monitoramento do condicionamento de matrizes de tabaco e suporte à definição de tempos de processo.

Agradecimentos

Somos gratos ao departamento BAT *Applied Chemometrics* por todo suporte.

Referências

- BARKER, M.; RAYENS, W. Partial least squares for discrimination. **Journal of Chemometrics**, 17(3), 166–173, 2003. DOI: 10.1002/cem.785.
- BREIMAN, L. Random Forests. **Machine Learning**, 45, 5–32, 2001. DOI: 10.1023/A:1010933404324.
- CARMINES, E. L.; GAWORSKI, C. L. Toxicological evaluation of glycerin as a cigarette ingredient. **Food and Chemical Toxicology**, 43(10), 1521–1539, 2005. DOI: 10.1016/j.fct.2005.04.010.
- CHOWDHURY, A. R.; BURNEY, U.; KING, D.; LEE, T.; HUTTER, D.; HUTTER, T.; Detection of Toxic Contaminants in Alcohol-Based Hand Sanitizers Using Infrared Spectroscopy. **Applied Spectroscopy Practica**, 1(2), 2023. DOI: 10.1177/27551857231204630.
- MA, J.; SUN, D.-W.; PU, H.; CHENG, J.-H.; WEI, Q. Advanced Techniques for Hyperspectral Imaging in the Food Industry: Principles and Recent Applications. **Annual Review of Food Science and Technology**, 10, 197–220, 2019. DOI: 10.1146/annurev-food-032818-121155.
- MARCELO, M. C. A. et al. Fast inline tobacco classification by near-infrared hyperspectral imaging and support vector machine-discriminant analysis. **Analytical Methods**, 11, 1966–1975, 2019. DOI: 10.1039/C9AY00413K.
- SAHA, D.; MANICKAVASAGAN, A. Machine learning techniques for analysis of hyperspectral images to determine quality of food products: A review. **Current Research in Food Science**, 4, 28–44, 2021. DOI: 10.1016/j.crfs.2021.01.002.
- TIAN, H.; GAO, Z.; HAN, L.; WAN, J.; YANG, X.; LI, M.; CHU, W.; LAI, M.; JI, X.; Analysis of Water Adsorption Capacity, Thermal Behavior, Antioxidant Activity, and Smoke Volume by Polyol Ester in Reconstituted Tobacco. **ACS Omega**, 9(44), 44336–44346. DOI: 10.1021/acsomega.4c05199.