BERNARDO SIQUEIRA COSTA BARBOSA

PREDICTIVE MODELING FOR PREDICTION OF TECHNOLOGICAL MATURITY AND MICROTERROIR ZONES OF GRAPEVINES IN SUBTROPICAL REGION IN BRAZIL

LAVRAS – MG
2026

BERNARDO SIQUEIRA COSTA BARBOSA

PREDICTIVE MODELING FOR PREDICTION OF TECHNOLOGICAL MATURITY AND MICROTERROIR ZONES OF GRAPEVINES IN SUBTROPICAL REGION IN BRAZIL

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-graduação em Agronomia/Fitotecnia, área de concentração em Produção Vegetal, para a obtenção do título de Mestre.

Orientador
Prof. Dr. Rafael Pio

Coorientador
Prof. Dr. Pedro Maranha Peche

LAVRAS – MG
2026

Ficha Catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).

Barbosa, Bernardo Siqueira Costa.
Predictive modeling for prediction of technological maturity and microterroir zones of grapevines in subtropical region in Brazil / Bernardo Siqueira Costa Barbosa. - 2026.
84 p. : il.

Orientador: Rafael Pio
Coorientador: Pedro Maranha Peche

Dissertação (Mestrado Acadêmico) - Universidade Federal de Lavras, 2026.
Bibliografia.

1. Viticultura de precisão. 2. Variabilidade espectral. 3. Vitis vinifera L.. I. Pio, Rafael. II. Maranha Peche, Pedro. III. Universidade Federal de Lavras. IV. Título.

BERNARDO SIQUEIRA COSTA BARBOSA

PREDICTIVE MODELING FOR PREDICTION OF TECHNOLOGICAL MATURITY AND MICROTERROIR ZONES OF GRAPEVINES IN SUBTROPICAL REGION IN BRAZIL

MODELAGEM PREDITIVA PARA A PREVISÃO DA MATURIDADE TECNOLÓGICA E DE ZONAS DE MICROTERROIR DE VIDEIRAS EM REGIÃO SUBTROPICAL DO BRASIL

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-graduação em Agronomia/Fitotecnia, área de concentração em Produção Vegetal, para a obtenção do título de Mestre.

APROVADA em 10 de julho de 2026.

Prof. Dr. Rafael Pio	UFLA
Prof. Dr. Pedro Maranha Peche	UFLA
Prof. Dra. Leila Aparecida Salles Pio	UFLA
Dr. Pedro Henrique Abreu Moura	EPAMIG

gov.br
Documento assinado digitalmente
RAFAEL PIO
Data: 24/09/2026 10:41:29-0300
Verifique em https://validar.iti.gov.br

Orientador
Prof. Dr. Rafael Pio

Coorientador
Prof. Dr. Pedro Maranha Peche

LAVRAS – MG
2026

À minha Família, em especial aos meus pais Valuce e Gabriela.

Dedico

AGRADECIMENTOS

À Universidade Federal de Lavras (UFLA) e ao Programa de Pós-Graduação em Agronomia/Fitotecnia pela estrutura concedida. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) – Código de Financiamento 001, da Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) e do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), aos quais agradeço pelo fomento à pesquisa.

Ao meu orientador, Prof. Dr. Rafael Pio, e ao coorientador, Prof. Dr. Pedro Maranha Peche, pela condução técnica, confiança e pelas oportunidades de desenvolvimento científico ao longo deste projeto. Aos professores Dra. Ana Claudia Costa, Dra. Leila Aparecida Salles Pio e Dr. Antônio Decarlos Neto, pelo suporte acadêmico e ensinamentos.

Aos meus pais, Valuce Antonioni Barbosa e Gabriela Siqueira Costa Barbosa, pela excelente educação, ensinamentos práticos e por fornecerem toda a base estrutural que me permitiu focar neste objetivo.

Às minhas irmãs, Giovana Siqueira Costa Barbosa e Manuela Siqueira Costa Barbosa, pelo suporte contínuo e presença essencial ao longo de todo esse processo.

Um agradecimento em especial à minha namorada, Maria Gabrielle dos Reis Morais, pelo apoio constante, pela parceria e pelos bons conselhos, que foram fundamentais durante todo este período.

Aos colegas do Setor de Fruticultura e alunos de iniciação científica, Jucimar, Eneidy, Talis e Denny, pelo auxílio direto na execução dos experimentos a campo e pelo excelente trabalho em equipe.

A todos que, direta ou indiretamente, colaboraram para a entrega deste trabalho.

ABSTRACT

Determining grape quality through conventional methods is a destructive and costly laboratory process that limits the understanding of spatial and temporal variability within vineyards. This study proposes to validate non-destructive predictive approaches, grounded in precision viticulture and machine learning, for monitoring technological maturity and delineating microterroir zones. The research focuses on the double-pruning regime (winter wine production) in subtropical regions of Brazil, investigated under two complementary perspectives of scale and data acquisition. In the first study, the focus was on interseasonal macroprediction (2024–2025) and microterroir zoning in a subtropical area. The integration of Growing Degree-Days (GDD) and Vegetation Indices (VIs) revealed the Random Forest algorithm as the most robust for the temporal prediction of Soluble Solids (SS, $R^2 = 0.73$) and Titratable Acidity (TA, $R^2 = 0.71$), although pH modeling exhibited performance degradation on independent datasets. Concurrently, K-Means unsupervised clustering integrated multi-year canopy vigor (via Google Earth Engine), topography, and soil proxies to delineate microterroir zones, achieving high efficacy in the stratification of macrophysiological traits (Variance Reduction of 33.62% for chlorophyll b). In the second study, a high-spatial-resolution predictive framework was developed for the same management system. The fusion of agrometeorological data with multispectral imagery acquired by an Unmanned Aerial Vehicle (UAV) significantly outperformed the isolated use of VIs. The XGBoost algorithm achieved the highest accuracy ($R^2 = 0.89$ for SS; $R^2 = 0.77$ for TA), demonstrating that GDD and cultivar determine the primary maturation dynamics, while VIs capture the refinement of spatial variability in vigor at a sub-metric scale. In summary, it is concluded that integrating climatic data with remote sensing platforms (orbital and aerial) provides a robust analytical framework for precision viticulture in double-pruning areas, optimizing the assessment of technological maturity, the management of microterroir zones, and harvest decisions.

Keywords: Precision Viticulture; Machine Learning; Winter Wines; Remote Sensing; Viticultural Zoning; Growing Degree-Days.

RESUMO

A determinação da qualidade da uva por métodos convencionais é um processo laboratorial destrutivo e oneroso, que limita a compreensão da variabilidade espacial e temporal nos vinhedos. Este trabalho propõe validar abordagens preditivas não destrutivas, fundamentadas em viticultura de precisão e aprendizado de máquina, para o monitoramento da maturação tecnológica e a delimitação de zonas de microterroir. A pesquisa concentra-se no regime de dupla poda (produção de vinhos de inverno) para regiões subtropicais no Brasil, investigado sob duas perspectivas complementares de escala e aquisição de dados. No primeiro estudo, focou-se na macropredição intersazonal (2024–2025) e no zoneamento de *microterroir* em área subtropical. A integração de Graus-Dia Acumulados (GDA) e Índices de Vegetação (IVs) revelou o algoritmo *Random Forest* como o mais robusto para a predição temporal de Sólidos Solúveis (SS, $R^2 = 0,73$) e Acidez Tartárica (AT, $R^2 = 0,71$), embora a modelagem do pH tenha apresentado degradação em dados independentes. Concomitantemente, o agrupamento não supervisionado *K-Means* integrou o vigor multianual do dossel (via *Google Earth Engine*), topografia e *proxies* de solo para delimitar zonas de *microterroir*, alcançando elevada eficácia na estratificação de características macrofisiológicas (Redução de Variância de 33,62% para clorofila *b*). No segundo estudo, desenvolveu-se uma estrutura preditiva de alta resolução espacial para o mesmo sistema de manejo. A fusão de dados agrometeorológicos com imagens multiespectrais obtidas por Veículo Aéreo Não Tripulado (VANT) superou significativamente o uso isolado de IVs. O algoritmo *XGBoost* obteve a maior acurácia ($R^2 = 0,89$ para SS; $R^2 = 0,77$ para AT), demonstrando que o GDA e a cultivar determinam a dinâmica primária de maturação, enquanto os IVs capturam o refinamento da variabilidade espacial do vigor em escala submétrica. Em síntese, conclui-se que a integração de dados climáticos com plataformas de sensoriamento remoto (orbital e aéreo) fornece um arcabouço analítico robusto para a viticultura de precisão em áreas de dupla poda, otimizando a avaliação da maturidade tecnológica, a gestão de zonas de *microterroir* e as decisões de colheita.

Palavras-chave: Viticultura de Precisão; Aprendizado de Máquina; Vinhos de Inverno; Sensoriamento Remoto; Zoneamento Vitícola; Graus-Dia Acumulados.

IMPACT INDICATORS

This dissertation presents potential and direct technological, economic, and social impacts, focusing on the territory of the viticultural regions of southeastern Brazil and subtropical areas producing winter wines. The directly benefited public encompasses rural producers, oenologists, and agronomists inserted in the local productive arrangement, whose field routines are optimized by reducing laboratory costs and time in determining grape maturity, characterizing a direct technological and economic impact. The research has a strong extensionist character by promoting the transfer of precision viticulture technologies from the academic environment of UFLA to external society, using partner rural properties as demonstrative and validation areas for algorithms and unmanned aerial vehicles, actively integrating the producer into agronomic solutions. This extension and research action involved the direct participation of 5 professors and 8 students, consolidating in situ scientific knowledge diffusion. In accordance with the guidelines of the National Extension Policy, the research impacts are primarily classified in thematic area 7, Technology and Production, for the development of predictive tools applied to agribusiness. Additionally, the obtained results align with the UN Sustainable Development Goals (SDGs), notably SDG 2 (Zero Hunger and Sustainable Agriculture), by promoting resilient and precision agricultural practices and SDG 9 (Industry, Innovation and Infrastructure), through the adoption of machine learning models and remote sensing in the modernization of the wine sector.

INDICADORES DE IMPACTO

A presente dissertação apresenta impactos potenciais e diretos de ordem tecnológica, econômica e social, com foco no território das regiões vitivinícolas do Sudeste brasileiro e áreas subtropicais produtoras de vinhos de inverno. O público diretamente beneficiado engloba produtores rurais, enólogos e agrônomos inseridos no arranjo produtivo local, cujas rotinas de campo são otimizadas pela redução de custos laboratoriais e de tempo na determinação da maturidade da uva, caracterizando um impacto tecnológico e econômico direto. O trabalho possui um forte caráter extensionista ao promover a transferência de tecnologias de viticultura de precisão do ambiente acadêmico da UFLA para a sociedade externa, utilizando propriedades rurais parceiras como áreas demonstrativas e de validação de algoritmos e veículos aéreos não tripulados, integrando ativamente o produtor nas soluções agronômicas. Esta ação extensionista e de pesquisa envolveu a participação direta de 5 docentes e 8 estudantes, consolidando a difusão do conhecimento científico in loco. Em consonância com as diretrizes da Política Nacional de Extensão, os impactos da pesquisa classificam-se primordialmente na área temática 7, Tecnologia e Produção, pelo desenvolvimento de ferramentas preditivas aplicadas ao agronegócio. Adicionalmente, os resultados obtidos alinham-se aos Objetivos de Desenvolvimento Sustentável (ODS) da ONU, destacando-se o ODS 2 (Fome Zero e Agricultura Sustentável), por promover práticas agrícolas resilientes e de precisão e o ODS 9 (Indústria, Inovação e Infraestrutura), mediante a adoção de modelos de aprendizado de máquina e sensoriamento remoto na modernização do setor vitivinícola.

LISTA DE SIGLAS

CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
Embrapa	Empresa Brasileira de Pesquisa Agropecuária
IBICT	Instituto Brasileiro de Informação em Ciência e Tecnologia
UFLA	Universidade Federal de Lavras

Sumário

PRIMEIRA PARTE

INTRODUÇÃO

A área mundial de vinhedos apresenta um declínio estrutural, com redução de 0,8% em 2025, totalizando sete milhões de hectares. Esta retração marca o sexto ano consecutivo de queda, impulsionada pelo arranque de videiras em países produtores de ambos os hemisférios. Em contrapartida, a viticultura brasileira diverge dessa tendência global ao registrar seu quinto ano consecutivo de expansão. A superfície cultivada no Brasil atingiu 91 mil hectares em 2025, um crescimento interanual de 9,6%, o que consolida a participação do país com 1,3% da área vitícola mundial (OIV, 2026).

No âmbito produtivo, a elaboração global de vinho foi estimada em 227 milhões de hectolitros em 2025, um incremento marginal de 0,6% em relação ao ano anterior. Este volume configura o terceiro ano consecutivo de baixa produção, reflexo da volatilidade climática e de ajustes estratégicos frente à retração da demanda internacional. Em contraste, a produção vitivinícola brasileira apresentou forte recuperação, alcançando 2,8 milhões de hectolitros no mesmo período. Esse resultado representa um salto de 80,6% frente à safra de 2024 e situa-se 15,8% acima da média dos últimos cinco anos, desempenho viabilizado por condições climáticas favoráveis nas principais regiões produtoras do país (OIV, 2026).

Neste contexto, a viticultura global tem testemunhado o surgimento de novas regiões produtoras que empregam técnicas de manejo inovadoras para obter vinhos de alta qualidade em climas anteriormente considerados marginais. A viticultura brasileira, particularmente na região Sudeste (área subtropical), tem sido impulsionada pela expansão da produção de vinhos finos devido à implementação da técnica de dupla poda (BRANT et al., 2020). Essa prática viabiliza a produção dos vinhos de inverno, invertendo o ciclo de colheita da videira da tradicional estação chuvosa de verão para o período mais seco de inverno, melhorando a qualidade do fruto sob condições climáticas tropicais (FAVERO et al., 2011) e promovendo a concentração de açúcares e compostos fenólicos nas bagas (REGINA et al., 2019).

O sucesso desta e de outras práticas vitícolas depende fundamentalmente da capacidade de determinar com precisão o momento ideal de colheita, ponto definido pelo conceito de maturação tecnológica, o qual representa o equilíbrio ideal entre açúcares, medidos como Sólidos Solúveis (°Brix), e a Acidez Tartárica (AT), parâmetros que governam diretamente o potencial alcoólico, a estabilidade e o perfil sensorial do vinho (PENSO et al., 2014). A determinação precisa do ponto de colheita é um fator decisivo na viticultura, pois a qualidade da uva no momento da colheita influencia diretamente o potencial do vinho final (CHEN et al., 2015). Tradicionalmente, o monitoramento para essa decisão é realizado por meio de amostragem manual de bagas no campo para análise laboratorial. Contudo, esse método apresenta limitações significativas: é destrutivo, oneroso, exige muita mão de obra e oferece baixa representatividade espacial. Dada a alta variabilidade intra-vinhedo que a maturação pode exibir, a amostragem pontual falha em capturar a verdadeira condição da cultura, podendo levar a decisões de colheita inadequadas que comprometem a qualidade e a homogeneidade do produto final (BRAMLEY, 2010a).

Para contornar essas limitações, a viticultura de precisão busca explorar a mais ampla gama de observações disponíveis para descrever a variabilidade espacial do vinhedo com alta resolução. Satélites têm sido utilizados na agricultura de precisão há mais de 40 anos, quando o Landsat 1 foi lançado em órbita em 1972 (MATESE; DI GENNARO, 2015). A primeira aplicação do sensoriamento remoto na agricultura de precisão ocorreu quando imagens Landsat de solo exposto foram utilizadas para estimar padrões espaciais no teor de matéria orgânica do solo (BHATTI; MULLA; FRAZIER, 1991). Baseada nesses preceitos, a Modelagem Preditiva dedica-se à predição temporal de atributos enológicos chave. Esta fase busca desenvolver e validar modelos preditivos robustos capazes de estimar atributos de maturação da uva, nomeadamente Concentração de Sólidos Solúveis, pH e Acidez Tartárica (KASIMATI et al., 2021). A predição baseia-se na integração de dados de sensoriamento remoto (índices de vegetação) e variáveis térmicas (Graus-Dia Acumulados) (PARKER et al., 2011). Esta capacidade preditiva pode ser validada com sucesso utilizando uma estação de crescimento subsequente e independente, demonstrando a generalização temporal dos modelos.

Entretanto, enquanto a modelagem preditiva aborda a previsão temporal dos atributos enológicos, ela deve ser acoplada à delimitação de zonas de *microterroir* para o manejo eficaz da variabilidade espacial dentro da parcela do vinhedo. Essa abordagem espacial concentra-se na caracterização da heterogeneidade estrutural por meio da integração de fatores ambientais e respostas vegetativas de longo prazo (ACEVEDO-OPAZO et al., 2008; PRIORI et al., 2013).

A hipótese central deste trabalho é que a fusão de preditores espaciais (IVs) e temporais (GDA) por meio de modelos de aprendizado de máquina pode superar as limitações das abordagens individuais, gerando estimativas mais precisas e robustas da maturidade tecnológica. Portanto, o objetivo primário deste estudo foi desenvolver modelos preditivos temporais para a maturação tecnológica da uva e delimitar zonas espaciais de *microterroir* sob condições subtropicais no Brasil, bem como desenvolver e avaliar modelos de aprendizado de máquina para a predição não destrutiva de Sólidos Solúveis (°Brix) e Acidez Tartárica cultivadas sob regime de dupla poda. Os objetivos específicos foram: (i) investigar se a inclusão do GDA como variável preditora melhora significativamente a acurácia dos modelos baseados em IVs; (ii) identificar os IVs mais influentes para a predição da maturação da uva; (iii) determinar o algoritmo de aprendizado de máquina de melhor desempenho para esta tarefa; e (iv) identificar as variáveis mais influentes para a geração de zonas de microterroir.

REFERÊNCIAS

ACEVEDO-OPAZO, C.; TISSEYRE, B.; GUILLAUME, S.; OJEDA, H. The potential of high spatial resolution information to define within-vineyard zones related to vine water status. *In*: PRECISION AGRICULTURE 2008, **Anais** [...]. [s.l: s.n.] p. 285–302. DOI: 10.1007/s11119-008-9073-1.

AGGOGERI, Francesco; PELLEGRINI, Nicola; TAGLIANI, Franco Luis. **Recent advances on machine learning applications in machining processes**. **Applied Sciences (Switzerland)**MDPI, , 2021. DOI: 10.3390/app11188764.

ANIĆ, Marina; KONTIĆ, Jasminka Karoglan; RENDULIĆ, Nera; ČARIJA, Mate; OSREČAK, Mirela; KAROGLAN, Marko; ANDABAKA, Željko. Evolution of Leaf Chlorophylls, Carotenoids and Phenolic Compounds during Vegetation of Some Croatian Indigenous Red and White Grape Cultivars. **Plants 2024, Vol. 13, Page 971**, *[S. l.]*, v. 13, n. 7, p. 971, 2024. DOI: 10.3390/PLANTS13070971. Disponível em: https://www.mdpi.com/2223-7747/13/7/971/htm. Acesso em: 27 jun. 2026.

ARNÓ, J.; MARTÍNEZ-CASASNOVAS, J. A.; RIBES-DASI, M.; ROSELL, J. R. Revisión. Viticultura de precisión. Líneas de investigación, retos y oportunidades del manejo sitio-específico en viña. **Spanish Journal of Agricultural Research**, *[S. l.]*, v. 7, n. 4, p. 779–790, 2009. DOI: 10.5424/SJAR/2009074-1092. Acesso em: 26 jun. 2026.

BACK, Alvaro José; BRUNA, Emílio Della. Soma térmica nas fases fenológicas da videira ‘Niágara Rosada’ cultivada em Urussanga, Santa Catarina. **Agropecuária Catarinense**, *[S. l.]*, v. 25, n. 3, p. 85–89, 2012. Disponível em: https://publicacoes.epagri.sc.gov.br/rac/article/view/670. Acesso em: 13 jun. 2025.

BARBOZA, Thiago Orlando Costa; FERRAZ, Marcelo Araújo Junqueira; PILON, Cristiane; VELLIDIS, George; VALERIANO, Taynara Tuany Borges; DOS SANTOS, Adão Felipe. Advanced Farming Strategies Using NASA POWER Data in Peanut-Producing Regions without Surface Meteorological Stations. **AgriEngineering**, *[S. l.]*, v. 6, n. 1, p. 438–454, 2024. DOI: 10.3390/AGRIENGINEERING6010027/S1. Disponível em: https://www.mdpi.com/2624-7402/6/1/27/htm. Acesso em: 13 jun. 2025.

BOIS, B.; WALD, L.; PIERI, Ph; VAN LEEUWEN, C.; COMMAGNAC, L.; CHERY, Ph; CHRISTEN, M.; GAUDILLÈRE, J. P.; SAUR, É. Estimating spatial and temporal variations in solar radiation within Bordeaux winegrowing region using remotely sensed data. **OENO One**, *[S. l.]*, v. 42, n. 1, p. 15–25, 2008. DOI: 10.20870/OENO-ONE.2008.42.1.829. Disponível em: https://oeno-one.eu/article/view/829. Acesso em: 27 jun. 2026.

BRAMLEY, R. G. V. Understanding variability in winegrape production systems 2. Within vineyard variation in quality over several vintages. **Australian Journal of Grape and Wine Research**, *[S. l.]*, v. 11, n. 1, p. 33–42, 2005. DOI: 10.1111/J.1755-0238.2005.TB00277.X. Disponível em: /doi/pdf/10.1111/j.1755-0238.2005.tb00277.x. Acesso em: 16 maio. 2026.

BRAMLEY, R. G. V. Precision Viticulture: managing vineyard variability for improved quality outcomes. **Managing Wine Quality: Viticulture and Wine Quality**, *[S. l.]*, p. 445–480, 2010. a. DOI: 10.1533/9781845699284.3.445. Disponível em: https://www.sciencedirect.com/science/chapter/edited-volume/abs/pii/B9781845694845500129. Acesso em: 4 jul. 2026.

BRAMLEY, R. G. V. Precision Viticulture: managing vineyard variability for improved quality outcomes. **Managing Wine Quality: Viticulture and Wine Quality**, *[S. l.]*, p. 445–480, 2010. b. DOI: 10.1533/9781845699284.3.445. Disponível em: https://www.sciencedirect.com/science/chapter/edited-volume/abs/pii/B9781845694845500129. Acesso em: 26 jun. 2026.

BRAMLEY, R. G. V.; HAMILTON, R. P. Terroir and precision viticulture: are they compatible ? **OENO One**, *[S. l.]*, v. 41, n. 1, p. 1–8, 2007. DOI: 10.20870/OENO-ONE.2007.41.1.855. Disponível em: https://oeno-one.eu/article/view/855. Acesso em: 5 out. 2025.

BRAMLEY, R. G. V.; OUZMAN, J.; BOSS, P. K. Variation in vine vigour, grape yield and vineyard soils and topography as indicators of variation in the chemical composition of grapes, wine and wine sensory attributes. **Australian Journal of Grape and Wine Research**, *[S. l.]*, v. 17, n. 2, p. 217–229, 2011. DOI: 10.1111/J.1755-0238.2011.00136.X. Acesso em: 5 out. 2025.

BRAMLEY, R. G. V.; OUZMAN, J.; THORNTON, C. Selective harvesting is a feasible and profitable strategy even when grape and wine production is geared towards large fermentation volumes. **Australian Journal of Grape and Wine Research**, *[S. l.]*, v. 17, n. 3, p. 298–305, 2011. DOI: 10.1111/J.1755-0238.2011.00151.X;WGROUP:STRING:PUBLICATION. Disponível em: /doi/pdf/10.1111/j.1755-0238.2011.00151.x. Acesso em: 24 jul. 2026.

BRAMLEY, R. G. V.; HAMILTON, R. P. **Hitting the zone - making viticulture more precise**. Urrbrae, S. Aust.: . Acesso em: 5 out. 2025.

BRANT, Luciana Alves Caldeira; DE SOUZA, Cláudia Rita; DA MOTA, Renata Vieira; FERNANDES, Fernanda de Paula; GONÇALVES, Mariana Gabriele Marcolino; DE MENEZES, Michele Duarte; PEREGRINO, Isabela; CURI, Nilton; REGINA, Murillo de Albuquerque. Macro scale analysis of syrah vineyards under winter growing cycles: Agronomical and ecophysiological responses. **Scientia Agricola**, *[S. l.]*, v. 78, n. 6, p. 1–9, 2020. DOI: 10.1590/1678-992x-2020-0055.

BREIMAN, Leo. Random forests. **Machine Learning**, *[S. l.]*, v. 45, n. 1, p. 5–32, 2001. DOI: 10.1023/A:1010933404324/METRICS. Disponível em: https://link.springer.com/article/10.1023/A:1010933404324. Acesso em: 14 jun. 2025.

CHAI, T.; DRAXLER, R. R. Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature. **Geoscientific Model Development**, *[S. l.]*, v. 7, n. 3, p. 1247–1250, 2014. DOI: 10.5194/GMD-7-1247-2014,. Acesso em: 14 jun. 2025.

CHEN, Shanshan; ZHANG, Fangfang; NING, Jifeng; LIU, Xu; ZHANG, Zhenwen; YANG, Shuqin. Predicting the anthocyanin content of wine grapes by NIR hyperspectral imaging. **Food Chemistry**, *[S. l.]*, v. 172, p. 788–793, 2015. DOI: 10.1016/J.FOODCHEM.2014.09.119. Disponível em: https://www.sciencedirect.com/science/article/abs/pii/S0308814614015040?via%3Dihub. Acesso em: 4 jul. 2026.

CHEN, Tianqi; GUESTRIN, Carlos. XGBoost: A Scalable Tree Boosting System. **Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**, *[S. l.]*, v. 13-17- August-2016, p. 785–794, 2016. DOI: 10.1145/2939672.2939785. Disponível em: https://arxiv.org/pdf/1603.02754. Acesso em: 14 jun. 2025.

CONRAD, O.; BECHTEL, B.; BOCK, M.; DIETRICH, H.; FISCHER, E.; GERLITZ, L.; WEHBERG, J.; WICHMANN, V.; BÖHNER, J. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. **Geoscientific Model Development**, *[S. l.]*, v. 8, n. 7, p. 1991–2007, 2015. DOI: 10.5194/GMD-8-1991-2015. Disponível em: www.geosci-model-dev.net/8/1991/2015/. Acesso em: 20 out. 2025.

CORTES, Corinna; VAPNIK, Vladimir; SAITTA, Lorenza. Support-vector networks. **Machine Learning 1995 20:3**, *[S. l.]*, v. 20, n. 3, p. 273–297, 1995. DOI: 10.1007/BF00994018. Disponível em: https://link.springer.com/article/10.1007/BF00994018. Acesso em: 14 jun. 2025.

CORWIN, D. L.; LESCH, S. M. Apparent soil electrical conductivity measurements in agriculture. **Computers and Electronics in Agriculture**, *[S. l.]*, v. 46, n. 1–3, p. 11–43, 2005. DOI: 10.1016/J.COMPAG.2004.10.005. Disponível em: https://www.sciencedirect.com/science/article/abs/pii/S0168169904001243?via%3Dihub. Acesso em: 27 jun. 2026.

DAUGHTRY, C. S. T.; WALTHALL, C. L.; KIM, M. S.; DE COLSTOUN, E. Brown; MCMURTREY, J. E. Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. **Remote Sensing of Environment**, *[S. l.]*, v. 74, n. 2, p. 229–239, 2000. DOI: 10.1016/S0034-4257(00)00113-9. Disponível em: https://www.researchgate.net/publication/222661408_Estimating_Corn_Leaf_Chlorophyll_Concentration_from_Leaf_and_Canopy_Reflectance. Acesso em: 10 mar. 2025.

DOBROWSKI, S. Z.; USTIN, S. L.; WOLPERT, J. A. Grapevine dormant pruning weight prediction using remotely sensed data. **Australian Journal of Grape and Wine Research**, *[S. l.]*, v. 9, n. 3, p. 177–182, 2003. DOI: 10.1111/J.1755-0238.2003.TB00267.X;PAGE:STRING:ARTICLE/CHAPTER. Disponível em: /doi/pdf/10.1111/j.1755-0238.2003.tb00267.x. Acesso em: 26 jun. 2026.

DRISSI, Rachid; GOUTOULY, Jean Pascal; FORGET, Dominique; GAUDILLERE, Jean Pierre. Nondestructive measurement of grapevine leaf area by ground normalized difference vegetation index. **Agronomy Journal**, *[S. l.]*, v. 101, n. 1, p. 226–231, 2009. DOI: 10.2134/AGRONJ2007.0167;SUBPAGE:STRING:ABSTRACT;REQUESTEDJOURNAL:JOURNAL:14350645;WEBSITE:WEBSITE:ACSESS.ONLINELIBRARY.WILE

Y.COM;WGROUP:STRING:PUBLICATION. Disponível em: /doi/pdf/10.2134/agronj2007.0167. Acesso em: 27 jun. 2026.

DRUSCH, M. et al. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. **Remote Sensing of Environment**, *[S. l.]*, v. 120, p. 25–36, 2012. DOI: 10.1016/J.RSE.2011.11.026. Disponível em: https://www.sciencedirect.com/science/article/abs/pii/S0034425712000636?via%3Dihub. Acesso em: 15 out. 2025.

ESCADAFAL RICHARD, Huete A. Etude des propriétés spectrales des sols arides appliquée à l'amélioration des indices de végétation obtenus par télédétection. **Comptes Rendus de l'Académie des Sciences**, *[S. l.]*, v. 312, n. Série 2, p. 1385–1391, 1991.

FAVERO, A. C.; ANGELUCCI DE AMORIM, D.; VIEIRA, R.; MOTA, D. A.; SOARES, A. M.; DE SOUZA, C. R.; DE ALBUQUERQUE REGINA, M.; REGINA, M. A. **Double-pruning of "Syrah" grapevines: a management strategy to harvest wine grapes during the winter in the Brazilian SoutheastVitis**. [s.l: s.n.].

FOUNTAS, Spyros; ANASTASIOU, Evaggelos; BALAFOUTIS, Athanasios; KOUNDOURAS, Stefanos; THEOHARIS, Serafeim; THEODOROU, Nikolaos. The influence of vine variety and vineyard management on the effectiveness of canopy sensors to predict winegrape yield and quality. **International Conference of Agricultural Engineering**, *[S. l.]*, 2014. Disponível em: www.eurageng.eu. Acesso em: 26 jun. 2026.

FRAGA, Helder; AMRAOUI, Malik; MALHEIRO, Aureliano C.; MOUTINHO-PEREIRA, José; EIRAS-DIAS, José; SILVESTRE, José; SANTOS, João A. Examining the relationship between the Enhanced Vegetation Index and grapevine phenology. **European Journal of Remote Sensing**, *[S. l.]*, v. 47, n. 1, p. 753–771, 2014. DOI: 10.5721/EUJRS20144743. Acesso em: 5 out. 2025.

GAVIOLI, Alan; DE SOUZA, Eduardo Godoy; BAZZI, Claudio Leones; SCHENATTO, Kelyn; BETZEK, Nelson Miguel. Identification of management zones in precision agriculture: An evaluation of alternative cluster analysis methods. **Biosystems Engineering**, *[S. l.]*, v. 181, p. 86–102, 2019. DOI: 10.1016/J.BIOSYSTEMSENG.2019.02.019. Disponível em: https://www.sciencedirect.com/science/article/abs/pii/S1537511018303295. Acesso em: 27 jun. 2026.

GIOVOS, Rigas; TASSOPOULOS, Dimitrios; KALIVAS, Dionissios; LOUGKOS, Nestor; PRIOVOLOU, Anastasia. Remote Sensing Vegetation Indices in Viticulture: A Critical Review. **Agriculture 2021, Vol. 11, Page 457**, *[S. l.]*, v. 11, n. 5, p. 457, 2021. DOI: 10.3390/AGRICULTURE11050457. Disponível em: https://www.mdpi.com/2077-0472/11/5/457/htm. Acesso em: 5 nov. 2024.

GITELSON, Anatoly A.; GRITZ, Yuri; MERZLYAK, Mark N. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. **Journal of Plant Physiology**, *[S. l.]*, v. 160, n. 3, p. 271–282, 2003. DOI: 10.1078/0176-1617-00887. Acesso em: 10 mar. 2025.

GOEL, N. S.; QIN, W. Influences of canopy architecture on relationships between various vegetation indices and LAI and Fpar: A computer simulation. **Remote Sensing Reviews**, *[S. l.]*, v. 10, n. 4, p. 309–347, 1994. DOI: 10.1080/02757259409532252. Disponível em: https://www.tandfonline.com/doi/abs/10.1080/02757259409532252. Acesso em: 10 mar. 2025.

GORELICK, Noel; HANCHER, Matt; DIXON, Mike; ILYUSHCHENKO, Simon; THAU, David; MOORE, Rebecca. Google Earth Engine: Planetary-scale geospatial analysis for everyone. **Remote Sensing of Environment**, *[S. l.]*, v. 202, p. 18–27, 2017. DOI: 10.1016/J.RSE.2017.06.031. Disponível em: https://www.sciencedirect.com/science/article/pii/S0034425717302900?via%3Dihub. Acesso em: 15 out. 2025.

GUYON, Isabelle; WESTON, Jason; BARNHILL, Stephen; VAPNIK, Vladimir. Gene selection for cancer classification using support vector machines. **Machine Learning**, *[S. l.]*, v. 46, n. 1–3, p. 389–422, 2002. DOI: 10.1023/A:1012487302797/METRICS. Disponível em: https://link.springer.com/article/10.1023/A:1012487302797. Acesso em: 13 nov. 2025.

HUETE, A.; DIDAN, K.; MIURA, T.; RODRIGUEZ, E. P.; GAO, X.; FERREIRA, L. G. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. **Remote Sensing of Environment**, *[S. l.]*, v. 83, n. 1–2, p. 195–213, 2002. DOI: 10.1016/S0034-4257(02)00096-2. Acesso em: 5 nov. 2024.

HUETE, A. R. A soil-adjusted vegetation index (SAVI). **Remote Sensing of Environment**, *[S. l.]*, v. 25, n. 3, p. 295–309, 1988. DOI: 10.1016/0034-4257(88)90106-X. Acesso em: 3 fev. 2025.

JOHNSON, L. F.; ROCZEN, D. E.; YOUKHANA, S. K.; NEMANI, R. R.; BOSCH, D. F. Mapping vineyard leaf area with multispectral satellite imagery. **Computers and Electronics in Agriculture**, *[S. l.]*, v. 38, n. 1, p. 33–44, 2003. DOI: 10.1016/S0168-1699(02)00106-0. Acesso em: 2 fev. 2025.

KASIMATI, Aikaterini; ESPEJO-GARCIA, Borja; VALI, Eleanna; MALOUNAS, Ioannis; FOUNTAS, Spyros. Investigating a Selection of Methods for the Prediction of Total Soluble Solids Among Wine Grape Quality Characteristics Using Normalized Difference Vegetation Index Data From Proximal and Remote Sensing. **Frontiers in Plant Science**, *[S. l.]*, v. 12, 2021. a. DOI: 10.3389/fpls.2021.683078.

KASIMATI, Aikaterini; ESPEJO-GARCIA, Borja; VALI, Eleanna; MALOUNAS, Ioannis; FOUNTAS, Spyros. Investigating a Selection of Methods for the Prediction of Total Soluble Solids Among Wine Grape Quality Characteristics Using Normalized Difference Vegetation Index Data From Proximal and Remote Sensing. **Frontiers in Plant Science**, *[S. l.]*, v. 12, 2021. b. DOI: 10.3389/fpls.2021.683078.

KATAOKA, T.; KANEKO, T.; OKAMOTO, H.; HATA, S. Crop growth estimation system using machine vision. **IEEE/ASME International Conference on Advanced Intelligent Mechatronics, AIM**, *[S. l.]*, v. 2, p. 1079–1083, 2003. DOI: 10.1109/AIM.2003.1225492. Acesso em: 3 fev. 2025.

KAUTH, R. J.; THOMAS, G. S. Purdue e-Pubs The Tasselled Cap-A Graphic Description of the Spectral-Temporal Development of Agricultural Crops as Seen by LANDSAT. *[S. l.]*, 1976. Disponível em: http://docs.lib.purdue.edu/lars_symphttp://docs.lib.purdue.edu/lars_symp/159. Acesso em: 24 out. 2025.

KODUR, S. Effects of juice pH and potassium on juice and wine quality, and regulation of potassium in grapevines through rootstocks (Vitis): a short review. **VITIS - Journal of Grapevine Research**, *[S. l.]*, v. 50, n. 1, p. 1–1, 2011. DOI: 10.5073/VITIS.2011.50.1-6. Disponível em: https://ojs.openagrar.de/index.php/VITIS/article/view/4052. Acesso em: 26 jun. 2026.

LAMB, David William. **Making sense of vineyard variability in Australia**. [s.l: s.n.]. Disponível em: https://www.researchgate.net/publication/275832528.

LOUHAICHI, Mounir; BORMAN, Michael M.; JOHNSON, Douglas E. Spatially located platform and aerial photography for documentation of grazing impacts on wheat. **Geocarto International**, *[S. l.]*, v. 16, n. 1, p. 65–70, 2001. a. DOI: 10.1080/10106040108542184. Disponível em: https://www.researchgate.net/publication/228906661_Spatially_Located_Platform_and_Aerial_Photography_for_Documentation_of_Grazing_Impacts_on_Wheat. Acesso em: 3 fev. 2025.

LOUHAICHI, Mounir; BORMAN, Michael M.; JOHNSON, Douglas E. Spatially located platform and aerial photography for documentation of grazing impacts on wheat. **Geocarto International**, *[S. l.]*, v. 16, n. 1, p. 65–70, 2001. b. DOI: 10.1080/10106040108542184. Disponível em: https://www.researchgate.net/publication/228906661_Spatially_Located_Platform_and_Aerial_Photography_for_Documentation_of_Grazing_Impacts_on_Wheat. Acesso em: 10 mar. 2025.

MATESE, Alessandro; DI GENNARO, Salvatore Filippo. **Technology in precision viticulture: A state of the art review**. **International Journal of Wine Research**Dove Medical Press Ltd, , 2015. DOI: 10.2147/IJWR.S69405.

MCBRATNEY, A. B.; WEBSTER, R. Choosing functions for semi-variograms of soil properties and fitting them to sampling estimates. **Journal of Soil Science**, *[S. l.]*, v. 37, n. 4, p. 617–639, 1986. DOI: 10.1111/J.1365-2389.1986.TB00392.X;SUBPAGE:STRING:ABSTRACT;REQUESTEDJOURNAL:JOURNAL:13652389A;WGROUP:STRING:PUBLICATION. Disponível em: /doi/pdf/10.1111/j.1365-2389.1986.tb00392.x. Acesso em: 22 nov. 2025.

MELO, Derlei D.; CUNHA, Isabella A.; AMARAL, Lucas R. Precision Zones: An Open-Source QGIS Plugin for Management-Zone Segmentation in Precision Agriculture. **AgriEngineering 2025, Vol. 7, Page 420**, *[S. l.]*, v. 7, n. 12, p. 420, 2025. DOI: 10.3390/AGRIENGINEERING7120420. Disponível em: https://www.mdpi.com/2624-7402/7/12/420/htm. Acesso em: 16 maio. 2026.

MEYER, G. E.; MEHTA, T.; KOCHER, M. F.; MORTENSEN, D. A.; SAMAL, A. TEXTURAL IMAGING AND DISCRIMINANT ANALYSIS FOR DISTINGUISHINGWEEDS FOR SPOT SPRAYING. **Transactions of the ASAE**, *[S. l.]*, v. 41, n. 4, p. 1189–1197, 1998. DOI: 10.13031/2013.17244. Disponível em: https://elibrary.asabe.org/azdez.asp?JID=3&AID=17244&CID=t1998&v=41&i=4&T=1. Acesso em: 3 fev. 2025.

MOLIN, José Paulo; FAULIN, Gustavo Di Chiacchio. Spatial and temporal variability of soil electrical conductivity related to soil moisture. **Scientia Agricola**, *[S. l.]*, v. 70, n. 1, p. 01–05, 2013. DOI: 10.1590/S0103-90162013000100001. Disponível em: https://www.scielo.br/j/sa/a/6pmd35WdvfqGDsKy8c6tq4R/?lang=en. Acesso em: 27 jun. 2026.

NETO, Joao Camargo. A combined statistical-soft computing approach for classification and mapping weed species in minimum -tillage systems. **ETD collection for University of Nebraska-Lincoln**, *[S. l.]*, 2004. Disponível em: https://digitalcommons.unl.edu/dissertations/AAI3147135. Acesso em: 3 fev. 2025.

OIV. **State of the World Wine Sector in 2025**. [s.l.] : International Organisation of Vine and Wine, 2026. Disponível em: www.oiv.int.

PARKER, A. K.; DE CORTÁZAR-ATAURI, I. G.; VAN LEEUWEN, C.; CHUINE, I. General phenological model to characterise the timing of flowering and veraison of Vitis vinifera L. **Australian Journal of Grape and Wine Research**, *[S. l.]*, v. 17, n. 2, p. 206–216, 2011. DOI: 10.1111/J.1755-0238.2011.00140.X;WGROUP:STRING:PUBLICATION. Disponível em: /doi/pdf/10.1111/j.1755-0238.2011.00140.x. Acesso em: 26 jun. 2026.

PENSO, Gener Augusto; CITADIN, Idemir; MARO, Luana Aparecida Castilho; SCARIOTTO, Silvia; CHICOKI, Alyne; AMBROSIO, Rodrigo. Evolução da maturação de uvas finas (Vitis vinifera L.), no município de Dois Vizinhos - Paraná. **Semina: Ciências Agrárias**, *[S. l.]*, v. 35, n. 6, p. 3085–3098, 2014. DOI: 10.5433/1679-0359.2014V35N6P3085. Disponível em: https://ojs.uel.br/revistas/uel/index.php/semagrarias/article/view/16269. Acesso em: 4 jul. 2026.

PIX4D SA. **Pix4Dmapper**. 2025. Disponível em: https://www.pix4d.com/. Acesso em: 27 jul. 2025.

PRIORI, S.; MARTINI, E.; ANDRENELLI, M. C.; MAGINI, S.; AGNELLI, A. E.; BUCELLI, P.; BIAGI, M.; PELLEGRINI, S.; COSTANTINI, E. A. C. Improving Wine Quality through Harvest Zoning and Combined Use of Remote and Soil Proximal Sensing. **Soil Science Society of America Journal**, *[S. l.]*, v. 77, n. 4, p. 1338–1348, 2013. DOI: 10.2136/sssaj2012.0376.

REGINA, M. de A.; FRÁGUAS, J. C.; ALVARENGA, A. A.; SOUZA, C. R. De; AMORIM, D. A. De; MOTA, R. V. Da; FÁVERO, A. C. Implantação e manejo do vinhedo para produção de vinhos de qualidade. *[S. l.]*, 2019. Disponível em: http://www.alice.cnptia.embrapa.br/handle/doc/541532. Acesso em: 4 jul. 2026.

REYES, Francesco; CASA, Raffaele; TOLOMIO, Massimo; DALPONTE, Michele; MZID, Nada. Soil properties zoning of agricultural fields based on a climate-driven spatial clustering of remote sensing time series data. **European Journal of Agronomy**, *[S. l.]*, v. 150, p. 126930, 2023. DOI: 10.1016/J.EJA.2023.126930. Disponível em: https://www.sciencedirect.com/science/article/pii/S1161030123001983?via%3Dihub. Acesso em: 27 jun. 2026.

RIENTH, Markus; TORREGROSA, Laurent; SARAH, Gautier; ARDISSON, Morgane; BRILLOUET, Jean Marc; ROMIEU, Charles. Temperature desynchronizes sugar and organic acid metabolism in ripening grapevine fruits and remodels their transcriptome. **BMC Plant Biology 2016 16:1**, *[S. l.]*, v. 16, n. 1, p. 164-, 2016. DOI: 10.1186/S12870-016-0850-0. Disponível em: https://link.springer.com/article/10.1186/s12870-016-0850-0. Acesso em: 26 jun. 2026.

ROSENQVIST, Ake; SHIMADA, Masanobu; ITO, Norimasa; WATANABE, Manabu. ALOS PALSAR: A pathfinder mission for global-scale monitoring of the environment. **IEEE Transactions on Geoscience and Remote Sensing**, *[S. l.]*, v. 45, n. 11, p. 3307–3316, 2007. DOI: 10.1109/TGRS.2007.901027. Disponível em: https://www.researchgate.net/publication/3205578_ALOS_PALSAR_A_Pathfinder_mission_for_global-scale_monitoring_of_the_environment. Acesso em: 20 out. 2025.

ROUSE, J. W. ,. Jr.; HAAS, R. H.; SCHELL, J. A.; DEERING, D. W. Monitoring vegetation systems in the Great Plains with ERTS. **NASA. Goddard Space Flight Center 3d ERTS-1 Symp., Vol. 1, Sect. A**, *[S. l.]*, 1974. . Acesso em: 2 fev. 2025.

SANTOS, H. G. Dos et al. Sistema Brasileiro de Classificação de Solos. *[S. l.]*, 2025. Disponível em: http://www.infoteca.cnptia.embrapa.br/handle/doc/1176834. Acesso em: 29 set. 2025.

SHINE, Philip; MURPHY, Michael D. Over 20 Years of Machine Learning Applications on Dairy Farms: A Comprehensive Mapping Study. **Sensors (Basel, Switzerland)**, *[S. l.]*, v. 22, n. 1, p. 52, 2021. DOI: 10.3390/S22010052. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC8747441/. Acesso em: 14 jun. 2025.

SMART, R. E.; DICK, Joy K.; GRAVETT, Isabella M.; FISHER, B. M. Canopy Management to Improve Grape Yield and Wine Quality - Principles and Practices. **South African Journal of Enology and Viticulture**, *[S. l.]*, v. 11, n. 1, p. 3–17, 1990. DOI: 10.21548/11-1-2232. Disponível em: https://www.journals.ac.za/sajev/article/view/2232. Acesso em: 26 jun. 2026.

SWEETMAN, C.; SADRAS, V. O.; HANCOCK, R. D.; SOOLE, K. L.; FORD, C. M. Metabolic effects of elevated temperature on organic acid degradation in ripening Vitis vinifera fruit. **Journal of Experimental Botany**, *[S. l.]*, v. 65, n. 20, p. 5975–5988, 2014. DOI: 10.1093/JXB/ERU343. Disponível em: https://dx.doi.org/10.1093/jxb/eru343. Acesso em: 26 jun. 2026.

TUCKER, Compton J. Red and photographic infrared linear combinations for monitoring vegetation. **Remote Sensing of Environment**, *[S. l.]*, v. 8, n. 2, p. 127–150, 1979. a. DOI: 10.1016/0034-4257(79)90013-0. Acesso em: 3 fev. 2025.

TUCKER, Compton J. Red and photographic infrared linear combinations for monitoring vegetation. **Remote Sensing of Environment**, *[S. l.]*, v. 8, n. 2, p. 127–150, 1979. b. DOI: 10.1016/0034-4257(79)90013-0. Acesso em: 10 mar. 2025.

WOEBBECKE, D. M.; MEYER, G. E.; VON BARGEN, K.; MORTENSEN, D. A. Color Indices for Weed Identification Under Various Soil, Residue, and Lighting Conditions. **Transactions of the ASAE**, *[S. l.]*, v. 38, n. 1, p. 259–269, 1995. DOI: 10.13031/2013.27838. Disponível em: https://elibrary.asabe.org/azdez.asp?JID=3&AID=27838&CID=t1995&v=38&i=1&T=1. Acesso em: 3 fev. 2025.

YUAN, Yifan; SHI, Bo; YOST, Russell; LIU, Xiaojun; TIAN, Yongchao; ZHU, Yan; CAO, Weixing; CAO, Qiang. Optimization of Management Zone Delineation for Precision Crop Management in an Intensive Farming System. **Plants**, *[S. l.]*, v. 11, n. 19, p. 2611, 2022. DOI: 10.3390/PLANTS11192611. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC9573654/. Acesso em: 22 nov. 2025.

ZARCO-TEJADA, P. J.; BERJÓN, A.; LÓPEZ-LOZANO, R.; MILLER, J. R.; MARTÍN, P.; CACHORRO, V.; GONZÁLEZ, M. R.; DE FRUTOS, A. Assessing vineyard condition with hyperspectral indices: Leaf and canopy reflectance simulation in a row-structured discontinuous canopy. **Remote Sensing of Environment**, *[S. l.]*, v. 99, n. 3, p. 271–287, 2005. DOI: 10.1016/J.RSE.2005.09.002. Disponível em: https://www.sciencedirect.com/science/article/abs/pii/S0034425705002865. Acesso em: 27 jun. 2026.

SEGUNDA PARTE

Microterroir Zones and Predictive Modeling of Grape Technological Ripening in Subtropical area in Brazil

B. BARBOSA[1], R. PIO[2], A. SANTOS[3], M. MENEZES[4], B. SILVA[5], O. COSTA[6], S. RODRIGUES[7], F. FILHO[8], C. RIBEIRO[9], P. PECHE[10]

1: *Federal University of Lavras (UFLA), master's student in Crop Science, Department of Agriculture (DAG), P.O. Box 3037, 37200-000, Lavras, Minas Gerais, Brazil; bernardosiqueira.agro@gmail.com*

2: *Federal University of Lavras (UFLA), Department of Agriculture (DAG), P.O. Box 3037, 37200-000, Lavras, Minas Gerais, Brazil; rafaelpio@ufla.br*

3: Federal University of Lavras (UFLA), Department of Agriculture (DAG), P.O. Box 3037, 37200-000, Lavras, Minas Gerais, Brazil; adao.felipe@ufla.br

4: Federal University of Lavras (UFLA), Department of Soil Science (DCS), *P.O. Box 3037, 37200-000, Lavras, Minas Gerais, Brazil*: *michele.menezes@ufla.br*

5: Federal University of Lavras (UFLA), Department of Soil Science (DCS), *P.O. Box 3037, 37200-000, Lavras, Minas Gerais, Brazil*: *brunom.silva@ufla.br*

6: Federal University of Lavras (UFLA), PhD student in Crop Science, Department of Agriculture (DAG), P.O. Box 3037, 37200-000, Lavras, Minas Gerais, Brazil; octavio.costa2@estudante.ufla.br.

7: Federal University of Lavras (UFLA), *PhD student in Soil Science,* Department of Soil Science (DCS), *P.O. Box 3037, 37200-000, Lavras, Minas Gerais, Brazil*: *sarahdieckman.a.r@gmail.com*

8: *Federal University of Lavras (UFLA), Department of Agriculture (DAG), P.O. Box 3037, 37200-000, Lavras, Minas Gerais, Brazil fabiano.filho1@estudante.ufla.br*

9: *Federal University of Lavras (UFLA), Department of Agriculture (DAG), P.O. Box 3037, 37200-000, Lavras, Minas Gerais, Brazil caio.ribeiro2@estudante.ufla.br*

10: Federal University of Lavras (UFLA), Department of Agriculture (DAG), P.O. Box 3037, 37200-000, Lavras, Minas Gerais, Brazil pedro.peche@ufla.br.

ABSTRACT

Traditional methods for determining wine grape quality rely on sample-based laboratory analyses, which are time-consuming, expensive, and destructive. To overcome these limitations, this study proposes a non-destructive approach based on precision viticulture and machine learning to predict grape maturation and delineate microterroir zones. We analyzed a comprehensive spatiotemporal dataset (n = 1116) across the 2024 and 2025 seasons, integrating vegetation indices (VIs), accumulated growing degree-days (GDD), and key physicochemical attributes, including soluble solids content (SSC), pH, and tartaric acidity (TA). Initial feature selection established GDD as the primary driver for all quality attributes, showing strong correlations with both SSC and TA. Building upon these temporal dynamics, we evaluated several machine learning models for inter-seasonal validation. The Random Forest (RF) algorithm emerged as the most robust model for macro-prediction, yielding high accuracy for SSC ($R^2 = 0.73$) and TA ($R^2 = 0.71$). Conversely, pH modeling proved highly challenging, exhibiting critical temporal degradation and weak generalizability across independent seasons. In parallel to the temporal predictions, unsupervised K-Means clustering (k = 3) was applied to define spatial zones by integrating multi-year canopy vigor, terrain attributes, and soil proxies. This approach effectively partitioned macro-physiological traits, such as chlorophyll *b*, and maximized variance reduction when combining annual insolation with soil properties. However, field validation highlighted a limitation, as the model captures broad spatial zones but struggles to stratify highly localized, sub-metric canopy variations. Ultimately, integrating climatic metrics with remote sensing provides a reliable framework for forecasting macro-temporal trends in SSC and TA, as well as for delineating robust microterroir zones. Nevertheless, improving inter-seasonal pH forecasting will require the future inclusion of specific soil chemical properties.

Keywords: Cabernet Sauvignon, Cabernet Franc, Precision Viticulture, Spectral variability

INTRODUCTION

The Brazilian viticulture, particularly in the Southeastern region (subtropical area) has been the expansion of fine wine production driven by the implementation of the double pruning technique (BRANT et al., 2020). This viticultural innovation aims to shift grape harvesting from the traditionally rainy summer season to the drier winter months, thereby improving fruit quality under tropical climatic conditions (FAVERO et al., 2011).

Among the cultivars benefiting from this approach, *Cabernet Sauvignon* and *Cabernet Franc* have demonstrated satisfactory agronomic performance and promising enological potential under the specific environmental conditions of southern Minas Gerais State. (REGINA et al., 2011)

Achieving uniform ripening, however, is made difficult by the spatial variability of the vineyard. Viticultural productivity is spatially inconsistent, both in terms of vegetative vigor (BRAMLEY; HAMILTON, 2007) (JOHNSON et al., 2003) and fruit quality (BRAMLEY; HAMILTON, 2005). This heterogeneity largely derives from variations in the edaphic and topographic attributes attributes of the site (BRAMLEY; OUZMAN; BOSS, 2011).

In the context of modern viticulture, this spatial heterogeneity presents a critical economic and operational challenge (MATESE; DI GENNARO, 2015). The rising costs of specialized agricultural labor demand more efficient management strategies. By identifying and delineating specific management zones, vineyard managers can adopt precision practices, such as selective step-wise harvesting. This approach not only mitigates the financial impact of labor inefficiencies but also ensures that grapes are harvested at their optimal technological maturity, maximizing the enological potential of each specific area (BRAMLEY; OUZMAN; THORNTON, 2011).

To address these challenges, predictive modelling is dedicated to the temporal prediction of key enological attributes. This phase seeks to develop and validate robust predictive models capable of estimating grape maturation attributes, namely Soluble Solids Concentration, pH, and Tartaric Acidity (KASIMATI et al., 2021a). The prediction is based on the integration of remote sensing data (vegetation indices) and thermal variables (Growing Degree Days) (PARKER et al., 2011). This predictive capability can be successfully validated using an independent, subsequent growing season, demonstrating the temporal generalization of the models.

While predictive modeling addresses the temporal forecasting of enological attributes, it must be coupled with the delineation of microterroir zones to effectively manage the spatial variability within the vineyard block. This spatial approach focuses on characterizing structural heterogeneity by integrating environmental factors and long-term vegetative responses (ACEVEDO-OPAZO et al., 2008), (PRIORI et al., 2013).

Despite the advances in precision agriculture, the application of data-driven tools remains scarce in the context of winter wine production. Filling this gap requires the development of a pioneering computational prototype, designed to translate complex agrometeorological and remote sensing data into practical decision-making tools for the viticultural sector.

Therefore, the objective of this study was to develop temporal predictive models for grape technological maturation and to delineate spatial microterroir zones under subtropical conditions in Brazil.

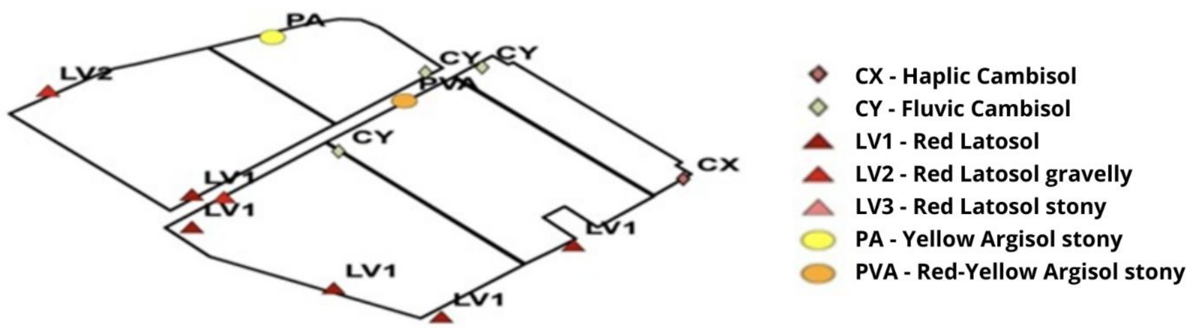
MATERIAL E METHODS

The experimental site was a commercial vineyard, located in Bom Sucesso, south of Minas Gerais State, Brazil (21°09', 44°53'; altitude: 865 m above sea level). The regional climate is classified as Cwa (humid subtropical with dry winters) according to Köppen, with mean annual temperature of 19.4 °C, annual rainfall of 1,529.7 mm, and relative humidity of 76.2%, with dry winters and rainy seasons between October and March, with greater intensity rain events between December and February (Alvares et al., 2013).

Vitis vinifera (L.) cultivar Cabernet Sauvignon (2.4 ha) and Cabernet Franc (1.4 ha) grapevines were planted in 2022, on rootstock IAC 766 'Campinas' and trained to a vertically shoot-positioned (VSP) trellis system, under a double pruning management regime, and drip irrigated. Rows were orientated northwest-southeast and row and vine spacing was 2.6 and 1.0 m, respectively. Production pruning was performed on 3 March 2024 and 3 March 2025 for Cabernet Sauvignon and on 17 February 2024 and 17 February 2025 for Cabernet Franc.

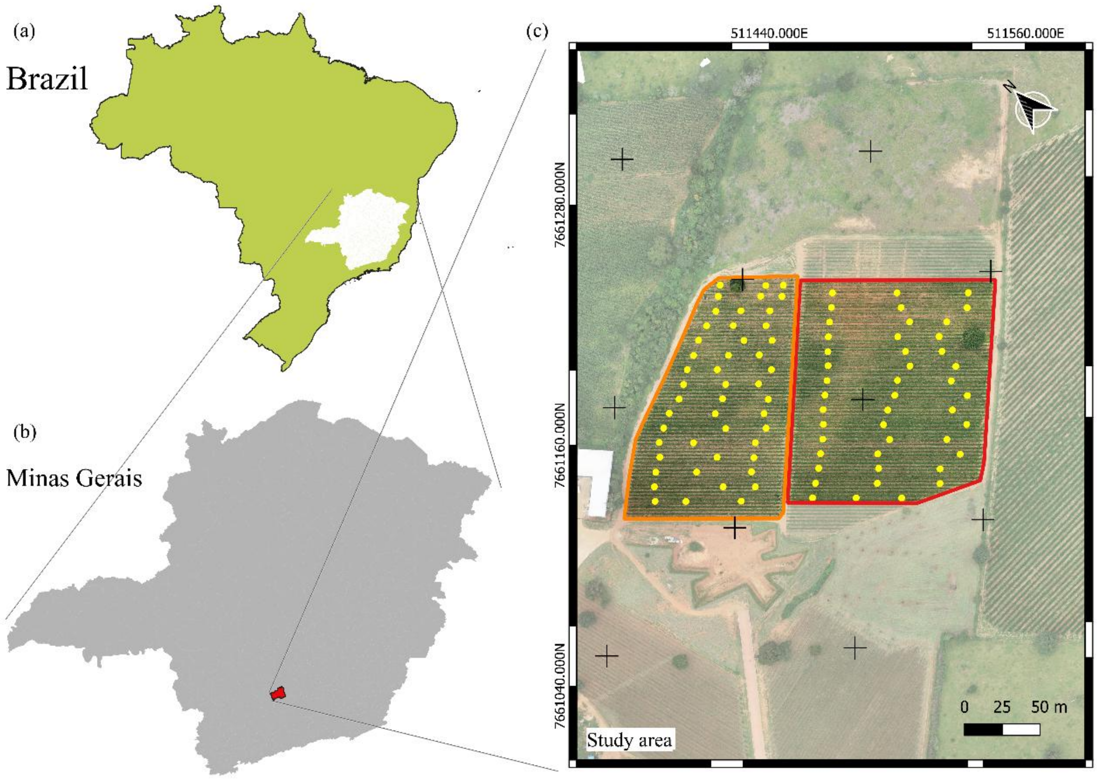
The vineyard area presents a heterogeneous distribution of soil classes, described according to the Brazilian Soil Classification System – SiBCS (SANTOS et al., 2025). The predominant soils are Red Latosols (LV1, LV2, and LV3), ranging from typical to stony and gravelly phases. In smaller proportions, Haplic Cambisols (CX) and Fluvic Cambisols (CY) occur. Additionally, Yellow Argisols stony (PA) and Red-Yellow Argisols stony (PVA) were identified (Figure1).

Figure 1 - Distribution of soil classes of the vineyard experimente.



A total of 93 georeferenced sampling points were established (45 for Cabernet Sauvignon and 48 for Cabernet Franc, enclosed within red and orange polygons, respectively). These points were used both for collecting grape ripening data in the field and for extracting spectral information from multispectral imagery (Figure 2).

Figure 2 - Location of the study area in the municipality of Bom Sucesso, south of Minas Gerais State, Brazil. (a) Map of Brazil highlighting the state of Minas Gerais;(b) Map of Minas Gerais highlighting the municipality of Bom Sucesso;(c) High-resolution satellite image of the vineyard experimente.



The data collection was carried out during the grape ripening phase, specifically after the onset of veraison. The sampling spanned six consecutive dates in and six consecutive dates in 2025, at weekly intervals up to harvest. At each georeferenced sampling point and on each date, 20 grape berries were manually harvested from three distinct vertical positions within the grape cluster (top, middle, and bottom), evenly distributed across both sides of the vertically shoot-positioned (VSP) trellis system and placed in labeled plastic bags immediately.

Following collection, the berries were homogenized and gently crushed to prevent seed damage and minimize the release of undesirable phenolic compounds. The resulting juice was subjected to laboratory analysis to quantify three key physicochemical parameters of grape maturity: soluble solids content, measured in degrees Brix (°Brix); pH, using a calibrated digital pH meter; and tartaric acidity (TA), results were subsequently converted into tartaric acid equivalents (g/L).

In parallel with field sampling, a series of high-resolution multispectral satellite images were acquired from Planet Labs Inc., each with a spatial resolution of 3 meters and containing surface reflectance data across eight spectral bands (coastal blue, blue, green I, green, yellow, red, red edge, and near-infrared), exported in GeoTIFF format. These images were selected to coincide precisely with the dates of field data collection to ensure temporal correspondence. The satellite imagery projected on WGS 84 / UTM zone 23S was subsequently processed in QGIS 3.4 (Free Software

Foundation, Boston, MA, USA) using the Raster Calculator tool to compute 16 vegetation indices for each sampling date and georeferenced location.

In addition to remote sensing and field measurements, a climatic variable accumulated degree-days (GDD) were computed following the method proposed by Villa Nova et al. (1972), using a base temperature of 10°C for grapevine development, as suggested by (BACK; BRUNA, 2012).

If Tm > Tb $$GD = \frac{TM+Tm}{2} - 10$$

If Tm < Tb < TM $$GD = \frac{(TM-10)^2}{2*(TM-Tm)}$$

If TM < Tb $$GD = 0$$

TM = daily maximum temperature; Tm = daily minimum temperature; Tb = base temperature (10°C)

Daily temperature data were obtained from the NASA POWER platform, aligned temporally with the six field sampling dates. The reliability of NASA POWER data for environmental and agrometeorological applications has been validated in previous studies (BARBOZA et al., 2024).

In this study, 16 vegetation indices were used, as detailed in Table 1. (Table 1).

Table 1 - Vegetation Indices (VIs) selected as predictors for the grape maturation attributes models, their mathematical equations, and original references.

V.I.	**Name**	**Formula**	**Ref.**
NDVI	Normalized Difference Vegetation Index	$NDVI = \frac{(NIR-R)}{(NIR+R)}$	(ROUSE et al., 1974)
NGRDI	Normalized Green-Red Difference Index	$NGRDI = \frac{(G-R)}{(G+R)}$	(TUCKER, 1979a)
SAVI	Soil-Adjusted Vegetation Index	$SAVI = \frac{(NIR-R)}{(NIR+R+L)} * (1+L)$	(HUETE, 1988)
RI	Redness Index	$RI = \frac{(R-G)}{(R+G)}$	(ESCADAFAL RICHARD, 1991)
EXG	Excess Green Index	$EXG = ((2*g) - r - b)$	(WOEBBECKE et al., 1995)
EXR	Excess Red Index	$EXR = (1.4*R) - G$	(MEYER et al., 1998)
EXGR	Excess Green minus Red Index	$EXGR = EXG - EXR$	(NETO, 2004)
GLI	Green Leaf Index	$GLI = \frac{((G-R)+(G-B))}{(G+R+G+B)}$	(LOUHAICHI; BORMAN; JOHNSON, 2001a)
CIVE	Color Index of Vegetation Extraction	$CIVE = ((0.441*R) - (0{,}811*G) + (0{,}385*B) + 18{,}78745)$	(KATAOKA et al., 2003)

PCD	Plant Cell Density Index	$PCD = (NIR/R)$	(LAMB, 2003)
PVR	Pigment Vegetation Ratio	$PVR = (G/R)$	(LAMB, 2003)
NDRE	Normalized Difference Red Edge Index	$\frac{NIR - RedEdge}{NIR + RedEdge}$	(TUCKER, 1979b)
GNDVI	Green Normalized Difference Vegetation Index	$\frac{NIR - G}{NIR + G}$	(LOUHAICHI; BORMAN; JOHNSON, 2001b)
GCI	Green Chlorophyll Index	$\frac{NIR}{G} - 1$	(GITELSON; GRITZ; MERZLYAK, 2003)
MCARI	Modified Chlorophyll Absorption in Reflectance Index	$((NIR - R) - 0.2 * (NIR - G)) * \frac{NIR}{R}$	(DAUGHTRY et al., 2000)
NLI	Non-Linear Index	$\frac{(NIR^2 - R)}{(NIR^2 + R)}$	(GOEL; QIN, 1994)

Table 1.Legend: L = 0,5

$\boldsymbol{g = \frac{G}{(R+G+B)}}$ $\boldsymbol{r = \frac{R}{(R+G+B)}}$ $\boldsymbol{b = \frac{B}{(R+G+B)}}$

To extract vegetation index values at the point scale, all satellite scenes were further processed in QGIS 3.4 (Free Software Foundation, Boston, MA, USA). To extract vegetation index values, a circular buffer with a 1.5-meter radius was applied to each georeferenced sampling point. Mean vegetation index values were extracted from each buffer area, ensuring spatial alignment between ground-based observations and remote sensing data. This procedure resulted in a temporally and spatially explicit dataset suitable for subsequent statistical analyses.

Vegetation indices, accumulated growing degree-days (GDD), and physicochemical measurements of grape maturity (soluble solids content (SSC), pH, and tartaric acidity) were integrated to construct the dataset. For each georeferenced sampling point, all predictor variables (vegetation indices and GDD) were temporally aligned and linked to the corresponding response variables (grape composition attributes). The resulting data matrix comprised 1116 observations (93 sampling points × 12 dates), each representing a unique spatiotemporal condition of the vineyard canopy and fruit development.

A preliminary descriptive statistical analysis was conducted to explore the potential of vegetative indexes and climatic variables in predicting grape quality parameters. In this exploratory phase, spearman correlation coefficients were computed to assess the linear associations between all predictor variables (vegetation indexes and accumulated degree-days) and the target attributes (soluble solids content, pH, tartaric acidity).

The initial phase of the machine learning (ML) modeling process involved the selection of the most relevant predictor variables. For this purpose, the Recursive Feature Elimination (RFE) method was applied, which iteratively assesses attribute importance (GUYON et al., 2002).

RFE was executed independently for each of the three target maturation variables: SSC (Soluble Solids Content), pH, and TA (Tartaric Acidity). The primary objective was to identify optimized predictor subsets tailored individually for each target attribute. This analysis utilized the complete dataset, encompassing both studied cultivars across the entire vineyard area (N=93 observation points). Following the RFE selection criteria, the top five predictor variables were retained for subsequent modeling and analysis.

To assess the model's ability to generalize to future seasons (temporal validation), this study adopted a sequential data splitting strategy. The dataset corresponding to the 2024 season was used for model fitting (training). The dataset from the 2025 season was held out entirely as an independent dataset for final testing and validation.

Consistent with methodological best practices (SHINE; MURPHY, 2021) hyperparameter tuning was conducted exclusively within the 2024 training set, employing an inner 10-fold CV process to select the optimal model configuration before the final validation on the 2025 data.

Predictive models were developed using three supervised machine learning techniques:

Random Forest: A supervised learning approach in which the ensemble learning method is used for regression. This combines numerous decision tree regressors into a single model trained on many data samples collected on the input feature using the bootstrap sampling method (BREIMAN, 2001).

Support Vector Regression: It is one of the most robust prediction methods. The (non-linear) model produced by this algorithm depends only on a subset of the training data because the cost function does not take into account any training data close to the model predictions (CORTES; VAPNIK; SAITTA, 1995).

Extreme Gradient Boosting (XGBoost): It is a high-performance, scalable machine learning algorithm based on decision tree ensembles. It incorporates a sparsity-aware learning strategy and an efficient weighted quantile sketch algorithm for handling structured and sparse data (CHEN; GUESTRIN, 2016).

The predictive performance of the models was evaluated using the coefficient of determination (R^2) (1), root mean square error (RMSE) (2), and mean absolute error (MAE) (3), calculated based on an independent validation subset. In addition, a 10-fold cross-validation was performed within the training data for each regression model to assess its generalization ability and ensure robustness (CHAI; DRAXLER, 2014)

Soil Electrical Conductivity (EC) is defined as the inverse of Electrical Resistivity (ER) (Rabello et al., 2014). Consequently, soil EC is obtained from the potential difference measured between electrodes as a result of the current emitted and received by the system. The operating principle is based on the relationship expressed in Equation (1):

$$CE\mathrm{a} = \frac{1}{\rho\alpha}$$

where CEa is the soil apparent electrical conductivity (S m^{-1}) and ρa is the soil apparent electrical resistivity (Ω m).

Data acquisition is performed in real time using the Fieldbox system, developed by the same company. Measurements were conducted with the device pulled by a tractor at an approximate speed of 5 km h^{-1}, moving along the vineyard interrows. Due to the spacing between discs, the effective reading depth corresponds to 0.30 m.

Apparent electrical conductivity (CEa) was calculated as the inverse of electrical resistivity (ρa), determined according to the four-electrode method described by Wenner (1915), as shown in Equation (2):

$$\rho\alpha = 2\pi\alpha\frac{V}{I}$$

The Soil Brightness Index (SBI) map was generated using a high-resolution (3m spatial resolution) Planet Labs satellite image acquired on March 22, 2022. This specific date was selected to ensure bare soil conditions within the vineyard block, minimizing interference from vegetation spectral signature. The SBI was calculated within the QGIS environment using the Raster Calculator tool and the following established spectral index formula:

$$SBI = \sqrt{\frac{(Blue^2 + Green^2 + Red^2 + NIR^2)}{4}} \text{(KAUTH; THOMAS, 1976)}$$

The Google Earth Engine (GEE) platform (GORELICK et al., 2017) was used to characterize the multiyear vegetative potential associated with each sampling point. A cloud-free median composite image was generated on the platform from the Sentinel-2 MSI surface reflectance dataset (Level-2A) (DRUSCH et al., 2012) covering the period from January 2023 to January 2025. Applying a median reducer to the time series enabled the synthesis of a single representative mosaic, reducing the influence of seasonal variability and atmospheric artifacts. From this composite image, the Enhanced Vegetation Index (EVI) was computed as an indicator of the spatial pattern of vegetation vigor.

The use of the Enhanced Vegetation Index (EVI) is recommended for vineyard monitoring because it captures vegetative growth and phenological stages of grapevines. Previous studies have demonstrated that EVI-derived metrics correlate significantly with critical stages such as flowering and veraison, and that multiyear EVI time series allow assessment of vegetative variability across years, supporting large-scale vineyard monitoring (FRAGA et al., 2014) .

$$EVI = 2.5 * \frac{(NIR - RED)}{(NIR + 6 * RED - 7.5 * BLUE + 1)} \qquad (HUETE, A. e\ colab., 2002)$$

Terrain attributes were extracted from the Digital Elevation Model (DEM) provided by the Alaska Satellite Facility, generated by the ALOS PALSAR sensor (ROSENQVIST et al., 2007) with a corrected spatial resolution of 12.5 meters (ALASKA SATELLITE FACILITY, 2015). Terrain insolation values were derived for the geographic coordinates of the polygon using the SAGA GIS platform (CONRAD et al., 2015), aiming to identify terrain-related differences.

To characterize the physiological status of the plants, two complementary attributes were measured at each sampling point. The relative leaf chlorophyll index, used as an indicator of nutritional status, was estimated with a portable meter (ClorofiLOG®, Falker) on three distinct and fully expanded leaves. NDVI was used as an indicator of vegetative vigor and was obtained with an active optical sensor (GreenSeeker®), positioned approximately 1 meter from the canopy side, with three independent readings per point. Physiological assessments (chlorophyll and NDVI) were conducted on April 17 and May 22, 2025. For all attributes, the final value assigned to each sampling point corresponded to the arithmetic mean of the respective triplicate measurements.

The delineation of microterroir zones was conducted in QGIS 3.4 (Free Software Foundation, Boston, MA, USA) using the Precision Zones plugin (MELO; CUNHA; AMARAL, 2025). Four distinct models were structured and compared based on different combinations of input variables. The biannual Enhanced Vegetation Index (EVI) time series was used as the baseline variable in all models to represent the spatial variability of vegetative vigor. Model differentiation followed a 2×2 factorial design: two Topography hypotheses (Elevation vs. Insolation) crossed with two Surface hypotheses (Electrical Conductivity vs. Soil Brightness). The biannual EVI was kept constant as a base vigor variable.

Models:

• Model 1 (Elevation + EC): Elevation, biannual EVI time series, and soil electrical conductivity.
• Model 2 (Insolation + EC): Annual terrain insolation, biannual EVI time series, and soil electrical conductivity.

•Model 3 (Elevation + SBI): Elevation, biannual EVI time series, and soil brightness.
• Model 4 (Insolation + SBI): Annual terrain insolation, biannual EVI time series, and soil brightness.

Principal Component Analysis (PCA) was employed in each approach to normalize variable contributions, eliminate multicollinearity, and weight the inputs according to their contribution to the total spatial variance of the study area (AGGOGERI; PELLEGRINI; TAGLIANI, 2021).

The selection of the number of components retained for clustering followed a quantitative criterion: the first two principal components (PCs) were selected, which consistently explained at least 70% of the total data variability across analyses. This approach preserves the most significant spatial patterns while minimizing noise associated with lower-order components.

The clustering stage, responsible for generating the zones, was performed using the unsupervised K-Means algorithm. Determination of the optimal number of clusters (k) was supported by a combined evaluation of the Elbow method and the Silhouette score. Based on these indicators, and considering agronomic interpretability and operational applicability, three management zones (k = 3) were adopted as the standard for all models, enabling direct comparison among the zoning approaches.

Subsequently, a post-processing step was applied to enhance spatial cohesion of the zone maps. A modal filter (majority filter) was used to smooth boundaries and remove salt-and-pepper noise. For this procedure, an analysis window with a radius of 5 was defined, corresponding to an 11×11-pixel neighborhood, in which the central pixel value is replaced by the mode (most frequent value) of its surroundings. The final output of this workflow consisted of microterroir maps with more homogeneous and operationally feasible zones.

To objectively compare the efficacy of the four models, a statistical validation was performed using the "Analysis" tool within the Precision Zones plugin.

This process involved overlaying the 93 ground-truth points onto the final zone maps of each of the four models. The plugin then calculated the Variance Reduction (VR%) for each of the six validation variables (Chl A, Chl B, and Greenseeker NDVI at two dates) (YUAN et al., 2022).

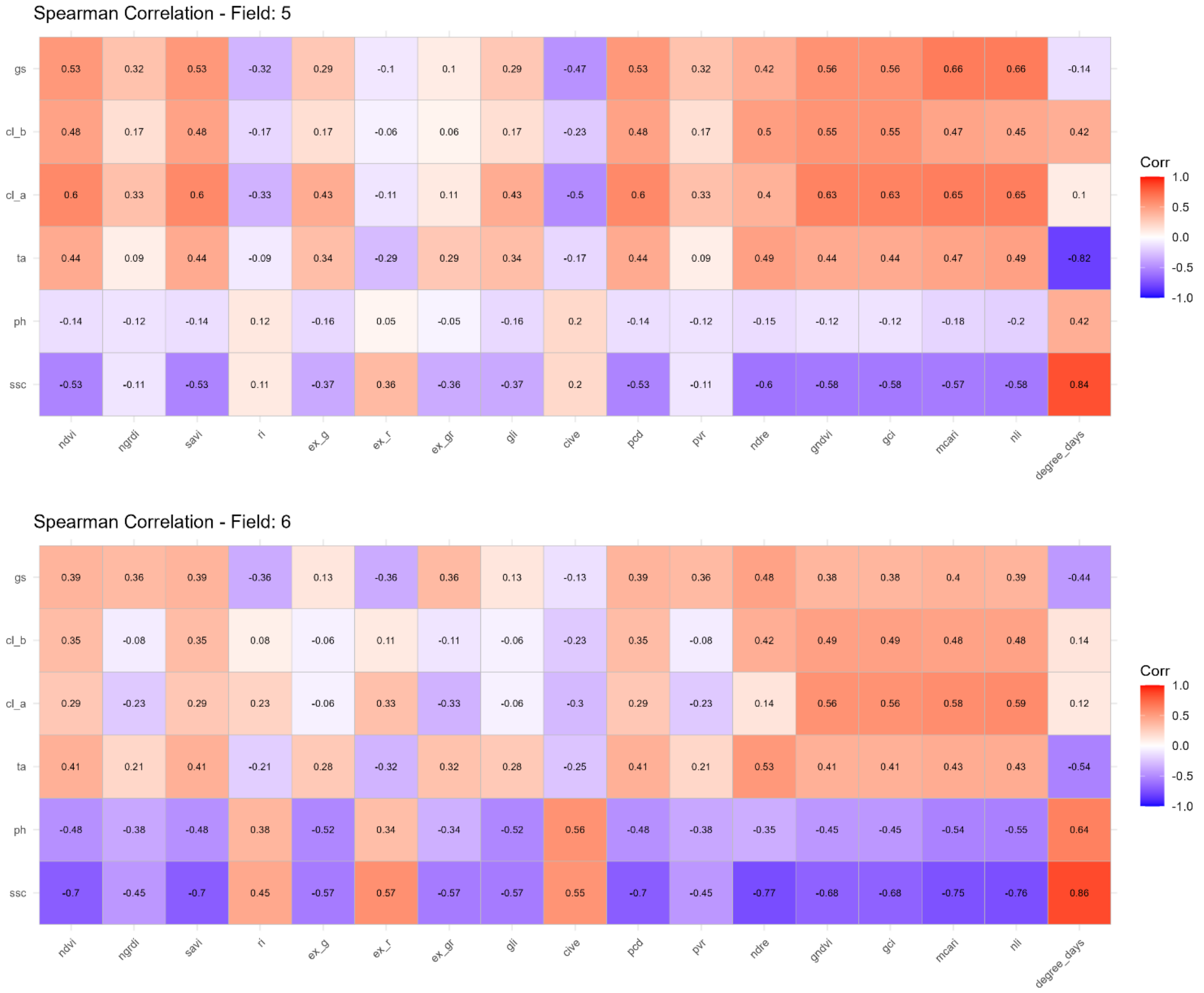
The VR% is a key metric of clustering efficacy, defined as:

$$VR\% = \left(\frac{SStotal - SSwithin}{SStotal}\right) * 100, MCBRATNEY\ e\ WEBSTER, 1986.$$

RESULTS

The Spearman correlation coefficients computed for Field 5 (Cabernet Sauvignon) and Field 6 (Cabernet Franc) (Figure 3) revealed strong and consistent relationships between predictor variables and the primary grape quality parameters.

Figure 3 - Spearman correlation matrices (heatmaps) illustrating the relationships between multispectral vegetation indices, thermal variables (growing degree days), and grape maturation attributes. (A) Field 5, corresponding to the Cabernet Sauvignon cultivar. (B) Field 6, corresponding to the Cabernet Franc cultivar. The color gradient denotes the strength and direction of the correlation coefficients (r_s), ranging from strong negative (dark blue) to strong positive (dark red) associations.



Accumulated degree-days (degree_days) emerged as the single strongest predictor for grape maturation across both fields. It exhibited a robust positive correlation with soluble solids content (ssc), with rs values of 0.84 in Field 5 and 0.86 in Field 6. Furthermore, degree days showed a moderate to strong positive correlation with pH (rs = 0.42 and rs = 0.64, for Fields 5 and 6, respectively) and a corresponding negative correlation with tartaric acidity (ta), which was strong in Field 5 (rs = -0.82) but moderate in Field 6 (rs = -0.54).

Conversely, the majority of vegetation indices demonstrated relationships that were inversely related to those of degree days. Soluble solids content (ssc) was strongly and negatively correlated with multiple VIs in both fields. The strongest negative associations were observed with indices such as nli (rs = -0.58; -0.76), ndre (rs = -0.60; -0.77), mcari (rs = -0.57; -0.75), and the normalized difference vegetation index (ndvi) (rs = -0.53; -0.70).

Tartaric acidity (ta) generally showed moderate positive correlations with the VIs. For instance, ndvi presented rs values of 0.44 (Field 5) and 0.41 (Field 6) against ta. Similarly, gndvi correlated positively with ta (rs = 0.44 and 0.41, respectively). The relationship between VIs and pH was less consistent across sites. In Field 5, most VIs showed negligible correlations with pH (e.g., ndvi, rs = -0.14). However, in Field 6, moderate negative correlations were observed, such as with nli (rs = -0.55) and ndvi (rs = -0.48).

The initial modeling phase consisted of identifying the most relevant predictors for each target variable using the Recursive Feature Elimination (RFE) method. The importance of each predictor was quantified by the percent increase in Mean Square Error (%IncMSE) when the variable's values were permuted.

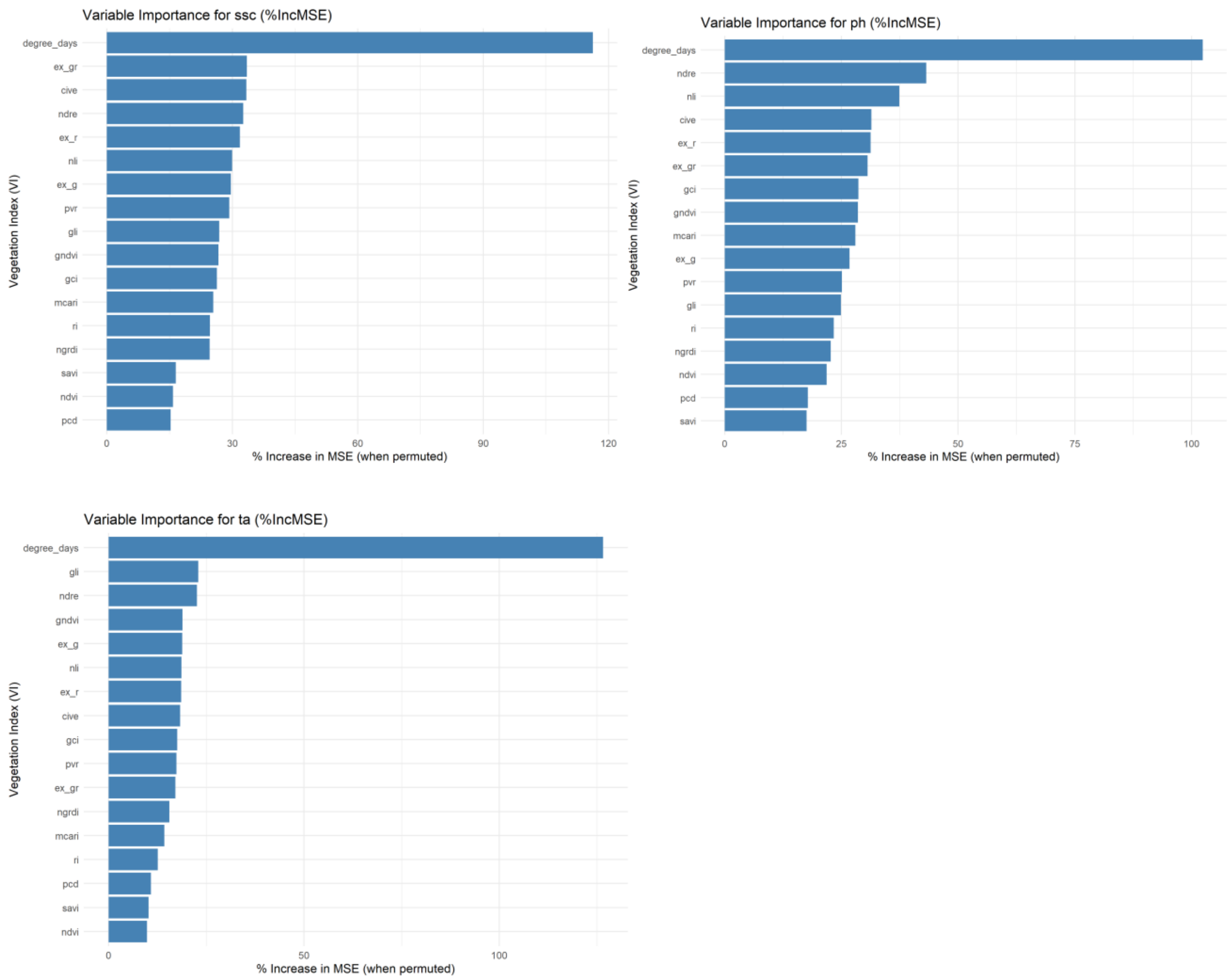
Across all three maturation targets, accumulated degree-days were identified as the single most dominant predictor by a substantial margin. For Soluble Solids Content (SSC), degree days showed overwhelming importance (approx. 118% IncMSE). A secondary group of predictors was identified among the vegetation indices, led by exgr (approx. 34%), cive (approx. 33%), ndre (approx. 32%), and exr (approx. 31%).

For pH, degree days also ranked first (approx. 105% IncMSE), though its dominance was less pronounced compared to SSC and TA. Notably, the ndre index (approx. 45%) was identified as the second-most important predictor, demonstrating an importance level nearly equal to that of degree days. This was followed by nli (approx. 38%), cive (approx. 30%) and exr (approx. 28%)

For Tartaric Acidity (TA), degree days was again the most critical predictor (approx. 122% IncMSE). The relative importance of the vegetation indices differed from SSC, with gli (approx. 20%), ndre (approx. 18%), gndvi (approx. 14%) and exg (approx. 14%).

As per the methodology, the top five ranked predictors for each specific target variable (SSC, TA, and pH) were selected from these results to proceed with the development of the final predictive models.

Figure 3 – Relative importance of thermal and multispectral predictors for estimating key enological targets, evaluated via the percent increase in Mean Squared Error (%IncMSE). (A) Soluble Solids Content (SSC). (B) pH. (C) Tartaric Acidity (TA). Accumulated degree-days consistently ranks as the dominant predictor across all three parameters.



Following the variable selection by RFE, the predictive performance of the machine learning models was evaluated for Soluble Solids Content (SSC), with results summarized in Table 3.

Table 3 - Statistical evaluation criteria (R^2, RMSE, and MAE) applied to the machine learning models for Soluble Solids Content (SSC) prediction, discriminating between the 2024 training and the 2025 independent testing phases.

Model	Dataset	R2	RMSE	MAE
RF	Test (2025)	0.73013	2.432747	2.018404
SVM	Test (2025)	0.688403	2.238039	1.861004
XGB	Test (2025)	0.627565	2.504457	2.0565
RF	Training (2024)	0.967324	0.512857	0.405478
XGB	Training (2024)	0.93832	0.698516	0.564665
SVM	Training (2024)	0.93383	0.721811	0.560771

The models demonstrated a high degree of fit to the 2024 training dataset. The Random Forest (RF) algorithm achieved the strongest fit ($R^2 = 0.967$, RMSE = 0.513), followed by XGBoost ($R^2 = 0.938$, RMSE = 0.699) and SVM ($R^2 = 0.934$, RMSE = 0.722). These low error metrics on the training set confirmed that all models successfully learned the data patterns from the 2024 season.

The primary assessment, however, was the temporal validation using the independent 2025 test dataset. As expected, a performance reduction was observed, reflecting inter-seasonal variability. In this validation, the SVM model emerged as the most robust predictor, achieving the lowest error metrics with an RMSE of 2.238 and an MAE of 1.861. While the Random Forest model explained the largest proportion of variance ($R^2 = 0.730$), its predictive accuracy was lower, indicated by a higher RMSE (2.433). The XGBoost model yielded the weakest performance on the independent test data ($R^2 = 0.627$, RMSE = 2.504).

Therefore, the SVM model was identified as the most effective for generalizing predictions of SSC to a future, unseen season.

For the prediction of pH, a similar pattern was observed during model training (Table 4). All algorithms demonstrated a strong fit to the 2024 data, with Random Forest achieving the highest R^2 (0.947) and lowest RMSE (0.064).

Table 4 - Statistical evaluation criteria (R^2, RMSE, and MAE) applied to the machine learning models for pH prediction, discriminating between the 2024 training and the 2025 independent testing phases.

Model	Dataset	R2	RMSE	MAE
RF	Test (2025)	0.3023433	0.24159356	0.20812643
SVM	Test (2025)	0.1357797	0.26553918	0.23202839
XGB	Test (2025)	0.0286009	0.33230761	0.26543588
RF	Training (2024)	0.9470950	0.06412163	0.04547285
XGB	Training (2024)	0.8976785	0.08748702	0.06830497
SVM	Training (2024)	0.8975135	0.08788017	0.06377096

However, the temporal validation on the 2025 test dataset revealed a significant degradation in predictive power for this variable. The Random Forest model provided the best, albeit weak, predictive performance, achieving the highest R^2 (0.302) and the lowest RMSE (0.242). The SVM ($R^2 = 0.136$, RMSE = 0.266) and XGBoost ($R^2 = 0.029$, RMSE = 0.332) models failed to generalize, with R^2 values approaching zero. These results indicate that the models, while effective for SSC, could not effectively predict pH across seasons based on the provided inputs.

For the prediction of Tartaric Acidity (TA), the models again demonstrated a robust fit to the 2024 training data (Table 5). The Random Forest algorithm provided the best training metrics ($R^2 = 0.889$, RMSE = 1.742).

Table 5 - Statistical evaluation criteria (R^2, RMSE, and MAE) applied to the machine learning models for Tartaric Acidity (TA) prediction, discriminating between the 2024 training and the 2025 independent testing phases.

Model	Dataset	R2	RMSE	MAE
RF	Test (2025)	0.7121591	3.217871	2.658605
SVM	Test (2025)	0.6729033	3.271454	2.661818
XGB	Test (2025)	0.5820201	3.467074	2.617628
RF	Training (2024)	0.8893676	1.742304	1.089571
XGB	Training (2024)	0.7884381	2.358262	1.554809
SVM	Training (2024)	0.7576251	2.551055	1.413271

In the temporal validation on the 2025 dataset, the models demonstrated a successful capacity to generalize, a result that contrasts sharply with the findings for pH. The Random Forest model yielded the strongest predictive performance, achieving the highest R^2 (0.712) and the lowest RMSE (3.218). The SVM ($R^2 = 0.673$, RMSE = 3.271) and XGBoost ($R^2 = 0.582$, RMSE = 3.467) models also showed predictive capability but were outperformed by the RF.

These findings suggest that both SSC and TA can be effectively modeled across seasons, while pH prediction remains a significant challenge.

Although all four models generated a consistent three-zone structure (k = 3), they exhibited systematic spatial differences in the position of zone boundaries and in the relative extent of each class. In all cases, the general pattern was maintained lower-value zones in the western portion (blue), intermediate values in the central region (red), and higher values in the eastern sector (yellow). However, the transitions between classes differed according to the combination of topographic, soil, and spectral variables included in each model.

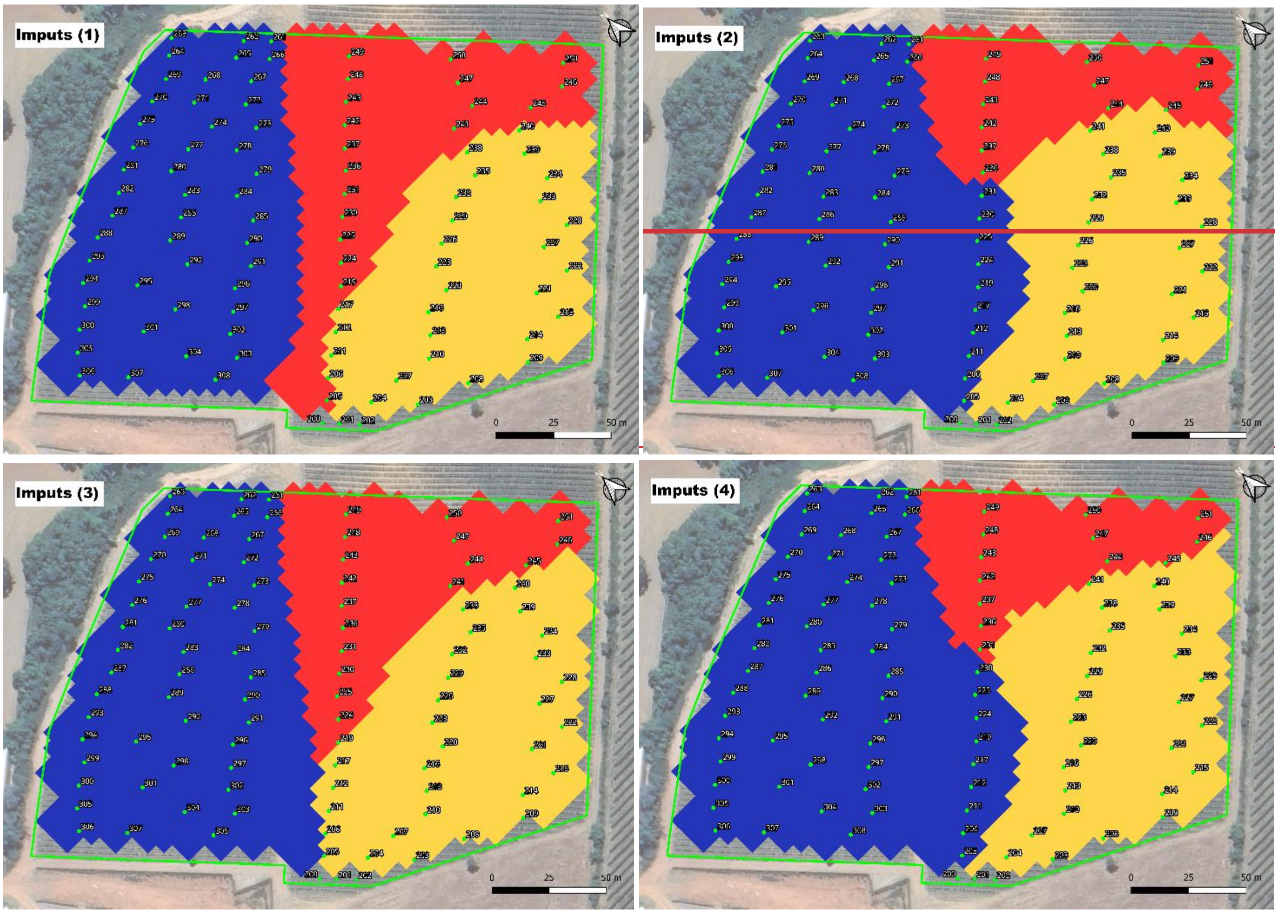
Model 1 (Elevation + EC) produced sharper and more linear boundaries, especially between the blue and red zones. This configuration reflects the influence of elevation, which introduced a strong topographic gradient that aligned with variations captured by soil electrical conductivity. The central region displayed relatively stable and well-defined boundaries, suggesting a strong coupling between terrain position and soil physical–chemical properties.

In Model 2 (Insolation + EC), the overall structure was preserved, but boundaries shifted slightly, particularly between the red and yellow zones. Replacing elevation with annual insolation generated smoother spatial transitions, mainly in the upper and central portions of the field. Compared to Model 1, there was a subtle expansion of the yellow zone toward the southeastern sector, indicating that spatial variability in energy input modulated the relationship between vegetation signals and soil conductivity;

Model 3 (Elevation + SBI) displayed more internally homogeneous zones, with reduced local variability and smoother polygonal transitions. Substituting EC with the Soil Brightness Index (SBI) refined the boundary definition, especially between the red and yellow zones. Because SBI is sensitive to texture, organic matter content, and soil exposure, this model captured finer-scale soil variability, resulting in less fragmentation and more coherent spatial structures relative to the EC-based models.

Model 4 (Insolation + SBI) presented the most continuous zoning pattern, with the smoothest boundaries and the highest internal consistency among zones. The combination of insolation with SBI reduced the minor shifts observed in the other models and produced a more gradual alignment between the blue and red zones, while further stabilizing the yellow zone in the eastern region. This suggests that the interaction between energy availability and optical soil properties enhances the spatial coherence of management zones.

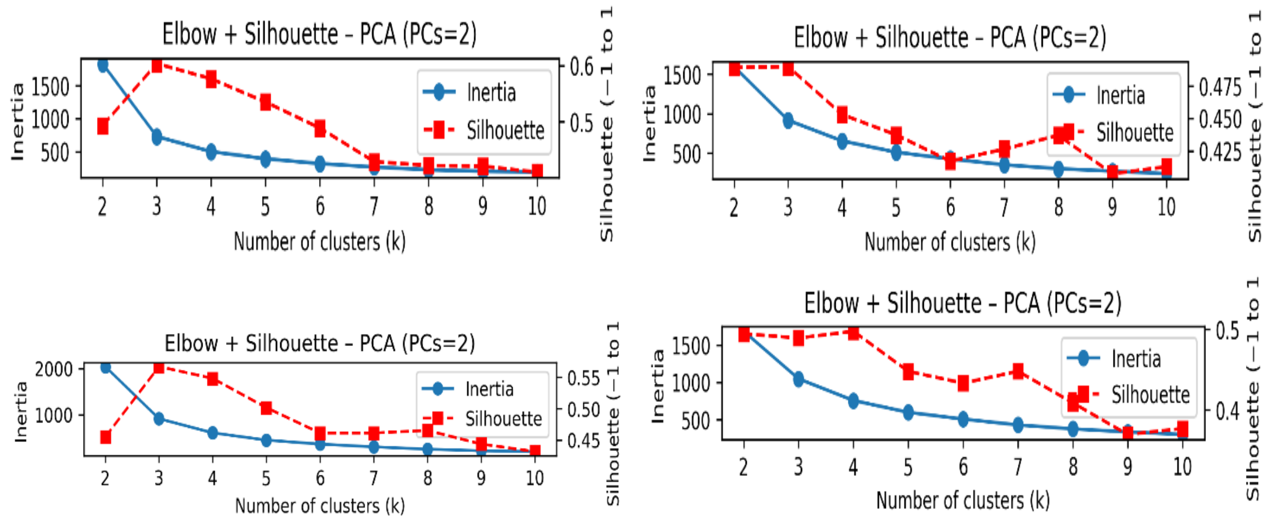
Figure 4 – Delineation of categorical microterroir zones ($k = 3$) using four different clustering models based on topographic, soil, and spectral inputs. Panels represent: (A) Model 1 (Elevation + EC); (B) Model 2 (Insolation + EC); (C) Model 3 (Elevation + SBI); and (D) Model 4 (Insolation + SBI). Map colors serve as categorical identifiers for the distinct microterroirs identified across the study area.



To ensure a standardized comparison across all four microterroir hypotheses, the optimal number of clusters (k) was statistically determined for each of the four models. The combined analysis of the Elbow Method (WCSS/Inertia) and the Average Silhouette Score was employed on the first two Principal Components (PCs) for each model.

In Model 1, the Silhouette coefficient reached its maximum at $k = 3$, decreasing steadily for higher values of k, whereas the Elbow curve suggested a milder inflection at $k = 4$. Model 2 showed a markedly different pattern, with the highest Silhouette value occurring at $k = 2$ and a progressive reduction for $k \geq 3$, indicating diminished cluster separability despite an Elbow trend pointing toward *k = 4*. Model 3 again demonstrated a clear Silhouette optimum at *k = 3*, accompanied by the largest decline in inertia between *k = 2* and *k = 3*, making it the most consistent model across both metrics. Model 4 showed comparable Silhouette values at *k = 2* and *k = 3*, with a slight advantage for *k = 2*, while the Elbow curve still indicated a potential inflection at *k = 3*.

Figure 5 - Combined optimization metrics for determining the appropriate number of management zones (*k*). The plots display the Within-Cluster Sum of Squares (Inertia) and Silhouette coefficients applied to the PCA-reduced datasets (PCs = 2). (A) Model 1. (B) Model 2. (C) Model 3. (D) Model 4.

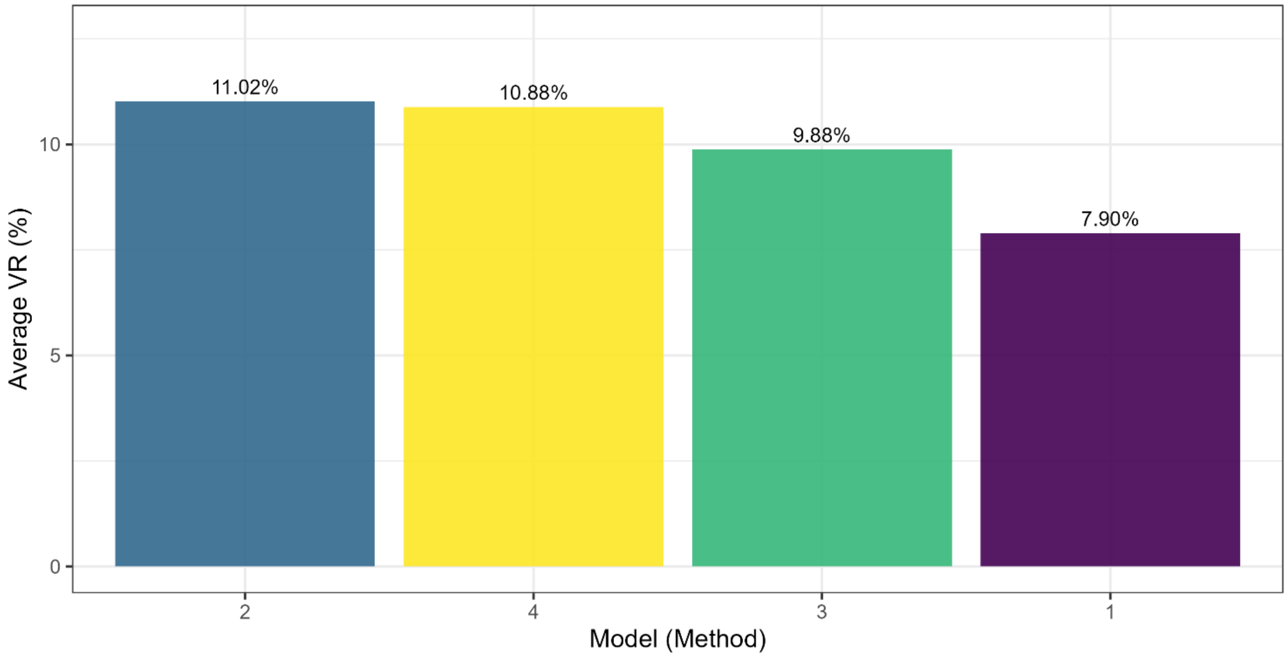


Collectively, the models differed primarily in the Silhouette-based optimal number of clusters: Models 1 and 3 favored *k = 3*, whereas Models 2 and 4 favored *k = 2*, suggesting that variations in preprocessing or model specification substantially altered the underlying cluster structure.

Given this overwhelming statistical convergence from both methods, and in alignment with the agronomic goal of creating practical 'High/Medium/Low' potential zones for viticultural management, k=3 was confirmed as the most robust, balanced, and statistically appropriate number of clusters for all four delineation models. This standardization allows for a direct and equitable comparison of the spatial patterns generated by each set of input variables.

The overall performance of each model, calculated as the mean VR% across all six validation variables, is presented in Figure 7.

Figure 6 – Ranking of the overall performance of the four zoning methods. The graph displays the average VR% calculated across all validation parameters for each model, highlighting Model 2 as the most statistically balanced.

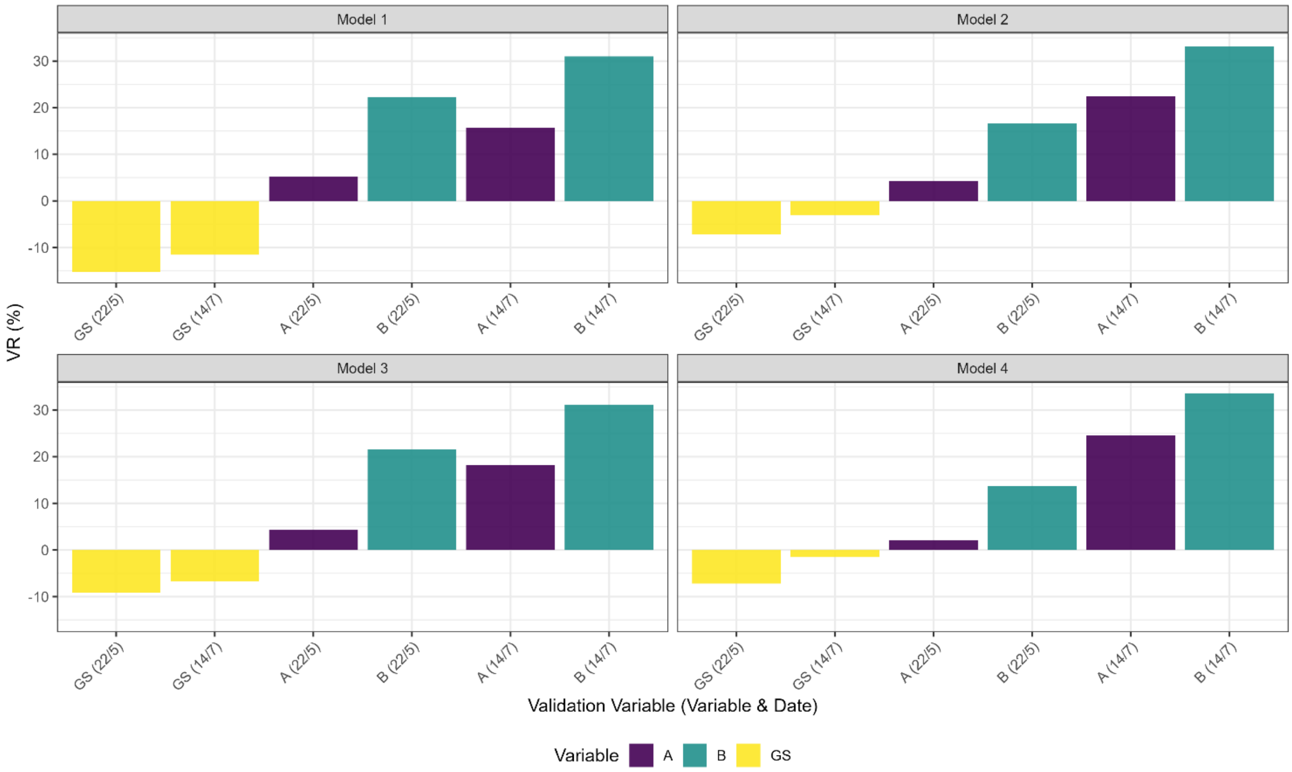


The analysis revealed distinct differences in model performance. Model 2 emerged as the most statistically balanced, achieving the highest average Variance Reduction (Average_VR = 11.02%). It was followed closely by Model 4 (Average_VR = 10.88%).

Conversely, Model 1 and Model 3 demonstrated significantly weaker clustering efficacy, with average VR values of 7.90% and 9.88%, respectively. While Model 2 provided the highest *average* performance, the near-equivalent score of Model 4 necessitates a more detailed analysis to determine the optimal model for practical application.

To understand the specific strengths and weaknesses of each model, a detailed analysis of VR% for each individual validation variable was conducted (Figure 8).

Figure 7 – Detailed Variance Reduction (VR%) analysis evaluating the specific clustering efficacy of the four zoning models across individual validation variables. The bar charts illustrate the VR% for chlorophyll a (A), chlorophyll b (B), and Normalized Difference Vegetation Index (NDVI) values acquired via a GreenSeeker active sensor (GS), evaluated on two distinct dates (May 22 and July 14). Panels correspond to: (A) Model 1; (B) Model 2; (C) Model 3; and (D) Model 4.



This granular analysis yielded two critical findings:

1. High Efficacy in Clustering Chlorophyll Variables: A significant outcome was the high VR% achieved for the Chlorophyll B (Variable B) variable measured on 14 july. All four models successfully partitioned this variable, with Model 4 achieving the highest efficacy (VR% = 33.62%), followed by Model 2 (VR% = 33.07%). Similarly, Model 4 also showed the best performance for Chlorophyll A (Variable A) on the same date (VR% = 24.57%).

This strong performance suggests that the input data used to generate the clustering models possess a high degree of spatial correlation with the canopy's physiological state, specifically Chlorophyll content, during this particular phenological stage (14 july). This finding validates the use of these input variables for creating zones predictive of chlorophyll content.

2. Systemic Failure to Model GreenSeeker (GS) NDVI: The most striking finding from the detailed analysis was the systemic failure of all four models to effectively partition the GreenSeeker (GS) NDVI data. As illustrated in Figure 19, the VR% for both GS variables (14 july and 22 may) were negative across all models.

A negative VR% indicates that the variance *within* the delineated zones was greater than the variance between the zones, implying that the clustering was counter-productive for this variable. This suggests a critical disconnect between the spatial input data used to create the models and the data captured by the proximal GreenSeeker sensor.

DISCUSSION

While some research has studied VIs in viticulture (GIOVOS et al., 2021), and the use of machine learning techniques for the estimation of grape quality (KASIMATI et al., 2021b) this study investigates an alternative non-destructive approach using precision viticulture techniques to predict grape maturation and delineate microterroir zones.

The correlation analysis reveals a distinct ecophysiological dynamic between the multispectral vegetation indices (VIs) and grape maturation attributes. Soluble solids content (SSC) exhibited a strong negative correlation with structural indices such as NLI, NDRE, MCARI, and NDVI. High values of these VIs indicate zones with high vegetative vigor and dense canopies (ZARCO-TEJADA et al., 2005). While vigorous vines exhibit high photosynthetic capacity, excessively dense canopies induce cluster shading, leading to reduced yield components and delayed fruit ripening, thereby compromising the final quality of the grapes and wine (SMART et al., 1990).

Consequently, high-vigor zones frequently present delayed technological maturation and lower sugar accumulation. This dynamic is corroborated by (FOUNTAS et al., 2014), who demonstrated that in vineyards with restricted management and irrigation, lower-vigor zones (lower NDVI) yield grapes with higher soluble solids and superior enological quality, such as elevated anthocyanin and phenolic content. Interestingly, while these authors found no significant spatial correlation between vegetative vigor indices and acidity, our results demonstrate a clear ecophysiological dynamic for this parameter.

In the present study, tartaric acidity (TA) showed moderate positive correlations with the VIs, as evidenced by NDVI (r_s = 0.44 in Field 5; r_s = 0.41 in Field 6) and GNDVI (r_s = 0.44 and 0.41, respectively). This direct relationship is driven by the microclimatic temperature within the canopy. It is well established that high temperatures alter key cellular mechanisms during ripening, accelerating the degradation of organic acids before the berries reach optimal maturity (RIENTH et al., 2016).

Therefore, more vigorous vines (associated with high VIs) develop denser canopies that provide shading and naturally reduce cluster temperatures. This thermal mitigation effectively slows down metabolic degradation, preserving higher acidity levels in high-vigor zones until harvest, a physiological response perfectly aligned with the literature (SWEETMAN et al., 2014).

Furthermore, the relationship between VIs and pH proved to be inconsistent across the evaluated fields, ranging from negligible to moderately negative. This instability occurs because grape pH is not strictly governed by canopy biomass or vegetative vigor, parameters effectively captured by remote sensing (DOBROWSKI; USTIN; WOLPERT, 2003). Instead, berry pH is heavily influenced by soil properties and water dynamics, particularly the uptake and accumulation of potassium from the soil into the berries. Because variations in soil exchangeable potassium strongly dictate berry pH

(KODUR, 2011) relying solely on canopy reflectance metrics is insufficient to model pH consistently across different spatial zones.

Despite the high coefficients of determination observed in the training and validation sets (R^2 ranging from 0.71 to 0.73), a rigorous analysis of the absolute errors reveals significant limitations for the immediate practical application of machine learning models in harvest decision-making. This performance aligns with recent findings by (KASIMATI et al., 2021b), who reported a maximum R^2 of 0.61 when predicting total soluble solids using a combination of proximal and remote sensing NDVI data across critical growth stages.

The RMSE values obtained in this study for soluble solids content (2.43 °Brix) and tartaric acidity (3.22 g L^{-1}) indicate that although the Random Forest algorithm captures the macro-temporal trend of grape maturation with excellent precision, it is heavily driven by growing degree-days (GDD) accumulation. This dependence is classically supported by the Spring Warming (SW) model, which establishes GDD as the fundamental metric to estimate grapevine phenological progression (PARKER et al., 2011).

Ultimately, the developed predictive models function as a robust Decision Support System (DSS) rather than a complete replacement for traditional analytical methodologies. By mapping and forecasting macro-temporal maturation trends weeks in advance, these tools provide critical preliminary monitoring that optimizes targeted field sampling, labor allocation, and winery reception logistics (ARNÓ et al., 2009); (BRAMLEY, 2010b). Nevertheless, as the estimated harvest window approaches, classical physicochemical analyses remain indispensable to determine the precise enological maturity and ensure the rigorous quality standards demanded by the fine wine market.

The spatial delineation results derived from the unsupervised K-Means clustering demonstrate a clear distinction between the scales of phenotypic variability within the vineyard. Determining the optimal number of clusters requires balancing statistical rigor with operational feasibility. Although the Silhouette coefficient favored a two-cluster structure ($k = 2$) for Models 2 and 4, we selected $k = 3$ to identify distinct *microterroir* zones (High, Medium, and Low potential). Constraining the delineation to $k = 2$ would mask intermediate gradients that are essential for variable-rate management and selective harvesting in viticulture.

This approach aligns with (GAVIOLI et al., 2019), who demonstrated that while statistical metrics frequently default to capturing only extreme field contrasts ($k = 2$), effective clustering must maximize variance reduction without inducing spatial fragmentation. By enforcing $k = 3$, Model 4 successfully balanced these criteria, capturing the necessary biological resolution of *microterroirs* while maintaining the spatial contiguity required for practical field operations.

Model 4 (Insolation + SBI) demonstrated considerable robustness in accommodating this operational standardization, generating the most continuous and spatially coherent boundaries among the three zones. By smoothing transitions and reducing the local fragmentation that predominantly affected models based on electrical conductivity (EC), Model 4 refined boundary delineation. As highlighted by (CORWIN; LESCH, 2005)and (MOLIN; FAULIN, 2013), while apparent soil EC is a standard proxy in precision agriculture, it integrates multiple deep soil profile properties (such as subsurface moisture, clay lenses, and salinity). This deep-profile integration often introduces high spatial fragmentation and weak or inconsistent correlations with the actual surface canopy vigor.

In contrast, combining surface optical properties with climate-driven topographical variables better captures functional spatial variability and improves spatial contiguity (REYES et al., 2023). Consequently, Model 4 effectively transformed a forced statistical convergence into a practical and highly efficient tool for the spatial management of the plot.

The statistical superiority of insolation-based models (Models 2 and 4) over those driven by absolute elevation (Models 1 and 3) reveals that local microclimatic dynamics exert a greater influence on the vineyard ecosystem than static topography alone. Elevation frequently serves in the literature as a generic proxy for temperature gradients; however, as demonstrated by (BOIS et al., 2008), in areas with complex terrain, annual insolation captures the effective solar radiation actually reaching the canopy. This available energy directly modulates water balance, grapevine development and berry ripening, overriding the simplified effect of altitude.

This structural validation is further consolidated by the capacity of Model 4 to stratify the physiological variables of the crop, achieving a substantial Variance Reduction for Chlorophyll b ($VR = 33.62\%$) and Chlorophyll a ($VR = 24.57\%$) during the July 14 sampling. Within the phenological dynamics of vineyards managed under winter harvest regimes via double pruning, mid-July coincides with critical, advanced stages of fruit ripening (FAVERO et al., 2011). It is during this period that the vegetative canopy reaches its maximum structural stability.

As observed by (ANIĆ et al., 2024) regarding leaf pigment evolution, this represents the exact phase where the concentration of chlorophylls a and b peaked during veraison and then decreased during grape ripening, fully expressing the cumulative environmental effect on pigment synthesis and degradation. The strong spatial correlation between the generated zoning and foliar chlorophyll levels validates the premise that the interaction between incident solar energy and optical soil surface properties governs the physiological vigor and photosynthetic activity of the vines. As established by (BRAMLEY, 2005) in the foundational concepts of precision viticulture, when management zones accurately capture spatial variations in canopy vigor, they transcend mere mathematical divisions, functioning instead as robust biological predictors of the vineyard's metabolic state leading up to harvest.

Conversely, the systemic failure of all models to stratify data from the proximal GreenSeeker sensor, which resulted in negative VR values in both May and July, exposes a clear incompatibility of physical and temporal scales. The proposed zoning is based on permanent edaphoclimatic factors and medium-spatial-resolution orbital sensors, reflecting the macro-structure and the stable baseline potential of the plot. In contrast, the GreenSeeker is an active proximal sensor with an extremely narrow, sub-metric field of view (DRISSI et al., 2009).

In discontinuous canopies, active proximal sensors primarily capture sub-metric architectural noise, such as foliage gaps and localized leaf orientation, rather than the baseline agronomic potential. This divergence highlights a complementary relationship rather than a contradiction. Macro-zones establish broad areas for structural management based on stable environmental factors, whereas proximal sensors should guide real-time, variable-rate applications to address the fine-scale variability that macro-models cannot resolve.

CONCLUSION

This study successfully achieved its objectives by developing temporal predictive models for grape technological maturation and delineating spatial microterroir zones to optimize winter wine production under subtropical conditions.

Regarding the temporal forecasting, Random Forest models driven by Growing Degree-Days (GDD) and multispectral vegetation indices effectively captured macro-temporal ripening trends for soluble solids and tartaric acidity. Although these algorithms do not replace traditional pre-harvest physicochemical analyses, they function as a practical Decision Support System (DSS) to optimize targeted field sampling and harvest logistics.

Spatially, the delineation of microterroir zones based on annual insolation and surface optical properties (SBI) outperformed traditional elevation and soil proxies. Enforcing a three-cluster structure ($k = 3$) proved essential, providing the optimal balance between statistical accuracy and the practical feasibility required to harvest distinct zones independently.

Ultimately, combining these stable microterroir zones with the temporal tracking of technological maturation establishes a structural baseline for precision viticulture under the double pruning regime. This integrated framework directly supports selective step-wise harvesting, ensuring grapes are picked at their optimal technological maturity while mitigating the operational costs of specialized agricultural labor.

ACKNOWLEDGMENTS:

We acknowledge the Federal University of Lavras (UFLA) for its institutional support and infrastructure, as well as the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and the Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) for the financial support and scholarships that were fundamental to conducting this study.

REFERENCES

ACEVEDO-OPAZO, C.; TISSEYRE, B.; GUILLAUME, S.; OJEDA, H. The potential of high spatial resolution information to define within-vineyard zones related to vine water status. *In*: PRECISION AGRICULTURE 2008, **Anais** [...]. [s.l: s.n.] p. 285–302. DOI: 10.1007/s11119-008-9073-1.

AGGOGERI, Francesco; PELLEGRINI, Nicola; TAGLIANI, Franco Luis. **Recent advances on machine learning applications in machining processes**. **Applied Sciences (Switzerland)**MDPI, , 2021. DOI: 10.3390/app11188764.

ANIĆ, Marina; KONTIĆ, Jasminka Karoglan; RENDULIĆ, Nera; ČARIJA, Mate; OSREČAK, Mirela; KAROGLAN, Marko; ANDABAKA, Željko. Evolution of Leaf Chlorophylls, Carotenoids and Phenolic Compounds during Vegetation of Some Croatian Indigenous Red and White Grape Cultivars. **Plants 2024, Vol. 13, Page 971**, *[S. l.]*, v. 13, n. 7, p. 971, 2024. DOI: 10.3390/PLANTS13070971. Disponível em: https://www.mdpi.com/2223-7747/13/7/971/htm. Acesso em: 27 jun. 2026.

ARNÓ, J.; MARTÍNEZ-CASASNOVAS, J. A.; RIBES-DASI, M.; ROSELL, J. R. Revisión. Viticultura de precisión. Líneas de investigación, retos y oportunidades del manejo sitio-específico en viña. **Spanish Journal of Agricultural Research**, *[S. l.]*, v. 7, n. 4, p. 779–790, 2009. DOI: 10.5424/SJAR/2009074-1092. Acesso em: 26 jun. 2026.

BACK, Alvaro José; BRUNA, Emílio Della. Soma térmica nas fases fenológicas da videira ‘Niágara Rosada’ cultivada em Urussanga, Santa Catarina. **Agropecuária Catarinense**, *[S. l.]*, v. 25, n. 3, p. 85–89, 2012. Disponível em: https://publicacoes.epagri.sc.gov.br/rac/article/view/670. Acesso em: 13 jun. 2025.

BARBOZA, Thiago Orlando Costa; FERRAZ, Marcelo Araújo Junqueira; PILON, Cristiane; VELLIDIS, George; VALERIANO, Taynara Tuany Borges; DOS SANTOS, Adão Felipe. Advanced Farming Strategies Using NASA POWER Data in Peanut-Producing Regions without Surface Meteorological Stations. **AgriEngineering**, *[S. l.]*, v. 6, n. 1, p. 438–454, 2024. DOI: 10.3390/AGRIENGINEERING6010027/S1. Disponível em: https://www.mdpi.com/2624-7402/6/1/27/htm. Acesso em: 13 jun. 2025.

BOIS, B.; WALD, L.; PIERI, Ph; VAN LEEUWEN, C.; COMMAGNAC, L.; CHERY, Ph; CHRISTEN, M.; GAUDILLÈRE, J. P.; SAUR, É. Estimating spatial and temporal variations in solar radiation within Bordeaux winegrowing region using remotely sensed data. **OENO One**, *[S. l.]*, v. 42, n. 1, p. 15–25, 2008. DOI: 10.20870/OENO-ONE.2008.42.1.829. Disponível em: https://oeno-one.eu/article/view/829. Acesso em: 27 jun. 2026.

BRAMLEY, R. G. V. Understanding variability in winegrape production systems 2. Within vineyard variation in quality over several vintages. **Australian Journal of Grape and Wine Research**, *[S. l.]*, v. 11, n. 1, p. 33–42, 2005. DOI: 10.1111/J.1755-0238.2005.TB00277.X. Disponível em: /doi/pdf/10.1111/j.1755-0238.2005.tb00277.x. Acesso em: 16 maio. 2026.

BRAMLEY, R. G. V. Precision Viticulture: managing vineyard variability for improved quality outcomes. **Managing Wine Quality: Viticulture and Wine Quality**, *[S. l.]*, p. 445–480, 2010. a. DOI: 10.1533/9781845699284.3.445. Disponível em:

https://www.sciencedirect.com/science/chapter/edited-volume/abs/pii/B9781845694845500129. Acesso em: 4 jul. 2026.

BRAMLEY, R. G. V. Precision Viticulture: managing vineyard variability for improved quality outcomes. **Managing Wine Quality: Viticulture and Wine Quality**, *[S. l.]*, p. 445–480, 2010. b. DOI: 10.1533/9781845699284.3.445. Disponível em: https://www.sciencedirect.com/science/chapter/edited-volume/abs/pii/B9781845694845500129. Acesso em: 26 jun. 2026.

BRAMLEY, R. G. V.; HAMILTON, R. P. Terroir and precision viticulture: are they compatible ? **OENO One**, *[S. l.]*, v. 41, n. 1, p. 1–8, 2007. DOI: 10.20870/OENO-ONE.2007.41.1.855. Disponível em: https://oeno-one.eu/article/view/855. Acesso em: 5 out. 2025.

BRAMLEY, R. G. V.; OUZMAN, J.; BOSS, P. K. Variation in vine vigour, grape yield and vineyard soils and topography as indicators of variation in the chemical composition of grapes, wine and wine sensory attributes. **Australian Journal of Grape and Wine Research**, *[S. l.]*, v. 17, n. 2, p. 217–229, 2011. DOI: 10.1111/J.1755-0238.2011.00136.X. Acesso em: 5 out. 2025.

BRAMLEY, R. G. V.; OUZMAN, J.; THORNTON, C. Selective harvesting is a feasible and profitable strategy even when grape and wine production is geared towards large fermentation volumes. **Australian Journal of Grape and Wine Research**, *[S. l.]*, v. 17, n. 3, p. 298–305, 2011. DOI: 10.1111/J.1755-0238.2011.00151.X;WGROUP:STRING:PUBLICATION. Disponível em: /doi/pdf/10.1111/j.1755-0238.2011.00151.x. Acesso em: 24 jul. 2026.

BRAMLEY, R. G. V.; HAMILTON, R. P. **Hitting the zone - making viticulture more precise**. Urrbrae, S. Aust.: . Acesso em: 5 out. 2025.

BRANT, Luciana Alves Caldeira; DE SOUZA, Cláudia Rita; DA MOTA, Renata Vieira; FERNANDES, Fernanda de Paula; GONÇALVES, Mariana Gabriele Marcolino; DE MENEZES, Michele Duarte; PEREGRINO, Isabela; CURI, Nilton; REGINA, Murillo de Albuquerque. Macro scale analysis of syrah vineyards under winter growing cycles: Agronomical and ecophysiological responses. **Scientia Agricola**, *[S. l.]*, v. 78, n. 6, p. 1–9, 2020. DOI: 10.1590/1678-992x-2020-0055.

BREIMAN, Leo. Random forests. **Machine Learning**, *[S. l.]*, v. 45, n. 1, p. 5–32, 2001. DOI: 10.1023/A:1010933404324/METRICS. Disponível em: https://link.springer.com/article/10.1023/A:1010933404324. Acesso em: 14 jun. 2025.

CHAI, T.; DRAXLER, R. R. Root mean square error (RMSE) or mean absolute error (MAE)? - Arguments against avoiding RMSE in the literature. **Geoscientific Model Development**, *[S. l.]*, v. 7, n. 3, p. 1247–1250, 2014. DOI: 10.5194/GMD-7-1247-2014,. Acesso em: 14 jun. 2025.

CHEN, Shanshan; ZHANG, Fangfang; NING, Jifeng; LIU, Xu; ZHANG, Zhenwen; YANG, Shuqin. Predicting the anthocyanin content of wine grapes by NIR hyperspectral imaging. **Food Chemistry**, *[S. l.]*, v. 172, p. 788–793, 2015. DOI: 10.1016/J.FOODCHEM.2014.09.119. Disponível em: https://www.sciencedirect.com/science/article/abs/pii/S0308814614015040?via%3Dihub. Acesso em: 4 jul. 2026.

CHEN, Tianqi; GUESTRIN, Carlos. XGBoost: A Scalable Tree Boosting System. **Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**, *[S.*

l.], v. 13-17- August-2016, p. 785–794, 2016. DOI: 10.1145/2939672.2939785. Disponível em: https://arxiv.org/pdf/1603.02754. Acesso em: 14 jun. 2025.

CONRAD, O.; BECHTEL, B.; BOCK, M.; DIETRICH, H.; FISCHER, E.; GERLITZ, L.; WEHBERG, J.; WICHMANN, V.; BÖHNER, J. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. **Geoscientific Model Development**, *[S. l.]*, v. 8, n. 7, p. 1991–2007, 2015. DOI: 10.5194/GMD-8-1991-2015. Disponível em: www.geosci-model-dev.net/8/1991/2015/. Acesso em: 20 out. 2025.

CORTES, Corinna; VAPNIK, Vladimir; SAITTA, Lorenza. Support-vector networks. **Machine Learning 1995 20:3**, *[S. l.]*, v. 20, n. 3, p. 273–297, 1995. DOI: 10.1007/BF00994018. Disponível em: https://link.springer.com/article/10.1007/BF00994018. Acesso em: 14 jun. 2025.

CORWIN, D. L.; LESCH, S. M. Apparent soil electrical conductivity measurements in agriculture. **Computers and Electronics in Agriculture**, *[S. l.]*, v. 46, n. 1–3, p. 11–43, 2005. DOI: 10.1016/J.COMPAG.2004.10.005. Disponível em: https://www.sciencedirect.com/science/article/abs/pii/S0168169904001243?via%3Dihub. Acesso em: 27 jun. 2026.

DAUGHTRY, C. S. T.; WALTHALL, C. L.; KIM, M. S.; DE COLSTOUN, E. Brown; MCMURTREY, J. E. Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. **Remote Sensing of Environment**, *[S. l.]*, v. 74, n. 2, p. 229–239, 2000. DOI: 10.1016/S0034-4257(00)00113-9. Disponível em: https://www.researchgate.net/publication/222661408_Estimating_Corn_Leaf_Chlorophyll_Concentration_from_Leaf_and_Canopy_Reflectance. Acesso em: 10 mar. 2025.

DOBROWSKI, S. Z.; USTIN, S. L.; WOLPERT, J. A. Grapevine dormant pruning weight prediction using remotely sensed data. **Australian Journal of Grape and Wine Research**, *[S. l.]*, v. 9, n. 3, p. 177–182, 2003. DOI: 10.1111/J.1755-0238.2003.TB00267.X;PAGE:STRING:ARTICLE/CHAPTER. Disponível em: /doi/pdf/10.1111/j.1755-0238.2003.tb00267.x. Acesso em: 26 jun. 2026.

DRISSI, Rachid; GOUTOULY, Jean Pascal; FORGET, Dominique; GAUDILLERE, Jean Pierre. Nondestructive measurement of grapevine leaf area by ground normalized difference vegetation index. **Agronomy Journal**, *[S. l.]*, v. 101, n. 1, p. 226–231, 2009. DOI: 10.2134/AGRONJ2007.0167;SUBPAGE:STRING:ABSTRACT;REQUESTEDJOURNAL:JOURNAL:14350645;WEBSITE:WEBSITE:ACSESS.ONLINELIBRARY.WILEY.COM;WGROUP:STRING:PUBLICATION. Disponível em: /doi/pdf/10.2134/agronj2007.0167. Acesso em: 27 jun. 2026.

DRUSCH, M. et al. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. **Remote Sensing of Environment**, *[S. l.]*, v. 120, p. 25–36, 2012. DOI: 10.1016/J.RSE.2011.11.026. Disponível em: https://www.sciencedirect.com/science/article/abs/pii/S0034425712000636?via%3Dihub. Acesso em: 15 out. 2025.

ESCADAFAL RICHARD, Huete A. Etude des propriétés spectrales des sols arides appliquée à l'amélioration des indices de végétation obtenus par télédétection. **Comptes Rendus de l'Académie des Sciences**, *[S. l.]*, v. 312, n. Série 2, p. 1385–1391, 1991.

FAVERO, A. C.; ANGELUCCI DE AMORIM, D.; VIEIRA, R.; MOTA, D. A.; SOARES, A. M.; DE SOUZA, C. R.; DE ALBUQUERQUE REGINA, M.; REGINA, M. A. **Double-pruning of “Syrah” grapevines: a management strategy to harvest wine grapes during the winter in the Brazilian SoutheastVitis**. [s.l: s.n.].

FOUNTAS, Spyros; ANASTASIOU, Evaggelos; BALAFOUTIS, Athanasios; KOUNDOURAS, Stefanos; THEOHARIS, Serafeim; THEODOROU, Nikolaos. The influence of vine variety and vineyard management on the effectiveness of canopy sensors to predict winegrape yield and quality. **International Conference of Agricultural Engineering**, *[S. l.]*, 2014. Disponível em: www.eurageng.eu. Acesso em: 26 jun. 2026.

FRAGA, Helder; AMRAOUI, Malik; MALHEIRO, Aureliano C.; MOUTINHO-PEREIRA, José; EIRAS-DIAS, José; SILVESTRE, José; SANTOS, João A. Examining the relationship between the Enhanced Vegetation Index and grapevine phenology. **European Journal of Remote Sensing**, *[S. l.]*, v. 47, n. 1, p. 753–771, 2014. DOI: 10.5721/EUJRS20144743. Acesso em: 5 out. 2025.

GAVIOLI, Alan; DE SOUZA, Eduardo Godoy; BAZZI, Claudio Leones; SCHENATTO, Kelyn; BETZEK, Nelson Miguel. Identification of management zones in precision agriculture: An evaluation of alternative cluster analysis methods. **Biosystems Engineering**, *[S. l.]*, v. 181, p. 86–102, 2019. DOI: 10.1016/J.BIOSYSTEMSENG.2019.02.019. Disponível em: https://www.sciencedirect.com/science/article/abs/pii/S1537511018303295. Acesso em: 27 jun. 2026.

GIOVOS, Rigas; TASSOPOULOS, Dimitrios; KALIVAS, Dionissios; LOUGKOS, Nestor; PRIOVOLOU, Anastasia. Remote Sensing Vegetation Indices in Viticulture: A Critical Review. **Agriculture 2021, Vol. 11, Page 457**, *[S. l.]*, v. 11, n. 5, p. 457, 2021. DOI: 10.3390/AGRICULTURE11050457. Disponível em: https://www.mdpi.com/2077-0472/11/5/457/htm. Acesso em: 5 nov. 2024.

GITELSON, Anatoly A.; GRITZ, Yuri; MERZLYAK, Mark N. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. **Journal of Plant Physiology**, *[S. l.]*, v. 160, n. 3, p. 271–282, 2003. DOI: 10.1078/0176-1617-00887. Acesso em: 10 mar. 2025.

GOEL, N. S.; QIN, W. Influences of canopy architecture on relationships between various vegetation indices and LAI and Fpar: A computer simulation. **Remote Sensing Reviews**, *[S. l.]*, v. 10, n. 4, p. 309–347, 1994. DOI: 10.1080/02757259409532252. Disponível em: https://www.tandfonline.com/doi/abs/10.1080/02757259409532252. Acesso em: 10 mar. 2025.

GORELICK, Noel; HANCHER, Matt; DIXON, Mike; ILYUSHCHENKO, Simon; THAU, David; MOORE, Rebecca. Google Earth Engine: Planetary-scale geospatial analysis for everyone. **Remote Sensing of Environment**, *[S. l.]*, v. 202, p. 18–27, 2017. DOI: 10.1016/J.RSE.2017.06.031. Disponível em: https://www.sciencedirect.com/science/article/pii/S0034425717302900?via%3Dihub. Acesso em: 15 out. 2025.

GUYON, Isabelle; WESTON, Jason; BARNHILL, Stephen; VAPNIK, Vladimir. Gene selection for cancer classification using support vector machines. **Machine Learning**, *[S. l.]*, v. 46, n. 1–3, p. 389–422, 2002. DOI: 10.1023/A:1012487302797/METRICS. Disponível em: https://link.springer.com/article/10.1023/A:1012487302797. Acesso em: 13 nov. 2025.

HUETE, A.; DIDAN, K.; MIURA, T.; RODRIGUEZ, E. P.; GAO, X.; FERREIRA, L. G. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. **Remote Sensing of Environment**, *[S. l.]*, v. 83, n. 1–2, p. 195–213, 2002. DOI: 10.1016/S0034-4257(02)00096-2. Acesso em: 5 nov. 2024.

HUETE, A. R. A soil-adjusted vegetation index (SAVI). **Remote Sensing of Environment**, *[S. l.]*, v. 25, n. 3, p. 295–309, 1988. DOI: 10.1016/0034-4257(88)90106-X. Acesso em: 3 fev. 2025.

JOHNSON, L. F.; ROCZEN, D. E.; YOUKHANA, S. K.; NEMANI, R. R.; BOSCH, D. F. Mapping vineyard leaf area with multispectral satellite imagery. **Computers and Electronics in Agriculture**, *[S. l.]*, v. 38, n. 1, p. 33–44, 2003. DOI: 10.1016/S0168-1699(02)00106-0. Acesso em: 2 fev. 2025.

KASIMATI, Aikaterini; ESPEJO-GARCIA, Borja; VALI, Eleanna; MALOUNAS, Ioannis; FOUNTAS, Spyros. Investigating a Selection of Methods for the Prediction of Total Soluble Solids Among Wine Grape Quality Characteristics Using Normalized Difference Vegetation Index Data From Proximal and Remote Sensing. **Frontiers in Plant Science**, *[S. l.]*, v. 12, 2021. a. DOI: 10.3389/fpls.2021.683078.

KASIMATI, Aikaterini; ESPEJO-GARCIA, Borja; VALI, Eleanna; MALOUNAS, Ioannis; FOUNTAS, Spyros. Investigating a Selection of Methods for the Prediction of Total Soluble Solids Among Wine Grape Quality Characteristics Using Normalized Difference Vegetation Index Data From Proximal and Remote Sensing. **Frontiers in Plant Science**, *[S. l.]*, v. 12, 2021. b. DOI: 10.3389/fpls.2021.683078.

KATAOKA, T.; KANEKO, T.; OKAMOTO, H.; HATA, S. Crop growth estimation system using machine vision. **IEEE/ASME International Conference on Advanced Intelligent Mechatronics, AIM**, *[S. l.]*, v. 2, p. 1079–1083, 2003. DOI: 10.1109/AIM.2003.1225492. Acesso em: 3 fev. 2025.

KAUTH, R. J.; THOMAS, G. S. Purdue e-Pubs The Tasselled Cap-A Graphic Description of the Spectral-Temporal Development of Agricultural Crops as Seen by LANDSAT. *[S. l.]*, 1976. Disponível em: http://docs.lib.purdue.edu/lars_symphttp://docs.lib.purdue.edu/lars_symp/159. Acesso em: 24 out. 2025.

KODUR, S. Effects of juice pH and potassium on juice and wine quality, and regulation of potassium in grapevines through rootstocks (Vitis): a short review. **VITIS - Journal of Grapevine Research**, *[S. l.]*, v. 50, n. 1, p. 1–1, 2011. DOI: 10.5073/VITIS.2011.50.1-6. Disponível em: https://ojs.openagrar.de/index.php/VITIS/article/view/4052. Acesso em: 26 jun. 2026.

LAMB, David William. **Making sense of vineyard variability in Australia**. [s.l: s.n.]. Disponível em: https://www.researchgate.net/publication/275832528.

LOUHAICHI, Mounir; BORMAN, Michael M.; JOHNSON, Douglas E. Spatially located platform and aerial photography for documentation of grazing impacts on wheat. **Geocarto International**, *[S. l.]*, v. 16, n. 1, p. 65–70, 2001. a. DOI: 10.1080/10106040108542184. Disponível em: https://www.researchgate.net/publication/228906661_Spatially_Located_Platform_and_Aerial_Photography_for_Documentation_of_Grazing_Impacts_on_Wheat. Acesso em: 3 fev. 2025.

LOUHAICHI, Mounir; BORMAN, Michael M.; JOHNSON, Douglas E. Spatially located platform and aerial photography for documentation of grazing impacts on wheat. **Geocarto International**, *[S.*

l.], v. 16, n. 1, p. 65–70, 2001. b. DOI: 10.1080/10106040108542184. Disponível em: https://www.researchgate.net/publication/228906661_Spatially_Located_Platform_and_Aerial_Photography_for_Documentation_of_Grazing_Impacts_on_Wheat. Acesso em: 10 mar. 2025.

MATESE, Alessandro; DI GENNARO, Salvatore Filippo. **Technology in precision viticulture: A state of the art review**. **International Journal of Wine Research**Dove Medical Press Ltd, , 2015. DOI: 10.2147/IJWR.S69405.

MCBRATNEY, A. B.; WEBSTER, R. Choosing functions for semi-variograms of soil properties and fitting them to sampling estimates. **Journal of Soil Science**, *[S. l.]*, v. 37, n. 4, p. 617–639, 1986. DOI: 10.1111/J.1365-2389.1986.TB00392.X;SUBPAGE:STRING:ABSTRACT;REQUESTEDJOURNAL:JOURNAL:13652389A;WGROUP:STRING:PUBLICATION. Disponível em: /doi/pdf/10.1111/j.1365-2389.1986.tb00392.x. Acesso em: 22 nov. 2025.

MELO, Derlei D.; CUNHA, Isabella A.; AMARAL, Lucas R. Precision Zones: An Open-Source QGIS Plugin for Management-Zone Segmentation in Precision Agriculture. **AgriEngineering 2025, Vol. 7, Page 420**, *[S. l.]*, v. 7, n. 12, p. 420, 2025. DOI: 10.3390/AGRIENGINEERING7120420. Disponível em: https://www.mdpi.com/2624-7402/7/12/420/htm. Acesso em: 16 maio. 2026.

MEYER, G. E.; MEHTA, T.; KOCHER, M. F.; MORTENSEN, D. A.; SAMAL, A. TEXTURAL IMAGING AND DISCRIMINANT ANALYSIS FOR DISTINGUISHINGWEEDS FOR SPOT SPRAYING. **Transactions of the ASAE**, *[S. l.]*, v. 41, n. 4, p. 1189–1197, 1998. DOI: 10.13031/2013.17244. Disponível em: https://elibrary.asabe.org/azdez.asp?JID=3&AID=17244&CID=t1998&v=41&i=4&T=1. Acesso em: 3 fev. 2025.

MOLIN, José Paulo; FAULIN, Gustavo Di Chiacchio. Spatial and temporal variability of soil electrical conductivity related to soil moisture. **Scientia Agricola**, *[S. l.]*, v. 70, n. 1, p. 01–05, 2013. DOI: 10.1590/S0103-90162013000100001. Disponível em: https://www.scielo.br/j/sa/a/6pmd35WdvfqGDsKy8c6tq4R/?lang=en. Acesso em: 27 jun. 2026.

NETO, Joao Camargo. A combined statistical-soft computing approach for classification and mapping weed species in minimum -tillage systems. **ETD collection for University of Nebraska-Lincoln**, *[S. l.]*, 2004. Disponível em: https://digitalcommons.unl.edu/dissertations/AAI3147135. Acesso em: 3 fev. 2025.

OIV. **State of the World Wine Sector in 2025**. [s.l.] : International Organisation of Vine and Wine, 2026. Disponível em: www.oiv.int.

PARKER, A. K.; DE CORTÁZAR-ATAURI, I. G.; VAN LEEUWEN, C.; CHUINE, I. General phenological model to characterise the timing of flowering and veraison of Vitis vinifera L. **Australian Journal of Grape and Wine Research**, *[S. l.]*, v. 17, n. 2, p. 206–216, 2011. DOI: 10.1111/J.1755-0238.2011.00140.X;WGROUP:STRING:PUBLICATION. Disponível em: /doi/pdf/10.1111/j.1755-0238.2011.00140.x. Acesso em: 26 jun. 2026.

PENSO, Gener Augusto; CITADIN, Idemir; MARO, Luana Aparecida Castilho; SCARIOTTO, Silvia; CHICOKI, Alyne; AMBROSIO, Rodrigo. Evolução da maturação de uvas finas (Vitis vinifera L.), no município de Dois Vizinhos - Paraná. **Semina: Ciências Agrárias**, *[S. l.]*, v. 35, n. 6,

p. 3085–3098, 2014. DOI: 10.5433/1679-0359.2014V35N6P3085. Disponível em: https://ojs.uel.br/revistas/uel/index.php/semagrarias/article/view/16269. Acesso em: 4 jul. 2026.

PIX4D SA. **Pix4Dmapper**. 2025. Disponível em: https://www.pix4d.com/. Acesso em: 27 jul. 2025.

PRIORI, S.; MARTINI, E.; ANDRENELLI, M. C.; MAGINI, S.; AGNELLI, A. E.; BUCELLI, P.; BIAGI, M.; PELLEGRINI, S.; COSTANTINI, E. A. C. Improving Wine Quality through Harvest Zoning and Combined Use of Remote and Soil Proximal Sensing. **Soil Science Society of America Journal**, *[S. l.]*, v. 77, n. 4, p. 1338–1348, 2013. DOI: 10.2136/sssaj2012.0376.

REGINA, M. de A.; FRÁGUAS, J. C.; ALVARENGA, A. A.; SOUZA, C. R. De; AMORIM, D. A. De; MOTA, R. V. Da; FÁVERO, A. C. Implantação e manejo do vinhedo para produção de vinhos de qualidade. *[S. l.]*, 2019. Disponível em: http://www.alice.cnptia.embrapa.br/handle/doc/541532. Acesso em: 4 jul. 2026.

REYES, Francesco; CASA, Raffaele; TOLOMIO, Massimo; DALPONTE, Michele; MZID, Nada. Soil properties zoning of agricultural fields based on a climate-driven spatial clustering of remote sensing time series data. **European Journal of Agronomy**, *[S. l.]*, v. 150, p. 126930, 2023. DOI: 10.1016/J.EJA.2023.126930. Disponível em: https://www.sciencedirect.com/science/article/pii/S1161030123001983?via%3Dihub. Acesso em: 27 jun. 2026.

RIENTH, Markus; TORREGROSA, Laurent; SARAH, Gautier; ARDISSON, Morgane; BRILLOUET, Jean Marc; ROMIEU, Charles. Temperature desynchronizes sugar and organic acid metabolism in ripening grapevine fruits and remodels their transcriptome. **BMC Plant Biology 2016 16:1**, *[S. l.]*, v. 16, n. 1, p. 164-, 2016. DOI: 10.1186/S12870-016-0850-0. Disponível em: https://link.springer.com/article/10.1186/s12870-016-0850-0. Acesso em: 26 jun. 2026.

ROSENQVIST, Ake; SHIMADA, Masanobu; ITO, Norimasa; WATANABE, Manabu. ALOS PALSAR: A pathfinder mission for global-scale monitoring of the environment. **IEEE Transactions on Geoscience and Remote Sensing**, *[S. l.]*, v. 45, n. 11, p. 3307–3316, 2007. DOI: 10.1109/TGRS.2007.901027. Disponível em: https://www.researchgate.net/publication/3205578_ALOS_PALSAR_A_Pathfinder_mission_for_global-scale_monitoring_of_the_environment. Acesso em: 20 out. 2025.

ROUSE, J. W. ,. Jr.; HAAS, R. H.; SCHELL, J. A.; DEERING, D. W. Monitoring vegetation systems in the Great Plains with ERTS. **NASA. Goddard Space Flight Center 3d ERTS-1 Symp., Vol. 1, Sect. A**, *[S. l.]*, 1974. . Acesso em: 2 fev. 2025.

SANTOS, H. G. Dos et al. Sistema Brasileiro de Classificação de Solos. *[S. l.]*, 2025. Disponível em: http://www.infoteca.cnptia.embrapa.br/handle/doc/1176834. Acesso em: 29 set. 2025.

SHINE, Philip; MURPHY, Michael D. Over 20 Years of Machine Learning Applications on Dairy Farms: A Comprehensive Mapping Study. **Sensors (Basel, Switzerland)**, *[S. l.]*, v. 22, n. 1, p. 52, 2021. DOI: 10.3390/S22010052. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC8747441/. Acesso em: 14 jun. 2025.

SMART, R. E.; DICK, Joy K.; GRAVETT, Isabella M.; FISHER, B. M. Canopy Management to Improve Grape Yield and Wine Quality - Principles and Practices. **South African Journal of**

Enology and Viticulture, *[S. l.]*, v. 11, n. 1, p. 3–17, 1990. DOI: 10.21548/11-1-2232. Disponível em: https://www.journals.ac.za/sajev/article/view/2232. Acesso em: 26 jun. 2026.

SWEETMAN, C.; SADRAS, V. O.; HANCOCK, R. D.; SOOLE, K. L.; FORD, C. M. Metabolic effects of elevated temperature on organic acid degradation in ripening Vitis vinifera fruit. **Journal of Experimental Botany**, *[S. l.]*, v. 65, n. 20, p. 5975–5988, 2014. DOI: 10.1093/JXB/ERU343. Disponível em: https://dx.doi.org/10.1093/jxb/eru343. Acesso em: 26 jun. 2026.

TUCKER, Compton J. Red and photographic infrared linear combinations for monitoring vegetation. **Remote Sensing of Environment**, *[S. l.]*, v. 8, n. 2, p. 127–150, 1979. a. DOI: 10.1016/0034-4257(79)90013-0. Acesso em: 3 fev. 2025.

TUCKER, Compton J. Red and photographic infrared linear combinations for monitoring vegetation. **Remote Sensing of Environment**, *[S. l.]*, v. 8, n. 2, p. 127–150, 1979. b. DOI: 10.1016/0034-4257(79)90013-0. Acesso em: 10 mar. 2025.

WOEBBECKE, D. M.; MEYER, G. E.; VON BARGEN, K.; MORTENSEN, D. A. Color Indices for Weed Identification Under Various Soil, Residue, and Lighting Conditions. **Transactions of the ASAE**, *[S. l.]*, v. 38, n. 1, p. 259–269, 1995. DOI: 10.13031/2013.27838. Disponível em: https://elibrary.asabe.org/azdez.asp?JID=3&AID=27838&CID=t1995&v=38&i=1&T=1. Acesso em: 3 fev. 2025.

YUAN, Yifan; SHI, Bo; YOST, Russell; LIU, Xiaojun; TIAN, Yongchao; ZHU, Yan; CAO, Weixing; CAO, Qiang. Optimization of Management Zone Delineation for Precision Crop Management in an Intensive Farming System. **Plants**, *[S. l.]*, v. 11, n. 19, p. 2611, 2022. DOI: 10.3390/PLANTS11192611. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC9573654/. Acesso em: 22 nov. 2025.

ZARCO-TEJADA, P. J.; BERJÓN, A.; LÓPEZ-LOZANO, R.; MILLER, J. R.; MARTÍN, P.; CACHORRO, V.; GONZÁLEZ, M. R.; DE FRUTOS, A. Assessing vineyard condition with hyperspectral indices: Leaf and canopy reflectance simulation in a row-structured discontinuous canopy. **Remote Sensing of Environment**, *[S. l.]*, v. 99, n. 3, p. 271–287, 2005. DOI: 10.1016/J.RSE.2005.09.002. Disponível em: https://www.sciencedirect.com/science/article/abs/pii/S0034425705002865. Acesso em: 27 jun. 2026.

TERCEIRA PARTE

horticulturae

Article

Prediction of Technological Maturity of Grapevines Under a Double Pruning System Using Data Fusion and Machine Learning

Octavio Pereira da Costa [1], Fabiano Luis de Sousa Ramos Filho [1], Bernado Siqueira Costa Barbosa [1], Rai Fernandes Queiroz Alves [1], Girley Valdes Fernandez [1], Matheus de Melo Amorim [2], Caio Canestri Ribeiro [2], Adão Felipe dos Santos [1,*], Rafael Pio [1] and Pedro Maranha Peche [1,*]

Academic Editor: Yang Liu

* Correspondence: adao.felipe@ufla.br (A.F.d.S.); pedro.peche@ufla.br (P.M.P.)

Received: 19 May 2026
Revised: 22 June 2026
Accepted: 26 June 2026
Published: 7 July 2026

[1] Department of Agriculture, Federal University of Lavras, Lavras 37200-202, Brazil; octavio.costa2@estudante.ufla.br (O.P.d.C.); fabiano.filho1@estudante.ufla.br (F.L.d.S.R.F.); bernardo.barbosa1@estudante.ufla.br (B.S.C.B.); rai.alves2@estudante.ufla.br (R.F.Q.A.); girley.fernandez1@estudante.ufla.br (G.V.F.); rafael.pio@ufla.br (R.P.)
[2] Crop Production, Federal University of Tocantins, Gurupi 77402-970, Brazil; matheusmelogpi21@gmail.com (M.d.M.A.); caio.ribeiro2@estudante.ufla.br (C.C.R.)

Abstract

The production of "winter wines" in south-eastern Brazil, enabled by the double pruning technique, requires precise assessment of grape technological maturity to ensure highquality outputs. However, conventional monitoring approaches are destructive, laborintensive, and limited in their ability to capture vineyard spatial variability. This study aimed to develop and validate a non-destructive predictive framework for Soluble Solids (°Brix) and Titratable Acidity (TA) by integrating spatial remote sensing data with temporal agrometeorological information. Multispectral imagery was acquired via an unmanned aerial vehicle in a vineyard cultivated with Sauvignon Blanc and Syrah, from which vegetation indices were derived and combined with Growing Degree-Days to train machine learning models, including Random Forest, Multilayer Perceptron, and XGBoost. The incorporation of agrometeorological data significantly improved predictive performance compared to models based solely on vegetation indices. Among the tested algorithms, XGBoost achieved the highest accuracy, with coefficients of determination of 0.89 for °Brix and 0.77 for TA, achieved by XGBoost on an independent hold-out test set. Model interpretability analysis indicated that Growing Degree-Days and cultivar were the primary drivers of maturation dynamics, while vegetation indices refined predictions by accounting for spatial variability in plant vigor. Overall, the proposed approach represents a promising proof-of-concept framework for non-destructive maturity monitoring in precision viticulture, supporting improved monitoring of grape maturation. However, multi-season validation across diverse vineyard conditions is required to confirm its generalizability and support its application as a routine decision-support tool.

Keywords: remote sensing; vegetation indices; precision viticulture; spatial variability; growing degree-days; UAV

1. Introduction

Global viticulture has witnessed the emergence of new production regions that employ innovative management techniques to achieve high-quality wines in climates previously considered marginal.. A notable example is the production of "winter-harvest wines" in southeastern Brazil, made possible by the double pruning technique. This practice inverts the grapevine's harvest cycle to the dry winter period, promoting the concentration of sugars and phenolic compounds in the berries [1]. The success of this and other viticultural practices fundamentally depends on the ability to precisely determine the optimal harvest time, a point defined by the concept of technological maturity. This stage represents the ideal balance between sugars, measured as Soluble Solids (°Brix), and Titratable Acidity (TA), parameters that directly govern the wine's alcoholic potential, stability, and sensory profile [2].

Accurately determining the harvest point is a decisive factor in viticulture, as the quality of the grapes at the time of picking directly influences the potential of the final wine [3]. Traditionally, monitoring relies on manual berry sampling and laboratory analysis, a destructive, time-consuming, and spatially limited approach that fails to capture within-vineyard variability in maturity, potentially leading to suboptimal harvest decisions [4]. To overcome these limitations, precision viticulture has emerged as a promising field employing remote sensing technologies for non-destructive, large-scale monitoring. Unmanned Aerial Vehicles (UAVs) equipped with multispectral sensors have become a powerful tool for assessing the spatial variability of the vegetative canopy [5], from which Vegetation Indices (VIs) are calculated as proxies for plant vigor, nutritional status, and water status [6]. However, VIs correlate inconsistently with grape quality parameters [7], as canopy vigor does not always directly translate to fruit maturation due to complex mediating factors such as water balance and source–sink relationships.

This reveals a fundamental gap: VIs provide a spatial snapshot of plant condition but lack the temporal dimension needed to capture phenological development. Studies using VIs as sole predictors have reported limited accuracy, with maximum R^2 values of 0.52 [8] and 0.61 [9] for Brix prediction, confirming that spatial vigor information alone cannot represent the cumulative, thermally driven nature of ripening. Growing Degree-Days (GDD), a well-established agrometeorological indicator that quantifies the accumulated thermal energy required for phenological stage transitions [10], offers a complementary temporal dimension. GDD-based approaches combined with machine learning have proven effective for predicting growth and yield in crops such as rice and sweet potato [11,12]; however, the synergistic integration of spatial (VIs) and temporal (GDD) data for grapevine technological maturity prediction remains a largely unexplored area.

The central hypothesis of this work is that fusing spatial (VIs) and temporal (GDD) predictors through machine learning models can overcome the limitations of individual approaches, generating more accurate and robust estimates of technological maturity. Therefore, the objective of this study was to develop and assess machine learning models for the non-destructive prediction of Soluble Solids (°Brix) and Titratable Acidity in 'Sauvignon Blanc' and 'Syrah' grapes grown under a double pruning regime. The specific objectives were: (i) to investigate whether the inclusion of GDD as a predictor variable significantly improves the accuracy of VI-based models; (ii) to identify the most influential VIs for predicting grape maturity; and (iii) to determine the best-performing machine learning algorithm for this task.

2. Materials and Methods

2.1. Study Area

The experiment was conducted during the 2024 winter harvest cycle, which commenced with production pruning in January 2024 and concluded with harvest in the winter (July/August) of 2024. The study area is a commercial *Vitis vinifera* vineyard located in the Macaia district, municipality of Bom Sucesso, Minas Gerais, Brazil. The area is situated at the geographical coordinates $21^\circ 9'12.78''$ S and $44^\circ 53'36.45''$ W, at an average altitude of 952 m (Figure 1). The local topography features an average slope of 13%. According to the Köppen–Geiger classification, the regional climate is classified as Cwa (humid subtropical with a dry winter and hot summer) [13], with an average temperature of the warmest month exceeding 22 $^\circ$C.

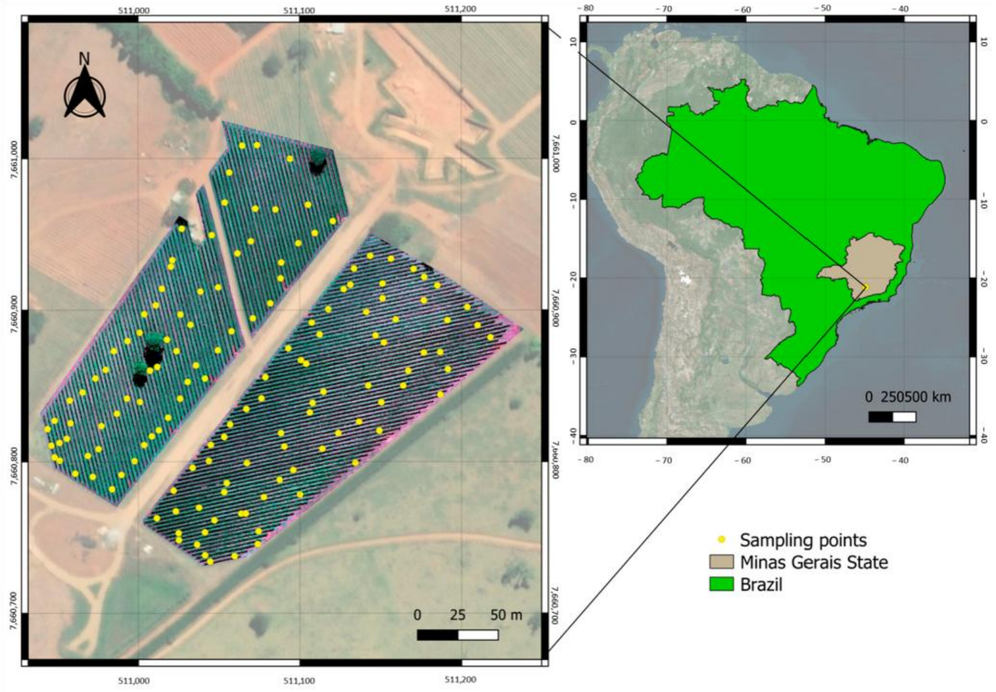


Figure 1. Geographical location of the experimental area in Bom Sucesso (MG, Brazil) and the spatial distribution of the sampling points within the vineyard, covering the 'Sauvignon Blanc' and 'Syrah' cultivars.

2.2. Vineyard Details

The experimental area covers 4.54 hectares, divided into two commercial blocks composed of two *Vitis vinifera* L. cultivars: 'Sauvignon Blanc' (white) and 'Syrah' (red). Both cultivars were grafted onto the 'IAC 766' rootstock.

The vineyard was established in 2021, with the vines being in their third year (third leaf) during the experimental cycle. The vines are trained on a vertical shoot positioning (VSP) trellis system. The blocks feature distinct row orientations: one aligned Northeast–Southwest and the other Northwest–Southeast. Planting spacing is 1 m between vines and 2.6 m between rows, resulting in a density of 3846 vines per hectare. Management for the winter cycle was initiated with production pruning carried out in January 2024, following the double pruning system [1]. The vineyard is equipped with a drip irrigation system for supplemental watering.

2.3. Sampling Design and Field Data Collection

To monitor ripening progression, 145 georeferenced sampling points were defined within the experimental area. Data collection was carried out during the grape ripening phase, commencing after the onset of veraison.

The 145 georeferenced sampling points were distributed across both cultivar blocks in proportion to their respective areas, with 70 points located in the 'Sauvignon Blanc' block and 56 points in the 'Syrah' block. Over six collection dates, this design generated a total dataset of 870 individual observations (145 points × 6 dates). No formal statistical outlier removal procedure was applied to the physicochemical data; samples compromised by berry damage or equipment malfunction during laboratory analysis were excluded on a case-by-case basis and replaced where feasible.

The final analytical dataset comprised 494 observations from 126 georeferenced sampling points over five collection dates. Descriptive statistics for the response variables were: Soluble Solids ranged from 7.6 to 21.7 °Brix (mean = 16.6 ± 2.87 °Brix; CV = 17.3%); Titratable Acidity ranged from 4.1 to 24.7 g/L (mean = 11.15 ± 4.46 g/L; CV = 40.0%).

Collections occurred on six consecutive dates in 2024 (14 June, 18 June, 26 June, 4 July, 10 July, and 17 July), at approximately weekly intervals leading to harvest. At each sampling point and on each date, 20 grape berries were manually harvested. The sampling was stratified, with berries collected from three distinct vertical positions within the grape clusters (top, middle, and bottom) and evenly distributed across both sides of the vertical shoot-positioned (VSP) trellis system. Samples were immediately placed in labeled plastic bags and transported under refrigeration.

2.4. Physicochemical Analysis (Technological Maturity)

Samples were processed at the Fruit Science Laboratory of the Federal University of Lavras (UFLA). The 20 berries from each point were homogenized and gently crushed manually to prevent seed damage. The resulting juice (must) was filtered and subjected to analysis of the following physicochemical parameters:

- Soluble Solids (SS): SS content was determined using a calibrated digital refractometer, with results expressed in °Brix.
- Titratable Acidity (TA): Total acidity was quantified following the standardized Adolf Lutz protocol [14]. The results were subsequently converted and expressed as tartaric acid equivalents (g/L) to ensure methodological consistency and comparability.

2.5. Multispectral Image Acquisition, Processing and Analysis

Multispectral image acquisition was performed on six dates throughout the ripening cycle, coinciding with the field sampling days. All flights were conducted between 11:00 AM and 1:00 PM (local time, UTC-3) under clear-sky conditions to ensure consistent solar illumination. An Unmanned Aerial Vehicle (UAV), model Mavic 3M (DJI, Shenzhen, China), was used. It was equipped with a Real-Time Kinematic (RTK) georeferencing system, aided by a D-RTK 2 base station (DJI, Shenzhen, China), to ensure high positional accuracy. The onboard multispectral sensor features four spectral bands: Green (G) (560 ± 16 nm), Red (R) (650 ± 16 nm), RedEdge (RE) (730 ± 16 nm), and Near-Infrared (NIR) (860 ± 26 nm).

Flight missions were planned and executed using DJI Pilot 2 software at an altitude of 80 m. Flight parameters included a speed of 5 m/s and 80% frontal and lateral overlap to ensure proper orthomosaic generation. Photogrammetric processing was performed using Pix4Dmapper v. 4.8.4 (Pix4D SA 2025). The images from each flight were imported into individual projects using the RTK metadata. Processing followed the 'Ag Multispectral' template, which includes image alignment, dense point cloud generation, and, finally, orthomosaic generation.

Radiometric calibration was applied during processing to convert Digital Number (DN) values to surface reflectance. To achieve this, images of a standard reflectance panel, captured on the ground before each flight, were imported and used by the software's calibration tool, ensuring data comparability across the different dates. At the end of processing for each date, multispectral reflectance orthomosaics were generated, containing all four bands (G, R, RE, NIR) with a final spatial resolution of 4.2 cm. The files were exported in GeoTIFF format using the WGS 84/UTM Zone 23S coordinate system.

The calibration panel used was the MicaSense Multispectral Calibration Reflectance

Panel with nominal per-band reflectance values of 48.9% (Green, 560 nm), 48.9% (Red, 650 nm), 48.8% (RedEdge, 730 nm), and 48.6% (NIR, 860 nm) as specified in the manufacturer's calibration reference (RP06-2051010-OB, MicaSense). Calibration certificates are available upon request from the corresponding authors. Regarding per-flight meteorological conditions, all flights were conducted under clear-sky conditions. All post-processing and image analysis steps were conducted in the R software environment v. 4.1.1 [15], supported by the 'terra' v. 1.8.70 [16], 'sf' v. 1.0.21 [17], and 'FIELDimageR' v. 0.6.2 [18] packages. First, to ensure perfect pixel-to-pixel overlap between collection dates, temporal co-registration of the six images was performed, aligning all orthomosaics to the image from the first flight date. Next, vegetation segmentation was carried out to remove soil pixels. A supervised classification method using the Random Forest (RF) algorithm was employed. Training polygons were manually digitized over representative 'canopy' and 'soil' areas using tools from the FIELDimageR v. 0.6.2 package [18]. From these, 400 random sample points per class were used to train the classifier using the superClass function from the 'RStoolbox' 1.0.2.2 package [19]. The resulting classification map was then used as a mask to extract exclusively vegetation pixels for subsequent analyses.

2.6. Feature Selection

The feature selection pipeline consisted of two sequential stages, aiming to obtain a subset of predictors that were both relevant (RFE) and independent (VIF). The entire process was conducted in the Python language v 3.9 using the scikit-learn library. Given the large number of existing Vegetation Indices (VIs), 20 VIs were initially calculated. The first selection stage employed Recursive Feature Elimination (RFE) [20]. RFE is a wrapper technique that uses an external estimator, in this case, the Random Forest
Regressor [21], to iteratively rank predictor importance. The estimator was configured with 100 trees (*n_estimators* = 100) and a fixed random state (*random_state* = 42) for reproducibility.

The RFE was executed independently for the target variables Soluble Solids (°Brix) and Titratable Acidity (g/L), following the removal of samples with missing target data and median imputation of missing predictors. The algorithm was set to select the subset containing the 10 most important features for each target, generating two preliminary lists of VIs.

To ensure model stability and remove predictor redundancy, the two 10-VI subsets generated by RFE were subjected to Variance Inflation Factor (VIF) analysis. The VIF was calculated for all variables in each subset. In an iterative process, the variable with the highest VIF value was identified; if this value was greater than 5, the feature was removed. The VIF was then recalculated for the remaining variables. This process was repeated until all VIs in the subset exhibited a VIF value below 5. This resulted in a final set of low-collinearity VIs for each target variable, which was then effectively used in the modeling stage.

2.7. Selected Vegetation Indices for Modeling

At the conclusion of the feature selection pipeline (RFE followed by VIF), the final sets of predictors for each target variable were defined. The process resulted in the selection of five Vegetation Indices (VIs) for the Soluble Solids (°Brix) model and five VIs for the Titratable Acidity (g/L) model. These two predictor sets were effectively used as inputs for the modeling stage. The details, mathematical equations, and original bibliographical references for each of these selected VIs are presented in Table 1.

Table 1. Vegetation Indices (VIs) selected as predic.tors for the Soluble Solids (°Brix) and Titratable Acidity (TA) models, their mathematical equations, and original references.

Index	Equation	Reference
Burn Area Index (BAI)	$\frac{1}{(0.1-Red)^2+(0.06-NIR)^2}$	[22]
Spectral Feature Depth Vegetation (SFDVI)	$\frac{Green+NIR}{2}-\frac{Red+RedEdge}{2}$	[23]
Green Difference Vegetation Index (GDVI)	$NIR-Green$	[24]
Chlorophyll Vegetation Index (CVI)	$NIR\times\frac{Red}{Green^2}$	[25]
Transformed Soil-Adjusted Vegetation Index (TSAVI)	$\frac{1.2\times(NIR-1.2\times Red-0.04)}{1.2\times NIR+Red-1.2\times0.04+0.08\times(1+1.2^2)}$	[26]
Plant Senescence Reflectance Index (PSRI)	$\frac{Red-Green}{RedEdge}$	[27]
Transformed chlorophyll absorption in reflectance (TCARI)	$3\times((RedEdge-Red)-0.2)\times(RedEdge-Green)\times(\frac{RedEdge}{Red})$	[28]

2.8. Growing Degree Days (GDD)

To quantify thermal accumulation during the experimental cycle, Growing DegreeDays (GDD) were calculated. Daily maximum (Tmax) and minimum (Tmin) air temperature data were obtained for the experiment's location from the NASA POWER platform [29].

Daily GDD was calculated using the methodology proposed by Villa Nova [10], which accounts for various scenarios where minimum temperatures may fall below the base temperature (Tb). A Tb of 10 °C was adopted, which is the standard for grapevine (*Vitis vinifera*). The equations used were:

$$\text{If Tm} > \text{Tb}: GOD = \frac{(TM+Tm)}{2} - Tb \qquad (1)$$

$$\text{If Tm} < \text{Tb} < \text{TM}: GOD = \frac{(TM+Tb)^2}{2\times(TM-Tm)} \qquad (2)$$

$$\text{If TM} < \text{Tb}: GOD = 0 \qquad (3)$$

where GDD represents the accumulated Growing Degree-Days (degrees °C × day) from the production pruning date to the respective field collection date; Tmax and Tmin are the daily maximum and minimum air temperatures (degrees °C), respectively; Tb is the base temperature, set at 10 degrees C for *Vitis vinifera* [10]; and n represents each individual day in the accumulation period.

GDD was accumulated daily, with the summation starting from the production pruning date. The cumulative GDD value from this start date until each respective field sampling date was then used as a predictor variable in the machine learning models.

2.9. Models

Three machine learning algorithms were selected to represent methodologically distinct paradigms: Random Forest (RF), a bagging ensemble of decision trees recognized for its robustness and built-in feature importance estimation [21]; XGBoost (XGB), a gradient-boosting framework with demonstrated superiority on structured tabular data; and Multi-layer Perceptron (MLP), a feedforward neural network capable of capturing complex non-linear relationships. This selection enables cross-paradigm benchmarking while covering the principal algorithmic approaches applied in precision agriculture. The categorical variable 'Cultivar' (Sauvignon Blanc/Syrah) was binaryencoded (0 = Sauvignon Blanc; 1 = Syrah) prior to model training. Feature standardization (mean = 0; standard deviation = 1) using StandardScaler was applied exclusively to the continuous predictors (the five selected VIs and, in the data fusion scenario, GDD); the binary cultivar encoding was excluded from standardization to preserve its categorical interpretation.

To predict the maturity variables (°Brix and Titratable Acidity), the optimized VI subsets were used as input for three machine learning algorithms: Random Forest (RF), XGBoost (XGB), and Multi-layer Perceptron (MLP). The workflow followed three stages: data preparation, hyperparameter optimization, and final evaluation.

First, the dataset was split into two subsets: 70% for training and 30% for testing (hold-out). The split was stratified by the 'Cultivar' variable to ensure representativeness in both sets. Next, the predictors in the training set were standardized (mean: 0, standard deviation: 1) using the StandardScaler. The scaling parameters derived from the training set were then applied to the test set, preventing data leakage.

To find the optimal hyperparameter combination for each algorithm, GridSearchCV was employed, combined with 5-fold cross-validation, applied exclusively to the training set (70%). Table 2 details the hyperparameter search space explored for each algorithm.

Table 2. Hyperparameter search space (optimization grid) explored via Grid Search for the Random Forest (RF), XGBoost (XGB), and Multi-layer Perceptron (MLP) algorithms.

Model	Hyperparameter	Values
Random Forest (RF)	n_estimators	100, 200, 300
	max_depth min_samples_split	None, 10, 20 2, 5
	min_samples_leaf	1, 2
	max_feautures	'sqrt', 'log2'
XGBoosting (XGB)	n_estimators	100, 200, 300
	learning_rate max_depth	0.05, 0.1 3, 5
	subsample	0.7, 1
	colsample_bytree	0.7, 1
Multi-layer Perceptron (MLP)	Hidden_layers_sizes	(50,), (100,), (50,50)
	Activation Alpha	'relu', 'tanh' 0.0001, 0.001

Learning_rate	'adaptative'

For the MLP algorithm, deeper architectures (more than two hidden layers) were not evaluated, as the limited dataset size (~870 observations) would render such configurations highly susceptible to overfitting; all selected MLP configurations exhibited stable convergence across cross-validation folds, as confirmed by monotonic training loss reduction. It should be noted that the stratified hold-out split was applied at the observation level (n = 870), meaning that the same georeferenced sampling location may appear in both the training and test sets at different collection dates. While temporal stratification across six phenologically distinct collection events partially mitigates this concern, the potential influence of spatial autocorrelation on the reported generalization performance is a recognized limitation. A spatially blocked cross-validation (GroupKFold, k = 5, grouped by sampling point identity) was performed and yielded: $R^2 = 0.865 \pm 0.017$ for °Brix and $R^2 = 0.819 \pm 0.042$ for TA (RMSE = 1.049 ± 0.042 °Brix and 1.881 ± 0.259 g/L, respectively). These results confirm that model performance is maintained under spatial blocking, indicating that the reported generalization capacity is not substantially inflated by spatial autocorrelation.

After identifying the best hyperparameters, the final model for each algorithm was re-trained on the entire training set (70%). The predictive performance of the optimized models was then definitively evaluated on the test set (30%), ensuring an unbiased estimate of their generalization capacity on unseen data.

2.10. Evaluation Metrics

The predictive performance of the optimized models was finally quantified on the test (hold-out, 30%) set, using the Coefficient of Determination (R^2, Equation (4)) and Root Mean Square Error (RMSE, Equation (5)) metrics.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (\hat{y}i - yi)^2}{\sum_{i=1}^{n} (yi - \overline{y})^2} \quad (4)$$

$$\mathbf{RMSE} = \sqrt{\sum_{i=1}^{n} \frac{(\hat{Y}i - Yi)^2}{n}} \quad \mathbf{(5)}$$

where $\hat{y}$ = *prediction value*; $\overline{y}$ = *mean value*; and y = *actual value*.

Furthermore, to understand how the best-performing model made its decisions and which predictor variables were most influential, the SHapley Additive exPlanations (SHAP) methodology was applied [30]. SHAP is a model-agnostic approach, grounded in game theory that calculates the marginal contribution of each predictor variable to each individual prediction. Importantly, SHAP values reflect internal model association patterns—the marginal contribution of each variable to individual predictions—and do not establish causal relationships between predictors and the underlying physiological processes of grape ripening.

The analysis of SHAP values, visualized through beeswarm summary plots, allows for detailed interpretation. This type of plot displays the distribution of SHAP values for each sample and variable, ranking the variables by their global importance. Moreover, the plot uses color to represent the feature's own value (high or low), making it possible to identify not only the magnitude of a predictor's impact on the prediction but also its direction. This analysis, applied to the test set, was crucial for validating whether the relationships learned by the model are agronomically coherent and for building confidence in the results. The complete methodology workflow is summarized in Figure 2.

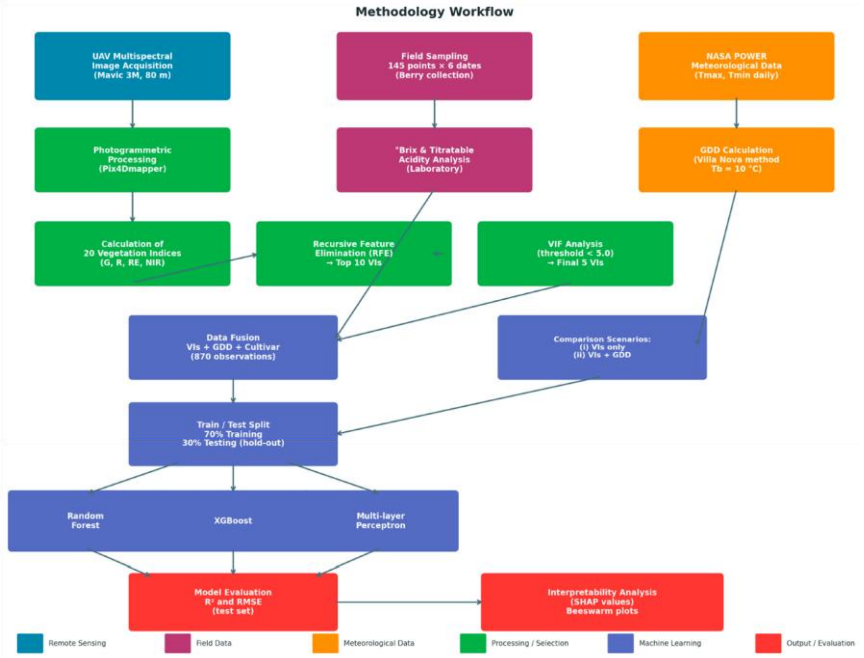


Figure 2. Schematic overview of the methodological workflow, from UAV image acquisition and field sampling through feature selection (RFE + VIF), data fusion (VIs + GDD + Cultivar), model training and evaluation, to SHAP interpretability analysis.

3. Results

3.1. Feature Selection Results

The first stage of the results consisted of filtering the 20 calculated Vegetation Indices (VIs), identifying the most relevant and independent subsets of predictors for modeling. The Recursive Feature Elimination (RFE) process ranked the 10 most promising VIs for each target variable, as detailed in Figure 3. Nine of these VIs were common to both targets (°Brix and Titratable Acidity): 'CVI', 'BAI', 'PSRI', 'GDVI', 'CIRE', 'REDVI', 'NDRE', 'RESR' and 'SFDVI'. The distinctions in the RFE selection were the 'TSAVI' index (selected only for °Brix) and 'TCARI' (selected only for Titratable Acidity).

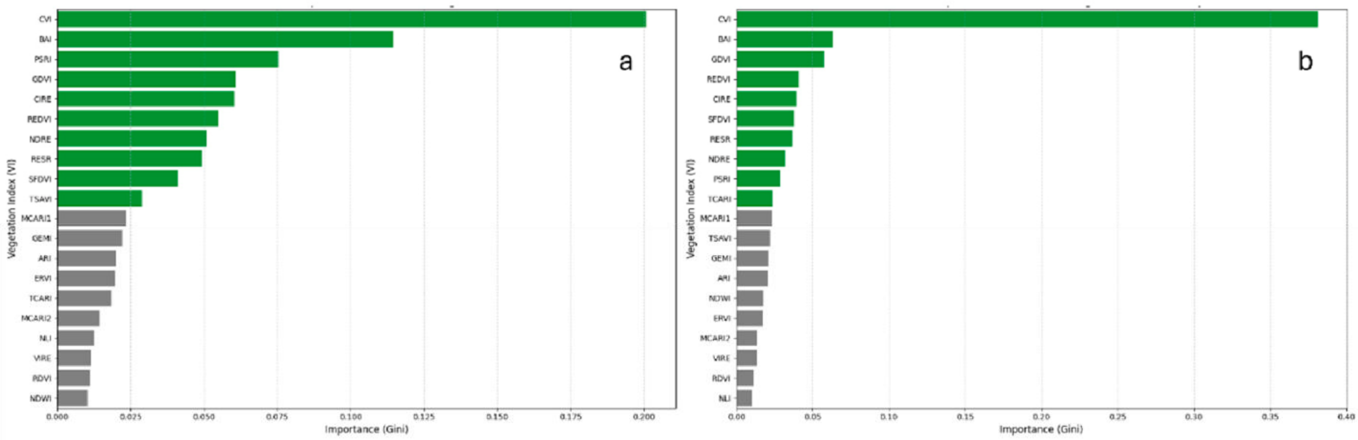


Figure 3. Results of the Recursive Feature Elimination (RFE), ranking the top 10 most important

Vegetation Indices (VIs) for predicting Soluble Solids (°Brix) (**a**) and Titratable Acidity (TA) (**b**). Green bars represent the top 10 selected VIs ranked by importance, while gray bars represent the remaining evaluated VIs.

In the next stage, these two 10-VI sets were subjected to Variance Inflation Factor (VIF) analysis to mitigate multicollinearity. An iterative process was applied to remove the variable with the highest VIF until all remaining predictors had a VIF < 5.0. Figure 4 illustrates the iterative elimination process. The VIF analysis revealed extreme collinearity in both sets, notably in 'RESR' (VIF = ∞) and 'NDRE' (VIF > 1800 for both targets), indicating that their information was almost entirely redundant. For the °Brix model, the 'PSRI' (VIF = 37.05), 'CIRE' (VIF = 24.5), and 'REDVI' (VIF = 9.8) VIs were also removed. For the Titratable Acidity model, 'CVI' (VIF = 27.54), 'CIRE' (VIF = 24.9), and 'REDVI' (VIF = 9.71) were removed.

At the end of the selection pipeline, two final and distinct sets of five VIs were obtained: the set for °Brix was composed of 'SFDVI', 'BAI', 'CVI', 'GDVI', and 'TSAVI', while the set for Titratable Acidity included 'SFDVI', 'BAI', 'GDVI', 'PSRI' and 'TCARI'. Three VIs ('SFDVI', 'BAI', 'GDVI') were selected as fundamental predictors for both variables. However, the models diverged in their specific predictors: 'CVI' and 'TSAVI' were uniquely selected for °Brix, while 'PSRI' and 'TCARI' were uniquely selected for Titratable Acidity. These two final, low-collinearity VI sets were then used for the modeling stage, both in isolation (IV-only models) and in fusion with the agrometeorological variable (GDD), to evaluate the impact of including the temporal predictor.

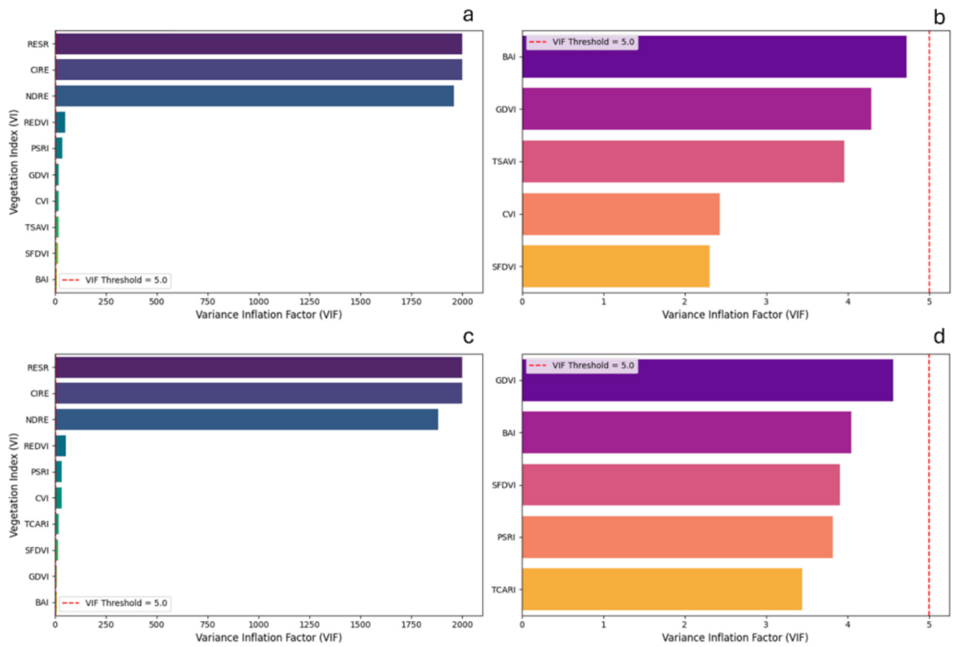


Figure 4. Results of the Variance Inflation Factor (VIF) analysis, illustrating predictor collinearity before (**a**,**c**) and after (**b**,**d**) the iterative elimination process (threshold VIF < 5.0). Panels (**a**,**b**) correspond to the °Brix model, and panels (**c**,**d**) correspond to the Titratable Acidity model.

3.2. Predictive Model Performance for Technological Maturity

The performance of the machine learning models was evaluated in two distinct scenarios: (i) using only the selected VIs as predictor variables and (ii) with the fusion of these predictors with the temporal variable GDD. The results demonstrate the crucial impact of data fusion for the prediction of technological maturity.

Although the Random Forest model fit the training data better for °Brix prediction, none of the three algorithms was able to create a viable predictive model using only the VIs in the test stage. The drastic reduction in accuracy (R^2) and increase in error (RMSE) between the training and test stages unequivocally demonstrate that these predictors, in isolation, do not provide sufficient information for the models to learn the relationship underlying technological maturity progression, leading them to memorize the specific noise of the training set instead of learning generalizable patterns.

For °Brix prediction, although the coefficient of determination on the training set for the XGBoost algorithm was 0.707, the performance on the test set was drastically reduced, with an R^2 of 0.378 (Figure 5b). The behavior of Random Forest was similar. It showed a strong fit to the training data, with an R^2 of 0.813 and RMSE of 1.254. However, in the test stage, the performance was equally unsatisfactory, with the R^2 dropping to 0.477 and the RMSE increasing to 1.957 (Figure 5a). The MLP showed a similar fit to the XGB model on the training data (R^2 of 0.737); however, it registered the lowest performance on the test set (R^2 of 0.303 and RMSE of 2.259) among the models tested (Figure 5c).

Upon integrating the Growing Degree-Days (GDD) variable into the predictor set (Figure 5d–f), the performance of all models for °Brix estimation exhibited significant improvements in generalization capacity, especially regarding accuracy, with error (RMSE) values below 1 °Brix for all models.

To isolate the contribution of each data type, a third scenario was evaluated using only GDD and Cultivar as predictors (GDD-only). For °Brix prediction, the GDD-only XGBoost model achieved $R^2 = 0.889$ on the test set (RMSE = 0.99 °Brix), closely approaching the performance of the full VI + GDD model. For Titratable Acidity, the GDD-only model yielded $R^2 = 0.844$ (RMSE = 1.818 g/L). The date-index proxy analysis—substituting GDD with a simple ordinal temporal index (1–5)—produced $R^2 = 0.815$ for °Brix and $R^2 = 0.821$ for TA, compared to 0.84 and 0.806 for the GDD-based models ($\Delta R^2 = 0.025$ and −0.015, respectively). These results indicate that the performance advantage of GDD over simple temporal ordering is marginal, particularly for TA prediction, and are discussed further in Section 4.

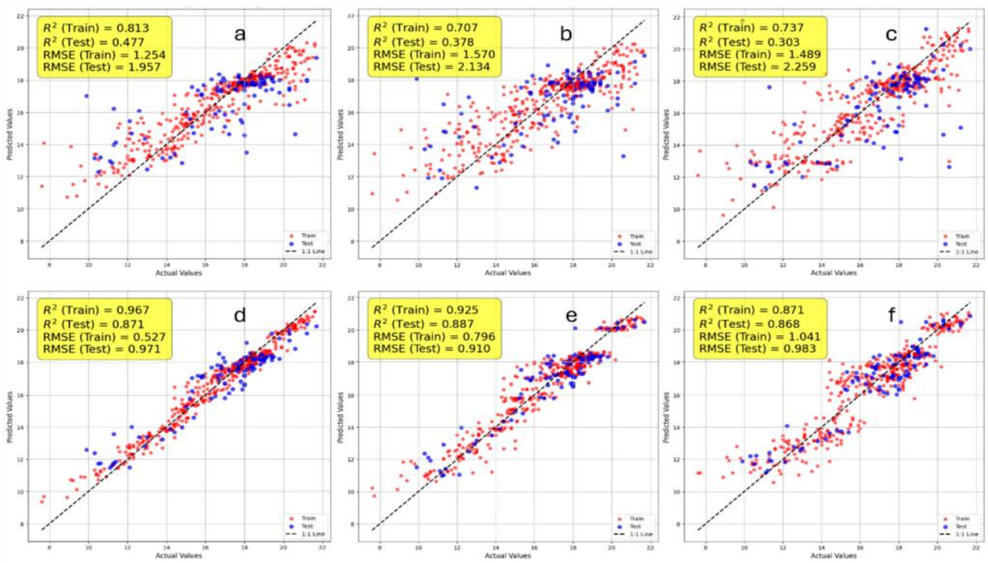


Figure 5. Scatter plots comparing the observed versus predicted ºBrix values for the Random Forest
(**a**,**d**), XGBoost (**b**,**e**), and MLP (**c**,**f**) models. The dashed line represents the 1:1 relationship. The upper panels (**a**–**c**) show results from models trained with VIs only; the lower panels (**d**–**f**) show results from models including GDD.

The XGBoost algorithm stood out with the best overall performance in predicting °Brix (Figure 5e). After being trained, it achieved an R^2 of 0.925 on the training set and, more importantly, maintained the most robust performance on the test set, with an R^2 of 0.887 and the lowest RMSE of 0.910. The Random Forest (RF) model showed a very competitive and close result (Figure 4d), with an R^2 of 0.967 on the training set and 0.871 on the test set, and an RMSE of 0.971. On the other hand, the MLP model (Figure 5f), which had shown inferior performance compared to the others in the previous scenario, also improved significantly when GDD values were added to the model, returning an error value (RMSE = 0.983) very close to that of the tree-based models. Furthermore, the MLP was the model with the lowest overfitting rate, with the training and test results showing a variation in only 0.003 in the coefficient of determination.
Similar to the performance found for °Brix prediction, the machine learning models that used only the VIs and 'Cultivar' as predictors did not produce generalizable estimates for Titratable Acidity (g/L). Although the tree-based algorithms fit the training data well (RF with $R^2 = 0.873$; XGBoost with $R^2 = 0.760$), their coefficients of determination on the test set were low, reaching only $R^2 = 0.451$ for RF and $R^2 = 0.445$ for XGBoost (Figure 6a,b). The MLP model showed the weakest fit on the training set ($R^2 = 0.687$) and the lowest performance on the test set ($R^2 = 0.444$) (Figure 6c). These results demonstrate the ineffectiveness of VIs alone for modeling the degradation of organic acids in grapes.
The inclusion of the GDD variable considerably improved the performance of all models. However, an overfitting trend persisted, with a reduction in accuracy (R^2) observed between the training and test data. This drop was 25.4% for the Random Forest model (0.969 train vs. 0.723 test), 17.1% for XGBoost (0.924 vs. 0.766), and 19.2% for the MLP (0.886 vs. 0.716) (Figure 6d–f). Despite this, all models with GDD achieved test R^2 values greater than 0.70, which did not occur in any model without GDD.
Regarding the error (RMSE) in predicting Titratable Acidity in the GDD-inclusive scenario, the result achieved by the XGBoost model stood out by obtaining the lowest error in the test stage with 1.632 g/L. The other models showed slightly higher errors, at 1.775 g/L for Random Forest and 1.796 g/L for the MLP.

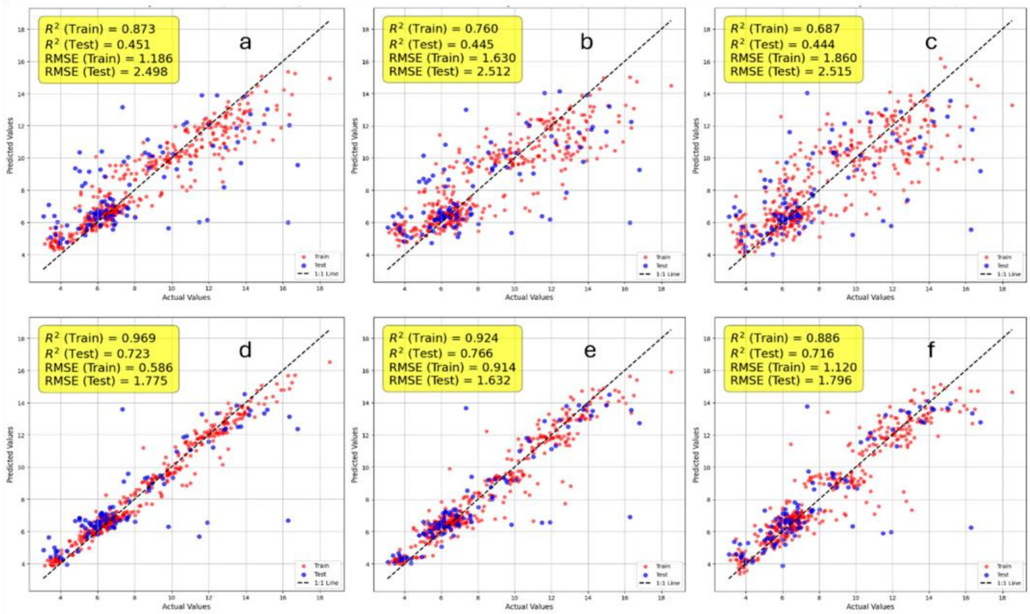


Figure 6. Scatter plots comparing the observed versus predicted Titratable Acidity (g/L) values for the Random Forest (**a**,**d**), XGBoost (**b**,**e**), and MLP (**c**,**f**) models. The dashed line represents the 1:1 relationship. The upper panels (**a**–**c**) show results from models trained with VIs only; the lower panels (**d**–**f**) show results from models including GDD.

The comparative analysis demonstrates that, although all three algorithms became reasonable predictors after data fusion, XGBoost showed clear superiority, with the highest R^2 (0.766) and the lowest RMSE (1.632 g/L). This result validates the study's central hypothesis regarding the importance of including climatic variables, but it also indicates that predicting Titratable Acidity is an inherently more complex task than predicting °Brix.

Statistical significance testing using the Wilcoxon signed-rank test on paired absolute residuals from the test set confirmed that XGBoost produced significantly lower errors than MLP for both °Brix ($p < 0.001$) and TA ($p < 0.001$). However, the error difference between XGBoost ($R^2 = 0.867$ for °Brix; $R^2 = 0.833$ for TA) and Random Forest ($R^2 = 0.865$ for °Brix; $R^2 = 0.846$ for TA) was not statistically significant ($p = 0.4054$ for °Brix; $p = 0.5039$ for TA), indicating that both tree-based ensemble approaches achieved equivalent predictive performance for this dataset. The best hyperparameters identified by the expanded XGBoost grid search were: learning_rate = 0.05, max_depth = 3, and n_estimators = 100 (for °Brix) and learning_rate = 0.05, max_depth = 3, and n_estimators = 100 (for TA).

3.3. Predictor Contribution Analysis (SHAP)

Interpretability analysis (SHAP) was applied to the °Brix models to understand each predictor's contribution. In all data fusion models (Figure 7d–f), GDD was confirmed as the predictor with the highest global influence. The SHAP summary plot reveals an agronomically coherent relationship: high GDD values (indicating greater thermal accumulation) generate a positive impact, increasing the predicted °Brix value, which is consistent with the known phenological rule of ripening, in which thermal accumulation progressively drives sugar concentration in the berry.

A fundamental change was observed in the models' internal behavior. In the GDDabsent (VI-only) scenarios, the tree-based models, Random Forest (Figure 7a) and XGBoost (Figure 7b), assigned the greatest weight to CVI. In contrast, the MLP (Figure 7c) prioritized BAI, although it distributed the importance more evenly among 'BAI', 'TSAVI', 'CVI', and 'SFDVI'. This suggests that, in the absence of GDD, the models attempted to use proxies for vigor and chlorophyll content (like CVI) to estimate maturity.

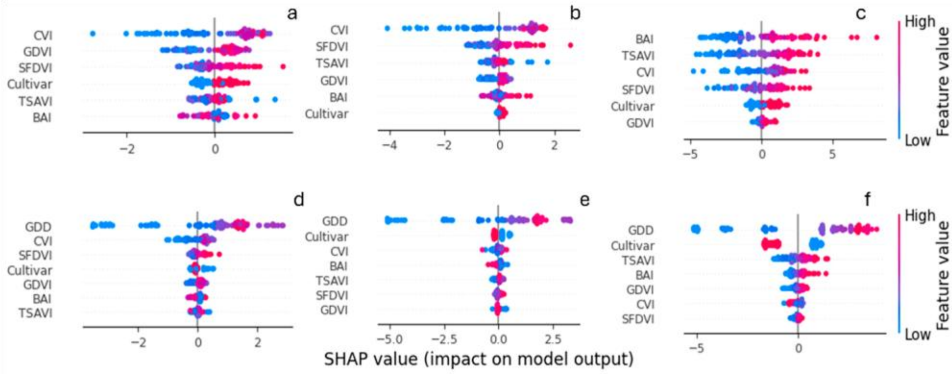


Figure 7. SHAP summary plots detailing the importance and impact of each variable on °Brix prediction. The upper panels (**a**–**c**) show the models without GDD; the lower panels (**d**–**f**) show the models with GDD. Models: Random Forest (**a**,**d**), XGBoost (**b**,**e**), and MLP (**c**,**f**). Each dot represents one observation. The dot color reflects the feature value magnitude: red indicates high feature values and blue indicates low feature values, following the standard SHAP color scale.

However, with the inclusion of GDD, the feature importance hierarchy changed dramatically. As seen in the SHAP plots for all three models (Figure 7d–f), the 'Cultivar' variable consistently emerged as the second-most influential predictor, surpassing all Vegetation Indices. The detailed analysis (beeswarm plot) reveals how the generalist model functions: for a given GDD value, the model systematically assigns a higher impact (higher predicted °Brix) to one cultivar (represented in red) than the other (in blue). This demonstrates that the model did not create separate logics but instead learned a base relationship (Brix vs. GDD) that is then systematically adjusted according to the cultivar identity. The SHAP analysis for the Titratable Acidity (TA) models reveals more complex behavior than that observed for °Brix (Figure 8). In the data fusion scenario (Figure 8d–f), GDD was consistently identified as the predictor with the highest impact across all three algorithms. As agronomically expected, the beeswarm plot demonstrates a strong inverse relationship: high GDD values (associated with advanced ripening stages) generate negative SHAP values, contributing to lower predicted acidity.

In the GDD-absent scenario, the models diverged. In the Random Forest (Figure 8a) and MLP (Figure 8c) models, the 'Cultivar' variable was the most influential factor, indicating that, lacking a temporal predictor, the grape's genetic identity was the most important information for determining acidity levels.

The most significant change occurred in the importance hierarchy after GDD was included. In the Random Forest (Figure 8d) model, 'Cultivar' remained the second-most important variable, following GDD, demonstrating that the model captured both the temporal trend and the baseline genetic difference between varieties. The MLP (Figure 8f) also maintained 'Cultivar' as highly relevant, following its tendency for greater balance among predictors.

In striking contrast, in the XGBoost (Figure 8e) model, the inclusion of GDD caused the

'Cultivar' variable to drop to the last position in importance. The XGBoost model attributed almost all predictive power to GDD, with the VIs acting as secondary modulators. This fundamental difference in how XGBoost and Random Forest handled the 'Cultivar' variable is a crucial insight into the internal workings of the algorithms for this specific task.

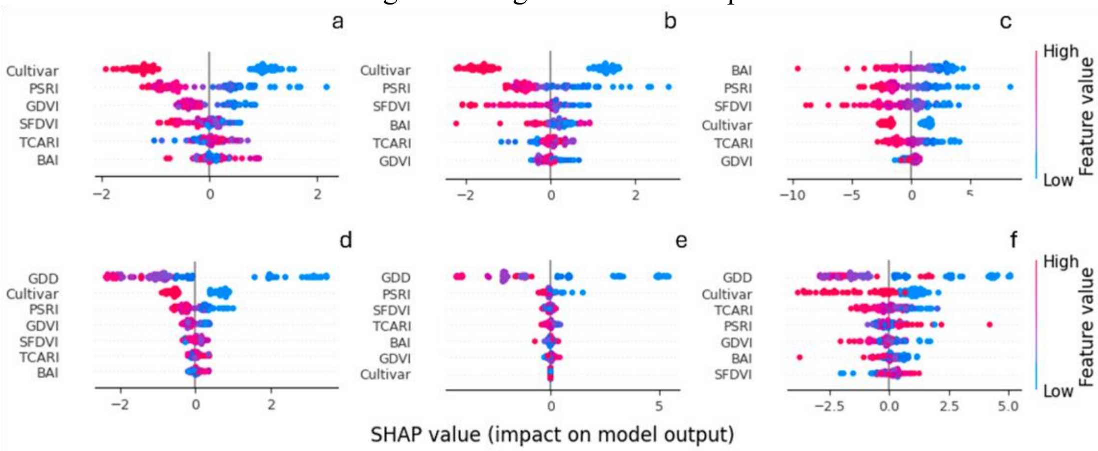


Figure 8. SHAP summary plots detailing the importance and impact of each variable on Titratable Acidity (g/L) prediction. The upper panels (**a**–**c**) show the models without GDD; the lower panels (**d**–**f**) show the models with GDD. Models: Random Forest (**a**,**d**), XGBoost (**b**,**e**), and MLP (**c**,**f**). Each dot represents one observation. The dot color reflects the feature value magnitude: red indicates high feature values and blue indicates low feature values, following the standard SHAP color scale.

Taken together, the SHAP interpretability analyses for °Brix and Titratable Acidity reinforce the relevance of the XGBoost model. It is evident that this model emerged not only as the most accurate predictor ($R^2 = 0.887$ for °Brix; $R^2 = 0.766$ for TA), but also as an interpretable model aligned with viticultural principles. The model demonstrated its ability to learn the opposing physiological dynamics (sugar accumulation and acid degradation) and to modulate both predictions based on cultivar identity. This approach validates the development of a single, generalist, and physiologically aware model for estimating technological maturity.

From a physicochemical perspective, the predominance of GDD across all models reflects its role as the primary integrator of veraison-associated metabolic processes. During grape ripening, sucrose imported from photosynthetically active leaves is hydrolysed by invertase into glucose and fructose, with this enzymatic activity being thermally regulated and tightly coupled to accumulated heat units [30]. The second-order importance of 'Cultivar' in the SHAP hierarchy is consistent with the distinct metabolic profiles of Sauvignon Blanc and Syrah, which differ substantially in the activity of malic enzyme and malate dehydrogenase—enzymes governing the rate of malate degradation and, consequently, the trajectory of Titratable Acidity decline during ripening [30]. The VIs, functioning as tertiary predictors, capture within-cultivar spatial heterogeneity in canopy photosynthetic capacity and chlorophyll content, which modulates the source–sink balance and thereby the rate at which individual vines deviate from the cultivar–GDD maturation trajectory. These patterns are consistent with known grapevine physiology; however, as SHAP values reflect model-internal associations rather than causal pathways, these interpretations should be regarded as hypothesis-generating rather than mechanistically conclusive.

4. Discussion

The central finding of this study is the unequivocal demonstration that Vegetation Indices (VIs) alone are insufficient predictors of grapevine technological maturity. Our results showed that, despite a rigorous feature selection process, the VI-only models failed to generalize, exhibiting low R^2 (0.30–0.47 for °Brix) and high overfitting. This finding corroborates the results of [8,9], who, despite testing different sensors and algorithms, also reported modest R^2 values (0.52 and 0.61, respectively). This reinforces our initial hypothesis: VIs provide a spatial "snapshot" of canopy vigor but lack the crucial temporal dimension that governs phenology. The inclusion of Growing Degree-Days (GDD) as a temporal predictor resolved this gap, transforming non-viable models into high-precision tools ($R^2 > 0.88$ for °Brix with XGBoost). GDD acted as the model's "backbone," providing the baseline trajectory of physiological development (sugar accumulation and acid degradation), while the VIs and the 'Cultivar' variable acted as secondary modulators, refining the prediction based on spatial vigor variability and genetics.

Regarding the analyses for selecting model inputs, collinearity among VIs is often overlooked, and models are generated with numerous VIs as predictors [9,31–33]. In most cases, only a relevance-focused feature selection (like RFE) is applied, without redundancy analysis. Many VIs are mathematical derivations that use the same spectral bands, resulting in informational redundancy. The application of a strict statistical criterion, such as a VIF threshold of 5.0 [34,35], was a crucial methodological step to mitigate this redundancy, ensuring the robustness and interpretability of the final model, following established recommendations for handling collinearity in multivariate explanatory models [36,37].

The result of this filtering was the non-selection of established indices like NDVI. VIs such as NDVI, NDRE, and GNDVI are widely used for predicting biomass and yield [5]. However, our results indicated that other VIs, such as CVI, SFDVI, and BAI, contributed more to predicting maturity. Although NDVI is an excellent indicator of general vigor, we suggest that its known saturation in dense canopies, such as grapevines, limited its ability to identify the subtle phenological changes associated with fruit ripening. This saturation phenomenon is well documented for vegetation canopies with a Leaf Area Index (LAI) exceeding approximately 3.0, beyond which NDVI asymptotically plateaus at values above 0.80 regardless of further increases in biomass or chlorophyll content [5,6]. Although a formal analysis of per-date reflectance distributions for this specific vineyard is deferred to the next revision stage, consistently elevated NDVI values throughout the ripening period are consistent with the known saturation behavior of this index in commercially managed, high-density grapevine canopies. This does not invalidate NDVI, which is indeed effective for yield prediction [38,39]. However, to estimate grape quality, the literature has already reported that it should be used in association with other data, which reinforces our data fusion approach [40,41].

Similarly, NDRE, which uses the red-edge band for a better estimate of chlorophyll content, was pre-selected by RFE. However, its subsequent removal due to an extremely high VIF (VIF > 1950) indicates that its information was already contained in other predictors, such as CVI and REDVI, which proved to be more relevant and less redundant. This reinforces that the choice of the best spectral predictors is highly dependent on the agronomic target and that a data-driven approach, such as the one we used (RFE + VIF), is superior to the simple adoption of conventional indices. It should also be acknowledged that the feature selection pipeline employed Random Forest as the base estimator for RFE, which may introduce a selection bias towards features that are particularly informative for tree-based ensemble methods, potentially yielding suboptimal feature subsets for XGBoost and MLP. Algorithm-specific or multi-learner feature selection strategies could be explored in future work to address this limitation.

Even after applying variable selection criteria and using only those of a spectral nature that contribute to predicting grape technological maturity, it is worth highlighting, among our study's main results, the low accuracy of machine learning models in predicting this maturity when VIs are used exclusively as input data. This result should be considered a point of caution for most growers, especially given the growth of service providers using drones for mapping linked to grapevine quality and justifying the use of these data as support for field decision-making. Thus, as demonstrated, regardless of the algorithm's architecture, the resulting models exhibited overfitting, combining good performance on the training data with insignificant predictive performance on the test set.

This low accuracy stems from a conceptual limitation wherein VIs provide spatial information on the canopy's physiological state but lack the temporal dimension that governs the ripening process. Ripening is a cumulative process, intrinsically linked to accumulated thermal energy (GDD) throughout the phenological cycle [42,43]. A vine can exhibit the same vigor, and therefore have the same VI value, at different developmental stages but with completely different fruit compositions. Without the temporal context provided by GDD, the models were unable to resolve this ambiguity, resulting in models with training R^2 values up to 0.81 (for RF), but with test R^2 values falling as low as 0.30 (for MLP) and never exceeding 0.48. Despite the low values found in our work using only VIs as input, [9] also reported low accuracy values when they tested different sensors to predict grape quality attributes.

The inclusion of GDD improved the performance of the models, especially in the test stage, validating the central hypothesis that data fusion is essential. The SHAP interpretability analysis was fundamental for decoding the internal workings of the final XGBoost model, revealing how this synergy operates and confirming that the model learned to replicate the physiological logic.

The interpretability analysis (SHAP) consistently revealed, across the three tested models, that GDD was the predictor with the greatest impact and potentiated the prediction of grape development, as also observed for other crops [11,12,44]. This result corroborates field observations, as GDD functions as a direct indicator of the grapevine's phenological development stage [45,46].

The SHAP results suggest a clear hierarchy: GDD and 'Cultivar' establish the baseline prediction, defining the expected maturity potential for that cycle stage and genetic variety. The VIs, in turn, act as secondary predictors. They modulate the baseline prediction by capturing spatial variability in the crop's physiological conditions, such as localized stresses or vigor differences, which cause a plant to deviate from the general trend expected for its GDD and cultivar. Furthermore, the importance of 'Cultivar' in the SHAP analysis demonstrates the model's ability to learn and quantify the genetically distinct ripening profiles between Sauvignon Blanc and Syrah, a fact aligned with physiological knowledge about sugar accumulation and acid degradation in different varieties [30].

However, an important methodological caveat applies to the interpretation of 'Cultivar' importance in the SHAP analyses: the two cultivars occupy spatially distinct blocks with different row orientations (Sauvignon Blanc: NE-SW; Syrah: NW-SE), meaning that 'Cultivar' is inherently confounded with block-level spatial factors, including canopy geometry, differential solar irradiance exposure, and localized microclimate variability. Consequently, the SHAP-attributed importance of 'Cultivar' reflects an undifferentiated composite of genetic ripening dynamics and spatial-environmental co-factors, and it is not possible within the current experimental design to attribute this contribution exclusively to genotypic differences in metabolic processes. Future studies should consider implementing cultivar-specific sub-models or incorporating spatial covariates such as topographic aspect and slope to disentangle these confounded effects.

However, the way this genetic influence interacts with remote sensing signals can be modulated by management factors, leading to sometimes divergent results in the literature. For instance, we observed that, under irrigation, VIs were less effective in predicting quality for Sauvignon Blanc compared to Syrah, demonstrating a complex interaction between cultivar and water management [47]. Conversely, the work by Santos [48] found robust correlations in *Vitis vinifera* L., suggesting the potential of VIs to replace conventional analyses in certain contexts. This apparent inconsistency in the literature does not invalidate the results presented here; on the contrary, it demonstrates that *Vitis vinifera* L. is extremely sensitive to local interactions among genotype, environment, and management. Therefore, the calibration of generalist models, as performed in the present study, is an indispensable step for the successful application of remote sensing tools in decision-making.

Among the evaluated models, XGBoost showed the highest accuracy, a result frequently observed in agricultural applications due to its efficiency in handling tabular data and complex non-linear relationships [49]. It is important to note, however, that all models (including RF and MLP) achieved high performance after the inclusion of GDD, suggesting that the quality of the predictor variables was more decisive than the choice of the specific algorithm.

Several methodological and contextual limitations of this study warrant explicit discussion. First, the dataset encompasses a single growing season (2024) at one commercial vineyard, constraining generalizability to season-specific climatic patterns and the particular agronomic conditions of this site. Second, GDD was accumulated over only six discrete collection events, making it nearly co-linear with the temporal ordering of sampling; a dateas-surrogate sensitivity analysis was performed: replacing GDD with an ordinal temporal index (1–5) yielded $R^2 = 0.815$ (°Brix) and $R^2 = 0.821$ (TA), compared to 0.84 and 0.806 with GDD ($\Delta R^2 = 0.025$ and -0.015). For TA, the date proxy marginally outperformed GDD, indicating that temporal ordering rather than thermodynamic accumulation may drive a portion of the predictive signal in this five-date dataset. Third, the stratified hold-out split did not account for spatial autocorrelation among the 145 georeferenced points across collection dates; a GroupKFold analysis stratified by point identity is planned. Fourth, the ‘Cultivar’ predictor is spatially confounded with block geometry and microclimate, limiting mechanistic interpretation of its SHAP-ranked importance. Fifth, the RFE pipeline employed Random Forest as the base estimator, potentially yielding suboptimal feature subsets for non-tree-based algorithms. Sixth, SHAP values reflect model-internal association patterns and do not establish causal relationships. Finally, soil properties, vine water status, and thermal heterogeneity within the vineyard were not incorporated as predictors, representing avenues for model improvement in future multi-sensor, multi-season studies.

5. Conclusions

The fusion of remote sensing data (Vegetation Indices) with the temporal variable Growing Degree-Days (GDD) proved to be a fundamental and highly effective approach for predicting Soluble Solids (°Brix) and Titratable Acidity in wine grapes grown under a double pruning system. The study validated its central hypothesis by demonstrating that the inclusion of GDD significantly improved accuracy, transforming models based only on VIs, which were shown to be incapable of generalization (test $R^2 < 0.48$), into robust prediction tools. Furthermore, the most influential and least redundant VIs for prediction were identified, such as 'CVI', 'SFDVI', and 'BAI', whereas traditional indices ('NDVI' and 'NDRE') which were removed due to saturation or high multicollinearity. Among the evaluated algorithms, XGBoost was determined to have the best predictive performance for both target variables, achieving a test set R^2 of 0.89 for °Brix and 0.77 for Titratable Acidity.

This work therefore reinforces that the fusion of temporal (GDD) and spatial (VI) data is a prerequisite for the accurate modeling of grapevine technological maturity. These findings constitute a promising proof-of-concept for non-destructive maturity prediction in precision viticulture; however, validation across multiple growing seasons and vineyard environments is essential to confirm the broader applicability and operational reliability of the proposed framework.

Author Contributions: O.P.d.C. contributed to conceptualization, methodology, formal analysis, data curation, and writing—original draft. F.L.d.S.R.F. contributed to investigation, software, and data processing. B.S.C.B. contributed to investigation, data curation, and writing—review and editing. R.F.Q.A. contributed to investigation and data curation. G.V.F. contributed to methodology, formal analysis, and writing—review and editing. M.d.M.A. contributed to methodology, formal analysis, and writing—review and editing. C.C.R. contributed to investigation, software, and data processing. P.M.P. contributed to supervision and methodology. A.F.d.S. contributed to supervision, conceptualization, project administration, and writing—review and editing. R.P. contributed to supervision. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding authors.

Acknowledgments: We acknowledge the Federal University of Lavras (UFLA) for its institutional support and infrastructure, and Vinícola Alma Gerais for their partnership and for providing the experimental area, as well as the Coordination for the Improvement of Higher Education Personnel (CAPES) Code 001, the National Council for Scientific and Technological Development (CNPq) and the Research Support Foundation of the State of Minas Gerais (FAPEMIG) for their financial support and scholarships, which were fundamental to conducting this study.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. de Albuquerque Regina, M.; Fráguas, J.C.; Aluarenga, Â.A.; de Souza, C.R.; de Amorim, D.A.; da Mote, R.V.; Fáuero, A.C. Implantação e manejo do vinhedo para produção de vinhos de qualidade. *Inf. Agropecuário* **2006**, *27*, 16–31.
2. Penso, G.A.; Citadin, I.; Maro, L.A.C.; Scarioto, S.; Chicock, A.; Ambrosio, R. Evolução da maturação de uvas finas (*Vitis vinifera* L.), no município de Dois Vizinhos—Paraná. *Semin. Ciências Agrárias* **2014**, *35*, 3085–3098. [CrossRef]
3. Chen, S.; Zhang, F.; Ning, J.; Liu, X.; Zhang, Z.; Yang, S. Predicting the anthocyanin content of wine grapes by NIR hyperspectral imaging. *Food Chem.* **2015**, *172*, 788–793. [CrossRef] [PubMed]
4. Bramley, R.G.V. 12—Precision Viticulture: Managing vineyard variability for improved quality outcomes. In *Managing Wine Quality*, 2nd ed.; Reynolds, A.G., Ed.; Woodhead Publishing: Cambridge, UK, 2022; pp. 541–586.
5. Giovos, R.; Tassopoulos, D.; Kalivas, D.; Lougkos, N.; Priovolou, A. Remote Sensing Vegetation Indices in Viticulture: A Critical Review. *Agriculture* **2021**, *11*, 457. [CrossRef]
6. Taskos, D.G.; Koundouras, S.; Stamatiadis, S.; Zioziou, E.; Nikolaou, N.; Karakioulakis, K.; Theodorou, N. Using active canopy sensors and chlorophyll meters to estimate grapevine nitrogen status and productivity. *Precis. Agric.* **2015**, *16*, 77–98. [CrossRef]

7. Song, J.; Smart, R.; Wang, H.; Dambergs, B.; Sparrow, A.; Qian, M.C. Effect of grape bunch sunlight exposure and UV radiation on phenolics and volatile composition of *Vitis vinifera* L. cv. Pinot noir wine. *Food Chem.* **2015**, *173*, 424–431. [CrossRef] [PubMed]
8. Lyu, H.; Grafton, M.; Ramilan, T.; Irwin, M.; Wei, H.-E.; Sandoval, E. Using Remote and Proximal Sensing Data and Vine Vigor Parameters for Non-Destructive and Rapid Prediction of Grape Quality. *Remote Sens.* **2023**, *15*, 5412. [CrossRef]
9. Kasimati, A.; Espejo-García, B.; Darra, N.; Fountas, S. Predicting Grape Sugar Content under Quality Attributes Using Normalized Difference Vegetation Index Data and Automated Machine Learning. *Sensors* **2022**, *22*, 3249. [CrossRef] [PubMed]
10. Nova, N.A.V. *Estimativa de Graus-Dia Acumulados Acima de Qualquer Temperatura Base, em Função das Temperaturas Máxima e Mínima*; Universidade de São Paulo, Instituto de Geografia: São Paulo, Brazil, 1972; Volume 30.
11. Liu, L.-W.; Lu, C.-T.; Wang, Y.-M.; Lin, K.-H.; Ma, X.; Lin, W.-S. Rice (*Oryza sativa* L.) Growth Modeling Based on Growth Degree Day (GDD) and Artificial Intelligence Algorithms. *Agriculture* **2022**, *12*, 59. [CrossRef]
12. Tedesco, D.; de Oliveira, M.F.; dos Santos, A.F.; Costa Silva, E.H.; de Souza Rolim, G.; da Silva, R.P. Use of remote sensing to characterize the phenological development and to predict sweet potato yield in two growing seasons. *Eur. J. Agron.* **2021**, *129*, 126337. [CrossRef]
13. Alvares, C.A.; Stape, J.L.; Sentelhas, P.C.; Gonçalves, J.d.M.; Sparovek, G. Köppen's climate classification map for Brazil. *Meteorol. Z.* **2013**, *22*, 711–728. [CrossRef]
14. Instituto Adolfo Lutz. *Normas Analíticas do Instituto Adolfo Lutz: Métodos Químicos e Físicos para Análise de Alimentos*, 3rd ed.; Instituto Adolfo Lutz: São Paulo, Brazil, 1985; Volume 1.
15. Team, R.C. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2025.
16. Hijmans, R.J.; Bivand, R.; Dyba, K.; Pebesma, E.; Sumner, M. *Terra: Spatial Data Analysis*; R package version 1.7-46; R Foundation for Statistical Computing: Vienna, Austria, 2023. Available online: https://CRAN.R-project.org/package=terra (accessed on 25 June 2026).
17. Pebesma, E. Simple Features for R: Standardized Support for Spatial Vector Data. *R J.* **2018**, *10*, 439–446. [CrossRef]
18. Matias, F.I.; Caraza-Harter, M.V.; Endelman, J.B. FIELDimageR: An R package to analyze orthomosaic images from agricultural field trials. *Plant Phenome J.* **2020**, *3*, e20005. [CrossRef]
19. Leutner, B.; Horning, N.; Leutner, M.B. *RStoolbox: Tools for Remote Sensing Data Analysis in R*; R Foundation for Statistical Computing: Vienna, Austria, 2022. Available online: https://CRAN.R-project.org/package=RStoolbox (accessed on 25 June 2026).
20. Guyon, I.; Weston, J.; Barnhill, S.; Vapnik, V. Gene Selection for Cancer Classification using Support Vector Machines. *Mach. Learn.* **2002**, *46*, 389–422. [CrossRef]
21. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
22. Chuvieco, E.; Martín, M.P.; Palacios, A. Assessment of different spectral indices in the red-near-infrared spectral domain for burned land discrimination. *Int. J. Remote Sens.* **2002**, *23*, 5103–5110. [CrossRef]
23. Baptista, G.d.M. Aplicação do Índice de Vegetação por Profundidade de Feição Espectral (SFDVI-Spectral Feature Depth Vegetation Index) em dados RapidEye. In *Proceedings of the XVII Simpósio Brasileiro de Sensoriamento Remoto–SBSR*; INPE: João Pessoa, Brazil, 2015; pp. 2277–2284.
24. Nigon, T.J.; Mulla, D.J.; Rosen, C.J.; Cohen, Y.; Alchanatis, V.; Rud, R. Evaluation of the nitrogen sufficiency index for use with high resolution, broadband aerial imagery in a commercial potato field. *Precis. Agric.* **2014**, *15*, 202–226. [CrossRef]
25. Vincini, M.; Frazzi, E.; D'Alessio, P. A broad-band leaf chlorophyll vegetation index at the canopy scale. *Precis. Agric.* **2008**, *9*, 303–319. [CrossRef]
26. Baret, F.; Guyot, G. Potentials and limits of vegetation indices for LAI and APAR assessment. *Remote Sens. Environ.* **1991**, *35*, 161–173. [CrossRef]
27. Merzlyak, M.N.; Gitelson, A.A.; Chivkunova, O.B.; Rakitin, V.Y. Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. *Physiol. Plant.* **1999**, *106*, 135–141. [CrossRef]
28. Haboudane, D.; Miller, J.R.; Pattey, E.; Zarco-Tejada, P.J.; Strachan, I.B. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sens. Environ.* **2004**, *90*, 337–352. [CrossRef]
29. Stackhouse, P.W.; Westberg, D.; Hoell, J.M.; Chandler, W.S.; Zhang, T. *Prediction of Worldwide Energy Resource (POWER)Agroclimatology Methodology-(1.0 Latitude by 1.0 Longitude Spatial Resolution)*; NASA Langley Research Center: Hampton, VA, USA, 2015.

30. Conde, C.; Silva, P.; Fontes, N.; Dias, A.C.; Tavares, R.M.; Sousa, M.J.; Agasse, A.; Delrot, S.; Gerós, H. Biochemical changes throughout grape berry development and fruit and wine quality. *Food* **2007**, *1*, 1–22.
31. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4765–4774.
32. dos Santos, A.F.; Corrêa, L.N.; Lacerda, L.N.; Tedesco-Oliveira, D.; Pilon, C.; Vellidis, G.; da Silva, R.P. High-resolution satellite image to predict peanut maturity variability in commercial fields. *Precis. Agric.* **2021**, *22*, 1464–1478. [CrossRef]
33. Jiang, J.; Atkinson, P.M.; Chen, C.; Cao, Q.; Tian, Y.; Zhu, Y.; Liu, X.; Cao, W. Combining UAV and Sentinel-2 satellite multi-spectral images to diagnose crop growth and N status in winter wheat at the county scale. *Field Crops Res.* **2023**, *294*, 108860. [CrossRef]
34. Abeysiriwardana, H.D.; Gomes, P.I.A. Integrating vegetation indices and geo-environmental factors in GIS-based landslidesusceptibility mapping: Using logistic regression. *J. Mt. Sci.* **2022**, *19*, 477–492. [CrossRef]
35. Weisberg, S. *Applied Linear Regression*; John Wiley & Sons: Hoboken, NJ, USA, 2005; Volume 528. [CrossRef]
36. Dormann, C.F.; Elith, J.; Bacher, S.; Buchmann, C.; Carl, G.; Carre, G.; Marquez, J.R.G.; Gruber, B.; Lafourcade, B.; Leitao, P.J.; et al. Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography* **2013**, *36*, 27–46. [CrossRef]
37. Liakos, K.G.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D. Machine learning in agriculture: A review. *Sensors* **2018**, *18*, 2674. [CrossRef] [PubMed]
38. Arab, S.T.; Noguchi, R.; Matsushita, S.; Ahamed, T. Prediction of grape yields from time-series vegetation indices using satellite remote sensing and a machine-learning approach. *Remote Sens. Appl. Soc. Environ.* **2021**, *22*, 100485. [CrossRef]
39. Martinez-Casasnovas, J.A.; Agelet-Fernandez, J.; Arno, J.; Ramos, M.C. Analysis of vineyard differential management zones and relation to vine development, grape maturity and quality. *Span. J. Agric. Res.* **2012**, *10*, 326–337. [CrossRef]
40. Acevedo-Opazo, C.; Tisseyre, B.; Guillaume, S.; Ojeda, H. The potential of high spatial resolution information to define withinvineyard zones related to vine water status. *Precis. Agric.* **2008**, *9*, 285–302. [CrossRef]
41. Santesteban, L.G.; Royo, J.B. Water status, leaf area and fruit load influence on berry weight and sugar accumulation of cv. 'Tempranillo' under semiarid conditions. *Sci. Hortic.* **2006**, *109*, 60–65. [CrossRef]
42. Ahmed, S.; Roberto, S.R.; Shahab, M.; Colombo, R.C.; Silvestre, J.P.; Koyama, R.; de Souza, R.T. Proposal of double-cropping system for 'BRS Isis' seedless grape grown in subtropical area. *Sci. Hortic.* **2019**, *251*, 118–126. [CrossRef]
43. Salama, A.-M.; Ezzat, A.; El-Ramady, H.; Alam-Eldein, S.M.; Okba, S.K.; Elmenofy, H.M.; Hassan, I.F.; Illés, A.; Holb, I.J. Temperate Fruit Trees under Climate Change: Challenges for Dormancy and Chilling Requirements in Warm Winter Regions. *Horticulturae* **2021**, *7*, 86. [CrossRef]
44. Chen, P.; Li, Y.; Liu, X.; Tian, Y.; Zhu, Y.; Cao, W.; Cao, Q. Improving yield prediction based on spatio-temporal deep learning approaches for winter wheat: A case study in Jiangsu Province, China. *Comput. Electron. Agric.* **2023**, *213*, 108201. [CrossRef]
45. Jones, G.V. Climate change in the western United States grape growing regions. In *Proceedings of the VII International Symposium on Grapevine Physiology and Biotechnology*; International Society for Horticultural Science: Davis, CA, USA, 2004; Volume 689, pp. 41–60.
46. Monteiro, H.S.A.; Tecchio, M.A.; Brito, S.d.N.S.; Neto, F.J.D.; Sánchez, C.A.P.C.; Alonso, J.C.; Feliciano, D.E.F.; Maniero, C.R.; Cunha, P.H.H.; Silva, M.d.S. Phenological Development, Thermal Requirement, and Quality of 'BRS Núbia' (*Vitis vinifera* L. x *Vitis labrusca* L.) Grapes on Different Rootstocks. *Horticulturae* **2025**, *11*, 466. [CrossRef]
47. Fountas, S.; Anastasiou, E.; Balafoutis, A.; Koundouras, S.; Theoharis, S.; Theodorou, N. The influence of vine variety and vineyard management on the effectiveness of canopy sensors to predict winegrape yield and quality. In Proceedings of the International Conference of Agricultural Engineering, Zurich, Switzerland, 6–10 July 2014; p. 8.
48. Santos, M.d.S.; Gebler, L.; Sebem, E. Correlation between vegetation indexes generated at *Vitis vinifera* L. and soil, plant and production parameters for emergency application in decision making. *Ciência Rural* **2022**, *52*, e20201037. [CrossRef]
49. Torres-Sánchez, R.; Martínez-Zafra, M.T.; Castillejo, N.; Guillamón-Frutos, A.; Artés-Hernández, F. Real-Time Monitoring System for Shelf Life Estimation of Fruit and Vegetables. *Sensors* **2020**, *20*, 1860. [CrossRef] [PubMed]