

CLASSIFICAÇÃO DA PRODUTIVIDADE DA CANA-DE-AÇÚCAR COM RANDOM FOREST INTEGRANDO ÍNDICES DE VEGETAÇÃO E VARIÁVEIS AMBIENTAIS

Victor Costa Leda¹

<https://orcid.org/0009-0003-3751-0119>

Bruna Soares Xavier de Barros²

<https://orcid.org/0000-0003-1859-1366>

Zacarias Xavier de Barros³

<https://orcid.org/0009-0002-5679-7955>

RESUMO

A estimativa da produtividade da cana-de-açúcar antes da colheita é importante para o planejamento das operações agrícolas e para a gestão da produção. O objetivo deste estudo foi avaliar o desempenho do algoritmo Random Forest na classificação da produtividade, utilizando diferentes combinações de dados espectrais, agronômicos e ambientais ao longo da safra. A área de estudo está localizada nos municípios de Pratânia e Lençóis Paulista, estado de São Paulo. Foram utilizadas imagens do sensor OLI, a bordo do satélite Landsat 8, para obtenção dos índices NDVI, EVI e MSAVI2 em períodos correspondentes a aproximadamente 90 e 180 dias antes da colheita, além da última imagem disponível antes da colheita. Também foram consideradas informações de solo, variedade, estágio da cultura, declividade e orientação de vertentes, combinadas em 12 configurações de entrada. As classificações foram realizadas por meio de validação cruzada em 10 grupos e avaliadas pelo percentual de classificações corretas e pelo índice Kappa. A configuração contendo todas as variáveis apresentou 87,9%, 87,38% e 82,93% de classificações corretas nos períodos inicial, intermediário e final da safra, respectivamente, com Kappa de 0,8. A configuração que utilizou os índices de vegetação obtidos aproximadamente 180 dias antes da colheita, juntamente com solo, variedade, estágio da cultura, declividade e aspecto, apresentou 85,2%, 84,82% e 81,58% de classificações corretas nos mesmos períodos, respectivamente, também com Kappa de 0,8. Dessa forma, a utilização dos índices de vegetação de uma única época permitiu manter resultados próximos aos obtidos com o conjunto completo, inclusive com informações espectrais adquiridas aproximadamente 180 dias antes da colheita.

Palavras-chave

Sensoriamento remoto; Aprendizado de máquina; Landsat 8; Agricultura de precisão; Planejamento agrícola.

Submetido em: 03/09/2026 – Aprovado em: 09/10/2026 – Publicado em: 09/10/2026

¹ Doutor, FCA/UNESP, Botucatu, SP, victorleda@gmail.com

² Pós-doutora, FCA/UNESP, Botucatu, SP, brunasxb@gmail.com

³ Professor Titular, FCA/UNESP, Botucatu, SP, zacarias.barros@unesp.br



SUGARCANE PRODUCTIVITY CLASSIFICATION WITH RANDOM FOREST INTEGRATING VEGETATION INDICES AND ENVIRONMENTAL VARIABLES

ABSTRACT

Estimating sugarcane productivity before harvest is important for planning agricultural operations and managing production. This study aimed to evaluate the performance of the Random Forest algorithm for productivity classification using different combinations of spectral, agronomic, and environmental data throughout the harvest season. The study area is located in the municipalities of Pratânia and Lençóis Paulista, São Paulo State, Brazil. Images from the OLI sensor onboard Landsat 8 were used to obtain the NDVI, EVI, and MSAVI2 vegetation indices for periods corresponding to approximately 90 and 180 days before harvest, in addition to the last available image before harvest. Information on soil type, sugarcane variety, crop stage, slope, and aspect was also considered and combined into 12 input configurations. The classifications were performed using 10-fold cross-validation and evaluated based on the percentage of correctly classified instances and the Kappa coefficient. The configuration containing all variables achieved 87.9%, 87.38%, and 82.93% correctly classified instances in the initial, intermediate, and final periods of the harvest season, respectively, with a Kappa coefficient of 0.8. The configuration using vegetation indices obtained approximately 180 days before harvest, together with soil type, variety, crop stage, slope, and aspect, achieved 85.2%, 84.82%, and 81.58% correctly classified instances in the same periods, respectively, also with a Kappa coefficient of 0.8. Thus, using vegetation indices from a single acquisition period maintained results close to those obtained with the complete dataset, including when spectral information was acquired approximately 180 days before harvest.

Keywords

Remote sensing; Machine learning; Landsat 8; Precision agriculture; Agricultural planning.

1 INTRODUÇÃO

A cana-de-açúcar apresenta grande importância econômica no Brasil, principalmente no estado de São Paulo, tanto pela produção de açúcar e etanol quanto pelo aproveitamento da biomassa para geração de energia. A obtenção de informações sobre a produtividade antes da colheita é importante nesse setor, pois pode auxiliar no planejamento das operações agrícolas e na organização da safra.

A produtividade da cultura está relacionada a múltiplos fatores que atuam ao longo do seu desenvolvimento. Características do solo e do relevo, variedade cultivada, estágio de desenvolvimento e condições ambientais podem resultar em diferenças de produtividade mesmo entre áreas próximas. A integração dessas informações torna-se uma alternativa para representar melhor a variabilidade existente nas áreas cultivadas.

Entre as possibilidades para obtenção de informações sobre as culturas agrícolas está o sensoriamento remoto, que permite acompanhar extensas áreas ao longo do tempo. Os dados espectrais podem ainda ser associados a informações agronômicas e ambientais, ampliando o conjunto de características consideradas na análise da produtividade.

Trabalhar com dados de naturezas distintas também exige métodos adequados para sua análise. Neste estudo, foi utilizado o algoritmo Random Forest para a classificação da produtividade da cana-de-açúcar, combinando informações espectrais com dados de solo, relevo, variedade e estágio de desenvolvimento da cultura. Outro aspecto considerado foi o momento de obtenção das informações espectrais. Os índices de vegetação foram avaliados em três períodos anteriores à colheita, possibilitando verificar se informações obtidas com maior antecedência poderiam manter resultados satisfatórios de classificação e, consequentemente, ter aplicação no planejamento agrícola.

Assim, o objetivo deste estudo foi avaliar o desempenho do algoritmo Random Forest na classificação da produtividade da cana-de-açúcar a partir de combinações de variáveis espectrais, agronômicas e ambientais, considerando índices de vegetação obtidos em períodos anteriores à colheita.

2 REVISÃO DE LITERATURA

2.1 Sensoriamento remoto aplicado à cana-de-açúcar

Acompanhar a produtividade da cana-de-açúcar em campo, talhão a talhão, é caro e lento em áreas extensas. O sensoriamento remoto contorna essa limitação: sensores multiespectrais registram a energia refletida pela vegetação ao longo de toda a safra, permitindo monitorar grandes áreas de cultivo a um custo consideravelmente menor do que o levantamento convencional.

Na cana-de-açúcar, imagens orbitais têm sido utilizadas para o mapeamento das áreas cultivadas, acompanhamento do desenvolvimento da cultura e obtenção de informações relacionadas à produtividade. Uma das vantagens desse tipo de dado é a possibilidade de observar uma mesma área em diversos momentos, acompanhando as modificações ocorridas durante o ciclo da cultura (SILVA et al., 2024).

Grande parte dessas aplicações utiliza informações extraídas da resposta espectral da vegetação. Os dados obtidos pelas imagens podem ser analisados isoladamente ou associados a características agronômicas e ambientais, dependendo da finalidade do estudo. Para a estimativa de produtividade, essa integração permite considerar tanto as condições apresentadas pela cultura nas imagens quanto outros fatores relacionados à área de produção.

Em revisão sistemática sobre a estimativa da produtividade da cana-de-açúcar por sensoriamento remoto, Silva et al. (2024) analisaram 72 trabalhos publicados entre 2017 e 2023. Foram encontradas distintas combinações de sensores, índices de vegetação e métodos de modelagem, com utilização dos dados orbitais tanto em modelos empíricos quanto em abordagens mecanísticas. O levantamento mostra também que não existe uma única configuração de dados aplicável a todas as condições de cultivo.

2.2 Índices de vegetação

Nem toda variação de reflectância captada por um satélite corresponde a mudança real na vegetação: solo exposto, sombra e umidade também alteram o sinal. Os índices de vegetação existem justamente para isolar esse ruído, combinando as bandas do vermelho e do infravermelho próximo de forma a realçar a resposta da cultura. Com destaque para três desses índices, NDVI, MSAVI2 e EVI, escolhidos por corrigirem, cada um à sua maneira, interferências do solo e da atmosfera sobre o sinal da vegetação.

O Índice de Vegetação por Diferença Normalizada (NDVI), proposto por Rouse et al. (1974), relaciona as reflectâncias do vermelho e do infravermelho próximo. É um dos índices mais empregados no monitoramento da vegetação, com valores que variam entre -1 e 1 e que podem ser relacionados às diferenças na quantidade e nas condições da cobertura vegetal.

Outros índices foram desenvolvidos buscando reduzir algumas interferências presentes na resposta espectral. O Índice de Vegetação Ajustado ao Solo Modificado 2 (MSAVI2) procura minimizar principalmente os efeitos da exposição do solo sobre a resposta da vegetação (Qi et al., 1994). Já o Índice de Vegetação Melhorado (EVI) incorpora ajustes relacionados às influências do solo e da atmosfera. Em condições de maior biomassa, pode apresentar maior sensibilidade às alterações do dossel em relação ao NDVI (HUETE et al., 2002).

A resposta dos índices também se modifica ao longo do desenvolvimento da cultura. Por esse motivo, o momento de aquisição das imagens pode ser tão importante quanto a escolha do próprio índice quando o objetivo é relacionar as informações espectrais à produtividade. Silva et al. (2024) mostram que índices de vegetação e informações obtidas em diferentes fases do ciclo continuam sendo empregados em estudos de estimativa da produtividade da cana-de-açúcar.

2.3 Variáveis agronômicas e ambientais associadas à produtividade

A produtividade da cana-de-açúcar está relacionada a fatores ambientais e às características da própria cultura. As condições climáticas exercem influência tanto durante o crescimento vegetativo quanto na maturação, enquanto características do solo interferem nas condições disponíveis para o desenvolvimento das plantas (MARCHIORI, 2004; TERAMOTO, 2003).

O relevo também pode interferir nas condições de cultivo e nas operações realizadas nos canaviais. Áreas com declividade suave, entre 2% e 5%, são consideradas adequadas ao cultivo, enquanto terrenos excessivamente planos podem apresentar limitações relacionadas à drenagem do solo (GOUVÊA, 2008).

A resposta da cultura às condições ambientais também varia ao longo de seu desenvolvimento. A cana-de-açúcar passa pelas fases de brotação, perfilhamento, crescimento dos colmos e maturação, com mudanças nas exigências da cultura ao longo do ciclo. O perfilhamento é influenciado por fatores como radiação solar, temperatura e umidade, enquanto, durante a maturação, a redução da disponibilidade hídrica contribui para diminuir o crescimento e favorecer maior concentração de açúcares nos colmos (MARCHIORI, 2004).

A associação de informações relacionadas ao solo, variedade, estágio da cultura, declividade e aspecto aos índices de vegetação permite incorporar à classificação fatores relacionados à produtividade. A utilização conjunta dessas variáveis também possibilita avaliar se as informações espectrais, isoladamente ou combinadas com características agronômicas e ambientais, são suficientes para discriminar as classes de produtividade.

2.4 Random Forest e classificação da produtividade

A produtividade da cana-de-açúcar depende de variáveis de natureza muito distinta entre si, categóricas como tipo de solo e variedade, e contínuas como os índices de vegetação. Isso exige um algoritmo capaz de lidar com essa mistura sem perder desempenho. O Random Forest atende a essa exigência: constrói um conjunto de árvores de decisão a partir de subconjuntos aleatórios das observações e das variáveis disponíveis, combinando os resultados individuais na classificação final (BREIMAN, 2001).

A possibilidade de reunir diversos tipos de informação é especialmente útil em estudos de produtividade agrícola. Características espectrais obtidas por sensores orbitais podem ser analisadas juntamente com dados de solo, relevo, clima, variedade e desenvolvimento da cultura, sem que essas variáveis precisem apresentar o mesmo comportamento ou natureza.

Luciano et al. (2021) utilizaram Random Forest para previsão da produtividade da cana-de-açúcar em escala local no Brasil, combinando imagens Landsat com informações agrônômicas e meteorológicas. A integração dessas fontes apresentou resultados superiores aos obtidos com conjuntos mais restritos de dados. O estudo também avaliou informações adquiridas meses antes da colheita, mostrando a possibilidade de antecipar a estimativa da produtividade.

Mais recentemente, Amaro et al. (2025) trabalharam com dados Sentinel-2 associados a informações agrônômicas e climáticas para estimar a produtividade da cana-de-açúcar antes da colheita. O estudo incluiu índices de vegetação e variáveis como tipo de solo, variedade e declividade, avaliando conjuntos de dados com Random Forest.

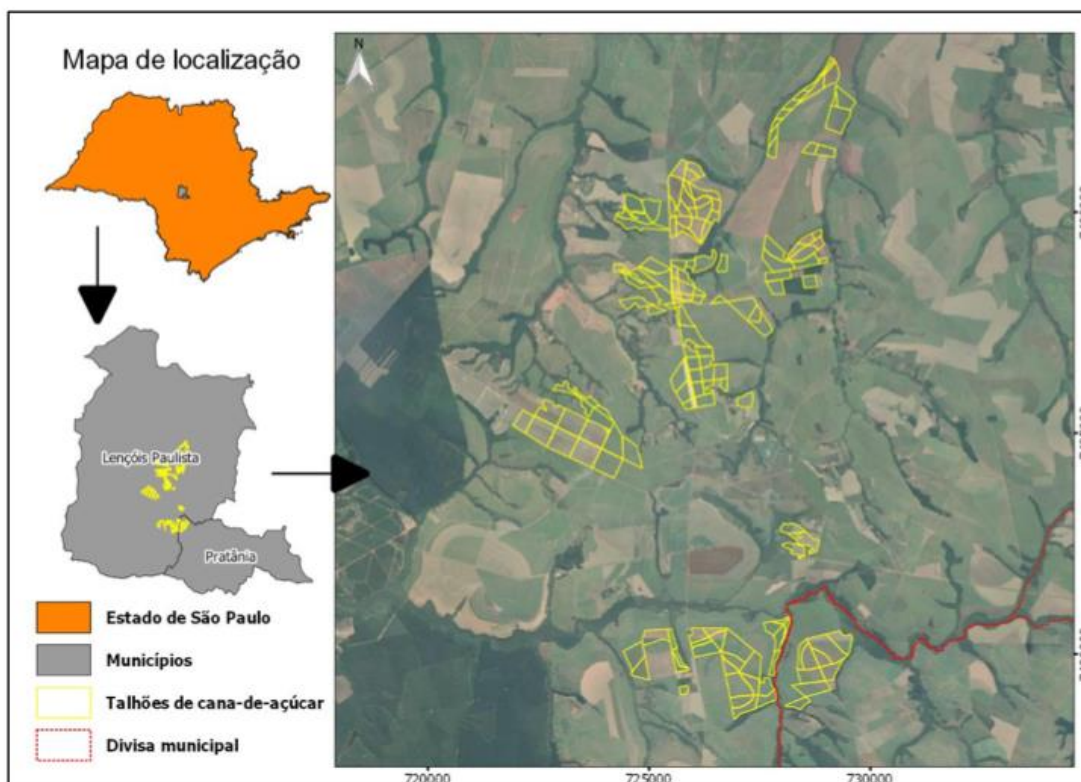
A composição do conjunto de variáveis passa, portanto, a ser uma questão importante nesse tipo de aplicação. Nem sempre a inclusão de todas as informações disponíveis é necessária para manter um desempenho satisfatório. Da mesma maneira, quando índices de vegetação são obtidos em diferentes momentos, torna-se possível avaliar até que ponto informações espectrais adquiridas com maior antecedência em relação à colheita podem ser utilizadas na classificação da produtividade.

3 METODOLOGIA

3.1 Área de estudo e dados utilizados

A área de estudo está localizada nos municípios de Pratânia e Lençóis Paulista, estado de São Paulo, entre as latitudes 22°38'33,14" S e 22°48'47,20" S e as longitudes 48°45'00,46" W e 48°52'10,96" W. A região está inserida no Planalto Ocidental Paulista, com altitudes entre aproximadamente 400 e 700 m (ROSS; MOROZ, 1997). Segundo a classificação de Köppen, o clima é do tipo Cwa em Pratânia e Aw em Lençóis Paulista (CEPAGRI, 2019).

Figura 1. Mapa de localização da área de estudo.



Fonte: Leda (2020).

Na área de estudo predominam Latossolo Vermelho Distrófico (LVd), Latossolo Amarelo Distrófico (LAd), Latossolo Vermelho Eutrófico (LVe) e Latossolo Vermelho-Amarelo Distrófico (LVAd). As informações pedológicas utilizadas foram provenientes de levantamento disponibilizado pelo grupo Zilor em 2018. Para a caracterização do relevo foram utilizadas cartas planialtimétricas produzidas pelo Instituto Geográfico e Cartográfico do Estado de São Paulo (IGC), na escala 1:10.000, datadas de 1978 e com equidistância vertical de 5 m entre as curvas de nível.

Foram utilizadas imagens multiespectrais do sensor Operational Land Imager (OLI), a bordo do satélite Landsat 8, órbita 221 e ponto 076, com resolução espacial de 30 m. As imagens foram obtidas por meio do portal EarthExplorer e correspondem às datas de 11/11/2013, 05/01/2014, 30/01/2014, 10/03/2014, 11/04/2014, 27/04/2014, 14/06/2014, 16/06/2014 e 11/09/2014. Para o cálculo dos índices de vegetação foram utilizadas as bandas 2, 4 e 5.

3.2 Processamento dos dados

O processamento das informações espaciais foi realizado em ambiente QGIS. As curvas de nível das cartas planialtimétricas foram georreferenciadas e vetorizadas, com a atribuição dos respectivos valores altimétricos.

A partir desses dados foi gerado o modelo digital de elevação e, posteriormente, os mapas de declividade e de orientação das vertentes, sendo esta última denominada, neste estudo, aspecto. Os valores foram extraídos em uma grade de 30 x 30 m, compatível com a resolução espacial das imagens Landsat. As informações de solo foram associadas aos talhões da área de estudo.

As imagens Landsat 8 foram utilizadas para o cálculo dos índices NDVI, EVI e MSAVI2. O banco de dados empregado nas classificações reuniu essas informações espectrais e as variáveis tipo de solo, variedade de cana-de-açúcar, estágio da cultura, declividade e aspecto.

Os dados de produtividade foram organizados em categorias com amplitude de 10 t ha⁻¹, abrangendo valores entre 50 e 120 t ha⁻¹. Para representar momentos da safra, a base foi dividida em três grupos, correspondentes aos períodos inicial, intermediário e final.

Em cada grupo foram consideradas três épocas de aquisição das imagens em relação à colheita: aproximadamente 180 dias antes (D-180), aproximadamente 90 dias antes (D-90) e a última imagem disponível antes da colheita (D-1). Para o período inicial, as datas foram 11/11/2013, 05/01/2014 e 11/04/2014, respectivamente. No período intermediário foram utilizadas as cenas de 30/01/2014, 27/04/2014 e 16/06/2014, enquanto no período final foram consideradas as datas de 10/03/2014, 14/06/2014 e 11/09/2014.

3.3 Classificação da produtividade

A classificação da produtividade foi realizada no software WEKA, a partir de arquivos no formato ARFF, utilizando o algoritmo *weka.classifiers.trees.RandomForest*. O banco de dados foi organizado em 12 configurações, permitindo avaliar o desempenho do algoritmo com combinações das variáveis disponíveis.

A configuração 1 reuniu solo, variedade e estágio da cultura, enquanto a configuração 2 acrescentou declividade e aspecto. Nas configurações 3, 4 e 5 foram utilizados solo, variedade, estágio, declividade e aspecto juntamente com os três índices de vegetação obtidos em D-1, D-90 e D-180, respectivamente. Nas configurações 6, 7 e 8 foram mantidos solo, variedade e estágio, associados aos índices de vegetação de cada uma das três épocas. As configurações 9, 10 e 11 combinaram declividade e aspecto aos índices de D-1, D-90 e D-180, respectivamente. A configuração 12 reuniu todas as variáveis e os índices de vegetação das três épocas.

Para avaliação do desempenho foi utilizada validação cruzada em 10 grupos (10-fold cross-validation). Em cada etapa, um dos grupos foi destinado à avaliação e os demais ao treinamento, repetindo-se o procedimento até que todos os grupos fossem utilizados na avaliação. Os resultados foram expressos pelo percentual de classificações corretas e pelo índice Kappa, utilizado para avaliar a concordância entre as classes observadas e aquelas atribuídas pelo modelo.

4 RESULTADOS

Foram avaliados três períodos da safra da cana-de-açúcar, denominados inicial, intermediário e final. Em cada período, o desempenho do algoritmo Random Forest foi analisado considerando as 12 configurações de variáveis estabelecidas.

4.1 Período inicial da safra

No período inicial foram utilizadas 4.263 observações, correspondentes ao mesmo número de pixels. Os resultados das classificações são apresentados na Tabela 1.

Tabela 1 – Resultados obtidos no período inicial de safra (outono) com a metodologia Random Forest e índice Kappa.

		Solo	Variedade	Estágio	Declividade	Aspecto	IVs D - 1	IVs D - 90	IVs D - 180	Classificações Corretas (%)	Classificações Incorretas (%)	Kappa
Análises	1	X	X	X						62,9	37,1	0,5
	2	X	X	X	X	X				79,4	20,6	0,7
	3	X	X	X	X	X	X			84,9	15,1	0,8
	4	X	X	X	X	X		X		84,9	15,1	0,8
	5	X	X	X	X	X			X	85,2	14,8	0,8
	6	X	X	X			X			68,2	31,8	0,6
	7	X	X	X				X		66,6	33,4	0,5
	8	X	X	X					X	69,2	30,8	0,6
	9				X	X	X			72,7	27,3	0,6
	10				X	X		X		71,4	28,6	0,6
	11				X	X			X	76,3	23,7	0,7
	12	X	X	X	X	X	X	X	X	87,9	12,1	0,8

Fonte: Adaptado de Leda (2020).

A análise 12, formada por todas as variáveis disponíveis, apresentou o maior percentual de classificações corretas, com 87,9%, e índice Kappa de 0,8. As análises 3, 4 e 5 apresentaram resultados próximos, com 84,9%, 84,9% e 85,2% de classificações corretas, respectivamente, todas com Kappa igual a 0,8.

A análise 1 apresentou 62,9% de classificações corretas e Kappa de 0,5. Na análise 2, que acrescentou declividade e aspecto às variáveis solo, variedade e estágio, o percentual de classificações corretas foi de 79,4%, com Kappa de 0,7.

Nas análises 6, 7 e 8, os percentuais variaram entre 66,6% e 69,2%, enquanto nas análises 9, 10 e 11 foram obtidos valores entre 71,4% e 76,3%.

Entre as análises com Kappa igual a 0,8, a análise 5, que utilizou os índices de vegetação obtidos aproximadamente 180 dias antes da colheita, apresentou 85,2% de classificações corretas, valor 2,7 pontos percentuais inferior ao observado na análise 12.

4.2 Período intermediário da safra

No período intermediário foram utilizadas 7.252 observações. Os resultados obtidos para as combinações de variáveis são apresentados na Tabela 2.

Tabela 2 – Resultados obtidos no período intermediário de safra (inverno) com a metodologia Random Forest e índice Kappa.

	Solo	Variedade	Estágio	Declividade	Aspecto	IVs D - 1	IVs D - 90	IVs D - 180	Classificações Corretas (%)	Classificações Incorretas (%)	Kappa
Análises	1	X	X	X					66,84	33,16	0,6
	2	X	X	X	X	X			79,62	20,38	0,7
	3	X	X	X	X	X			85,14	14,86	0,8
	4	X	X	X	X	X	X		85,04	14,96	0,8
	5	X	X	X	X	X		X	84,82	15,18	0,8
	6	X	X	X		X			69,26	30,74	0,6
	7	X	X	X			X		69,76	30,24	0,6
	8	X	X	X				X	69,39	30,61	0,6
	9			X	X	X			72,57	27,43	0,7
	10			X	X		X		71,46	28,54	0,6
	11			X	X			X	69,93	30,07	0,6
	12	X	X	X	X	X	X	X	87,38	12,62	0,8

Fonte: Adaptado de Leda (2020).

A análise 12 apresentou o maior percentual de classificações corretas, com 87,38%, e Kappa de 0,8. As análises 3, 4 e 5 alcançaram 85,14%, 85,04% e 84,82%, respectivamente, também com Kappa igual a 0,8.

A análise 1 apresentou 66,84% de classificações corretas e Kappa de 0,6, enquanto a análise 2 atingiu 79,62% e Kappa de 0,7. Nas análises 6, 7 e 8, os percentuais variaram entre 69,26% e 69,76%. As análises 9, 10 e 11 apresentaram valores entre 69,93% e 72,57%.

A diferença entre as análises 12 e 5 foi de 2,56 pontos percentuais. As análises 3, 4 e 5 também apresentaram percentuais próximos, apesar de utilizarem índices de vegetação correspondentes a períodos variados anteriores à colheita.

4.3 Período final da safra

No período final foram utilizadas 4.594 observações, correspondentes aos pixels empregados para extração dos índices de vegetação e das informações associadas.

Tabela 3 – Resultados obtidos no período final de safra (primavera) com a metodologia Random Forest e índice Kappa.

	Solo	Variedade	Estágio	Declividade	Aspecto	IVs D - 1	IVs D - 90	IVs D - 180	Classificações Corretas (%)	Classificações Incorretas (%)	Kappa
Análises	1	X	X	X					61,06	38,94	0,5
	2	X	X	X	X				79,99	20,01	0,7
	3	X	X	X	X	X			82,06	17,94	0,8
	4	X	X	X	X		X		80,78	19,22	0,8
	5	X	X	X	X			X	81,58	18,42	0,8
	6	X	X	X		X			67,50	32,50	0,6
	7	X	X	X			X		64,50	35,50	0,6
	8	X	X	X				X	65,83	34,18	0,6
	9			X	X	X			68,87	31,13	0,6
	10			X	X		X		68,39	31,61	0,6
	11			X	X			X	64,58	35,42	0,6
	12	X	X	X	X	X	X	X	82,93	17,07	0,8

Fonte: Adaptado de Leda (2020).

A análise 12 apresentou o maior percentual de classificações corretas, com 82,93%, e Kappa de 0,8. As análises 3, 4 e 5 alcançaram 82,06%, 80,78% e 81,58%, respectivamente, mantendo o mesmo valor de Kappa.

A análise 1 apresentou 61,06% de classificações corretas e Kappa de 0,5. A análise 2 alcançou 79,99% e Kappa de 0,7. Nas análises 6 a 11, os percentuais de classificações corretas variaram entre 64,50% e 68,87%, com Kappa igual a 0,6.

A diferença entre as análises 12 e 5 foi de 1,35 ponto percentual, inferior às diferenças verificadas nos períodos inicial e intermediário.

Considerando os três períodos, a análise 12 apresentou 87,9%, 87,38% e 82,93% de classificações corretas nos períodos inicial, intermediário e final, respectivamente.

Para a análise 5, os valores foram de 85,2%, 84,82% e 81,58%. Em ambas as análises, o índice Kappa permaneceu em 0,8 nos três períodos avaliados.

5 DISCUSSÃO

Os resultados mostraram que o desempenho do Random Forest variou de acordo com a composição do conjunto de dados utilizado. Em todos os períodos da safra, os maiores percentuais de classificações corretas foram obtidos pela análise 12, que reuniu solo, variedade, estágio da cultura, declividade, aspecto e os índices de vegetação das três épocas. Entretanto, as análises 3, 4 e 5, que mantiveram as variáveis agronômicas e de relevo e utilizaram os índices de uma época, apresentaram resultados próximos ao conjunto completo e Kappa igual a 0,8.

Esse comportamento indica que a utilização conjunta de informações de diferentes naturezas foi mais adequada do que algumas das combinações mais restritas avaliadas. A análise 1, formada por solo, variedade e estágio, apresentou resultados inferiores nos três períodos. A análise 2, que acrescentou declividade e aspecto a esse conjunto, apresentou percentuais maiores de classificações corretas. Entretanto, como as variáveis foram avaliadas em conjuntos, esses resultados não permitem atribuir individualmente o aumento de desempenho à declividade ou ao aspecto, nem estabelecer a importância isolada de cada atributo.

A integração entre informações espectrais e outros dados relacionados à cultura também tem sido utilizada em estudos de estimativa da produtividade da cana-de-açúcar. Luciano et al. (2021) empregaram Random Forest combinando imagens Landsat com informações agronômicas e meteorológicas e verificaram melhor desempenho quando as fontes de dados foram utilizadas conjuntamente. Embora naquele estudo a produtividade tenha sido estimada de forma contínua, enquanto no presente trabalho foi realizada sua classificação em categorias, os resultados convergem quanto à utilidade da integração de tipos distintos de informação.

A composição do conjunto de variáveis também foi avaliada por Amaro et al. (2025), que utilizaram imagens Sentinel-2 associadas a informações agronômicas e climáticas na estimativa da produtividade da cana-de-açúcar. Os autores observaram que a seleção de variáveis pode reduzir o conjunto de entrada sem necessariamente comprometer o desempenho do modelo. No presente estudo, comportamento semelhante pode ser observado nas análises 3, 4 e 5, que utilizaram menos informações espectrais que a análise 12 e mantiveram percentuais próximos de classificações corretas.

Entre essas combinações, a análise 5 apresenta uma característica particularmente relevante para a aplicação dos resultados. Nessa análise, foram utilizados os índices NDVI, EVI e MSAVI2 obtidos aproximadamente 180 dias antes da colheita, juntamente com solo, variedade, estágio da cultura, declividade e aspecto.

Mesmo utilizando somente essa época de aquisição das informações espectrais, foram obtidos 85,2%, 84,82% e 81,58% de classificações corretas nos períodos inicial, intermediário e final, respectivamente, com Kappa igual a 0,8 em todos eles. As diferenças em relação à análise 12 foram de 2,70, 2,56 e 1,35 pontos percentuais.

O momento de aquisição das imagens constitui uma questão importante na utilização do sensoriamento remoto para estimativa de produtividade, pois a resposta espectral da vegetação se modifica ao longo do ciclo da cultura. A revisão realizada por Silva et al. (2024) mostra que estudos de produtividade da cana-de-açúcar têm utilizado períodos variados do ciclo, sensores e combinações de índices de vegetação, sem uma única configuração aplicável a todas as condições de cultivo.

Nesse sentido, o resultado mais relevante deste estudo não está apenas no maior percentual de acerto obtido pela análise 12, mas na proximidade observada quando foram utilizadas informações espectrais adquiridas com maior antecedência. Nas condições avaliadas, a utilização de uma única época de aquisição das imagens, aproximadamente 180 dias antes da colheita, permitiu manter desempenho próximo ao obtido com as três épocas avaliadas. A possibilidade de utilizar informações espectrais com essa antecedência pode contribuir para a obtenção antecipada de informações sobre a produtividade e ampliar sua aplicação no planejamento das operações agrícolas.

Os resultados devem ser considerados no contexto das condições avaliadas neste estudo, que compreendeu uma única área e uma única safra. Além disso, a produtividade foi analisada em classes com amplitude de 10 t ha⁻¹, de modo que os resultados se referem à capacidade de discriminação dessas classes e não à estimativa de valores contínuos de produtividade. A avaliação foi realizada por validação cruzada em 10 grupos, sem a utilização de um conjunto externo e independente de validação. Estudos futuros envolvendo outras áreas, safras e condições de cultivo poderão contribuir para avaliar a estabilidade do desempenho observado e a possibilidade de aplicação do modelo em diferentes condições.

6 CONCLUSÃO

O algoritmo Random Forest apresentou desempenho satisfatório na classificação da produtividade da cana-de-açúcar, destacando-se pela capacidade de integrar informações espectrais, agronômicas e ambientais. Os resultados demonstraram que a utilização conjunta de todas as variáveis proporcionou os maiores percentuais de classificações corretas nos três períodos da safra analisados, evidenciando a importância da combinação de diferentes fontes de informação para melhorar o desempenho do modelo.

Entretanto, a utilização dos índices de vegetação obtidos aproximadamente 180 dias antes da colheita, associados às informações de solo, variedade, estágio da cultura, declividade e aspecto, apresentou resultados próximos aos obtidos com o conjunto completo de variáveis, mantendo índice Kappa de 0,8 nos períodos inicial, intermediário e final.

Dessa forma, verificou-se que uma única época de aquisição de imagens pode ser suficiente para alcançar resultados satisfatórios nas condições avaliadas, reduzindo a necessidade de utilizar imagens de diferentes períodos. Essa possibilidade favorece a obtenção antecipada de informações sobre a produtividade, contribuindo para o planejamento das operações agrícolas, a organização da colheita e a gestão dos recursos disponíveis. Conclui-se, portanto, que a metodologia apresenta potencial para apoiar a tomada de decisões no setor sucroenergético, sendo recomendáveis estudos complementares para avaliar sua aplicação em diferentes condições de cultivo.

REFERÊNCIAS

- AMARO, R. P.; TODOROFF, P.; CHRISTINA, M.; DUFT, D. G.; LUCIANO, A. C. S. Performance evaluation of Sentinel-2 imagery, agronomic and climatic data for sugarcane yield estimation. **Computers and Electronics in Agriculture**, v. 237, 110522, 2025. DOI: 10.1016/j.compag.2025.110522.
- BREIMAN, L. Random forests. **Machine Learning**, v. 45, p. 5-32, 2001. DOI: 10.1023/A:1010933404324.
- CENTRO DE PESQUISAS METEOROLÓGICAS E CLIMÁTICAS APLICADAS À AGRICULTURA – CEPAGRI. **Clima dos municípios paulistas**. Campinas: UNICAMP, 2019. Disponível em: http://www.cpa.unicamp.br/outras-informacoes/clima_muni_305.html. Acesso em: 3 jan. 2019.
- GOUVÊA, J. R. F. **Mudanças climáticas e a expectativa de impactos na cultura de cana-de-açúcar na região de Piracicaba-SP**. 2008. Dissertação (Mestrado em Agronomia) – Universidade de São Paulo, Piracicaba, 2008.
- 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**, v. 83, n. 1-2, p. 195-213, 2002. DOI: 10.1016/S0034-4257(02)00096-2.
- LEDA, V. C. **Modelagem da produtividade por meio de dados edafoclimáticos e orbitais**. 2020. 72 f. Tese (Doutorado em Agronomia – Energia na Agricultura) – Faculdade de Ciências Agrônômicas, Universidade Estadual Paulista, Botucatu, 2020.
- LUCIANO, A. C. S.; PICOLI, M. C. A.; DUFT, D. G.; ROCHA, J. V.; LEAL, M. R. L. V.; LE MAIRE, G. Empirical model for forecasting sugarcane yield on a local scale in Brazil using Landsat imagery and random forest algorithm. **Computers and Electronics in Agriculture**, v. 184, 106063, 2021. DOI: 10.1016/j.compag.2021.106063.
- MARCHIORI, L. F. S. **Influência da época de plantio e corte na produtividade da cana-de-açúcar**. 2004. 277 p. Tese (Doutorado) – Escola Superior de Agricultura Luiz de Queiroz, Piracicaba, 2004.
- QI, J.; CHEHBOUNI, A.; HUETE, A. R.; KERR, Y. H.; SOROOSHIAN, S. A modified soil adjusted vegetation index. **Remote Sensing of Environment**, v. 48, n. 2, p. 119-126, 1994. DOI: 10.1016/0034-4257(94)90134-1.

ROSS, J. L. S.; MOROZ, I. C. **Mapa geomorfológico do Estado de São Paulo: escala 1:500.000**. São Paulo: Departamento de Geografia-FFLCH-USP/IPT/FAPESP, 1997.

ROUSE, J. W.; HAAS, R. H.; SCHELL, J. A.; DEERING, D. W. Monitoring vegetation systems in the Great Plains with ERTS. In: EARTH RESOURCES TECHNOLOGY SATELLITE-1 SYMPOSIUM, 3., 1973, Washington, D.C. **Proceedings [...]**. Washington, D.C.: NASA, 1974. v. 1, p. 309-317.

SILVA, N. R. F.; CHAVES, M. E. D.; LUCIANO, A. C. S.; SANCHES, I. D.; ALMEIDA, C. M.; ADAMI, M. Sugarcane yield estimation using satellite remote sensing data in empirical or mechanistic modeling: a systematic review. **Remote Sensing**, v. 16, n. 5, 863, 2024. DOI: 10.3390/rs16050863.

TERAMOTO, E. R. **Avaliação e aplicação de modelos de estimativa de produção de cana-de-açúcar (*Saccharum spp.*) baseados em parâmetros do solo e clima**. 2003. 86 f. Tese (Doutorado em Agronomia) – ESALQ/USP, Piracicaba, 2003.