

IDENTIFICAÇÃO DE CULTURA AGRÍCOLA DO TOMATE BASEADO EM IMAGENS MSI/SENTINEL-2

Angela Gabrielly Pires Silva¹, João Victor dos Santos Sobrinho², Giovanna Mota Borges Campos², Édipo Henrique Cremon², Max Well de Oliveira Rabelo², e Marina Alberti Macedo²

¹Programa de Pós-Graduação em Sensoriamento Remoto - PGSER, Divisão de Observação da Terra e Geoinformática - DIOTG, Instituto Nacional de Pesquisas Espaciais – INPE, Av. dos Astronautas, 1758 – 12227-010, São José dos Campos – SP, angela.pires@inpe.br, ²Instituto Federal de Goiás - IFG – Campus Goiânia, Grupo de Estudos em Geomática - GEO, R. 75, 46 - Centro, Goiânia - GO, 74055-110, jvictorss202@gmail.com, giovannacampos214@gmail.com, edipo.cremon@ifg.edu.br, maxwell.rabelo@ifg.edu.br, marina.macedo@ifg.edu.br

RESUMO

Este trabalho teve por objetivo identificar as áreas ocupadas pela cultura agrícola de tomate utilizando técnicas de classificação supervisionada aplicadas às imagens orbitais do sensor MSI (Multispectral Instrument) a bordo dos satélites Sentinel-2A e 2B. A área de estudo foi o município de Cristalina, localizado no leste do estado de Goiás. A classificação foi realizada na plataforma do Google Earth Engine (GEE) com base em amostras com ocorrência do plantio do tomate e sua ausência, divididas em 70% para treinamento/calibração e 30% para validação (teste). No treinamento foram utilizados os algoritmos de classificação *Random Forest* (RF) e *Gradient Tree Boost* (GTB). O RF teve melhor desempenho preditivo (F-score = 0,9745) em relação ao GTB (F-score = 0,9680). Este trabalho indica que é possível realizar o mapeamento da cultura do tomate por sensoriamento remoto orbital com boa exatidão.

Palavras-chave – *Solanum lycopersicum*, Google Earth Engine, Random Forest, Classificação supervisionada, aprendizado de máquina.

ABSTRACT

The main goal of this work was to identify the tomato crop based on supervised classification in MSI (Multispectral Instrument) imagery from the Sentinel-2A and 2B satellites. Study area was Cristalina City, located in the east of the Goiás State (Brazil). The classification was performed on the Google Earth Engine (GEE) platform based on samples with occurrence of tomato planting and its absence, divided into 70% for training/calibration and 30% for validation (test). In the training the Random Forest (RF) and Gradient Tree Boost (GTB) classification algorithms were used. RF had better predictive performance (F-score = 0.9745) in relation to GTB (F-score = 0.9680). This work indicates that it is possible to perform tomato crop mapping by orbital remote sensing with good accuracy.

Key words – *Solanum lycopersicum*, Google Earth Engine, Random Forest, Supervised classification, machine learning.

1. INTRODUÇÃO

O Brasil é um importante competidor no mercado de commodities agrícolas e um forte exportador de tomates

(*Solanum lycopersicum*), produzindo o equivalente a 3,9 milhões de toneladas da cultura, distribuída em uma área de 54.801 hectares no período da safra 2021 [1]. Por sua vez, Goiás é o principal estado produtor e exportador desta cultura agrícola [2].

Por ser uma cultura altamente suscetível ao ataque de pragas e doenças, a cultura do tomateiro tem um período específico no qual é permitido o cultivo. No estado de Goiás o período é estabelecido pela Instrução Normativa nº06/2011 publicada pela Agência Goiana de Defesa Agropecuária [3]. De acordo com o documento, o semeio de tomate rasteiro só poderá ser realizado de primeiro de janeiro a primeiro de junho de cada ano. Logo, o desenvolvimento de técnicas de coleta de dados e métodos que colaborem para identificar esse tipo de cultura, sem a necessidade de enviar equipes até o local, apresentando dados recentes e periodicamente atualizados é de suma importância no setor da agricultura [4].

A área total de cultivo de tomate é variável, devido à rotação sazonal e anual do cultivo, e é influenciada por acordos comerciais entre os setores de produção e agroindústrias. A quantidade de tomate plantado pode ser utilizada pelas associações de produtores para receber apoio econômico de agências governamentais e financeiras. As estimativas precisas do tomate permitem melhorias na organização do ciclo de transformação industrial, tais dados iniciais sobre a produção de tomate podem estabilizar o preço de mercado, pois a incerteza sobre a produção esperada pode causar especulação [5].

Uma das aplicações de interesse estratégico para o levantamento de informações territoriais é o uso de imagens orbitais para o mapeamento de coberturas agrícolas, uma vez que, imagens tomadas destas plataformas conseguem imagear vastas áreas com qualidade e periodicidade a um baixo custo operacional. Imagens multiespectrais provenientes de sensores ópticos são o insumo principal nesses grandes projetos de mapeamento de culturas agrícolas. Essas imagens tiveram um importante salto tecnológico, e dentro deste contexto é possível destacar o sensor MSI (*Multispectral Instrument*) a bordo dos satélites Sentinel-2A e 2B da ESA (Agência Espacial Européia).

A cultura do tomate pode ser considerada um plantio de fácil observação e tem um período de plantio relativamente curto, tais condições facilitarão a obtenção de dados provenientes dos métodos de imageamento. Parte-se do pressuposto de que as conjunções de todos os aspectos citados justificam a aplicação das técnicas de sensoriamento remoto para a detecção da cultura de tomate no estado

de Goiás, promovendo, assim, na indústria, a identificação de produtores e consequentemente a comercialização do produto de forma rápida e eficaz, e, na gestão ambiental, o direcionamento de ações de políticas públicas adequadas tais como a fiscalização e o monitoramento das condições ambientais do local de plantio e entorno. O objetivo deste trabalho foi identificar as áreas ocupadas pela cultura agrícola de tomate utilizando técnicas de classificação supervisionada aplicadas às imagens do sensor MSI/Sentinel-2.

2. MATERIAL E MÉTODOS

2.1. Área de estudo

Como área de estudo foi definido o município de Cristalina (Figura 1) localizado na região leste do estado de Goiás. A altitude do município, sua temperatura amena e a qualidade do solo têm permitido que a região tenha um papel importante na agricultura do país. Cristalina tem a agricultura como principal setor de sua economia. A agricultura é bastante diversificada, sendo o maior produtor de trigo, sorgo, feijão, batata-inglesa, cebola, alho e tomate do estado [6].

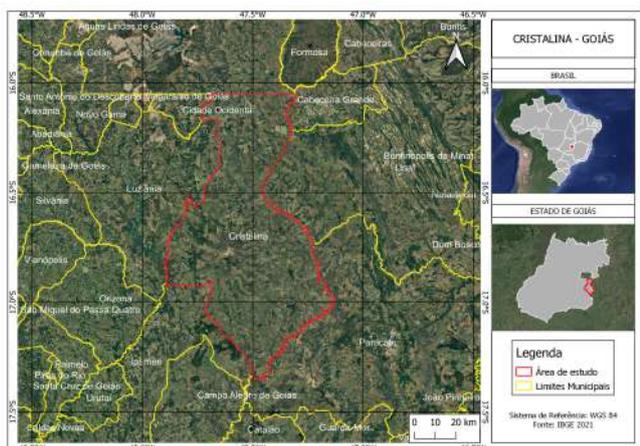


Figura 1: Localização da área de estudo.

2.2. Materiais

Para a realização do estudo foi necessária a aquisição das imagens do sensor MSI/Sentinel-2 com todas as bandas multiespectrais. A missão Sentinel-2 conta com dois satélites gêmeos (Sentinel-2A e 2B) voando na mesma órbita, mas em fase de 180°, fornecendo dados com revisita com 5 dias [7]. O sensor MSI/Sentinel-2 possui 13 bandas espectrais com resolução espacial variando de 60, 20 e 10m (Tabela 1). Todo o processamento das imagens foi realizado na plataforma Google Earth Engine [8] e as figuras finais foram elaborados no *software* QGIS [9]. As imagens do sensor MSI utilizadas estavam corrigidas em reflectância de superfície.

Neste trabalho foi definido o período entre 20/06/2020 e 30/07/2020. A escolha desse intervalo de tempo está relacionada ao período de vazio sanitário para o tomate e ao período compreendido pela semeadura, estabelecimento da cultura em campo e colheita [3]. Também foram utilizadas amostras de locais de plantio de tomate da Agrodefesa do estado de Goiás, cedidos pelo CREA/GO.

Resolução Espacial (m)	Número da banda	Nome da banda
10	B02	Blue
	B03	Green
	B04	Red
	B08	NIR
20	B05	Red Edge 1
	B06	Red Edge 2
	B07	Red Edge 3
	B08A	Red Edge 4
	B11	SWIR 1
	B12	SWIR 2
60	B01	Aerossol
	B09	Water Vapor
	B10	Cirrus

Tabela 1: Bandas espectrais do sensor MSI/Sentinel-2 de acordo com a resolução espacial [7].

2.3. Métodos

Os procedimentos metodológicos realizados no estudo são apresentados na Figura 2.

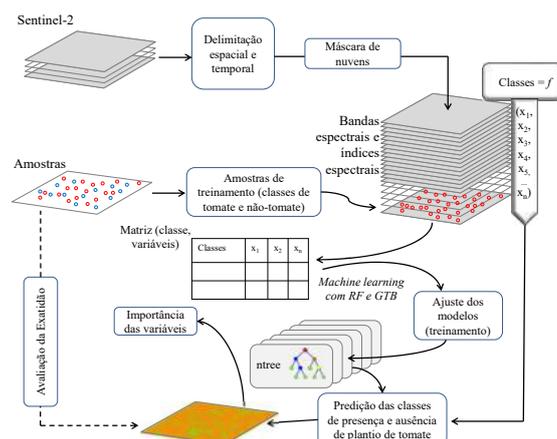


Figura 2: Fluxograma das etapas utilizadas nesse trabalho.

Para as imagens MSI utilizou-se um filtro de nuvem de 5% e para o período analisado foi considerado um filtro redutor temporal de mediana. Ao todo foram utilizadas 1000 amostras, sendo essas utilizadas para treino e validação em que parte foram disponibilizadas pelo CREA/GO (amostras de plantio de tomate) e outra parte foram adquiridas por interpretação visual (outras classes).

Para a classificação supervisionada foram utilizados dois algoritmos baseados em árvore de decisão: *Random Forest* (RF) e *Gradient Tree Boost* (GTB), onde 70% das amostras foram para treinamento e calibração do modelo e 30% para validação da classificação (teste). As métricas de validação utilizadas foram a exatidão global (acurácia), F-score, especificidade e sensibilidade. Para o mapeamento de plantações de tomates foram utilizadas 21 variáveis, com as bandas espectrais (Tabela 1) e índices espectrais (Tabela 2).

Ainda foram analisadas quais variáveis foram as mais importantes na classificação no qual permite ranquear a importância dos algoritmos em valores normalizados entre 0 e 100.

Índice	Fórmula
NDVI	$(B08 - B04) / (B08 + B04)$
EVI2	$2,5 * ((B08 - B04) / (B08 + 2,4 * B04 + 1))$
CAI	$B12 / B11$
NDWI	$(B08 - B11) / (B08 + B11)$
GCVI	$(B08 / B03 - 1)$
Hall Cover	$(-B04 * 0,017 - B08 * 0,007 - B12 * 0,079 + 5,22)$
PRI	$(B02 - B03) / (B02 + B03)$
SAVI	$(1 + L) * (B08 - B04) / (B08 + B04 + 0,5)$

Tabela 2: Índices espectrais utilizados no estudo.

3. RESULTADOS

Todo o código usado no GEE para este trabalho pode ser acessado pelo endereço: <https://bitly.com/beRBuuNp>. Analisando as métricas de validação de ambos os algoritmos classificadores (Tabela 3), pode-se perceber que os algoritmos RF e GTB tiveram desempenhos semelhantes quando comparados com a exatidão global, porém o RF teve uma vantagem na métrica de especificidade, ou seja, o algoritmo RF teve uma capacidade maior em indicar corretamente áreas em que tinha ausência de plantações de tomate quando comparado com GTB. Analisando a métrica de sensibilidade que indica a capacidade de indicar presença do cultivo de tomates pode-se notar que os classificadores tiveram resultados similares.

Algoritmo	RF	GTB
Exatidão Global	0,9589	0,9486
F-score	0,9745	0,9680
Sensibilidade	0,9703	0,9742
Especificidade	0,9107	0,8475

Tabela 3: Métricas de desempenho dos algoritmos RF e GTB.

A área classificada com a presença da cultura de tomate no município de Cristalina-GO, no período analisado, foi de 28351,153 ha com o algoritmo RF e 22221,716 ha com o GTB, correspondendo à 4,61% e 3,61% do território total do município, respectivamente (Figura 3).

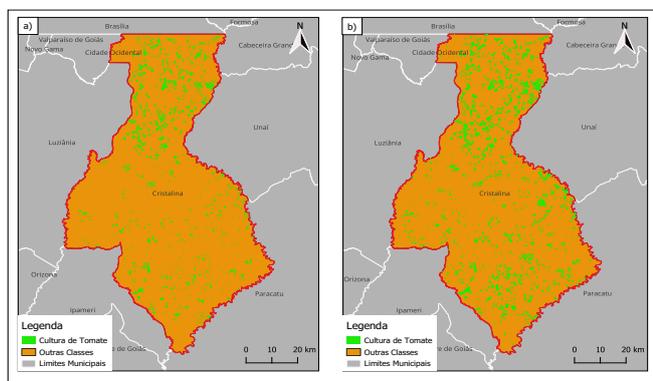


Figura 3: Classificação de cultura de tomate utilizando os algoritmos a) RF e b) GTB.

Categorizando as variáveis por importância (Figura 4) tem-se como as cinco variáveis mais importantes para o algoritmo RF a banda espectral de vapor de água (B9), os índices espectrais Hall Cover e NDWI, duas bandas posicionadas na

borda do vermelho (B7 e B8A) e a banda do infravermelho próximo. Já as cinco variáveis mais importantes para o GTB foram o índice PRI, as bandas do Blue (B02) e de vapor de água (water vapor - B09) e os índices espectrais NDWI e Hall Cover.

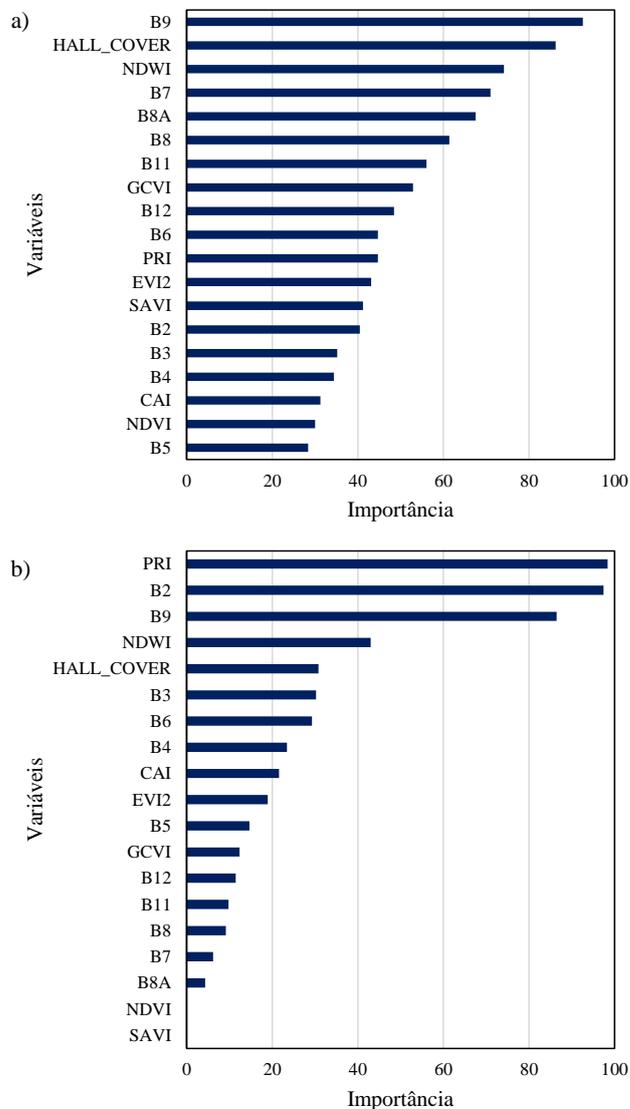


Figura 4: Importância das variáveis espectrais para a classificação utilizando os classificadores a) RF e b) GTB.

4. DISCUSSÃO

Quando comparado a área mapeada pelos classificadores RF e GTB com os dados da Produção Agrícola Municipal [1] para o mesmo ano de 2020, é possível constatar que a área plantada por tomate no município foi de 3500 ha, ou seja, os classificadores indicaram uma área mais de seis vezes maior ao reportado em dados oficiais.

Sobre o desempenho entre o RF e GTB, a abordagem de classificação por árvore de decisão utilizando *boosting*, encaixam-se sucessivamente árvores de decisão rasas, cada uma tendo em conta o erro de classificação das árvores anteriores para construir um conjunto forte de aprendizes

fracos. Em contraste, o RF consiste em muitas árvores de decisão profundas, mas descorrelacionadas, construídas com diferentes amostras dos dados [10].

Na literatura, comparações entre RF e GTB o primeiro apresentou desempenho melhor em relação ao segundo [10, 11]. Além do desempenho entre os algoritmos, o gráfico de importância de variáveis elucidou a diferença em como o RF e o GTB utilizam as variáveis preditivas. O GTB teve uma concentração de importância em um pequeno número de variáveis enquanto o RF teve uma distribuição mais homogênea entre as variáveis. Isso se deve pelo fato que no algoritmo RF cada árvore de decisão é independente, então se as variáveis têm uma correlação maior, a importância é dividida entre elas. Já no GTB as árvores são construídas sucessivamente, então a primeira árvore tem maior importância e as demais que são construídas sobre a árvore anterior são altamente correlacionadas e consequentemente pela repetição de informações há uma menor contribuição das demais variáveis [11].

5. CONCLUSÕES

Neste trabalho foram analisados dois métodos diferentes para a classificação de áreas de plantio de tomate com base em imagens MSI/Sentinel-2. O primeiro método foi baseado no algoritmo RF e o segundo método no GTB. O algoritmo RF obteve desempenho ligeiramente superior ao GTB.

Com base nos resultados conclui-se que é possível realizar o mapeamento da cultura do tomate por sensoriamento remoto orbital via imagens do sensor MSI/Sentinel-2 com boa acurácia. Com isso, é possível ter ciência de áreas produtoras de tomates que constavam nos dados para controle fitossanitário. Adicionalmente, empresas de crédito agrícola podem fazer uso dessas informações para monitoramento das áreas das propriedades que receberam crédito envolvendo o cultivo do tomate.

6. REFERÊNCIAS

- [1] Instituto Brasileiro de Geografia e Estatística (IBGE). *Levantamento Sistemático da Produção Agrícola*. Instituto Brasileiro de Geografia e Estatística (IBGE), 2021.
- [2] Companhia Nacional de Abastecimento CONAB. Tomate: análise dos indicadores da produção e comercialização no mercado mundial, brasileiro e catarinense. *Compêndio de estudos Conab*, 21:11–12, 2020.
- [3] Agrodefesa. *Instrução Normativa nº 6 de 14 de Junho de 2011*. Agrodefesa, 2011.
- [4] J. Inglada, M. Arias, B. Tardy, O. Hagolle, S. Valero, D. Morin, G. Dedieu, G. Sepulcre, S. Bontemps, P. Defourny, et al. Assessment of an operational system for crop type map production using high temporal and spatial resolution satellite optical imagery. *Remote Sensing*, 7(9):12356–12379, 2015.
- [5] G. V. Laurin, C. Belli, R. Bianconi, P. Laranci, and D. Papale. Early mapping of industrial tomato in central and southern Italy with Sentinel 2, aerial and Rapideye additional data. *The Journal of Agricultural Science*, 156(3):396–407, 2018.
- [6] IMB. *Painéis Municipais – Sínteses Municipais*. Instituto Mauro Borges, 2022.
- [7] ESA Sentinel. User handbook. *ESA Standard Document*, 64, 2021.
- [8] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore. Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote sensing of Environment*, 202:18–27, 2017.
- [9] QGIS Development Team. *QGIS Geographic Information System*. Open Source Geospatial Foundation, 2022.
- [10] C. Krauss, X. A. Do, and N. Huck. Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *European Journal of Operational Research*, 259(2):689–702, 2017.
- [11] E. A. Freeman, G. G. Moisen, J. W. Coulston, and B. T. Wilson. Random forests and stochastic gradient boosting for predicting tree canopy cover: comparing tuning processes and model performance. *Canadian Journal of Forest Research*, 46(3):323–339, 2016.