

# CLASSIFICAÇÃO DOS ESTÁGIOS DE CORTE DA CANA-DE-AÇÚCAR A PARTIR DE IMAGENS SENTINEL-2 E APRENDIZAGEM DE MÁQUINA

Hugo De Melo Fukuzava<sup>1</sup>, Ana Cláudia dos Santos Luciano<sup>2</sup>

<sup>1</sup> Universidade de São Paulo, Av. Pádua Dias, 11 - Agronomia, Piracicaba - SP, 13418-900, hugo.melo.fukuzava@usp.br; <sup>2</sup> Universidade de São Paulo Av. Pádua Dias, 11 - Agronomia, Piracicaba - SP, 13418-900 analuciano@usp.br

## RESUMO

Devido as grandes extensões do território brasileiro e da importância da cana-de-açúcar, é necessário o monitoramento da cultura. O objetivo do trabalho foi utilizar séries temporais de Sentinel-2 e algoritmos de aprendizado de máquina para identificar diferentes estágios da cana-de-açúcar: cana planta; cana soca e cana reforma. Para isso, foi utilizado o pacote *Satellite Image Time Series* do software R e o algoritmo Random Forest. Foram selecionadas as bandas do vermelho (R), verde (G), azul (B) e a borda do vermelho (Rededge). A calibração do algoritmo de classificação foi feita a partir de amostras de referência de mapas de uso e cobertura da terra. Criou-se três modelos distintos, o primeiro com as bandas RGB, o segundo com RGB+Rededge e, o terceiro apenas Rededge. A acurácia da classificação RGB foi 86.5%, RGB+Rededge 88.9 e Rededge 88.3%. A combinação das bandas RGB com o Rededge trouxe ganhos na identificação da cana soca e reforma.

**Palavras-chave** – índices de vegetação, mapeamento, sensoriamento remoto, aprendizado de máquina

## ABSTRACT

*Due to the large extensions of the Brazilian territory and the importance of sugarcane, it is necessary to monitor the sugarcane crop. The objective of this work was to use Sentinel-2 time series and machine learning algorithms to identify different stages of sugarcane: plant cane; ratoon cane and sugarcane reform. The Satellite Image Time Series package of the R software and Random Forest algorithm was used for this. The red (R), green (G), blue (B), and rededge bands were selected. The calibration of the classification algorithm was done from reference samples of land use and land cover maps. Three distinct models were created, the first with the RGB bands, the second with RGB+Rededge, and the third with Rededge only. The accuracy of the RGB classification was 86.5%, RGB+Rededge 88.9, and Rededge 88.3%. The combination of the RGB bands with Rededge brought gains in the identification of ratoon and ratoon reform.*

**Keywords** — vegetation index, mapping, remote sensing, machine learning

## 1. INTRODUÇÃO

A cana-de-açúcar tem grande destaque no agronegócio e na economia brasileira, tendo aproximadamente 83 milhões de hectares plantados [1], os quais representam cerca de 10% de toda área agrícola do país. Apesar da vasta área plantada, a tendência ainda é de expansão, com projeção de aumento das áreas em um milhão de hectares para a safra 2022/2023 [2]. Quando olhamos mais especificamente para o estado de São Paulo, a região é responsável por 30% do Produto Interno Bruto de todo país (PIB) e, destes 10% correspondem somente ao setor agropecuário. A cana-de-açúcar vem expandindo na região sudeste para o oeste do estado desde 2013 e a sua área plantada está substituindo as áreas de pastagens no estado [3], aumentando assim sua importância não só na economia de São Paulo como também na economia nacional.

Pesquisas sobre o uso dos dados de sensoriamento remoto e técnicas de monitoramento da cobertura e do uso do solo vêm sendo amplamente aplicadas à cana-de-açúcar, dado sua evidente importância, gerando inovações e avanços tanto na automação desses processos quando nas metodologias para aplicá-los [4]. Dentre as aplicações do sensoriamento remoto, está o monitoramento de extensas áreas agrícolas, que colabora para obtenção de informações úteis para formulação de políticas socioeconômicas, análise de dados ambientais e também previsões da produção futura. Uma importante ferramenta para medir a capacidade produtividade da cana é entender qual o corte da cultura. Por exemplo, a cana em reforma e a cana planta terão desempenhos de produtividade diferentes, por isso, é importante identificar os diferentes estágios de desenvolvimento da cultura para um monitoramento mais preciso.

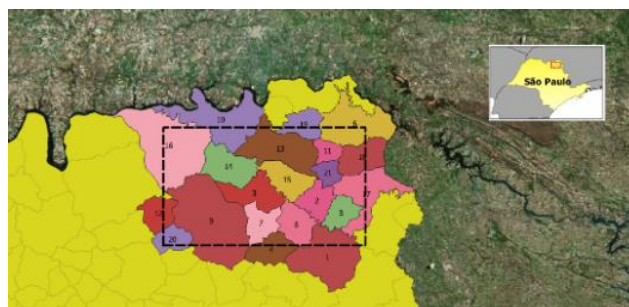
Para a identificação da cana-de-açúcar e seus estágios de corte é possível utilizar séries temporais de imagens e algoritmos de classificação, como os de aprendizado de máquina [5]. Diversos estudos têm utilizado algoritmos de aprendizado de máquina e imagens de satélite para mapeamento de áreas agrícolas em escalas locais e regionais [6,7,8,9] e, os resultados indicam avanços em termos de redução de tempo, custo e recursos quando comparado aos métodos tradicionais que utilizam de forma intensa levantamentos em campo e recursos humanos.

Neste contexto, o presente trabalho tem por objetivo utilizar séries temporais de imagens Sentinel-2 e algoritmos de aprendizado de máquina para identificar diferentes

estágios de desenvolvimento da cana-de-açúcar: cana planta; cana soca e cana reforma na região nordeste do estado de São Paulo.

## 2. MATERIAL E MÉTODOS

A área de estudo tem aproximadamente 670.000 hectares e, está situada na região nordeste do estado de São Paulo. (Figura 1).



**Figura 1: Área de estudo composta por 21 municípios no Estado de São Paulo.**

Para a classificação das áreas agrícolas foram utilizadas imagens do satélite Sentinel-2 obtidas pelo R através do repositório em nuvem do serviço AWS (*Amazon Web Services*). Foram utilizadas as bandas vermelho (R), verde (G), azul (B) e RedEdge para o ano de 2017. Além disso, foram obtidas amostras de uso e cobertura da terra contendo 6 classes: 1) agricultura anual; 2) áreas naturais; 3) pastagens; 4) áreas de cana soca (segundo ao sexto corte); 5) cana planta e 6) áreas de cana em reforma. As amostras possuem localização espacial com latitudes e longitudes referentes a uma data de início (01-01-2016) e fim (01-06-2017) do ciclo agrícola. As amostras de cana-de-açúcar foram obtidas de mapas de referência do projeto Canasat e, as amostras de áreas agrícola, pastagem e natural por meio dos mapeamentos dos projetos TerraClass e LAPIG 2015.

As amostras foram balanceadas, para isso foram aplicadas duas técnicas sendo estas down sampling e aplicação do algoritmo *Self Organization Map* (SOM). O SOM é um algoritmo de agrupamentos de dados, o qual consiste em uma análise envolvendo redes neurais.

Para a classificação foi utilizado o pacote de código aberto SITS [10] (*Satellite Image Time Series*) implementado no software estatístico R. Esse pacote permite o uso de ferramentas estatísticas de aprendizagem de máquina com análise de séries temporais de imagens incluindo métodos como discriminação linear e análise quadrática, Suport Vector Machine, Random Forest, Boosting, Deep Learning, Convolução Neural.

Foram criados três modelos de classificação, considerando os seguintes conjuntos de dados: 1) as bandas RGB, 2) combinação das bandas RGB + Rededge e, 3) apenas a banda Rededge. Para o treinamento dos modelos foram

utilizadas 2100 amostras de referência, destas 1680 foram utilizadas para o treinamento e 420 para testes sendo distribuídas entre as diferentes classes e aplicado uma validação cruzada com  $k = 10$ . Para a classificação foi utilizado algoritmo Random Forest (RF) treinado com parâmetro de 100 árvores. Os modelos foram avaliados em relação a acurácia global e as acurácias do produtor e usuário.

## 3. RESULTADOS

A acurácia global da classificação obtida pelos 3 modelos testados foi de 86.5%, 88.9% e 88.3% para o modelo RGB, RGB+rededge e rededge, respectivamente.

O modelo com combinação das bandas RGB+Rededge foi o que apresentou maior acurácia global. Apesar disso, a classe de cana planta apresentou menores acurácias do produtor com este modelo. O modelo que apresentou melhores acurácias do produtor foi o Rededge (Tabela 1). No modelo Rededge a classe em estágio de reforma obteve aproximadamente 90% de acertos comparado a 80.5% no RGB e RGB+Rededge. O mesmo ocorreu com a acurácia do usuário no estágio de reforma, sendo RGB e RGB+Rededge com 80.5% e 86.8% para 94.4% no Rededge.

Acurácia do Produtor			
Classes	RGB (%)	RGB+Rededge (%)	Rededge (%)
Cana Planta	68.18	61.36	67.21
Cana Soca	72.73	76.36	76.91
Reforma	80.49	80.49	91.07
Acurácia do Usuário			
Classes	RGB (%)	RGB+Rededge (%)	Rededge (%)
Cana Planta	78.95	79.41	74.55
Cana Soca	76.92	75.00	69.77
Reforma	80.49	86.84	94.44
Acurácia Global			
	86.47	88.89	88.33

**Tabela 1: Valores de acurácia global, do produtor e do usuário para classificação random forest com os diferentes modelos (RGB, RGB+Rededge, Rededge).**

Os estágios de cana soca e cana planta apresentaram valores semelhantes para todas as classificações ficando acima de 61%, com médias próximas a 70% para a acurácia do produtor e usuário (Tabela 1).

## 4. DISCUSSÃO

Os resultados destacam uma boa separação entre as diferentes classes de cana-de-açúcar; áreas agrícolas e as outras classes de uso e cobertura da terra. É verificado que a inserção da banda rededge traz uma melhoria ao classificador na identificação de novas áreas de reforma. Tal fato pode ser explicado pelo rededge ser um ótimo identificador de

clorofila [11] o que pode explicar este comportamento, uma vez que a reforma é a retirada da cana de açúcar do campo. Quando nos referimos a ganhos em termos de processamento temos que RGB+Rededge necessitou do dobro em dados para se obter os mesmos resultados. Para maioria das classes a acurácia do produtor e do usuário apresentaram os mesmos valores com exceção do rededge no qual teve um aumento de cerca de 10% na acurácia do produtor. Resultados semelhantes foram obtidos em outros trabalhos, com acurácia global acima de 90% para a classificação de cana-de-açúcar[12].

## 5. CONCLUSÕES

Os resultados mostraram que o uso do algoritmo de aprendizado de máquina Random Forest em conjunto com imagens do satélite Sentinel-2 e a banda Rededge possibilita boa discriminação entre os diferentes estágios de desenvolvimento da cana-de-açúcar. A utilização da banda Rededge foi essencial para discriminação das classes de cana reforma e cana plana.

Por fim, a proposta neste trabalho pode trazer melhorias na criação de mapas de uso e cobertura da terra para áreas de cana-de-açúcar e, especialmente, para o monitoramento de estágios de desenvolvimento da cultura (cana-planta, cana soca e cana reforma) e culturas agrícolas, auxiliando no monitoramento da taxa de renovação dos canaviais.

## 6. REFERÊNCIAS

[1] IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. LSPA - Levantamento Sistemático da Produção Agrícola. Rio de Janeiro: IBGE, 2022. Disponível em: <https://www.ibge.gov.br/estatisticas/economicas/agricultura-epecuaria/9201-levantamento-sistematico-da-producao-agricola.html?=&t=resultados>. Acesso em: 29 jan. 2022.

[2] CONAB – COMPANHIA NACIONAL DE ABASTECIMENTO. - Acompanhamento da safra brasileira de cana-de-açúcar | v.8 – safra 2021/22, nº3 terceiro levantamento | novembro 2021. Disponível em: <https://www.conab.gov.br/infoagro/safra/cana/boletim-da-safra-de-cana-de-acucar>. Acesso em: 12 fev. 2022.

[3] R. T. Nonato, S. R. D. M. Oliveira.. Técnicas de mineração de dados para identificação de Áreas com cana-de-açúcar em imagens Landsat 5. Engenharia Agrícola,33:1268 – 1280.

[4] R. Simoes, G. Camara, G. Queiroz, F. Souza, P. R. Andrade, L. Santos, A. Carvalho, K. Ferreira. Satellite Image Time Series Analysis for Big Earth Observation Data. Remote Sensing, [S.L.], v.13, n. 13, p. 2428, 22 jun. 2021.

[5] A. C. S. Luciano, M. C. A. Picolli, J. V. Rocha, H. C. J. Franco, G. M. Sanchez, M. R. L. V Leal, G. L. Maire. Generalized space-

time classifiers for monitoring sugarcane areas in Brazil. Remote Sensing of Environment. <https://doi.org/10.1016/j.rse.2018.06.017>.

[6] M. C. A. Picolli, G. Camara, I. Sanchez, R. Simoes, Carvalho, A., Coutinho, A., Esquerdo, J., Antunes, J., Begotti, R. A., Arvor, C. Almeida. (2018). Big earth observation time series analysis for monitoring Brazilian agriculture. ISPRS Journal of Photogrammetry and Remote Sensing, August, 1–12. <https://doi.org/10.1016/j.isprsjprs.2018.08.007>.

[7] J. Xiong, P. S. Thenkabail, M. K. P. Gummateluguntla, J. Poehnel, R. G. Congalton, K. Yadav, D. Thau, (2017). Automated cropland mapping of continental Africa using Google Earth Engine cloud computing. ISPRS Journal of Photogrammetry and Remote Sensing, 126, 225–244. <https://doi.org/10.1016/j.isprsjprs.2017.01.019>

[8] L. Zhong, L. Hu, L. Yu, P. Gong, G. S. Biging, (2016). Automated mapping of soybean and corn using phenology. ISPRS Journal of Photogrammetry and Remote Sensing, 119, 151–164. <https://doi.org/10.1016/j.isprsjprs.2016.05.014>

[9] B. F. T. Rudorff, D. A. Aguiar, W. F. Silva, L. M. Sugawara, M. Adamii, M. A. Moreira. Studies on the Rapid Expansion of Sugarcane for Ethanol Production in São Paulo State (Brazil) Using Landsat Data. Remote Sensing, [S.L.], v. 2, n. 4, p. 1057-1076, 9 abr. 2010. MDPI AG. <http://dx.doi.org/10.3390/rs2041057>.

[10] R. Simoes, G. Camara, G. Queiroz, F. Souza, P. R. Andrade, L. Santos, A. Carvalho, K. Ferreira. “Satellite Image Time Series Analysis for Big Earth Observation Data”. Remote Sensing, 13: 2428, 2021. doi:10.3390/rs13132428.

[11] C. H. W. Souza, R. A. C. Lamparelli. D. D. D. Justina, J. V. Rocha, Estudo da Banda Rededge do satélite Rapideye na discriminação da cobertura vegetal, SSB, 3612, 2015

[12] A. C. S. Luciano, H. M. Fukuzava, Classificação de áreas de cana-de-açúcar no Estado de São Paulo a partir de séries temporais de imagens Landsat e algoritmos de aprendizagem de máquina, SIICUSP, 1, 2021