

ESTIMATIVA DE CARBONO TOTAL NO SOLO A PARTIR DE IMAGENS MULTIESPECTRAIS UTILIZANDO ALGORITMOS DE APRENDIZADO DE MÁQUINA

Rayssa Santos Barbosa¹, Maria Cecília Lemes Santos¹, Bruna Ferreira dos Santos², Filipe Vieira da Silva², George Deroco Martins³, Marcela Amorim Melo⁴, Angélica Cristina Moreira Tomas⁴, Verônica Martins de Lima Ferreira⁴, Daiane Gabriela Ribeiro⁵, Edmar Isaías de Melo⁵

¹Engenharia de Agrimensura e Cartográfica, Universidade Federal de Uberlândia (UFU), Monte Carmelo – MG, Brasil, (rayssabsantos, maria.lemes)@ufu.br; ²Programa de Pós-graduação em Agricultura e Informações Geoespaciais, Universidade Federal de Uberlândia (UFU), Monte Carmelo – MG, Brasil, (lipesilva, brufe)@ufu.br; ³Instituto de Geografia, Universidade Federal de Uberlândia (UFU), deroco@ufu.br; ⁴Instituto de Ciências Agrárias, Universidade Federal de Uberlândia (UFU), Uberlândia – MG, Brasil, (marcela.melo, angelica.tomas, veronica.ferreira)@ufu.br; ⁵Instituto de Química, Universidade Federal de Uberlândia (UFU), Uberlândia – MG, Brasil, (dgribeiro,emelo)@ufu.br.

RESUMO

O carbono no solo é um indicador crucial da fertilidade e da sustentabilidade dos ecossistemas agrícolas. Este estudo teve como objetivo estimar o carbono total no solo por meio de dados multiespectrais e algoritmos de aprendizado de máquina, fornecendo uma alternativa eficiente às análises laboratoriais tradicionais. Amostras de solo foram coletadas sob diferentes sistemas de uso em propriedades localizadas nos municípios de Monte Carmelo, Patrocínio e Serra do Salitre, MG, Brasil, e imagens proximais foram capturadas com a câmera Mapir Survey 3W. Os dados radiométricos extraídos das bandas espectrais foram utilizados como variáveis de entrada em modelos preditivos. Três algoritmos de aprendizado de máquina foram comparados: Random Forest, Multilayer Perceptron e Random Tree. O algoritmo Random Forest apresentou o melhor desempenho, alcançando R² de 0,79 e RMSE de 30%, demonstrando potencial na estimativa de carbono a partir de dados de sensoriamento remoto. Esses resultados destacam a aplicabilidade do aprendizado de máquina como ferramenta eficiente na análise de solos e no monitoramento ambiental, facilitando a adoção de práticas agrícolas sustentáveis.

Palavras-chave — Teor de carbono, Sensoriamento Remoto, *Random Forest*, Imagens Multiespectrais.

ABSTRACT

Soil carbon is a crucial indicator of the fertility and sustainability of agricultural ecosystems. This study aimed to estimate total soil carbon using multispectral data and machine learning algorithms, providing an efficient alternative to traditional laboratory analyses. Soil samples were collected under different land use systems on properties located in the municipalities of Monte Carmelo, Patrocínio and Serra do Salitre, MG, Brazil, and proximal images were captured with the Mapir Survey 3W camera. The radiometric data extracted from the spectral bands were used as input variables in predictive models. Three machine learning

algorithms were compared: Random Forest, Multilayer Perceptron and Random Tree. The Random Forest algorithm presented the best performance, reaching R² of 0.79 and RMSE of 30%, demonstrating potential in estimating carbon from remote sensing data. These results highlight the applicability of machine learning as an efficient tool in soil analysis and environmental monitoring, facilitating the adoption of sustainable agricultural practices.

Keywords — *Carbon content, Remote Sensing, Random Forest, Multispectral Imaging.*

1. INTRODUÇÃO

O solo é um recurso natural essencial para a sustentabilidade da produção agrícola, armazenando nutrientes indispensáveis para o desenvolvimento das plantas e atuando como um dos maiores reservatórios de carbono na superfície terrestre [1]. A presença de carbono no solo é fundamental para indicar a fertilidade e a saúde dos ecossistemas, além de ter um papel relevante na mitigação das mudanças climáticas [2].

A determinação do teor de carbono é geralmente realizada por meio de análises laboratoriais, como a oxidação via dicromato de potássio. No entanto, essas técnicas envolvem custos elevados e são laboriosas, o que limita sua aplicação em larga escala [3].

O sensoriamento remoto apresenta uma alternativa promissora, permitindo a obtenção de informações agronômicas por meio da análise de imagens e identificando variações nos atributos químicos do solo. Em particular, o uso de algoritmos de aprendizado de máquina para processar dados multiespectrais oferece um caminho inovador para estimar o carbono no solo de maneira indireta e eficiente, associando informações radiométricas a medições laboratoriais. [4].

Dessa forma, este estudo teve como objetivo estimar o carbono total em solos de diferentes usos da terra, utilizando dados radiométricos obtidos por meio de imagens capturadas com a câmera Mapir Survey 3W e processadas por algoritmos de aprendizado de máquina. Foram comparados os algoritmos Random Forest, Multilayer Perceptron e Random

Tree para avaliar sua utilização na predição dos valores de carbono, destacando-se o Random Forest como a metodologia mais eficiente para essa aplicação.

2. MATERIAL E MÉTODOS

Neste estudo, foram utilizadas 207 amostras de solo coletadas em propriedades localizadas nos municípios de Monte Carmelo, Patrocínio e Serra do Salitre (Figura 1). As amostras são originadas de diversos usos do solo, como cafeeiro, cerrado, milho, pastagem, eucalipto, soja, restauração, pomar, sorgo, Integração Lavoura Pecuária Floresta (ILPF) e Sistema Agroflorestal (SAF).

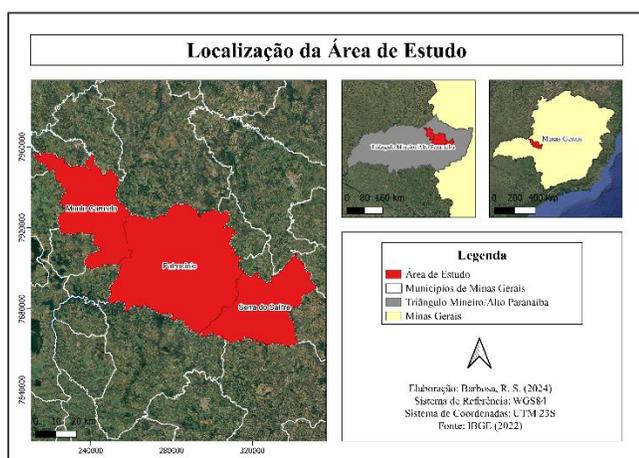


Figura 1. Mapa de localização contendo os municípios onde foram coletadas as amostras.

Dessa forma, realizou-se a captura de imagens proximais utilizando a câmera Mapir Survey 3W, equipada com três bandas espectrais compreendidas pelas faixas do vermelho, verde e infravermelho. Para tomada das imagens, a câmera foi fixada em uma haste, enquanto as amostras foram posicionadas sobre fundo preto, visando a padronização do processo (Figura 2).



Figura 2. Processo de captura das imagens por meio da câmera Mapir Survey 3W.

O processamento das imagens ocorreu no software ENVI Basic, onde foram coletados dados radiométricos através da ferramenta ROI Tool. Neste processo foi confeccionada uma ROI (*Region Of Interest*) de referência em formato circular, abrangendo aproximadamente 50.000 pixels (Figura 3).

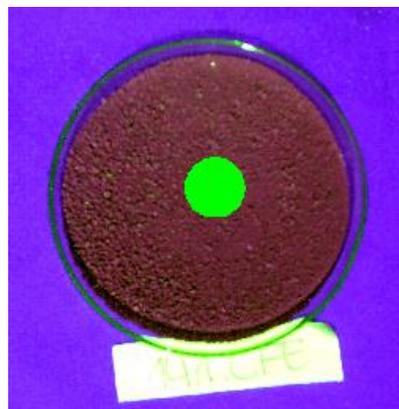


Figura 3. Coleta de dados radiométricos no ENVI, onde em verde encontra-se a ROI referência e abaixo uma amostra de solo.

Assim, extraiu-se a mesma quantidade de pixels de todas as imagens, e esses dados foram exportados para o Excel com objetivo de calcular a média do conjunto de dados de cada amostra, de modo que, ao final do processo, cada amostra apresentasse uma média representativa. No ambiente Excel, também foram calculados índices de vegetação (IVs) a partir das bandas disponíveis, com intuito de melhorar o modelo preditivo (Tabela 1).

Índice	Fórmula	Referência
NDVI	$\frac{(NIR - RED)}{(NIR + RED)}$	[5]
NDCCI	$\frac{(NIR - GREEN)}{(NIR + GREEN)}$	[6]
GSAVI	$\frac{(NIR - GREEN)}{(NIR + GREEN + 0,5)} (1,5)$	[7]
GOSAVI	$\frac{(NIR - GREEN)}{(NIR + GREEN + 0,16)}$	[7]

Tabela 1. IVs utilizados para confecção do modelo.

Após a elaboração das 207 médias, os valores radiométricos foram associados aos respectivos teores de carbono total obtidos em laboratório e divididos em 80% para treinamento e 20% para validação, de forma aleatória, conforme método *Holdout* [8], resultando em 166 amostras para treinamento e 41 amostras para validação. Em seguida, ambos os conjuntos de dados foram convertidos para o formato padrão do *software* Weka (.arff), que é utilizado para processamento de dados a partir de algoritmos de

aprendizado de máquina e avaliação do desempenho, sem a utilização de quaisquer códigos de programação [9].

Sendo assim, o arquivo correspondente a 80% das amostras foi adicionado na aba inicial do software Weka, enquanto o arquivo de 20% foi inserido na aba *Classify* por meio da ferramenta *Supplied test set*. Foram testados diversos algoritmos de aprendizado de máquina, incluindo *Random Forest*, *Random Tree* e *Multilayer Perceptron*. Os resultados de cada algoritmo foram avaliados com base no Coeficiente de Determinação (R^2) e na Raiz do Erro Quadrático Médio (RMSE), permitindo identificar o algoritmo com o melhor desempenho na estimativa do carbono total nas amostras analisadas.

3. RESULTADOS E DISCUSSÃO

A Tabela 2 apresenta os resultados da aplicação dos algoritmos de aprendizado de máquina *Random Forest*, *Multilayer Perceptron* e *Random Tree* na estimativa de carbono total no solo. Os indicadores utilizados para avaliar o desempenho dos modelos foram o **Coeficiente de Determinação (R^2)** e a **Raiz do Erro Quadrático Médio (RMSE)**.

Algoritmo	RMSE	R^2
Random Forest	30%	0,7922
Multilayer Perceptron	38%	0,6343
Random Tree	39%	0,6642

Tabela 2. Resultados obtidos para cada algoritmo, sendo apresentados a Raiz do Erro Quadrático Médio (RMSE) e o Coeficiente de Correlação (R^2)

O algoritmo *Random Forest* destacou-se como o mais eficiente entre os modelos testados, alcançando um **R^2 de 0,79** e um **RMSE de 30%**, indicando forte correlação entre os valores preditos e observados e menor margem de erro. A robustez do *Random Forest* decorre de sua capacidade de lidar com dados complexos e não lineares, característica essencial na análise de variáveis ambientais como o teor de carbono no solo. Estudos anteriores, como o de [10], também demonstraram a superioridade do *Random Forest* na estimativa de biomassa, reforçando a aplicabilidade desse algoritmo em contextos agrônômicos e ambientais.

Os algoritmos *Multilayer Perceptron* e *Random Tree* apresentaram desempenho inferior. O *Multilayer Perceptron* obteve um **R^2 de 0,63** e **RMSE de 38%**, revelando uma menor precisão na predição dos valores de carbono. O *Random Tree*, por sua vez, obteve **R^2 de 0,66** e **RMSE de 39%**, apresentando uma leve melhora em relação ao *Multilayer Perceptron*, mas ainda distante do desempenho alcançado pelo *Random Forest*.

Os resultados indicam que, embora todos os algoritmos tenham potencial para estimar o carbono no solo, o ***Random Forest*** é o mais recomendado, considerando sua precisão e

baixa margem de erro. Sua eficácia reforça o uso de aprendizado de máquina como uma abordagem promissora para a análise de variáveis agrônômicas e ambientais, facilitando tomadas de decisão mais eficientes e sustentáveis.

Para uma aplicação visual do modelo estabelecido, foi selecionada uma amostra representativa de cafeeiro para submetê-la ao processo de estimativa utilizando o algoritmo ***Random Forest***. O valor de carbono total determinado em laboratório para essa amostra foi de **59,4 (g/kg)**. Dessa forma, observou-se que a classe predominante na área amostral inclui o valor real, corroborando a eficácia do modelo em capturar as características espectrais relacionadas ao teor de carbono da amostra (Figura 4).

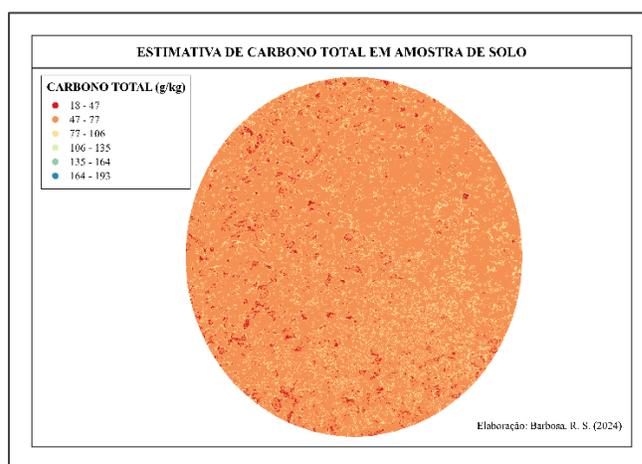


Figura 4. Estimativa de carbono total para amostra de solo de cafeeiro.

4. CONCLUSÕES

Este estudo demonstrou o potencial do aprendizado de máquina na estimativa do carbono total no solo a partir de dados radiométricos extraídos de imagens multiespectrais. Entre os algoritmos avaliados, o *Random Forest* apresentou o melhor desempenho, alcançando um R^2 de 0,79 e um RMSE de 30%, destacando-se como uma ferramenta eficiente para predição de carbono com alta precisão. A robustez deste algoritmo justifica sua aplicação na análise de variáveis agrônômicas, especialmente em situações que demandam a integração de dados ambientais complexos.

Os algoritmos *Multilayer Perceptron* e *Random Tree* apresentaram desempenhos inferiores, com maiores erros e menor correlação entre os dados preditos e observados, indicando que modelos mais simples ou com menor capacidade de generalização podem ter limitações em estudos desse tipo.

Com base nos resultados alcançados, algumas direções futuras podem ser exploradas para aprimorar este tipo de estudo. A expansão do conjunto de dados, com a coleta de amostras em diferentes regiões e condições climáticas, pode contribuir para a melhoria da generalização dos modelos, ampliando sua aplicabilidade.

A incorporação de dados hiperespectrais pode aumentar a precisão dos modelos, fornecendo maior riqueza de informações espectrais para distinguir variações sutis de carbono no solo. Além disso, o monitoramento temporal, utilizando séries temporais de imagens, permitiria o acompanhamento contínuo da dinâmica do carbono no solo, favorecendo análises mais robustas de longo prazo. A integração de outras variáveis agrônômicas, como umidade, textura do solo e biomassa, pode aumentar ainda mais a precisão e a utilidade dos modelos.

Outra perspectiva relevante é a automatização e o desenvolvimento de plataformas operacionais. Implementar esses algoritmos em plataformas automatizadas, como drones e satélites, pode facilitar a adoção dessa metodologia por agricultores e pesquisadores, ampliando sua aplicação em larga escala. Essas perspectivas reforçam a relevância do aprendizado de máquina e do sensoriamento remoto no manejo sustentável dos solos, auxiliando na tomada de decisões agrônômicas e ambientais. A continuidade dessas pesquisas pode fornecer uma base científica sólida para o desenvolvimento de práticas agrícolas mais eficientes e sustentáveis.

5. AGRADECIMENTOS

Este trabalho foi realizado por meio do apoio financeiro de bolsa de iniciação científica fornecida pela Fundação de Amparo à Pesquisa do Estado de Minas Gerais – FAPEMIG. Também contou com a colaboração financeira por meio de parceria Público Privada Com o Consórcio Cerrado das Águas - CCA e Grupo de Extensão e Pesquisa em Recursos Naturais e Agrícolas – RENAGRI (Diário oficial da União, Seção 3, nº 96, 22 de maio 2023).

6. REFERÊNCIAS

- [1] M. M. G. Neto *et al.* Predição do Teor de Carbono Total em Solos de Áreas Experimentais de Integração Lavoura-Pecuária por Meio da Espectroscopia NIR, *Embrapa*, v. 201, pp. 1-6, 2012. Disponível em: <https://ainfo.cnptia.embrapa.br/digital/bitstream/item/79195/1/ct-201.pdf>. Acesso em: 12 out. 2024.
- [2] W. Bettiol *et al.* Entendendo a matéria orgânica do solo em ambientes tropical e subtropical. Embrapa, Brasília, 2023. Disponível em: <https://www.alice.cnptia.embrapa.br/alice/bitstream/doc/1153147/1/Bettiol-Entendendo-materia-organica-2023.pdf>. Acesso em: 28 out. 2024.
- [3] P. C. Teixeira *et al.* Manual de Métodos de Análise de Solo, Embrapa, Brasília, pp. 359-367, 2017. Disponível em: <https://ainfo.cnptia.embrapa.br/digital/bitstream/item/194453/1/Pt-3-Cap-1-Carbono-organico.pdf>. Acesso em: 13 out. 2024.
- [4] A. C. de C Berardi, J. de M. Naime, A. V. de Resende, L. H. Bassoi, R. Y. Inamasu. Agricultura de precisão: resultados de um novo olhar, Embrapa, Brasília, p. 58-73, 2014. Disponível em: <https://www.alice.cnptia.embrapa.br/alice/bitstream/doc/1153147/1/Bettiol-Entendendo-materia-organica-2023.pdf>. Acesso em: 28 out. 2024.
- [5] J. W. Rouse, R. H. Haas, J. A. Schell, and D. W. Deering. Monitoring vegetation systems in the Great Plains with ERTS. *NASA Special Publication*, pp. 351-309, 1974.
- [6] P. Lynch, L. Blesius, E. Hines. Classification of Urban Area Using Multispectral Indices for Urban Planning, *Remote Sensing*, 2020; v. 12(15):2503. <https://doi.org/10.3390/rs12152503>.
- [7] R.P. Sripada, R.W. Heiniger, J.G. White, A. Meijer. Aerial Color Infrared Photography for Determining Early In-Season Nitrogen Requirements in Corn. *Agronomy Journal*, v. 98, pp. 968-977, 2006. <http://dx.doi.org/10.2134/agronj2005.0200>.
- [8] J. P. Z. Cunha. Um estudo comparativo das técnicas de validação cruzada aplicadas a modelos mistos. Dissertação, São Paulo, 2019. Disponível em: https://www.teses.usp.br/teses/disponiveis/45/45133/tde-26082019-220647/publico/Dissertacao_JoaoPauloZanola.pdf. Acesso em: 28 out. 2024.
- [9] E. Frank, M. A. Hall, and I. H. Witten. The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", *Morgan Kaufmann*, Fourth Edition, 2016. Disponível em: https://ml.cms.waikato.ac.nz/weka/Witten_et_al_2016_appendix.pdf. Acesso em: 13 out. 2024.
- [10] V. H. Nisieimon. Estimativa de biomassa de pastagem utilizando técnicas de seleção de atributos e aprendizagem de máquina a partir de imagens do Sentinel-2. Trabalho de Conclusão de Curso, Campinas, 2020. Disponível em: <https://repositorio.unicamp.br/Busca/Download?codigoArquivo=553484&tipoMidia=0>. Acesso em: 14 out. 2024.