

# APLICAÇÃO DE APRENDIZADO POR TRANSFERÊNCIA PARA O MAPEAMENTO DE SISTEMAS INTENSIVOS DE PRODUÇÃO AGROPECUÁRIA POR SENSORIAMENTO REMOTO EM LARGA ESCALA

Patrick Calvano Kuchler<sup>1</sup>, Rodrigo Ferraz<sup>2</sup>, Luan Porto Vieira<sup>3</sup>, Sofia Martins Lage<sup>2</sup>, Mateus Benchimol<sup>4</sup>, Margareth Simões<sup>2,4</sup>

<sup>1</sup>Professor Adjunto do Departamento de Geografia Física (DGF) e do Programa de Pós-graduação em Geografia (PPGEO) da Universidade do Estado do Rio de Janeiro (UERJ). Rua São Francisco Xavier, 524, Pavilhão João Lyra Filho, 4 Andar, Bairro Maracanã, RJ, CEP: 20550-013, e-mail: [geocalvano@gmail.com](mailto:geocalvano@gmail.com); <sup>2</sup>Embrapa Solos, Rua Jardim Botânico, 1024, Jardim Botânico, RJ, e-mail: {rodrigo.demonte, margareth.simoes}@embrapa.br e [sofmlage@gmail.com](mailto:sofmlage@gmail.com); <sup>3</sup>Departamento de Engenharia de Sistemas e Computação, UERJ/FEN, Faculdade de Engenharia, Universidade do Estado do Rio de Janeiro (UERJ), Bairro Maracanã, RJ, CEP: 20550-013, e-mail: [luanportovieira@gmail.com](mailto:luanportovieira@gmail.com); <sup>4</sup>Programa de Pós-graduação em Meio Ambiente (PPGMA), Doutorado Interdisciplinar, Universidade do Estado do Rio de Janeiro (UERJ), Rua São Francisco Xavier, 524, Pavilhão João Lyra Filho, 12 Andar, Bloco F, Sala 12005, Bairro Maracanã, RJ, CEP: 20550-013, e-mail: [mateusbenchimol@hotmail.com](mailto:mateusbenchimol@hotmail.com)

## RESUMO

O aprendizado por transferência (*transfer learning*) é uma técnica de *machine learning* que permite reaproveitar conhecimento adquirido em um domínio para tarefas em um novo, sendo útil para classes complexas com difícil aquisição de dados, como os sistemas integrados Lavoura-Pecuária (ILP). Este estudo investiga a aplicabilidade da técnica de Adaptação de domínio, baseada em TL, para otimizar a classificação de ILP em grandes áreas, aplicando modelos treinados em uma região ou ano específico para outras áreas e períodos distintos com características particulares. Observou-se que tanto o TL espacial quanto o temporal melhoraram significativamente a acurácia, mesmo em diferentes regimes climáticos e anos com características variadas. Em regiões com poucas amostras, os índices de acurácia mostraram grande melhoria, enquanto áreas com mais amostras apresentaram ganhos menores. Os resultados indicam o potencial de generalização dos modelos para o monitoramento de ILP em amplas áreas agrícolas como subsídio às políticas públicas setoriais.

**Palavras-chave** — transferência de aprendizagem, aprendizagem de máquina, *random forest*, adaptação de domínio, e sistemas integrados.

## ABSTRACT

*Transfer learning is a machine learning technique that enables reusing knowledge acquired in one domain for tasks in a new domain, making it valuable for complex classes with difficult data acquisition, such as Integrated Crop-Livestock (ICL) systems. This study investigates the applicability of TL to optimize ICL classification over large areas by applying models trained in a specific region or year to different areas and periods with distinct characteristics. Both spatial and temporal TL significantly*

*improved accuracy, even across varying climate regimes and years with unique traits. In regions with few samples, accuracy indices showed considerable improvement, while areas with more samples displayed smaller gains. The results highlight the potential of model generalization for ICL monitoring over extensive agricultural areas, supporting sectoral public policies.*

**Key words** — *transfer learning, machine learning, random forest, domain adaptation, integrated systems.*

## 1. INTRODUÇÃO

Os sistemas de Integração Lavoura-Pecuária (ILP) combinam cultivos agrícolas e pecuária em uma mesma área, promovendo o consórcio e/ou a sucessão de diversas culturas ao longo do ano, incluindo a pastagem como componente essencial. Esses sistemas são fundamentais para a agricultura sustentável.

A expansão das plataformas de sensoriamento remoto aumentou o volume de dados disponíveis, possibilitando séries temporais densas e o processamento paralelo de grandes volumes, caracterizados como Big Earth Observation Data [1-4]. Isso tem avançado o mapeamento de grandes áreas, embora o desafio de obter dados rotulados de campo persista.

Ao contrário de classes estáveis, como pastagens e culturas perenes, sistemas ILP e outros com cultivos sequenciais apresentam alta variabilidade anual, dificultando a coleta direta ou indireta de dados em campo devido à necessidade de ampla amostragem em diferentes períodos. Nessas situações, entrevistas com produtores e cooperativas têm sido estratégias eficazes [5-9], embora limitadas pela dependência de uma rede ampla de contatos [10,11]. Nesse contexto, o aprendizado por transferência (TL) surge como uma solução para superar esses desafios, possibilitando a adaptação de modelos a áreas com dados limitados através do aproveitamento de informações de domínios com maior

disponibilidade de dados [12,13,14]. A Adaptação de Domínio (DA) dentro do conceito de TL pode ser considerado como uma técnica de *machine learning* que visa transferir o conhecimento adquirido em um domínio de origem para realizar tarefas relacionadas em um domínio-alvo [15]. Ao aproveitar o conhecimento aprendido no domínio de origem, o DA/TL pode ajudar a superar os desafios associados à limitação de dados rotulados no domínio-alvo [16, 17].

eficácia dos princípios de AD/TL no mapeamento de ILP, aplicando adaptações de domínios espaciais e temporais para maximizar a acurácia e robustez dos modelos em regiões e períodos com poucas amostras rotuladas. As hipóteses são: (1) informações de domínios espaciais permitem transferir conhecimento entre regiões climáticas distintas, reduzindo a necessidade de dados locais, e (2) dados de domínios temporais distintos podem mitigar variações anuais, permitindo aplicar modelos treinados a períodos subsequentes sem perda significativa de acurácia.

## 2. MATERIAL E MÉTODOS

A implementação do TL ocorreu em Mato Grosso, importante produtor de gado e soja no Brasil. Com mais de 900.000 km<sup>2</sup>, o estado abrange três biomas — Amazônia, Cerrado e Pantanal —, oferecendo um cenário diversificado e desafiador para o mapeamento de classes complexas.

A base de teste incluiu dados de levantamentos de campo, entrevistas com produtores, parcerias com consultores, visitas à Embrapa e amostras do estudo de Câmara et al. [1]. Esse conjunto gerou 891 polígonos com área média de 96 hectares, distribuídos em 46 municípios e abrangendo oito safras (2011/2012 a 2018/2019), com 90% das amostras concentradas entre 2014/2015 e 2018/2019.

Para avaliar o impacto da distribuição geográfica nos resultados estaduais, consideraram-se as seis regiões climáticas do zoneamento ecológico-econômico do estado [18], baseadas em 20 anos de dados e na duração da precipitação no trimestre mais chuvoso. As abordagens de aprendizado por transferência incluíram dimensões espaciais e temporais, implementadas em duas etapas principais.

### 2.1. Transferência Temporal

Avaliamos a acurácia da classificação de ILP em diferentes anos usando duas estratégias. A primeira, tradicional, utiliza 2/3 das amostras de um único ano para treinamento e o restante para validação. Na segunda abordagem, combinamos amostras de múltiplos anos para treinar o modelo e validamos em anos específicos não incluídos no treinamento. Essa abordagem permite testar o quanto o modelo aprende com padrões temporais comuns dos ILP, aprimorando a adaptabilidade para mudanças fenológicas e de precipitação [16,17].

### 2.2. Transferência Espacial

Utilizando um zoneamento ecológico do estado de estudo, as amostras foram agrupadas em regiões climáticas. Para avaliar o TL entre essas zonas, com diferentes condições clima, a abordagem tradicional dividiu amostras de cada área em 2/3 para treinamento e 1/3 para validação local. Já na segunda estratégia, as amostras de todas as regiões foram combinadas para criar uma base de dados abrangente, utilizada para classificar ILP em cada área específica.

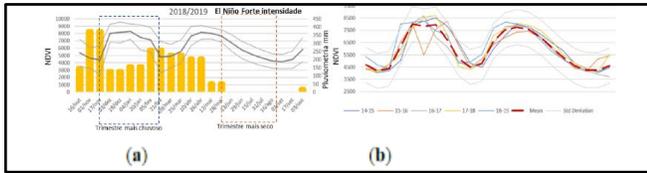
### 2.3. Classificação

Para a classificação, utilizamos o algoritmo *Random Forest* (RF), aplicando 100 árvores. Foram utilizados dados do MOD13Q1, produto NDVI com intervalo de 16 dias em uma série temporal entre outubro do ano base a outubro do ano base +1. Os dados coletados em campo e a legenda adotada, contaram com 3 classes - ILP Soja-milho-pastagem e Soja\_Pastagem), Soja-Cereais (milho, milheto, girassol, sorgo) e Soja-Algodão. Para as análises dos resultados de TL nos diferentes cenários, utilizamos os índices de Acurácia Global (AG) e o valor de *fscore* para a classe ILP.

## 3. RESULTADOS

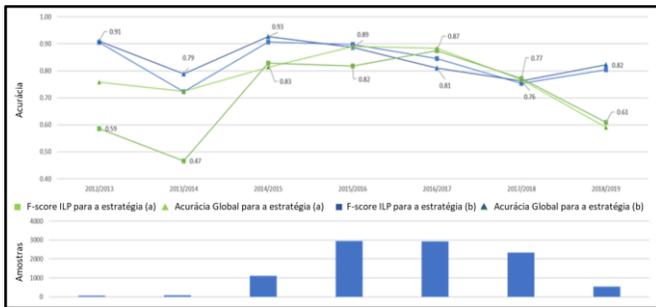
Para a análise do aprendizado por transferência temporal, foi inicialmente estudada a variabilidade intra-anual das séries temporais, utilizando a média dos valores de NDVI do sistema ILP e os registros mensais de precipitação coletados nas estações meteorológicas do Instituto Nacional de Meteorologia [20] entre 2014 e 2019.

A análise das séries temporais de NDVI revelou uma variabilidade intra-anual significativa no perfil da classe ILP, especialmente durante a estação chuvosa, de outubro a março, que corresponde ao período de cultivo da soja. Essa variabilidade é acentuada em anos com alta intensidade dos fenômenos El Niño e La Niña, apresentando dois picos de crescimento no NDVI em 2015/2016, possivelmente associados à heterogeneidade da precipitação e à interferência atmosférica, além de um platô prolongado em 2018/2019. Em contraste, os perfis de NDVI mostraram-se mais homogêneos durante o trimestre seco, de junho a agosto, mesmo em anos influenciados por *El Niño* e *La Niña*. Como exemplo, a figura 1 apresenta a série temporal do ano-safra 2018-2019 sobreposta com o regime de precipitação, além do quadro com a média de todos os anos-safra.



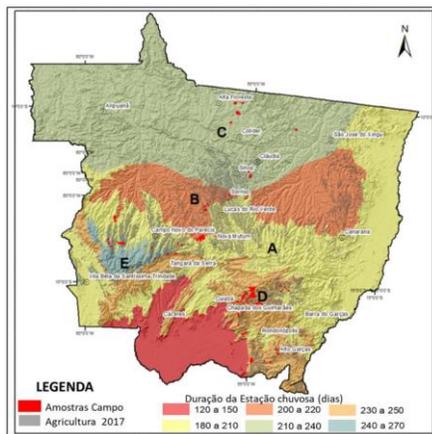
**Figura 1. Séries temporais NDVI média para todo o MT: (a) STIS do ano-safra 2018/2019; (b) STIS de todos os anos**

Os resultados da classificação no TL Temporal, mostram que a acurácia global e o *f*score são maiores para a estratégia (b), ou seja, quando são utilizados os dados dos cinco anos principais. Os resultados apresentados na Figura 2 sugerem que a precisão está mais relacionada à quantidade de amostras do que à variabilidade da precipitação anual.



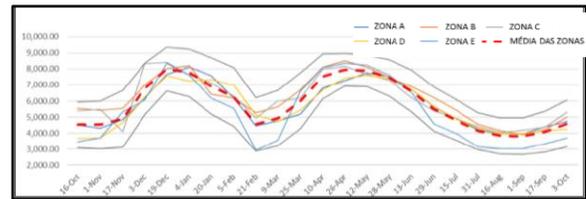
**Figura 2. Acurácia global e *f*score para cada ano e para as duas estratégias de amostragem temporal (o número de amostras disponíveis por ano é representado por histogramas na figura inferior)**

Em anos com poucas amostras (2012/2013, 2013/2014 e 2018/2019), dados combinados de cinco anos aumentaram significativamente a acurácia global e o *f*-score. Nos anos com mais amostras (2015/2016, 2016/2017 e 2017/2018), esses índices pouco variaram, independentemente do número de anos no treinamento. O Transfer Learning Espacial começou com uma análise visual de curvas de NDVI.



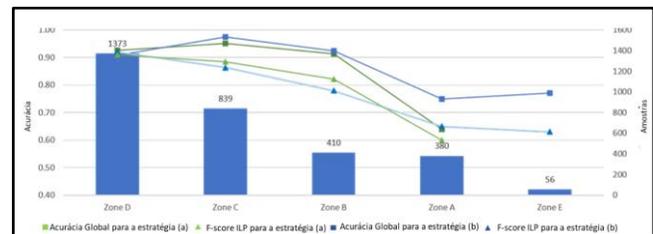
**Figura 3. Regiões Climáticas**

A Zona E apresenta a curva de NDVI mais 'ruidosa', com maior amplitude de variação, mas é também a área com o menor número de amostras (56). Durante o segundo ciclo, correspondente à estação seca, as Zonas C e D exibem comportamentos semelhantes, embora possuam uma quantidade maior de amostras. Diferentemente dos padrões observados na análise da variabilidade temporal, a dispersão e simetria dos valores de NDVI por região não seguem uma tendência clara. Na Zona A, a amplitude dos valores é maior na estação chuvosa, enquanto nas Zonas B, C e E observa-se uma maior diversidade de valores durante a estação seca. Na Zona D, a amplitude se mantém relativamente constante ao longo do ano (Figura 4)



**Figura 4. Variabilidade espacial das amostras de iLP e regiões climáticas do estado de Mato Grosso.**

A análise mostra que maior quantidade de amostras em uma zona climática resulta em acurácia global e *f*score mais elevados, tanto nos dados (a) quanto em (b). As áreas B, C e D apresentaram valores semelhantes de acurácia e *f*score em ambas as estratégias, como ilustrado na figura 5.



**Figura 5. Acurácia global e *f*score da classe iLP, obtidos para as 5 regiões climáticas**

Fica evidente ainda que na zona A, quando todos os anos são usados para treinamento (estratégia b), há uma melhoria significativa na acurácia global

#### 4. DISCUSSÃO

Os resultados indicam que o Transfer Learning (TL) temporal aumentou significativamente a acurácia da classificação em anos sem dados rotulados, especialmente ao usar dados de múltiplos anos no treinamento. Essa abordagem foi eficaz em anos com alta variabilidade de precipitação, permitindo ao modelo adaptar-se a diferentes ciclos de crescimento e generalizar para períodos não amostrados, o que é crucial em regiões agrícolas com clima irregular. O TL espacial também foi vantajoso ao usar dados de regiões com diferentes condições climáticas e ambientais, melhorando a classificação de sistemas ILP em áreas com

pouca amostragem local. Essa implementação resultou em maior acurácia em comparação à abordagem tradicional, evidenciando o potencial do TL espacial para lidar com variabilidades regionais e adaptar modelos a novas áreas, reduzindo a necessidade de coleta intensiva de dados de campo.

## 5. CONCLUSÕES

Este estudo confirma que a adaptação de domínio, tanto temporal, quanto espacial, técnicas de aprendizado por transferência, é uma estratégia viável e eficaz para o mapeamento de sistemas de Integração Lavoura-Pecuária em contextos onde há escassez de dados rotulados. Os resultados indicam que o TL temporal e espacial contribuíram para uma acurácia aprimorada e para a generalização dos modelos, o que reduziu o impacto das variações climáticas e fenológicas sobre a classificação de uso da terra. Em anos com elevada variabilidade de precipitação e em regiões com condições climáticas diversas, o TL permitiu uma maior adaptação dos modelos, mostrando-se uma ferramenta robusta para a classificação precisa de sistemas ILP.

## 6. REFERÊNCIAS

- [1] CÂMARA, Gilberto et al. Land cover change maps for Mato Grosso State in Brazil: 2001-2017 (Version 2), 10.1594/PANGAEA.895495. Disponível em: <https://doi.pangaea.de/10.1594/PANGAEA.895495>.
- [2] TAMIMINIA, H., et al. A comprehensive review of random forest applications in remote sensing. *IEEE Geoscience and Remote Sensing Magazine*, 8(4):123-145, 2020.
- [3] LI, Songnian et al. Geospatial big data handling theory and methods: A review and research challenges. *ISPRS Journal of Photogrammetry and Remote Sensing*, 115:119–133, 2016. DOI: 10.1016/j.isprsjprs.2015.10.012.
- [4] FERREIRA, Karine R. et al. Earth Observation Data Cubes for Brazil: Requirements, Methodology and Products. *Remote Sensing*, 12(24):4033, 2020. DOI: 10.3390/rs12244033.
- [5] ARVOR, Damien; JONATHAN, Milton; MEIRELLES, Margareth Simões Penello; DUBREUIL, Vincent; DURIEUX, Laurent. Classification of MODIS EVI time series for crop mapping in the state of Mato Grosso, Brazil. *International Journal of Remote Sensing*, 32(22):7847–7871, 2011. DOI: 10.1080/01431161.2010.531783.
- [6] CHEN, Yaoliang et al. Mapping croplands, cropping patterns, and crop types using MODIS time-series data. [s.l.: s.n.]. 69. DOI: 10.1016/j.jag.2018.03.005.
- [7] KASTENS, Jude H.; BROWN, J. Christopher; COUTINHO, Alexandre Camargo; BISHOP, Christopher R.; ESQUERDO, Júlio César D. M. Soy moratorium impacts on soybean and deforestation dynamics in Mato Grosso, Brazil. *PLOS ONE*, 12(4), 2017. DOI: 10.1371/journal.pone.0176168.
- [8] KUCHLER, Patrick Calvano; BÉGUÉ, Agnès; SIMÕES, Margareth; GAETANO, Raffaele; ARVOR, Damien; FERRAZ, Rodrigo P. D. Assessing the optimal preprocessing steps of MODIS time series to map cropping systems in Mato Grosso, Brazil. *International Journal of Applied Earth Observation and Geoinformation*, 92:102150, 2020. DOI: 10.1016/j.jag.2020.102150.
- [9] KUCHLER, Patrick C. Monitoring Complex Integrated Crop–Livestock Systems at Regional Scale in Brazil: A Big Earth Observation Data Approach. *Remote Sensing*, 14(8):1648, 2022.
- [10] BÉGUÉ, Agnès et al. Remote Sensing and Cropping Practices: A Review. *Remote Sensing*, 10(1):99, 2018. DOI: 10.3390/rs10010099.
- [11] BELLÓN, Beatriz; BÉGUÉ, Agnès; SEEN, Danny Lo; ALMEIDA, Claudio Aparecido De; SIMÕES, Margareth. A Remote Sensing Approach for Regional-Scale Mapping of Agricultural Land-Use Systems Based on NDVI Time Series. *Remote Sensing*, 9:600, 2017. DOI: 10.3390/rs9060600.
- [12] MA, Y., et al. Transfer learning in environmental remote sensing. *Remote Sensing of Environment*, 301:113924, 2024.
- [13] TUIA, D., PERSELLO, C., BRUZZONE, L. Domain adaptation for the classification of remote sensing data: an overview of recent advances. *IEEE Geosci. Remote Sens. Mag.*, 4:41–57, 2016. DOI: 10.1109/MGRS.2016.2548504.
- [14] ZHU, Q., Shen, F., Shang, P., Pan, Y., Li, M., 2020. Hyperspectral remote sensing of phytoplankton species composition based on transfer learning. *Remote Sens.* 12 <https://doi.org/10.3390/rs12030364>
- [15] PAN, S.J., YANG, Q. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.*, 22:1345–1359, 2010. DOI: 10.1109/TKDE.2009.191.
- [16] GADIRAJU, K.K., VATSA vai, R.R. Comparative analysis of deep transfer learning performance on crop classification. In: *Proceedings of the 9th ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data, BIGSPATIAL 2020*. Association for Computing Machinery, Inc. DOI: 10.1145/3423336.3431369.
- [17] LU, G. et al. *AGI for Agriculture*, 2023.
- [18] SANTOS, R. D., et al. Zoneamento ecológico-econômico do estado de São Paulo. *Embrapa Solos*, 2000.
- [20] INMET. Dados das estações meteorológicas, 2019. Disponível em: <https://portal.inmet.gov.br/>. Acesso em: 25 jul. 2020..