

Classificação de Monoculturas de Café via Redes Neurais com Redução de Dados utilizando Fusão Espectral de Imagens Sentinel

Fernanda Esteves Coelho Chaves, Renato da Rocha Lopes e João Marcos Travassos Romano

Resumo— Este trabalho propõe uma abordagem de fusão de imagens multiespectrais do Sentinel-2 e de radar de abertura sintética (synthetic aperture radar, SAR) do Sentinel-1 para classificar áreas cafeeiras em Machado-MG, utilizando redes neurais dos tipos U-Net e long short-term memory (LSTM). As bandas foram combinadas por técnicas como média, principal component analysis (PCA), discrete wavelet transform (DWT) e pirâmide espectral, visando à redução de dimensionalidade com preservação de informações relevantes. A U-Net obteve os melhores resultados, com acurácia de 92,3% e índice Kappa de 0,78 usando PCA, destacando a eficiência da fusão espectral no mapeamento agrícola.

Palavras-Chave— Fusão de imagens, Imagens multiespectrais, imagens Sentinel, Classificação de Áreas Cafeeiras

Abstract— This work proposes an approach for fusing multispectral Sentinel-2 images and radar data (synthetic aperture radar) from Sentinel-1 to classify coffee-growing areas in Machado, Brazil, using universal network (U-Net) and long short-term memory (LSTM) neural networks. The image bands were combined using techniques such as averaging, principal component analysis (PCA), discrete wavelet transform (DWT), and spectral pyramid, aiming to reduce dimensionality while preserving relevant information. U-Net achieved the best results, with 92.3% accuracy and a Kappa index of 0.78 using PCA, highlighting the effectiveness of spectral fusion in agricultural mapping.

Keywords— Image fusion, Multispectral images, Sentinel images, Coffee Areas Classification

I. INTRODUÇÃO

Com os avanços tecnológicos no campo dos sensores ao longo dos anos, observou-se um aumento significativo na quantidade de dados gerados por esses dispositivos. Para lidar com esse grande volume de informações provenientes de diferentes fontes, a fusão de sensores surgiu como uma estratégia eficaz para condensá-las, facilitando sua análise e interpretação [1]. A partir da fusão dos dados obtidos de fontes diversas, torna-se viável gerar um novo conjunto de dados, resultante da combinação de informações obtidas por sensores de naturezas distintas.

Dada a ampla variedade de aplicações que envolvem a fusão de sensores com imagens adquiridas em diferentes comprimentos de onda, consolidou-se uma subárea específica denominada fusão de imagens. Essa abordagem tem como

principal objetivo combinar imagens previamente adquiridas a fim de gerar uma nova imagem com maior resolução espacial, espectral ou temporal, ou ainda viabilizar sua reconstrução e/ou restauração [2]. Adicionalmente, a fusão de imagens pode ser empregada para realçar, revelar ou identificar objetos na cena que, de forma isolada, seriam invisíveis ou de difícil detecção. Para isso, diversos algoritmos têm sido propostos, os quais podem ser organizados, de maneira geral, em três grandes categorias: fusão direta de imagens, fusão baseada em decomposição piramidal e fusão baseada na transformada discreta de wavelet (do inglês, DWT, *Discrete Wavelet Transform*) [3], [4].

Ainda na década de 1980, começaram a surgir abordagens voltadas à realização da fusão de imagens no domínio da frequência, destacando-se, entre elas, o uso da transformada piramidal. Essa técnica mostrou-se promissora por sua capacidade de representar informações em diferentes escalas, possibilitando a identificação de variações abruptas de contraste, bem como a localização precisa de detalhes relevantes tanto no domínio espacial quanto no domínio espectral [3]. A decomposição piramidal permite isolar componentes de alta e baixa frequência, o que facilita a combinação seletiva de informações provenientes de diferentes imagens, contribuindo significativamente para a geração de imagens fusionadas com maior riqueza de detalhes e melhor qualidade perceptual. Essa abordagem tem sido amplamente aplicada em áreas nas quais a fidelidade das informações estruturais e texturais é essencial [5].

A DWT pode ser interpretada como um caso particular de decomposição piramidal. Além de analisar o conteúdo em frequência das imagens de entrada em múltiplas escalas, também fornecem informações sobre a localização temporal dessas frequências, ou seja, permitem a identificação dos momentos em que determinados componentes espectrais ocorrem [6]. Essa capacidade de análise espaço-frequência torna as DWTs especialmente eficazes na preservação de bordas e detalhes sutis das imagens, aspectos essenciais em tarefas de fusão onde a integridade estrutural e espectral das imagens de origem deve ser mantida. Ademais, sua natureza de multirresolução permite adaptar a análise a diferentes níveis de detalhe, tornando-a adequada para uma ampla gama de aplicações, como detecção de mudanças, realce de imagens e segmentação em sensoriamento remoto [7].

No setor agrícola, a tecnologia de fusão de imagens surge como uma ferramenta promissora, ao fornecer informações integradas que possibilitam uma compreensão mais abrangente

Fernanda Esteves Coelho Chaves, DECOM, Universidade Estadual de Campinas, Campinas-SP, e-mail: f215835@unicamp.br. O trabalho foi realizado no âmbito do Centro de Pesquisa Aplicada em Inteligência Artificial BIOS – Brazilian Institute of Data Science, apoiado pelo processo nº 2020/09838-0, Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).

das condições agrícolas. Uma aplicação específica nessa área foi desenvolvida por Barrero e Perdomo [8], que utilizaram a DWT de Haar [9] com o objetivo de detectar plantas daninhas em lavouras de arroz com 50 dias de crescimento. O método proposto realiza a fusão de imagens RGB (do inglês *red*, *green* e *blue*) de alta resolução com imagens multiespectrais (MS), gerando um produto híbrido (RGB-MS) com características aprimoradas de discriminação entre culturas e ervas daninhas. Essas imagens fusionadas foram então submetidas a uma rede neural para classificação, sendo comparadas também às imagens RGB e multiespectrais originais. Os melhores resultados de detecção obtidos por Barreto e Perdomo foram a partir das imagens fusionadas, evidenciando o potencial da técnica no contexto agrícola.

Já o trabalho apresentado em [10] propõe métodos eficientes para registro, fusão e redução de dimensionalidade de imagens multissensor, com destaque para técnicas baseadas em wavelets. Os resultados demonstram melhor desempenho e menor custo computacional em relação a abordagens tradicionais, como análise de componentes principais (do inglês, PCA, *principal component analysis*), otimizando computacionalmente a análise de dados remotos com diferentes resoluções.

O avanço do *big data*, o acesso crescente a imagens de alta resolução e o uso de plataformas em nuvem têm impulsionado a agricultura orientada por dados [11]. Nesse contexto, técnicas de inteligência artificial vêm sendo aplicadas ao mapeamento de culturas agrícolas via sensoriamento remoto. Adrian et al. [12], por exemplo, empregaram a fusão de dados multitemporais do Sentinel-1 e Sentinel-2 para melhorar a classificação de culturas com algoritmos de aprendizado de máquina. Contudo, o alto custo computacional desses modelos tem motivado a busca por soluções mais eficientes.

Considerando que o Brasil é atualmente o maior produtor e exportador de café do mundo [13], uma aplicação relevante e estratégica no contexto nacional refere-se à identificação e classificação de áreas cultivadas com café. Essa aplicação tem o potencial de contribuir para o planejamento e a gestão da produção cafeeira. Estudos voltados para a classificação de cafezais no Brasil foram conduzidos por Coltri et al. [14] e Cordeiro et al. [15], ambos utilizando imagens de satélite, embora sem a incorporação de técnicas de fusão de imagens em suas abordagens.

Apesar desses avanços, há lacunas em aplicações que integrem dados de diferentes faixas espectrais na agricultura brasileira. A fusão de imagens, combinando dados ópticos e de radar, aprimora a precisão dos modelos de IA, mesmo em condições adversas, além de reduzir a dimensionalidade e o volume de dados processados, otimizando a eficiência computacional. Assim, propomos aplicar a fusão de imagens e a classificação por meio de IA em regiões cafeeiras, visando uma gestão agrícola mais eficiente e sustentável.

O presente trabalho está estruturado da seguinte forma: na Seção II descrevemos em detalhes a metodologia empregada, incluindo os procedimentos de aquisição e processamento dos dados, os métodos de fusão de imagens utilizados e as configurações das redes neurais e seus treinamentos. Na Seção III apresentamos e analisamos os resultados obtidos. Por fim, a Seção IV contém as considerações finais e conclusões

decorrentes do estudo.

II. METODOLOGIA

A. Aquisição e Pré-processamento dos Dados

As imagens utilizadas neste trabalho foram obtidas por meio da plataforma Sentinel Hub [16], mais especificamente por meio da sua interface gráfica, o EO Browser [17]. Esta ferramenta permite a visualização, manipulação e obtenção de imagens de satélite de forma prática e eficiente. A plataforma oferece acesso direto a dados das missões Sentinel, do programa Copernicus, com filtros por cobertura de nuvens, período e composição espectral.

Já a área de estudo corresponde ao município de Machado, localizado no sul de Minas Gerais. O município, por sua vez, apresenta uma alta produção cafeeira, sendo um importante polo da agricultura regional. Além disso, o relevo do município é predominantemente montanhoso e ondulado, o que influencia diretamente na distribuição espacial das culturas, bem como na qualidade e na forma das imagens ópticas, já que podem apresentar sombreamento devido ao relevo.

Adicionalmente, foram selecionadas cenas com baixa cobertura de nuvens e boa qualidade radiométrica. As bandas espectrais do Sentinel-2 foram extraídas nas resoluções de 10 m e 20 m, sendo estas últimas reamostradas para 10 m, a fim de garantir a consistência espacial entre as bandas. Os dados foram exportados no formato GeoTIFF, organizados e empilhados para compor os tensores multibanda utilizados como entrada nos modelos de classificação ou nos métodos de fusão. Dessa forma, foram obtidas 12 bandas do Sentinel-2 [B1 (*coastal aerosol*), B2 (*blue*), B3 (*green*), B4 (*red*), B5 (*red edge 1*), B6 (*red edge 2*), B7 (*red edge 3*), B8 (*NIR*), B8A (*narrow NIR*), B9 (*water vapor*), B11 (*SWIR 1*), B12 (*SWIR 2*)] e 2 bandas do Sentinel-1 [VV (*Vertical-Vertical*) e VH (*Vertical-Horizontal*)], resultando em 14 bandas de 1831 x 2500 pixels cada.

Vale ressaltar que os valores ausentes foram preenchidos com a média de cada banda, e os dados foram padronizados para média zero e desvio padrão unitário. Por fim, as máscaras de referência utilizadas para a segmentação foram obtidas a partir dos dados disponibilizados pela Companhia Nacional de Abastecimento (CONAB), referentes ao mapeamento agrícola do ano de 2017. Esses dados foram processados para gerar imagens binárias, nas quais os pixels são rotulados como 1 (presença da classe de interesse) ou 0 (ausência), de acordo com as áreas identificadas nos mapeamentos oficiais [18]. Tais características evidenciam a abordagem de treinamento e classificação pixel-a-pixel utilizada nesse trabalho.

B. Métodos de Fusão de Dados

A fim de obter novos dados relevantes provenientes do Sentinel-1 e do Sentinel-2 para a tarefa de classificação, foram empregados diferentes métodos para a fusão espectral.

Na fusão por média, a dimensionalidade do Sentinel-2 foi reduzida agrupando bandas espectralmente semelhantes [bandas do espectro visível (B2, B3 e B4), *red edge* (B5, B6 e B7), infravermelho médio (B11 e B12) e infravermelho próximo (B8 e B8A)]. As bandas B1 e B9 foram mantidas

separadas por sua especificidade, enquanto no Sentinel-1, as polarizações VV e VH foram fundidas por média simples. O resultado foi uma imagem final com 7 bandas.

A abordagem baseada em PCA utilizou todas as 14 bandas combinadas de Sentinel-1 e Sentinel-2. Após o agrupamento das bandas em uma matriz 3D, a imagem tridimensional foi transformada em uma matriz bidimensional para a aplicação da decomposição PCA, gerando componentes ortogonais ordenados por variância explicada. As 8 primeiras componentes principais foram selecionadas para compor a imagem final reduzida, preservando aproximadamente 99,83% da variância total dos dados. Esse número foi escolhido como um compromisso entre a redução de dimensionalidade e a preservação da informação original.

Para a DWT com base Haar, as 14 bandas empilhadas foram tratadas como um cubo hiperespectral (matriz 3D). A DWT unidimensional foi aplicada pixel a pixel ao longo do eixo espectral, com 2 níveis de decomposição. Essa transformação gerou coeficientes em diferentes escalas espectrais, dos quais os 8 com maior variância ao longo dos pixels da imagem foram selecionados, sendo reorganizados posteriormente no formato original. Uma versão complementar da DWT foi realizada utilizando a wavelet Daubechies (db6), com a mesma estratégia: decomposição espectral ao longo dos pixels, cálculo da variância dos coeficientes e seleção dos 8 mais informativos para compor a imagem final.

A Pirâmide Laplaciana Espectral foi aplicada à imagem resultante da fusão dos sensores Sentinel-2 e Sentinel-1. Primeiramente, as bandas espectrais foram suavizadas com filtro gaussiano 1D e submetidas a um processo iterativo de redução e reconstrução espectral, semelhante à construção de uma pirâmide de escalas. A cada nível, as bandas eram subamostradas espectralmente e interpoladas de volta ao tamanho original, permitindo o cálculo de resíduos espectrais, que realçam os detalhes perdidos. Os resíduos formaram a pirâmide laplaciana, com o último nível representando uma versão suavizada do espectro. Para a imagem final, foram selecionados resíduos de níveis mais profundos e o último *lowpass*, gerando uma representação compacta e multiescala da informação espectral.

C. Arquiteturas de Redes Utilizadas

Foram utilizadas duas arquiteturas principais de redes neurais: a LSTM (do inglês, *Long Short-Term Memory*) [19] e a U-Net (do inglês, *Universal Network*) [20], [21].

Para a abordagem baseada em redes recorrentes, foi implementada uma arquitetura Bidirectional LSTM com o objetivo de capturar possíveis padrões espectrais e relacionamentos interbandas, mesmo que não necessariamente temporais. Inicialmente, os dados foram organizados em um vetor com n atributos por amostra, correspondentes às bandas espectrais dos sensores Sentinel-2 e Sentinel-1, fundidas previamente por diferentes técnicas. Como as redes LSTM exigem entradas tridimensionais, foi adicionada uma dimensão extra para simular uma sequência temporal. A rede foi composta por 2 blocos LSTM bidirecionais com normalização em lote e *dropout*, seguidos de 1 camada totalmente conectada (*dense*) com

ativação ReLU e outra de saída com ativação sigmoide para a classificação binária. O treinamento foi realizado utilizando o otimizador Adam e a função de perda *binary crossentropy*, com validação em conjunto separado e estratégia de *early stopping* para evitar sobreajuste.

A arquitetura U-Net clássica foi utilizada para segmentação binária das imagens multiespectrais do Sentinel-1 e Sentinel-2, que possuem 14 bandas espectrais. As imagens foram normalizadas com o *StandardScaler* e divididas em patches de 256 x 256 pixels, com um stride de 128, para alimentar o modelo. A U-Net consiste em 2 caminhos principais: o de contração, que extrai características de alto nível, e o de expansão, que reconstrói a segmentação com alta resolução. Ao final, a U-Net produziu uma saída binária que segmenta áreas de interesse nas imagens.

Além disso, para lidar com o desbalanceamento de classes (já que há muito mais pixels que não são regiões de plantio de café), os modelos foram treinados com a função de perda ponderada (do inglês, *weighted binary crossentropy*) [22].

D. Aplicação dos Métodos e Treinamento dos Modelos

Para cada método de fusão avaliado, os dados correspondentes foram individualmente divididos em conjuntos de treinamento, validação e teste, seguindo a proporção de 60%, 20% e 20%, respectivamente. A U-Net foi treinada separadamente com as imagens resultantes de cada método de fusão, compostas pelas respectivas bandas empilhadas, utilizando função de perda binária e otimizador Adam. A LSTM, por sua vez, foi treinada com sequências de valores espectrais ao longo do tempo, também de forma independente para cada conjunto fundido, mantendo o mesmo conjunto de classes. Como referência comparativa, também foram realizados treinamentos com os dados originais, compostos pelas 14 bandas dos sensores Sentinel-1 e Sentinel-2 empilhadas, sem aplicação de qualquer técnica de fusão. Esses dados de entrada foram utilizados tanto na U-Net quanto na LSTM, permitindo estabelecer uma base de comparação direta com os resultados obtidos a partir dos diferentes métodos de fusão espectral.

A principal métrica de avaliação foi a acurácia geral, complementada pelo índice Kappa de Cohen [23] e por análise visual das previsões. A acurácia indica a proporção de acertos do modelo em relação ao total de amostras, sendo útil para medir o desempenho global. No entanto, em contextos com classes desbalanceadas, essa métrica pode não refletir corretamente a performance do modelo. Por isso, também foi utilizado o índice Kappa, que mede a concordância entre as previsões e os valores reais, descontando o efeito do acaso. Valores próximos de 1 indicam alta concordância, o que torna o Kappa uma métrica essencial para avaliar modelos de segmentação binária.

III. RESULTADOS

A acurácia global entre a máscara verdadeira e a predição usando as redes neurais classificadoras U-Net e LSTM de referência (sem o uso de métodos de fusão) é igual a 92% para ambas, conforme visto na Tabela I. Ambas as redes utilizaram o mesmo conjunto de dados tanto para o treino quanto para a predição. Já o índice Kappa para a rede U-Net é igual a 0,78

e para a rede LSTM é igual a 0,75. Tais valores indicam uma concordância substancial entre os resultados do classificador e os rótulos reais, i. e., os classificadores em questão apresentam um bom desempenho na identificação correta das classes, mesmo considerando a chance de acertos ao acaso, dados ruidosos e classes desbalanceadas.

TABELA I
RESULTADOS DO ÍNDICE KAPPA E ACURÁCIA PARA DIFERENTES MÉTODOS DE FUSÃO COM U-NET E LSTM

Método	U-Net		LSTM	
	Kappa	Acurácia	Kappa	Acurácia
U-Net	0,778	0,922	-	-
LSTM	-	-	0,754	0,922
Média	0,677	0,898	0,697	0,889
PCA	0,777	0,923	0,710	0,906
Piramidal	0,682	0,889	0,544	0,868
DWT Haar	0,745	0,910	0,681	0,902
DWT Daubechies	0,696	0,892	0,557	0,877

TABELA II
COMPARATIVO ENTRE AS REDES U-NET E LSTM COM BASE NA ACURÁCIA E ÍNDICE KAPPA MÉDIOS

Rede	Acurácia Média	Kappa Médio
U-Net	0,922	0,778
LSTM	0,922	0,754

Em seguida, conforme visto na Tabela I e na Tabela III, os métodos de fusão de imagens PCA, DWT Haar e Média apresentaram as maiores métricas de avaliação (tanto individualmente, quanto pelos valores médios entre as duas redes) para a tarefa de classificação de regiões cafeeiras analisadas nesse trabalho. Os métodos em questão utilizaram um número reduzido de dados, mantendo os valores das métricas avaliadas próximos aos valores de referência. Dessa forma, com uma representação espectral reduzida em 8, 8 e 7 bandas, respectivamente, é possível obter resultados semelhantes àqueles utilizados como base com 14 bandas (12 bandas ópticas e 2 bandas SAR). Por fim, analisando o desempenho geral das redes, tem-se que a rede neural U-Net apresentou uma acurácia média igual à rede LSTM. No entanto, a U-Net apresentou um maior índice Kappa, conforme pode ser visto na Tabela II.

TABELA III
RESULTADOS MÉDIOS DE ACURÁCIA E ÍNDICE KAPPA PARA DIFERENTES MÉTODOS DE FUSÃO

Método de Fusão	Acurácia Média	Kappa Médio
U-Net	0,922	0,778
LSTM	0,922	0,754
Média	0,893	0,687
PCA	0,914	0,744
Piramidal	0,879	0,613
DWT Haar	0,906	0,713
DWT Daubechies	0,885	0,627

Os resultados obtidos estão diretamente relacionados tanto

à natureza das redes utilizadas quanto às características dos dados empregados. Como os dados se referem a uma única imagem adquirida em um único dia, a variação temporal é praticamente inexistente, o que limita significativamente a capacidade da rede LSTM de extrair informações temporais relevantes. Além disso, nesse contexto específico, os dados espectrais, como texturas, contornos e combinações de bandas, se mostram mais importantes para a tarefa de classificação do que informações temporais. Por essa razão, a U-Net, por ser uma rede convolucional especializada na extração de padrões espaciais e espectrais locais, tende a apresentar melhor desempenho. Essa vantagem é ainda mais acentuada considerando-se que as características do relevo da região de Machado (marcado por um terreno ondulado a montanhoso) influenciam diretamente a distribuição das culturas agrícolas e a resposta espectral das imagens. Essa complexidade espacial favorece a atuação de redes convolucionais como a U-Net, que conseguem capturar tais variações com maior eficácia. Como exemplo, as Figuras 1 e 2 apresentam, respectivamente, as regiões de café reais e as regiões de café preditas pelo modelo U-Net, utilizando a técnica de fusão PCA. Além disso, a Figura 3 destaca, em amarelo, as áreas onde ocorreram classificações errôneas em relação aos rótulos de referência.

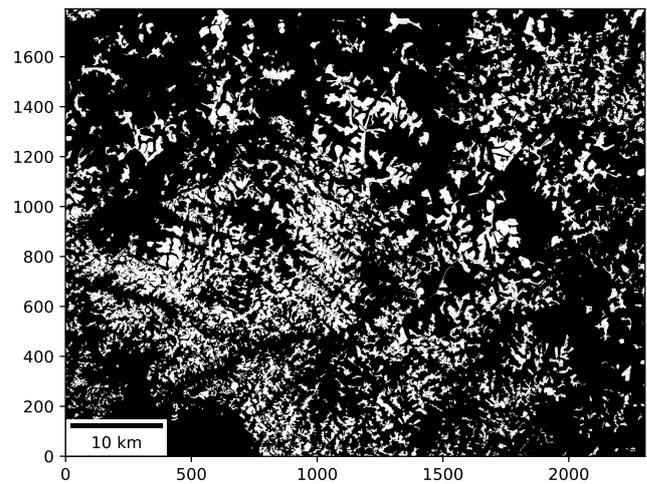


Fig. 1. Verdade terrestre da região de Machado-MG com as áreas de plantio de café em destaque (em branco).

Os resultados indicam que a combinação de redes convolucionais como a U-Net com técnicas de fusão de imagens oferece uma abordagem robusta e eficiente para a classificação de áreas agrícolas, especialmente em regiões com características complexas de relevo e cultivo.

IV. CONCLUSÕES

Este trabalho apresentou a utilização da técnica de fusão de imagens multiespectrais e SAR aplicadas à classificação de áreas cafeeiras em Machado-MG, utilizando redes neurais U-Net e LSTM. Entre os métodos avaliados, a combinação da U-Net com a técnica de fusão via PCA apresentou os melhores resultados, i.e., acurácia de 92,3% e índice Kappa de 0,78, evidenciando o potencial da redução dimensional com preservação de informações relevantes. A U-Net apresentou

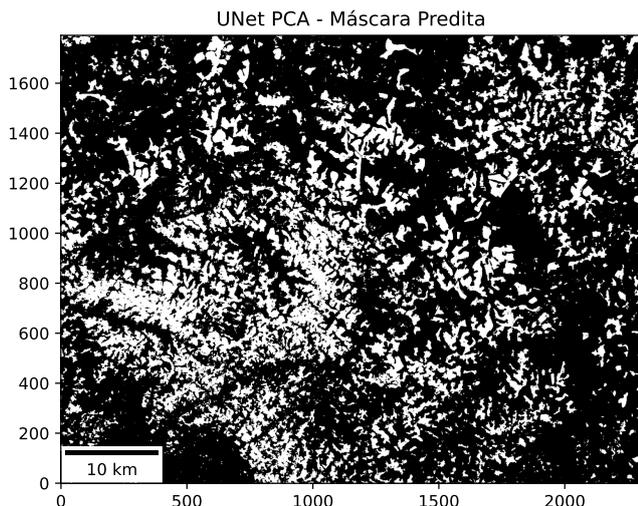


Fig. 2. Regiões preditas (em branco) pelo algoritmo de fusão PCA e rede classificadora U-Net da região de Machado-MG.

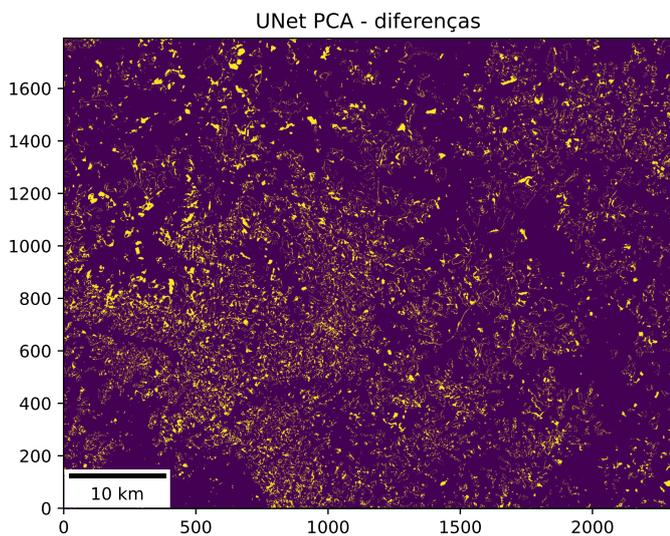


Fig. 3. Diferenças (em amarelo) entre a verdade terrestre e as regiões preditas utilizando a técnica de fusão PCA e a rede classificadora U-Net.

melhores resultados em relação à LSTM, reforçando a importância de abordagens convolucionais em contextos predominantemente espectrais e espaciais, mas com baixa variabilidade temporal, como é o caso das imagens de sensoriamento remoto utilizadas. Os achados reforçam a viabilidade do uso de técnicas de fusão espectral e inteligência artificial para o mapeamento agrícola, contribuindo para uma gestão mais eficiente e sustentável da produção cafeeira, possibilitando uma menor complexidade computacional devido à compactação dos dados. Trabalhos futuros podem explorar séries temporais, inclusão de dados meteorológicos e dados referentes aos diferentes estágios de plantio, bem como a inclusão de outros índices como o Índice de Vegetação por Diferença Normalizada (do inglês, NDVI, *Normalized Difference Vegetation Index*) com o objetivo de enriquecer a representação das condições vegetativas e potencialmente aprimorar o desempenho dos modelos

de classificação utilizando-se a fusão de imagens para uma compactação dos dados.

AGRADECIMENTOS

Agradecemos à pesquisadora Dra. Renata Ribeiro do Valle Gonçalves pelo suporte técnico e científico durante a realização deste trabalho, e ao Centro de Pesquisas Meteorológicas e Climáticas Aplicadas à Agricultura (CEPAGRI) pelo apoio institucional.

REFERÊNCIAS

- [1] M. L. Fung, M. Z. Chen, and Y. H. Chen, "Sensor fusion: A review of methods and applications," in *2017 29th Chinese Control And Decision Conference (CCDC)*. IEEE, 2017, pp. 3853–3860.
- [2] T. Stathaki, *Image fusion: algorithms and applications*. Elsevier, 2011.
- [3] S. Krishnamoorthy and K. Soman, "Implementation and comparative study of image fusion algorithms," *International Journal of Computer Applications*, vol. 9, no. 2, pp. 25–35, 2010.
- [4] S. Bedi and R. Khandelwal, "Comprehensive and comparative study of image fusion techniques," *International Journal of Soft Computing and Engineering (IJSCE)*, vol. 3, no. 1, pp. 2231–2307, 2013.
- [5] W. Ma, K. Wang, J. Li, S. X. Yang, J. Li, L. Song, and Q. Li, "Infrared and visible image fusion technology and application: A review," *Sensors*, vol. 23, no. 2, p. 599, 2023.
- [6] D. Sundararajan, *The discrete Fourier transform: theory, algorithms and applications*. World Scientific, 2001.
- [7] H. Li, S. Liu, Q. Duan, and W. Li, "Application of multi-sensor image fusion of internet of things in image processing," *Ieee Access*, 2018.
- [8] O. Barrero and S. A. Perdomo, "RGB and multispectral UAV image fusion for gramineae weed detection in rice fields," *Precision Agriculture*, vol. 19, pp. 809–822, 2018.
- [9] P. Porwik and A. Lisowska, "The Haar-wavelet transform in digital image processing: its status and achievements," *Machine Graphics and Vision*, vol. 13, no. 1/2, pp. 79–98, 2004.
- [10] J. Le Moigne, "Multi-sensor image registration, fusion and dimension reduction," *Online journal of space communication*, 2003.
- [11] M. Chi, A. Plaza, J. A. Benediktsson, Z. Sun, J. Shen, and Y. Zhu, "Big data for remote sensing: Challenges and opportunities," *Proceedings of the IEEE*, vol. 104, no. 11, pp. 2207–2219, 2016.
- [12] J. Adrian, V. Sagan, and M. Maimaitijiang, "Sentinel SAR-optical fusion for crop type mapping using deep learning and Google Earth Engine," *ISPRS Journal of Photogrammetry and Remote Sensing*, 2021.
- [13] D. Nakazone and M. S. M. Saes, "O agronegócio café do Brasil no mercado internacional," *Revista FAE Business*, vol. 9, pp. 40–42, 2004.
- [14] P. P. Coltri *et al.*, "Classificação de áreas de café em Minas Gerais por meio do novo algoritmo QMAS em imagem espectral Geoeye-1," *Simpósio Brasileiro de Sensoriamento Remoto*, 2011.
- [15] R. L. Cordeiro *et al.*, "QuMinS: Fast and scalable querying, mining and summarizing multi-modal databases," *Information Sciences*, 2014.
- [16] Sentinel Hub, <https://www.sentinel-hub.com>, Sinergise Solutions d.o.o., a Planet Labs Company. Acesso em: abr. 2025.
- [17] EO Browser, <https://apps.sentinel-hub.com/eo-browser/>, Sinergise Solutions d.o.o., a Planet Labs Company. Acesso em: abr. 2025.
- [18] Companhia Nacional de Abastecimento (CONAB). Mapeamentos agrícolas. 2017. <https://portaldeinformacoes.conab.gov.br/mapamentos-agricolas.html>. Acesso em: abr. 2025.
- [19] N. Kussul, M. Lavreniuk, and L. Shumilo, "Deep recurrent neural network for crop classification task based on sentinel-1 and sentinel-2 imagery," in *IGARSS*. IEEE, 2020.
- [20] S. Wei *et al.*, "Multi-temporal sar data large-scale crop mapping based on u-net model," *Remote Sensing*, 2019.
- [21] F. H. Wagner *et al.*, "Using the u-net convolutional network to map forest types and disturbance in the atlantic rainforest with very high resolution images," *Remote Sensing in Ecology and Conservation*, 2019.
- [22] M. R. Rezaei-Dastjerdehei, A. Mijani, and E. Fatemizadeh, "Addressing imbalance in multi-label classification using weighted cross entropy loss function," in *2020 27th national and 5th International Iranian Conference on Biomedical Engineering (ICBME)*. IEEE, 2020.
- [23] Y. Cai, H. Lin, and M. Zhang, "Mapping paddy rice by the object-based random forest method using time series sentinel-1/sentinel-2 data," *Advances in Space Research*, vol. 64, no. 11, pp. 2233–2244, 2019.