

# MAPEAMENTO DO PARQUE CAFEIEIRO EM RIO BANANAL-ES POR MEIO DE TÉCNICAS DE SENSORIAMENTO REMOTO ORBITAL E MACHINE LEARNING

Renato Corrêa Taques<sup>1</sup>; Fernando Soares de Oliveira<sup>2</sup>; Samuel de Assis Silva<sup>3</sup>

**Resumo** – A previsão de safra é essencial para a gestão eficiente da produção agrícola, contribuindo para a redução da incerteza do mercado e estabilidade econômica. Nesse contexto, a estimativa precisa da área colhida é crucial para projeção da safra cafeeira, já que a área cultivada impacta diretamente no potencial produtivo regional. Métodos tradicionais de mapeamento, como a vetorização manual por fotointerpretação, são custosos e demorados, enquanto algoritmos tradicionais de classificação geralmente apresentam baixa acurácia. Métodos mais avançados, integrando sensoriamento remoto e aprendizado de máquina (*machine learning*), podem fornecer resultados mais precisos e detalhados. Este estudo objetivou desenvolver e avaliar estratégias para estimar a área colhida de café no município de Rio Bananal, localizado no norte do Espírito Santo, utilizando imagens de satélite e técnicas avançadas de classificação. O processamento das imagens foi realizado no ambiente *Google Earth Engine* (GEE), com o algoritmo Random Forest, implementado em JavaScript. Foram utilizadas quatro imagens Sentinel-2A, adquiridas no primeiro semestre de 2024, das quais foram extraídas dez bandas espectrais originais, 14 índices espectrais derivados e bandas probabilísticas do produto *Dynamic World V1*, formando as variáveis preditoras do modelo. Para o treinamento e validação do modelo, foram selecionadas aleatoriamente 501 regiões, classificadas quanto ao uso e cobertura do solo por meio de fotointerpretação de imagens PlanetScope (jan. a abr. de 2024), a partir das quais foram gerados 8 mil pontos amostrais. A acurácia foi avaliada por validação cruzada, utilizando a matriz de confusão para calcular as acurácias global, do usuário e do produtor. O modelo estimou uma área de 21.226 hectares cultivados com café para 2024, comparável aos resultados de estudos anteriores realizados pelo Instituto Jones dos Santos Neves (20.084 ha, em 2019–2020) e ao Levantamento Sistemático da Produção Agrícola do IBGE (18.100 ha, em 2023). A classificação apresentou conformidade global de 92%, e, especificamente para a classe café, as acurácias do usuário e do produtor foram de 87% e 90%, respectivamente. Os resultados demonstram que a abordagem utilizada possui elevado potencial para estimar áreas cultivadas de café com alta precisão, podendo contribuir para a previsão de safras e planejamento regional. Contudo, sugere-se a utilização de séries temporais mais extensas, a inclusão de amostras obtidas diretamente em campo para minimizar eventuais erros decorrentes da fotointerpretação e aumentar ainda mais a confiabilidade das estimativas.

**Palavras-chaves:** classificação supervisionada; *Coffea canephora*; monitoramento agrícola; computação em nuvem.

<sup>1</sup>M.Sc. Solos e Nutrição de Plantas, Pesquisador do Incaper, renato@incaper.es.gov.br

<sup>2</sup>D.Sc. Ciência Florestal, Pesquisador do Incaper

<sup>3</sup>Engenharia Agrícola, Professor do CCAE-Ufes

## MAPPING THE COFFEE FIELD IN RIO BANANAL-ES USING ORBITAL REMOTE SENSING AND MACHINE LEARNING TECHNIQUES

**Abstract** – Harvest forecasting is essential for the efficient agricultural production management, contributing to reducing market uncertainty and enhancing economic stability. In this context, accurate estimates of the harvested area are crucial for projecting the coffee harvest, as the cultivated area directly impacts regional production potential. Traditional mapping methods, such as manual vectorization through photointerpretation, are costly and time-consuming, while traditional classification algorithms generally show low accuracy. More advanced methods, that integrate remote sensing and machine learning, can provide results that are more accurate and more detailed. This study aimed to develop and evaluate strategies for estimating the harvested coffee area in the municipality of Rio Bananal, located in northern Espírito Santo, using satellite imagery and advanced classification techniques. Image processing was performed on Google Earth Engine (GEE) using the Random Forest algorithm implemented in JavaScript. Four Sentinel-2A images acquired in the first half of 2024 were used, from which ten original spectral bands, 14 derived spectral indices, and probabilistic bands from the Dynamic World V1 product were extracted, forming the model's predictor variables. For model training and validation, 501 regions were randomly selected, classified by land use and land cover through photointerpretation of PlanetScope images (January to April 2024), from which 8,000 sample points were generated. Accuracy was assessed through cross-validation, using the confusion matrix to calculate overall, user, and producer accuracies. The model estimated an area of 21,226 hectares of coffee cultivation by 2024, comparable to the results of previous studies conducted by the Jones dos Santos Neves Institute (20,084 hectares in 2019–2020) and the IBGE's Systematic Survey of Agricultural Production (18,100 hectares in 2023). The classification showed an overall accuracy of 92%, and, specifically for the coffee class, the user and producer accuracies were 87% and 90%, respectively. The results demonstrate that the approach used has great potential for estimating coffee cultivation areas with high precision, potentially contributing to crop forecasting and regional planning. However, it is recommended to use longer time series and include samples obtained directly from the field to minimize potential errors resulting from photointerpretation and further increase the reliability of the estimates.

**Keywords:** supervised classification; *Coffea canephora*; agricultural monitoring; cloud computing.

## MAPEO DEL PARQUE CAFETERO EN RIO BANANAL-ES MEDIANTE TÉCNICAS DE TELEDETECCIÓN ORBITAL Y APRENDIZAJE AUTOMÁTICO (MACHINE LEARNING)

**Resumen** – La previsión de la cosecha es esencial para la gestión eficiente de la producción agrícola, contribuyendo a la reducción de la incertidumbre del mercado y a la estabilidad económica. En este contexto, la estimación precisa de la superficie cosechada es crucial para la proyección de la cosecha de café, ya que la superficie cultivada tiene un impacto directo en el potencial productivo regional. Los métodos de mapeo tradicionales, como la vectorización manual por fotointerpretación, son costosos y consumen mucho tiempo, mientras que los algoritmos de clasificación tradicionales generalmente tienen baja precisión. Los métodos más avanzados, que integran la teledetección y el aprendizaje automático (*machine learning*), pueden proporcionar resultados más precisos y detallados. Este estudio tuvo como objetivo desarrollar y evaluar estrategias para estimar el área cosechada de café en el municipio de Rio Bananal, ubicado en el norte de Espírito Santo, utilizando imágenes de satélite y técnicas avanzadas de clasificación. El procesamiento de las imágenes se realizó en el entorno Google Earth Engine (GEE), con el algoritmo Random Forest, implementado en JavaScript. Se utilizaron cuatro imágenes Sentinel-2A, adquiridas en el primer semestre de 2024, de las cuales se extrajeron diez bandas espectrales originales, 14 índices espectrales derivados y bandas probabilísticas del producto Dynamic World V1, formando las variables predictoras del modelo. Para el entrenamiento y la validación del modelo, se seleccionaron aleatoriamente 501 regiones, clasificadas según el uso y la cobertura del suelo mediante la fotointerpretación de imágenes PlanetScope (enero a abril de 2024), a partir de las cuales se generaron 8000 puntos de muestreo. La precisión se evaluó mediante validación cruzada, utilizando la matriz de confusión para calcular la precisión global, la precisión del usuario y la precisión del productor. El modelo estimó una superficie de 21 226 hectáreas cultivadas con café para 2024, comparable a los resultados de estudios anteriores realizados por el Instituto Jones dos Santos Neves (20 084 hectáreas, en 2019–2020) y al Levantamiento Sistemático de la Producción Agrícola del IBGE (18 100 hectáreas en 2023). La clasificación presentó una precisión global del 92 %, específicamente para la clase de café, las precisiones del usuario y del productor fueron del 87 % y del 90 %, respectivamente. Los resultados demuestran que el enfoque utilizado tiene un gran potencial para estimar con gran precisión las áreas cultivadas de café, lo que puede contribuir a la previsión de cosechas y a la planificación regional. Sin embargo, se sugiere el uso de series temporales más extensas y la inclusión de muestras obtenidas directamente en el campo para minimizar posibles errores derivados de la fotointerpretación y aumentar aún más la fiabilidad de las estimaciones.

**Palabras clave:** clasificación supervisada; *Coffea canephora*; monitorización agrícola; computación en la nube.

## INTRODUÇÃO

O Espírito Santo se destaca como um dos principais polos cafeeiros do Brasil, ocupando a liderança nacional na produção de café conilon (*Coffea canephora*) e a segunda posição geral na produção de café, atrás apenas de Minas Gerais. Em 2024, o Espírito Santo encerrou a safra com um volume total de 13,9 milhões de sacas, representando um crescimento significativo em relação ao ciclo anterior. Para 2025, a estimativa é que sejam produzidas 16 milhões de sacas, que representa um crescimento de 18,2% na produção total de café do estado. A maior parte dessa produção corresponde ao conilon, evidenciando a importância dessa variedade para a economia capixaba (Conab, 2025).

A estimativa precisa da área colhida é essencial para prever a safra cafeeira, pois permite o planejamento adequado da logística de escoamento, a definição de políticas públicas e a tomada de decisões estratégicas pelos produtores e agentes do mercado. Métodos tradicionais de levantamento, baseados em questionários e entrevistas, apresentam limitações em abrangência e precisão, principalmente em regiões com extensa área territorial e sistemas produtivos diversificados. Esses métodos enfrentam ainda dificuldades relacionadas ao custo, à escalabilidade e à frequência das atualizações.

A estimativa da área de plantio de café, por meio de técnicas de sensoriamento remoto e aprendizado de máquina, tem se mostrado uma abordagem eficiente para o monitoramento agrícola em larga escala. Esses métodos permitem integrar informações espaciais e espectrais, possibilitando identificar e quantificar com maior precisão as áreas cultivadas, especialmente em regiões heterogêneas como o Espírito Santo.

O sensoriamento remoto refere-se à aquisição de informações sobre a superfície terrestre sem contato direto, utilizando sensores que capturam a energia refletida ou emitida pelos objetos. Imagens obtidas por satélites, como as da missão Sentinel-2, fornecem séries temporais de dados multiespectrais com alta resolução, fundamentais para discriminar alvos agrícolas (Jensen, 2006; ESA, 2015).

Aprendizado de máquina é um ramo da inteligência artificial que desenvolve algoritmos capazes de aprender padrões a partir de dados, sem necessidade de programação explícita. Entre os métodos amplamente utilizados para classificação de uso e cobertura do solo, está o *Random Forest* (RF), algoritmo de aprendizado supervisionado baseado na construção de múltiplas

árvores de decisão, sendo amplamente utilizado na classificação de áreas agrícolas por sua robustez e eficácia ao lidar com dados espectrais e derivados (Breiman, 2001).

O processamento em nuvem, como viabilizado pela plataforma *Google Earth Engine* (GEE), revolucionou a análise geoespacial, oferecendo acesso facilitado a vastos conjuntos de imagens e alta capacidade computacional remota. Essa abordagem possibilita o desenvolvimento de modelos escaláveis e replicáveis, com maior frequência de atualização e abrangência espacial (Gorelick et al., 2017).

Diversos estudos têm demonstrado o potencial da combinação entre imagens de satélite e aprendizado de máquina para estimar áreas cultivadas com alta precisão.

Schneider et al. (2025) destacaram a eficácia da utilização de séries temporais Sentinel-2 combinadas com coerência espacial, obtendo acurácia superior a 90% na detecção e mapeamento de café. Kelley, Pitcher e Bacon (2018) alcançaram acurácia global de 90,5% na classificação de áreas complexas com café sob sombra na Nicarágua, utilizando imagens Landsat 8 e o algoritmo *Random Forest* no GEE. Além disso, Tridawati, Aidarna e Wikantika (2020) também reforçaram a importância da integração de dados multitemporais e multirresolução, variáveis texturais e topográficas, alcançando acurácia global de, aproximadamente, 79,3% no mapeamento de áreas cafeeiras. Esses estudos exemplificam a confiabilidade e a eficiência alcançáveis com o uso integrado dessas tecnologias para o monitoramento agrícola.

Além das aplicações em monitoramento agrícola, destaca-se a crescente importância do mapeamento das áreas de produção de café para atendimento às exigências de mercados internacionais, como o Regulamento da União Europeia para Produtos Livres de Desmatamento (EUDR – *European Union Regulation for Deforestation-Free Products*). O EUDR requer que commodities agrícolas exportadas para a União Europeia, como o café, sejam livres de desmatamento e estejam em conformidade com a legislação ambiental local, incluindo o Código Florestal Brasileiro. Dessa forma, torna-se essencial mapear com precisão as lavouras e verificar a regularidade ambiental das propriedades, reforçando o papel do sensoriamento remoto e da análise geoespacial na certificação e rastreabilidade da produção (União Europeia, 2023).

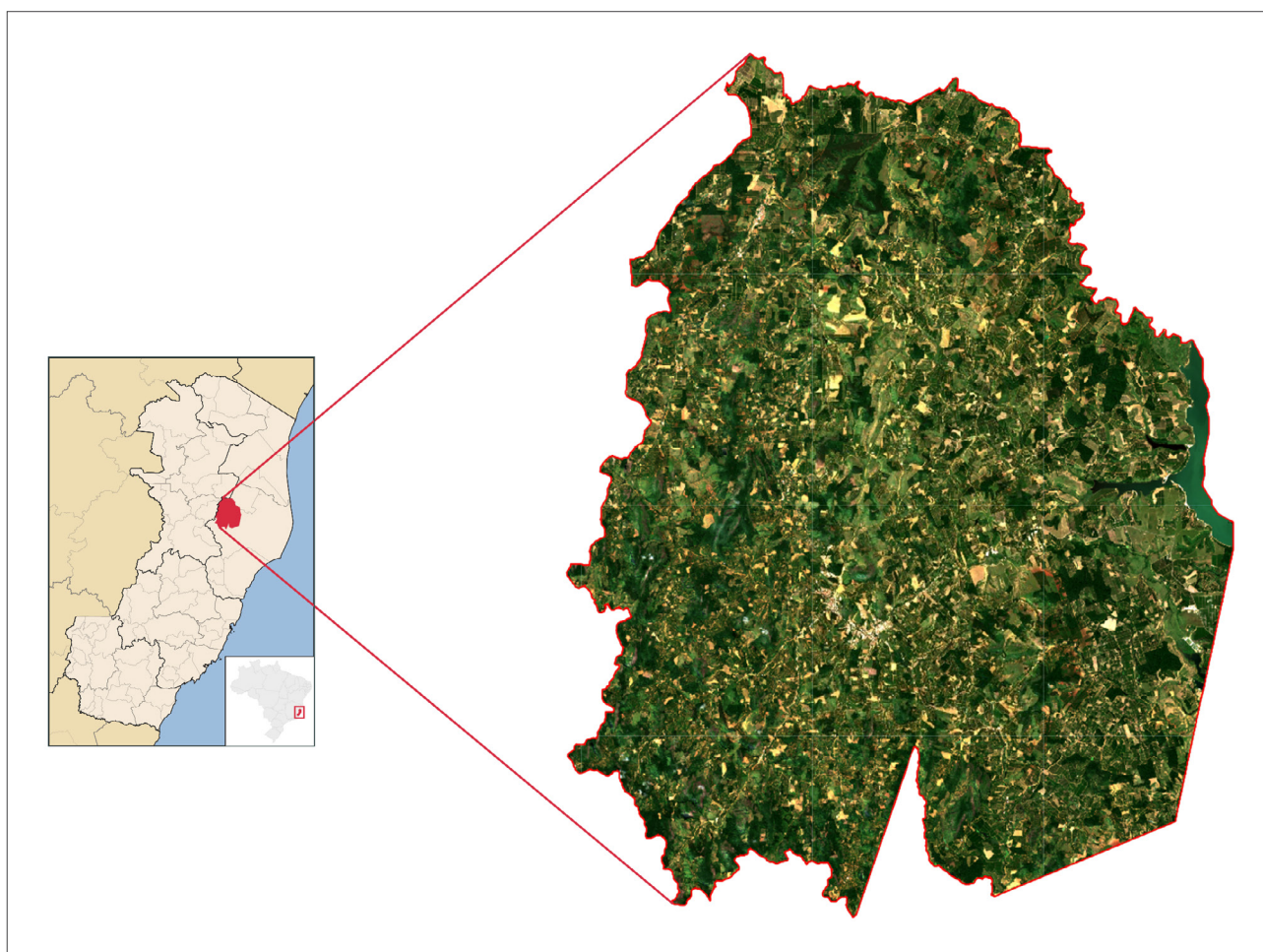
Diante desse contexto, este artigo propôs-se a abordar o desafio de estimar a área colhida de café por meio da aplicação integrada de técnicas de sensoriamento remoto e aprendizado de máquina. Foram utilizadas imagens multiespectrais Sentinel-2A e índices derivados, aliados às bandas probabilísticas do produto Dynamic World V1, para compor um conjunto de variáveis de entrada ao algoritmo Random Forest. Essa abordagem permitiu desenvolver e testar um modelo de classificação voltado à identificação dos talhões de café em produção na safra de 2024, explorando o potencial do processamento em nuvem disponibilizado pela plataforma Google Earth Engine. Ao adotar essa estratégia, o estudo busca demonstrar a viabilidade técnica e científica de métodos automatizados e escaláveis para o mapeamento agrícola, contribuindo para o aprimoramento de sistemas de monitoramento e previsão de safra no estado do Espírito Santo.

## MATERIAL E MÉTODOS

### ÁREA DE ESTUDO

O estudo foi conduzido no município de Rio Bananal, situado na região Norte do Espírito Santo. Com uma área territorial de, aproximadamente, 645,4 km<sup>2</sup>, o município corresponde a cerca de 0,98% do território estadual (Figura 1). Localiza-se a cerca de 175 km da capital, Vitória, fazendo parte da microrregião do Rio Doce (IBGE, 2024).

O relevo de Rio Bananal é formado por colinas e áreas suavemente onduladas, com altitudes que variam de 95 metros até mais de 700 metros na região sudoeste do município. O clima predominante é tropical quente e úmido, apresentando temperatura média anual de 24,2 °C. Rio Bananal está totalmente inserido no bioma da Mata Atlântica e possui cerca de 18% de sua área coberta por vegetação nativa, conforme dados do Atlas da Mata Atlântica do Estado do Espírito Santo: 2007–2008/2012–2015 (Sossai, 2018).



**Figura 1** – Localização do município de Rio Bananal, Espírito Santo, Brasil.

**Fonte:** Adaptado de Wikipédia (2025).

A economia local é predominantemente agrícola, com destaque para a produção de café conilon (*Coffea canephora*), principal atividade econômica do município. A expressiva representatividade de Rio Bananal como uma das principais regiões produtoras de café do Espírito Santo justifica sua escolha como área de estudo.

## IMAGENS DE SATÉLITE

Foram utilizadas duas fontes principais de dados orbitais no desenvolvimento do presente estudo: imagens do satélite PlanetScope e do sensor MSI (*MultiSpectral Instrument*) a bordo do Sentinel-2A. Cada uma dessas fontes foi selecionada em função de suas características espectrais, espaciais e temporais, adequadas aos objetivos de fotointerpretação e modelagem.

As imagens PlanetScope, operadas pela empresa Planet Labs Inc., possuem resolução espacial de, aproximadamente, 5 metros, com cobertura multiespectral em quatro bandas: R (vermelho); G (verde); B (azul); e NIR (infravermelho próximo). Com alta frequência de revisita, podendo gerar imagens diárias dependendo da localização, essas imagens foram utilizadas para a etapa de fotointerpretação visual das feições, permitindo a identificação precisa das classes de uso e cobertura do solo (Planet, 2019). Foram selecionadas cenas adquiridas entre os meses de janeiro e abril de 2024, priorizando imagens com baixa cobertura de nuvens.

As imagens Sentinel-2A, por sua vez, foram utilizadas como base para a classificação automática. Esse satélite integra o programa europeu Copernicus, operado pela Agência Espacial Europeia (ESA), e conta com o sensor MSI, que fornece 13 bandas espectrais com resoluções espaciais de 10 m, 20 m e 60 m, distribuídas nos espectros visível, infravermelho próximo (NIR) e infravermelho de ondas curtas (SWIR). As imagens apresentam um ciclo de revisita de 5 dias em combinação com o Sentinel-2B, o que permite a geração de séries temporais consistentes para o monitoramento agrícola (ESA, 2015).

A composição da imagem Sentinel-2A utilizada neste estudo foi obtida a partir de uma coleção de cenas multiespectrais (*ImageCollection*) correspondente ao período de janeiro a junho de 2024. Inicialmente, foram aplicados filtros espaciais e temporais para selecionar apenas as imagens que cobrem a área de interesse (município de Rio Bananal) e que apresentam menos de 1% de cobertura de nuvens. Em seguida, foram

selecionadas as bandas espectrais de interesse (B2 a B12) e aplicadas funções de correção de escala e cálculo de índices espectrais. Por fim, as cenas selecionadas foram combinadas por meio do operador de mediana, que sintetiza os valores dos pixels ao longo da série temporal, reduzindo a influência de ruídos e eventuais artefatos atmosféricos, resultando em uma imagem composta representativa do período agrícola analisado.

## COLETA E PREPARAÇÃO DOS DADOS

A construção do banco de dados de amostras para o treinamento e validação do modelo de classificação teve início com a delimitação manual de polígonos representativos das classes de uso e cobertura do solo na área de estudo. Essa etapa foi realizada por meio de fotointerpretação visual de imagens PlanetScope.

As atividades de fotointerpretação e edição dos polígonos foram executadas no *software* QGIS Desktop, versão 3.38, utilizando como base o conjunto de imagens corrigidas e georreferenciadas previamente inseridas no sistema. A interpretação foi feita por analistas com experiência em mapeamento agrícola e conhecimento prévio da região, o que contribuiu para a acurácia na delimitação dos polígonos. Foram atribuídas informações a cada feição vetorial criada sobre a classe de uso e cobertura do solo.

Para garantir a representatividade espacial das amostras e evitar viés de seleção, foram definidos 501 polígonos distribuídos aleatoriamente ao longo da área de estudo, respeitando a proporção das principais classes observadas (Tabela 1). As amostras foram organizadas em um *shapefile* vetorial com atributos associados, como ID, classe temática e coordenadas. Esse conjunto de dados serviu como base para a extração de pontos amostrais e posterior modelagem na plataforma GEE.

**Tabela 1** – Distribuição dos polígonos amostrais por classe de uso e cobertura do solo utilizados no mapeamento

(continua)

CLASSE	DESCRIÇÃO	Nº DE AMOSTRAS	ÁREA HECTARES
Café (CF)	Talhões de café com vegetação vigorosa, organizada em fileiras.	120	82,76
Corpo d'água (AG)	Incluem rios, lagos, reservatórios e áreas inundadas permanentes.	51	58,97
Floresta (FL)	Fragmentos de vegetação natural preservada, com dossel fechado.	83	259,11

(conclusão)			
CLASSE	DESCRIÇÃO	Nº DE AMOSTRAS	ÁREA HECTARES
Formação natural não florestal (NF)	Inclui vegetações naturais de menor porte, como campos nativos e de altitude, áreas de arbustos, e de afloramento rochoso.	29	61,35
Outras lavouras (OL)	Engloba cultivos de espécies temporárias e perenes.	31	24,97
Pastagem (PA)	São caracterizadas pela presença de vegetação herbácea cultivada ou natural, destinada à alimentação de animais de criação.	67	129,37
Silvicultura (SI)	Áreas de plantios florestais homogêneos, como eucalipto, seringueira e pinus.	51	97,08
Área não vegetada (NV)	Áreas com ausência de cobertura vegetal, como solos expostos, áreas urbanizadas, e áreas de mineração.	69	77,66
<b>TOTAL</b>		<b>501</b>	<b>791,27</b>

## PROCESSAMENTO DAS IMAGENS E MODELAGEM DE DADOS

O processamento dos dados e a construção do modelo de classificação foram realizados na plataforma Google Earth Engine, a qual permite acesso a grandes volumes de imagens de satélite e oferece infraestrutura computacional em nuvem para análises geoespaciais em larga escala (Gorelick et al., 2017).

Foram inicialmente gerados 8 mil pontos amostrais, distribuídos em quantidades iguais entre as classes de uso e cobertura do solo, com o objetivo de evitar desequilíbrios no treinamento do modelo. Cada ponto herdou automaticamente o rótulo da classe correspondente, permitindo sua associação direta à variável resposta da classificação.

Em seguida, foram selecionadas imagens Sentinel-2A do primeiro semestre de 2024, todas com menos de 1% de cobertura de nuvens. Para cada cena, foram utilizadas dez bandas espectrais originais do sensor MSI (B2, B3, B4, B5, B6, B7, B8, B8A, B11 e B12). Além dessas bandas, foram calculados 14 índices espectrais (Tabela 2).

**Tabela 2** – Índices espectrais e relação com classes de uso e cobertura do solo

(continua)

ÍNDICE ESPECTRAL	PRINCIPAL APLICAÇÃO	CLASSES DE USO E COBERTURA RELACIONADAS
NRI – Nitrogen Reflectance Index	Utilizado para estimar a quantidade de nitrogênio na vegetação, sendo útil para avaliar a fertilidade do solo e o estado nutricional das plantas.	Lavouras, pastagens
GI – Greenness Index	Mede a intensidade da cor verde da vegetação, auxiliando na distinção entre áreas de vegetação densa e degradada.	Vegetação densa, lavouras
GLI – Green Leaf Index	Indica a quantidade de folhas verdes na cobertura vegetal, sendo útil para estudos de biomassa e produtividade agrícola.	Floresta, lavouras
NDVI – Normalized Difference Vegetation Index	Índice amplamente utilizado para monitoramento da vegetação, calculado a partir da diferença normalizada entre as bandas do infravermelho próximo (NIR) e vermelho (RED).	Vegetação em geral
EVI – Enhanced Vegetation Index	Variante do NDVI, reduzindo os efeitos da atmosfera e da refletância do solo, sendo mais sensível em áreas de vegetação densa.	Floresta, áreas agrícolas
LAI – Leaf Area Index	Representa a área foliar da vegetação por unidade de solo, sendo essencial para estudos de crescimento e produtividade agrícola.	Vegetação arbórea e agrícola
NDWI – Normalized Difference Water Index	Aplicado para identificação e monitoramento de corpos d'água, utilizando a diferença entre as bandas do verde (GREEN) e infravermelho próximo (NIR).	Corpos d'água
MNDWI – Modified NDWI	Variante do NDWI, que melhora a separação entre corpos d'água e superfícies urbanas, utilizando a banda do SWIR em substituição ao NIR.	Áreas urbanas, corpos d'água
GCVI – Green Chlorophyll Vegetation Index	Indica a concentração de clorofila na vegetação, auxiliando na análise da saúde das plantas.	Floresta, lavouras

(conclusão)

ÍNDICE ESPECTRAL	PRINCIPAL APLICAÇÃO	CLASSES DE USO E COBERTURA RELACIONADAS
PRI - Photochemical Reflectance Index	Relacionado à eficiência do uso da luz pelas plantas e à sua resposta ao estresse ambiental.	Vegetação sob estresse
NBR - Índice de Diferença da Queimadura Normalizado	Utilizado para identificar áreas afetadas por queimadas ou incêndios florestais.	Áreas não vegetadas, vegetação nativa
SAVI - Soil-Adjusted Vegetation Index	Similar ao NDVI, mas ajustado para minimizar a influência do solo em áreas com pouca vegetação.	Solo exposto, vegetação rala
AVI - Agricultural Vegetation Index	Específico para monitoramento da vegetação agrícola, proporcionando melhor sensibilidade a variações de crescimento.	Áreas agrícolas
BSI - Bare Soil Index	Destinado à detecção de áreas de solo exposto, útil para estudos de desertificação e degradação ambiental.	Áreas desmatadas, solos expostos

Como complemento, o modelo de classificação incorporou também as bandas de probabilidade do produto Dynamic World V1 (DW). Esse produto, gerado a partir de modelos de aprendizado profundo aplicados a imagens Sentinel-2, fornece para cada pixel a probabilidade de pertencimento a oito classes de uso e cobertura do solo, a saber: *water* (água), *trees* (vegetação arbórea), *grass* (vegetação herbácea), *flooded vegetation* (vegetação alagada ou inundada), *crops* (cultivo agrícola), *shrub and scrub* (vegetação arbustiva ou matagal), *built* (construções ou áreas urbanizadas), e *bare* (solo exposto) (Brown et al., 2022). A inclusão dessas bandas de probabilidade como variáveis auxiliares enriquece o conjunto de dados de entrada, permitindo ao algoritmo de *machine learning* identificar padrões mais sutis e melhorar a separação entre classes temáticas, especialmente em áreas com assinaturas espectrais semelhantes ou coberturas vegetais heterogêneas.

A preparação dos dados envolveu a extração dos valores das bandas espectrais, dos índices derivados e das probabilidades do DW para cada ponto amostral, compondo o conjunto com 32 variáveis preditoras, fornecendo uma base abrangente para a modelagem (Equação 1).

$$\text{Classe} = f(B2, B3, B4, B5, B6, B7, B8, B8A, B11, B12, NRI, GI, GLI, NDVI, EVI, LAI, NDWI, MNDWI, GCVI, PRI, NBR, SAVI, AVI, BSI, \textit{water}, \textit{trees}, \textit{grass}, \textit{flooded\_vegetation}, \textit{crops}, \textit{shrub\_and\_scrub}, \textit{built}, \textit{bare}) \quad (1)$$

Nesta expressão, Classe refere-se à categoria de uso e cobertura do solo atribuída a cada pixel ou ponto amostral. A função *f* representa o modelo de classificação (neste caso, Random Forest), que utiliza como variáveis preditoras:

- Dez bandas espectrais originais do sensor Sentinel-2 MSI: B2, B3, B4, B5, B6, B7, B8, B8A, B11 e B12.
- Catorze índices espectrais derivados: NRI, GI, GLI, NDVI, EVI, LAI, NDWI, MNDWI, GCVI, PRI, NBR, SAVI, AVI e BSI.
- Oito bandas de probabilidade do produto Dynamic World V1: *water*, *trees*, *grass*, *flooded vegetation*, *crops*, *shrub and scrub*, *built* e *bare*.

O modelo de classificação analisou os valores das 32 variáveis de entrada para cada unidade espacial, aprendendo padrões espectrais e espaciais que permitem distinguir as classes temáticas da área de estudo. O algoritmo Random Forest (RF) foi treinado para reconhecer as assinaturas espectrais de cada classe de uso e cobertura do solo, utilizando 250 árvores de decisão, cinco variáveis avaliadas por divisão, no mínimo quatro amostras por nó terminal e fração amostral de 0,6 por árvore. Essa configuração buscou otimizar o desempenho do modelo, assegurando equilíbrio entre precisão e capacidade de generalização.

Após o ajuste dos parâmetros, o modelo RF treinado foi aplicado à coleção de imagens contendo as variáveis preditoras, resultando na classificação dos pixels de toda a área de estudo. Para melhorar a qualidade espacial do resultado obtido, foram aplicados filtros morfológicos (erosão seguida de dilatação), visando remover pequenos agrupamentos isolados de pixels classificados erroneamente e consolidar áreas homogêneas das classes mapeadas.

## VALIDAÇÃO CRUZADA E ACURÁCIA DA CLASSIFICAÇÃO

A avaliação da qualidade do modelo de classificação é etapa fundamental para garantir a confiabilidade dos resultados obtidos. Neste estudo, o desempenho do modelo RF foi aferido por meio da análise da matriz de confusão, da acurácia global, da acurácia do usuário, da acuracidade do produtor, indicadores amplamente utilizados em estudos de sensoriamento remoto e ciência dos dados (Congalton, 1991; Kelley et al., 2018; MapBiomass, 2025; Tridawati et al., 2020).

A matriz de confusão consiste em uma tabela que relaciona as classificações preditas pelo modelo com as classes de referência observadas nos dados de validação. Cada elemento da matriz representa a frequência de ocorrência entre classes verdadeiras e preditas, permitindo identificar acertos e erros de classificação.

A acurácia global ( $ag$ ) é o percentual de pontos de validação corretamente classificados, calculada pela razão entre o número de acertos e o número total de amostras, conforme a Equação 2:

$$ag = \frac{\sum_{i=1}^k x_{ii}}{N} \quad (2)$$

em que  $x_{ii}$  é o número de amostras corretamente classificadas para a classe  $i$ ,  $k$  é o número de classes e  $N$  é o total de amostras.

A acuracidade do usuário ( $au$ ) indica a probabilidade de um pixel classificado em determinada classe realmente pertencer a essa classe no terreno, sendo sensível aos erros de comissão, de acordo com a Equação 3:

$$au = \frac{x_{ii}}{\sum_{j=1}^k x_{ij}} \quad (3)$$

em que  $x_{ij}$  é o número de amostras da classificadas como  $i$  na classe de referência  $j$ .

A acurácia do produtor ( $ap$ ) mede a probabilidade de um pixel de referência de uma classe ser corretamente classificado, sendo sensível aos erros de omissão, conforme a Equação 4:

$$ap = \frac{x_{ii}}{\sum_{j=1}^k x_{ji}} \quad (4)$$

em que  $x_{ji}$  é o número de amostras da classe de referência  $i$  classificadas como  $j$ .

Para aumentar a consistência da avaliação, foi empregado o procedimento de validação cruzada  $k$ -fold, com  $k = 10$ , no qual o conjunto de dados de validação foi particionado aleatoriamente em dez subconjuntos de

tamanho semelhante. O modelo foi ajustado dez vezes, cada vez utilizando nove subconjuntos para treinamento e um para validação. Uma matriz de confusão foi obtida em cada rodada e, ao final, somou-se os valores das matrizes para o cálculo dos indicadores. Esse procedimento reduz o risco de viés decorrente da divisão específica dos dados e fornece estimativas mais estáveis de desempenho do modelo (Kohavi, 1995).

## RESULTADOS E DISCUSSÃO

O modelo desenvolvido neste estudo estimou uma área total de 21.226 hectares cultivados com café no município de Rio Bananal para a safra de 2024, representando cerca de 33% da área territorial total do município (Figura 2). Esse resultado demonstra a relevância econômica e territorial que a cafeicultura possui na região.

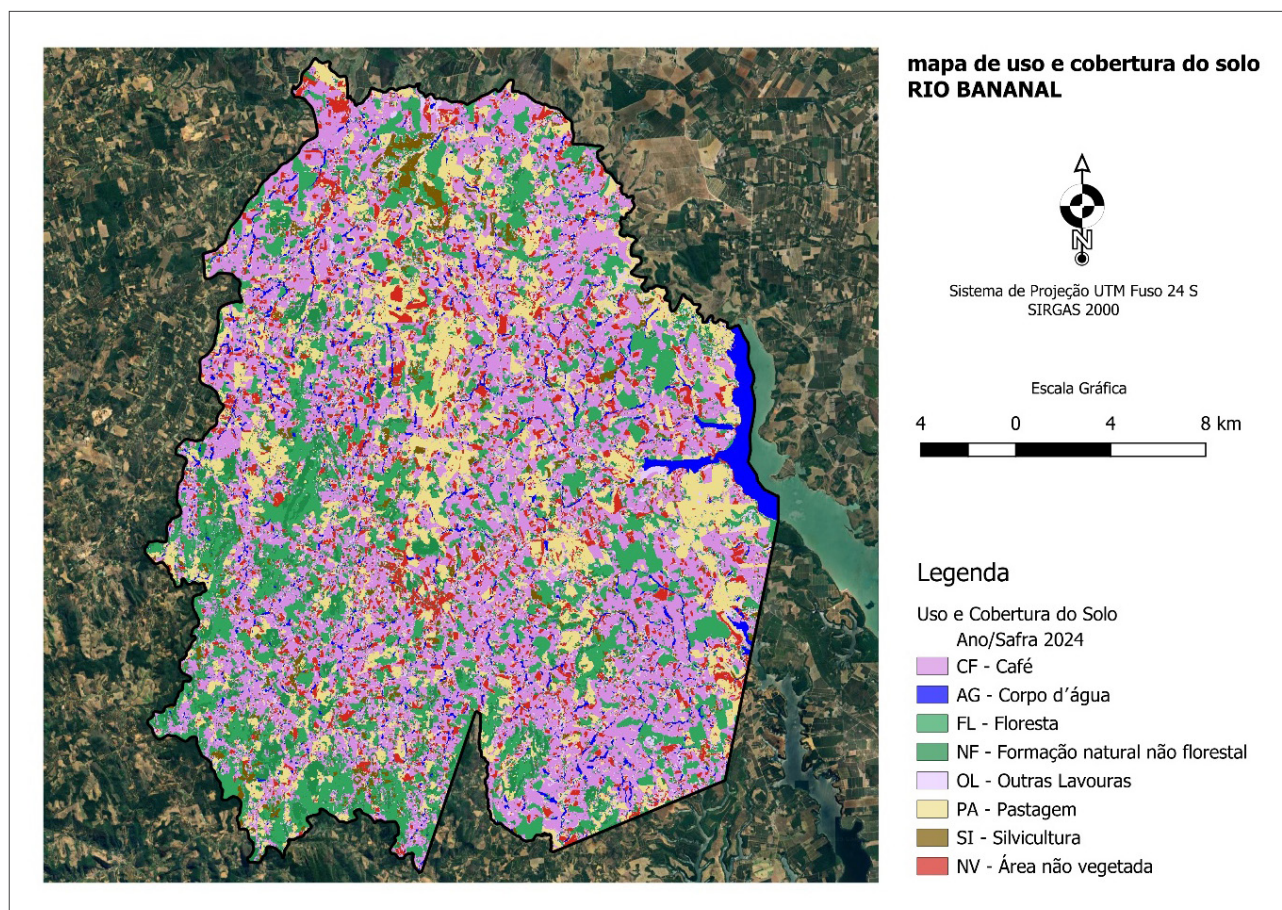
Para efeito comparativo, o mapeamento de uso e cobertura do solo realizado pelo Instituto Jones dos Santos Neves (IJSN), no período 2019–2020, registrou uma área cultivada com café de, aproximadamente, 20.084 hectares (IJSN, 2024). Considerando a proximidade temporal entre os levantamentos, observa-se uma relativa estabilidade nas áreas destinadas ao café, com ligeiro aumento identificado pelo presente estudo, possivelmente associado à expansão das áreas produtivas.

Em contrapartida, os dados provenientes do Levantamento Sistemático da Produção Agrícola (LSPA) de 2023, conduzido pelo IBGE, estimaram uma área colhida de café em Rio Bananal em torno de 18.100 hectares (IBGE, 2025). Essa discrepância pode ser explicada pelas diferenças metodológicas entre os estudos, uma vez que o levantamento do IBGE se baseia principalmente em entrevistas e declarações de produtores, enquanto o presente estudo utilizou métodos objetivos de classificação de imagens de satélite com elevada resolução espacial e temporal.

A matriz de confusão da validação cruzada para o mapeamento das classes de uso e cobertura do solo apresenta uma acurácia global de 92,6%, indicando um desempenho robusto do modelo RF na classificação das classes de mapeamento. Os valores de acurácia do usuário e do produtor (Tabela 3) apresentaram índices elevados, com destaque para corpo d'água (AG) e floresta (FL), ambas com acurácia do produtor acima de 95%, demonstrando a eficácia do modelo na distinção dessas coberturas, que apresentam assinaturas

espectrais bem definidas. Resultados similares foram encontrados por Kelley, Pitcher e Bacon (2018), ao utilizarem o Random Forest em imagens Landsat 8 para mapear áreas de café sob sombra em paisagens

complexas da Nicarágua, alcançando 90,5% de acurácia global, e por Schneider et al. (2025), que aplicaram séries temporais Sentinel-2 com coerência espacial para detectar cafezais, obtendo acurácia superior a 90%.



**Figura 2** – Mapa de uso e cobertura do solo do município de Rio Bananal-ES, de 2024.

**Tabela 3** – Indicadores da acurácia da classificação

CLASSE	ACURÁCIA DO USUÁRIO (%)	ACURÁCIA DO PRODUTOR (%)
Café (CF)	87,0	90,7
Corpo d'água (AG)	99,2	96,6
Floresta (FL)	91,7	95,7
Formação natural não florestal (NF)	91,1	92,1
Outras lavouras (OL)	94,6	92,1
Pastagem (PA)	93,2	95,4
Silvicultura (SI)	93,9	87,7
Área não vegetada (NV)	91,0	90,7

A classe café (CF) apresentou acurácia do usuário de 87,0% e acuracidade do produtor de 90,7%. Esses valores indicam que o modelo possui boa capacidade tanto de identificar corretamente as áreas de café (produtor)

quanto de evitar a inclusão de áreas de outras classes como café (usuário). Ainda que a acurácia do usuário para café seja ligeiramente inferior às demais classes, o desempenho é considerado satisfatório, especialmente

diante dos desafios associados à similaridade espectral entre o café e outras culturas perenes ou áreas de vegetação secundária.

Os valores obtidos nesse estudo são comparáveis, e, em alguns casos, superiores aos relatados na literatura para mapeamentos de uso e cobertura do solo baseados em sensoriamento remoto e aprendizado de máquina, como os apresentados por Schneider *et al.* (2025) e Kelley, Pitcher e Bacon (2018), reforçando a aplicabilidade e eficácia do modelo desenvolvido para o mapeamento da cafeicultura regional.

### CONSIDERAÇÕES FINAIS

O estudo realizado demonstrou a eficácia da utilização de sensoriamento remoto orbital e técnicas de aprendizado de máquina (Random Forest) na estimativa da área cultivada com café conilon no município de Rio Bananal, Espírito Santo. O método adotado proporcionou resultados detalhados, destacando-se pela acurácia global elevada (92,6%) e pela coerência com informações oficiais previamente publicadas.

Contudo, algumas limitações e restrições devem ser consideradas. Apesar da solidez dos resultados, variações fenológicas da cultura, semelhança espectral entre café e outras culturas perenes ou vegetação secundária e a presença de áreas de sombreamento nas imagens podem afetar pontualmente a precisão da classificação.

Para futuros estudos, sugere-se a utilização de séries temporais mais extensas de imagens Sentinel-2, cobrindo diferentes fases fenológicas da cultura, o que poderia melhorar a distinção entre classes espectralmente similares. Recomenda-se também a incorporação de dados auxiliares obtidos em campo, como informações detalhadas sobre práticas agrícolas ou manejo das lavouras, visando melhorar a precisão do modelo. Além disso, a aplicação integrada de técnicas avançadas de processamento de imagens e aprendizado profundo (em inglês, *deep learning*) pode ser investigada como alternativa para aumentar ainda mais a precisão e a confiabilidade dos resultados obtidos.

### CONTRIBUIÇÃO DOS AUTORES

Renato Corrêa Taques coletou e processou os dados, analisou os resultados e escreveu o artigo. Fernando Soares de Oliveira e Samuel de Assis Silva revisaram o artigo.

### CONFLITO DE INTERESSES

Os autores declaram não haver conflito de interesses.

### AGRADECIMENTOS

Agradecemos à Secretaria da Agricultura, Abastecimento, Aquicultura e Pesca (Seag), à Organização Internacional Ítalo-Latino-Americana (IILA) e à Fundação de Amparo à Pesquisa e Inovação do Espírito Santo (Fapes) pelo apoio financeiro ao projeto, assim como à Universidade Federal do Espírito Santo (Ufes) e ao Incaper pelo suporte institucional e infraestrutura disponibilizada para a realização deste trabalho.

### REFERÊNCIAS

- BREIMAN, L. Random forests. **Machine Learning**, v. 45, p. 5–32, 2001. DOI:10.1023/A:1010933404324.
- BROWN, C. F. *et al.* Dynamic World, near real-time global 10 m land use land cover mapping. **Scientific Data**, v. 9, p. 251, 2022. DOI:10.1038/s41597-022-01307-4 2.
- CONAB – Companhia Nacional de Abastecimento. **Acompanhamento da safra brasileira: café – safra 2025 – segundo levantamento**, maio 2025. Brasília: Conab, 2025. Disponível em: <https://www.gov.br/conab/pt-br>. Acesso em: 4 jun. 2025.
- CONGALTON, R. G.; GREEN, K. **Accuracy Assessment of Remotely Sensed Data: Principles and Practices**. 3. ed. Boca Raton: CRC Press, 2019.
- ESA – European Space Agency. Sentinel-2 User Handbook. **ESA Standard Document**, 2015. Disponível em: [https://sentinels.copernicus.eu/documents/247904/685211/Sentinel-2\\_User\\_Handbook](https://sentinels.copernicus.eu/documents/247904/685211/Sentinel-2_User_Handbook). Acesso em: 4 jun. 2025.
- GORELICK, N. *et al.* Google Earth Engine: Planetary-scale geospatial analysis for everyone. **Remote Sensing of Environment**, v. 202, p. 18–27, 2017. DOI:10.1016/j.rse.2017.06.031.
- IBGE – Instituto Brasileiro de Geografia e Estatística. **Cidades e Estados: Rio Bananal (ES)**. Rio de Janeiro: IBGE. Disponível em: <https://www.ibge.gov.br/cidades-e-estados/es/rio-bananal.html>. Acesso em: 22 ago. 2024.
- IBGE – Instituto Brasileiro de Geografia e Estatística. **Produção Agrícola Municipal**. Rio de Janeiro: IBGE. Disponível em: <https://www.ibge.gov.br/estatisticas/economicas/agricultura-e-pecuaria.html>. Acesso em: 23 jun. 2025.
- IJSN – Instituto Jones dos Santos Neves. **Mapeamento Uso do Solo ES – 2019–2020**. Vitória: IJSN. Disponível em: [https://ide.geobases.es.gov.br/layers/geonode:ijsn\\_map\\_uso\\_solo\\_es\\_2019\\_2020](https://ide.geobases.es.gov.br/layers/geonode:ijsn_map_uso_solo_es_2019_2020). Acesso em: 2 ago. 2024.
- JENSEN, J. R. **Remote Sensing of the Environment: An Earth Resource Perspective**. 2. ed. Boston: Pearson, 2006.

KELLEY, L. C.; PITCHER, L.; BACON, C. Using Google Earth Engine to map complex shade-grown coffee landscapes in Northern Nicaragua. **Remote Sensing**, v. 10, n. 6, p. 952, 2018. DOI: 10.3390/rs10060952.

KOHAVI, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. *In*: **International Joint Conference On Artificial Intelligence – IJCAI**, 14. Montréal, 1995. p. 1137-1145.

MAPBIOMAS. **Projeto MapBiomias Brasil**. [S.l.]: MapBiomias, [s.d.]. Disponível em: <https://brasil.mapbiomas.org/>. Acesso em: 4 jun. 2025.

PLANET. **Planet Imagery Product Specifications**. São Francisco, CA: Planet Labs Inc., 2019. Disponível em: <https://assets.planet.com/docs/combined-imagery-product-spec-april-2019.pdf>. Acesso em: 3 jun. 2025.

SCHNEIDER, B. O.; ALVES, M. C. M.; FERREIRA, V. A. Coffee crop detection and mapping using Sentinel-2 data and spatial coherence. **Theoretical and Applied Engineering**, v. 9, n. 3, p. 1-10, 2025. DOI: 10.31422/taae.v9i3.64.

SOSSAI, M. F. **Atlas da Mata Atlântica do estado do Espírito Santo: 2007-2008/2012-2015**. Cariacica: Iema, 2018.

TRIDAWATI, A.; AIDARNA, H.; WIKANTIKA, K. Mapping the distribution of coffee plantations from multi-resolution, multi-temporal, and multi-sensor data using a Random Forest algorithm. **Remote Sensing**, v. 12, n. 23, p. 3933, 2020. DOI: 10.3390/rs12233933.

UNIÃO EUROPEIA. Regulamento (UE) 2023/1115 do Parlamento Europeu e do Conselho, de 31 de maio de 2023. **Jornal Oficial da União Europeia**, L 150, p. 206-247, 2023. Disponível em: <https://eur-lex.europa.eu/eli/reg/2023/1115/oj/eng>. Acesso em: 3 jun. 2025.

WIKIPÉDIA. **Localização de Rio Bananal no Espírito Santo**. Wikimedia Commons, 2025. Disponível em: [https://pt.wikipedia.org/wiki/Rio\\_Bananal](https://pt.wikipedia.org/wiki/Rio_Bananal). Acesso em: 15 set. 2025.