

IV SUSTENTARE & VII WIPIS
WORKSHOP INTERNACIONAL
Sustentabilidade, Indicadores e Gestão de Recursos Hídricos
de 16 a 18 de novembro de 2022

EVENTO GRATUITO TOTALMENTE ONLINE

Realização: SUSTENTARE FUD CAMPINAS WIPES IRL-OP

Apoio: Agência das Relações PCJ COMITÊS PCJ

ANÁLISE DA DINÂMICA DO USO E COBERTURA DA TERRA UTILIZANDO O CLASSIFICADOR RANDOM FOREST PARA O DISTRITO DE BENTO RODRIGUES / MARIANA / MG

Fabício Ribeiro Garcia, Universidade Federal do Rio de Janeiro, fabricao.garcia@coc.ufrj.br
Otto Corrêa Rotunno Filho, Universidade Federal do Rio de Janeiro, otto@coc.ufrj.br

Resumo

As informações sobre o uso e cobertura do solo desempenham um papel vital em muitos aspectos sociais, econômicos, políticos e ambientais, visto que estas permitem subsidiar planos de ações. O processamento em nuvem tem se tornado cada vez mais difundido para análise e obtenção de dados ambientais de uso e cobertura do solo. Neste sentido, o objetivo deste trabalho é realizar a classificação de uso e cobertura do solo, para os anos de 2014, 2016 e 2021, para a região de Bento Rodrigues, distrito do município de Mariana/MG, região a qual passou grandes alterações ambientais devido ao rompimento da barragem de Fundão em 2015. O estudo analisou a aplicação do algoritmo de classificação supervisionada Random Forest (RF) através do processamento em nuvem do Google Earth Engine (GEE). A metodologia demonstrou um grande potencial, sendo uma ferramenta versátil na análise de dados ambientais, facilitando a identificação do uso e cobertura do solo, apresentando índices de concordância acima de 90% para todos os períodos analisados.

Palavras-chave: Classificação supervisionada, Machine Learning, Google Earth Engine.

1. Introdução

O uso e cobertura do solo é uma das informações mais relevantes na interpretação de uma imagem de satélite, pois permite a identificação e distribuição espacial do tipo de cobertura que reveste o solo conforme sua natureza. Tal conhecimento é de grande importância ao planejamento por contribuir na orientação da ocupação da paisagem, respeitando sua capacidade de suporte e/ou sua estabilidade/vulnerabilidade.

A cobertura do solo é um fator básico e importante, a qual conecta várias partes do ambiente, neste sentido, o sensoriamento remoto vem se tornando uma ferramenta muito útil e indispensáveis para o monitoramento dessa dinâmica de uso e ocupação do solo, visto que este tem proporcionado a obtenção de dados atualizados com maior agilidade que os métodos tradicionais de observação em campo.

Desse modo, o objetivo deste trabalho foi avaliar o desempenho do classificador supervisionado Random Forest por meio de imagens Landsat através do processamento em nuvem do Google Earth Engine (GEE) para a região de Bento Rodrigues, distrito do município de Mariana/MG, região a qual passou grandes alterações ambientais devido ao rompimento da barragem de Fundão em 2015.



2. Fundamentação teórica

A classificação da cobertura do solo com base em imagens de satélite é uma área que tem se observado um grande avanço das tecnologias de aquisição de imagens de sensoriamento remoto, além disso, a popularização da computação em nuvem vem promovendo a utilização de uma quantidade cada vez maior de dados de sensoriamento remoto (Anil Gavade & Vijay S. Rajpurohit, 2021).

Os métodos de classificação são divididos em dois grupos: os supervisionados e os não supervisionados. A classificação supervisionada tem maior precisão, pois utiliza dados amostrais dos tipos de ocupação do solo que se pretendem levantar e classificar (Oliveira et al., 2014). Os métodos supervisionados usam amostras das classes detectadas para treinar um classificador e, em seguida, usam o classificador treinado para prever os atributos de cada pixel em uma imagem de sensoriamento remoto (Zhao et al., 2021).

Os fatores mais importantes que afetam a precisão dos resultados dos classificadores de imagens são as amostras de treinamento, que devem fornecer informações suficientes para treinar o modelo. Os algoritmos de aprendizado profundo têm recebido maior interesse no campo do processamento de imagens de sensoriamento remoto nos últimos anos. (Anil Gavade & Vijay Rajpurohit, 2021).

Dentre os classificadores mais utilizados recentemente, pode-se destacar o Random Forest, o qual é um algoritmo de aprendizagem supervisionada. O Random Forest é uma técnica de classificação não paramétrica que gera um conjunto de árvores de decisão e estima a importância dos recursos de entrada (David L. Miller, 2022). Segundo NETO (2014), o algoritmo Random Forest origina várias árvores de decisão por meio de conjuntos de atributos escolhidos de forma aleatória em relação ao conjunto original.

3. Metodologia

O trabalho foi desenvolvido para na região do distrito de Bento Rodrigues/ Mariana/ MG; utilizou-se a plataforma do Google Earth Engine (GEE) para processamento dos produtos de sensoriamento remoto Landsat 8. Foram realizadas filtragens dos conjunto de dados, sendo para a filtragem temporal de cada ano analisado, utilizou-se todas as imagens presentes de cada período, a fim de se obter a melhor composição com ausência de nuvens das cenas, e então um melhor resultado para a classificação. Aplicou-se classificador supervisionado Random Forest para os anos de 2014, 2016 e 2021; por fim realizou-se a acurácia do método a partir da avaliação do índice Kappa. A Figura 1 apresenta o fluxograma de desenvolvimento deste trabalho.



IV SUSTENTARE & VII WIPIS

WORKSHOP INTERNACIONAL

Sustentabilidade, Indicadores e Gestão de Recursos Hídricos

de 16 a 18 de novembro de 2022

EVENTO
GRATUITO
TOTALMENTE
ONLINE

Realização:



SUSTENTARE
PUC MINAS



WIPES
UFFPR

Apoio:



Agência das Ráguas PCJ



COMITÊS PCJ

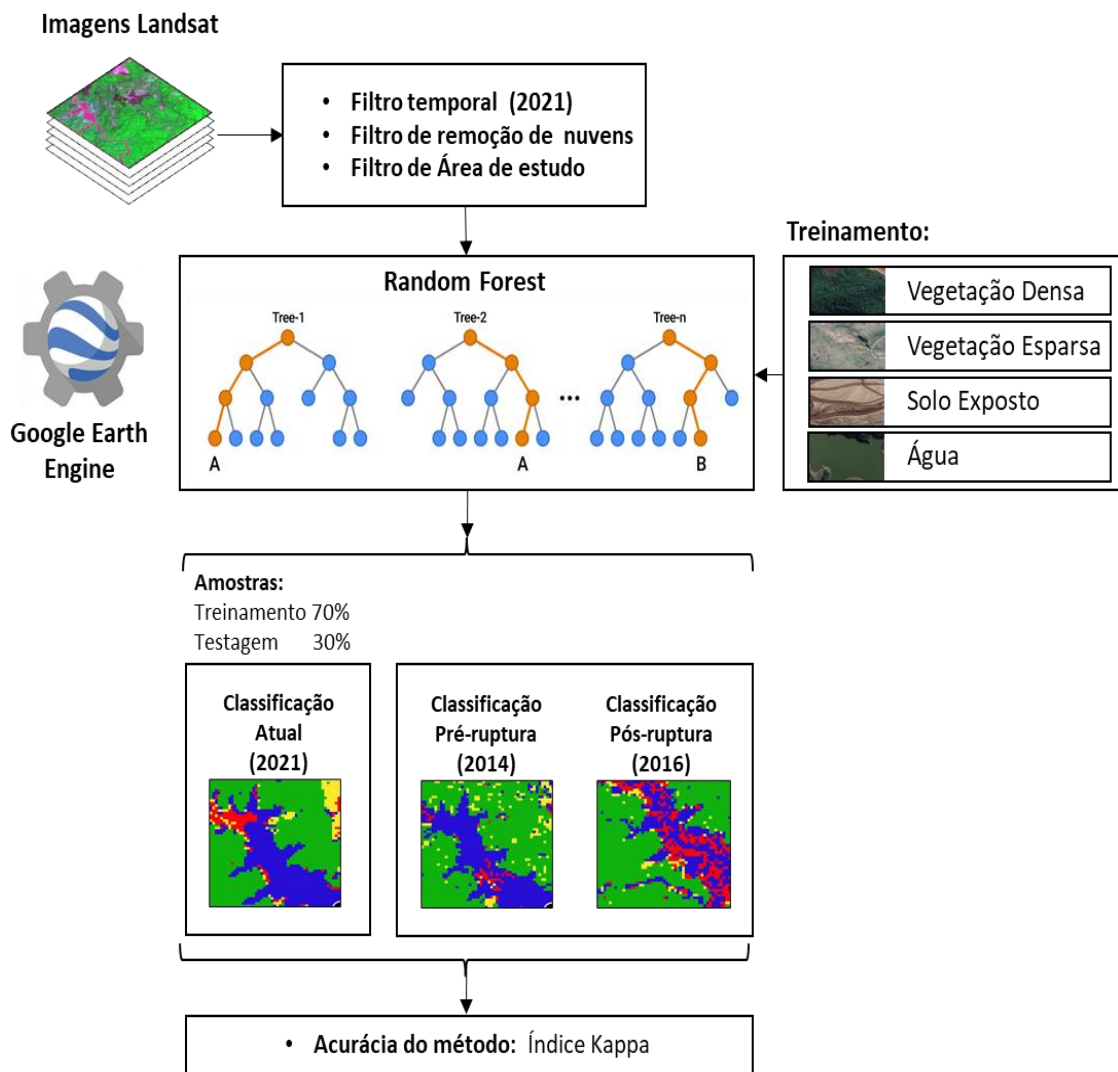


Figura 1: Fluxo metodológico de desenvolvimento do trabalho.

3.1 Área de Estudo

A área de estudo foi a região de Bento Rodrigues, o qual é um subdistrito do município de Mariana/MG. Em 2015, ano da ruptura da barragem da Samarco, Bento Rodrigues possuía uma população de aproximadamente 600 habitantes, os quais ocupavam cerca de 200 imóveis. No subdistrito de Bento Rodrigues está localizada a jusante das barragens de rejeitos de mineração Fundão, Germano e de Santarém, operadas pela empresa mineradora Samarco, conforme Figura 2.



IV SUSTENTARE & VII WIPIS

WORKSHOP INTERNACIONAL

Sustentabilidade, Indicadores e Gestão de Recursos Hídricos

de 16 a 18 de novembro de 2022

EVENTO GRATUITO
TOTALMENTE ONLINE

Realização:



SUSTENTARE PUCMINAS



WIPIS

Apoio:



Agência das Rocias PCJ



COMITÊS PCJ

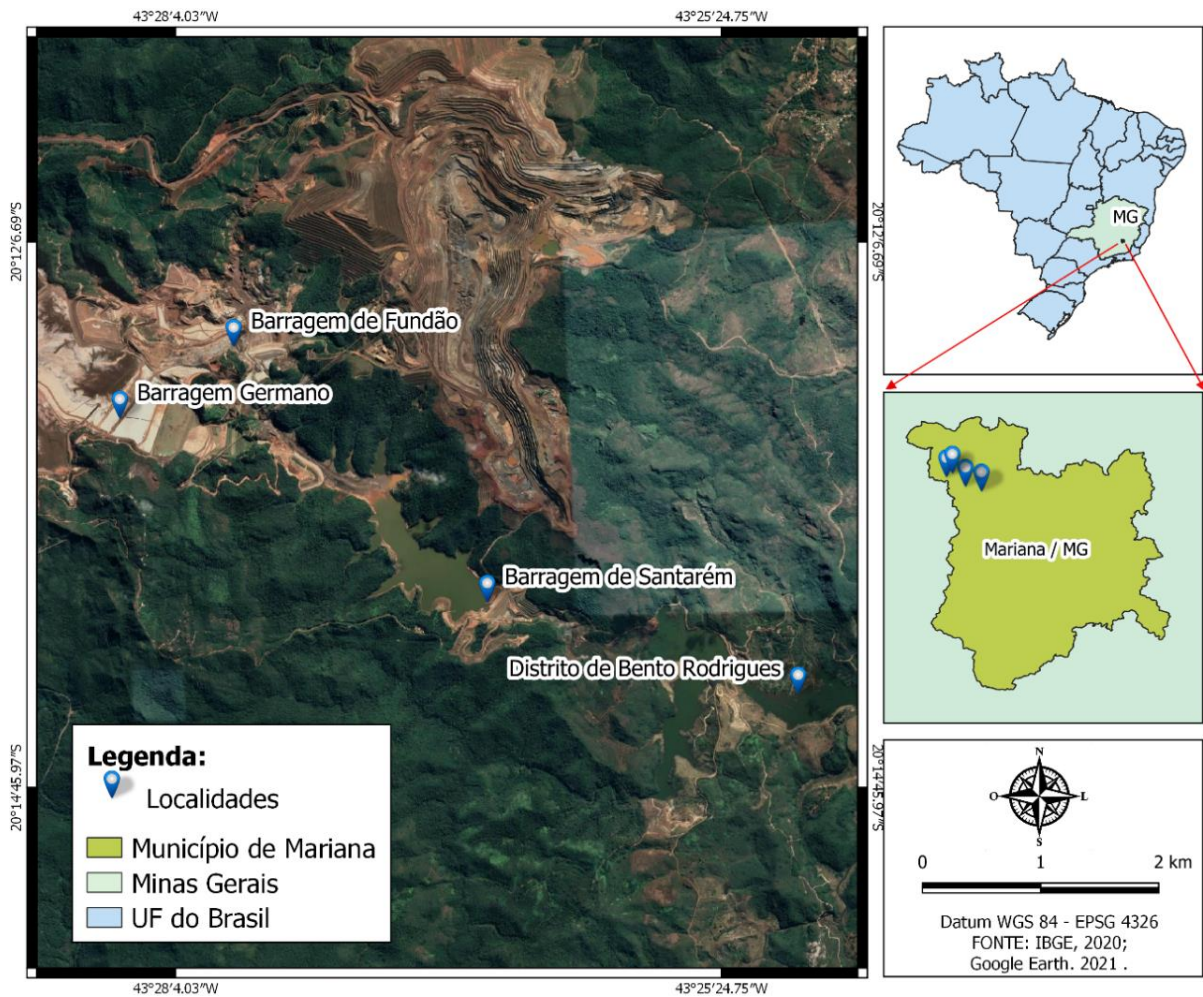


Figura 2: Subdistrito de Bento Rodrigues / Mariana / MG.

Na figura 2 observa-se ainda alterações resultantes do evento do dia 5 de novembro de 2015, quando a barragem do Fundão, da empresa Samarco Mineração S.A. rompeu, assim resultando no maior desastre socioambiental do País.

3.2 – Base de Dados Utilizada

Foram utilizadas as imagens Landsat 8, especificamente o produto LANDSAT/LC08/C01/T1_RT (coleção 1, camada 1, Tempo real), do sensor OLI (Operational Land Imager), constituída pelas bandas 2, 3, 4, 5, 6 e 7, disponibilizada gratuitamente pelo United States Geological Survey (USGS) no site <http://earthexplorer.usgs.gov/>, e acessadas

pela plataforma do Google Earth Engine. Na tabela 1 são apresentadas as bandas do Landsat 8 utilizadas.

Tabela 1. Bandas espectrais do sensor OLI do satélite Landsat 8

Banda	Resolução (m)	Comprimento de Onda (nm)
B2 - Blue	30	450 - 510
B3 - Green	30	530 - 590
B4 - Red	30	640- 670
B5 - Infravermelho próximo	30	850 - 880
B6 - Infravermelho médio 1	30	1570 - 1650
B7 - Infravermelho médio 2	30	2110 - 2290

3.3 Classificador Random Forest

Um mapeamento de uso e cobertura do solo visa basicamente interpretar as características de determinada região. O algoritmo Random Forest necessita ser alimentado por conjuntos de feições de uso e cobertura do solo, para tanto, é necessário identificar as classes presentes. Para este trabalho, foram identificadas 4 classes, sendo estas: vegetação densa, vegetação esparsa, solo exposto e corpos hídricos, como apresentado na figura 3 a seguir.

Tabela 2 – Classes de uso e cobertura do solo identificadas

CLASSE	DESCRIÇÃO	AMOSTRA	COR
Vegetação Densa	Região de floresta densa sem espaçamento		Verde
Vegetação Esparsa	Região de floresta esparsa, campos e gramíneas		Amarelo
Solo Exposto	Solo exposto por ação antrópica		Vermelho
Água	Corpos d'água, rios e lagos		Azul

Essa etapa foi realizada a partir de uma classificação por fotointerpretação visual- a partir de uma escala média de visualização de 1:500- utilizando como camada base as imagens do Google Earth do ano 2021, de onde foram selecionadas as amostras para cada classe de cobertura.

A análise para a classificação inicial partiu dos dados do ano de 2021, onde para realizar o treinamento do algoritmo foram definidas quatro classes principais, sendo elas vegetação densa,

IV SUSTENTARE & VII WIPIS
WORKSHOP INTERNACIONAL
Sustentabilidade, Indicadores e Gestão de Recursos Hídricos
 de 16 a 18 de novembro de 2022

EVENTO GRATUITO TOTALMENTE ONLINE

Realização: SUSTENTARE PUC CAMPINAS

Apoio: Agência das Rocias PCJ

COMITÊS PCJ

vegetação esparsa, solo exposto e corpos hídricos, de onde foram coletadas 100 amostras para cada classe de cobertura, totalizando 400 amostras distribuídas pela imagem de satélite, como apresentado na Figura 3 a seguir.

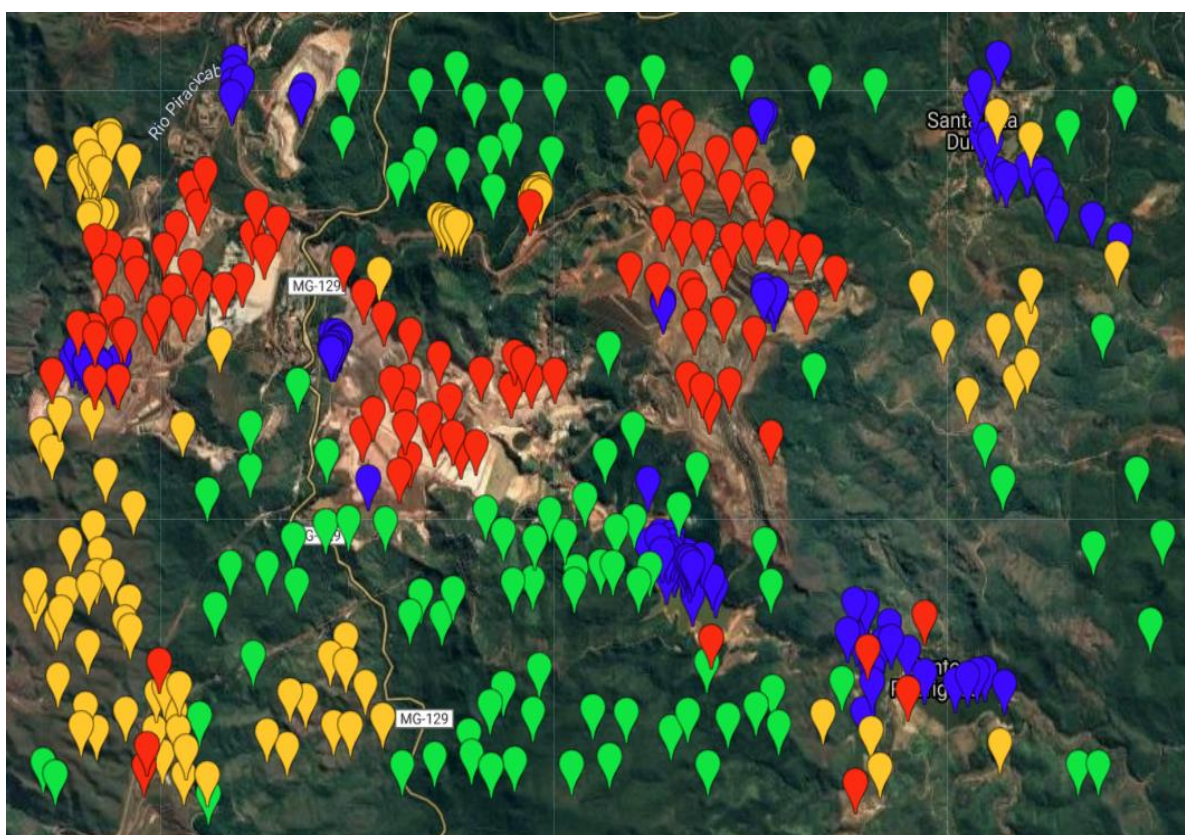


Figura 3: Pontos de amostragem de treinamento e teste.

Do total do conjunto de amostras da imagem do ano de 2021, foram utilizadas 70% para o treinamento e 30% para testes do classificador, e então classificação das imagem referente aos anos de 2014, 2016 e 2021.

4. Resultados

Nas Figura 4 a 6 estão apresentados os resultados da classificação de uso e cobertura do solo para o distrito de Bento Rodrigues pelo Randon Forest. Pela análise visual do resultado da figura 4, da classificação para o ano de 2014 (pré-ruptura das barragens) é possível notar a predominância da classe de vegetação densa; a classe de água caracterizando o reservatório da



IV SUSTENTARE & VII WIPIS

WORKSHOP INTERNACIONAL

Sustentabilidade, Indicadores e Gestão de Recursos Hídricos

de 16 a 18 de novembro de 2022

EVENTO
GRATUITO
TOTALMENTE
ONLINE

Realização:



SUSTENTARE
PUC-MINAS



WIPIS
2007-2019

Apoio:



Agência das Ráguas PCJ



COMITÊS PCJ

barragem de contenção de Santarém, bem como a identificação da classe de solo exposto, referente às áreas de processos de mineração das barragens de rejeito de Germano e Fundão; e a classe de vegetação esparsa concentrada mais a leste da cena, no entorno do distrito de Bento Gonçalves apresentando uma concordância com a Figura 2.

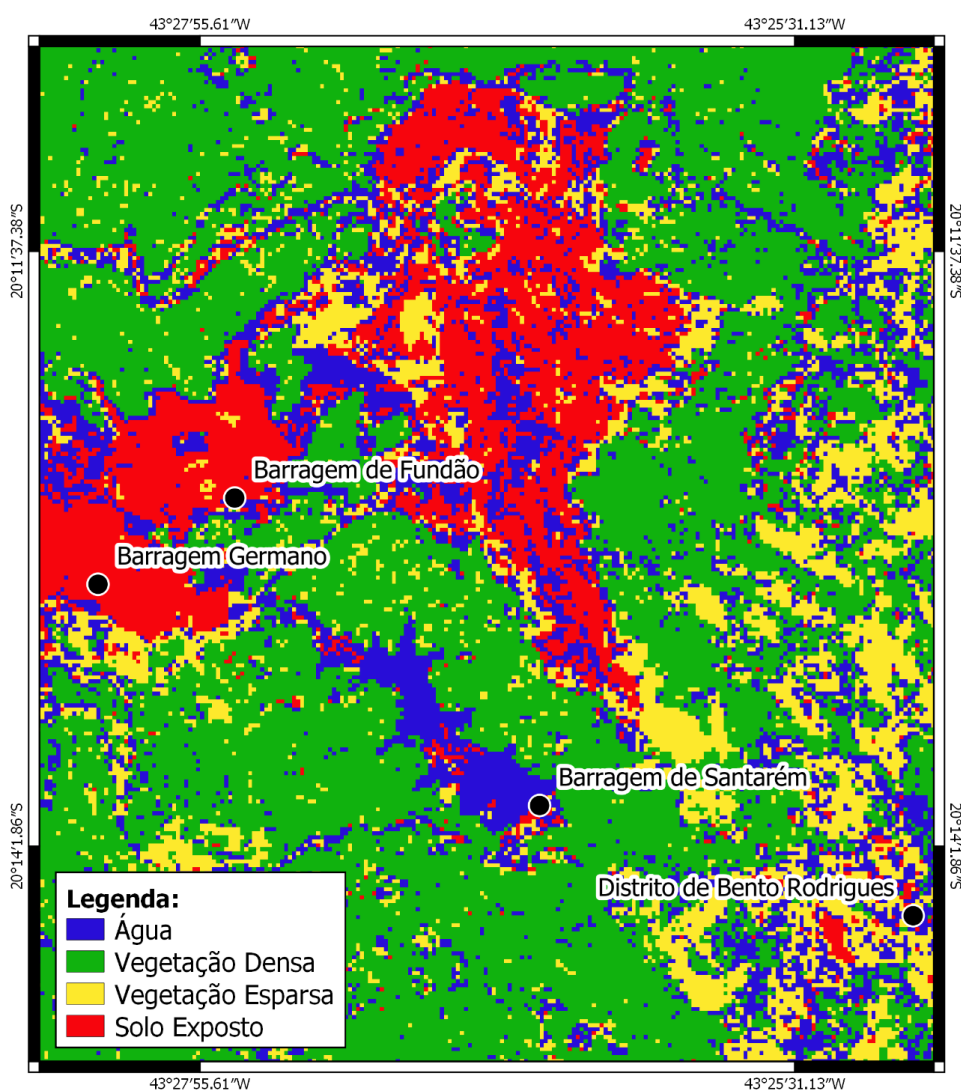


Figura 4: Classificação de uso e cobertura do solo para 2014.

Pela análise visual do resultado da figura 5, da classificação para o ano de 2016 (pós-ruptura das barragens) é possível notar significativas alterações a jusante das barragens de re-

IV SUSTENTARE & VII WIPIS

WORKSHOP INTERNACIONAL

Sustentabilidade, Indicadores e Gestão de Recursos Hídricos

de 16 a 18 de novembro de 2022

EVENTO GRATUITO TOTALMENTE ONLINE

Realização:

Apoio:

jeito rejeitos Germano e Fundão. Observa-se ainda alterações na região do reservatório da barragem de Santarém, onde há um aumento das suas dimensões, e a identificação da classe de solo exposto, o que pode ter sido resultando do deslocamento de material de rejeito das barragens a montante desta, como apresentado na Figura 5 a seguir.

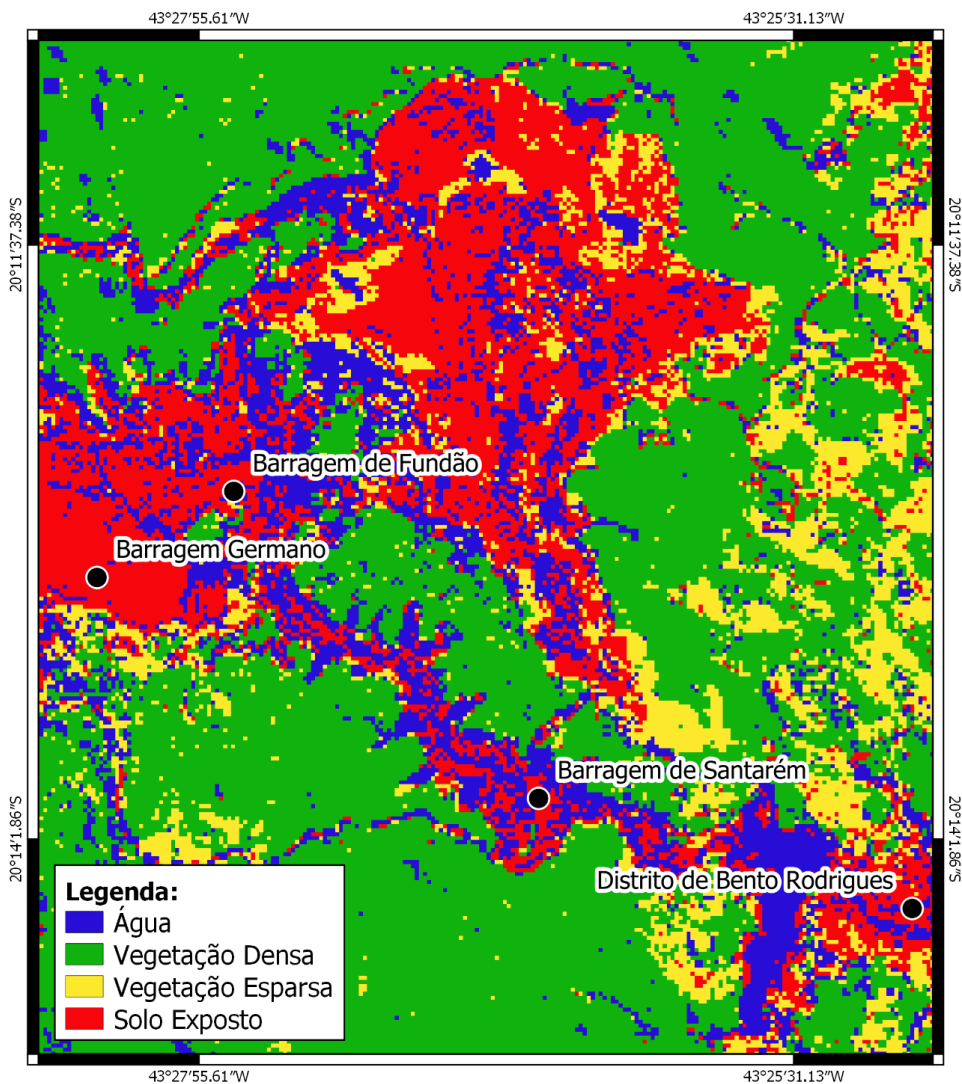


Figura 5: Classificação de uso e cobertura do solo para 2016.

Da classificação do ano de 2016 observa-se ainda uma significativa alteração na região mais próxima ao distrito de Bento Rodrigues, principalmente pelo surgimento de uma grande



IV SUSTENTARE & VII WIPIS

WORKSHOP INTERNACIONAL

Sustentabilidade, Indicadores e Gestão de Recursos Hídricos

de 16 a 18 de novembro de 2022

EVENTO
GRATUITO
TOTALMENTE
ONLINE

Realização:




Apoio:




área da classe de água e uma concentração de solo exposto, classes essa que não eram notadas como predominantes na imagem de 2014, anterior ao evento de rupturas das barragens.

Da Figura 6 abaixo, classificação para o ano de 2021 (6 anos pós-ruptura das barragens) nota-se identificação das classes bem definidas, bem como a estabilização da massa de água na região de Bento Rodrigues, a qual foi identificada na imagem de 2016.

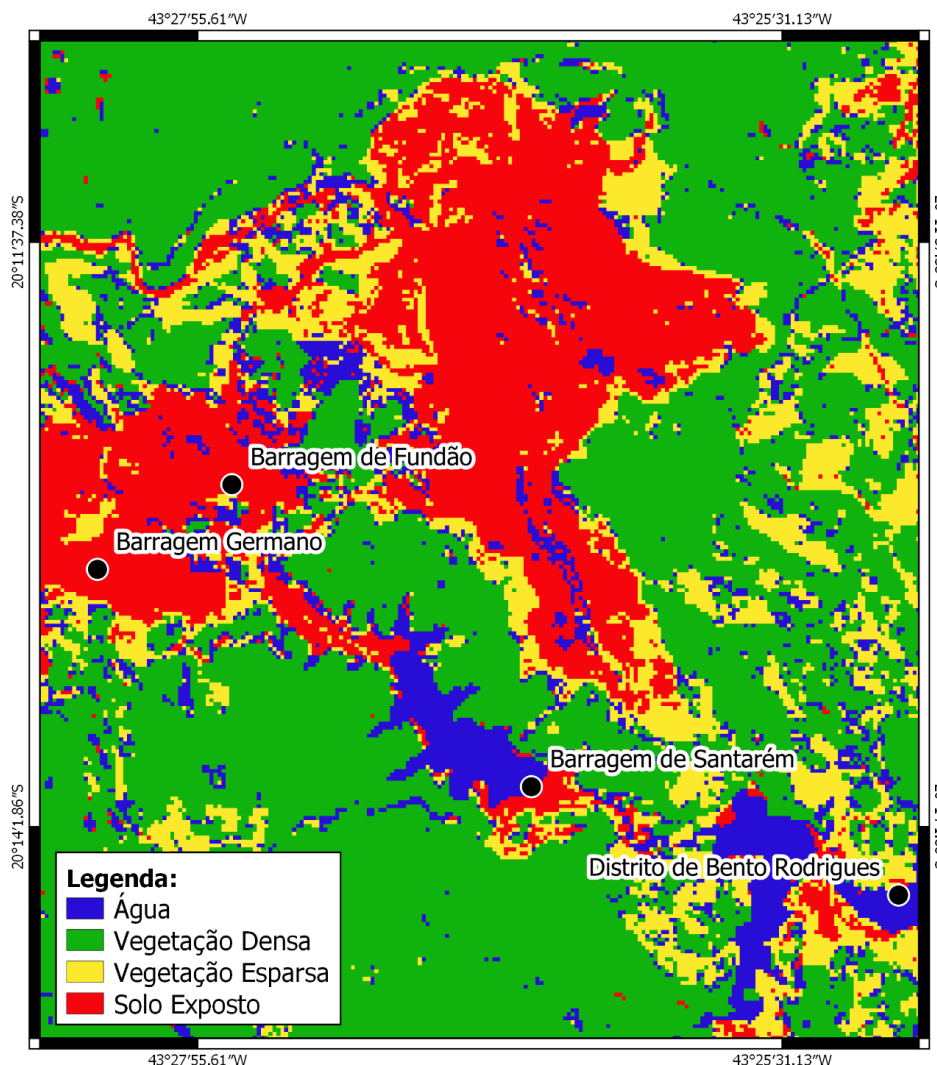


Figura 6: Classificação de uso e cobertura do solo para 2021.

Para a avaliação da metodologia utilizou-se inicialmente a determinação da matriz de erro para avaliar as divergências ocorridas entre as classes temáticas. A matriz de confusão permite



observar na diagonal dos pares concordantes dos pontos na comparação entre o resultado da classificação a partir da imagem de satélite e as classes coletadas em campo. Já as demais células da matriz de confusão indicam o total dos pares que não há concordância dos pontos na comparação. A seguir nas figuras 7a, 7b e 7c são apresentadas as matrizes de erro para os anos de 2014, 2016 e 2021, respectivamente.

Classe Classificada	Classe Real			
	Água	Vegetação Densa	Vegetação Esparsa	Solo Exposto
Água	72	1	0	4
Vegetação Densa	1	65	1	0
Vegetação Esparsa	2	2	63	1
Solo Exposto	3	1	1	63

Classe Classificada	Classe Real			
	Água	Vegetação Densa	Vegetação Esparsa	Solo Exposto
Água	61	1	1	0
Vegetação Densa	0	61	1	0
Vegetação Esparsa	0	0	70	1
Solo Exposto	8	1	2	60

Classe Classificada	Classe Real			
	Água	Vegetação Densa	Vegetação Esparsa	Solo Exposto
Água	61	1	1	0
Vegetação Densa	0	61	1	0
Vegetação Esparsa	0	0	70	1
Solo Exposto	8	1	2	60

A análise direta dos pares acima indica no total de 280 pontos verificados, a concordância em 263 pontos, sendo equivalente a 93,9% do total de pontos verificados, e a não concordância em 17 pontos, o que corresponde a aproximadamente 6,0% do total de pontos verificados, como apresentado na Figura 7a. Observando os resultados, percebe-se que as classes que apresentaram os maiores erros de não concordância foram as de solo exposto e vegetação densa, com cerca de 41,2% para ambas, assim sendo que deveriam ser assinaladas à esta classe, porém sendo assinaladas a outras.

Para o anos de 2016, na figura 7b, a análise direta dos pares acima indica no total de 267 pontos verificados, a concordância em 252 pontos, sendo equivalente a 94,4% do total de pontos verificados, e a não concordância em 15 pontos, o que corresponde a aproximadamente 5,6% do total de pontos verificados. Dos resultados, percebe-se que a classe que apresentou o maior erro de não concordância foi a de solo exposto com cerca de 53,33%, seguida da classe de vegetação densa com 33,33%, e por fim a classe de vegetação esparsa com 13,33% do total de amostras deveriam ser assinaladas a uma classe, porém sendo assinaladas a outras.

Para o resultado da classificação para o ano de 2021, Figura 7c, destaca-se que este foi o que apresentou os melhores resultados de concordâncias dos pares de classes observadas e classificadas pelo algoritmo. Esse resultado pode ser explicado devido ao fato desta imagem ter sido utilizada como base para coleta de amostras e treinamento do algoritmo Randon Forest. Neste sentido, para o ano de 2021 a análise indica que no total de 264 pontos verificados foi obtida uma concordância em 256 pontos, sendo equivalente a 96,97% do total de pontos verificados, e a não concordância em 8 pontos, o que corresponde a aproximadamente 3,0% do total de pontos verificados. Dos resultados, percebe-se que a classe que apresentou o maior erro de não concordância foi a de vegetação densa com cerca de 62,5%, seguida da classe de solo



exposto com 37,5%, e por fim a classe de vegetação esparsa, a qual não apresentou valores discordantes para a classificação.

Após a etapa de determinação da matriz de confusão para cada ano, realizou-se a determinação do índice Kappa para a avaliação do nível de concordância entre os conjuntos de dados observados e comparados aos classificados. Determinando-se o índice global Kappa para verificar a consistência global, assim sendo encontrado para o ano de 2021 um valor de 0,96, enquanto para 2016 foi encontrado o valor de 0,92, e 0,91 para o ano 2014. Segundo Landis & Koch (1977), os quais caracterizaram diferentes faixas para os valores Kappa, segundo o grau de concordância, assim sugerindo que valores no intervalo compreendido entre 0,81 e 1,0 são bastante representativos e sendo assim possuem um nível de concordância considerada como perfeita.

5. Conclusões

Observa-se que o algoritmo Random Forest apresentou um bom desempenho de concordância de classes, assim reforçando seu potencial para classificação do uso e cobertura do solo.

O Google Earth Engine apresentou o grande poder de processamento da plataforma de computação em nuvem, bem como, através do algoritmo aprendizado de máquina Random Forest foi possível grande volume de imagens Landsat 8, amostras de treinamento.

6. Agradecimentos

Os autores agradecem à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) pelas bolsas de mestrado do primeiro autor.

7. Referências bibliográficas

ALEMOHAMMED, SH; MCCOLL, KA; KONINGS, AG; ENTEKHABI, D.; STOFFELEN, A.. **Characterization of precipitation product errors across the United State using multiplicative triple collection**. Hydrol Earth Syst Sci 2015, v19:3489–3503.

ANIL B. GAVADE & VIJAY S. RAJPUROHIT. **Systematic analysis of satellite image-based land cover classification techniques: literature review and challenges**, International Journal of Computers and Applications, 2021. 43:6, 514-523. DOI: <https://doi.org/10.1080/1206212X.2019.1573946>.

BIGIARINI M, NAUDITT A, BIRKEL C, VERBIST K, RIBBE L. **Temporal and spatial evaluation of satellite based rainfall estimates across the complex topographical and climatic gradients of Chile**. Hydrol Earth Syst Sci. 2017. v21:1295–1320.

IV SUSTENTARE & VII WIPIS
WORKSHOP INTERNACIONAL
Sustentabilidade, Indicadores e Gestão de Recursos Hídricos
 de 16 a 18 de novembro de 2022

EVENTO GRATUITO TOTALMENTE ONLINE

Realização: SUSTENTARE FUD CAMPINAS WIPES IBC-UFOP

Apoio: Agência das Bacias PCJ COMITÊS PCJ

DAVID L. MILLER, MICHAEL ALONZO, SUSAN K. MEERDINK, MICHAEL A. ALLEN, CHRISTINA L. TAGUE, DAR A. ROBERTS, JOSEPH P. MCFADDEN, **Seasonal and interannual drought responses of vegetation in a California urbanized area measured using complementary remote sensing indices**, ISPRS Journal of Photogrammetry and Remote Sensing, Volume 183, 2022, Pages 178-195, ISSN 0924-2716, <https://doi.org/10.1016/j.isprsjprs.2021.11.00>.

LANDIS, J. RICHARD. GARY G. KOCH. **The Measurement of Observer Agreement for Categorical Data Biometrics**, vol. 33, no. 1, [Wiley, International Biometric Society], 1977, pp. 159–74, <https://doi.org/10.2307/2529310>.

NETO, C. Di G. **Potencial de técnicas de mineração de dados para o mapeamento de áreas cafeiras**. INPE, São José dos Campos, 2014.

ZHAO, X .; HONG, D .; GAO, L .; ZHANG, B .; CHANUSSOT, J. **Transferable Deep Learning from Landsat Data Series for National Land-Cover Mapping with Noisy Labels: A Case Study of China**. Remote Sens. 2021 , 13 , 4194. <https://doi.org/10.3390/rs13214194>.