# MAPIA30M: METODOLOGIA DE CLASSIFICAÇÃO DE ÁREA QUEIMADA USANDO INTELIGÊNCIA ARITIFICAL UTILIZANDO O CONJUNTO DE DADOS DO PROGRAMA QUEIMADAS

*Paulo W. P. da Cunha*[1], *Júlia Sousa Gayotto*[1], *Fabricio Galende Marques de Carvalho*[1], *Fabiano Morelli*[1] *e Jonatas Leon*[1]

[1]Instituto Nacional de Pesquisas Espaciais - INPE, Av. dos Astronautas, 1.758, São Josédos Campos/SP - Brasil,
paulo.cunha@inpe.br, juliagayotto@gmail.com, fabricio.galende@inpe.br, fabiano.morelli@inpe.br, jonatas.leon@inpe.br

## RESUMO

A classificação de áreas queimadas no território brasileiro, conduzida pelo Programa Queimadas, atualmente depende de processos manuais de auditoria. Com o objetivo de automatizar esse processo, foi desenvolvido um algoritmo de classificação baseado em Random Forest, um método de inteligência artificial, para analisar dados específicos dessas áreas. Este artigo descreve a metodologia aplicada e apresenta os resultados obtidos, demonstrando o potencial desse algoritmo para a automação da classificação de áreas queimadas.

***Palavras-chave*** **–** Classificação, Área Queimada, Inteligência Artificial, Automatização, Random Forest, Sensoriamento Remoto.

## *ABSTRACT*

*The Burned area classification across Brazilian territory, carried out by the Program Queimadas, currently relies on manual auditing processes. To automate this procedure, a classification algorithm based on Random Forest, an artificial intelligence method, was developed to analyze specific data from these burned areas. This article describes the applied methodology and presents the results, demonstrating the potential of this algorithm for automating the classification of burned areas.*

***Key words*** **–** *Classification, Burned Areas, Artificial Intelligence, Automation, Random Forest, Remote Sensing.*

## 1. INTRODUÇÃO

O Programa Queimadas do INPE [1] monitora as áreas queimadas no Bioma Cerrado, mas atualmente depende de auditorias manuais, um processo demorado e sujeito a erros. Para solucionar este problema, pesquisadores desenvolveram um algoritmo de classificação baseado em inteligência artificial, especificamente o método Random Forest. Este algoritmo utiliza dados de satélites, como Landsat 8 e 9, para identificar áreas queimadas com base em alterações em índices espectrais, como o NDVI e o NBR.

O objetivo principal é automatizar a classificação de áreas queimadas, tornando o processo mais eficiente e preciso. O algoritmo foi treinado com dados da órbita/ponto 220/065 do satélite Landsat no bioma Cerrado. Esses dados foram escolhidos porque já haviam sido auditados, permitindo a validação do algoritmo. O modelo foi então testado em três conjuntos de dados adicionais da mesma órbita/ponto em datas diferentes.

Os resultados mostraram que o algoritmo Random Forest é altamente preciso na classificação de áreas queimadas, com métricas como acurácia, precisão, revocação e F1-Score próximas a 100 por cento. Isso significa que o algoritmo pode identificar corretamente a maioria das áreas queimadas e reduzir significativamente a necessidade de auditorias manuais.

## 2. MATERIAL E MÉTODOS

O mapeamento de áreas queimadas foi conduzido pelo INPE por meio do subprojeto MapIA30, desenvolvido para identificar áreas queimadas utilizando imagens de resolução média de 30 metros. Com a integração dos dados dos satélites Landsat 8 e 9, o projeto permitiu reduzir o intervalo temporal para o mapeamento de áreas queimadas para 8 dias.

O Programa Queimadas identifica áreas queimadas analisando alterações em índices espectrais ao longo do tempo, com ênfase no NDVI e NBR, fundamentais para detectar mudanças na vegetação e localizar cicatrizes de queimadas. Inicialmente, são mapeadas as áreas com alterações que podem ter diversas causas, incluindo incêndios. [2] Em seguida, é realizada uma etapa de classificação para identificar especificamente as mudanças atribuídas a queimadas, tarefa que ainda requerem auditorias manuais atualmente. Esses índices são calculados a partir das imagens captadas pelos satélites, e a frequência de revisita de cada órbita define o intervalo entre as capturas da mesma área.

Visando automatizar essa análise e reduzir a necessidade de auditorias manuais, foi desenvolvido um algoritmo de classificação de áreas queimadas utilizando aprendizado de máquina, especificamente o algoritmo Random Forest, que emprega uma série de árvores de decisão para identificar padrões associados a áreas queimadas. [3] O principal objetivo é automatizar a classificação de áreas queimadas, substituindo o processo manual de avaliação por um algoritmo de inteligência artificial mais eficiente e preciso.

Os dados utilizados para o treinamento e teste do algoritmo de classificação correspondem à órbita-ponto 220/065 do satélite Landsat, registrados em 22/09/2018, no Bioma Cerrado. Cada objeto é representado por polígonos com atributos associados, obtidos por meio de análises estatísticas dos valores dos pixels que o compõem. As imagens originais contêm as informações de reflectância espectral de cada pixel, e os atributos derivados incluem a mediana de cada banda espectral, além de índices espectrais como NDVI [4], NBR [5], BAI [6], GEMI [7], NBR SWIR [8] e MIRBI [9],

calculados a partir das bandas espectrais dos satélites Landsat 8 e 9.

A escolha desse conjunto de dados baseou-se no fato de serem dados previamente auditados, permitindo, assim, a validação dos rótulos atribuídos pelo algoritmo desenvolvido e a avaliação de sua precisão. Dessa base, foram selecionadas 7.193 instâncias para o processo de classificação, sendo 6.504 polígonos classificados como áreas queimadas e 689 como áreas não queimadas. Os índices utilizados foram analisados por meio da técnica de **Feature Importance**, que permitiu selecionar as variáveis mais relevantes para a classificação das áreas queimadas.

O processo iniciou com o pré-processamento dos dados, incluindo a verificação de valores ausentes e a remoção de colunas com baixo desempenho, conforme apontado pela análise de importância das variáveis (**Feature Importance**). Em seguida, os dados foram separados em conjuntos de treinamento e teste, priorizando um balanceamento entre as classes, para otimizar a precisão do modelo.

Após o balanceamento das classes, o modelo foi treinado com os dados e, em seguida, testado com o conjunto de testes. O modelo treinado foi então armazenado em um arquivo no formato pickle, recurso essencial em diversas aplicações de ciência de dados e aprendizado de máquina. Este tipo de arquivo permite a serialização e desserialização eficiente de objetos Python, como modelos treinados, dicionários, listas e arrays. Dessa forma, dados complexos podem ser salvos de maneira compacta e recuperados posteriormente, preservando sua estrutura original.

O modelo desenvolvido foi testado para avaliar sua capacidade de classificar dados não auditados, simulando um cenário real. Para essa avaliação, utilizamos um conjunto de dados composto por três bases correspondentes à órbita/ponto 220/065, capturadas nas datas de 26/07/2023, 11/08/2023 e 04/09/2023.

Foi definida uma função para classificar cada instância com base na probabilidade prevista. Se a probabilidade for maior ou igual a 0,7, a instância é classificada como "queimada". Para probabilidades entre 0,3 e 0,7, a classificação atribuída é "dúvida". Por fim, para valores inferiores a 0,3, a instância é categorizada como "não queimada".

## 3. RESULTADOS

Para avaliar a precisão e a confiabilidade dos modelos de classificação de áreas queimadas, é fundamental utilizar várias métricas de desempenho. As principais métricas baseadas no modelo Random Forest incluem: Acurácia (Accuracy), Precisão (Precision), Revocação (Recall), F1-Score e Matriz de Confusão.

A Figura 1 apresenta a matriz de confusão, que ilustra os resultados obtidos no teste do algoritmo, baseado no modelo treinado. Nesta matriz, são exibidas as quantidades de acertos e erros nas previsões feitas pelo modelo.

Com essas informações, foram calculadas as métricas de Acurácia, Precisão, Revocação e F1-Score, cujos resultados estão detalhados na Tabela 1. Essas métricas nos permitem quantificar a performance do modelo desenvolvido.

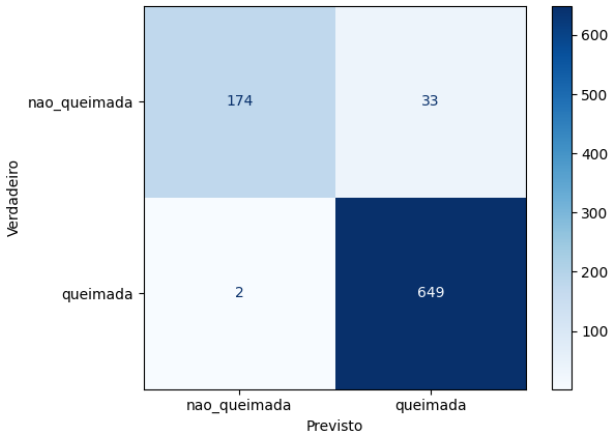



**Figura 1: Matriz de confusão com o dataset 22/09/2018**

| Acurácia | Precisão | Recall | F1 |
|---|---|---|---|
| 0.9592 | 0.9516 | 0.9969 | 0.9737 |

**Tabela 1: Valores das metricas obtidas no treinamento.**

Já na Tabela 2, encontram-se os resultados das classificações das outras três bases analisadas com o modelo desenvolvido.

| Dataset | Queimadas | Não Queimadas | Dúvidas |
|---|---|---|---|
| **26/07/2023** | 2357 | 222 | 71 |
| **11/08/2023** | 2277 | 1412 | 3994 |
| **04/09/2023** | 3342 | 597 | 537 |

**Tabela 2: Classificação das bases de dados não rotuladas.**

Quanto as bases de dados posteriormente analisadas pelo algoritmo desenvolvido, é possível visualizar na Figura 2 que nos datasets, grande parte dos poligonos foram classificados sem nenhuma necessidade de auditória manual.

Apesar da alta quantidade de áreas classificadas como **dúvidas** e **não queimadas** no dataset de 11/08/2023, a análise dos polígonos indica que o algoritmo desenvolvido foi eficaz na identificação e classificação dos muitos ruídos presentes, separando bem as queimadas das demais classes, conforme ilustrado na Figura 3.

## 4. DISCUSSÃO

O uso de algoritmos de inteligência artificial, como o Random Forest, para a classificação automatizada de áreas queimadas no Brasil, mostra-se promissor, ainda que enfrente desafios específicos. Diversos estudos analisam esse potencial, explorando tanto as oportunidades quanto as dificuldades envolvidas.

A automação da classificação, como a proposta, pode reduzir de maneira significativa o tempo e os recursos necessários para identificar áreas queimadas. O uso de algoritmos como o Random Forest já demonstrou uma acurácia de 95,92 por cento na identificação de áreas queimadas em conjuntos de dados de teste, o que sugere uma capacidade robusta de identificar essas regiões com alta precisão.

Outro ponto positivo é a redução de auditorias manuais. Com um sistema automatizado, a necessidade de revisão

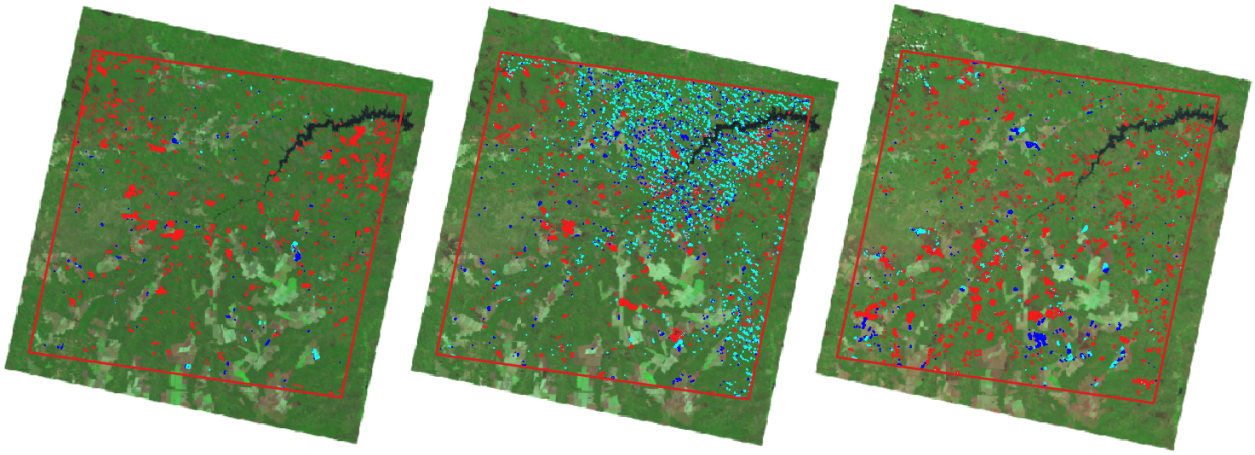


**Figura 2: Resultado da classificação dos 3 datasets: queimadas (vermelho), dúvidas (ciano) e não queimadas (azul)**

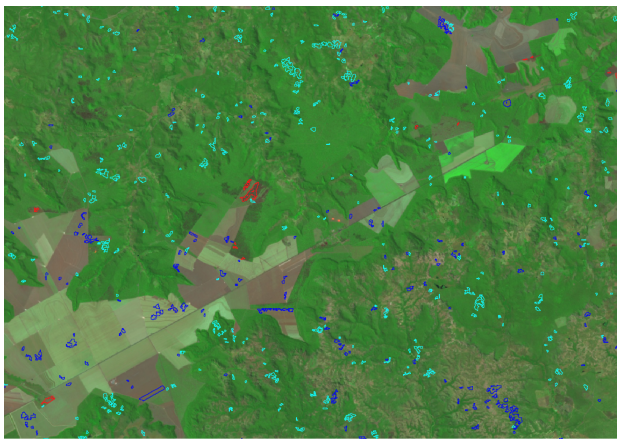


**Figura 3: Detalhes da eficiencia da classificação do algoritmo desenvolvido.**

manual é diminuída, permitindo que os especialistas concentrem-se em áreas de maior complexidade ou incerteza. O algoritmo é capaz de classificar áreas com alto grau de certeza, enquanto apenas as áreas classificadas como “dúvidas ”, requerem auditoria manual, otimizando o tempo dos analistas e minimizando o esforço necessário.

No entanto, a aplicação desses algoritmos enfrenta alguns desafios. A definição da classe “não queimadas ”é um ponto crítico. Dada a diversidade de paisagens e usos do solo no Brasil, áreas modificadas por desmatamento, agricultura ou outras práticas podem apresentar características espectrais semelhantes às de áreas queimadas, o que frequentemente leva a falsos positivos. Isso destaca a complexidade em estabelecer fronteiras precisas para essa classe.

A ambiguidade dos dados é outro fator desafiador, até mesmo para especialistas, pois a presença de nuvens, sombras e variações na vegetação pode dificultar a distinção entre áreas queimadas e não queimadas, introduzindo incertezas na classificação. Ainda, ocorrem erros de comissão em áreas agrícolas. Atividades como o preparo do solo e a colheita podem gerar padrões espectrais semelhantes aos de áreas queimadas, especialmente em um sistema de classificação binária (queimada/não queimada). Esse problema reforça a necessidade de incluir mais classes e refinar os atributos utilizados pelos algoritmos, de modo a aumentar a precisão e a especificidade da classificação.

Para enfrentar esses desafios e explorar o potencial do uso de algoritmos de classificação automatizada, algumas recomendações podem ser consideradas. A expansão da base de dados com uma diversidade maior de amostras de diferentes tipos de áreas queimadas e não queimadas poderia melhorar a robustez e a precisão do modelo, já que o algoritmo seria exposto a uma gama mais ampla de variáveis e cenários.

Outro aspecto importante seria o refinamento dos limiares de confiança. Ajustar os limites de probabilidade para classificação (por exemplo, o limiar de 70 por cento utilizado no estudo) pode ajudar a aprimorar a separação entre as categorias "queimada", "dúvida"e "não queimada", permitindo um mapeamento mais detalhado e menos ambíguo das áreas de interesse.

Por fim, a incorporação de novas classes e atributos pode melhorar substancialmente o desempenho do algoritmo. A introdução de classes adicionais, como "agricultura"e "desmatamento", e o desenvolvimento de novos atributos para capturar as características espectrais específicas de cada classe podem auxiliar o algoritmo a distinguir com mais precisão áreas queimadas de outras formas de alteração do solo.

## 5. CONCLUSÕES

Visando aprimorar o desempenho do modelo, planejamos expandir a base de dados utilizada no treinamento, construindo uma base de conhecimento robusta para uma classificação automática de áreas queimadas. Esse aprimoramento, juntamente com a avaliação de novos limiares de confiança para a classificação, deverá reduzir ainda mais a quantidade de imagens classificadas como "dúvidas", minimizando a necessidade de auditorias manuais.

Entre os principais desafios está a definição da classe "não queimadas", que inclui áreas alteradas por atividades de mudança na cobertura da terra, como desmatamento e práticas agrícolas. Áreas agrícolas, especialmente durante o preparo do solo ou colheita, podem apresentar características espectrais semelhantes às de áreas queimadas, resultando em

falsos positivos. Esse fato evidenciou a limitação de uma classificação binária, que pode não capturar adequadamente toda a complexidade dos eventos mapeados.

Outro obstáculo a ambiguidade dos dados, que dificulta a classificação de algumas áreas, até mesmo para especialistas. Nessas situações, áreas de classificação incerta são designadas como "não queimadas", o que contribui para a variabilidade dentro dessa classe.

O estudo demonstra o potencial do uso de algoritmos de inteligência artificial, como o Random Forest, para a classificação automatizada de áreas queimadas no Brasil. O modelo desenvolvido se mostrou robusto e preciso, com alta acurácia na identificação de áreas queimadas.A implementação desse modelo pode otimizar o processo de monitoramento de áreas queimadas, liberando recursos para outras tarefas e permitindo uma resposta mais rápida a incêndios.

Pesquisas futuras devem focar na redução de erros de comissão em áreas agrícolas, solo exposto, desmatamento, preparo do solo e colheita. Sugerimos ampliar o número de classes e definir atributos relevantes para melhorar a precisão do algoritmo, aumentando sua capacidade de diferenciar essas categorias e explorar integralmente as oportunidades oferecidas pela inteligência artificial nesse campo.

## 6. REFERÊNCIAS

[1] A. W. Setzer, N. J. Ferreira, and F. Morelli. O programa queimadas do inpe. In A. W. Setzer and N. J. Ferreira, editors, *Queimadas e incêndios florestais: mediante monitoramento orbital*, pages 23–52. São Paulo, 2021.

[2] A. E. Melchiori, A. W. Setzer, F. Morelli, R. Libonati, P.A. Cândido, and S. C. Jesús. A landsat-tm/oli algorithm for burned areas in the brazilian cerrado – preliminary results. pages 1302–1311, 11 2014.

[3] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, October 2001.

[4] J.W. Rouse, R.H. Haas, J.A. Schell, and D.W. Deering. Monitoring vegetation systems in the great plains with erts. In *3rd ERTS Symposium*, pages 309–317, 1973.

[5] C. Key and N. Benson. *Landscape Assessment (LA) Sampling and Analysis Methods*, pages LA 1–51. 01 2005.

[6] E. Chuvieco and M. Martín. Cartografía de grandes incendios forestales en la península ibérica a partir de imágenes noaa-avhrr. *Serie Geográfica*, 7, 01 1998.

[7] B. Pinty and M. Verstraete. Gemi: A non-linear index to monitor global vegetation from satellites. *Vegetatio*, 101:15–20, 07 1992.

[8] S. Liu, Y. Zheng, M. Dalponte, and X. Tong. A novel fire index-based burned area change detection approach using landsat-8 oli data. *European Journal of Remote Sensing*, 53:104–112, 01 2020.

[9] S. Trigg and S. Flasse. An evaluation of diverent bi-spectral spaces for discriminating burned shrub-savannah. *International Journal of Remote Sensing*, 22, 2001.