

AVALIAÇÃO DA ACURÁCIA DA CLASSIFICAÇÃO ARBÓREA A PARTIR DE UMA CENA CBERS 4A UTILIZANDO RANDOM FOREST

GABRIELLI BEATRIZ DOS SANTOS¹
PAULO ROBERTO DA SILVA RUIZ¹

RESUMO

Neste estudo é explorada a utilização do satélite CBERS (*China-Brazil Earth Resources Satellite*) para classificações automáticas da cobertura arbórea urbana com as imagens disponibilizadas pelo CBERS 4A, com resolução espacial de até 2 metros em 4 bandas espectrais (azul, verde, vermelho e infravermelho próximo), no município de Adamantina, localizada no estado de São Paulo, Brasil. O uso de adequado algoritmo de Aprendizado de Máquina o qual foi escolhido o *Random Forest* é baseado em árvores de decisão que combina múltiplas árvores para melhorar a precisão e controlar o problema de *overfitting*, juntamente com o sensoriamento remoto e a inteligência artificial que como ferramentas contribuem nos mapeamentos da cobertura arbórea com rapidez e precisão, oferecendo uma visão abrangente, identificando os padrões de vegetação, as áreas verdes, distribuição e saúde das árvores nos diferentes ambientes urbanos. A metodologia da pesquisa consiste em Pré-Processamento, Segmentação, Amostras Classes, Extração atributos, Classificação Arbórea e a Avaliação da qualidade, com auxílio dos softwares *eCognition Developer 8.7* e o *Weka 3.8.6* (*Waikato Environment for Knowledge Analysis*) comum em Aprendizado de Máquina. Os resultados evidenciam se o algoritmo escolhido é capaz de diferenciar e classificar a imagem corretamente sem nenhuma possibilidade de *overfitting*, tendo como análise o Índice de Kappa, Matriz de confusão e o Mapa de Verdade de Solo. Portanto essa análise irá mostrar se os dados do satélite CBERS 4A é recomendado em mapeamentos urbanos e nos planejamentos.

Palavras-chave: Aprendizado de máquina; Arborização urbana; Classificação de imagens; Cobertura arbórea; *Random forest*; Sensoriamento remoto.

ABSTRACT

In this study, the use of the CBERS (*China-Brazil Earth Resources Satellite*) for automatic classifications of urban tree cover is explored using images provided by CBERS 4A, which offers spatial resolution up to 2 meters in 4 spectral bands (blue, green, red, and near-infrared), in the municipality of Adamantina, located in the state of São Paulo, Brazil. The selected Machine Learning algorithm, *Random Forest*, is based on decision trees that combine multiple trees to improve accuracy and control the problem of *overfitting*. Along with remote sensing and artificial intelligence, these tools contribute to mapping tree cover quickly and accurately, providing a comprehensive view, identifying vegetation patterns, green areas, tree distribution, and tree health in different urban environments. The research methodology consists of Pre-Processing, Segmentation, Sample Classes, Attribute Extraction, Tree Classification, and Quality Evaluation, with the help of the software *eCognition Developer 8.7* and *Weka 3.8.6* (*Waikato Environment for Knowledge Analysis*), which are common in Machine Learning. The results highlight whether the chosen algorithm is capable of correctly differentiating and classifying the image without any possibility of *overfitting*, using the Kappa Index, Confusion Matrix, and Ground Truth Map for analysis. Therefore, this analysis will show whether the CBERS 4A satellite data is recommended for urban.

Keywords: Image classification; Machine learning; Random forest; Remote sensing; Tree cover; Urban forestry.

¹ Faculdade de Tecnologia de Adamantina (Fatec), Adamantina-SP, Brasil

1 INTRODUÇÃO

Nos ambientes urbanos são encontrados diversos desafios no planejamento para atender às demandas crescentes de qualidade de vida e distribuição de serviços, devido ao rápido crescimento de sua população, a qual superou o índice rural globalmente (UNPD, 2018). Nesse sentido, a arborização urbana auxilia na melhoria das condições ambientais da população, equilibrando as áreas construídas com o ambiente natural e fazendo parte da história da cidade (Londe; Mendes, 2014). Áreas urbanas com cobertura arbórea relevante beneficiam-se de sombreamento, redução de temperaturas, proteção de bacias hidrográficas, prevenção de erosão e outros aspectos positivos (Buckeridge, 2015; Scott et al., 1999; Fischer e Steed, 2008; Pradella, 2015; Costa Filho, 2009).

O monitoramento da vegetação urbana por Sensoriamento Remoto (SR) é utilizado para o mapeamento e atualização cartográfica das áreas urbanas, disponibilizando imagens de alta resolução espacial e espectral proporcionadas por satélites modernos (Myint et al., 2011). No entanto, o processamento de grandes volumes de dados requer recursos e técnicas de extração de informações para tomada de decisão (Witten et al., 2011). Nesse contexto, a classificação supervisionada é uma técnica de análise de imagens de alta resolução, envolvendo etapas como segmentação, obtenção e extração de amostras, e treinamento de algoritmos (Ruiz, 2017).

A utilização do SR para o mapeamento arbóreo fornece uma base sólida onde para o poder público propor melhorias na cobertura arbórea, contribuindo com a gestão urbana sustentável ao proporcionar o desenvolvimento de estratégias eficazes de planejamento e conservação ambiental. Para a realização de um mapeamento eficiente, o qual apresente alta acurácia, é necessário adotar imagens com resolução espacial centimétrica e bandas espectrais além do visível. Autores consagrados na literatura de SR, como Jensen (2011), Myint et al. (2011) e Blaschke et al. (2014), evidenciam que a alta resolução espacial é necessária para separar os elementos urbanos, que em sua maioria são pequenos, já a resolução espectral diferencia as classes urbanas, as quais possuem alta semelhança espectral, como por exemplo classificar vegetação arbórea e rasteira.

Satélites que possuam sensores com altas resoluções espaciais e espectrais em sua maioria são comerciais. Destacam-se os satélites norte-americanos

Geoeye, WorldView-2, 3 e 4 e chineses como o Jilin-1, os quais possuem custos elevados e negociados em dólares. A utilização de dados espaciais para realizar mapeamentos urbanos podem ser inviáveis para municípios pequenos e médios que sofrem com limitação orçamentária. Assim, este trabalho objetiva contribuir com esses desafios avaliando a viabilidade da utilização de imagens orbitais gratuitas do satélite sino-brasileiro CBERS (*China-Brazil Earth Resources Satellite*) em classificações automáticas da cobertura arbórea urbana.

As imagens CBERS 4A possuem resolução espacial de até 2 metros em 4 bandas espectrais (azul, verde, vermelho e infravermelho próximo) e distribuídas de maneira gratuita pelo Instituto Nacional de Pesquisas Espaciais (INPE). Aliado a essas características é necessário adotar um adequado algoritmo de Aprendizado de Máquina (AM) para prover classificações automáticas do solo urbano. Dessa forma, este trabalho justifica-se por possibilitar avaliar tanto o dado espacial como os algoritmos de classificação, contribuindo com a extração de informação de um tipo específico de dados que pode ser aplicado em diversas áreas dos setores público e privado.

Por fim a hipótese deste trabalho é a seguinte: É possível realizar classificações automáticas da cobertura arbórea a partir dos dados CBERS 4A com acurácia capaz de auxiliar atividades de planejamento urbano?

1.1. Objetivos

O objetivo geral deste trabalho é analisar a acurácia da classificação supervisionada da cobertura arbórea urbana a partir de imagens do satélite CBERS 4A, identificando suas potencialidades e limitações.

Para isso, os objetivos específicos contemplam:

- Realizar o pré-processamento da imagem.
- Definir as classes de cobertura do solo.
- Recolher amostras de treinamento, definir os atributos e realizar o treinamento do modelo de classificação.
- Verificar a acurácia da classificação por meio do mapa de verdade de solo.
- Confeccionar os mapas de cobertura arbórea.

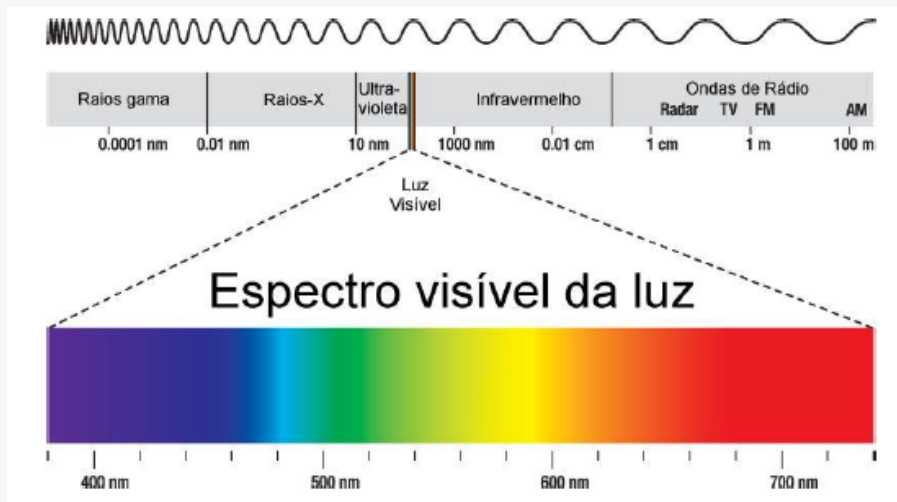
2 FUNDAMENTAÇÃO TEÓRICA

Planos de arborização urbana, considerada florestas urbanas, são importantes componentes bióticos para a cidade em termos climáticas, melhorando a qualidade do ar e o microclima urbano. A arborização urbana pode ser dividida em bosques, praças, parques, calçadas de vias públicas, sendo uma vegetação próxima da população. Por isso sofre com a falta de planejamento, cuidado e fiscalização dos órgãos públicos. Além disso, árvores em locais inadequados acabam afetando a população e despertam ações inadequadas como por exemplo danos em encanamento, calhas, quedas de folhas/frutos e galhos em locais onde circulam pedestres, automóveis, aumentando risco de acidentes. Por isso, a arborização urbana requer planejamento tanto em sua localização como conservação.

O SR e a Inteligência Artificial são ferramentas úteis para auxiliar no desenvolvimento dos planos de arborização, pois contribuem com mapeamentos da cobertura arbórea com maior rapidez e precisão, oferecendo uma visão abrangente da cobertura arbórea, a qual permite a identificação de padrões de vegetação e a quantificação da área ocupada por árvores (Londe e Mendes et al., 2014). O SR multiespectral é uma ferramenta útil para o estudo e monitoramento da cobertura vegetal urbana. Ele permite a obtenção de informações detalhadas sobre a distribuição espacial das áreas verdes e a saúde das árvores em diferentes ambientes urbanos (Brasil, 2002).

O SR multiespectral consiste no imageamento em diversas bandas espectrais. Conforme a Figura 1, a radiação eletromagnética solar atinge a superfície terrestre em diferentes comprimentos de onda, os quais determinam as bandas espectrais. Os sensores acoplados aos satélites são capazes de capturar a radiação refletida pelos objetos terrestres em uma faixa espectral, que pode compreender várias bandas espectrais. Satélites com sensores multiespectrais podem gerar imagens composta por várias bandas, por exemplo do azul até o infravermelho distante. Quanto maior o número de bandas, melhor a resolução espectral do satélite e, por consequência, melhor para os algoritmos de AM diferenciarem os objetos em uma classificação automática (Ruiz, 2017). Conforme Scott et al. (1999), a região espectral do infravermelho são sensíveis às características biofísicas das plantas, como vigor e condição foliar, pois é nessa faixa espectral que a vegetação reflete a maior parte da radiação incidente.

Figura 1 – Espectro Eletromagnético da radiação



Fonte: Matavali (2021)

Por sua vez, a resolução espacial de um sensor consiste em sua capacidade de distinguir objetos, ou seja, quanto menor for o objeto, menor deverá ser o pixel da imagem para distingui-lo (Blaschke et al., 2014). Sensores com resoluções métricas são utilizados para detecção de mudanças, mapeamento de desmatamento e agrícola, ou seja, que abrangem grandes áreas. Já em aplicações urbanas, devido as áreas e os objetos serem menores são necessários sensores com resolução espacial centimétrica (Ruiz, 2017).

O programa CBERS é uma parceria entre Brasil e China, no setor técnico-científico espacial. No Brasil praticamente todas as instituições do meio ambiente utilizam imagens desse programa, pois é importante no desmatamento e queimadas na Amazônia Legal e cerrado, monitoramento hídrico, crescimento urbano e ocupação do solo (INPE, 2014). O satélite utilizado neste trabalho é o CBERS-4A, o qual possui diversas câmeras, uma delas é a Multiespectral e Pancromática de Ampla Varredura (WPM). Ela é capaz de obter imagens nas bandas espectrais pancromática (450 – 900 nanômetros - nm) com 2 metros de resolução espacial e nas multiespectrais com 8 metros de resolução espacial: azul (450 – 520 nm), verde (520 – 590 nm), vermelho (630 – 690 nm) e infravermelho próximo (770 – 890 nm).

A classificação de imagens é o processo de extrair informação das áreas em questão. Para mapear as áreas de interesse da superfície terrestre é necessário utilizar as imagens para classificar os objetos, criando as classes de interesse. Dessa forma, é possível extrair seus padrões, que serão utilizados como parâmetros de treinamento pelos algoritmos de AM. Para isso, são necessárias a

extração de amostras de treinamento, podendo possuir rótulos (classes) para a classificação supervisionada ou sem rótulos, no caso, para a classificação não supervisionada (Ruiz, 2017).

3 METODOLOGIA

A área de estudo localiza-se na cidade de Adamantina – SP (Figura 2), com a coordenada central de 21°41'06"S, 51°04'22"W. Essa área foi escolhida por possuir uma grande diversidade de área verde, com a presença do Parque Caldeira, compreendendo uma parte da região noroeste de Adamantina, dentre outros bairros, destacam-se a Vila Joaquina, Vila Industrial, Parque Eldorado e Jd. Portugal. O trabalho foi realizado a partir de uma imagem orbital do satélite CBERS-4A, obtida em setembro de 2023, com resolução espacial de até 2 metros em 4 bandas espectrais (Figura 4).

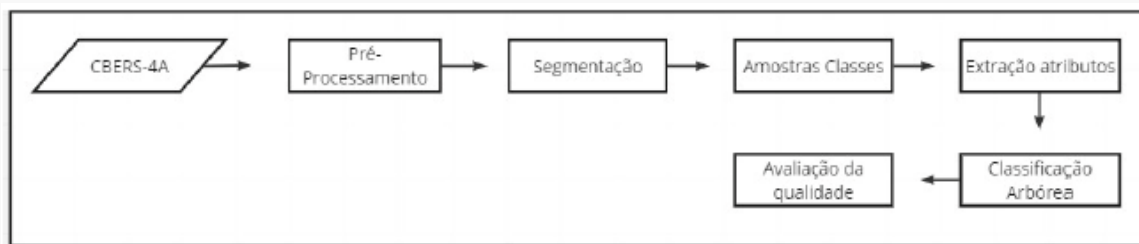
A Figura 3 apresenta a sequência metodológica adotada neste trabalho, descrito na sequência.

Figura 2 – Localização da área de estudo



Fonte: Autoria Própria (2024)

Figura 3 – Passos metodológicos adotados neste trabalho



Fonte: Autoria Própria (2024)

O pré-processamento consiste na conversão dos dados de nível de cinza para radiância, consistindo em uma transformação física para obter o valor da incidência da radiação solar refletida pelos objetos para o espaço, os quais foram captados pelos sensores do satélite. A seguir é aplicada a correção atmosférica, que considera o quanto a contaminação da atmosfera, com gases e partículas, afetou os dados captados pelos sensores. Dessa forma, o dado é transformado para a grandeza física de reflectância de superfície. Em seguida é realizada a fusão de bandas, a qual combina a melhor resolução espacial da banda pancromática com as bandas multiespectrais para sintetizar uma nova imagem multiespectral com melhor resolução espacial (Jensen, 2011).

Foi adotada a classificação supervisionada por regiões, para isso, após o pré-processamento a imagem foi segmentada por meio do Algoritmo Multirresolução (Multiresolution Segmentation), disponível no software eCognition Developer 8.7. A imagem foi ajustada para falsa cor, onde os canais RGB receberam respectivamente as bandas infravermelha próxima, vermelha e verde, desta forma a vegetação se destaca em cor vermelha, facilitando sua diferenciação com os demais elementos da imagem (Figura 4).

Figura 4 – Imagem em falsa cor R(B4) G(B3) B(B2)



Fonte: Autoria Própria (2024)

A seguir, foram selecionados 714 segmentos característicos das classes, a serem utilizadas como amostras de treinamento, distribuídas entre: Árvore (147), Grama (295) e Outro (272). Optou-se em adotar essas classes devido à grande semelhança espectral entre Árvore e Grama, sendo necessário separá-las e entender o grau de confusão entre elas. Além de agrupar os demais alvos na classe denominada como Outro, a qual engloba diversos materiais como asfalto, telhados, solo exposto, calçadas, galpões, dentre outros.

De cada segmento foram extraídos 15 atributos, sendo a média e desvio padrão da radiância em cada banda espectral, além dos Atributos Customizáveis (ACs), do brilho e dos índices de vegetação *Normalized Difference Vegetation Index* (NDVI, em português: Índice de Vegetação por Diferença Normalizada) e *Soil-Adjusted Vegetation Index* (SAVI, em português: Índice de Vegetação Ajustado ao Solo). Os atributos foram obtidos através da aritmética de bandas apresentadas na Tabela 1, “Bn” referem-se às bandas espectrais da imagem.

Após o recolhimento das amostras, procedeu-se à etapa de treinamento a partir do algoritmo *Random Forest* para a construção do modelo de classificação dos dados. Essa etapa ocorreu no ambiente *Weka 3.8.6* (*Waikato Environment for Knowledge Analysis*), um software conhecido de Aprendizado de Máquina (*Machine Learning*) escrito em *Java*, desenvolvido na Universidade Waikato na Nova Zelândia. Contém uma coleção de ferramentas de visualização e diversos algoritmos para solucionar problemas que demandam mineração e predição de dados (Kulkarni; Kulkarni, 2016).

Tabela 1 – Atributos utilizados nas classificações

Atributo	Equações	Referência
Brilho	$Brilho = \frac{\sum_{i=1}^n banda\ i}{n}$	Definiens (2007)
AC 1	$AC1 = \frac{Brilho + B2}{B3}$	Leonardi (2010)
AC 2	$AC2 = \frac{Brilho + B2}{B4}$	Ruiz (2017)
AC 3	$AC3 = Brilho + B2 - B3$	Ruiz (2017)
AC 4	$AC4 = Brilho + B2 - B4$	Ruiz (2017)
NDVI	$NDVI = \frac{B4 - B3}{B4 + B3}$	Rouse et al. (1974)
SAVI	$SAVI = \frac{1.5 * (B4 - B3)}{B4 + B3 + 1.5}$	Huete (1988)

Fonte: Autoria Própria (2024)

O algoritmo *Random Forest* é um método de aprendizado de máquina baseado em árvores de decisão que combina múltiplas árvores para melhorar a precisão da previsão e controlar o problema do *overfitting*. Seu princípio fundamental baseia-se no uso de um conjunto de árvores de decisão, onde cada árvore é construída a partir de um subconjunto aleatório dos dados de treinamento e um subconjunto aleatório dos atributos. Essa abordagem, conhecida como "*bootstrap aggregating*" ou "*bagging*", ajuda a reduzir a variância do modelo. Ao treinar várias árvores de decisão independentes e combinar suas previsões, o *Random Forest* obtém uma média das previsões individuais, o que tende a ser mais robusto e preciso do que uma única árvore de decisão (Breiman, 2001).

Outro princípio adotado no *Random Forest* é a aleatoriedade incorporada na seleção dos atributos em cada nó das árvores de decisão. Durante a construção de cada árvore, em vez de considerar todas as características para determinar o melhor ponto de divisão, o algoritmo seleciona aleatoriamente um subconjunto dessas características. Essa aleatoriedade adicional ajuda a diversificar as árvores e diminuir a correlação entre elas, o que é essencial para melhorar a precisão do modelo final. Ao fazer isso, o algoritmo garante que as árvores sejam suficientemente diferentes entre si, mitigando o risco de *overfitting* aos dados de treinamento (Breiman, 2001).

A combinação das previsões das múltiplas árvores de decisão é realizada através de um processo de votação para problemas de classificação e média para problemas de regressão. No caso de classificação, cada árvore emite um voto para a classe predita, e a classe com o maior número de votos é a predição final do modelo. Para problemas de regressão, as previsões de todas as árvores são simplesmente médias. Esse processo de agregação contribui para a estabilidade e a precisão do *Random Forest*, tornando-o um algoritmo poderoso e popular para diversas aplicações de aprendizado de máquina.

A avaliação da qualidade das classificações foi realizada por meio de comparações com o mapa de verdade de solo desenvolvido. Para isso, foram gerados pontos aleatórios na área da imagem. A partir deles, foi realizado o cruzamento espacial para obter a classe presente no mapa de verdade de campo e na classificação, utilizando de 2500 a 3000 pontos aleatórios. O resultado é uma tabela de classes associando cada ponto à classe presente no mapa verdade e na classificação. De posse destas tabelas, foi gerada a matriz de confusão da classificação e, a partir dela, as métricas de acurácia.

Por meio da matriz de confusão, é possível obter o cálculo da exatidão global, índice Kappa, índice Kappa condicional, bem como os erros de omissão e de comissão de cada classe. Os erros de omissão referem-se à relação entre

o número de amostras classificadas corretamente da classe k e o número total de amostras de referência da classe k . Já os erros de comissão, correspondem à relação entre o número de amostras classificadas corretamente da classe k e o número total de amostras classificadas da classe k . Por meio do índice Kappa, Landis e Koch (1977), para categorização por meio da tabela de referência (Tabela 2). Os Autores informam, que suas divisões são arbitrárias e ambíguas, entretanto de muita utilidade como uma padronização para a discussão dos resultados alcançados (Ruiz, 2017).

Tabela 2 – Categorização do Índice Kappa

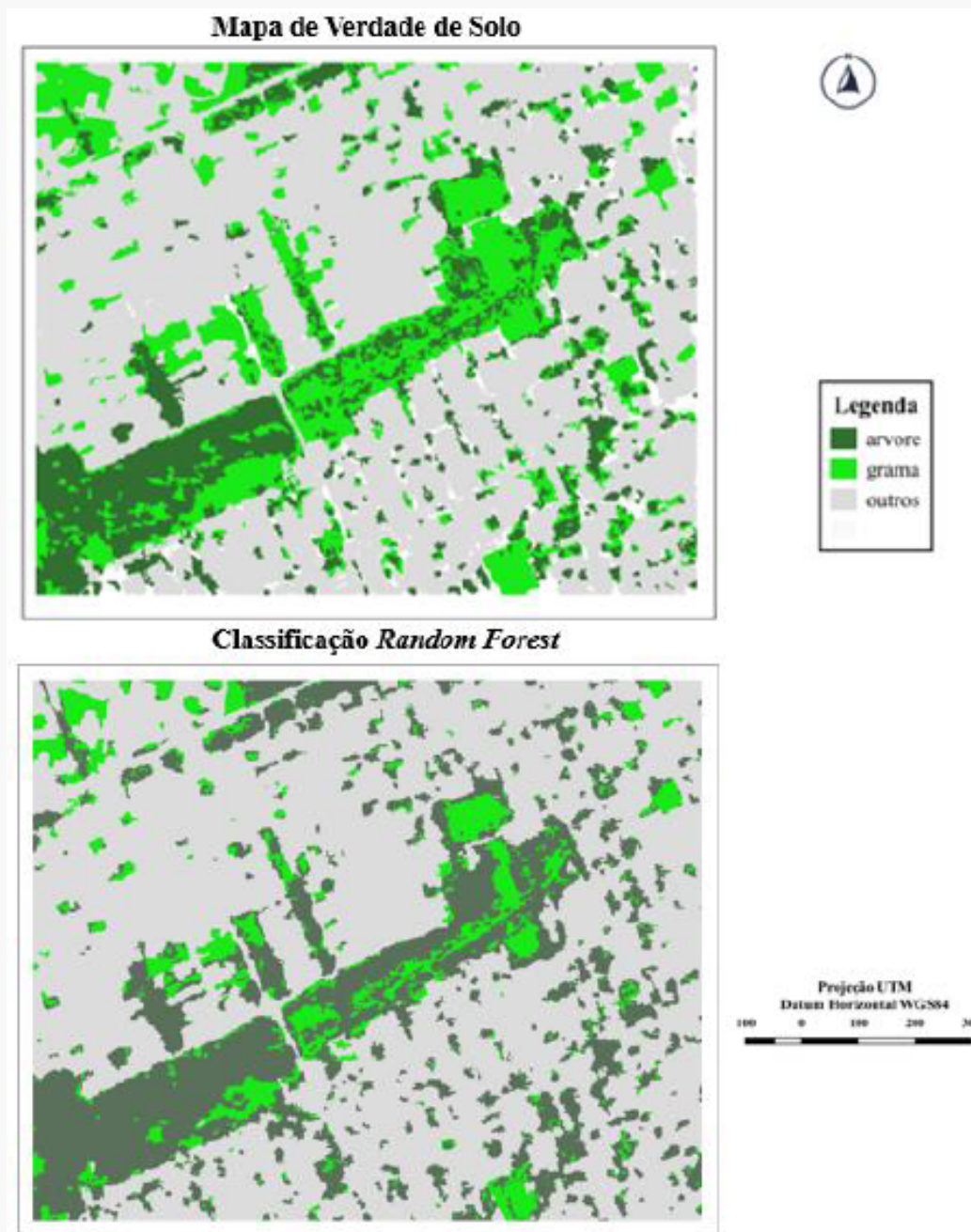
Valores	Categorização
<0	Péssima
0 – 0,2	Ruim
0,21 – 0,4	Razoável
0,41 – 0,6	Bom
0,61 – 0,8	Muito Bom
0,81 – 1,0	Excelente

Fonte: Landis e Koch (1977)

4 RESULTADOS E DISCUSSÕES

A Figura 5 apresenta o mapa de verdade de solo e o resultado da classificação utilizando o algoritmo *Random Forest*. O mapa de verdade foi elaborado empiricamente com o auxílio da imagem de satélite. É possível verificar que há muitos erros nas classes Árvore e Grama, ficando visível as confusões nessas classes, sobretudo na área de parques e bosques.

Figura 5 – Mapa de verdade de solo e Classificação Random Forest



Fonte: Autoria Própria (2024)

Para realizar a validação dos resultados foram utilizados 2771 pontos distribuídos aleatoriamente nas imagens a fim de verificar a correspondência entre as classes. A classificação apresentou índice de exatidão global de 83,22%, com índice Kappa de 0,68. Verifica-se na Matriz de Confusão (Figura 6) muitas confusões entre as classes Árvore e Grama.

Figura 6 – Matriz de confusão da classificação *Random Forest*

		Referencia		
		Árvore	Gramma	Outro
Classificação	Árvore	382	275	65
	Gramma	5	229	15
	Outro	32	73	1695

Fonte: Autoria Própria (2024)

A partir da análise da avaliação por classe (Tabela 3) é possível verificar que as classes *Árvore* e *Outro* apresentam as maiores exatidões do produtor, ou seja, o modelo classificou corretamente as amostras conforme a referência. Em contrapartida, a classe *Árvore* apresenta uma baixa exatidão do usuário, verificando-se que a classificação não é confiável para esta classe, manifestando-se no índice Kappa do usuário. Por outro lado, quando se analisa os índices da classe *Gramma*, observa-se o inverso, ou seja, baixa exatidão do produtor e alta exatidão do usuário.

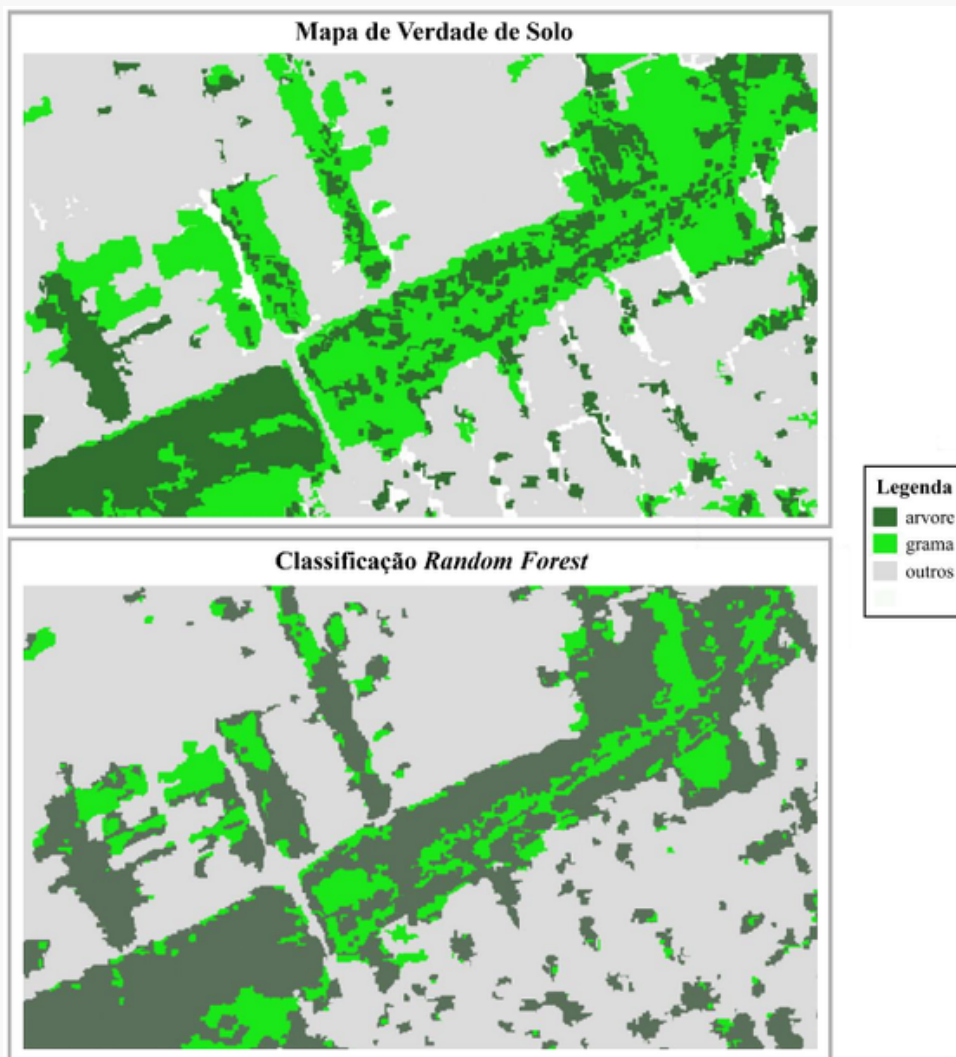
Tabela 3 – Avaliação da qualidade por classe

Avaliação	Árvore	Gramma	Outro
Exatidão do produtor da classe	0,91	0,40	0,95
Erro de omissão da classe	0,09	0,60	0,05
Exatidão do usuário da classe	0,53	0,92	0,94
Erro de comissão da classe	0,47	0,08	0,06
Índice <i>Kappa</i> usuário	0,42	0,89	0,75
Índice <i>Kappa</i> produtor	0,87	0,32	0,79

Fonte: Autoria Própria (2024)

A matriz de confusão ajuda a entender os índices condicionais por classe. Das 577 amostras de referência da classe Grama, 229 foram realmente classificadas como Grama, já 275 foram classificadas erroneamente como Árvore. Isto evidencia que a exatidão do produtor desta classe é muito baixa. Por outro lado, este fato também afeta a classe Árvore, pois ela foi muito classificada como Grama, impactando em sua baixa exatidão do usuário. Esta observação fica evidente quando analisamos a distribuição espacial da classificação no mapeamento (Figura 7). No Parque Caldeira observa-se que a classe Árvore tomou o lugar da classe Grama, ocasionando em extensas áreas classificadas erroneamente, evidenciando-se uma grande confusão entre essas classes.

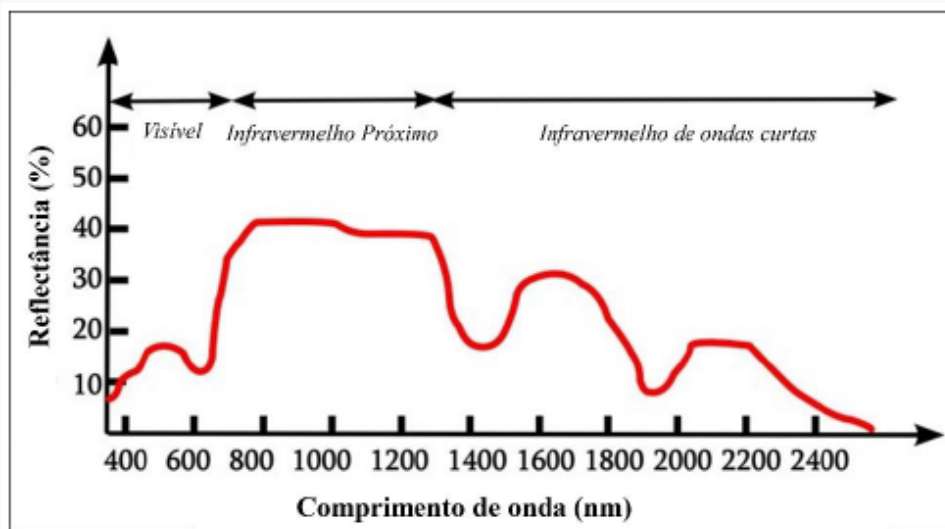
Figura 7 – Detalhe do mapa de verdade e da classificação na região do Parque Caldeira



Fonte: Autoria Própria (2024)

As confusões verificadas, sobretudo entre *Árvore* e *Gramma*, são justificadas pelo comportamento espectral dos dois alvos (Ponzoni; Shimabukuro, 2010). A vegetação apresenta sua maior reflectância na região do infravermelho (Figura 8). Portanto, para ser possível diferenciar estas classes é preciso aumentar as bandas nessa região do espectro eletromagnético, ou seja, utilizar um sensor com melhor resolução espectral. Satélites com mais bandas espectrais são capazes de separar adequadamente diversas classes de vegetação, sendo o caso dos comerciais *WorldView-2* (Anjos, 2016; Ruiz, 2017) e *WorldView-3* (Ruiz, 2017), *Geoeye-2* (Piracicaba, 2020).

Figura 8 – Comportamento espectral da vegetação



Fonte: Adaptado de Ponzoni e Shimabukuro (2010)

Apesar das limitações verificadas na classificação com os dados CBERS 4A, esse sensor é promissor para aplicações de mapeamento urbano. A classificação realizada neste trabalho apresentou índice Kappa de 0,68, categorizada por Landis e Koch (1977) como muito boa. Sobretudo, há possibilidades de melhoria utilizando outros algoritmos de AM e métodos de classificação, no intuito de alcançar acurácias mais elevadas a fim de utilizá-lo em aplicações reais de planejamento urbano.

5 CONCLUSÕES

O presente trabalho realizou uma classificação supervisionada por meio do algoritmo de AM *Random Forest* de uma região da cidade de Adamantina

utilizando dados do satélite CBERS 4A. Foram adotadas três classes de cobertura do solo urbano, sendo elas: Árvore, Grama e Outros a fim de avaliar o potencial dos dados para o mapeamento da cobertura arbórea desta área. Em sua condução, foi adotado o método de classificação supervisionada com a obtenção de amostras das classes utilizadas na etapa de treinamento para a construção do modelo de classificação. Foram extraídos 15 atributos para serem utilizados na construção das árvores de decisão, construída a partir de 714 amostras de treinamento.

O modelo de classificação desenvolvido foi capaz de classificar toda a imagem, obtendo índice de exatidão global de 83,22%, com índice Kappa de 0,68. Foram verificadas confusões entre as classes espectralmente semelhantes, Árvore e Grama, as quais apresentaram quedas significativas nas exatidões do produtor e usuário. A partir da distribuição espacial das classes no mapeamento foi verificado o elevado erro de comissão, ou seja, inclusão de Árvore em áreas relativas à classe Grama, sendo esta afetada pelo erro de omissão.

É importante mencionar que mesmo com as confusões verificadas, os dados do satélite CBERS 4A podem ser adotados em mapeamentos urbanos para auxiliar atividades de planejamento urbano, respondendo a hipótese deste trabalho. Sobretudo, é necessária a realização de refinamentos em amostras e a adoção combinada de métodos de classificação para melhoria da acurácia, possibilitando adotá-lo em mapeamentos.

Recomenda-se para trabalhos futuros:

- o refinamento das amostras e extração de outros índices espectrais para a etapa de treinamento do modelo de árvore de decisão;
- a avaliação de outros algoritmos de classificação, como Redes Neurais Artificiais, com arquiteturas de *Deep Learning* associado à área de Visão Computacional;
- a avaliação e desenvolvimento de novos métodos de classificação de imagens, como a classificação hierárquica em árvore de decisão que particiona a imagem e classifica por partes a imagem total;
- a avaliação da copa das árvores em campo para avaliar os limites da classificação com dados CBERS 4A para subsidiar trabalhos automáticos de contagem de árvores em calçadas, por exemplo.

AGRADECIMENTOS

O autor agradece ao Conselho Nacional de Desenvolvimento Científico e

Tecnológico (CNPQ) pela concessão da Bolsa de Iniciação Científica e Tecnológica, concedida no período de: 09/2023 e 08/2024, sob o processo de nº 164470/2023-9.

REFERÊNCIAS

ANJOS, C. S. **Classificação de áreas urbanas com imagens multiespectrais e hiperespectrais utilizando métodos não-paramétricos**. Tese (Doutorado em Sensoriamento Remoto) — Instituto Nacional de Pesquisas Espaciais, São José dos Campos - SP, 2016.

BLASCHKE, T.; HAY, G. J.; KELLY, M.; LANG, S.; HOFMANN, P.; ADDINK, E.; QUEIROZ FEITOSA, R.; MEER, F.; WERFF, H.; COILLIE, F. V.; TIEDE, D. **Geographic object-based image analysis - towards a new paradigm**. **ISPRS Journal of Photogrammetry and Remote Sensing**: official publication of the International Society for Photogrammetry and Remote Sensing (ISPRS), v. 87, n. 100, p. 180–191, jan. 2014.

BRASIL. Ministério das Cidades. **Diretrizes para a Política Nacional de Desenvolvimento Urbano**. Brasília, 2002.

BREIMAN, L. **Random forests**. *Machine Learning*, v. 45, n. 1, p. 5-32, 2001.

BUCKERIDGE, Marcos. **Árvores urbanas em São Paulo: planejamento, economia e água**. *Estudos Avançados*, v. 29, p. 85-101, 2015.

COSTA FILHO, J. M. **Arborização urbana e ilhas de calor: uma revisão de conceitos e aplicações**. *Ambiente Construído*. Fortaleza. v. 9, n. 2, p. 75-98, 2009.

DEFINIENS. **Definiens Developer 7**: Reference Book. Munique: Definiens AG, 2007. 195 p. Disponível em: <<http://www.ecognition.cc/download/ReferenceBook.pdf>>. Acesso em: 05 mai. 2024.

FISCHER, B. C.; STEED, R. L. **Understanding the relationship between watershed condition and urban tree cover in Portland, Oregon, USA**. Urban Forestry & Urban Greening, Amsterdam, v. 7, n. 1, p. 21-33, 2008.

HUETE, A. R. **A soil-adjusted vegetation index (SAVI)**. Remote Sensing of Environment, v. 25, p. 53-70, 1988.

INPE - Instituto Nacional de Pesquisas Espaciais. (2020). **Satélite Sino-Brasileiro de Recursos Terrestres (CBERS)**. Disponível em: <<http://www.cbears.inpe.br>>. Acesso em 13 de maio de 2024.

JENSEN, J. R. **Sensoriamento remoto do ambiente: uma perspectiva em recursos terrestres**. São José dos Campos: Parêntese, 2011. Tradução José Carlos Neves Epiphany (coordenador).

KULKARNI, E. G.; KULKARNI, R. B. WEKA powerful tool in data mining. **International Journal of Computer Applications** (0975 – 8887). National Seminar on Recent Trends in Data Mining. 2016.

LANDIS, J. R.; KOCH, G. G. **The measurement of observer agreement for categorical Data**. Biometrics, v. 33, n. 1, p.159–174, 1977. DOI 10.2307/2529310. Disponível em: <<http://www.jstor.org/stable/pdf/2529310.pdf?acceptTC=true>>. Acesso em: 05 mai. 2024.

LEONARDI, F. **Abordagens cognitivas e mineração de dados aplicadas a dados ópticos orbitais e de laser para classificação de cobertura do solo urbano**. Dissertação (Mestrado) — Instituto Nacional de Pesquisas Espaciais, São José dos Campos - SP, 2010.

LONDE, V.; MENDES, T. **Arborização urbana: uma análise de sua importância na cidade de Teresina, Piauí**. GeoTextos, Salvador, v. 10, n. 2, p. 350-370, 2014.

MATAVELI, G. **Satélites e sensores: curso de uso escolar de sensoriamento remoto no estudo do meio ambiente**. Instituto Nacional de Pesquisas Espaciais (INPE), 20 jul. 2021. Satélites e sensores, p. 1-89. Disponível em: http://www.dsr.inpe.br/DSR/educacao/uso-escolar-sensoriamento-remoto/apresentacoes-2021-arquivos/20jul21_guilhermemataveli-compactado.pdf. Acesso em: 15 mai. 2024.

MYINT, S. W.; GOBER, P.; BRAZEL, A.; GROSSMAN-CLARKE, S.; WENG, Q. **Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery**. Remote Sensing of Environment, Nova York, v. 115, n. 5, p. 1145-1161, 2011.

PIRACICABA - Secretaria Municipal de Defesa do Meio Ambiente. **Plano Municipal de Arborização Urbana**. Piracicaba, SP: Secretaria Municipal de Defesa do Meio Ambiente. 2020. Disponível em: < https://130d0c4c-ab3e-edc810804d80417aba96.filesusr.com/ugd/9804b1_2fcc86c1000846e8a2d2f5d430f418af.pdf >. Acesso em 07 Jun. 2024.

PONZONI, F. J.; SHIMABUKURO, Y. E. **Sensoriamento remoto no estudo da vegetação**. São José dos Campos: Parêntese, 2010, 127p.

PRADELLA, A. C. P. **O papel das árvores na conservação do solo em áreas urbanas: revisão de literatura**. Ambiente Construído, Porto Alegre, v. 15, n. 2, p. 157-168, 2015.

ROUSE, J. W.; HAAS, R. H.; SCHELL, J. A.; DEERING, D. W.; HARLAN, J. C. **Monitoring the vernal advancement of retrogradation of natural vegetation**. Greenbelt: NASA/GSFC, 1974. 371 p. Disponível em: <<http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19750020419.pdf>>. Acesso em: 10 mai. 2023.

RUIZ, P. R. S. **Classificação da cobertura do solo urbano usando árvores de decisão a partir de cenas WorldView-2 e WorldView-3 para diferentes níveis de legenda**. Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2017, 181 p.

SCOTT, K. I.; SIMPSON, J. R.; MCPHERSON, E. G. Effects of tree cover on parking lot microclimate and vehicle emissions. **Journal of Arboriculture**, Champaign, v. 25, n. 3, p. 129-142, 1999.

UNITED NATIONS POPULATION DIVISION (UNPD). **World urbanization prospects: the 2018 revision**. Nova York, 2018.

WITTEN, I. H.; FRANK, E; HALL, M. A. **Data mining: practical machine learning tools and techniques**. 3. ed. San Francisco: Morgan Kaufmann, 2011. 664p.