

**11ª Jornada Científica e  
Tecnológica do IFSULDEMINAS**  
& **8º Simpósio de  
Pós-Graduação**

## **CLASSIFICAÇÃO DE IMAGENS E HIPERESPECTRAIS POR MEIO DE ÁRVORES DE DECISÃO**

**João V. P. RISSATTI<sup>1</sup>; Camila S. dos ANJOS<sup>2</sup>**

### **RESUMO**

O sensoriamento remoto hiperespectral visa distinguir e identificar os materiais presentes na superfície da Terra por meio dos comportamentos específicos de cada material ao longo do espectro eletromagnético. Tal distinção e identificação é realizada por diferentes algoritmos de classificação de imagens. Entretanto, não existe um classificador perfeito, visto que todo algoritmo apresenta erros de classificação. Com o uso de imagens orbitais e aéreas de alta resolução espacial e espectral, a identificação dos materiais presentes no ambiente urbano é obtida por classificações cada vez mais precisas. Desta maneira, os métodos avançados de classificação de imagens precisam ser avaliados para responder a dúvida de qual método seria o mais preciso. O objetivo deste trabalho é realizar classificação uma imagem hiperespectral, através dos métodos de classificação C4.5 e Random Forest e analisar a qualidade das classificações, identificando qual método gerou a classificação com maior detalhamento do local de interesse. O método Random Forest apresentou uma acurácia global superior ao C4.5 na identificação dos materiais presentes no ambiente urbano.

#### **Palavras-chave:**

Imagens aéreas; Random forest; Classificação C4.5.

### **1. INTRODUÇÃO**

Desde os tempos antigos até os dias de hoje, nota-se uma frequente evolução dos sistemas sensores. Atualmente esses sistemas estão nos proporcionando a capacidade de identificarmos os mais variados alvos presentes na superfície da Terra (ANJOS, 2016).

A melhoria das resoluções espacial e espectral dos sensores trouxe grandes benefícios, os quais têm ajudado muito, tanto na identificação de pequenos e variados alvos quanto na separação desses alvos conforme a análise do comportamento dos diferentes alvos avaliados.

Esses benefícios trouxeram principalmente uma facilidade a mais para que este trabalho fosse realizado, no qual será utilizado um sensor de alta resolução espacial e espectral e, os métodos de classificação C4.5 e Random Forest com o intuito de classificar toda uma área urbana.

### **2. MATERIAL E MÉTODOS**

O objeto de estudo corresponde-se a uma imagem aérea situada na universidade de Houston e seus arredores, localizada no estado do Texas, representada pela Figura 1. Imagem obtida pelo

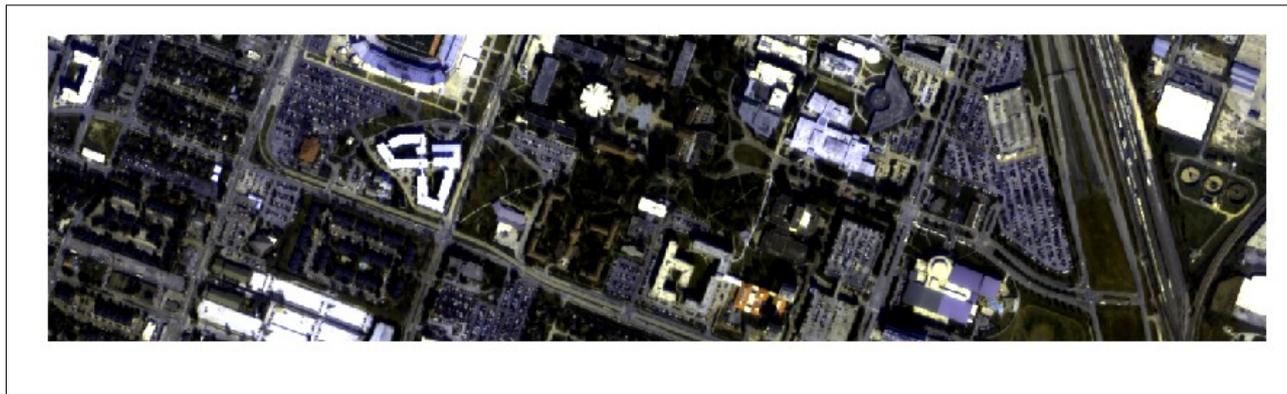
---

<sup>1</sup>Bolsista PIBIC/CNPq, IFSULDEMINAS – *Campus* Inconfidentes. E-mail: vjrissatti@gmail.com.

<sup>2</sup>Orientador, IFSULDEMINAS – *Campus* Inconfidentes. E-mail: camila.lacerda@ifsuldeminas.edu.br.

sensor hiperespectral ITRES CASI 1500, que cobre uma faixa espectral de 380-1050 nm com 48 bandas espectrais, e uma resolução espacial de 1 metro.

Figura 1 – Imagem aérea de Houston e seus arredores.



Fonte: Autor.

Para a melhor distinção dos materiais será realizada a extração de atributos, etapa essencial para extrair informações importantes do conjunto de dados que poderá ser utilizada mais avante, na classificação da imagem.

As transformações foram realizadas utilizando-se o software ENVI 5.0. Após isso foram gerados os índices de vegetação por meio da ferramenta Band Math do mesmo software, sendo estes índices o SAVI (Soil Adjusted Vegetation Index) e NDVI (Normalized Difference Vegetation Index).

Após a obtenção dos atributos de MNF (Minimum Noise Fraction), PCA (Principal Component Analysis) e as já citadas SAVI e NDVI, foi gerado um banco de dados no software Ecognition 64, onde a imagem foi segmentada, de forma que posteriormente foram realizadas a coleta de amostras para cada uma das classes. A seguir essas amostras selecionadas foram exportadas em formato CSV (Comma Separated Values).

Logo em seguida, estes arquivos CSV foram importados pelo software livre de mineração de dados Weka, no qual os arquivos importados foram classificados pelos algoritmos C4.5 (QUINLAN, 1993) e Random Forest (BREIMAN, 2001).

Com as classificações prontas, foi realizada a validação da qualidade das classificações no mesmo software Weka, a fim de avaliar qual dos dois métodos de classificação apresentou melhores resultados de classificação para os conjuntos de dados.

### 3. RESULTADOS E DISCUSSÕES

A seguir apresentam-se os resultados dos dois métodos de classificação, sendo utilizados os algoritmos C4.5 e Random Forest para a mesma imagem hiperespectral. No total utilizaram-se 151

atributos para que ambas as classificações fossem feitas, apresentados Tabela 1.

Tabela 1 – Atributos selecionados pelo software Weka para as classificações das imagens.

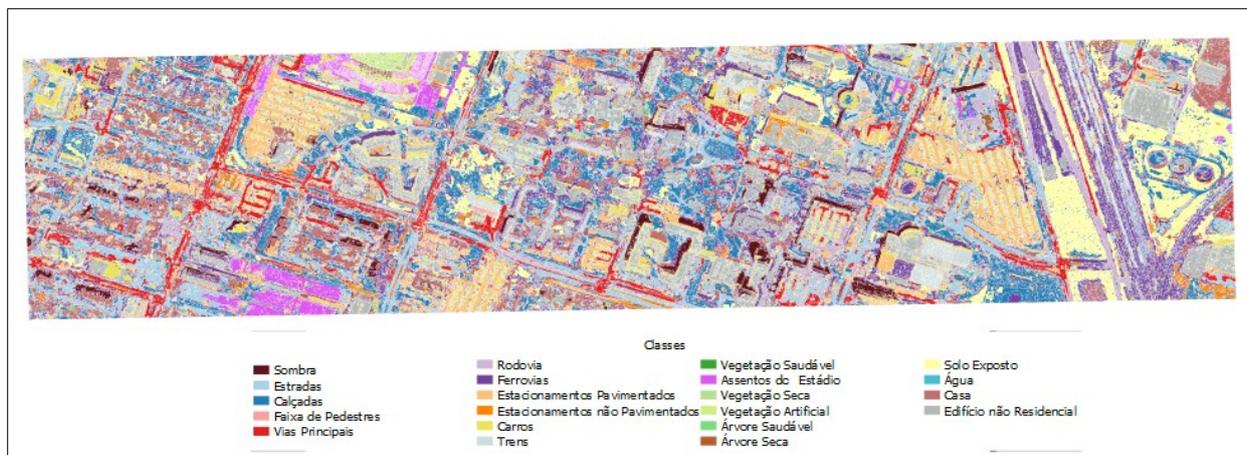
Mean Layer	Class Name	Brightness	Área	Density	Max. Diff	Total Atrib.
146	1	1	1	1	1	151

Fonte: Autor.

Com os atributos apresentados na Tabela 1, percebe-se que o atributo mais utilizado para as classificações em, ambos os algoritmos, foi o Mean Layer, que seria simplesmente a média espectral de cada banda espectral da imagem.

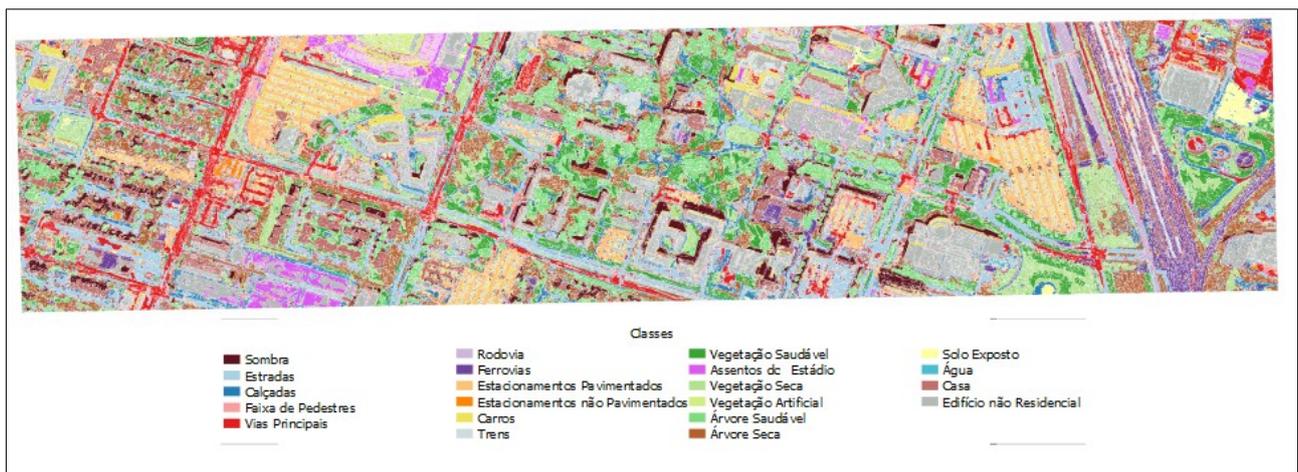
As Figuras 2 e 3 apresentam os resultados das classificações realizadas através dos métodos C4.5 e Random Forest.

Figura 2: Classificação da imagem hiperespectral com o método C4.5.



Fonte: Autor.

Figura 3: Classificação da imagem hiperespectral com o método Random Forest.



Fonte: Autor.

A classificação da imagem através do método C4.5 apresentou uma exatidão global de 73,27% e um índice kappa de 0,71, sendo categorizado segundo Landis e Koch(1977), como de

qualidade sendo muito boa. Enquanto a classificação pelo método Random Forest apresentou uma exatidão global de 84,72 % e um índice kappa de 0,84 sendo categorizado segundo Landis e Koch(1977), como sendo uma classificação de qualidade excelente.

Ambas as classificações obtiveram exatidão global superior a 70%, ou seja, apresentaram resultados satisfatórios. Porém, avaliando a matriz de confusão, concluiu-se que houve grande confusão entre as classes Estradas e Principais Vias, sendo essa confusão muito maior no método C4.5 que no método de classificação Random Forest. Tal confusão era esperada visto que o material das duas classes possuem a mesma composição físico-química.

A classificação que obteve melhores resultados, portanto, foi a classificação realizada pelo algoritmo Random Forest, visto que sua avaliação tanto com relação a exatidão global quanto ao índice kappa foi superior ao método C4.5.

#### 4. CONCLUSÕES

Para se atingir uma compreensão da realidade, apresenta-se dois objetivos: classificar uma imagem hiperespectral através de dois métodos de árvores de decisão, e, após isso, apresentar uma análise sobre os resultados dos dois métodos, verificando qual possibilita discriminar os detalhes dos alvos de interesse do ambiente urbano de forma mais análogo a realidade.

De acordo com os resultados obtidos considerando-se os métodos Random Forest e C4.5 para classificação de imagens hiperespectrais, o melhor método de árvores de decisão é o Random Forest, devido a sua maior qualidade de árvores de decisão que o método C4.5.

Em futuras pesquisas, pode-se apresentar novas análises com outros métodos de classificação comparados ao método Random Forest, visto que, o mesmo foi comprovado como método mais preciso das árvores de decisão.

#### REFERÊNCIAS

ANJOS, C. S. **Classificação de Áreas Urbanas com Imagens Multiespectrais e Hiperespectrais utilizando Métodos Não-Paramétricos** - 381p. Tese (Doutorado) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2016.

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

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.

QUINLAN, J. R. **C4.5: programs for machine learning**. San Mateo: Morgan Kaufmann Publishers, 1993, 300p.