

**11ª Jornada Científica e  
Tecnológica do IFSULDEMINAS**

**& 8º Simpósio de  
Pós-Graduação**

## **AVALIAÇÃO DE ALGORITMOS DE SELEÇÃO DE ATRIBUTOS**

**Taiane A. F. CARVALHO<sup>1</sup>; Camila S. dos ANJOS<sup>2</sup>; Jadson M. da SILVA<sup>3</sup>**

### **RESUMO**

Nesse trabalho foi testada a qualidade de uma classificação supervisionada, usando diferentes métodos de seleção de atributos. Para esse estudo utilizou-se de uma imagem multiespectral WorldView-3 de alta resolução espacial da capital da Suécia, Estocolmo. Foram gerados 267 atributos derivados da imagem original, e então foi realizada a seleção dos atributos pelos algoritmos Cfs Subset Eval, Classifier Attribute Eval, GainRatio Attribute Eval, InfoGain Attribute Eval e Wrapper Subset Eval disponíveis no *software* WEKA 3.8. Após foi realizada a classificação dos diferentes conjuntos de dados. Para a validação foi realizada a comparação do índice kappa gerado por cada classificação, que demonstrou que a imagem classificada após a seleção de atributos pelo método Cfs Subset Eval apresentou maior exatidão global.

### **Palavras-chave:**

Algoritmos; Classificação supervisionada; Processamento de imagens.

### **1. INTRODUÇÃO**

Análise de imagens é uma das principais tarefas envolvidas em um sistema de classificação. Seu objetivo é obter informações suficientes para distinguir entre diferentes regiões de interesse. Normalmente, o processo de classificação é baseado nos níveis de cinza ou cores, características de forma e textura. Diversas aplicações utilizam técnicas de classificação para reconhecimento de objetos em imagens, tais como análise de imagens multiespectrais e interpretação de fotografias aéreas (SCHWARTZ e PEDRINI, 2003).

Porém, nos casos em que a tarefa de classificação deve ser feita considerando dados pertencentes a espaços de grande dimensão e nos casos em que os atributos disponíveis para caracterizar cada amostra não esclarecem de forma óbvia o que diferencia um padrão pertencente a uma classe de outro pertencente a outra classe, o ser humano vai encontrar muitas dificuldades para executar a classificação. Sendo assim, a automatização do processo de classificação passa a ser de grande interesse e sua viabilidade aumenta conforme cresce o poder de processamento e memória dos computadores (SEMOLINI et al, 2002).

A grande dimensionalidade do espaço de atributos pode causar um alto custo de processamento e prejudicar a exatidão da análise das cenas. O termo seleção de atributos refere-se aos algoritmos que selecionam o melhor subconjunto a partir de um conjunto de atributos dado que

---

1 Discente, IFSULDEMINAS – *Campus* Inconfidentes. E-mail: taianeapcarvalho@gmail.com.br.

2 Docente, IFSULDEMINAS – *Campus* Inconfidentes. E-mail: camila.lacerda@ifsuldeminas.edu.br.

3 Discente, IFSULDEMINAS – *Campus* Inconfidentes. E-mail: jadson.silva@alunos.ifsuldeminas.edu.br.

conduza a um menor erro na classificação (SANTOS et al, 2005).

Este trabalho busca avaliar a exatidão da classificação realizada com um grande número de atributos e com os atributos selecionados por diferentes algoritmos de seleção de atributos Cfs Subset Eval, Classifier Attribute Eval, GainRatio Attribute Eval, InfoGain Attribute Eval e Wrapper Subset Eval disponíveis no *software* WEKA 3.8.

### 3. MATERIAL E MÉTODOS

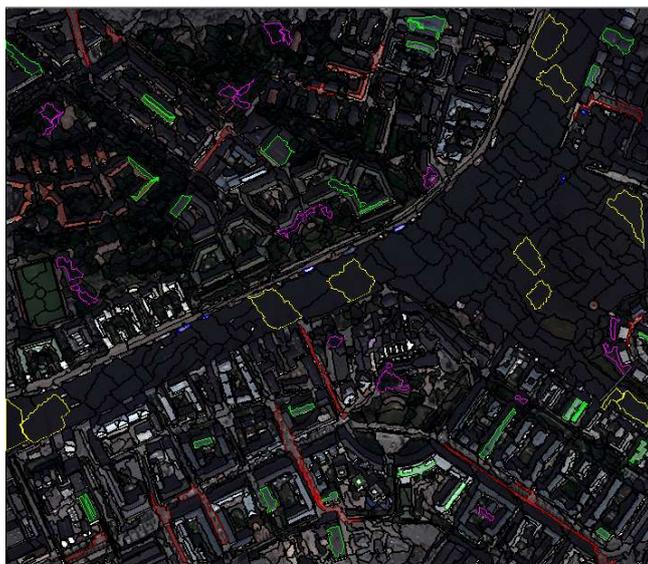
A imagem utilizada nesse trabalho foi obtida no site *DigitalGlobe*, na cidade de Estocolmo capital da Suécia (Figura 1). A imagem possui 8 bandas multiespectrais que envolvem o visível e infravermelho próximo com resolução espacial de 1,6 m, e possui a banda pancromática com resolução espacial de 0,4 m.



**Figura 1 – Imagem WorldView-3 sobre parte da cidade de Estocolmo (data).**

A fim de obter uma imagem com alta resolução espacial e alta resolução espectral realizou-se a fusão pelo método de componentes principais. Com o objetivo de incluir dados derivados da imagem original foram extraídos os atributos Normalized Difference Vegetation Index (NDVI) e as componentes principais para a imagem.

Em seguida foi realizada a segmentação da imagem (Figura 2). Então foi realizada a seleção das amostras para as classes determinadas (Rua, Água, Construção, Vegetação, Barco). Além disso foi realizada a extração de atributos derivadas dos segmentos como, por exemplo: média, desvio padrão, área, comprimento, moda, entre outros. Totalizando 267 atributos.



**Figura 2 – Imagem segmentada e amostras**

Após a extração de atributos foi realizada a classificação no *software* Weka 3.8 pelo algoritmo Neive Bayes da imagem usando todos os atributos. Então foi realizada a seleção de atributos pelos algoritmos Cfs Subset Eval, Classifier Attribute Eval, GainRatio Attribute Eval, InfoGain Attribute Eval e Wrapper Subset Eval disponíveis no *software* WEKA 3.8. E então foi realizada a classificação pelo método NeiveBayes utilizando apenas os atributos selecionadas por cada algoritmo.

A validação do trabalho foi realizada comparando o índice kappa gerado por todas as classificações.

#### **4. RESULTADOS E DISCUSSÕES**

Os algoritmos de seleção de atributos Classifier Attribute Eval, Gain Ratio Attribute Eval, Info Gain Attribute Eval, realiza o ranqueamento dos melhores para os piores atributos, então foi determinado a utilização dos atributos que possuíam relevância maior que 0.

Depois de definidos os atributos para cada método de seleção, foi realizada a classificação pelo método Neive Bayes, obtendo então o índice kappa para cada classificação (Tabela 1).

Com pode-se notar na Tabela 1, um dos algoritmos de seleção de atributos foi mais eficiente na classificação, com um kappa de aproximadamente 4% maior.

Algoritmo de seleção	Número de Atributos	Kappa
Cfs Subset Eval	27	0,7329
Classifier Attribute Eval	95	0,6893
Completo	267	0,6893
Gain Ratio Attribute Eval	89	0,6893
Info Gain Attribute Eval	89	0,6893
Wrapper Subset Eval	7	0,6889

*Tabela 1 – Comparação entre os índices kappa para os diferentes algoritmos.*

Três dos algoritmos utilizados apresentaram o mesmo índice kappa que o gerado pelo conjunto completo, entretanto destaca-se que o esforço computacional para a classificação usando os algoritmos de seleção de atributos é muito menor já que o conjunto selecionado pelos algoritmos Classifier Attribute Eval, Gain Ratio Attribute Eval, Info Gain Attribute Eval possui um terço do tamanho do conjunto completo e o conjunto selecionado pelo algoritmo Wrapper Subset Eval é 38 vezes menor que o conjunto completo.

## 5. CONCLUSÕES

Quando se trabalha com um conjunto grande de dados (*Big Data*) é necessário realizar uma seleção de atributos devido ao alto custo de processamento, como foi visto nesse trabalho é possível reduzir o conjunto de dados em até 38 vezes mantendo a mesma qualidade da classificação de imagens gerada pelo conjunto completo. E com esses algoritmos de seleção de atributos, consegue se trabalhar com dados específicos para um determinado trabalho.

## REFERÊNCIAS

- SANTOS, Joelma Carla; DE FREITAS OLIVEIRA, João Ricardo; DUTRA, Luciano Vieira. Uso de Algoritmos Genéticos na Seleção de Atributos para Classificação de Regiões. In: **GeoInfo**. 2005. p. 253-261.
- SCHWARTZ, William Robson; PEDRINI, Hélio. Método para classificação de imagens baseada em matrizes de coocorrência utilizando características de textura. **III Colóquio Brasileiro de Ciências Geodésicas, Curitiba PR, Brasil**, p. 1, 2003.
- SEMOLINI, Robinson et al. Support vector machines, inferência transdutiva e o problema de classificação. 2002.