



Análise de Integração de Mineradores de Dados com a Plataforma InterIMAGE – Qual a Melhor Solução?

Analysis of Integration of Data Mining with the InterIMAGE Platform -
What is the best solution?

Rodrigo Rodrigues Antunes¹
Edilson de Souza Bias²
Ricardo Seixas Brites³
Gilson A. O. P. Costa⁴

Recebido em novembro de 2017.
Aprovado em outubro de 2018.

RESUMO

O objetivo principal da pesquisa foi avaliar a integração de vários sistemas mineradores de dados open source com o sistema de classificação baseado em objetos InterIMAGE. Os sistemas mineradores utilizados nesta pesquisa foram: SIPINA, RapidMiner Studio, KNIME Analytics Platform, Orange Canvas e WEKA. Para este estudo foi utilizada a imagem do sensor GeoEye-1, de 2013, da área urbana do município de Goianésia, Goiás, Brasil. O sistema InterIMAGE possibilitou a segmentação multiresolução e a coleta de amostras (com os respectivos atributos) de cada classe definida. Os resultados foram avaliados por meio da métrica de acurácia Kappa, demonstrando-se por meio de bons resultados a viabilidade de integração do sistema InterIMAGE com os mineradores utilizados. Pelos resultados ficou demonstrado que a quantidade de regras influencia diretamente o tempo de processamento da classificação no InterIMAGE mas não no resultado (acurácia da classificação). O resultado de cada classificação foi satisfatório, de acordo com o índice Kappa de cada classificação: SIPINA-ID3 (66% Kappa), SIPINA-C45 (68% Kappa), SIPINA-GID3 (70% Kappa), SIPINA-Assistent86 (70% Kappa), SIPINA-CHAID (70% Kappa), RapidMiner (77% Kappa), KNIME (73% Kappa) Orange Canvas (81% Kappa) e WEKA (78% Kappa). Além disso, para analisar a significância de cada resultado da classificação por meio dos índices utilizou-se um teste estatístico (teste Z).

¹Universidade de Brasília (UnB), Instituto de Geociências (IG), Brasília-DF, Brasil. E-mail: rodrigorantunes@hotmail.com, edbias@unb.br, brites@unb.br

²Universidade do Estado do Rio de Janeiro (UERJ), Departamento de Informática e Ciência da Computação Rio de Janeiro-RJ, Brasil. E-mail: gilson.costa@ime.uerj.br

PALAVRAS-CHAVE: OBIA. InterIMAGE. Data Mining. Kappa. teste z.chave.

ABSTRACT

The main objective of the research was to evaluate the integration of several open-source data mining software packages with the object-based classification system InterIMAGE. The data mining systems used in this research were: SIPINA, RapidMiner Studio, KNIME Analytics Platform, Orange Canvas and WEKA. We used a GeoEye-1 sensor image from 2013, covering the urban area of the city of Goianésia, Goiás, Brazil. The open source system InterIMAGE was used for the multiresolution segmentation and sample collection, for each class of interest. The results were evaluated through the comparison of the Kappa Index and demonstrate the viability of the integration of the data mining systems with InterIMAGE. They showed that the total number of rules directly influenced the classification processing time in InterIMAGE but not the quality of the results (accuracy of classification). The result of each classification was satisfactory, according to the Kappa accuracy index: SIPINA-ID3 (66% Kappa), SIPINA-C45 (68% Kappa), SIPINA-GID3 (70% Kappa), SIPINA-CHAID (70% Kappa), RapidMiner (77% Kappa), KNIME (73% Kappa) Orange Canvas (81% Kappa) and WEKA (78% Kappa). In order to analyze the significance of each classification result, by means of the Kappa indices a statistical test (Z test) was carried out.

KEYWORDS: OBIA. InterIMAGE. Data Mining. Kappa. z test.

* * *

Introdução

Um dos principais usos de imagens de sensoriamento remoto (SR) é a extração de informações referente à cobertura da terra (GRIPPA et al., 2016). Do ponto de vista científico, para se extrair informações de imagens de SR é necessário estabelecer métodos de análise com regras claras e coerentes, que possam ser reproduzidas por qualquer analista no processo de interpretação. Essas regras estabelecem critérios para a análise dos elementos que compõem a paisagem, e se baseiam em propriedades como cor, tonalidade, textura, estrutura e homologia (MENESES; SANO, 2012).

A evolução dos sensores de alta resolução espacial expôs as limitações das técnicas tradicionais de classificação pixel-a-pixel (BLASCHKE, 2001).

Vários dos aspectos que envolvem a análise de imagens de SR, notadamente de imagens de alta resolução espacial, não podem ser

considerados utilizando-se apenas informações provenientes de pixels individuais, uma vez que estas não representam o contexto geográfico (BLASCHKE; BURNETT; PEKKARINEN, 2004). A técnica de OBIA – Object Based Image Analysis permite a inclusão de informações adicionais nos processos de classificação e modelagem, relativas a segmentos, ou objetos de imagens, que são agrupamentos contíguos de pixels. Essas informações podem estar relacionadas com estatísticas sobre os conjuntos de pixels que compõem os objetos, sobre a textura e forma dos objetos, ou sobre relações topológicas entre objetos (BLASCHKE, 2013).

Métodos baseados em OBIA possuem geralmente três etapas principais: segmentação de imagem; extração de atributos, ou propriedades dos segmentos; e classificação. A segmentação de imagem é procedimento para particionar uma área da imagem inteira em segmentos (ou objetos de imagem), que são grupos de pixels com valores espectrais homogêneos.

Diversos autores têm abordado e discutido os conceitos de objeto nas classificações que utilizam OBIA. Furtado et al. (2012) descreve que segmentar uma imagem, consiste em determinar agrupamentos de pixels adjacentes. Castilla e Hay (2008) descrevem o segmento como uma região discreta, coerente internamente e diferente do seu contexto. Autores como Navulur (2007) e Camargo et al. (2009) afirmam que após a segmentação a imagem está constituída de objetos, possibilitando a extração de atributos como: textura, forma, relações topológicas, entre outros.

Blaschke (2009) realiza uma abrangente revisão da literatura sobre OBIA, onde foram selecionados 145 artigos, 84 capítulos de livros e quase 600 artigos de conferências. Sobre a segmentação, esclarece que ela pode ser dividida em quatro categorias: baseada em pontos; baseada em borda; baseada em região e combinada. Blaschke (2009) citando Hay e Castilla (2008), descreve que não importa qual dos métodos é aplicado, pois a segmentação fornecerá os blocos de construção de análise de imagem, baseada em objetos.

Após a segmentação e análise de segmentos representativos de cada classe de objetos de interesse, a classificação é realizada com base em segmentos de imagem. Em OBIA, a qualidade da segmentação influencia diretamente o resultado da classificação de uma imagem SR (BLASCHKE, 2003; DORREN et al., 2003; MEINEL; NEUBERT, 2004; ADDINK et al., 2007).

Uma ferramenta de destaque no contexto de classificação de imagens de alta resolução espacial é o sistema InterIMAGE, plataforma livre baseada em conhecimento para interpretação automática de imagens (COSTA et al., 2010), que possui operadores para aplicar a OBIA no processo de classificação. Entretanto, é importante ressaltar que versão atual do InterIMAGE possui limitação para processar e classificar imagem superior a 9 Megapixels (3000 x 3000 pixels), e uma nova plataforma, chamada InterCloud – InterIMAGE Cloud Platform, uma reformulação do atual sistema concebida para processar grandes volume de dados em clusters de computadores físicos ou virtuais, vem sendo desenvolvida (FERREIRA et al., 2017).

No processo de interpretação de imagem, executado pelo sistema InterIMAGE, a definição de limiares para uma regra de decisão é importante para que se atinja uma boa acurácia na classificação. Estas regras podem ser definidas contemplando-se, além de valores espectrais, propriedades texturais, morfológicas e topológicas de objetos de imagens.

A utilização do InterIMAGE, integrada com um sistema de mineração de dados, para auxiliar a definição de regras e valores de limiares, tem apresentado bons resultados, sendo que, em diversos trabalhos publicados até o momento, o sistema minerador WEKA (GHOLAP, 2012) foi utilizado. Neste contexto, é interessante testar e avaliar a integração de outros sistemas mineradores de dados e seus algoritmos de classificação com o InterIMAGE, de modo a disponibilizar ao analista alternativas de mineradores eficientes para o desenvolvimento de projetos.

Otukei e Blaschke (2010) fizeram comparação entre várias técnicas de classificação de mineração de dados para avaliar alterações no uso do solo

através de OBIA. Entre as técnicas, foram testadas: Árvore de Decisão; Máquinas de Vetores de Suporte; e Máxima Probabilidade. Os autores concluíram que o desempenho do processo de classificação foi melhor com a técnica de Árvore de Decisão.

Neste trabalho, a técnica de classificação de Árvore de Decisão (BREIMAN et al., 1984) foi investigada no contexto de OBIA. Foram analisados diversos algoritmos, implementados nos seguintes pacotes para a mineração de dados: SIPINA, RAPIDMINER STUDIO, KNIME ANALYTICS PLATFORM, ORANGE CANVAS e WEKA. Os resultados produzidos pelos vários algoritmos, em termos de regras de decisão e valores de limiares foram inseridos no sistema InterIMAGE, para a classificação baseada em objetos, tendo sido analisadas as acurácias das respectivas classificações.

A presente pesquisa se justifica pelo fato de o software WEKA ser o único minerador de dados utilizado e identificado nos artigos que desenvolvem aplicações com OBIA, em combinação com o sistema InterIMAGE, conforme pode-se observar nos trabalhos:

- a) Análise de imagem baseada em objeto e mineração de dados aplicadas à classificação do uso do solo urbano por quadra em imagens WorldView-2 (CARVALHO et al., 2013);
- b) Classificação de imagem baseada em objeto (OBIA) utilizando índices de vegetação (KAWASHIMAET et al., 2013);
- c) Classificação da Cobertura da Terra, utilizando os Programas Livres: InterIMAGE, WEKA e QuantumGIS (NASCIMENTO et al., 2013);
- d) Geobias e mineração de dados na classificação da cobertura do solo urbano em São Luís (MA) com imagens Worldview-2 e o sistema InterIMAGE (SOUSA; KUX, 2014);
- e) Desenvolvimento de técnica para monitoramento do cadastro urbano baseado na classificação orientada a objetos. Estudo de caso: Município de Goianésia, Goiás (ANTUNES et al., 2015);

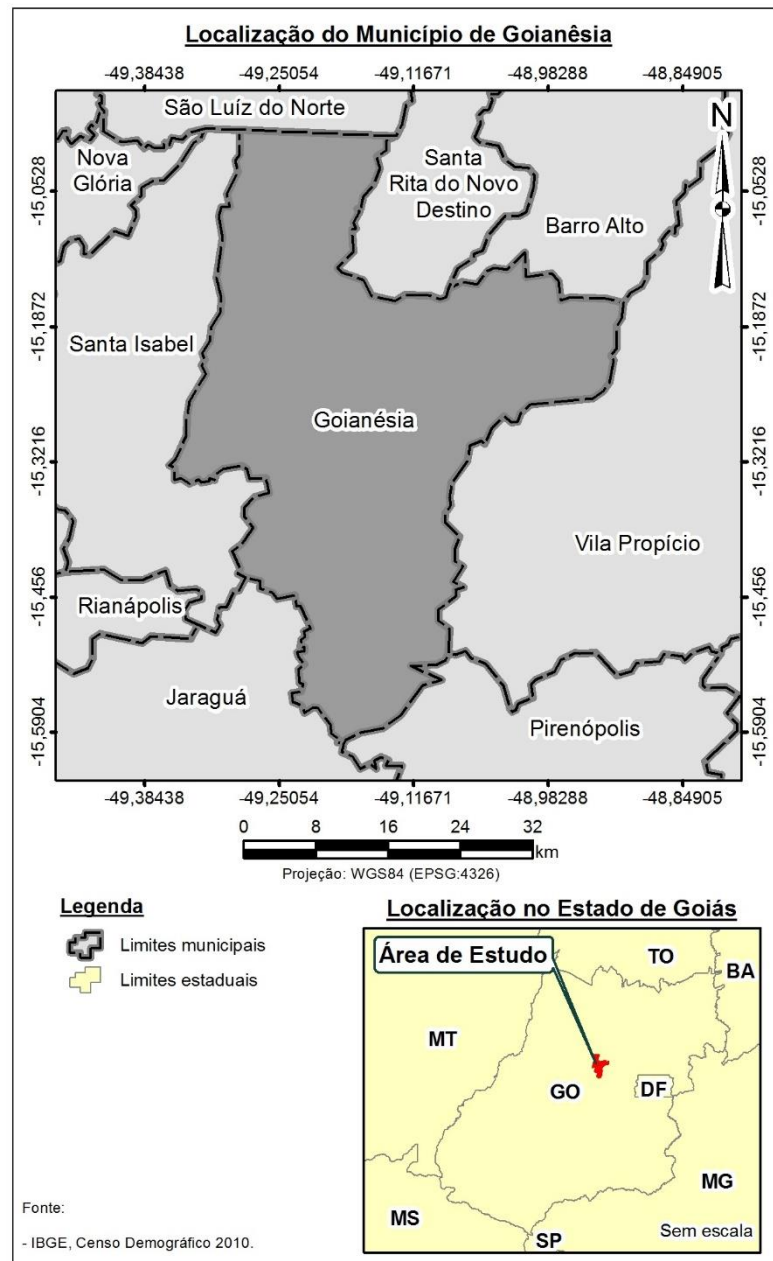
- f) Técnicas de mineração de dados aplicadas à classificação do estágio sucessional da vegetação em áreas de floresta ombrófila mista (SOTHE et al., 2016);
- g) Mapeamento da cobertura da terra do município de Raposa (Ma) utilizando imagens Worldview-II, o aplicativo InterIMAGE e mineração de dados (MENEGHETTI et al., 2014);
- h) Aplicação de mineração de dados e técnicas GEOBIA para análise de susceptibilidade ao fogo no Parque Nacional Itatiaia, no Brasil (DE SOUSA et al., 2014);
- i) Uso do sistema InterIMAGE para a identificação de alvos urbanos em imagens do satélite Worldview II (PASSO et al., 2013);
- j) Classificação orientada a objeto em associação às ferramentas reflectância acumulada e mineração de dados (DE GRANDE et al., 2017);
- k) Aplicação da Análise de Imagens Baseada em Objetos como Ferramenta de Monitoramento do Cadastro Urbano em Pequenos Municípios (BIAS et al., 2014); e
- l) Análise do nível de legenda de classificação de áreas urbanas empregando imagens multiespectrais e hiperespectrais com os métodos árvore de decisão C4.5 e floresta randômica (DOS ANJOS et al., 2017).

Assim sendo, o objetivo do presente trabalho foi avaliar os principais pacotes de software mineradores de código aberto, além do WEKA, e seus algoritmos de indução de árvore de decisão, de forma a avaliar a integração com o sistema InterIMAGE na classificação baseada em objetos de imagens de SR.

1 Área de Estudo e Materiais

A imagem de SR utilizada nos experimentos recobre uma área localizada no município de Goianésia (Figura 1), na parte centro-norte do estado de Goiás, Brasil.

Figura 1- Localização da área de estudo – Goianésia, Goiás, Brasil.



Fonte: Elaborado pelos autores.

Trata-se de uma imagem do sensor GeoEye-1, originalmente com uma banda pancromática, com resolução espacial de 50 cm, e quatro bandas multiespectrais (azul, verde, vermelho e infravermelho), cada uma com resolução espacial de 2 metros.

Nos experimentos utilizou-se recorte fusionado (*pansharpened*) da região central da cidade de Goianésia, com 977 por 1531 pixels (ANTUNES et al., 2015).

Os pacotes de software mineradores de dados utilizados nesta pesquisa foram:

- a) SIPINA é um software livre e acadêmico desenvolvido pela Universidade de Lyon, França, voltado à mineração de dados e indicado para a indução de árvores de decisão (*Classification Trees*). Teve sua primeira versão distribuída em 1995 (KAUR; SINGH, 2013) e, de acordo com Rakotomalala (2008), o mesmo implementa vários algoritmos de aprendizado supervisionado, com uma construção interativa e visual de árvore de decisão. Os principais algoritmos de indução de árvore de decisão disponíveis e testados no SIPINA foram: ID3, C45, GID3, Assistant86 e CHAID. O algoritmo ID3 (Iterativo DiChaudomiser 3) foi originalmente desenvolvido por J. Ross Quilan (1986), na Universidade de Sydney, Austrália. Nesse algoritmo, a escolha dos atributos a serem utilizados pela árvore se dá a partir de informações de entropia e ganho de informação. Para medir o ganho de informação de um atributo se utiliza o conceito de entropia, da Teoria da Informação. O valor da entropia corresponde à impureza do atributo, sendo o ganho de informação a variação da impureza. Quanto menor o valor da entropia, menor a incerteza e mais utilidade tem o produto para a classificação (WILGES et al., 2010). O algoritmo C.45 foi proposto em 1993, novamente por Ross Quinlan, para superar as limitações do algoritmo ID3 como, por

exemplo, a dificuldade de tratar um grande número de valores (HSSINA et al., 2014).

O algoritmo GID3 é uma generalização do ID3 e C45, de Quilan, onde algumas folhas de um processo de separação podem ser mescladas. A ideia é destacar as folhas mais interessantes (FAYYAD, 1994). O algoritmo Assistant86 é outra melhoria do ID3, de Quilan. Nesse algoritmo há um critério de melhoria do ganho da informação e vários parâmetros que permitem controlar o tamanho da árvore (CESTNIK; KONONENKO; BRATKO, 1987).

O algoritmo CHAID é uma versão aprimorada do algoritmo AID de Morgan e Sonquist (1963). Segundo Kass (1980), as particularidades do CHAID são: (1) A utilização da estatística qui-quadrado como no critério de divisão; (2) a fusão de algumas folhas que vêm do mesmo nó. Este algoritmo é amplamente difundido em softwares comerciais. Por outro lado, é menos implementado em software livre;

- b) RapidMiner Studio é um sistema de código aberto, desenvolvido na linguagem Java, para concepção de processos analíticos avançados com aprendizagem de máquina, mineração de dados, mineração de texto, análise preditiva e análise de negócios (RAPIDMINER, 2017). A primeira versão do RapidMiner Studio surgiu em 2001 com nome de YALE (Yet Another Learning Environment), sistema desenvolvido pelo departamento de inteligência artificial da Universidade de Dortmund, da Alemanha (FAULHABER, 2007). A versão atual (7.0) do RapidMiner Studio permite desenvolver o processo de análise de dados por meio de vários operadores disponíveis em um ambiente de visualização gráfica;
- c) O KNIME Analytics Platform é um sistema de código aberto, baseado na plataforma Eclipse (Java) com possibilidade de customizar e implementar novos módulos em um ambiente visual (KNIME, 2017). Esta modularidade permite a aplicação do KNIME

Analytics Platform em ambientes de produção comercial, ensino e pesquisa. O desenvolvimento do KNIME Analytics Platform começou em 2004, por um grupo de desenvolvedores na Universidade de Konstanz, Alemanha. O objetivo era atender demandas de processamento e análise de grande quantidade de dados de uma indústria farmacêutica. Em 2006, foi lançada a primeira versão deste minerador. Atualmente, encontra-se na versão 3.1.2. O software permite trabalhar de forma compartilhada e colaborativa (KNIME Team Space e KNIME Server Lite), suporta escalabilidade sob demanda em nuvem (KNIME Cloud Server), possui extensão para Big Data (Hadoop) (KNIME Big Data Extensions e KNIME Big Data Connectors) e pode ser executado em ambiente de processamento em cluster (KNIME Cluster Execution) (KNIME, 2017);

- d) Orange Canvas, foi desenvolvido na Linguagem Python. É um conjunto abrangente de software baseado em componentes para mineração de dados e aprendizagem de máquina, desenvolvido no Laboratório de Bioinformática, da Faculdade de Computação e Ciência da Informação da Universidade de Ljubljana, Eslovênia, juntamente com a comunidade de código aberto (DEMSAR; CURK; ERJAVEC, 2013). O Orange Canvas trabalha com blocos de construção de fluxos chamado Widgets para análise de dados que são montados no ambiente de programação visual do sistema. Os Widgets são agrupados em classes de acordo com sua função.
- e) WEKA (Waikato Environment for Knowledge Analysis) é um software de mineração de dados de código aberto desenvolvido na linguagem Java. É desenvolvido na Universidade de Waikato (Nova Zelândia) e teve a sua primeira versão lançada em 1997. Este sistema possui um conjunto de algoritmos de aprendizado de máquina para tarefas de mineração de dados. Os algoritmos podem ser aplicados diretamente a um conjunto de dados ou chamado a

partir de seu próprio código Java. Vários são os algoritmos de indução de árvore de decisão disponíveis WEKA, entre eles J48, REPTree, Decision Stump, RandomTree e outros. Para este trabalho foi testado e analisado o J48, que é a implementação do algoritmo em C4.5 no minerador WEKA (GHOLAP, 2012).

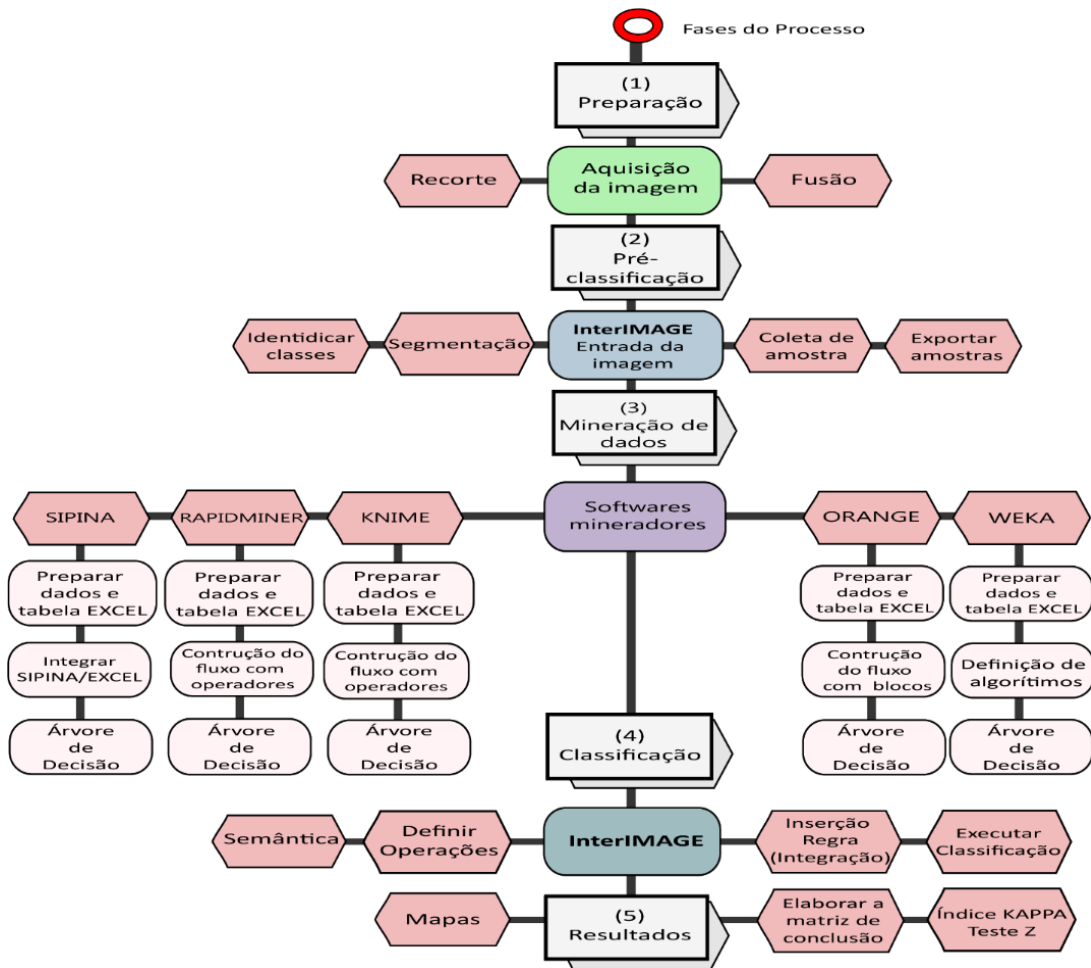
Para a segmentação e classificação baseada em objetos foi utilizado o sistema InterIMAGE 1.43 (InterIMAGE, 2010).

Para a leitura de dados matriciais, vetoriais e confecção de mapas foi utilizado o software QuantumGIS, Versão 2.18.1 64 bits (QGIS BRASIL, 2017).

3 Metodologia

A Figura 2 representa as etapas metodológicas cumpridas para alcançar os resultados da pesquisa.

Figura 2-Esquema das etapas do trabalho.



Fonte: Elaborado pelos autores.

3.1 Fase (1): Preparação

A imagem original GeoEye-1 foi fusionada (*pansharpened*), utilizando-se o método GS – Gram-Schmidt, o software foi o ENVI 5.0. Posteriormente por meio do software Quantum GIS foi realizado recorte da área central, em uma dimensão de 977x1531 pixels.

3.2 Fase (2): Pré-classificação

Nessa fase foram definidas as classes de interesse: telhado metálico, telhado de amianto, telhado de cerâmica clara, telhado de cerâmica escura, piscina, vegetação, solo exposto, piso de concreto, vias urbanas e sombra.

A segmentação da imagem foi desenvolvida por meio do operador TA_Baatz_Segmenter do InterIMAGE, adaptado de Baatz e Shape (1999), gerando diferentes segmentações para diferentes classes, de acordo com os parâmetros apresentados na Tabela 1.

Tabela 1- Parâmetros utilizados na segmentação.

Classes / n° de amostras	Banda de entrada	Peso das bandas	Compacidade	Cor	Escala
Metálico / 25	0,1,2,3	1,1,1,1	0.8	0.4	90
Amianto / 25	0,1,2,3	1,1,1,1	0.7	0.5	80
Cerâmica clara / 20	0,1,2,3	1,1,1,1	0.5	0.5	80
Cerâmica escura / 20	0,1,2,3	1,1,1,1	0.5	0.5	70
Piscina / 30	0,1,2,3	1,1,1,1	0.8	0.4	60
Solo exposto / 30	0,1,2,3	1,1,1,1	0.5	0.5	60
Piso de concreto / 25	0,1,2,3	1,1,1,1	0.5	0.5	60

Fonte: Elaborado pelos autores.

Após a segmentação, foram coletadas amostras de cada classe definida, utilizando a janela *samples editor* do InterIMAGE. Os seguintes atributos foram selecionados e exportados: brightness, mean1, mean2, mean3, mean4, maxPixVal1, maxPixVal2, maxPixVal3, maxPixVal4, minPixVal1, minPixVal2, minPixVal3, minPixVal4, ratio1, ratio2, ratio3, ratio4, bandDiv41, bandDiv43, mean1, mean2, mean3, mean4, area, angle, compactness, squareness, and roundness.

Para as classes não apresentadas na Tabela 1 (vegetação, vias urbanas e sombra) não foram coletadas amostras, pois foram utilizados os seguintes operadores do InterIMAGE: TA_NDVI_Segmenter para vegetação, TA_ShapeFile_Import para vias urbanas e TA_Arithmetic para sombra.

3.3 Fase (3): Mineração de Dados

Utilizando a planilha Excel, foi possível preparar os dados para leitura do minerador (pré-processamento), permitindo iniciar o processo de mineração de dados com cada sistema minerador (SIPINA, RapidMiner Studio, KNIME Analytics Platform, Orange Canvas e WEKA).

3.4 Fase (4): Classificação – InterIMAGE

Na classificação baseada em objetos, os segmentos são vistos como os objetos e auxiliam na busca de identificação das diferentes características que definem um objeto. Estes objetos têm características geográficas como forma, comprimento e entidades topológicas (NAVULUR, 2007; CAMARGO et al., 2009). Para Orlando e La Rosa (2014), estes atributos formam uma base de conhecimento para os objetos, que podem ser chamados no processo de classificação.

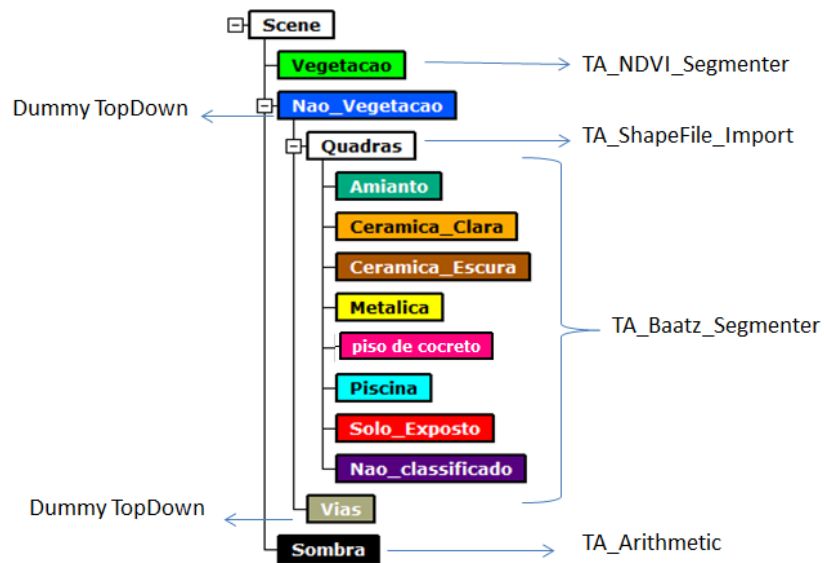
O InterIMAGE permitiu a criação da rede semântica das classes de acordo com o objetivo do trabalho e a janela TopDown Decision permitiu criar um conjunto de expressões, denominadas regras de decisão. Estas expressões definem o conhecimento estruturado e explícito do analista e são utilizadas pelo sistema no processo de interpretação (InterIMAGE, 2010). De acordo com Costa et al. (2010), no InterIMAGE, o conhecimento explícito sobre os objetos, extraível a partir da própria cena, é modelado em uma rede semântica definida pelo usuário por meio de uma interface gráfica. A rede semântica é uma forma alternativa de representação de classe, que apresenta a maneira pela qual as diferentes categorias são conectadas com as respectivas classes de interesses e como estão espacialmente relacionadas com estas mesmas classes. Elas modelam e armazenam o conhecimento especializado (MAVRANTZA; ARGIALAS, 2008).

Um conjunto de operadores com funções é oferecido pelo InterIMAGE e, para este trabalho, foram utilizados: TA_Baatz_Segmenter,

TA_Arithmetic, TA_ShapeFile_Import, Dummy TopDown e TA_NDVI_Segmenter (InterIMAGE Wiki, 2014).

Na Figura 3 são apresentados a rede semântica e o operador utilizado em cada classe.

Figura 3-Rede semântica e operadores utilizados no trabalho. Sistema classificador InterIMAGE.



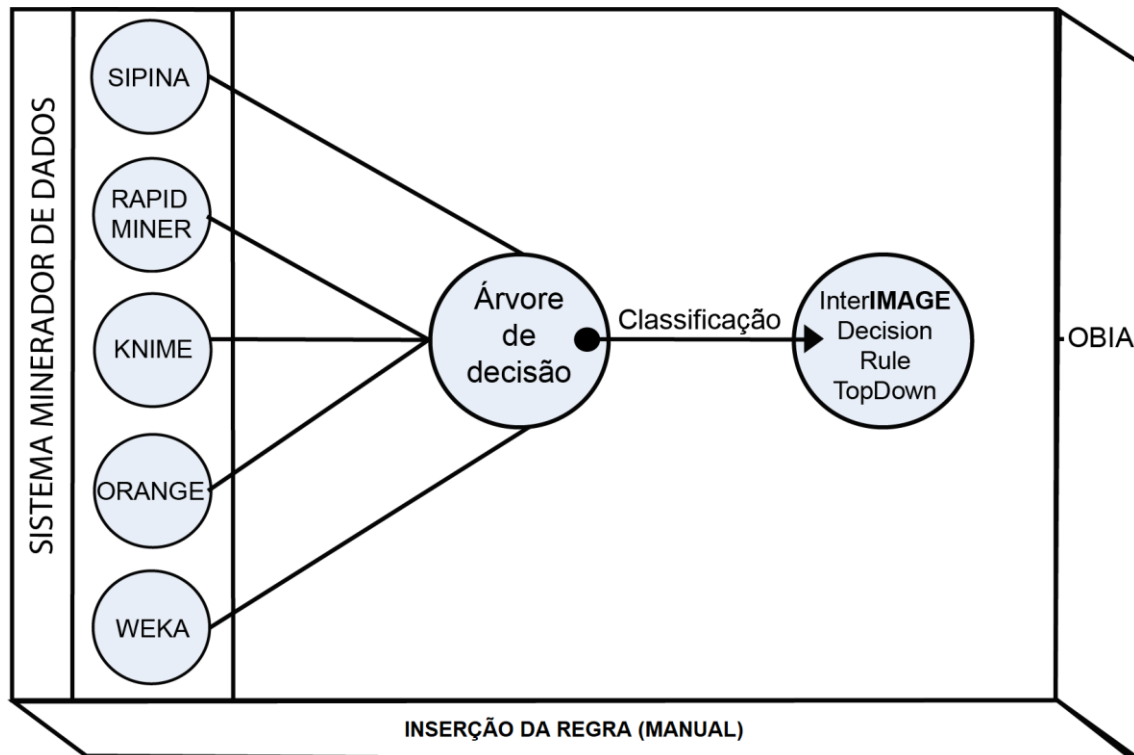
Fonte: Elaborado pelos autores.

No campo da Inteligência Artificial, a representação do conhecimento pode ser por meio de uma rede semântica, onde a informação é representada como um conjunto de nós conectados um ao outro por um conjunto de arcos rotulados que representam relações entre os nós (RICH, 1988). Para Brachman (1983), rede semântica é um grafo rotulado e direcionado formado por um conjunto de nós representando os objetos e por um conjunto de arcos representando as relações entre os objetos.

A regra de decisão com valores de limiares definidos pelos algoritmos dos mineradores SIPINA, RapidMiner Studio, KNIME Analytics Platform, Orange Canvas e WEKA foram inseridos no classificador InterIMAGE por meio do operador TopDown Decision Rule, conforme ilustrado na figura 04. Essa inserção de dados foi feita manualmente, o que representa a

possibilidade da inserção de erro por parte do operador, sendo um dos fatores limitadores da utilização de mineradores de dados para o sistema InterIMAGE.

Figura 4-Integração dos mineradores SIPINA, RapidMiner Studio, KNIME Analytics Platform, Orange Canvas e WEKA com o InterIMAGE para classificação.



Fonte: Elaborado pelos autores.

4 Análise da Qualidade das Classificações

A análise da acurácia temática foi aplicada de acordo com as seguintes avaliações: quantificação do número de amostras, distribuição aleatória de pontos de checagem, investigação por interpretação visual, composição da matriz de confusão e índice de concordância Kappa. Por meio de uma distribuição multinomial, foi calculado o número de amostras. A unidade amostral utilizada foi o pixel. Detalhes referentes aos procedimentos podem ser consultados em Passo et al. (2013) e Antunes et al. (2014).

De acordo com Cadena (2011), normalmente duas estratégias podem ser usadas no sentido de calcular o número de amostras necessárias para computar a acurácia: a distribuição binomial e a multinomial. O modelo binomial é apenas apropriado para computar o número de amostras necessário para uma única classe visto que este simplesmente faz a distinção entre classificação correta e incorreta e não leva em conta o conjunto de classes analisadas. Por outro lado, na distribuição multinomial, o processo de validação não é uma questão apenas de certo ou errado, mas o erro deve ser classificado em relação ao total de possibilidades (ou conjunto de classes), de modo que o uso dessa distribuição implica no conhecimento a priori do número de classes e suas proporções no mapa.

Segundo Congalton e Green (1999), a criação de matriz de erros não é simplesmente questão de correto ou incorreto (o caso binomial), e sim questão de qual ou quais categorias de erros estão confusas. Portanto, o uso de uma distribuição binomial para a determinação do tamanho da amostra para uma matriz de confusão não é apropriado e sim uma distribuição multinomial, já que esta leva em consideração o número de classes.

Neste trabalho, após o processo de classificação e da geração de um mapa temático, foi criada, uma matriz de confusão para analisar a qualidade (testar a exatidão das aplicações dos diversos mineradores), aplicando-se o índice Kappa. Para analisar a significância de cada resultado da classificação utilizou-se teste estatístico (teste Z).

A quantidade de amostras foi definida a partir da Equação (1), que gerou um quantitativo de 664 amostras (CONGALTON; GREEN, 1999).

$$N = \frac{B \prod_i (1 - \Pi_i)}{b^2_i} \quad (1)$$

Onde:

N = número de amostras;

B = obtido da tabela de distribuição qui-quadrado;

1 Grau de liberdade e $1 - \alpha / k$;

Π_i = no mapa, a proporção da classe com maior ocorrência;

$1 - \alpha$ = grau de confiança;

K = número de classes;

b = erro admissível.

O índice de acurácia Kappa (COHEN, 1960) é expresso de acordo com a Equação (2).

$$K = \frac{P_o - P_c}{1 - P_c} \quad (2)$$

P_o = Precisão Global (Proporção de unidades que concordam); e

P_c = Proporção de unidades que concordam por coincidência, representada pela Equação (3):

$$P_c = \frac{\sum_{i=1}^M n_{i+} n_{+i}}{N^2} \quad (3)$$

M = número de classes;

n_{i+} = total de elementos classificados para categoria i;

n_{+i} = total de elementos de referência amostrados para uma categoria i; e

N = número total de amostras.

De acordo com Amorim et al. (2016), para cada índice Kappa é possível calcular um intervalo de confiança por meio da variância da amostra (σ^2) e o método comumente utilizado na comunidade do sensoriamento remoto é o teste “Z”.

A significância de um único índice tem como finalidade determinar se o nível de acerto da classificação e os dados de referência são significativamente maiores que zero. O teste feito para dois índices traduz se realmente existe diferença significativa entre os dois índices testados (CONGALTON; GREEN, 1999).

A variância da amostra de uma Matriz de Confusão pode ser calculada a partir do método de Delta (AMORIM et al., 2016), de acordo com as equações: 4, 5, 6, 7, 8, 9 e 10.

$$\sigma_k^2 = \frac{1}{n} \left[\frac{\theta_1 \cdot (1-\theta_1)}{(1-\theta_2)^2} + \frac{2 \cdot (1-\theta_1) \cdot (2\theta_1 \theta_2 - \theta_3)}{(1-\theta_2)^3} + \frac{(1-\theta_1)^2 \cdot (\theta_4 - 4\theta_2^2)}{(1-\theta_2)^4} \right] \quad (4)$$

$$\theta_1 = \frac{1}{n} \sum_{i=1}^c X_{ii} \quad (5)$$

$$\theta_2 = \frac{1}{n^2} \sum_{i=1}^c X_{i+} X_{+i} \quad (6)$$

$$\theta_3 = \frac{1}{n^2} \sum_{i=1}^c X_{ii} (X_{i+} + X_{+i}) \quad (7)$$

$$\theta_4 = \frac{1}{n^3} \sum_{i=1}^c \sum_{j=1}^c X_{ij} (X_{j+} + X_{+j})^2 \quad (8)$$

Com a variância de todos os índices Kappa calculada, os testes de significância podem ser efetivados utilizando as equações (9) e (10) que se seguem (AMORIM et al., 2016):

$$Z = \frac{k}{\sqrt{\sigma^2}} \quad (9)$$

$$Z_{1-2} = \frac{k_2 - k_1}{\sqrt{\sigma_{k_2}^2 + \sigma_{k_1}^2}} \quad (10)$$

O teste z para os índices Kappa das classificações foi realizado a 95% de significância. Quando $z > 1,96$, o teste é significativo, rejeita-se a hipótese de nulidade, podendo concluir que existe diferença estatística entre os valores calculados (OLIVEIRA et al., 2015).

Com o número de amostras definidas para a área de teste, uma pesquisa de ponto aleatório (centroide do pixel) foi aplicada através da ferramenta vetorial Quantum GIS. Posteriormente, a interpretação visual individual foi verificada (Figura 5), auxiliada pela imagem final da classificação. Os pontos verificados foram classificados em uma matriz de confusão que possibilitou calcular o índice de concordância Kappa.

Figura 5-Pesquisa de pontos aleatórios (765 amostras). Sistema Quantum GIS.

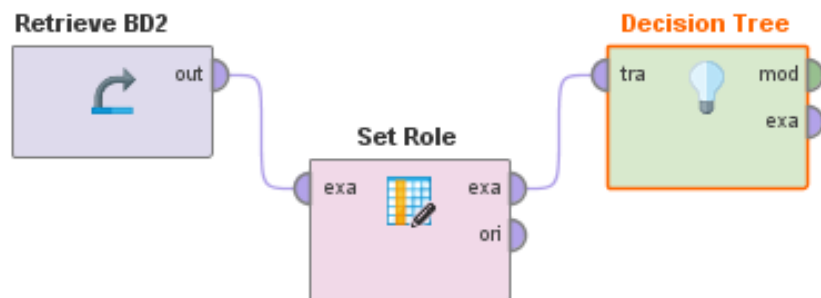


Fonte: Elaborado pelos autores.

5 Resultados e Discussão

No sistema minerador RapidMiner, é necessário utilizar diferentes operadores para construir o fluxo da classificação. Os seguintes operadores foram utilizados neste trabalho (Figura 6): Retrieve, Set Role e Decision Tree (RAPIDMINER STUDIO, 2017).

Figura 6-Operadores do RapidMiner Studio utilizados neste trabalho.

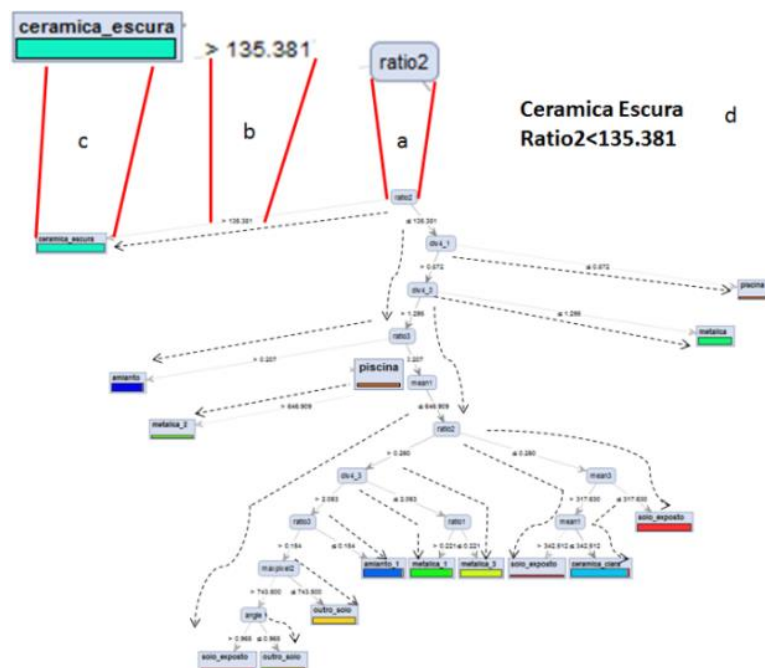


Fonte: Elaborado pelos autores.

- **Retrieve:** este operador lê os dados a partir de um diretório de armazenamento de dados. Vários formatos de dados podem ser lidos pelo minerador: csv, xls, bancos de dados SQL, etc. Também é possível ler dados em armazenamento distribuído (na nuvem), pois possui interfaces: Amazon S3, Dropbox, Sales Force, Twitter e Zapier.
- **Set Role:** este operador é usado para mudar a função de um ou mais atributo. No presente trabalho o atributo class é definido como atributo chave e os demais (Compacity, Angle, Squareness, Circleness, Brightness, Entropy, Maxpixelvalue, Mean, Minpixelvalue, Ratio e Bandmeandiv) como atributos regulares.
- **Decision Tree:** Induz uma árvore de decisão para classificação de dados nominais e numéricos.

A estrutura da árvore de decisão gerada pelo sistema minerador de dados RapidMiner Studio é mostrada na Figura 7. Nas Figuras 7a e 7b encontram-se os exemplos de atributo (razão2) e o valor de limiar (> 135.381) da classe de cerâmica escura (Figura 7c) que são mostrados no primeiro nó da árvore. Na Figura 7d representa-se a regra da classe cerâmica escura.

Figura 7-Árvore de decisão gerada pelo sistema RapidMiner Studio.



Fonte: Elaborado pelos autores.

Outros operadores para classificação são disponibilizados no RapidMiner: Decision Stump, Random Tree e Random Forest. Essas árvores não foram aplicadas neste trabalho, pois cada operador nesse sistema possui especialidade para o manuseio do conjunto de dados (RAPIDMINER, 2017):

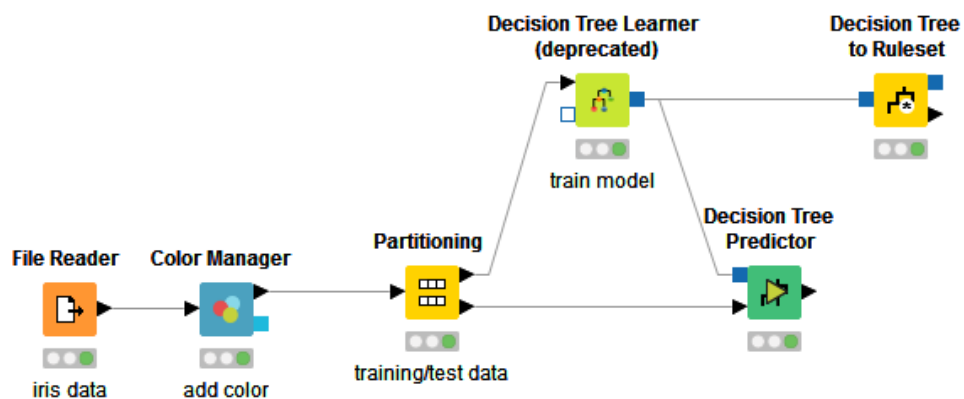
- **Operador ID3** - apenas aplicado em dados nominais para a classificação;
- **Operador CHAID** - apenas aplicado em dados nominais para classificação;
- **Operador Decision Stamp** - pode ser aplicado em dados nominais e numerais, mas gera apenas uma única divisão na árvore, o que não é aplicado a este trabalho;
- **Operador Random Tree** - este operador não utiliza todos os dados das amostras para construir a árvore de decisão. Ele utiliza apenas um subconjunto aleatório de atributos para dividir a árvore. Sendo assim, algumas classes podem ficar sem regra e valores de limiares para inserir no InterIMAGE;
- **Operador Random Forest** - este operador gera um conjunto aleatório de árvores (floresta). As árvores geradas têm a mesma característica da árvore

do operador Random Tree, ou seja, é formado por um conjunto aleatório de dados, podendo-se descartar classes no processo de classificação.

Como o RapidMiner, é necessário utilizar diferentes operadores para construir o fluxo de treinamento no minerador de dados KNIME. Os operadores do KNIME utilizados neste trabalho foram os seguintes (Figura 8): Retrieve, Set Role e Decision Tree (RAPIDMINER STUDIO, 2017).

O fluxo e os operadores utilizados do KNIME Analytics Platform são apresentados na Figura 8:

Figura 8-Fluxo e operadores utilizados no sistema KNIME Analytics Platform.



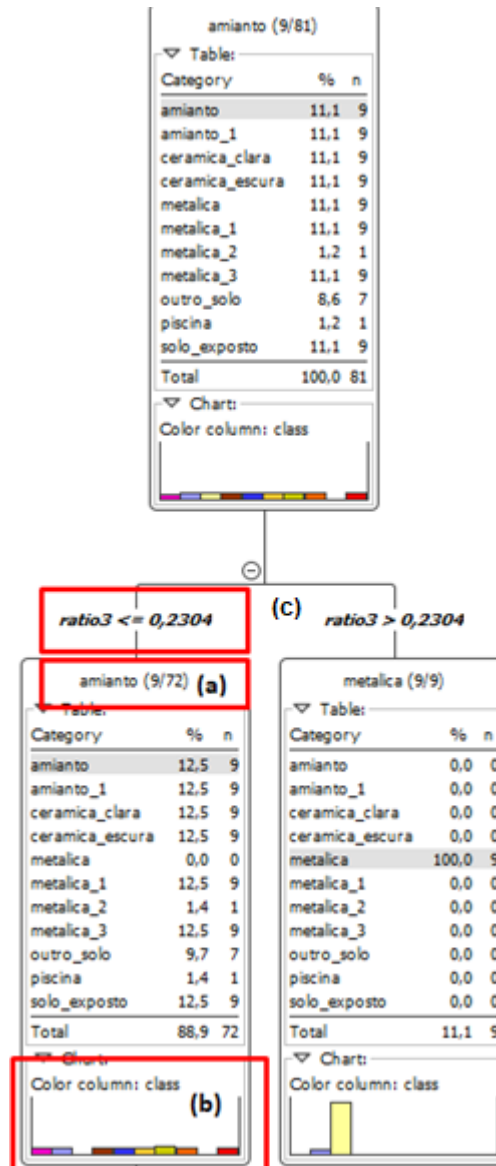
Fonte: Elaborado pelos autores.

- **File Reader** - este nó é utilizado para ler dados de um repositório. Pode ser ler vários formatos e extensões. Para este projeto foi lido em extensão CSV;
- **Color Manager** - define cores para as classes que farão parte da árvore de decisão;
- **Partitioning**: divide a tabela de entrada em duas partições, permitindo criar dados de treinamento e dados de teste;
- **Decision Tree Learner** - este nó induz uma árvore de decisão de classificação na memória principal. Tem como base o algoritmo C45 de Quilan;
- **Decision Tree Predictor** - este nó induz uma árvore de decisão;

- **Decision Tree RoleSet** - converte o modelo de árvore de decisão em tabela contendo as regras de forma textual.

Porção da árvore de decisão induzida pelo KNIME é apresentada na Figura 9.

Figura 9-Árvore de decisão gerada pelo sistema KNIME.



Fonte: Elaborado pelos autores.

Na Figura 9a, a classe (amianto), que representa o nó e a quantidade de amostras (9/72) utilizadas para partição. A Figuras 9b representa as cores de cada classe na árvore e a Figura 9c é o atributo com a regra e valores de

limiar ($\text{ratio} \leq 0,2304$). O valor entre parênteses (Figura 9a), representado por (9/72), indica que as 9 amostras utilizadas na partição são consideradas confiáveis da classe (metálica) no nó e (72) é o número total de amostras usado para este nó.

O sistema minerador de dados SIPINA disponibiliza cinco algoritmos de árvore de decisão: ID3, C45, GID3, ASSISTENT 86 e CHAID. A maioria, à exceção do CHAID, é constituída por extensões ao ID3 de Quilan (1986).

Como o RapidMiner e KNIME, é necessário utilizar diferentes operadores para construir o fluxo de treinamento no minerador de dados Orange Canvas. Os detalhes do funcionamento e integração do sistema minerador de dados Orange Canvas com o InterIMAGE, utilizando a mesma imagem para entrada de dados e configurações, podem ser consultados em Antunes et al. (2016).

O WEKA possui uma biblioteca de algoritmos de aprendizagem de máquina já testada e comparada para classificação do uso do solo, de acordo com os resultados apresentados por Sato et al. (2013). Detalhes do funcionamento e integração do sistema minerador de dados WEKA com o InterIMAGE utilizando a mesma imagem para entrada de dados e configurações neste trabalho podem ser consultados em Antunes et al. (2015).

Não foi possível a inserção de dados gerados pelos mineradores diretamente no InterIMAGE. Todos os limiares gerados nas árvores de decisão pelos diversos mineradores testados neste trabalho foram inseridos individualmente, de forma manual, no InterIMAGE, por meio da interface TopDown Decision Rule.

Na Tabela 2 pode-se observar o desempenho (tempo) das classificações com base em diferentes regras, aplicadas às árvores de decisão de cada classificação. É possível que diferentes mineradores, usando o mesmo algoritmo de classificação, apresentam desempenhos diferentes em razão dos números de regras.

Tabela 2- Diferentes mineradores de dados com diferentes números de regras e tempo da classificação baseada em objeto no InterIMAGE.

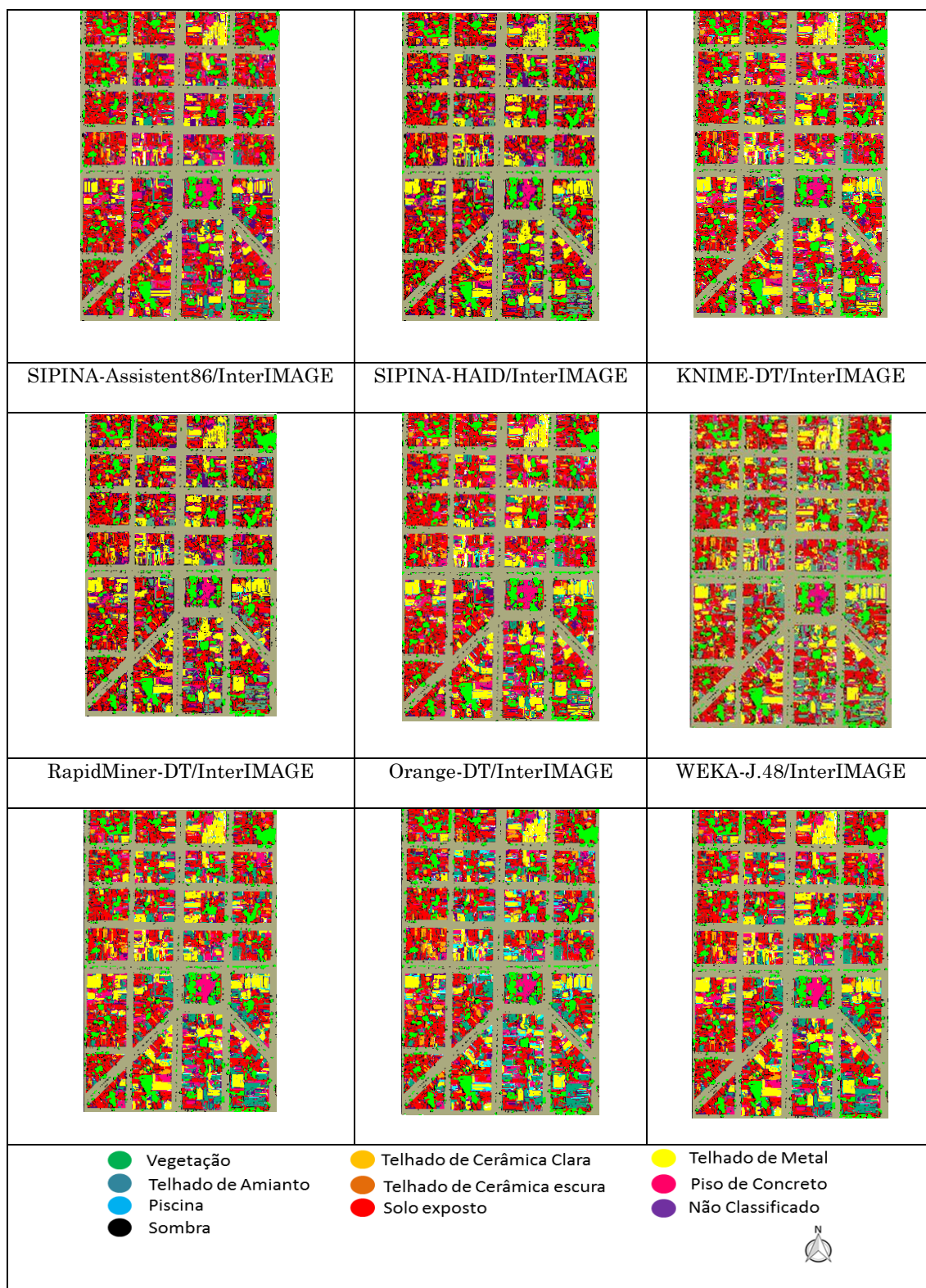
Sistema	Algoritmo	Regras (und.)	Tempo de classificação
SIPINA	Decision Tree (ID3)	16	1018 s
SIPINA	Decision Tree (C45)	14	872 s
SIPINA	Decision Tree (GID3)	16	993 s
SIPINA	Decision Tree (Assistent86)	11	735 s
SIPINA	Decision Tree (CHAID)	20	1150 s
RapidMiner	Decision Tree (C45)	14	872 s
KNIME	Decision Tree (C45)	08	455 s
Orange Canvas	Decision Tree (C45)	15	963 s
WEKA	Decision Tree (J48)	13	900 s

Fonte: Elaborado pelos autores.

Na Figura 10 pode-se visualizar o resultado de cada classificação utilizando os diferentes mineradores de dados.

Figura 10-Resultado da classificação (OBIA). Integração de sistemas classificadores e InterIMAGE. Sistema InterIMAGE.

SIPINA-ID3/InterIMAGE	SIPINA-C45/InterIMAGE	SIPINA-GID3/InterIMAGE
-----------------------	-----------------------	------------------------



Fonte: Elaborado pelos autores.

Foi aplicada a matriz de confusão para uma análise de concordância da classificação final para cada classificação. A avaliação dos resultados da classificação foi feita baseada no método de concordância KAPPA (BIAS et al.,

2014; ANTUNES et al., 2015; PASSO et al., 2013). A Tabela 3 apresenta exemplo da aplicação da Matriz de Confusão com o minerador KNIME.

Tabela 3-Matriz de confusão para a classificação KNIME e InterIMAGE.

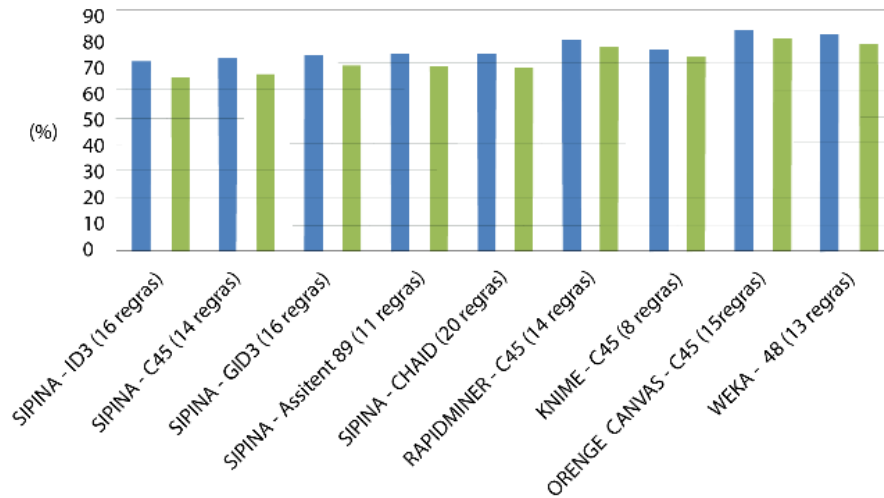
	Amianto	Cerâmica	Metálica	Vias	Solo Exposto	Pavimento de concreto	Sombra	Vegetação	Piscina	Não classificad	Total
Amianto	65				6	5				2	78
Cerâmica		8	14		94						116
Metálica	1		63		1	6				3	74
Vias				181							181
Solo Exposto			4		48						52
Pavimento de concreto	10		4			22				1	37
Sombra	3		3				2			1	9
Vegetação								113			113
Piscina									2	2	4
Não classificado											0
Total	79	8	88	181	149	33	2	113	2	9	-

Fonte: Elaborado pelos autores.

As maiores confusões apresentadas em todas as matrizes apareceram nas classes de telhado de cerâmica e solo exposto, havendo dificuldades na separação, por apresentarem as mesmas características espectrais. Os resultados tiveram a influência do aumento da proporção das classes Vias e Vegetação na classificação. No entanto, apresentaram melhor separação e com pouca confusão.

A porcentagem da exatidão total para cada classificação por meio do índice de exatidão Kappa é apresentado no Gráfico 1.

Gráfico 1-Resultado da classificação (Kappa) do InterIMAGE por meio da integração com os mineradores: SIPINA, RapidMiner Studio, KNIME Analytics Platform, Orange Canvas e WEKA.



Fonte: Elaborado pelos autores.

A quantidade de regras inseridas no InterIMAGE não interferiu no resultado da classificação baseada em objetos. Uma análise (Gráfico 1) pode ser feita por meio do resultado da integração KNIME/InterIMAGE com 8 regras, que atingiu acurácia de 73% para o índice Kappa e o da integração SIPINA (CHAID)/InterIMAGE, com 20 regras, com acurácia 70% para o índice Kappa, ou seja, a integração SIPINA(CHAD)/InterIMAGE utilizando maior quantidade de regras (20 regras) não obteve melhor resultado que KNIME/InterIMAGE com menor quantidade de regras (8 regras). O resultado do teste de significância (teste z) é apresentado na Tabela 4.

Tabela 4-Resultados do índice Kappa e do teste z para cada classificação do InterIMAGE por meio da integração com os mineradores: SIPINA, RapidMiner Studio, Knime Analytics Platform, Orange Canvas e WEKA.

SOFTWARE	SIPINA	SIPINA	SIPINA	SIPINA	SIPINA	RAPID MINER	KNIME	ORANGE CANVAS	WEKA
ALGORITMO	ID3	C45	GID3	Assist86	CHAID	C45	C45	C45	J48
EXATIDÃO GLOBAL (%)	70	72	73	73	73	79	75	83	81
KAPPA	0,66	0,68	0,70	0,70	0,70	0,77	0,73	0,81	0,78
VARIANCIA KAPPA	0,000406	0,000406	0,000382	0,000391	0,000387	0,000344	0,000387	0,000268	0,00032
SIPINA ID3 (16 regras)	calc z 32,25	0,72	1,42	1,42	1,42	4,02	2,48	4,62	4,44
SIPINA C45 (14 regras)	0,72	calc z 35,56	0,74	0,73	0,73	3,40	1,84	5,20	3,84
SIPINA GID3 (16 regras)	1,42	0,74	calc z 35,29	0,0	0,0	2,60	1,08	4,31	3,01
SIPINA Assist86 (11 regras)	1,42	0,73	0,00	calc z 34,85	0,0	2,58	3,00	4,28	2,99
SIPINA CHAID (20 regras)	1,42	0,73	0,00	0,00	calc z 35,05	2,59	1,08	4,29	3,00
RAPIDM C45 (14 regras)	4,02	3,40	2,60	2,58	2,59	calc z 40,97	1,48	1,62	0,39
KNIME C45 (8 regras)	2,48	1,84	1,08	1,07	1,08	1,48	calc z 40,05	3,12	1,87
ORANGE C45 (15 regras)	5,77	5,20	4,31	4,28	4,29	1,62	3,12	calc z 48,78	1,23
WEKA J48 (13 regras)	4,44	3,84	3,01	2,99	3,00	0,39	2,05	1,23	calc z 42,80

Fonte: Elaborado pelos autores.

O teste Z, o qual teve como base os valores e a variância do índice Kappa, permitiu a comparação entre os resultados das diferentes classificações com o InterIMAGE e mineradores de dados. Os valores foram organizados na Tabela 4 cuja diagonal principal equivale ao resultado do Teste Z para um único índice Kappa e os demais valores mostram o resultado da correlação entre dois índices.

Ao analisar a Tabela 4, observa-se que as classificações que resultaram em índice 0,0 (zero), não possuem significância, pois possuem o mesmo índice Kappa, causando assim uma inexistência de dados para o Teste Z. É o caso, por exemplo, da classificação InterIMAGE/SIPINA (GID3) e InterIMAGE/SIPINA (CHAD), ambos com índices 73% Kappa. Como estes resultados de classificação não possuem um resultado de significância, não

faz sentido à comparação entre eles e a correlação não foi considerada nos cálculos.

O valor Z foi “tabelado” ao nível de 5% de probabilidade, igual a 1,96. Levando em consideração o valor Z calculado conforme apresentado na Tabela 4 e maior que 1,96, consideramos que o resultado e diferença entre as classificações foram significativas, concluindo-se que o resultado das classificações é estatisticamente diferente, é o caso dos valores apresentado de cor vermelha na Tabela 4.

A maior significância dos resultados foi entre Orange Canvas/SIPINA (ID3) que atingiu um valor de 4,62 e WEKA/SIPINA (ID3) que atingiu um valor de 4,44. As classificações do InterIMAGE utilizando os algoritmos do SIPINA (ID3, C45, GID3, Assistent 86 e CHAID) tiveram valores de significância abaixo de 1,96, demonstrando que não houve diferença estatística.

6 Conclusão

Os resultados apresentados mostraram a possibilidade de trabalhar com outros mineradores de dados, além do WEKA, para um processo de classificação baseada em objetos, por meio do sistema InterIMAGE. Outros mineradores de código aberto, além do WEKA, utilizados neste trabalho foram: SIPINA, RapidMiner Studio, KNIME Analytics Platform e Orange Canvas.

Todos os sistemas mineradores de dados testados nesta pesquisa foram fáceis de operar, mesmo com interfaces e visualizações diferentes.

O SIPINA não apresentou um bom resultado utilizando seus algoritmos para este trabalho. O melhor índice de acurácia atingido foi 70% Kappa (GID3, Assistent 86 e CHAID) e os resultados das classificações OBIA com seus algoritmos não teve diferença estatística entre eles, de acordo com o teste Z.

Orange Canvas (81% Kappa) e WEKA (78% Kappa) apresentaram os melhores resultados de classificação e índices de significância entre eles foram 1,96, ou seja, não houve diferença estatística, possibilitando ao analista OBIA escolher o mais interessante para seu projeto.

Independentemente de sistema minerador de dados ou algoritmo utilizado, a quantidade de regras influencia o tempo de processamento do InterIMAGE mas não influencia o resultado (acurácia) da classificação baseada em objetos. Sistemas mineradores diferentes, utilizando o mesmo algoritmo (C.45), podem apresentar uma estrutura de árvore diferente devido a vários fatores, como características internas dos softwares de mineração de dados, diferentes amostragens (com diferentes efeitos sobre a acurácia) e outros.

O InterIMAGE demonstrou ser um software eficiente para uma classificação automática de imagem baseada em objetos. Além da limitação do tamanho da imagem (3.000 x 3.000), também se observa a ausência de operadores para análises de vizinhança e de contexto, que são essenciais em OBIA, o que traria grande melhoria para o sistema e conseqüentemente para a classificação.

A falta de atualização do software desde 2014, principalmente por falta de investimento, não trouxe limitações para os processos de classificação, haja vista que todos os operadores continuam funcionando e a versão para Windows 7 (1.43), tem-se apresentado estável até o momento.

Agradecimentos

Os autores reconhecem o apoio prestado pela CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior).

Referências

ADDINK, E. A.; DE JONG, S. M.; PEBESMA, E. J. The importance of scale in object-based mapping of vegetation parameters with hyperspectral imagery.

- Photogrammetric Engineering e Remote Sensing**, v. 73, n. 8, 2007, pp. 905-912.
- AMORIM, R. M.; ALBARICI, F. L.; DEL PINO, M. A. I. T. Mapa digital do escoamento superficial por meio de imagens de sensor remoto na sub-bacia do Rio Moji Guaçu-MG. **Revista Brasileira de Geografia Física**, v. 9, n. 3, 2016, pp. 881-896.
- ANTUNES, R. R.; BIAS, E.S.; BRITES R.S. e COSTA, A.O.P. Desenvolvimento de técnica para monitoramento do cadastro urbano baseado na classificação orientada a objetos. Estudo de caso: Município de Goianésia, Goiás. **Revista Brasileira de Cartografia**, v. 67, n. 2, Brasília, 2015, pp. 357-372.
- ANTUNES, R. R.; BIAS, E.S.; BRITES R.S. e COSTA, A.O.P. Integration of open-source tools for object-based monitoring of urban targets. In: **GEOBIA 2016: Solutions and synergies**. University of Twente Faculty of Geo-Information and Earth Observation (ITC). 2016.
- BAATZ, M.; SCHAPE, A. Object-oriented and multi-scale image analysis in semantic networks. **The 2nd International Symposium: Operationalization of Remote Sensing, New Methodologies**, 16-20 August 1999.
- BIAS, E.S.; ANTUNES, R. R.; BRITES R.S. e COSTA, A.O.P. Application of imagery analysis based on objects as a tool for monitoring the urban cadaster in small municipalities. **South Eastern European Journal of Earth Observation and Geomatics**. Greece: Aristotle University of Thessaloniki, 2014.
- BLASCHKE, T. What's wrong with pixels? Some recent developments interfacing remote sensing and GIS. **GeoBIT/GIS**, v. 6, 2001, pp. 12-17.
- BLASCHKE, T. **Object-based contextual image classification built on image segmentation. Advances in techniques for analysis of remotely sensed data**. IEEE Workshop on, 2003, pp. 113-119.
- BLASCHKE, T. Object-based image analysis: a new paradigm remote sensing? **ASPRS, Anual Conference**. Baltimore, Maryland, 2013.
- BLASCHKE, T.; BURNETT, C.; PEKKARINEN, A. Image segmentation methods for object-based analysis and classification. In: **Remote sensing image analysis: Including the spatial domain**. Netherlands, 2004, pp. 211-236.
- BRACHMAN, R. J. What IS-A Is and isn't: an analysis of taxonomic links in semantic networks. **Computer**, v. 16, 1983, pp. 30-36.

- BREIMAN, L.; Friedman, J. Stone, C; e Olshen, R.A. **Classification and regression trees**. CRC Press, 1984.
- CADENA, G. T. **Classificação dos tipos de pavimentos das vias urbanas a partir de imagem de alta resolução espacial por meio de análise orientada a objeto**. [Dissertação de mestrado em Ciências Cartográficas, Universidade Estadual Paulista, p. 114]. Presidente Prudente, SP, 2011.
- CAMARGO, F.F.; FLORENZANO, T.G.; ALMEIDA, C.M.; OLIVEIRA, C.G. **Geomorphological Mapping Using Object-Based Analysis and ASTER DEM in the Paraíba do Sul Valley, Brazil**. *International Journal of Remote Sensing*, v. 30, n. 24, 2009, pp. 6613-6620.
- CARVALHO, M.V.A.; KUX, H.J.H., FLORENZANO, T.G. Análise de imagem baseada em objeto e mineração de dados aplicadas à classificação do uso do solo urbano por quadra em imagens WorldView-2. Anais XVI Simpósio Brasileiro de Sensoriamento Remoto - SBSR, Foz do Iguaçu, PR, Brasil, abril de 2013, INPE.
- CASTILLA, G.; HAY, G J. Image objects and geographic objects. In: BLASCHKE, T.; LANG, S.; HAY, G.J. (Eds.) *Object-Based Image Analysis Spatial Concepts for Knowledge-Driven Remote Sensing Applications*. Berlin: Springer, 2008. pp. 91-110.
- CESTNIK, B.; KONONENKO, I.; BRATKO, I. ASSISTANT 86: a knowledge elicitation tool for sophisticated users. **Proc. of the 2nd European Working Session on Learning**, 1987, pp. 31-45.
- COHEN, J. A. Coefficient of agreement for nominal scales. *Educational and measurement*, v. XX, n. 1, 1960, pp. 37-46.
- CONGALTON, R. G.; GREEN, K. *Assessing the accuracy of remotely sensed data: principles and practices*. Boca Raton-USA: Lewis Publisher, 1999.
- COSTA, G. A. O. P.; FEITOSA, R.Q.; FONSECA, L.M.G.; OLIVEIRA, D.A.B. FERREIRA, R.S.; e CASTEJON, R.F. Knowledge-based interpretation of remote sensing data with the InterIMAGE System: major characteristics and recent developments. **GEOBIA 2010**. Gent, Belgium, 2010.
- DE GRANDE, T. O.; DE ALMEIDA, T.; CICERELLI, R. E. Classificação orientada a objeto em associação às ferramentas reflectância acumulada e mineração de dados. *Pesquisa Agropecuária Brasileira*, v. 51, n. 12, 2017, pp. 1983-1991.

- DEMSAR, J.; CURK, T.; DE SOUSA, G. M., FERNANDES, M. C., e COSTA, G. A. O. P. (2015). Classificação da susceptibilidade a ocorrência de incêndios através de mineração de dados e GEOBIA. *Revista Brasileira de Cartografia*, 67(3).
- ERJAYEC, A. Orange: data mining toolbox in Python; *Journal of Machine Learning Research*, Aug/2013, pp. 2349-2353.
- DORREN, L. K. A.; MAIER, B.; SEIJMONSBERGEN, A. C. Improved landsat-based forest mapping in steep mountainous terrain using object-based classification. **Forest Ecology and Management**, v. 183. Elsevier, Amsterdam, 2003, pp. 31-46.
- DOS ANJOS, C.S.; ALMEIDA, C.S.; GALVÃO, L.S.; FILHO, C.R.S.; LACERDA, M.C.; e Prati, R.C. Análise do nível de legenda de classificação de áreas urbanas empregando imagens multiespectrais e hiperespectrais com os métodos árvore de decisão c4.5 e floresta randômica. *Boletim de Ciências Geodésicas*, v. 23, n. 2, 2017, Curitiba, Brasil.
- FAULHABER, A. **AI Tools**: a short introduction to Yale. Department of Mathematics and Computer Science. Saarbrücken, Germany: Saarland University, 2007.
- FAYYAD, U. **Branching on attribute values in decision tree generation**. AI Group, M/S 5253660, Jet Propulsion Laboratory, California Institute of Technology. Pasadena, CA. 1994.
- FERREIRA, R. S. **InterIMAGE cloud platform**: the architecture of a distributed platform for automatic, object-based image interpretation. [Tese de doutorado em engenharia elétrica pela Pontifícia Universidade Católica do Rio de Janeiro – PUC-Rio]. Rio de Janeiro, 2015.
- FERREIRA, R. S.; BENTES, C.; COSTA, G.O.A.P.; OLIVEIRA, D.A.B. A set of methods to support object-based distributed analysis of large volumes of Earth observation data. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (JSTARS)**, v. 10, n. 2, 2017, pp. 681-690.
- FURTADO, L. F. A., FRANCISCO, C. N., ALMEIDA, C. M. Análise de Imagem Baseada em Objeto para Classificação das Fisionomias da Vegetação em Imagens de Alta Resolução. *Geociências*, v. 32, n.3, 2013, pp.441-451.

- GHOLAP, J. **Performance tuning of j48 algorithm for prediction of soil fertility**. Pune, Maharashtra, India: Dept. of Computer Engineering. College of Engineering, 2012.
- GRIPPA, T.; LENNERT, M.; e BEAUMONT, B. An open-source semi-automated processing chain for urban OBIA classification. In: **GEOBIA 2016: Solutions and Synergies**. 14 September 2016. University of Twente Faculty of Geo-Information and Earth Observation (ITC).
- HAY, G.J., Castilla, G., 2008. Geographic Object-Based Image Analysis (GEOBIA): A new name for a new discipline. In: Blaschke, T., Lang, S., Hay, G. (Eds.), *Object Based Image Analysis*. Springer, Heidelberg, Berlin, New York, pp. 93_112.
- HSSINA, B., MERB OUHA, A., EZZIKOURI, H. ERRITALI, M. A. A comparative study of decision tree ID3 and C4.5. (IJACSA) International Journal of Advanced Computer Science and Applications, Special Issue on Advances in Vehicular Ad Hoc Networking and Applications. TIAD laboratory, Computer Sciences Department, Faculty of sciences and techniques. Sultan Moulay Slimane University, Morocco, 2014.
- INTERIMAGE WIKI. **Attributes description**. Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio) e Instituto Nacional de Pesquisa Espacial (INPE). 2014. Site: http://wiki.dpi.inpe.br/doku.php?id=interimage:attributes_description Acessado em junho de 2017.
- INTERIMAGE. **Manual do usuário**. 2010. Site: <http://www.lvc.ele.puc-rio.br/projects/interimage/pt-br/documentacao/> Acessado em abril de 2016.
- KASS, G. An exploratory technique for investigating large quantities of categorical data. **Applied Statistics**, v. 29, n. 2, 1980, pp. 119-127.
- KAWASHIMAET, R. S., SALIM, A., COSTA, W. F., SILVA, J. R., COUTINHO FILHO, O. S., e QUINTANILHA, J. A. **Classificação de imagem baseada em objeto (OBIA) utilizando índices de vegetação**. Anais XVI Simpósio Brasileiro de Sensoriamento Remoto - SBSR, Foz do Iguaçu, PR, Brasil, 13 a 18 de abril de 2013, INPE.
- KAUR, A.; SINGH, S. Classification and selection of best saving service for potential investors using decision tree – Data mining algorithms. **International**

Journal of Engineering and Advanced Technology (IJEAT), v. 2, n. 4, April 2013.

KNIME ANALYTICS PLATFORM. Open for innovation. 2017. Switzerland. Site: <<https://www.knime.org/>> Acessado em junho de 2017.

NASCIMENTO, A., RUBIM, I. B., PEREIRA, E. G. S., BARROS, R. S., e RICHTER, M. (2013). Classificação da cobertura da terra, utilizando os programas livres: InterImage, WEKA e QuantumGIS. Anais XVI Simpósio Brasileiro de Sensoriamento Remoto - SBSR, Foz do Iguaçu, PR, Brasil, 13 a 18 de abril de 2013, INPE.

MAVRANTZA, O.; ARGIALAS, D. An object-oriented image analysis approach for the identification of geologic lineaments in a sedimentary geotectonic environment. In: BLASCHKE, T.; LANG, S.; HAY, G. J. **Object-based image analysis**. Berlin, German. Springer, 2008.

MEINEL, G.; NEUBERT, M. A comparison of segmentation programs for high resolution remote sensing data. Commission VI in **Proceeding of XXth ISPRS Congress**. International Society for Photogrammetry and Remote Sensing. Istanbul, Turkey, 2004.

MENESES, P. R.; SANO, E. Classificação pixel a pixel de imagens. In: MENESES, P. R.; ALMEIDA, T. **Introdução ao processamento de imagens de sensoriamento remoto**. 1. ed., v. 1. Brasília: CNPq, 2012.

MENEGHETTI, G. T. e Kux, H. J. H. (2014). Mapeamento da cobertura da terra do município de Raposa (Ma) utilizando imagens Worldview-II, o aplicativo InterIMAGE e mineração de dados. *Revista Brasileira de Cartografia*, 1(66/2).

MORGAN, J. N.; SONQUIST, J. A. Problems in the analysis of survey data, and a proposal. **Journal of the American Statistical Association**, n. 58, 1963, pp. 415-434.

NAVULUR, K. Multispectral image analysis using the object-oriented paradigm. Boca Raton: **Taylor e Francis Group**, 2006. 165 p. 31.

OLIVEIRA, F. P. D.; FILHO, F.E.I.; SOUSA, A.L. e SOARES, V.P. Mapeamento de florestas monodominadas por Myracrodruon urundeuva com imagens TM-Landsat 5 e Rapideye. **Floresta e Ambiente**, v. 22, n. 3, 2015, pp. 322-333.

ORLANDO, Pietro; LA ROSA, Emanuele. Object oriented methodology for change detection technique: the case of Scopello-Sicily. **South-Eastern European**

- Journal of Earth Observation and Geomatics.** Grécia: Aristotle University of Thessaloniki, 2014.
- OTUKEI, J. R.; BLASCHKE, T. Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms. **International Journal of Applied Earth Observation and Geoinformation.** Elsevier, 2010.
- PASSO, D. P.,BIAS, E.S.;BRITES, R.S. e COSTA, G.A.O.P. Uso do sistema InterIMAGE para a identificação de alvos urbanos em imagens do satélite Worldview II. **Revista Brasileira de Cartografia**, v. 6, 2013.
- QGIS BRASIL. **Comunidade de usuários QGIS Brasil.** 2015. Disponível em: <<http://qgisbrasil.org/>> Acesso em: 14 abr. 2017.
- QUILAN, J. **Induction of decision trees**, v. 1. Boston (USA): Machine Learning, Kluwer Academic Publishers, 1986, pp. 81-106.
- RAKOTOMALALA, R. **Introduction of a decision tree using SIPINA.** Tutorial. Departamento de Informática e Estatística. University Lyon. France, 2008.
- RAPIDMINER STUDIO. **Manual.** Boston, USA. 2014. Site: <<http://docs.RapidMiner.com/downloads/RapidMiner-v6-user-manual.pdf>> Acessado em julho de 2017.
- RICH, E. **Inteligência artificial.** São Paulo: McGraw-Hill, 1988.
- SATO, L. Y.; SHIMABUKURO, Y.E. KUPLIC, T.M.; e GOMES, V.C.F Análise comparativa de algoritmos de árvore de decisão do sistema WEKA para classificação do uso e cobertura da terra. **XVI Simpósio Brasileiro de Sensoriamento Remoto**, 2013, pp. 2353-2360.
- SOUZA, U. D. V., e KUX, H. J. H. Geobias e mineração de dados na classificação da cobertura do solo urbano em São Luís (MA) com imagens Worldview-2 e o sistema InterIMAGE. **Revista Brasileira de Cartografia** (2014). N. 66/3: pp. 433-445.
- SOTHE, C.; SCHIMALSKI, M.B.; LIESENBERG, V.; ALMEIDA, C.M.; SOUZA, C.F. E SOUZA, J.B. Técnicas de mineração de dados aplicadas à classificação do estágio sucessional da vegetação em áreas de floresta ombrófila mista. **Scientia Forestalis**, volume 44, n. 112, 2016, pp. 863-873.
- WILGES, B.; MATEUS, G.; NASSAR, S.; BASTOS, R. Avaliação da aprendizagem por meio de lógica de fuzzy validado por uma Árvore de Decisão ID3. **Novas**

Tecnologias na Educação, v. 8, n. 3. Centro interdisciplinar de novas tecnologias na educação – CINTED. Universidade Federal do Rio Grande do Sul – UFRGS. Dez./2010.