

**Classificação do Uso da Terra e Cobertura Vegetal  
Utilizando Técnicas de Mineração de Dados***Land-use and land-cover classification using data mining techniques***Rodrigo Cesário Justino**

Universidade Federal de Alfenas

rodrigofriend@bol.com.br

**Marcos Roberto Martines**

Universidade Federal de São Carlos

mmartines@ufscar.br

**Fernando Shinji Kawakubo**

Universidade de São Paulo

fsk@usp.com

Recebido (*Received*): 16/11/2016  
DOI: 10.11606/rdg.v33i0.122888Aceito (*Accepted*): 21/07/2017

**Resumo:** O objetivo deste estudo é apresentar o mapeamento do uso da Terra e cobertura vegetal utilizando técnicas de mineração de dados. A área teste selecionada corresponde à bacia do Rio São Tomé, localizada na região de Alfenas, sul do estado de Minas Gerais. Para a realização deste estudo, foram utilizadas imagens multiespectrais geradas pelo sensor *Linear Imaging Self-Scanner* (LISS III) a bordo do satélite *Indian Resource Satellite* (IRS). A classificação das imagens foi feita utilizando o aplicativo *Geographic Data Mining Analyst* (GeoDMA) que possui algoritmos para segmentação de imagens, extração de atributos, seleção de feições e classificação. A análise dos resultados foi feita de maneira visual e quantitativa por meio de uma matriz de confusão. Para a geração desta matriz, 3000 pontos de referência foram coletados de maneira totalmente aleatória. Os resultados desta pesquisa mostraram o grande potencial da técnica de mineração de dados na classificação do uso da Terra e cobertura vegetal. As principais vantagens observadas de sua utilização foram: (1) permitiu incorporar no processo de classificação um grande número de variáveis tanto espectrais, espaciais quanto atributos do relevo; (2) simplificou a interpretação das regras de classificação por meio dos resultados das árvores de decisão; (3) gerou uma classificação consistente.

**Abstract:** This study aims at mapping the land-use and land-cover (LULC) types using data mining (DM) techniques. The area selected for study is the São Tomé river watershed, located in the region of Alfenas, south of the State of Minas Gerais. We performed this study using multispectral imagery taken by the Linear Imaging Self-Scanner (LISS III) instrument, onboard the Indian Resource Satellite (IRS). The image classification was carried out using the Geographic Data Mining Analyst (GeoDMA), a toolbox that has algorithms for image segmentation, feature extraction, feature selection and classification. The analysis of the results was made by visual fashion, and quantitatively, by means of an error matrix generation. To this end, 3000 reference points was collected in a totally random manner. The results found in this study highlighted the great potential of DM for classifying the LULC types. Some of the key advantages noted in the use of DM include: the possibility of incorporating in the classification process a large number of variables, which include both spectral and spatial attributes, and relief information; it simplified the image interpretation through a decision tree outcome, and; it generated a consistent classification result.

**Keywords:** Remote Sensing; GeoDMA; Segmentation; Decision Tree.

**Palavras-chave:** Sensoriamento Remoto; GeoDMA; Segmentação; Árvore de Decisão.

## 1. INTRODUÇÃO

Uma das principais aplicações do sensoriamento remoto voltadas para o planejamento agrícola e ambiental é o mapeamento do uso da Terra e cobertura vegetal (MUTCV). Tradicionalmente, o MUTCV é feito de maneira visual utilizando os elementos de reconhecimento da fotogrametria interpretativa, como cor, textura, forma, contexto, etc. (CERON; DINIZ, 1966). Apesar dos produtos cartográficos gerados por meio da fotointerpretação apresentarem alto nível de fidedignidade, tal metodologia apresenta alguns empecilhos, pois dependendo do nível de detalhe do mapeamento, a técnica adota critérios subjetivos no processo de decisão, o que dificulta a sistematização de regras. Além disso, os mapeamentos realizados por meio da fotointerpretação são muitas vezes demorados e custosos.

Em razão das dificuldades mencionadas, a utilização de técnicas computacionais como forma de dinamizar o mapeamento é crucial. As técnicas de classificação digital de imagens de satélite têm se sofisticado ao longo dos anos com significativas melhorias no processo de reconhecimento de padrões. Uma das abordagens mais promissoras na atualidade é a Classificação de Imagens Orientada a Objeto (CIOO). A CIOO tem como unidade de análise grupos de pixels delimitados por segmentos - ao invés de pixels isolados como na classificação estatística convencional. Esses segmentos geralmente incorporam um amplo conjunto de informações que incluem atributos espectrais, texturais e espaciais (BLASCHKE, 2010) que são usados para treinar ou construir uma base de conhecimento para a classificação de imagens.

A principal característica da classificação baseada no conhecimento é que tal modelo incorpora a complexidade do raciocínio cognitivo herdada da fotointerpretação, porém, materializada na forma computacional. Assim, os mesmos elementos de reconhecimento utilizados na fotointerpretação (cor, textura, forma e contexto) podem ser utilizados para orientar a classificação de determinado tipo de uso ou cobertura vegetal. Os critérios adotados são bastante variados e dependem do conhecimento do especialista e do nível de complexidade da classe mapeada.

Diante do exposto, Vieira *et al.* (2012) afirmam que uma das etapas mais importantes da classificação baseada no conhecimento é a construção do modelo de interpretação de imagem (“conhecimento”). Neste modelo, o analista deve ter uma exata noção dos melhores atributos a serem utilizados na classificação. Apesar do grande potencial, o modelo baseado no conhecimento também não apresenta apenas vantagens, pois definir qual é o melhor atributo para separar uma determinada classe nem sempre é uma tarefa trivial de ser executada. A dificuldade se torna ainda maior quando um número bastante grande de variáveis é considerado no processo de classificação.

Uma alternativa viável e promissora para geração automática de uma estrutura de conhecimento é a adoção de técnicas de mineração de dados (SILVA *et al.*, 2008; VIEIRA *et al.*, 2012). Na visão de Fayyad *et al.* (1996), a mineração de dados refere-se a uma etapa particular de um processo maior conhecido como “Descoberta do Conhecimento em Banco de Dados”, ou simplesmente KDD (*Knowledge Discovery in Databases*). O KDD tem como objetivo identificar e entender os padrões dos dados potencialmente úteis que podem ser utilizados para resolver determinados problemas, como por exemplo, selecionar os principais elementos a serem empregados no mapeamento das áreas cobertas por florestas nativas. Para isto, o KDD engloba várias etapas desde a preparação dos dados que serão analisados, passando pela mineração (aplicação de análise e de algoritmos voltados para a descoberta do conhecimento), avaliação e refinamento do “conhecimento” (FAYYAD *et al.*, 1996).

A adoção de técnicas de mineração de dados em classificação de imagens de satélite ainda é relativamente pouco disseminada e necessita de estudos mais aprofundados como forma de avaliar as suas potencialidades e limitações. Assim, o presente trabalho tem como objetivo apresentar a utilização da técnica de mineração de dados aplicada ao MUTCV em um estudo de caso localizado no sul de Minas Gerais. Atributos espectrais, texturas, espaciais e topográficos serão explorados no processo de classificação. O resultado do mapeamento será importante tanto para analisar o desempenho da técnica empregada quanto para caracterizar o arranjo do uso da Terra e cobertura vegetal da bacia hidrográfica selecionada.

## 2. ÁREA DE ESTUDO

A área de estudo corresponde à bacia do rio São Tomé localizada no sul de Minas Gerais. O rio São Tomé constitui um importante afluente do reservatório de Furnas, drenando uma área de aproximadamente 34 mil hectares, compreendendo parte dos municípios de Machado, Serrania e Alfenas (SILVA *et al.*, 2011). A nascente do rio São Tomé localiza-se no município de Machado a cerca de 1200 metros de altitude. A geomorfologia da região está incluída na unidade morfoclimática do domínio das regiões serranas de morros mamelonares caracterizadas por morros arredondados em formato de meia laranja (AB' SABER, 2003). O clima da região é subtropical classificado como Cwa, segundo Köppen (PEEL *et al.*, 2007), com verão úmido e inverno seco. A vegetação típica é formada por floresta estacional semidecidual ombrófila mista e cerrado (RADAMBRASIL, 1983). Com relação ao uso da Terra, a região destaca-se pelo plantio do café e atividade pecuária. Mais recentemente, observa-se também o avanço crescente na região do plantio da cana-de-açúcar.

## 3. MATERIAIS E MÉTODOS

Para a realização do presente estudo, quatro etapas principais foram adotadas. Inicialmente, foi gerado um Modelo Digital do Terreno (MDT) com o objetivo de incorporar as informações do relevo no processo de classificação. Em seguida, iniciou-se a etapa de pré-processamento das imagens multiespectrais do satélite *Indian Resource Satellite (IRS-P6 ou Resourcesat-1)* visando corrigir as distorções geométricas presentes nas imagens e compatibiliza-las com os produtos derivados do MDT. Finalizada esta etapa, iniciou-se a segmentação das imagens e o MUTCV utilizando as técnicas de mineração de dados. Por fim, os resultados do mapeamento foram analisados por meio da geração de uma matriz de erro. Detalhes dos procedimentos adotados são descritos a seguir.

### 3.1. Derivações de Informações do Terreno

Foram utilizadas como base cartográfica quatro cartas topográficas na escala 1:50 000 publicadas pelo Instituto Brasileiro de Geografia e Estatística (IBGE), folhas: Areado (SF-23-V-D-I-4), Alfenas (SF-23-I-I-3), Campestre (SF-23-V-D-IV-2) e Machado (SF-23-I-III-1). Estas cartas encontram-se em extensão DGN (um formato do *software MicroStation* da empresa *Bentley Systems*). Inicialmente, as informações planimétricas (drenagem e sistema viário) e altimétricas (curvas de nível e pontos cotados) foram organizadas em forma de planos de informações (PI). Para cada PI, foram realizadas as devidas correções com relação a consistência topológica dos segmentos, “limpeza” das feições e agregação de atributos.

A partir dos PIs de curvas de nível e pontos cotados, foi gerado um Modelo Digital do Terreno (MDT), que consiste numa representação matemática contínua que descreve as variações relacionadas à altitude do terreno. No presente estudo, optou-se pela geração do MDT utilizando um interpolador linear que se baseia na distância de *Borgefors* (GORTE; KOOLHOVEN, 1990). Após a construção do MDT, os seguintes produtos cartográficos foram derivados: hipsometria, declividade, orientação das vertentes e hierarquia de sub-bacias hidrográficas seguindo a metodologia de Strahler (1952).

### 3.2. Imagens de Satélite e Pré-Processamento

As imagens do sensor *Linear Imaging Self-Scanner (LISS III)* do satélite *Indian Resource Satellite (IRS-P6 ou Resourcesat-1)* foram baixadas do sítio do Instituto Nacional de Pesquisas Espaciais (INPE). O LISS III é um sensor do tipo *Pushbroom* que possui quatro faixas espectrais operando nos intervalos do comprimento de onda ( $\lambda$ ) do verde (0.52-0.59  $\mu\text{m}$  (b2)), vermelho (0.62-0.68  $\mu\text{m}$  (b3)), infravermelho próximo (0.77-0.86  $\mu\text{m}$  (b4)) e infravermelho médio (1.55-1.70  $\mu\text{m}$  (b5)). As imagens possuem resolução espacial de 23,5 metros no terreno quantificada em 7 bits e com tempo de revisita de 24 dias (CHANDER *et al.*, 2008).

A cena utilizada foi adquirida no dia 31 de julho de 2010 correspondendo à época do período seco. Esta época foi escolhida com o objetivo de evitar possíveis erros de classificações com a presença de outros cultivos, como pasto verde e milho (PRADO *et al.*, 2016). Ambas as classes mencionadas se fazem presentes na época chuvosa, e experiências anteriores com classificações realizadas nesse período mostraram confusões frequentes envolvendo principalmente as classes de cana-de-açúcar, milho, pasto verde e café em estágio de formação.

Com relação aos procedimentos operacionais, a primeira etapa adotada no tratamento das imagens consistiu na correção geométrica das imagens (CRÓSTA, 2002). A correção geométrica foi realizada por meio da identificação de pontos de controle na imagem utilizando como referência a base cartográfica definida no projeto. Foram selecionados no total 15 pontos de controle distribuídos ao longo da área de estudo. Utilizando a transformação afim, o erro de georreferenciamento foi inferior a 1,0 pixel. Em seguida, as imagens foram reamostradas com espaçamento da grade de 20 metros utilizando o interpolador espacial vizinho mais próximo.

É importante mencionar que pelo fato das imagens terem sido adquiridas em condições atmosféricas de céu limpo, nenhum procedimento relacionado a correção atmosférica foi adotado no presente estudo.

### 3.3. Segmentação de Imagens e Mineração de Dados

Esta etapa foi realizada utilizando o aplicativo *Geographic Data Mining Analyst* (GeoDMA), desenvolvido por Korting *et al.* (2009; 2013). O GeoDMA integra métodos de análise de imagens com técnicas de mineração de dados funcionando como um *plugin* do TerraView.

Para realizar o particionamento das imagens, adotou-se o algoritmo de segmentação multiresolução (BAATZ *et al.*, 2004). De acordo com Benz *et al.* (2004), a segmentação por multiresolução é uma técnica de fusão de regiões do tipo *bottom up* que se inicia com um pixel-objeto. Nos passos seguintes, objetos menores são fundidos com objetos maiores em uma sequência de etapas que visa minimizar a heterogeneidade do tamanho dos segmentos gerados, garantindo assim uma distribuição espacial regular dos objetos. No GeoDMA, três parâmetros são definidos: a cor, a compacidade e a escala. Enquanto a cor e a compacidade medem a heterogeneidade dos objetos, a escala destina-se ao controle do tamanho dos objetos.

Finalizada a etapa de segmentação, iniciou-se a fase de extração dos atributos. Além das bandas multiespectrais, também foram incluídas na análise as informações derivadas do MDT bem como as propriedades espaciais dos segmentos.

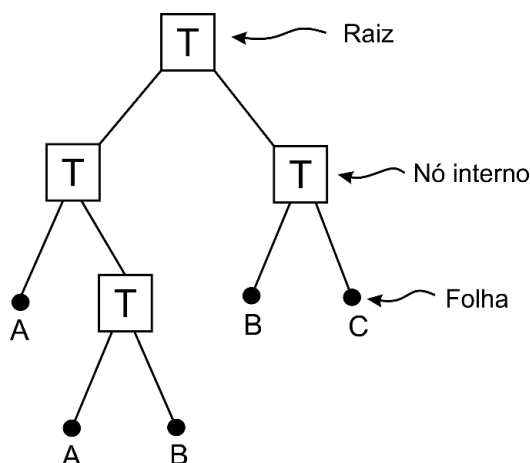
As informações topográficas (hipsometria, declividade, orientação das vertentes e hierarquia de sub-bacias hidrográficas) foram consideradas porque observações em campo constataram um forte controle do relevo no ordenamento territorial da bacia. Já as formas dos segmentos são importantes porque permitem diferenciar principalmente as classes ditas naturais das classes antrópicas. Geralmente, as classes naturais, como as manchas de mata, se distribuem na paisagem de formas irregular, enquanto as plantações de eucalipto, apresentam formas mais geométricas. Existem casos, porém, que esta regra não é válida. Manchas de eucalipto abandonado, por exemplo, podem apresentar formas mais irregulares que se assemelham as áreas de mata. Para este caso, outros atributos (espectrais ou texturais) devem ser incorporados no processo de classificação para permitir a diferenciação entre as classes mencionadas.

Foram exploradas, no total, 19 métricas: amplitude, dissimilaridade, entropia, homogeneidade, média, moda, desvio padrão, ângulo, área, área caixa, círculo, ajuste elíptico, dimensão fractal, raio, comprimento, perímetro, razão entre perímetro e área, ajuste retangular e largura. Detalhes das descrições das métricas descritas podem ser obtidos em Korting *et al.* (2013).

Após a extração dos atributos, iniciou-se a fase de amostragem. Foram selecionadas amostras de 7 classes de uso da Terra e cobertura vegetal: Corpo d'Água, Café Formação (café na fase inicial de desenvolvimento ou em recuperação da folhagem após o manejo da poda), Café Produção (café na

fase produtiva ou que foi submetido recentemente ao manejo da poda), Cana-de-Açúcar, Mata, Pastagem e Solo Exposto. Para a classificação, utilizou-se o algoritmo C4.5 desenvolvido por Quinlan (1993) que gera como resultado uma árvore de decisão (AD).

A AD consiste em uma hierarquia de nós internos e folhas conectadas por ramos. Cada nó interno corresponde uma variável que é utilizada para a classificação. O primeiro nó interno é conhecido como raiz e os demais são chamados de intermediários. A rotulação das classes ocorre no nível das folhas. A **Figura 1** ilustra um modelo esquemático de AD binária. O processo de subdivisão inicia-se pelo nó raiz e se estende até as folhas, sendo que a quantidade de nós internos irá depender da complexidade das classes mapeadas.



**Figura 1:** Modelo de classificação por AD. Cada caixa corresponde a um nó onde é aplicado um determinado teste representado por “T” para a rotulação da classe ou geração de novas subdivisões com os respectivos nós internos. Fonte: adaptado de Friedl; Brodley (1997).

Para a escolha dos melhores atributos a serem utilizados na classificação, utilizou-se uma métrica (“T”) que se baseia na entropia (uma medida de textura) e no ganho de informação. Esta métrica tem como objetivo minimizar o grau de impureza dos nós. Ou seja, a impureza é nula em um determinado nó da árvore se todos os exemplos nele pertencerem a mesma classe. Analogamente, o grau de impureza é máximo no nó se houver o mesmo número de exemplos para cada classe possível (ZUBEN, ATTUX, 2010). Para determinar a eficácia da condição de teste realizada, é necessário comparar o grau de entropia do nó-pai (antes da divisão) com o grau de entropia dos nós-filhos (após a divisão). O atributo que gerar uma maior diferença de ganho de informação é escolhido como condição de teste (ibidem).

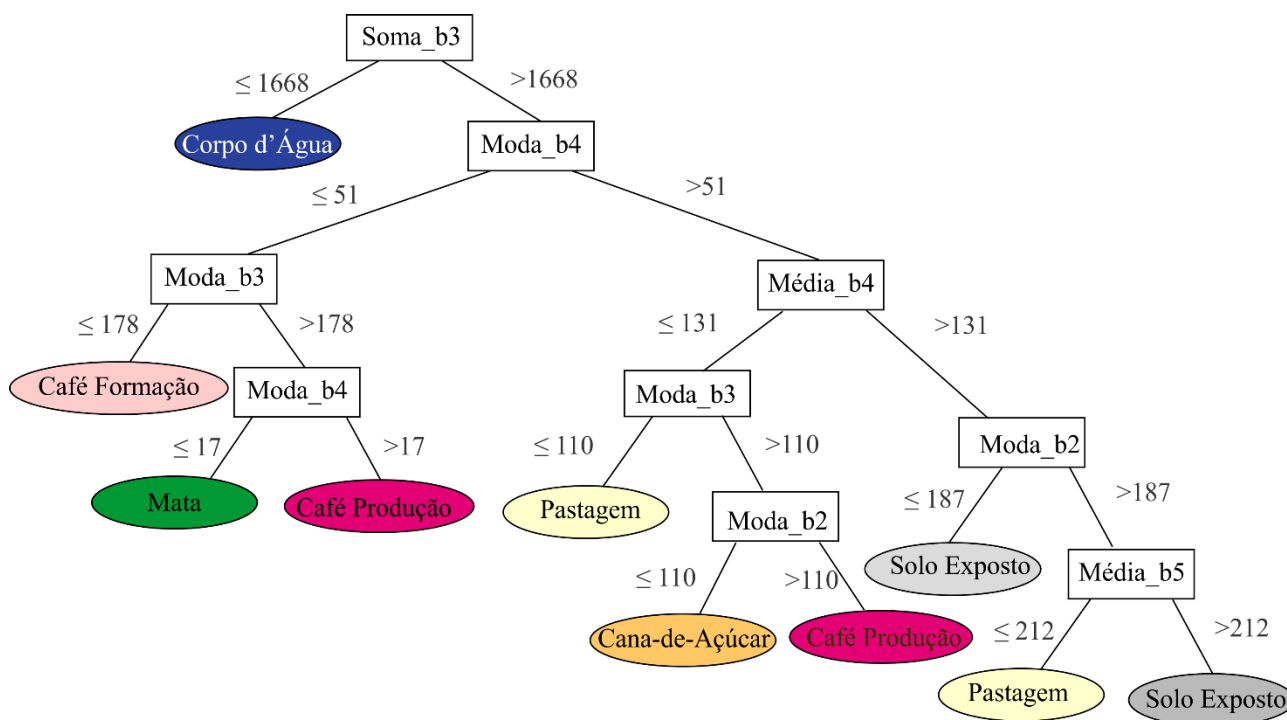
### 3.4. Análise de Exatidão

A análise da exatidão da classificação foi feita por meio de uma Matriz de Confusão (MC). Na MC, os resultados da classificação são confrontados com as amostras de referência. Os valores distribuídos na diagonal principal desta matriz correspondem aos acertos obtidos no mapeamento. Já os valores dispostos fora da diagonal principal correspondem às confusões de classificação que podem estar associadas à inclusão ou exclusão de classes (CONGALTON, 1991). O erro inclusão (também referido como erro de comissão) é responsável pela superestimativa de uma determinada classe “A”, pois pixels pertencentes a uma classe “B” foi erroneamente rotulado como sendo “A”. Por sua vez, o erro de exclusão (também conhecido como erro de omissão) é responsável pela subestimativa do mapeamento, pois pixels que deveriam ter sido classificados como “A” foram erroneamente atribuídos a “B”.

Para a construção da MC foram coletados 3000 pontos de referência distribuídos aleatoriamente na área de estudo. A checagem destes pontos foi feita com informações obtidas em trabalhos de campo e com auxílio das imagens de alta resolução do *Google Earth*. Foram extraídas porcentagens de acertos globais (proporção de pontos dispostos na diagonal principal da matriz) e individuais, bem como os respectivos erros associados com o objetivo de avaliar as causas das confusões. Também foi utilizado o índice *kappa* como referência de qualidade do mapeamento. O índice *kappa* é considerado mais confiável que a exatidão global por considerar em um único índice todos os valores da matriz de confusão (CONGALTON; GREEN, 1993).

#### 4. RESULTADOS E DISCUSSÃO

Após vários testes de segmentação, adotou-se como parâmetro de segmentação os seguintes valores: 40 (compacidade), 30 (escala) e 70 (cor), gerando no total 2499 objetos. A AD utilizada para a classificação é ilustrada na **Figura 2**. Conforme pode ser observado, todas as bandas foram utilizadas na construção da árvore, o que demonstra a importância da complementaridade das informações extraídas das imagens IRS LISS III no MUTCV. Apenas três métricas foram utilizadas: a soma, a média e principalmente a moda das propriedades espectrais. Enquanto as métricas de soma e de média foram utilizadas uma única vez cada no processo de subdivisão, a moda foi utilizada sete vezes, demonstrando a sua importância no processo de classificação. Por outro lado, nenhum atributo espacial ou topográfico foi selecionado, apesar da importância desses elementos no mapeamento do uso da Terra e da cobertura vegetal.



**Figura 2:** Árvore de decisão utilizando o algoritmo C4.5.

As classes Pastagem, Solo exposto e Café Produção foram identificadas duas vezes ao longo das ramificações da AD. Isto aconteceu por dois motivos principais: a AD adota um modelo estritamente não paramétrico, portanto, nenhuma suposição com relação a um modelo de distribuição de probabilidade é adotada; as classes citadas apresentam grande heterogeneidade espectral.

A classe Pastagem inclui tanto a vegetação rasteira com baixa atividade fotossintética (pastos secos) quanto os pastos sujos e banhados. A classe Solo Exposto também apresenta grande diversidade espectral, pois inclui os solos secos (e as suas tipologias) e os solos preparados para o cultivo, além dos afloramentos rochosos. Já a classe Café Produção engloba tanto o café em plena atividade produtiva (caracterizado por índices mais altos de área foliar e de sombra de dossel) quanto o café velho ou podado (caracterizado pela maior presença de galhos secos e menor abundância de folhas verdes e de sombra).

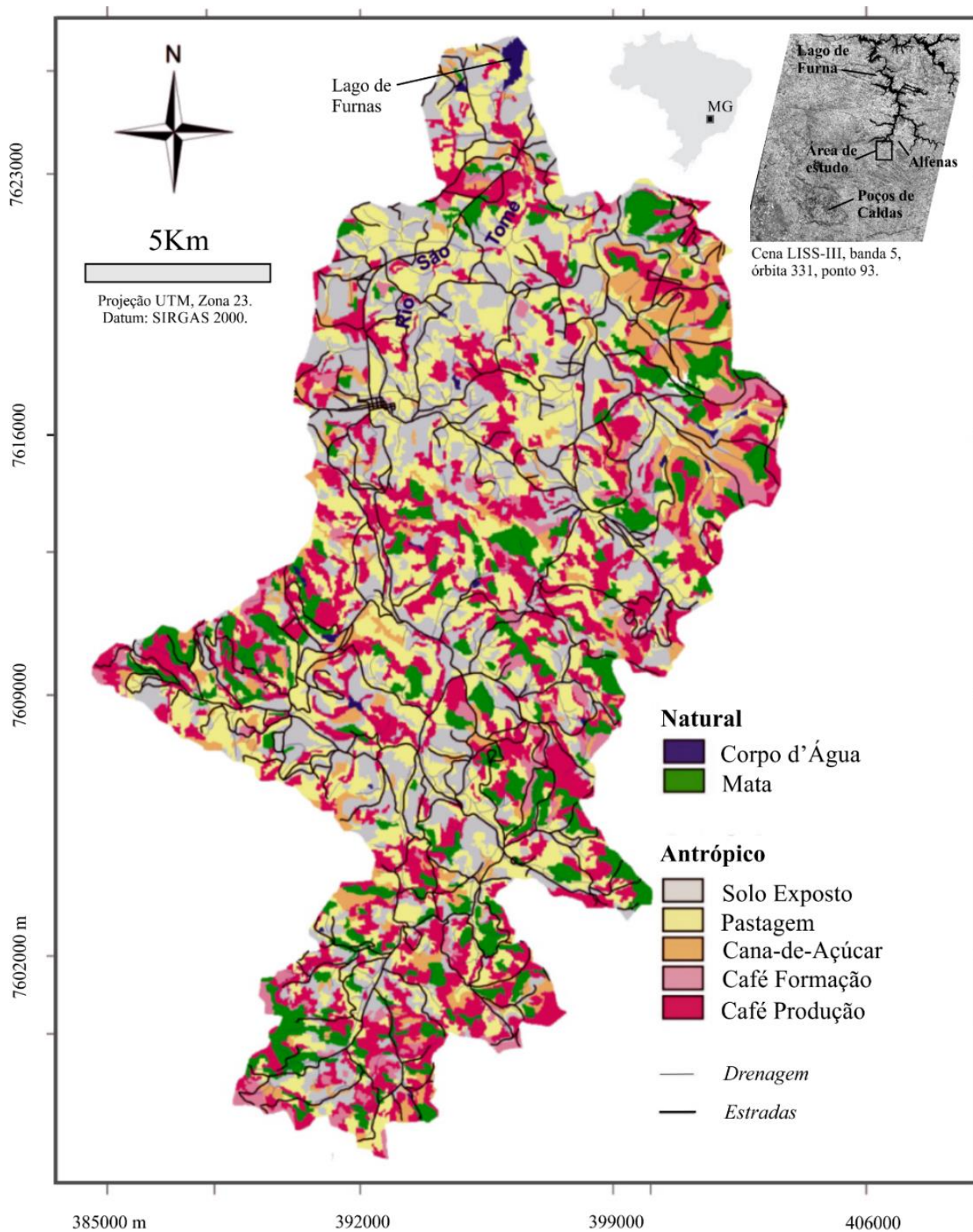
Ao analisar de maneira geral o resultado do mapeamento (**Figura 3**), observa-se que a matriz da bacia é formada por pastagem. Observa-se uma expressiva presença de solo exposto em razão da confusão com os pastos secos. As matas ciliares são praticamente inexistentes e as manchas, fragmentadas, localizam-se principalmente nas áreas de maior declive do terreno. O café, classificado em Formação e Produção, é o principal cultura agrícola da bacia. Essa cultura ocupa principalmente as cotas mais altas do relevo, sendo cultivada em vertentes mais íngremes adjacentes às manchas de mata e em topos de morro. Por fim, os plantios de cana-de-açúcar concentram-se na porção nordeste da bacia por se tratar de uma área mais plana que favorece o cultivo mecanizado.

A matriz de confusão (**Tabela 1**) mostra que os resultados de exatidão e de confiabilidade alcançaram resultados expressivos. A média da exatidão e de confiabilidade foi 75% e 73%, respectivamente. A exatidão global alcançou um valor de 73%. Outro valor utilizado como referência de qualidade do mapeamento foi o índice *Kappa*, que alcançou o valor de 0,66, sendo enquadrada na categoria de desempenho bom (0,61-0,80) (LANDIS; KOCH, 1977).

As classes que tiveram as maiores exatidões individuais foram a Cana-de-Açúcar (91%) seguida pelo Solo Exposto (85%) e Mata (85%). No caso do café, vários estudos já realizados mostram a dificuldade em separar cultivos de café de outros usos em decorrência de sua alta diversidade espectral (MOREIRA *et al.*, 2004; DO PRADO *et al.*, 2016). Dependendo do estágio fenológico do cultivo, idade, manejo e condições topográficas, o comportamento espectral do café pode ser semelhante à mata (café produção), ao pasto verde ou cana-de-açúcar (café formação ou café produção em terrenos declivosos) ou pastagem (café velho ou podado). A exatidão da classe Pastagem (69%) foi comprometida pela grande confusão com o Solo Exposto. Por fim, a classe Corpo d'Água (63%) foi confundida com a Mata e Café Produção em virtude do sombreamento existentes nestas classes. Com relação à confiabilidade individual, a classe Corpo d'Água foi a que apresentou a maior confiabilidade (100%), seguido pelo Café Produção (86%) e Café Formação (85%). A classe de menor confiabilidade foi a Cana-de-Açúcar (38%) em decorrência principalmente da confusão com o Café Produção.

Por fim, com relação a importância do MUTCV da bacia do Rio São Tomé, esta constitui-se um importante afluente da represa de Furnas. O conhecimento do uso da Terra é crucial para propor medidas que visem a conservação da qualidade da água e da vida útil do reservatório. Isto porque tanto a quantidade de nutrientes carreados pelos rios quanto a quantidade de sedimentos erodidos e depositados no fundo dos reservatórios estão diretamente relacionados com atividades agrícolas que são praticadas na bacia.

Além do aspecto ambiental e de manutenção da atividade do reservatório, as informações levantadas pelo MUTCV são de grande importância para o planejamento agrícola da região. O sul de Minas Gerais destaca-se no cenário nacional pela produção de café e pecuária leiteira. O conhecimento de área plantada de café em fase produtiva e em fase de desenvolvimento, por exemplo, pode auxiliar os órgãos responsáveis pelas estimativas anuais de produção do café a melhorar as estimativas de previsão de safra. Isso porque dependendo do estágio fenológico do cultivo, diferentes valores de produtividade são observados. A classe Café Produção, por exemplo, corresponde os talhões no auge da fase produtiva, enquanto o Café Formação ainda se encontra em fase de crescimento, e por conseguinte, de ascensão na produção de grãos. Estas informações levantadas por meio do mapeamento podem ser combinadas com os métodos tradicionais de estimativas de safra (baseados em questionários) como forma de tornar a metodologia mais objetiva, especialmente na fase de coleta de dados.



**Figura 3:** Mapa de uso da Terra e cobertura vegetal gerado a partir de classificação por Mineração de Dados.



**Tabela 1:** Matriz de confusão dos resultados da classificação

Classificação → Referências ↓	1	2	3	4	5	6	7	Total	Acur. (%)
Café formação (1)	57	13	2	0	0	7	0	79	72
Café produção (2)	8	219	48	37	7	35	0	354	61
Cana-de-açúcar (3)	0	0	34	2	1	0	0	37	91
Pastagem (4)	0	5	1	106	38	2	0	152	69
Solo exposto (5)	0	1	2	29	188	0	0	220	85
Mata (6)	2	13	1	2	0	108	0	126	85
Corpo d'água (7)	0	3	0	1	0	3	12	19	63
Total	67	254	88	177	234	155	12	987	
Conf. (%)	85	86	38	59	80	69	100		

Média da acurácia (Acur.)=  $(72+61+91+69+85+85+63)/7 = 75,1\%$ ;  
Média da confiabilidade (Conf.)=  $(85+86+38+59+80+69+100)/7 = 73,8\%$   
Exatidão Global=  $(57+219+34+106+188+108+12)/987 = 73,3\%$   
Kappa = 0,66

#### 4. CONCLUSÃO

A técnica de mineração de dados mostrou-se eficiente no MUTCV em área com grande diversidade de uso como no sul de Minas Gerais. A exatidão global alcançada foi de 73% com índice *kappa* de 0,66, o que a enquadra na categoria de desempenho bom. Apesar de considerar na mineração de dados atributos espectrais, espaciais e topográficos, apenas informações espectrais foram selecionadas na classificação. Todas as bandas do sensor LISS III foram selecionadas, o que reforça o potencial dessas imagens LISS-III na identificação e classificação do uso da Terra e cobertura vegetal. Tal resultado, todavia, não negligencia a importância da forma dos segmentos no processo de classificação nem dos atributos do relevo, pois este último exerce papel muitas vezes condicionante na distribuição espacial dos elementos da paisagem. Por fim, é importante também salientar que a técnica de mineração de dados constitui em um instrumento de auxílio ao geógrafo na identificação de regras e padrões, porém, ela não substitui etapas importantes do mapeamento, como a realização de trabalhos de campo e o exercício da fotointerpretação. Isso porque a “geração automática” das regras é derivada de amostragens feitas com base em informações coletadas em campo e analisadas em laboratório.

#### AGRADECIMENTOS

Os autores expressão os seus agradecimentos à CAPES pela concessão da bolsa de mestrado.

#### REFERÊNCIAS

- AB'SÁBER, A. N. **Os domínios de natureza no Brasil: potencialidades paisagísticas**. São Paulo: Ateliê Editorial, 2ª ed., 2003.
- BAATZ, M.; BENZ, U.; DEGHANI, S.; HEYNEN, M.; HÖLTJE, A.; HOFMANN, P.; LINGENFELDER, I.; MIMLER, M.; SOHLBACH, M.; WEBER, M.; WILLHAUCK, G. **User Guide 4 (eCognition)**. Germany: Definiens, 2004.
- BLASCHKE, T. Object based image analysis for remote sensing. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 65, n.1, p. 2-16, 2010.

- CERON, A. O.; DINIZ, J. A. F. O uso de fotografias aéreas na identificação das formas de utilização agrícola da terra. *Revista Brasileira de Geografia*, n. 2, p.161-173, 1966.
- CHANDER, G.; COAN, M. J.; SCARAMUZZA, P. L. Evaluation and comparison of the IRS-P6 and the Landsat sensors. *IEEE Transactions on Geoscience and Remote Sensing*, v. 46, n. 1, p. 209-221, 2008.
- CONGALTON, R. G. A Review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, v. 37, p. 35-46, 1991.
- CONGALTON, R. G.; GREEN, K. A Practical look at the sources of confusion in error matrix generation. *Photogrammetric Engineering & Remote Sensing*, v. 59, n. 5, p. 641-644, 1993.
- CRÓSTA, A. P. **Processamento digital de imagens de sensoriamento remoto**. Campinas, SP: IG/UNICAMP. Ed. rev. 4ª Reimpressão. 2002.
- DO PRADO, D. F. C.; HAYAKAWA, E. H.; KAWAKUBO, F. S. **Mapeamento do cultivo de café no sul de Minas Gerais utilizando imagens Landsat-5 TM e variáveis topográficas**. Revista do Departamento de Geografia USP, v. 31, p. 43-57, 2016.
- FAYYAD, U.; PIATETSKY-SHAPIO, G.; SMYTH, P. From data mining to knowledge discovery in databases. *AI Magazine*, v.17, n. 3, p.37-54, 1996.
- FRIEDL, M. A.; BRODLEY, C. E. Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment*, v.61, n.3, p.399-408, 1997.
- GORTE, B. G. H.; KOOLHOVEN, W. Interpolation between isolines based on the Borgfors distance transform. *ITC Journal*, v.3, p. 245-247, 1990.
- KORTING, T. S.; FONSECA, L. M. G.; ESCADA, M. I. S.; CÂMARA, G. GeoDMA – Um sistema para mineração de dados de sensoriamento remoto. In: Simpósio Brasileiro de Sensoriamento Remoto, 14.; 2009, Natal. *Anais...* São José dos Campos: INPE, 2009.
- KORTING, T. S.; FONSECA, L. M. G.; CÂMARA, G. GeoDMA – Geographic data mining analyst. *Computers & Geosciences*, v.57, p.133-145, 2013.
- LANDIS, J. R.; KOCH G. G. The measurement of observer agreement for categorical data. *Biometrics*, v. 33, p.159–174, 1977.
- MOREIRA, M. A.; ADAMI, M.; RUDORFF, B. F. T. Análise espectral e temporal da cultura do café em imagens Landsat-5. *Pesquisa Agropecuária Brasileira*, v.39. n.3, p.223-231. 2004.
- PEEL, M. C.; FINLAYSON, B. L.; MCMAHON, T. A. Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences*, v.11, p.1633-1644, 2007.
- PRADO, D. F. C.; HAYAKAWA, E. H.; KAWAKUBO, F. S. mapeamento do cultivo de café no sul de Minas Gerais utilizando imagens Landsat-5 TM e variáveis topográficas. *Revista do Departamento de Geografia*, v. 31, 43-57, 2016.
- QUINLAN, J. R. **C4. 5: Programs for machine learning**. [S.l.]: Morgan kaufmann, 1993.
- RADAMBRASIL. **Levantamento dos recursos naturais: geologia, geomorfologia, pedologia, vegetação e uso potencial da terra. Folhas SF. 23 / 24, Rio de Janeiro/Vitória**. Rio de Janeiro: Ministério das Minas e Energia, 1983. 780p.
- SILVA, A. C; SILVA, I. C.; PIMENTA, C. J.; TANURE, E. L.; CAMPOS, J. C.; MIRANDA, J. M.; PIMENTA, M. E. S. G.; LANDGRAF, P. R. C.; MENDONÇA, R. S.; PEREIRA, R. C. Desenvolvimento sustentável da bacia hidrográfica do rio São Tomé. Gestão de Extensão e Estudos Independentes: *Revista do I Fórum de Extensão Universitária*, p.1-9, 2011.
- SILVA, M. P. S.; CÂMARA, G.; ESCADA, M. I. S.; SOUZA, R. C. M. Remote-sensing image mining: detecting agents of land-use change in tropical forest areas. *International Journal of Remote Sensing*, v.29, n.16, p.4803-4822, 2008.

STRAHLER, A. N. Hypsometric (area-altitude) – analysis of erosion al topography. **Geological Society of America Bulletin**, v.63, n.10, p.1117-1142, 1952.

VIEIRA, M. A.; FORMAGGIO, A. R.; RENNÓ, C. D.; ATZBERGER, C.; AGUIAR D. A.; MELLO, M. P. Object based image analysis and data mining applied to a remotely sensed Landsat time-series to map sugarcane over large areas. **Remote Sensing of Environment**, v.123, n. 553-562, 2012.

ZUBEN, F. J. V.; ATTUX, R. R. F. 2010. **Árvores de decisão**. Campinas: DCA/FEEC/Unicamp.