

Classificação automática de áreas cafeeiras utilizando imagens de sensoriamento remoto e técnicas de mineração de dados.

Cesare Di Girolamo Neto ¹
Leila Maria Garcia Fonseca ¹
Thales Sehn Korting ¹
Ieda Del'Arco Sanches ¹
Isaque Daniel Rocha Eberhardt ¹
Hugo do Nascimento Bendini ¹
Rennan de Freitas Bezerra Marujo ¹
Kleber Tranbaquini ¹

¹ Instituto Nacional de Pesquisas Espaciais - INPE
Caixa Postal 515 - 12227-010 - São José dos Campos - SP, Brasil
{cesare, hbendini, rennan, kleber, isaque, ieda}@dsr.inpe.br, {thales, leila}@dpi.inpe.br

Abstract. Coffee is the main crop produced in the southern of Minas Gerais state, Brazil, and techniques for estimating the area used for this crop are being intensely investigated in order to produce reliable yield estimates. Coffee trees have a similar spectral pattern to forest, making it difficult to automatically distinguish these land use types. This study evaluated the Random Forest and Decision Tree algorithms for an automatic classification of coffee areas in municipality of Machado, Minas Gerais, Brazil. First, the data were preprocessed by creating gray level masks in each of the 11 bands of a Landsat-8 image. Then the Random Forest and Decision Trees were trained and applied on the image in order to verify its potential for discriminating coffee areas. Lastly, the analysis and validation of the results were conducted using as reference one map manually classified. The Kappa index and the overall accuracy were used to assess the quality of the models tested. The Random Forest classifier presented better results than the Decision Trees, with an accuracy of 84.13% and a Kappa index of 0.6, which is more accurate when compared to previous studies. We also provide a list of bands that were not suitable for this type of classification.

Palavras-chave: *Random forests, decision trees, land cover, agriculture, Landsat-8, Florestas aleatórias, árvores de decisão, cobertura da terra, agricultura, Landsat-8.*

1. Introdução

A análise do uso da terra consiste em determinar como uma área de interesse é utilizada, permitindo a caracterização das interações antrópicas com o meio ambiente. O conhecimento da dinâmica de transformação do uso da terra mostra-se cada vez mais importante, a fim de analisar a forma pela qual determinado espaço está sendo ocupado. No Brasil, grandes transformações em seu espaço foram decorrentes do desenvolvimento do setor agrícola (HELFAND e RESENDE, 2000). Com levantamentos do uso e cobertura da terra, os padrões de organização do espaço podem ser compreendidos e podem-se observar diversas consequências, como o uso inadequado do solo (ROSA, 2009).

Para obtenção e análise dos padrões do uso da terra, mostra-se importante atrelar dados advindos de sensores remotos (BRANNSTROM et al., 2008). Esta tecnologia permite gerar mapas de cobertura do solo, os quais têm grande importância por demonstrarem, a partir da interpretação de imagens de satélites, áreas ocupadas por pastagem, culturas agrícolas, cursos de rios e outras feições. Neste sentido, a cultura do café, embora esta apresente variações no comportamento espectral, por causa de fatores como espaçamento, idade e época do ano, ela pode ser identificada e mapeada em imagens de média resolução espacial, com boa precisão de mapeamento (MOREIRA et al., 2004).

O mapeamento de áreas cafeeiras torna-se ainda mais atrativo dado que o Brasil é atualmente o maior produtor de café do mundo, sendo que em 2012 foi responsável por cerca de 37% da produção mundial, o equivalente a 55,9 milhões de sacas de 60 Kg. Com cerca de

60% da produção destinada ao mercado externo, os ganhos anuais do país com a exportação deste grão chegaram próximos a US\$ 6 bilhões (USDA, 2013). O café é cultivado em 12 estados brasileiros, sendo que o maior produtor é Minas Gerais, com aproximadamente 51% da produção nacional (MINISTÉRIO DA AGRICULTURA, 2013).

O estado de Minas Gerais teve sua área cafeeira mapeada por Moreira et al. (2007). Este trabalho visou mapear, manualmente, todo o estado e identificar áreas cafeeiras e não cafeeiras para imagens do sensor TM (*Thematic Mapper*) do satélite Landsat-5 para o ano de 2006. Dentre outras conclusões obtidas pelos autores, está a dificuldade de identificação de áreas cafeeiras com áreas de cerrado e eucalipto. De maneira geral, a cultura cafeeira apresenta resposta espectral bastante complexa, em função da variabilidade de seus diversos parâmetros, tais como declividade, espaçamento entre plantas, estado vegetativo, estágio fenológico, sombreamento, manejo e dentre outros (VIEIRA et al., 2006). Adami et al. (2009), avaliaram a exatidão do mapeamento realizado por Moreira et al. (2007), quanto à identificação de áreas cafeeiras. Os autores concluíram que 95% do mapeamento foi feito corretamente, mais precisamente para as regiões sul e sudeste de Minas Gerais, essa taxa de acerto foi de 99%. Tal exatidão foi atribuída à característica de cultivo da região.

Todavia, uma das maiores dificuldades encontradas pelos usuários de dados de sensoriamento remoto reside na disponibilidade de um método rápido e preciso para obter mapas de uso e cobertura da terra. O mapeamento manual, apesar de ser uma técnica precisa, é praticamente inviável quando aplicado a áreas muito extensas, pela demanda de tempo (VASCONCELOS e NOVO, 2004).

Uma alternativa aos métodos manuais é o uso de métodos de classificação automática, com metodologias que permitem a classificação automática pixel a pixel, por exemplo. A metodologia computacional de mineração de dados demonstra ter alto potencial de aplicação em estudos relacionados ao mapeamento de áreas cafeeiras, sendo utilizadas diversas técnicas de classificação, como redes neurais artificiais (ANDRADE et al., 2011; 2013), máquinas de vetores suporte (BISPO et al., 2014) e árvores de decisão (DELGADO et al., 2012).

O crescente uso da técnica de mineração de dados chamada de florestas aleatórias, ou, *Random Forests*, tem chamado a atenção em estudos relacionados à identificação do uso da terra (RODRIGUEZ-GALIANO et al., 2012). Esta técnica também pode ser considerada uma das mais precisas quando comparada a outras, como redes neurais artificiais e máquinas de vetores suporte (CARUANA et al., 2008). As florestas aleatórias ainda são computacionalmente muito efetivas, além de evitarem sobreajuste (*overfitting*) e serem pouco sensíveis a ruídos (BREIMAN, 2001).

Estas observações evidenciam a importância de pesquisas capazes de gerar informações que possam ser utilizadas para fornecer suporte ao mapeamento de áreas cafeeiras por métodos computacionais. Sendo assim, o objetivo deste trabalho é avaliar o potencial dos algoritmos de *Random Forests* e árvores de decisão para o mapeamento automático de áreas cafeeiras para o município de Machado/MG.

2. Metodologia de trabalho

A metodologia usada para desenvolver este trabalho foi a CRISP-DM (*CRoss Industry Standard Process for Data Mining*) (CHAPMAN et al., 2000). Esta divide o ciclo de vida de um projeto em seis fases: compreensão do domínio, entendimento dos dados, preparação dos dados, modelagem, avaliação e distribuição. A metodologia é cíclica, sendo que a sequência lógica entre as fases não é rígida, sendo comum, e quase sempre necessário, voltar e avançar entre diferentes fases.

2.1. Entendimento dos dados

Os dados utilizados foram coletados com o auxílio do pesquisador Maurício Alves Moreira do Instituto Nacional de Pesquisas Espaciais. Estes dados referem-se a uma máscara com a classificação manual de café para a região sul de Minas Gerais. O município selecionado para o estudo foi Machado, por ser um dos principais produtores do estado e já ter sido alvo de outros estudos relacionados à classificação automática de café (Figura 1).

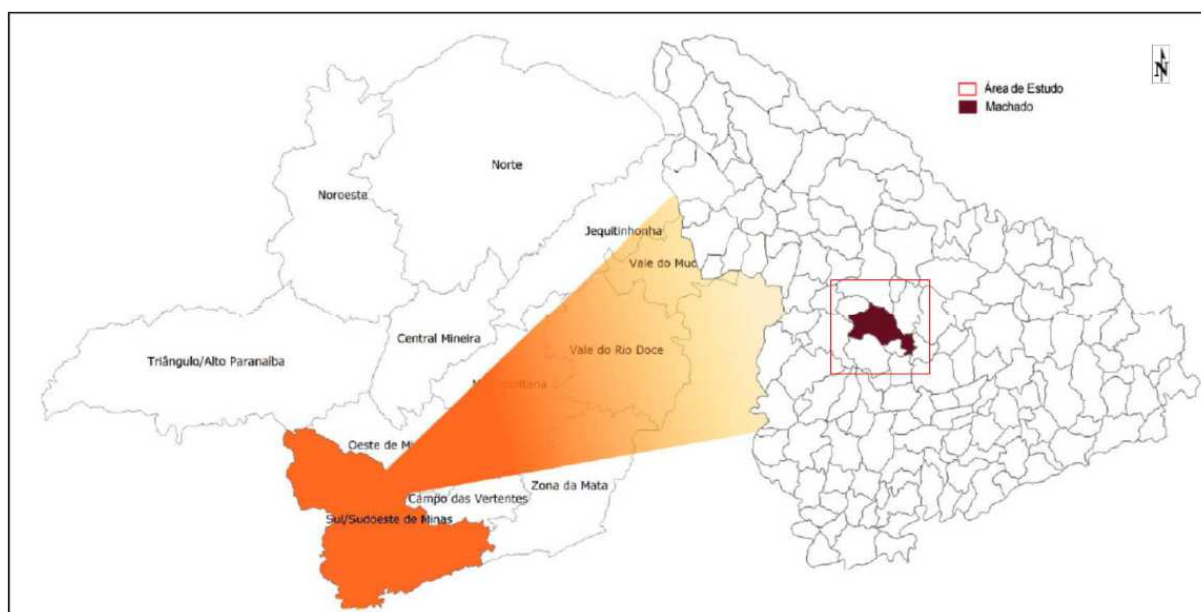


Figura 1: Área de estudo – município de Machado/MG (Adaptado de ANDRADE, 2011).

O conjunto de dados obtido foi composto por três arquivos:

- Uma imagem com 11 bandas do sensor OLI (*Operational Land Imager*) do satélite Landsat-8 para a região sul de Minas Gerais datada de 08/02/2014.
- Um shapefile contendo os estados de Minas Gerais.
- Uma imagem temática com classificação de áreas de café para a mesma região da imagem.

2.2. Preparação dos dados

Esta fase seguiu os seguintes passos:

Passo 1 - Reunião dos dados brutos em um sistema de informação geográfica

Foram utilizados os softwares SPRING, versão 5.1.8 (CÂMARA et al., 1996) e o ENVI (ENVI, 2009) para inserir as imagens e o shapefile.

Passo 2 - Criação de atributos de nível de cinza

O segundo passo da preparação dos dados foi criar os atributos de nível de cinza para cada uma das bandas da imagem. Foi feito um recorte da área de estudos para o município de Machado e a reamostragem das bandas 8, 10 e 11 para a resolução de 30x30m (igual a das demais bandas). Com este recorte foi realizada uma modelagem numérica de terreno, a fim de

se obter o nível de cinza dos pixels nesta região. Este procedimento foi repetido para cada uma das bandas.

Passo 3 - Extração dos atributos para modelagem

O terceiro passo foi extrair os atributos das imagens para criar um arquivo que pudesse ser utilizado no software WEKA, versão 3.6.11 (HALL et al., 2009). Este software é o responsável por realizar o processo de mineração de dados e extração de padrões. Foram adaptados scripts na linguagem computacional C++ para a extração destes dados do software SIG, resultando em um arquivo do tipo .csv, com os atributos necessários para a mineração de dados.

O atributo meta ou a variável dependente foi a classe nomeada *café* ou *não-café*. Ela foi obtida de acordo com a classificação manual realizada por Moreira para cada pixel dentro da área de estudo. Na imagem classificada também foram mapeadas classes relacionadas ao perímetro urbano de Machado e também pequenas áreas classificadas como *café podado*. No processo de poda o café sofre, normalmente, uma redução de sua área foliar a fim de aumentar a sua produtividade, entretanto a poda pode ser de diversas formas, passando por procedimentos como recepa, decote, esqueletamento e desponte (PROCAFÉ, 2013). Como pode ocorrer uma mudança no comportamento espectral nestas áreas a classe de café podado foi excluída do conjunto de treinamento. O mesmo foi realizado para a classe urbana, seguindo procedimento análogo a Andrade et al., (2013).

Os atributos preditivos partiram do valor de cinza de cada pixel. Uma relação entre as bandas foi utilizada para gerar os índices de vegetação como EVI e NDVI (Tabela 1). O conjunto de dados preparado para uma imagem totalizou 645.785 registros.

Tabela 1: Atributos utilizados no conjunto de dados.

Nome	Tipo	Significado
BANDA 1	Numérico	Nível de cinza dos pixels na banda 1.
BANDA 2	Numérico	Nível de cinza dos pixels na banda 2.
BANDA 3	Numérico	Nível de cinza dos pixels na banda 3.
BANDA 4	Numérico	Nível de cinza dos pixels na banda 4.
BANDA 5	Numérico	Nível de cinza dos pixels na banda 5.
BANDA 6	Numérico	Nível de cinza dos pixels na banda 6.
BANDA 7	Numérico	Nível de cinza dos pixels na banda 7.
BANDA 8	Numérico	Nível de cinza dos pixels na banda 8.
BANDA 9	Numérico	Nível de cinza dos pixels na banda 9.
BANDA 10	Numérico	Nível de cinza dos pixels na banda 10.
BANDA 11	Numérico	Nível de cinza dos pixels na banda 11.
CLASSE	Nominal	Classe do pixel: <i>café</i> ou <i>não-café</i>

2.3. Modelagem

A partir do conjunto de dados brutos foi realizado um procedimento de amostragem sem reposição. Tal processo permite que os algoritmos de mineração tratem bases de dados pela redução do número de casos avaliados sem que haja perdas significativas na qualidade das predições (WITTEN et al., 2011). Este procedimento foi realizado mantendo o equilíbrio das classes em 50% cada uma. Classes balanceadas tendem a produzir melhores resultados do que

uma distribuição diferente (WEISS e PROVOST, 2001). O conjunto reamostrado totalizou 161.446 registros.

A partir do conjunto de dados preparado e balanceado, foram utilizados os métodos de seleção de atributos. Estes métodos têm como finalidade reduzir o número de atributos do conjunto, melhorando o desempenho dos modelos gerados e proporcionando um ganho computacional. Foram aplicados quatro métodos: *Wrapper* (JOHN e KOHAVI, 1997), *Correlation Feature Selection* (CFS) (HALL, 1999), *Infogain* e *Gainratio* (WITTEN et al., 2011).

Para gerar as *Random Forests* foi necessário calibrar alguns parâmetros do algoritmo, visando melhor taxa de acerto na classificação. Foi avaliado o número de árvores em cada floresta, sendo que este parâmetro foi variado até 100, conforme recomendado por Breiman (2001). Além deste, a profundidade de cada árvore também foi ajustada, sendo mantida a opção padrão do software, que é uma profundidade ilimitada. Já para as árvores de decisão foi variado o parâmetro de número de objetos por folha, variando-se de 7500 até 250. Este atributo funciona como uma forma de aumentar ou reduzir a profundidade da árvore, similar a um atributo de poda (WITTEN et al., 2011).

3. Resultados e discussão

Os classificadores foram gerados a partir do conjunto de treinamento (161.446 registros) e posteriormente avaliado no conjunto de teste, contendo os demais pixels para o município de Machado. Os resultados da classificação encontram-se na Tabela 2.

Tabela 2: Taxas de acerto e índice Kappa para a classificação automática de áreas cafeeiras por *Random Forests* e árvores de decisão para o conjunto de teste.

	Árvores de decisão			
Objetos por folha	7500	1500	750	250
Taxa de acerto	76,10%	75,93%	77,47%	76,90%
Kappa	0,38	0,40	0,42	0,42
	Conjunto de teste			
Árvores na Floresta	5	10	25	100
Taxa de acerto	81,70%	81,00%	83,97%	84,13%
Kappa	0,54	0,54	0,60	0,60

As árvores de decisão mostram-se classificadores menos eficientes que as *Random Forests*, onde se notou uma estabilização na taxa de acerto ao redor de 77% e 84%, respectivamente. A árvore de decisão gerada com 750 registros apresentou um tempo de processamento 23% maior quando comparada a de 1500 registros, todavia o ganho na taxa de acerto foi 1,54 pontos percentuais. Para as *Random Forests* notou-se que a floresta gerada com 25 árvores não mostrou diferença considerável na classificação em relação a de 100 árvores. Já o seu custo computacional foi menor, sendo que o tempo de processamento foi de 76% inferior ao tempo utilizado para gerar a floresta com 100 árvores.

Os resultados obtidos com as árvores de decisão não foram melhores do que os trabalhos obtidos em literatura. O valor de índice Kappa obtido com esse classificador, no melhor caso, foi de 0,42, sendo bastante inferior ao valor de 0,55 obtido por Andrade et al. (2013) para regiões montanhosas.

No caso das *Random Forests*, foi possível obter índice Kappa igual a 0,60, melhor valor obtido no presente estudo, o que pode ser considerado um avanço na área em relação aos resultados de trabalhos anteriores. Andrade et al. (2013) também atingiram este índice, entretanto, foi necessário realizar uma separação de áreas com relevo acentuado e plano.

Como este estudo considerou todo o município de Machado (extensão territorial maior) e não fez distinção do relevo das áreas, as *Random Forests* podem ser consideradas promissoras para classificação automática de áreas cafeeiras.

A Figura 2 mostra o mapa temático de classificação automática com o mapa de classificação manual (referência) sobrepostos. Este procedimento permitiu a identificação de áreas cafeeiras que foram classificadas incorretamente. Podem-se observar dois pontos (1 e 2) na Figura 2A que apresentaram uma grande quantidade de áreas cafeeiras classificadas incorretamente. Nestes locais do estado há algumas cadeias de morros e, conseqüentemente, uma declividade que influenciou na precisão da classificação. Ainda na Figura 2A, outros dois pontos (3 e 4) a sudeste do estado obtiveram alta taxa de classificação incorreta. Estes locais apresentaram outras culturas agrícolas e áreas de mata, as quais foram confundidas com áreas cafeeiras.

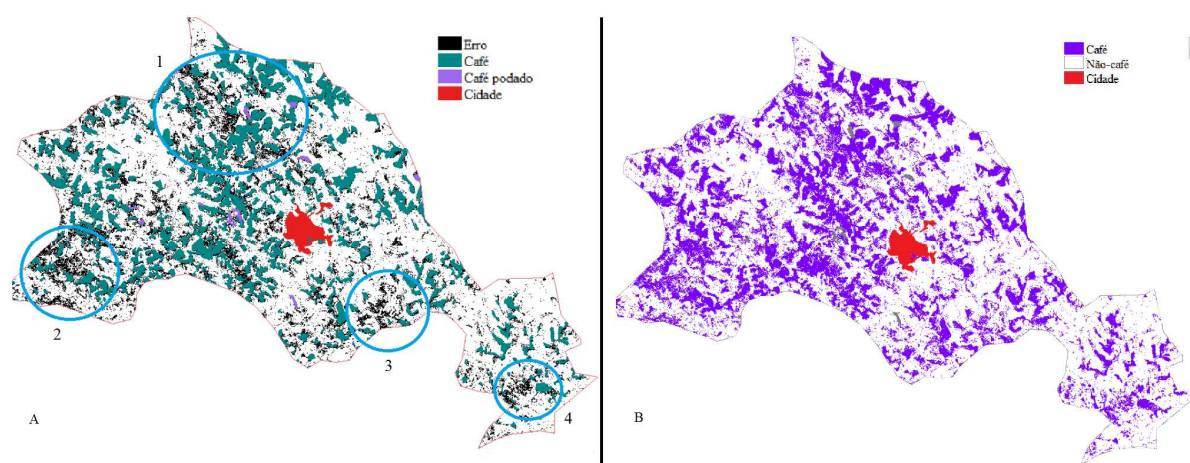


Figura 2: A) Mapa com a classificação automática de café contendo os erros de classificação e as quatro principais regiões onde eles ocorreram (destacados com círculos azuis na figura). B) Mapa temático com a classificação de áreas cafeeiras pelo algoritmo de *Random Forests*.

Os métodos de seleção de atributos não apresentaram ganhos de taxa de acerto e índice Kappa na classificação pelo algoritmo de *Random Forests*, apenas um ganho computacional no tempo de indução dos classificadores. O método *Wrapper* selecionou apenas um atributo, que foi a banda 8 (pancromática), entretanto seu classificador obteve taxa de acerto de 59,8% e índice Kappa de 0,22. Este caso mostrou que a utilização apenas dos pixels da banda pancromática como atributos não é adequada para estudos desta natureza. O método CFS selecionou quatro atributos para classificação, as bandas 2, 3 5 e 8. O desempenho também foi inferior aos modelos da Tabela 2, sendo que sua taxa de acerto foi de 78,3% e índice Kappa de 0,49.

Já os métodos de *Infogain* e *Gainratio* não retornam um conjunto, eles apenas indicam quais atributos são menos relevantes no conjunto de dados. Estes métodos mostraram que a exclusão das bandas 7, 9 e 10 não afetaram o desempenho dos classificadores gerados com 100 árvores na floresta. A taxa de acerto foi de 84,0% e o índice Kappa se manteve em 0,6. O tempo de indução dos classificadores sem as bandas mencionadas foi 20% inferior com relação ao classificador com todas as bandas.

4. Conclusão

A maioria dos classificadores automáticos de imagens espectrais avalia apenas a resposta espectral dos alvos. Quando o alvo estudado, como a cultura cafeeira, apresenta tamanha complexidade, onde fatores ambientais, fenológicos e de manejo interferem na resposta

espectral, tais classificadores não apresentam resultados satisfatórios. Assim, o algoritmo de mineração de dados de *Random Forests* pode ser considerado uma alternativa à classificação manual de áreas cafeeiras.

Os resultados alcançados no presente trabalho, para a região cafeeira de Machado-MG, igualam-se aos encontrados por Andrade et al. (2013) com redes neurais artificiais, onde houve uma distinção de áreas com relevo elevado. Ao se trabalhar com toda a área do município de Machado, as *Random Forests* podem ser consideradas classificadores que tendem a superar o desempenho das redes neurais. Todavia, seria necessário conduzir mais investigações sobre como esta técnica iria classificar toda a área do município. Apesar de ainda não ter sido incorporado ao conjunto de atributos nenhum outro parâmetro a não ser o espectral, a metodologia proposta em ambos os trabalhos separou ambientes urbanos, a fim de permitir uma melhor classificação.

Sugere-se como proposta de continuidade do trabalho ampliar o número de classes a fim de treinar um modelo com menor variabilidade espectral. Pode-se ainda trabalhar com uma classificação baseada em segmentação ou mesmo avaliar outros algoritmos de classificação.

Agradecimentos

Ao Professor Maurício Alves Moreira, pelo fornecimento da imagem de referência. Ao Professor Antônio Roberto Formaggio, pelos conselhos e ensinamentos em sensoriamento remoto agrícola.

5. Referências Bibliográficas

Adami, M.; Moreira, M. A.; Barros, M. A.; Martins, V. A. Avaliação da exatidão do mapeamento da cultura do café no Estado de Minas Gerais In: Simpósio Brasileiro de Sensoriamento Remoto (SBSR), 14., 2009, Natal. **Anais...**, São José dos Campos: INPE, 2009. Artigos, p. 1-8. On-line. Disponível em: <<http://www.dsr.inpe.br/laf/cafesat/artigos/AvaliacaoExatidaoMapeamentoCafe.pdf>>. Acesso em: 02 abr. 2014.

Andrade, L.N. **Redes Neurais Artificiais aplicadas na identificação automática de áreas cafeeiras em imagens de satélite**. 2011. 87p. Dissertação (Mestrado em Ciência da Computação) – Departamento de ciência da computação, Universidade Federal de Minas Gerais, 2011,

Andrade, L.N.; Vieira, T.G.C.; Lacerda, W.S.; Davis Junior, C.A. Redes Neurais Artificiais (RNA) aplicadas à classificação de áreas cafeeiras na região de Três Pontas-MG. In: Simpósio Brasileiro de Sensoriamento Remoto, 15., 2011, Curitiba. **Anais...** São José dos Campos: INPE, 2011. Artigos, p. 7603-7610. On-line. Disponível em: <<http://www.dsr.inpe.br/sbsr2011/files/p0606.pdf>>. Acesso em: 07 abr. 2014.

Andrade, L.N.; Vieira, T.G.C.; Lacerda, W.S.; Volpato, M.M.L.; Davis Junior, C.A. Aplicação de redes neurais artificiais na classificação de áreas cafeeiras da região de Machado (MG). **Coffee Science**, v.8, n.1, p.78-90, 2013.

Bispo, R.C.; Lamparelli, R.A.C.; Rocha, J.V. Using fraction images derived from modis data for coffee crop mapping. **Engenharia Agrícola**, v.34, n.1, p.102–111, 2014.

Brannstrom, C.; Jepson, W.; Fillipi, A. M.; Redo, D.; Xu, S.; Ganesh, S. Land change in the Brazilian Savanna (Cerrado), 1986-2002: Comparative analysis and implications for land-use policy. **Land use policy**, v.25, p.579-595, 2008.

Breiman, L. Random forests. **Machine Learning Journal**. Hingham, v.45, p.5–32, 2001.

Camara, G.; Souza, R. C. M.; Freitas, U. M.; Garrido, J. SPRING: Integrating remote sensing and GIS by object-oriented data modelling. **Computers & Graphics**. v. 20, n.3, p. 395-403, 1996.

Caruana, R.; Karampatziakis, N.; Yessenalina, A. An empirical evaluation of supervised learning in high dimensions. In: International conference on Machine learning, 25., 2008, Helsinki. **Proceedings...** Helsinki:ACM, 2008. p. 96-103. On-line. Disponível em: < <http://cms.brookes.ac.uk/research/visiiongroup/>

talks/rg_august_09_normalized_cuts/eval_of_supervised_learning_in_high_dimensions.pdf>. Acesso em: 07 abr. 2014.

Chapman, P.; Clinton, J.; Kerber, R.; Khabaza, T.; Reinartz, T.; Shearer, C.; Wirth, R. **CRISP-DM 1.0: step-by-step data mining guide**. Illinois: SPSS, 2000. 74 p.

Delgado, R. C.; Sedyama, G. C.; Lima, E. L.; Oliveira, E. C.; Rodrigues, R.A. Evolução espaço-temporal do café no estado de Minas Gerais por meio do classificador de árvore de decisão. **Enciclopédia biosfera**, v.8, n.15, p. 904-913, 2012.

Envi. ENVI Reference Guide. 2009 ed. IDL, 2009. Disponível em:
<http://www.exelisvis.com/portals/0/pdfs/envi/Reference_Guide.pdf> Acesso em: 25.jun.2014.

Hall, M. A. **Correlation-based feature selection for machine learning**. 1999. 178p. Tese (Doutorado em Ciência da Computação) – Department of Computer Science, University of Waikato, Nova Zelândia. 1999.

Hall, M. A.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; Witten, I. H. The WEKA Data Mining Software: An Update; **SIGKDD Explorations**. v.11, n.1, p. 10-18, 2009.

Helfand, S. M.; Rezende, G. C. **Padrões Regionais de Crescimento da produção de grãos no Brasil e Papel da Região Centro-Oeste**. Rio de Janeiro: IPEA, 2000. 20 p. (ISSN 1415-4765).

John, G. H.; Kohavi, R. Wrappers for feature subset selection. **Artificial Intelligence**. v.97, n.1-2, p.273-324, 1997.

Ministério da agricultura. Ministério da Agricultura, Pecuária e abastecimento. <www.agricultura.gov.br>, acesso em 26.jan.2013.

Moreira, M. A.; Adami, M.; Rudorff, B. F. T. Análise espectral e temporal da cultura do café em imagens Landsat. **Pesquisa Agropecuária Brasileira**, v.39, n.3, p.223-231, 2004.

Moreira, M. A.; Barros, M. A.; Rosa, V. G. C.; Adami, M. Tecnologia de informação: imagens de satélite para o mapeamento de áreas de café de Minas Gerais. **Informe Agropecuário**, v.28, n.241, p.27-37, 2007.

PROCAFÉ. Fundação PROCAFÉ Minas Gerais.
<<http://www.fundacaoprocafe.com.br/sites/default/files/publicacoes/pdf/revista/Coffea11.pdf>>, Acesso em 20.mar.2013.

Rodriguez-Galiano, V. F.; Ghimire, B.; Rogan, J.; Chica-Olmo, M.; Rigol-Sanchez, J.P. An assessment of the effectiveness of a random forest classifier for land-cover classification. **ISPRS Journal of Photogrammetry and Remote Sensing**. v.67, p.93-104, 2012.

Rosa, R. **Introdução ao sensoriamento remoto**. 7.ed. Uberlândia: EDUFU, 2009. 264 p.

USDA. **United States Department Of Agriculture**. <www.usda.gov>, acesso em 15.02.2013.

Vasconcelos, C. H.; Novo, E. M. L. M. Mapeamento do uso e cobertura da terra a partir da segmentação e classificação de imagens – fração solo, sombra e vegetação derivadas do modelo linear de mistura aplicado a dados do sensor TM/Landsat5, na região do reservatório de Tucuruí – PA. **ACTA Amazonica**. v.34, n.3, p. 487-493, 2004.

Vieira, T. G. C.; Aalves, H. M. R.; Lacerda, M. P. C.; Veiga, R. D.; Epiphanyo, J. C. N. Crop parameters and spectral response of coffee (coffea arábica l.) areas within the state of Minas Gerais, Brazil. **Coffee Science**, v.1, n.2, p. 111-118, 2006.

Weiss, G. M.; Provost, F. **The effect of class distribution on classifier learnig**: an empirical study. New Brunswick: Rutgers University, 2001, 26 p. (Technical Report ML-TR-44).

Witten, I. H.; Frank, E.; Hall, M. A. **Data mining**: practical machine learning tools and techniques. 3ed. San Francisco: Morgan Kaufmann, 2011. 416 p.