

COMPARAÇÃO ENTRE CLASSIFICADORES DE IMAGEM ORBITAL APLICADOS NO MAPEAMENTO DA BACIA VACACAÍ MIRIM, RS

LUIS FERNANDO CHIMELO RUIZ
RICARDO SIMÃO DINIZ DALMOLIN
ALEXANDRE TEN CATEN
ALESSANDRO MIOLA
SILVIA BEATRIZ ALVES ROLIM

Estudante do Curso Superior de Tecnologia em Geoprocessamento, Bolsista PIBIC, Colégio Politécnico da UFSM, Av. Roraima, CEP 97105-900, Santa Maria, RS, e-mail(ruiz.ch@gmail.com). Professor do Departamento de Solos da Universidade Federal de Santa Maria (UFSM), Bolsista CNPq, Santa Maria, RS, e-mail(dalmolin@ufsm.br). Professor da Universidade Federal de Santa Catarina campus Curitibanos (UFSC), e-mail (acaten@yahoo.com.br). Doutorando em Engenharia Agrícola, Professor do Colégio Politécnico da UFSM, e-mail (engenheiroacm@gmail.com). Professora da Universidade Federal do Rio Grande Sul (UFRGS), e-mail (silvia.rolim@ufrgs.br).

RESUMO

Um importante fator para o manejo dos recursos naturais é o conhecimento do uso da terra. Assim, este trabalho tem como objetivo comparar três métodos de classificação de imagem: máxima verossimilhança, distância de Mahalanobis e árvore de decisão. A imagem utilizada foi do satélite RESOURCESAT-1, sensor LISS III da Bacia Vacacaí Mirim na região central do estado do Rio Grande do Sul (RS). As classes de uso da terra investigadas foram: soja, arroz, solo exposto, floresta, campo e água. Utilizou-se as variáveis elevações, declividade, índice de vegetação por diferença normalizada, índice de vegetação ajustado ao solo, bandas do vermelho (0.62-0.68 μm - VERM), infravermelho próximo (0.77-0.86 μm - IVP) e infravermelho médio (1.55-1.70 μm - IVM) para discriminar essas classes. Para a análise de concordância entre os dados gerados e a verdade terreno, foi calculado o índice kappa e a exatidão global. Além disso, foi atribuída uma qualidade ao índice kappa. O classificador árvore de decisão, utilizando todas as variáveis, obteve os resultados de 86.52% para kappa e de 88.82% exatidão global.

Palavras chave: métodos de classificação, árvore de decisão, mapeamento do uso da terra

ABSTRACT

An important factor in the management of natural resources is the knowledge of land use. Thus, this study aims to compare three methods of image classification: maximum likelihood, Mahalanobis distance and decision tree. The image used was of the sensor LISS III onboard of the ResourceSat-1 Platform, of Vacacaí Mirim Basin in the central portion of Rio Grande do Sul State (RS). The classes of land use investigated were: soy, rice, bare soil, forest, field and water. In order to discriminate these classes, we used the following variables: elevation, slope, vegetation index by the normalized difference, vegetation index adjusted for soil, the red (0.62-0.68 μm), near infrared (0.77-0.86 μm) and mid-infrared (1.55-1.70 μm) bands. For the analysis of agreement between the data generated and the ground truth, we calculated the overall accuracy and kappa index. Furthermore, it was assigned a quality index for kappa. As a result, the decision tree classifier using all variables, obtained 86.52% to index kappa and 88.82% for the and overall accuracy.

Key words: classification methods, decision tree, mapping of land use

1 INTRODUÇÃO

O avanço de áreas agrícolas causa impactos ambientais devido ao desmatamento de florestas e à ocupação dos campos nativos. O uso agrícola desordenado pode contribuir para os processos erosivos e para a desertificação, bem como para o empobrecimento de solos agrícolas. Desse modo, faz-se necessário quantificar e qualificar o uso da terra para que os tomadores de decisão possam agir em áreas prioritárias.

O sensoriamento remoto pela análise digital da imagem possibilita a obtenção de informações sobre a superfície terrestre. Isso o torna uma ferramenta primária para a aquisição de mapas do uso e cobertura da terra (BARBOSA, 2009).

A análise digital de uma imagem de satélite pode estar em dois grupos: análise visual e análise digital. Na análise digital, existem as técnicas de classificação pelos métodos: supervisionado, não supervisionado e híbrido (NOVO, 1992).

A classificação de imagem consiste em determinar uma regra de decisão para que as células da imagem (pixels) possam ser atribuídas a classes de ocupação, possibilitando um processo de mapeamento menos subjetivo e permitindo repetições (NOVO, 1992).

Sendo assim, este trabalho tem por objetivo analisar os classificadores: máxima verossimilhança, distância de Mahalanobis e árvore de decisão na discriminação das classes de ocupação: floresta, soja, arroz, solo exposto, campo e água. Para tanto, foram avaliados os valores de índice kappa, qualidade do índice kappa e exatidão global.

2. MATERIAL E MÉTODOS

2.1 Área de estudo

A bacia hidrográfica do rio Vacai Mirim está localizada no sul do Brasil, no centro do Estado do Rio Grande do Sul, como ilustra a Figura 1. Tem aproximadamente 115347 hectares e está envolvida pelas coordenadas geográficas 53° 46' 30" e 53° 49' 29" de longitude oeste e 29° 36' 55" e 29° 39' 50" de latitude sul, datum SIRGAS2000. Na cabeceira do rio Vacacai Mirim, localiza-se o reservatório de água do Departamento Nacional de Obras e Saneamento DNOS/CORSAN que é responsável por cerca de 40% do abastecimento público da cidade de Santa Maria. A geomorfologia do local é caracterizada pelo Planalto dos Campos Gerais, com uso da terra predominantemente para o cultivo de soja e campo nativo. Rebordo do Planalto Sul-Riograndense ocupado por floresta de Mata Atlântica, e pela Depressão Central, onde estão as áreas com arroz irrigado, campo nativo e lavouras de soja. Problemas ambientais caracterizam essa bacia, apresentando conflitos de uso da água, taxas de urbanização, áreas com alta degradação ambiental.

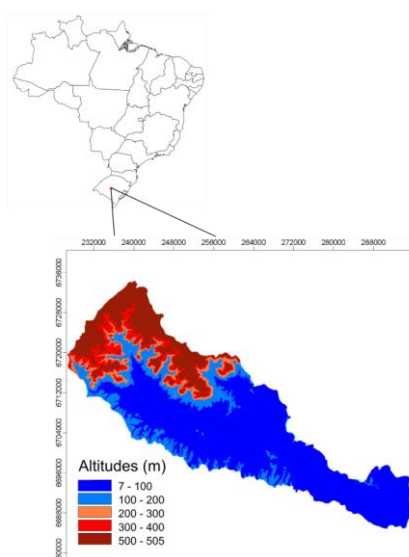


Figura 1 – Mapa de localização da área de estudo, coordenada UTM, Fuso: 22, datum SIRGAS2000.

Para o desenvolvimento deste trabalho, primeiramente, foram realizadas duas visitas à área de estudo, registrando a campo o uso da terra a partir de fotos horizontais georreferenciadas, como representado na Figura 2.

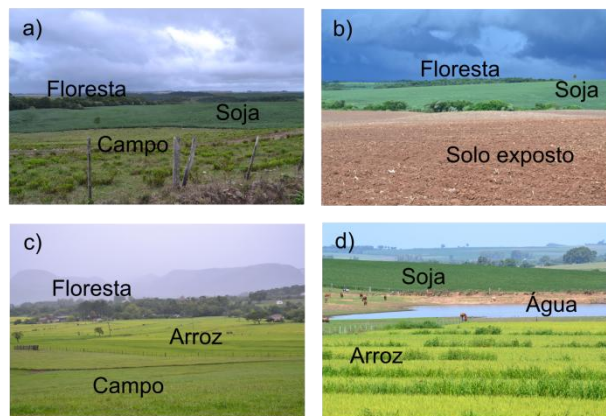


Figura 2 – Fotos horizontais das classes de uso da terra. a) e b) Classes do Planalto dos Campos Gerais. c) e d) Classes da Depressão.

2.2 Classificadores de imagem orbital

2.2.1 Máxima verossimilhança

O algoritmo máxima verossimilhança avalia a variância e a covariância das classes determinadas na área de treinamento para classificar um pixel desconhecido. É assumido que essas classes possuem distribuição normal, ou seja, Gaussiana. Com isso, os padrões espectrais dessas classes podem ser descritos por um valor de média e pela matriz de covariância. Assim, a partir desses parâmetros, é possível calcular a probabilidade estatística de o pixel pertencer a uma das classes de uso da terra (LILLESAND; KIEFFER, 2004).

2.2.2 Distância de Mahalanobis

A distância de Mahalanobis possui algumas semelhanças com o classificador máxima verossimilhança, pois nesse algoritmo também assume-se a hipótese que os dados possuem uma distribuição normal. Entretanto, essa distância admite que todas as covariâncias das classes sejam iguais, apenas diferenciadas pelo vetor de médias. Além disso, para o pixel pertencer a uma determinada classe, ele deve possuir a menor distância de Mahalanobis (RICHARDS, 1986).

2.2.3 Árvore de decisão

A técnica por árvore decisão se distingue das anteriores devido à possibilidade de utilizar imagens com distribuições não gaussianas e que possuem ruídos, também pode adotar-se tipos de dados diferentes dos numéricos (QUINLAN, 1993). A estrutura da árvore de decisão é composta por um nó principal, que apresenta

uma única classe, um nó interno (nó-decisão) e a partir disso tem-se os nós filhos. Esses nós são teste de condições e, assim, podem ser implementados em programas de Sistemas de Informações Geográficas (SIG).

Na sequência, o estudo foi dividido em três etapas: (i) preparações das fontes de dados com a aquisição da imagem orbital e do modelo digital do terreno (MDT), (ii) produção das variáveis de superfície, (iii) a criação e acurácia do mapa de uso da terra.

2.3 Preparações das fontes de dados com a aquisição da imagem orbital e do MDT.

A imagem orbital utilizada foi a do satélite RESOURCESAT-1, sensor LISS III, disponibilizada no catálogo de imagens do Instituto Nacional de Pesquisas Espaciais (INPE) (<http://www.dgi.inpe.br/CDSR/>). A cena utilizada foi a de 12 de fevereiro de 2010. O motivo da escolha dessa cena é devido à resolução espacial de 23.5 metros, cobertura total e poucas nuvens sobre a área de estudo. Essa imagem sofreu correção geométrica e da atmosfera.

Na construção do MDT, foram utilizadas as curvas de nível na escala 1:50000 das cartas topográficas do Serviço Geográfico Brasileiro (SGB). No programa SAGA-GIS, as curvas de nível foram interpoladas pela técnica *Thin Plate Spline (TIN)* e salva com a mesma resolução espacial da imagem orbital (<http://www.saga-gis.org/>).

2.4 Produção das variáveis de superfície

A imagem utilizada foi composta pelas bandas do vermelho (0.62-0.68 μm - VERM), infravermelho próximo (0.77-0.86 μm - IVP) e infravermelho médio (1.55-1.70 μm - IVM).

A partir das bandas da imagem, foi gerado o índice de vegetação por diferença normalizada (NDVI), expressa na Equação 1, e o índice de vegetação ajustado ao solo (SAVI) ilustrada na Equação 2.

Apesar de o NDVI ser um índice bastante usado, ele apresenta algumas limitações referentes à interferência devido à cor do solo e dos efeitos de umidade (JENSEN, 2009).

$$\text{NDVI} = \frac{\text{IVP} - \text{VERM}}{\text{IVP} + \text{VERM}} \quad (1)$$

Em virtude disso, foi calculado o SAVI que utiliza a constante L, um fator de ajuste de subtrato do dossel. O Huete (1988) recomenda o valor de L = 0.5, pois engloba uma maior variação de condições de vegetação.

$$\text{SAVI} = \frac{(1+L) * (\text{IVP} - \text{VERM})}{\text{IVP} + \text{VERM} + L} \quad (2).$$

A partir do MDT, foram gerados os atributos de terreno: elevação (ELEV) e declividade (DECL). Justificado pela relação de suas características de distribuição espacial com os cultivos de soja e arroz. As

áreas de várzeas encontram-se nas baixas declividades, e o cultivo de soja encontra-se nas áreas de maior elevação.

2.5 Criação e acurácia do mapa de uso da terra.

A área de treinamento foi gerada no programa QGIS 1.7.0 pelas ferramentas de edição (SHERMAN et al., 2011). As classes de uso e cobertura da terra vetorizadas na imagem foram: água, solo exposto, arroz, soja, campo e floresta, obtendo as respectivas áreas 136.4 ha, 203.7 ha, 118.3 ha, 118.7 ha, 289.3 ha. Tomou-se o cuidado em coletar amostras puras e distantes das bordas das feições.

As classificações pelos métodos distância de Mahalanobis e máxima verossimilhança foram obtidas pelo módulo *Imagery-Classification* do SAGA-GIS. Para cada técnica, foram efetuadas duas classificações. Na primeira foram inseridas apenas as três bandas, e na segunda, além das três bandas, foram informados os atributos do terreno e os índices NDVI e SAVI.

Na classificação por árvore de decisão, o primeiro passo foi extrair os valores das variáveis de superfície com área de treinamento, essa área é a mesma utilizada para as classificações anteriores. Esse procedimento foi executado com a opção *Grid Values to Shapes* do SAGA-GIS, possibilitando armazenar os valores das variáveis VER, IVP, IVM, NDVI, SAVI, ELEV e DECL em um novo arquivo vetorial. A tabela desse arquivo foi exportada para o formato texto separado por vírgula, contendo 20815 linhas e oito colunas.

A montagem da árvore de decisão foi realizada no programa de mineração de dados Weka 3.7 (<http://www.cs.waikato.ac.nz/ml/weka/>). O algoritmo árvore de decisão especificado no programa foi o J48. O teste de acurácia e índices estatísticos da árvore foram dados pelo conjunto de treinamento. Também foi selecionada a opção verdadeira para a poda da árvore (*reducedErrorPruning=True*). Essas configurações foram necessárias para diminuir a extensão e complexidade da árvore.

Para converter a AD em um formato possível de ser interpretado pelo SAGA-GIS, utilizou-se o programa ADtoSIG (RUIZ et al., 2011). O arquivo de saída do ADtoSIG foi usado na ferramenta *Grid Calculator* do SAGA-GIS, gerando a classificação da imagem por árvore de decisão. Esses procedimentos foram efetuados para duas classificações, a primeira que contém todos os atributos, e a segunda apenas as bandas da imagem. Nesse caso, a escolha de qual atributo utilizar foi feita no próprio programa Weka.

Para analisar a concordância dos métodos de classificação, foram gerados 500 pontos aleatórios sobre a imagem. Em cada ponto verificou-se o uso da terra, assim originando a verdade terreno. A verdade terreno foi confrontada com os métodos de classificações, gerando o índice Kappa. Também foi calculado o índice de exatidão global para as seis classificações, esses cálculos foram efetuados na linguagem de programação Python através das bibliotecas GDAL e Numpy.

(<http://pypi.python.org/pypi>). O nível de exatidão do índice kappa foi estabelecido conforme a Tabela 1.

Tabela 1: Classificação da exatidão, conforme o valor do índice kappa (LANDIS; KOCH, 1977).

Índice Kappa (K)	Qualidade
$K = 0,2$	Ruim (R)
$0,2 < K = 0,4$	Razoável (Ra)
$0,4 < K = 0,6$	Bom (B)
$0,6 < K = 0,8$	Muito Bom (MB)
$K = 0,8$	Excelente (E)

3 RESULTADOS E DISCUSSÃO

3.1 Tabela com as áreas das classes para cada classificador

Na Tabela 2 estão os valores das áreas de uso em porcentagem para cada método de classificação. A última coluna contém o cálculo do desvio padrão (DP) das áreas de cada classe. O uso área urbana e sombra/nuvem estão com valores zeros porque foram vetorizados sobre a imagem, assim não havendo diferença nas áreas. As classes soja e água foram as que tiveram menores valores de DP, sugerindo que os classificadores quantificaram satisfatoriamente essas classes. Já os usos da terra floresta e arroz tiveram suas áreas superestimadas, como se observa na Figura. 3, para a classe floresta (3c e 3e) e para classe a classe arroz (3d e 3f). Desse modo, os resultados variaram e, consequentemente, foram obtidos valores DP mais altos. No uso campo foi observado o oposto, pois tais áreas foram subestimadas pelos métodos de classificação, resultando em um valor de DP de 6.92 %.

Tabela 2: Áreas em porcentagem e o desvio padrão entre elas para as imagens temáticas resultante.

Usos da terra	A	B	C	D	E	F	DP
Soja	11.11	10.52	6.13	7.7	7.09	8.34	1.96
Arroz	8.4	8.47	6.64	13.76	10.46	15.38	3.40
Solo exposto	9.4	10.5	11.07	22.47	10.42	15.32	4.99
Água	1.36	1.51	1.18	1.26	1.28	1.44	0.12
Floresta	25.37	27.29	45.7	27.95	37.6	26.77	8.11
Campo	38.45	35.8	23.37	20.95	27.24	26.84	6.92
Área urbana	1.95	1.95	1.95	1.95	1.95	1.95	0.00
Nuvem/sombra	3.96	3.96	3.96	3.96	3.96	3.96	0.00
Total (%)	100	100	100	100	100	100	

A:AD, B: AD bandas, C: MAHA, D: MAHA banda, E: MaxVer, F: MaxVer bandas.

3.2 Avaliações do índice kappa, exatidão global e qualidade do kappa.

A exatidão dos mapas resultantes foi comprovada pelo índice kappa, exatidão global e a qualidade dos

índices kappa. Uma primeira análise possível é que, quando os classificadores utilizaram apenas as bandas da imagem, seus resultados foram menores ou iguais aos valores dos mapas com todas as variáveis. Outro resultado foi que o classificador AD utilizando as bandas resultou em um melhor índice kappa em relação aos classificadores MAHA e MaxVer, quando usaram todas as variáveis ou mesmo utilizando só as bandas da imagem. E, na classificação por AD, a partir das variáveis calculadas foi obtido o melhor valor de índice kappa e, consequentemente, maior valor para a exatidão global. As qualidades dos índices kappa, representado na linha três da Tabela 3, foram de excelente a muito bom, mas, quando se visualiza a Figura. 3, percebe-se grandes diferenças entre os mapas classificados.

Tabela 3: Valores dos índices de exatidão

	A	B	C	D	E	F
Kappa(%)	86.52	81.82	68.15	68.90	75.06	72.20
Exatidão Global(%)	88.82	85.03	74.25	74.25	79.64	77.05
Qualidade	E	E	MB	MB	MB	MB

A:AD, B: AD bandas, C: MAHA, D: MAHA banda, E: MaxVer, F: MaxVer bandas.

Resultado semelhante de índice kappa para o método árvore de decisão foi encontrado por Figueiredo e Carvalho (2006), Tais autores, avaliaram a exatidão do mapeamento do uso da terra em Capixaba no estado do Acre, utilizando as variáveis das bandas do sensor TM/Landsat 5, imagens fração derivadas do modelo de mistura espectral e o NDVI. As classes utilizadas foram floresta, capoeira, pasto alto, pasto baixo, solo e água, com resultados alcançando valores de índices kappa de 90% para o método árvore de decisão, 89% para o classificador MaxVer e 79% para o ISODATA.

3.3 Análise visual dos mapas temáticos

Analisando a Figura 3, observam-se áreas na depressão central com solo exposto, onde, para o mês da cena, deveria existir lavouras de arroz. No entanto, como a região sofreu fortes precipitações, que ocasionaram um acúmulo de sedimentos nas partes mais baixas do relevo, essas as áreas de arroz foram classificadas como solo exposto.

Na área do Planalto dos Campos Gerais, as classificações MaxVer e MAHA realizadas somente com as bandas da imagem superestimaram as classes arroz em áreas de floresta e campo. Isto ocorreu devido ao sombreamento do relevo, fazendo com que as respostas espectrais dessas classes apresentassem um comportamento espectral semelhante ao de lavouras de arroz.

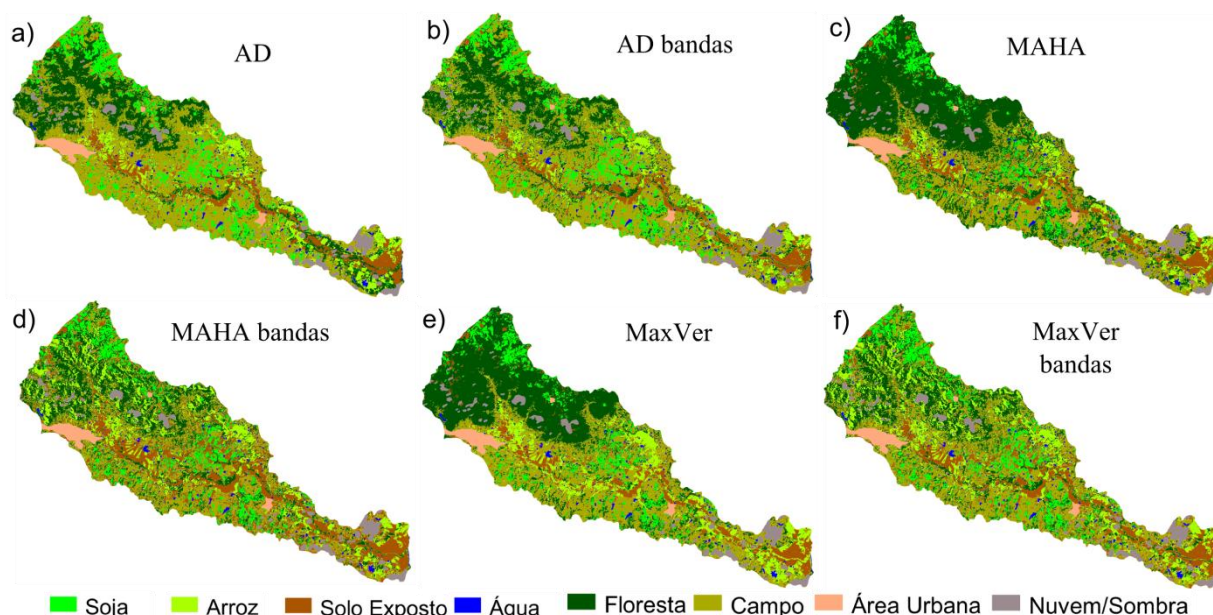


Figura 3 – Resultados dos mapas temáticos dos diferentes métodos de classificação.

4. CONCLUSÃO

A classificação de imagem orbital com o melhor resultado foi o método por árvore de decisão, obtendo o índice kappa 86.52 % e uma exatidão global de 88.82 %. Outro resultado relevante é que, quando os classificadores utilizaram todas as variáveis para discriminar as classes, os valores dos índices de concordância tiveram melhores resultados. Quando analisadas as imagens temáticas resultantes dos diferentes métodos, observa-se que houve grandes diferenças, mas o índice de qualidade do kappa não apresenta bom desempenho para trazer essa informação devido a categorizar a qualidade do kappa por intervalos.

AGRADECIMENTOS

Ao CNPq pela bolsa de iniciação científica, bolsa PQ concedida aos autores e aporte financeiro via Edital Universal. Laboratório de Sensoriamento Remoto Geológico do Programa Pós-Graduação em Sensoriamento Remoto da UFRGS pelo apoio.

REFERÊNCIAS

BARBOSA, A. P. **Comparação de métodos de classificação de imagens, visando o gerenciamento de áreas citrícolas**. Botucatu, 2009. 65 f. Dissertação (Mestrado em Agronomia/Energia na Agricultura) – Faculdade de Ciências Agrônômicas, Universidade Estadual Paulista.

FIGUEIREDO, S. M. de M.; CARVALHO, L. M. T. de. **Avaliação da exatidão do mapeamento da cobertura da terra em Capixaba, Acre, utilizando classificação por árvore de decisão**. Cerne, v. 12, n. 1, 38-47 p., jan./mar. 2006.

HUETE, A.R., **A Soil-adjusted Vegetation Index (SAVI)**, Remote Sensing Environ, 25: 295-309 p, 1988.

JENSEN, J. R. **Sensoriamento Remoto do Ambiente: uma perspectiva em recursos terrestres**. 2ed. São José dos Campos: Parêntese, 604 p, 2009

LILLESAND, T. M.; KIEFER, R. W. **Remote sensing and image interpretation**. 2. ed. Chichester: John Wiley & Sons, 750 p., 1994.

NOVO, E. M. L. M. **Sensoriamento remoto: princípios e aplicações**. Ed. São Paulo: Edgard Blücher Ltda, 307 p, 1992

RUIZ, L. F. C.; ten CATEN, A.; DALMOLIN, R. S. D. **ADtoSIG Software de Conversão**. Universidade Federal de Santa Maria, v.1, 2011

WEKA, disponível em: <<http://www.cs.waikato.ac.nz/ml/weka/>>. Acesso 10 de março de 2012.

Catálogo INPE disponível em <<http://www.dgi.inpe.br/CDSR/>>. Acesso 30 de março de 2012.

RICHADS, J. A. **Remote sensing digital image analysis: an introduction**. Berlin: Springer-Verlag, 1986. 281 p.
SAGA – GIS disponível em <<http://www.saga-gis.org/>>. Acesso 3 de abril de 2012

QUINLAN, J. R. C4.5: **Programs for Machine Learning**. Morgan Kaufmann. California. p235, 1993.

SHERMAN GE, SUTTON T, BLAZEK R, HOLL S, DASSAU O, MORELY B, MITCHELL T AND LUTHMAN L. 2011. **Quantum GIS User Guide – Version 1.7 “Wroclaw”**. Disponível em: http://download.osgeo.org/qgis/doc/manual/qgis-1.7.0_user_guide_en.pdf. Acesso em: 1 de Abril de 2012

Bibliotecas Python disponível em <http://pypi.python.org/pypi>>. Acesso 4 de Abril de 2012.