

USO E COBERTURA DA TERRA APOIADOS EM ALGORITMOS BASEADOS EM APRENDIZADO DE MÁQUINA: O CASO DE MARIANA - MG

Natália Peixoto Gaiad¹; Ana Paula Marques Martins²; Aline Bernarda Debastiani²;
Ana Paula Dalla Corte³; Carlos Roberto Sanquetta³

¹Mestranda do Programa de Pós Graduação em Engenharia Florestal, Universidade Federal do Paraná (npgaiad@gmail.com);

²Doutoranda do Programa de Pós Graduação em Engenharia Florestal, Universidade Federal do Paraná;

³Prof. Dr do Departamento de Ciências Florestais, Universidade Federal do Paraná

Recebido em: 08/04/2017 – Aprovado em: 10/06/2017 – Publicado em: 20/06/2017
DOI: 10.18677/EnciBio_2017A99

RESUMO

Técnicas aplicadas a dados de sensoriamento remoto vêm sendo cada vez mais utilizadas para melhorar a interpretação e análise do uso e cobertura da terra. Uma forma de análise desses dados é a classificação das imagens de satélite, que pode ser dividida em supervisionada e não-supervisionada. O objetivo deste trabalho foi avaliar classificadores supervisionados, como: árvores de decisão (5 algoritmos), Redes Neurais Artificiais (RNA) e *Support Vector Machine* (SVM), bem como, determinar a acuracidade e precisão na detecção das classes de uso e cobertura da terra. A área de estudo foi o município de Mariana, Minas Gerais, local onde ocorreu o rompimento da barragem da mineradora Samarco e com isso o deslizamento de muitos sedimentos. Por meio de uma imagem de satélite da série Landsat 8, sensor OLI, a classificação de imagens foi feita aplicando os seguintes algoritmos de classificação supervisionada: RNA, SVM, bem como os algoritmos de árvores de decisão: J48, Random Tree, Random Forest, REPTree e LMT. O algoritmo que obteve o melhor desempenho foi o SVM, com maior valor de exatidão global (98,323%) e maior índice Kappa (0,979), já que este resulta em uma única solução ótima; com a matriz de confusão foi possível notar que a classe de Sombra e de Lama obtiveram maior erro na classificação, tendo como classes de maior acerto Água e Vegetação.

PALAVRAS-CHAVE: Árvores de decisão, RNA, SVM.

LAND-USE AND LAND-COVER BASED ON MACHINE LEARNING ALGORITHMS : THE CASE OF MARIANA – MG

ABSTRACT

Techniques applied to remote sensing data have been increasingly used to improve analysis and interpretation of land-use and land-cover. Satellite images classification is one of these techniques, it may be supervised or non-supervised. In this paper we evaluated supervised classifiers such as decision trees (five algorithms), Artificial Neural Networks (ANN) and Support Vector Machine (SVM). We also determined the accuracy and precision of the land-use and land-cover class detection. The area

analyzed was in Mariana, Minas Gerais, where a dam belonging to Samarco broke causing mining sediments to spread over the area. We used Landsat 8, sensor OLI, which were classified by the following supervised classification algorithms: ANN, SVM, as well as decision trees (J48, Random Tree, Random Forest, REPTree and LMT). SVM outperformed the other algorithms, presenting the greater global accuracy (98,323%) and higher kappa index (0,979), since it results in an unique optimal solution; with the confusion matrix it was possible to determine that Shadow and Mud were the classes that obtained the highest classification error, and Water and Vegetation as the classes with lower classification error.

KEYWORDS: ANN, SVM, Decision Trees.

INTRODUÇÃO

A tragédia de Mariana está sendo considerada o pior desastre ambiental da história do Brasil, desastre este que gerou uma quantidade enorme de lama de rejeitos deixando várias pessoas desabrigadas, causou mortes, atingiu muitas cidades e danificando totalmente a estrutura do Rio Doce, que deságua no mar do Espírito Santo (MENDONÇA, 2016). Nesse contexto, tecnologias de sensoriamento remoto possuem relevante potencial de aplicação nas atividades de resposta a desastres ambientais, pois, após o impacto, torna-se imprescindível o conhecimento da extensão e quantificação dos danos ocorridos na área afetada para que seja possível realizar ações e mobilização de recursos para mitigação dos mesmos (SAUSEN & LACRUZ, 2015).

Ademais, o Sensoriamento Remoto ganha destaque na avaliação e análise do uso e cobertura da terra, dados orbitais de sensoriamento remoto revolucionaram esses estudos por fornecer informações sintetizadas sobre o uso e cobertura da terra em determinado momento e local (ATTRI et al., 2015). Nesse contexto, a classificação digital de imagens de sensoriamento remoto tem tornado-se uma das informações que permite melhor gestão de áreas afetadas por ações antrópicas drásticas com confiabilidade e rapidez.

Classificação de imagens pode ser definida, segundo KAVZOGLU & COLKESEN (2009), como atribuir a cada pixel da imagem original uma classe previamente definida, de acordo com a reflectância espectral dos objetos. Os métodos de classificação podem ser divididos, basicamente, em supervisionados e não supervisionados. O método supervisionado pode ainda ser dividido entre paramétricos e não paramétricos.

Os classificadores paramétricos, no entanto, necessitam do conhecimento prévio da distribuição estatística dos dados a classificar, o que na prática, é difícil de atingir em muitos casos (PETROPOULOS et al., 2012). Como uma alternativa, algoritmos não-paramétricos vêm sendo utilizados para superar esse problema, dentre os quais pode-se citar as Redes Neurais Artificiais (RNAs) e o *Support Vector Machine* (SVM) (PEREIRA & CENTENO, 2013). Essas técnicas, também conhecidas como algoritmos baseados em aprendizado de máquina (AM), na qual inclui as árvores de decisão, vêm sendo muito investigadas na atualidade e têm sido consideradas por muitos pesquisadores como técnicas promissoras para a classificação digital de imagens (ANDRADE et al., 2014).

Diante do exposto e tendo em vista a necessidade de classificação do uso da terra e detectar a técnica com maior acurácia e precisão, o objetivo deste estudo foi avaliar o desempenho de árvores de decisão (5 algoritmos) frente as Redes

Neurais Artificiais (RNA) e *Support Vector Machine* (SVM). Também fez parte do trabalho uma análise acerca das maiores confusões espectrais na classificação.

MATERIAL E MÉTODOS

Caracterização da área de estudo

A área de estudo corresponde ao subdistrito de Bento Rodrigues, um dos locais mais afetados após o rompimento da barragem da mineradora Samarco em 2015. É um subdistrito de Santa Rita Durão, no município de Mariana, Minas Gerais. Bento Rodrigues está localizado a 35 km do centro de Mariana e a 124 km de distância da capital do Estado, Belo Horizonte. O município de Mariana situa-se na região sudeste do Estado de Minas Gerais, latitude 20°22'40" S e longitude 43°24'58" W, altitude de 712 m e área de 1196,7 km² (Figura 1).

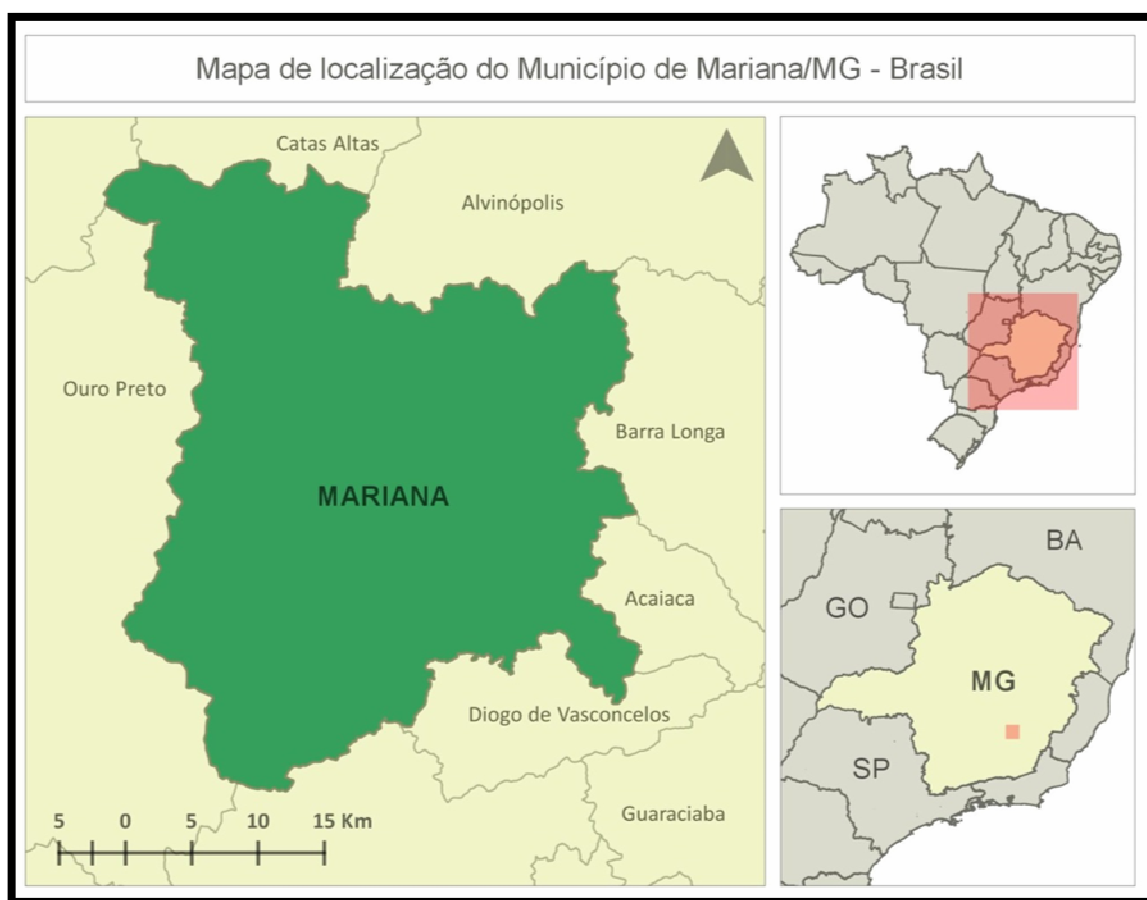


FIGURA 1. Localização do município de Mariana, onde se situa o subdistrito de Bento Rodrigues.

Fonte: Os autores, (2017).

O subdistrito caracteriza-se pela intensa atividade de extração mineral e consequência disso, estão ali localizadas as barragens de rejeitos de mineração denominadas Fundão e de Santarém, operadas pela empresa mineradora Samarco. Após o rompimento da barragem do Fundão no dia 5 de novembro de 2015 houve uma mudança drástica da paisagem local, atingindo centenas de moradores das vilas de Bento Rodrigues e Paracatu de Baixo conforme notícias divulgadas na mídia.

Obtenção da imagem Landsat

Foi utilizada uma imagem da série Landsat 8, sensor OLI (*Operational Land Imager*), datada de 31 de janeiro de 2016 (três meses após o rompimento), bandas 1, 2, 3, 4, 5, 6, 7 e 9, tendo órbita 217 e ponto 74. A imagem adquirida é do catálogo do Serviço Geológico dos Estados Unidos (<http://earthexplorer.usgs.gov/>). A imagem foi inicialmente convertida em radiância no topo da atmosfera para posterior correção atmosférica pelo módulo FLAASH (*Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes*) para eliminar os efeitos do vapor d'água, oxigênio, espalhamento de aerossol e outros efeitos adjacentes. Foi realizado o recorte da imagem visando contemplar apenas a área de interesse a ser estudada. Todo o processamento foi desenvolvido no software ENVI 5.2.

Classificação de imagens

Para a etapa de classificação de imagem, iniciou-se a coleta de polígonos de amostras das áreas de interesse dentro da imagem, denominados ROIs. Foram obtidas sete classes de usos da terra, sendo: Sombra, Nuvem, Rocha, Vegetação, Lama, Solo/Mineração e Água.

Essas classes de uso da terra foram interpretadas visualmente por meio da resposta espectral de cada uma, adicionadas pela informação da coloração e forma e também com a ajuda de imagens acessadas no Google Earth, possibilitando maior eficiência na identificação das mesmas, principalmente envolvendo mineração, lama, água e rocha; sendo que a primeira era visualizada como uma grande porção de área limpa com diferentes níveis de altitude. Lama e água foram diferenciadas na obtenção de polígonos em que a água se caracterizava por uma coloração bem escura, em sua forma mais limpa, tendo sua resposta espectral quase preta, se distinguindo da cor marrom avermelhada da lama. As outras classes seguiram estes métodos para a classificação das mesmas.

Os algoritmos de classificação supervisionada avaliados foram: *Support Vector Machine* (SVM), Redes Neurais Artificiais (RNA), e os algoritmos de árvores de decisão: J48, Random Tree, Random Forest, REPTree e LMT; sendo essas classificações realizadas com todas as bandas.

O método de classificação pelo algoritmo SVM, segundo NASCIMENTO et al. (2009), é um classificador que visa a determinação de limites de decisão com o intuito de permitir a separação ótima entre as classes por meio da minimização dos erros. Este algoritmo é considerado um classificador supervisionado e não paramétrico.

RNA's podem ser definidas como técnicas computacionais que possuem um modelo matemático inspirado em neurônios de organismos inteligentes que adquirem informações por meio dos treinamentos efetuados no decorrer do processo de condução (MOREIRA et al., 2013). Uma vantagem do uso dessa técnica é a possibilidade de manipular grandes quantidades de dados que não assumem uma distribuição normal (ATKINSON & TATNALL, 1997). A RNA utilizada foi constituída de apenas uma camada oculta e taxa de aprendizado definida como 0,01.

Outros métodos de classificação utilizados, além dos classificadores supervisionados, foram os algoritmos de árvore de decisão. Esses algoritmos resultam em classificadores que podem prever e revelar classes baseadas em valores dos atributos de um conjunto de dados. Tem como resultado árvores de

classificação por meio de um conjunto de dados de treinamento, subdividindo as amostras em subconjuntos de acordo com a classe de cada um (QUINLAN, 1993).

Ainda de acordo com QUINLAN (1993), o algoritmo J48 tem como critério que, a cada nó, é escolhido um atributo que melhor subdivide as amostras em conjuntos homogêneos e que a escolha do atributo para a subdivisão é o ganho de informação obtida. Foi utilizado o critério de poda da árvore para que essa não fosse demasiadamente grande. Já para o algoritmo Random Tree, o mesmo leva em consideração um número x de atributos, aleatoriamente escolhidos em cada nó, onde esta classificação não faz cortes na árvore de decisão (GIASSON et al., 2013).

Segundo NETO (2014), o algoritmo Random Forest segue um objetivo diferente dos outros comumente utilizados. Este tem como finalidade originar várias árvores de decisão por meio de conjuntos de atributos escolhidos de forma aleatória em relação ao conjunto original. Ainda de acordo com esse autor, estes atributos finais são denominados como uma amostragem de *bootstrap*, amostragem essa que há reposição, melhorando a interpretação dos dados. A partir da quebra desses dados e a formação de novos subconjuntos é que se dá a formação da árvore de decisão (LORENZETT & TELÖCKEN, 2016).

Sobre o algoritmo REPTree, ainda segundo GIASSON et al., (2013), essa árvore de decisão criada usa informações de ganho e variância, sendo alterada de acordo com a poda de erro reduzido. O último algoritmo de árvore de decisão utilizado para avaliar o desempenho dos classificadores foi o LMT (*Logistic Model Trees*). Esse, de acordo com FRIZZARINI & LAURETTO (2013), tem como objetivo criar árvores baseadas em uma regressão logística com o intuito de selecionar um atributo que seja relevante dentro dos dados existentes, sendo este um processo de repetição.

Para avaliar os algoritmos de classificação supervisionada, foram obtidos o índice Kappa (K) e o índice de Exatidão Global (G), descritos nas Equações 1 e 2, respectivamente.

$$G = \frac{\sum_{i=1}^M n_{ii}}{N} = P_0 \quad (1)$$

$$K = \frac{P_0 - P_e}{1 - P_e} \quad (2)$$

Onde: n_{ii} = Número de pixels da diagonal principal da Matriz de Confusão de cada classe; N = Número de classes; P_0 = Concordância total ou Exatidão global; P_e : concordância ao acaso esperada; P_r : concordância esperada. Foi utilizada a estatística Z à 5% de probabilidade de erro para testar a significância estatística de cada classificador e da diferença entre os mesmos (CONGALTON & GREEN, 1998) (Equação 3).

$$Z = \frac{K_2 - K_1}{\sqrt{\sigma^2 K_2 + \sigma^2 K_1}} \quad (3)$$

Onde: K_1 : índice Kappa da imagem 1; K_2 : índice Kappa da imagem 2; σ^2 : variância do índice Kappa (Equação 4).

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

Em que:

$$\theta_1 = \frac{1}{X} \sum_{i=1}^r X_{1+} X_{1+}; \quad \theta_2 = \frac{1}{X^2} \sum_{i=1}^r X_{1+} X_{1+};$$

$$\theta_3 = \frac{1}{X^2} \sum_{i=1}^r X_{ij} (X_{1+} X_{+1}); \quad \theta_4 = \frac{1}{X^3} \sum_{i=1}^r \sum_{j=1}^r X_{ij} (X_{j+} X_{+1})^2$$

RESULTADOS E DISCUSSÃO

Na Tabela 1, têm-se os valores obtidos para os índices de Exatidão Global, índice Kappa e variância do Kappa para todos os classificadores utilizados e seguidos do *ranking* de acordo com o melhor desempenho. O índice Kappa de todos os classificadores utilizados foi superior a 0,91, resultados estes considerados como excelentes ao considerar a classificação de LANDIS & KOCH (1977).

É notável que dentre os algoritmos utilizados para classificação da imagem Landsat 8, o algoritmo SVM foi o que resultou em um melhor desempenho, obtendo o maior valor de exatidão global (98,323%), maior índice Kappa (0,979) e menor variância do Kappa (0,0000153).

TABELA 1. Resultados obtidos por meio da classificação Supervisionada com diferentes classificadores.

Classificadores	Exatidão Global	Índice Kappa	Variância do Kappa	Rank
SVM	98,323	0,9790 [*] A	0,0000153	1
RNA	97,305	0,9660 [*] B	0,0000243	2
LMT	95,329	0,9420 [*] C	0,0000407	3
Random Forest	95,269	0,9415 [*] C	0,0000410	4
Random Tree	94,551	0,9330 [*] CD	0,0000468	5
REPTree	93,952	0,9250 [*] DE	0,0000518	6
J48	92,814	0,9113 [*] E	0,0000598	7

^{*}significativo a 5% de probabilidade de erro. Letras iguais na vertical são consideradas significativamente iguais a 5% de probabilidade de erro.

Todos os classificadores são significativamente maiores que zero de acordo com a estatística Z. Ainda de acordo com essa estatística, o índice Kappa do algoritmo SVM é significativamente diferente dos demais algoritmos, apresentando-se superior por resultar no maior Kappa e menor variância, sendo seguido pela RNA e posteriormente pelos algoritmos de árvore de decisão: LMT, Random Forest e Random Tree. Dentre esses algoritmos de árvore de decisão mencionados, o Random Tree pode ser considerado estatisticamente igual ao algoritmo REPTree, que por sua vez é igual ao algoritmo J48, o qual propiciou o pior resultado quanto a todos os parâmetros de desempenho.

O resultado do algoritmo SVM como melhor desempenho pode ser atribuído ao fato de que o mesmo resulta em uma única solução ótima, enquanto que a RNA, algoritmo seguido de melhor desempenho, pode apresentar muitas soluções de acordo com os pesos e configurações iniciais do processamento (BURGES, 1998). Esse resultado poderia ser modificado se o enfoque for encontrar uma configuração

ótima dos parâmetros da RNA, fazendo com que este último resulte no melhor classificador. De acordo com HUANG et al. (2002), este apresenta uma comparação do algoritmo SVM com outros classificadores, tais quais MAXVER, redes neurais e árvores de decisão, finalizando com o resultado de que o SVM possui altos níveis de robustez e acurácia comparado a todos os outros.

Corroborando com os resultados obtidos, WERE et al. (2015) utilizaram os algoritmos SVM para regressão, Random Forest e RNA para o mapeamento de estoques de carbono orgânico no solo, em uma Reserva Florestal, Quênia, África, e concluíram que o SVM se mostrou superior aos demais. O Random Forest apresentou resultado de 95,27% de exatidão global, e 0,941 de índice Kappa, sendo estes valores superiores aos resultados relatados por TATSUMI et al., (2015), obtendo valores de exatidão global de 81% e índice Kappa de 0,70, que usou este algoritmo para a classificação do uso da terra no sul do Peru e apontou como um resultado de alta performance.

Na Tabela 2 constam os valores da matriz de confusão para o algoritmo de melhor desempenho. A matriz de confusão é uma forma de representar as condições entre a verdade de campo e o resultado do processo de classificação, ou seja, as linhas correspondem à classe correta e as colunas à classe estimada no processo de classificação (CONGALTON, 1991).

TABELA 2. Matriz de confusão do algoritmo SVM.

	Sombra	Nuvem	Rocha	Vegetação	Lama	Solo/Mineração	Água
Sombra	212 (94,64%)	0 (0%)	0 (0%)	3 (0,55%)	0 (0%)	0 (0%)	0 (0%)
Nuvem	0 (0%)	212 (99,07%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Rocha	0 (0%)	0 (0%)	225 (98,68%)	0 (0%)	0 (0%)	3 (1,10%)	0 (0%)
Vegetação	12 (5,36%)	1 (0,47%)	3 (1,32%)	538 (99,45%)	0 (0%)	0 (0%)	0 (0%)
Lama	0 (0%)	0 (0%)	0 (0%)	0 (0%)	161 (96,99%)	0 (0%)	0 (0%)
Solo/Mineração	0 (0%)	1 (0,47%)	0 (0%)	0 (0%)	5 (3,01%)	269 (98,90%)	0 (0%)
Água	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	25 (100%)
Total	224 (100%)	213 (100%)	228 (100%)	541 (100%)	166 (100%)	272 (100%)	25 (100%)

De acordo com a matriz de confusão da classificação com o algoritmo SVM, a classe sombra apresentou maior confusão entre as classes utilizadas, sendo que 94,64% dos pixels foram corretamente classificados, ou seja, correspondem realmente a essa classe, sendo observada confusão apenas com a classe de vegetação (5,36%). Esse fato pode ser justificado pelo relevo declivoso do local, o que gera sombra nas encostas das montanhas, as quais geralmente são cobertas por vegetação.

Para a classe Nuvem, 99,07% dos pixels foram classificados corretamente, apresentando apenas uma pequena confusão (1 pixel cada) com as classes Vegetação e Solo/Mineração. Já para a classe Rocha, o algoritmo propiciou 98,68% de acerto, apresentando confusão apenas com a classe Vegetação (1,32%). O algoritmo SVM classificou corretamente 99,45% da classe de vegetação, apresentando 0,55% de confusão com a classe Sombra. A classe Lama foi classificada corretamente em 96,99% dos pixels, apresentando confusão apenas com a classe Solo/Mineração, o que é compreensível devido ao fato da lama ser proveniente de rejeitos da própria mineração. A classe Solo/Mineração foi classificada erroneamente em 3 pixels (1,1%), obtendo, conseqüentemente, 98,9%

de acerto do algoritmo SVM. Apenas a classe água obteve 100% de acerto na classificação, sendo isso explicado pelo reduzido número de pixels coletados para a classificação desta classe (25 pixels).

Em relação aos algoritmos SVM e RNA, estes exigem um maior tempo de processamento e são considerados sistemas de “caixas pretas”, devido à dificuldade na interpretação de como encontraram determinados resultados (RUIZ et al., 2014). Ao contrário disso, tem-se as árvores de decisão que possibilitam uma melhor interpretação dos resultados, já que estas constroem árvores com as regras da classificação.

CONCLUSÃO

Diante dos resultados obtidos, foi permitido constatar que todos os algoritmos utilizados apresentaram desempenho considerado bom. O melhor desempenho foi observado para o algoritmo SVM (índice Kappa de 0,979 e exatidão global de 98,323%), seguido da RNA e das árvores de decisão. Entre as árvores de decisão, o algoritmo LMT foi superior, seguido do Random Forest e Random Tree, esses não diferindo significativamente.

Em relação a confusão espectral dos alvos na matriz de confusão do algoritmo SVM, frente à classificação, foi possível detectar maior quantidade de erros nas classes de Sombra e Lama, e aquelas com maiores acertos foram Água, seguido de Vegetação.

Apesar do algoritmo SVM apresentar melhor desempenho, com a utilização das árvores de decisão, o problema da confusão espectral poderia ser minimizado, já que estas são construídas por regras de fácil interpretação, e com isso possível de mudanças na coleta e organização dos dados utilizados para o treinamento destes algoritmos.

REFERÊNCIAS

ANDRADE, A. C; FRANCISCO, C. N.; ALMEIDA, C. M. Desempenho de classificadores paramétrico e não paramétrico na classificação da fisionomia vegetal. **Revista Brasileira de Cartografia**, v. 66, n.2, p. 349-363, 2014.

ATKINSON, P. M.; TATNALL, A. R. L. Introduction: neural networks in remote sensing. **International Journal of Remote Sensing**, v.18, n.4, p.699- 709, 1997. Disponível em: <<http://dx.doi.org/10.1080/014311697218700>>. doi: 10.1080/014311697218700

ATTRI, P.; CHAUDHRY, S.; SHARMA, B. Remote Sensing & GIS based Approaches for LULC Change Detection – A Review. **International Journal of Current Engineering and Technology**, v.5, n.5. 2015.

BURGES, C. J. C. **A Tutorial on Support Vector Machines for Pattern Recognition**. Data Mining and Knowledge Discovery, 2, p.121-167, 1998. Disponível em: <<http://dx.doi.org/10.1023/A:1009715923555>>. doi: 10.1023/A:1009715923555

CONGALTON, R. G. A review of assessing the accuracy of classifications of remotely sensed data. **Remote Sensing of Environment**, v. 49, n.12, p. 1671-1678, 1991. Disponível em: <[https://dx.doi.org/10.1016/0034-4257\(91\)90048-b](https://dx.doi.org/10.1016/0034-4257(91)90048-b)>. doi: 10.1016/0034-4257(91)90048-b

CONGALTON, R. G.; GREEN, K. **Assessing the accuracy of remotely sensed data: principles and practicecs**. New York: Lewis Publishers, 1998. 137p.

FRIZZARINI, C.; LAURETTO, M. S. **Proposta de um Algoritmo para indução de árvores de classificação para dados desbalanceados**. Escola de Artes, Ciências e Humanidades – Universidade de São Paulo (EACH-USP). São Paulo, 2013.

GIASON, E.; HARTEMINK, A. E.; TORNQUIST, C. G.; TESKE, R.; BAGATINI, T. Avaliação de cinco algoritmos de árvores de decisão e três tipos de modelos digitais de elevação para mapeamento digital de solos a nível semidetalhado na Bacia do Lageado Grande, RS, Brasil. **Ciência Rural**, Santa Maria, v.43, n.11, p.1967-1973, nov/2013. Disponível em: <<http://dx.doi.org/10.1590/S0103-84782013001100008>>. doi: 10.1590/S0103-84782013001100008

HUANG, C.; DAVIS, L. S.; TOENSHEND, J. R. G. An assessment of support vector machines for land cover classification. **International Journal of Remote Sensing**, v. 23, n. 4, p. 725-749, 2002. Disponível em: <<http://dx.doi.org/10.1080/01431160110040323>>. doi: 10.1080/01431160110040323

KAVZOGLU, T., COLKESEN, I., A Kernel Functions Analysis for Support Vector Machines for Land Cover Classification. **International Journal of Applied Earth Observation and Geoinformation**, 11, p. 352–359, 2009.

LANDIS, J.; KOCH, G. G. The measurements of agreement for categorical data. **Biometrics**, v.33, n.3, p.159-179, 1977. Disponível em: <<http://dx.doi.org/10.2307/2529310>>. doi: 10.2307/2529310

LORENZETT, C. D. C; TELÖCKEN, A. V. **Estudo comparativo entre os algoritmos de Mineração de Dados Random Forest e J48 na tomada de decisão**. Curso de Ciência da Computação. Universidade de Cruz Alta (UNICRUZ). Rio Grande do Sul, 2016.

MENDONÇA, H. Desastre Mariana. Em: El País. Disponível em: <http://brasil.elpais.com/tag/desastre_mariana/a>. Acesso em 10 de março de 2017.

MOREIRA, G. F.; FERNANDES, R. B. A.; FERNANDES FILHO, E. F.; VIEIRA, C. A. O.; SANTOS, K. A. Classificação automatizada do uso e cobertura do solo usando imagens Landsat. **Revista Brasileira de Geografia Física**, v. 06, n.1, p.058-065. 2013.

NASCIMENTO, R. F. F.; ALCÂNTARA, E. H.; KAMPEL, M.; STECH, J. L.; NOVO, E. M. L. M.; FONSECA, L. M. G. O algoritmo Support Vector Machine (SVM): avaliação da separação ótima de classes em imagens CCD-CBERS-2. In: Simpósio Brasileiro de Sensoriamento Remoto (SBSR), 14., 2009, Natal, PA. **Anais...**São José dos Campos: INPE, p. 2079-2086, 2009.

NETO, C. Di G. **Potencial de técnicas de mineração de dados para o mapeamento de áreas cafeeiras**. INPE, São José dos Campos, 2014.

PEREIRA, G. H. A.; CENTENO, J. A. S. Utilização de Support Vector Machine para classificação multiclases de imagens Landsat TM+. **Anais...** XVI Simpósio Brasileiro de Sensoriamento Remoto - SBSR, Foz do Iguaçu, 13 a 18 de abril de 2013, INPE, p.2322-2328.

PETROPOULOS, G. P.; ARVANITIS, K.; SIGRIMIS, N. Hyperion Hyperspectral Imagery Analysis Combined with Machine Learning Classifiers for Land Use/Cover Mapping. **Expert Systems with Applications**, 39, p. 3800-3809, 2012.

QUINLAN, J.R. **C4.5: programs for machine learning**. Sydney, Australia: Morgan Kaufmann Publishers, 1993. 302 p.

RUIZ, L. F. C.; ten CATEN, A.; DALMOLIN, R. S. D. Árvore de decisão e a densidade mínima de amostras no mapeamento da cobertura da terra. **Ciência Rural**, Santa Maria, v.44, n.6, p.1001-1007, 2014.

SAUSEN, T. M.; LACRUZ, M. S. P. **Sensoriamento Remoto para Desastres**. São Paulo: Oficina de Textos, 2015. 285 p.

TATSUMI, K.; YAMASHIKI, Y.; TORRES, M. A. C.; TAPE, C. L. R. Crop classification of upland fields using Random forest of time-series Landsat 7 ETM+ data. **Computers and Electronics in Agriculture**, v.115, p.171-179, 2015. Disponível em: < <http://dx.doi.org/10.1016/j.compag.2015.05.001>>. doi: 10.1016/j.compag.2015.05.001

WERE, K.; BUI, D. T.; DICK, O. B.; SINGH, B. R. A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afrotropical landscape. **Ecological Indicators**, v.52, p.394-403, 2015. Disponível em: < <http://dx.doi.org/10.1016/j.ecolind.2014.12.028>>. doi: 10.1016/j.ecolind.2014.12.028