METHOD FOR LANDSLIDES IDENTIFICATION AT THE SAO PAULO STATE COAST, BRAZIL

Luiz Augusto MANFRÉ[1], Eduardo Jun SHINOHARA[2], Janaina Bezerra SILVA[1], Raquel Nogueira Del Pintor SIQUEIRA[3] , Mariana Abrantes GIANNOTTI[1] and José Alberto QUINTANILHA[4]

(1) Escola Politécnica da Universidade de São Paulo – Departamento de Engenharia de Transportes – Laboratório de Geoprocessamento. Av. Prof. Almeida Prado, Travessa 2, n° 83 – Cidade Universitária - São Paulo – SP. CEP: 05508-900. Endereços eletrônicos: luizmanfre@gmail.com; janas@usp.br.
(2) Centro Universitário SENAC. Av. Engenheiro Eusébio Stevaux, 823 - Santo Amaro – São Paulo – SP. CEP 04696-000. Endereço eletrônico: edjun.shinohara@gmail.com.
(3) Programa de Pós-graduação em Ciências Ambientais. Instituto de Energia e Ambiente - IEE/USP. Prédio da Divisão de Ensino e Pesquisa, 2º Andar – Av. Prof. Luciano Gualberto, 1289 - Cidade Universitária – São Paulo – SP. CEP: 05508-900. Endereço eletrônico: raquelsiq@gmail.com.
(4) Escola Politécnica da Universidade de São Paulo – Departamento de Engenharia de Transportes. Av. Prof. Almeida Prado, Travessa 2, n° 83 – Cidade Universitária - São Paulo – SP. CEP: 05508-900. Endereços eletrônicos: jaquinta@usp.br

ABSTRACT – Satellite images are an important tool to map natural disaster, mainly debris flow. The Support Vector Machines (SVM) algorithm has been used to classify the natural disaster, obtaining good results, although some images present shadows and mists which difficult the classification. Some enhancements minimize those problems facilitating the classification process. This paper aims to present a method to classify debris flow areas near to an important road of the Sao Paulo State coast, Brazil, using LANDSAT images. Maximum Likelihood Classification (MLC) and SVM algorithms were applied. Due to the shadows the classification points huge debris flow areas. To neutralize the influence of shadows, Normalized Difference Vegetation Index (NDVI) was employed which turns easier to sample the training areas and perform the classification. MLC algorithm cannot be applied in case of a unique band, SVM can. So SVM is performed for the enhancement of classification and better results are observed with the combined methods SVM/NDVI. The overlay of this classification and Digital Terrain Model confirms the coincidence of debris flow event and classification. This method was very effective to the area now studied and may be useful to debris flow mapping.
Keywords: Debris flow, Remote Sensed Images, Numeric Difference Vegetation Index, Support Vector Machine.

RESUMO – Imagens de satélites são importantes ferramentas para mapear desastres naturais, principalmente deslizamento. O algoritmo Support Vector Machines (SVM) tem sido utilizado para classificar desastres naturais, apresentando resultados muito bons. Porém, algumas imagens apresentam sombras e brumas que dificultam o processo de classificação. Alguns realces minimizam esses problemas e facilitam a classificação. Esse artigo tem como objetivo apresentar um método de classificar um deslizamento próximo a uma importante rodovia do Estado de São Paulo, Brasil, utilizando imagens LANDSAT. Os algoritmos Máxima Verossimilhança (MLC) e SVM foram aplicados. Devido às sombras, a classificação inclui áreas erroneamente na de deslizamento. Para neutralizar a influência das sombras, o Índice de Diferença de Vegetação Normalizado (NDVI) foi aplicado, o que facilitou o treinamento e o processo de classificação. O algoritmo MLC não pode ser aplicado em apenas uma banda, o que é possível com o SVM. Assim, o SVM foi aplicado no realce NDVI, obtendo-se resultados muito superiores. A sobreposição da classificação com um Modelo Digital de Terreno permitiu a confirmação das áreas de deslizamento. O método apresentado foi muito eficiente para a área de estudo e pode ser replicado no mapeamento de deslizamentos.
Palavras-chave: Deslizamentos, Imagens de Sensoriamento Remoto, Índice de Diferença Numérica da Vegetação, Support Vector Machine.

INTRODUCTION

The support vector machine (SVM) is a binary classification technique that performs automatic split data in two classes. This algorithm uses only the support vectors or the most external samples of each class for classification (Vapnik, 1995). It describes an optimal separating hyperplane with the greater distances to the margin delineated by support vectors (Burges, 1998). However, the data is commonly distributed as non-linear way, restricting for SVM use. To solve this problem, Boser et al. (1992) propose the use of Kernel functions to project the data to a higher dimensional space and allow the description of the separating hyper plane.

The use of this algorithm in the land cover classification using orbital images has increased in the last decade. Many authors have used this technique to monitor and identify land use changes. It has shown good results with medium and high spatial resolution images. Using 15 meters resolution ASTER images, Szuster et al, (2011) compared the results of coastal area classifications with SVM, Maximum Likelihood Classification (MLC) and Artificial Neural Nets (ANN). SVM produced the better results, classifying with higher accuracy than the other two tools.

Zhao & Liu (2010) used Landsat and CBERS (China-Brazil Earth Research Satellite) images to monitor land use changes in Hanoi (Vietnam) using SVM. Similarly Lizarazo (2008) also employed SVM to classify the urban land use, showing significant results. According to Yao et al, 2008 SVM algorithm is the most accurate for the image classifications.

Foody & Mather (2006) and Pal & Mather (2006) highlighted that SVM classifier has the advantage of a lower sample effort comparing to other classifiers. The authors tested this assumption for an agricultural area at the United Kingdom. The results showed that if the support vectors of each class are well sampled, there is no need of a great amount of training areas.

Also, the application of this algorithm to identify natural disaster areas has been tested and confirmed it viability. Petropoulos et al. (2010) and Petropoulos et al. (2011) evaluated burnt areas in Greece, using ASTER 15 meters spatial resolution and Landsat 30 meters spatial resolution. The results showed a good accuracy for SVM method. Yilmaz (2010) compared SVM, ANN, logistic regression and conditional probability to identify landslides in Turkey. The better results were acquired for SVM and ANN methods, which, according to authors, are the most susceptible models to the pixels variability.

This paper aims to evaluate SVM algorithm and propose a method for its use in the identification of landslides near to an important road in the Sao Paulo State coast, Brazil, using Landsat images taken in 2000 since a great debris flow was observed in this year.

METHODOLOGY

Study area

In December of 1999, a severe event of debris flow at Pilões watershed region (Figure 1) was observed. The watershed is situated in Serra do Mar slope, an area that encompasses two cities: São Bernardo do Campo and São Vicente at São Paulo State, Brazil. The debris flow event was a consequence of four days of heavy rain (230 mm). This rain affected 700 meters of the Anchieta road around the kilometer 41 (Ogura, 2006). The area presents high precipitation values during all year (over 3000 mm/year) and great intensity of rains between November and March. In rainy months the precipitation average can reach 1000 mm monthly (Wolle, 1998).

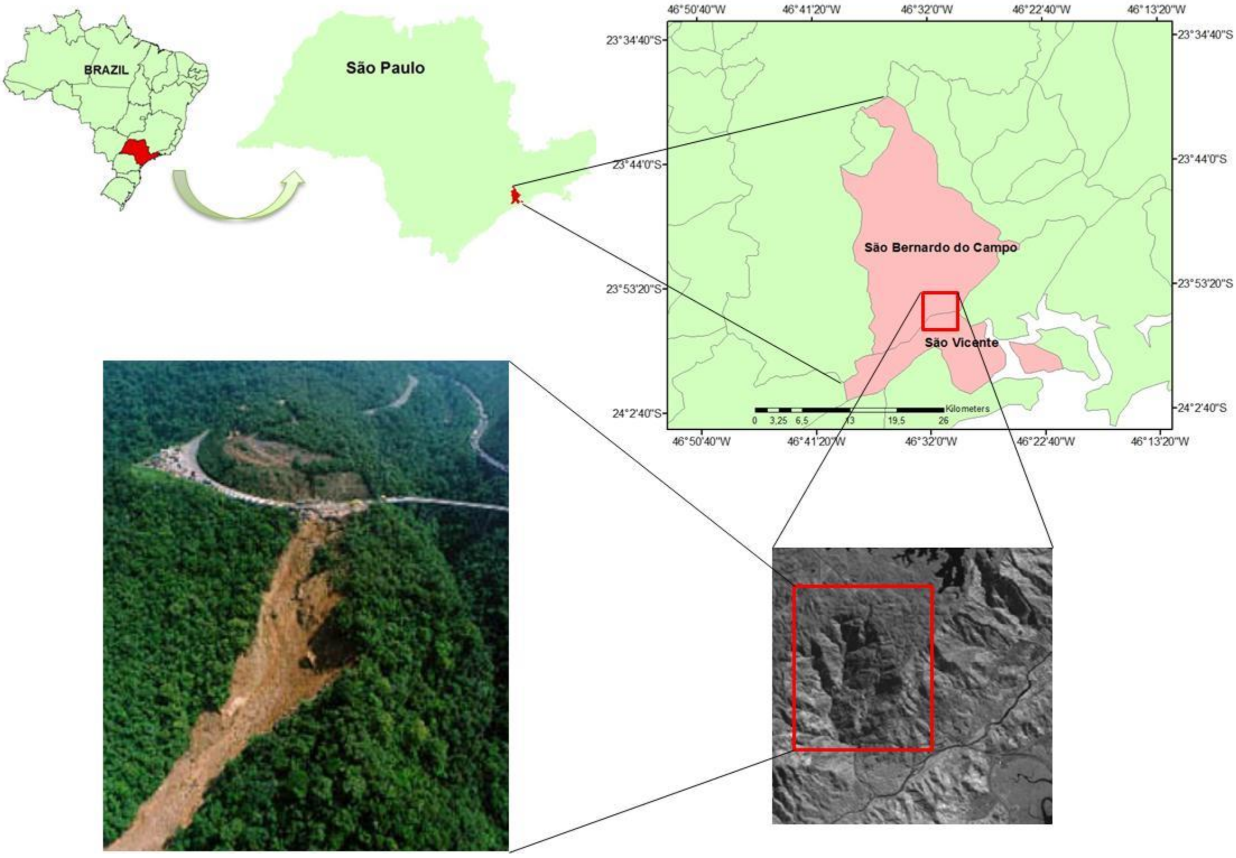

Figure 1. Study area location and debris flow picture. Adapted from Ogura (2009).

Procedures

Landsat images (30m) taken in June, 2000 are used in this study. The reason for that choice is because in this month this area presents lower precipitation values. For other periods, high precipitation levels are observed. In the most of rainy months the images present many clouds and shadows which disturb its image quality and classification. The year 2000 was selected due to the severe debris flow event occurred in December of 1999 at 41 Km of Anchieta highway.

SVM classification was performed in ENVI version 4.8 (ITT, 2009). According to Keuchel et al., (2003); LI and Liu, (2010), a success in SVM classification depends on the choice of the kernel function. Following the literature (Keuchel et al., 2003; Carrao et al., 2008; Knorn et al., 2009; Huang et al., 2008; Kuemmerle et al., 2009) and considering the theoretical assumptions, we chose the Radial Basis Function (RBF) land cover classification and change detection.

In this study, to detect debris scars normalized difference vegetation index (NDVI) enhancement is applied on Landsat images. It measures the surface reflectance and provides an estimate of the vegetation growth and biomass (Hall et al. 1995).Given as:

$$NDVI = \frac{NIR - RED}{NIR + RED}$$

Where,

NDVI is the Normalized difference vegetation index.

NIR is Infrared band.

RED is the band 3 of the TM instrument.

Samples of pure pixels of each class collected from NDVI enhancement are the training area. According to the literature (Foody & Mathur, 2006; Giles et al., 2006) this sampling procedure presents similar results from samples of boundaries class pixels. By the algorithm, it points the best support vectors for the class separation and sampling the boundaries would imply in less training areas. However, the debris areas at the image are very small and the sampling of the boundaries of this class is not allowed. Therefore, units of pixels of each class in the image are proportionally sampled.

SVM algorithm is developed to distinguish the classes: disaster event, water, vegetation, urban. Like Lizarazo (2008), which compared directly SVM classification and Segmentation-based classification (and provided similar results), we opted to not make the segmentation before the classification.

Also, for comparative purpose, the same training areas are classified by Maximum Likelihood Classification (MLC). This algorithm quantitatively evaluates both the variance and covariance of the category spectral response pattern (Lillesand & Kiefer, 2002) assuming the distribution of data points to be Gaussian (Bayarsaikhan et al., 2009) described by mean vector and covariance matrix.

As MLC cannot be applied in a case of a unique band and consequently NDVI cannot be calculated, the bands 3 and 4 of the TM instrument are selected and then SVM and MLC are performed. In both cases, the accuracy percentage of each class (confusion matrix), Kappa index are calculated.

The Kappa index is calculated using image interpretation to define the ground truth. We chose the pixels randomly. The overall accuracy and commission and omission errors were compared for each proposal methodology.

RESULTS AND DISCUSSIONS

The classification results for the SVM and MLC using bands 3 and 4 and SVM by NDVI enhancement are shown in the Figures 2, 3 and 4, respectively. Analyzing them we noticed that the SVM/NDVI classified a small amount of areas as debris flow, comparing to the other two methods. Considering the preliminary knowledge of the area, we overlay the classifications on a high space resolution satellite image, in order to evaluate the identification of the Pilões river disaster, near to the Anchieta Road. The overlay of the SVM/NDVI and the GeoEye Image (2000) is shown in Figure 5, and in perspective in Figure 6. A goodness of fit of the debris flow and the classification is observed. However, the other two methods (MLC and SVM using bands 3 and 4) did not fit as well, since they have classified correctly the debris flow areas but also classified wrongly other land cover as debris flow.

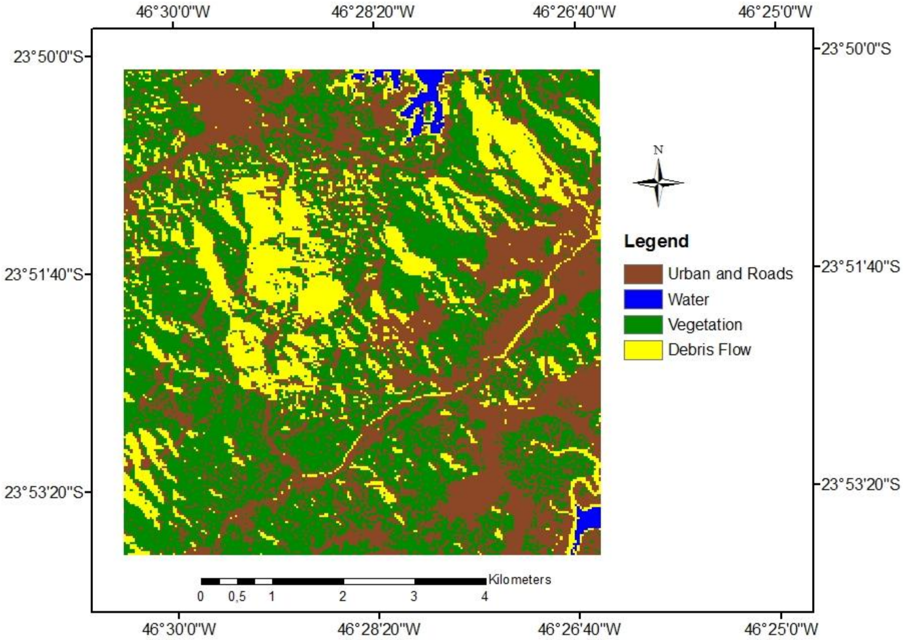

Figure 2. MLC based on 2000 Landsat image for the study area.

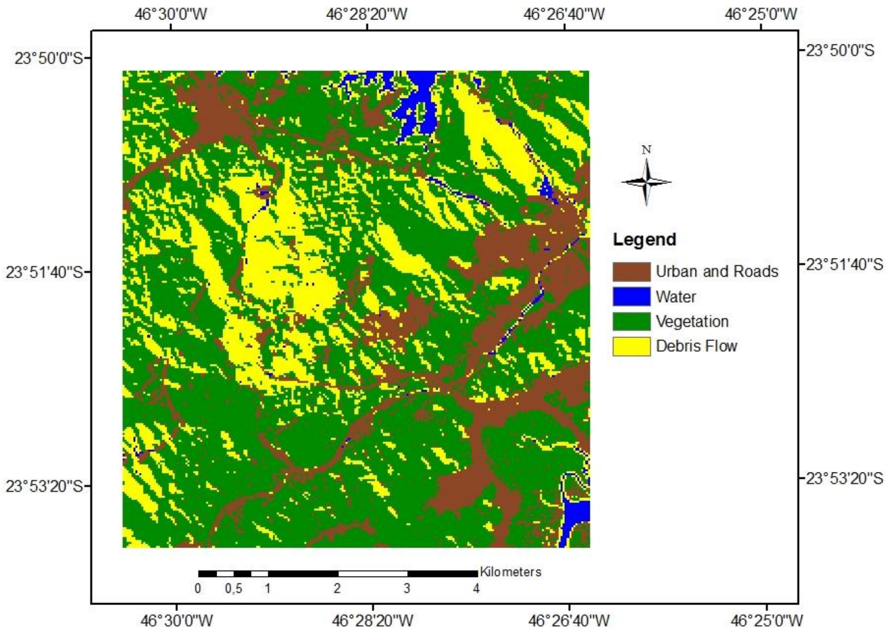

Figure 3. SVM Classification using bands 3 and 4 based on 2000 Landsat image for the study area.

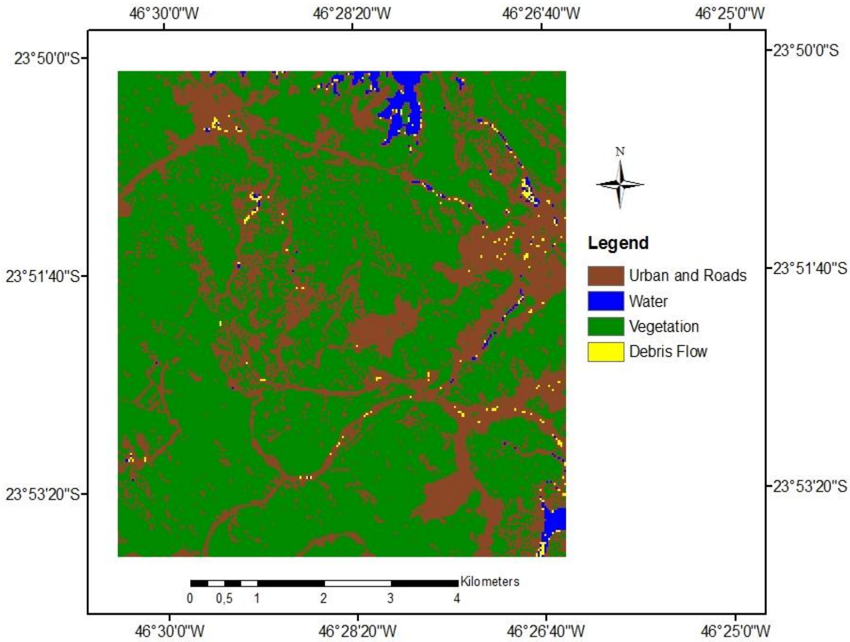

Figure 4. SVM Classification using NDVI enhancement based on 2000 Landsat image for the study area.

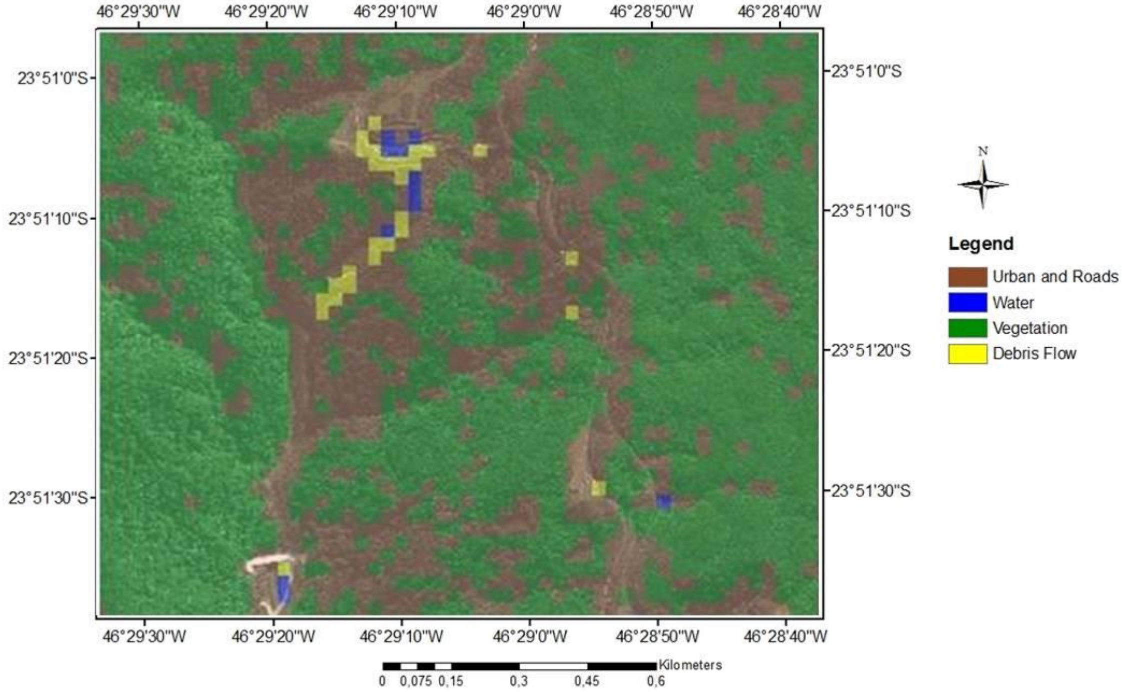

Figure 5. High resolution image and SVM/NDVI classification overlay.

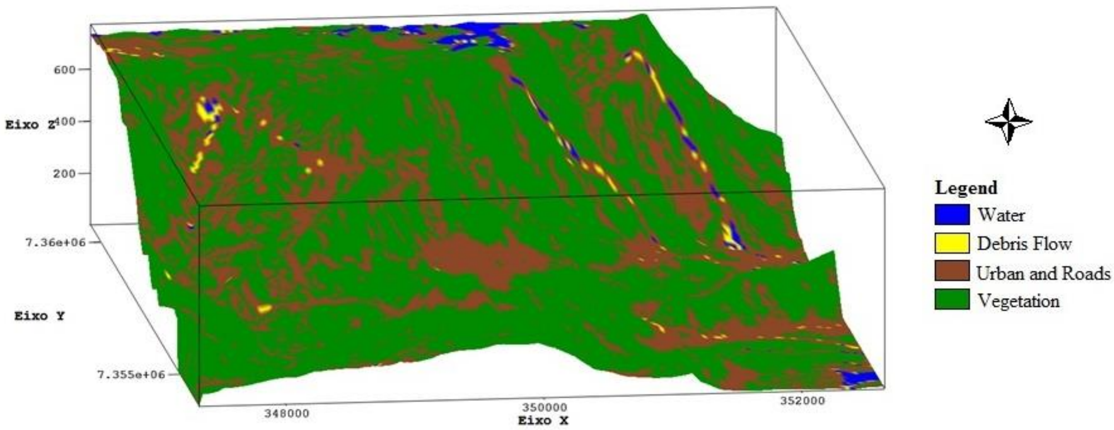

Figure 6. Debris flow perspective vision for SVM/NDVI classification and the digital terrain model overlay.

The performance of the methods of SVM using bands 3 and 4, SVM by the NDVI enhancement and MLC using bands 3 and 4 were analyzed by the coefficient of Overall Accuracy, the confusion matrix and the Kappa coefficient. The results are summarized in Table 1 which indicates a great advantage for the SVM/NDVI classification method.

Table 1. Kappa index and overall accuracy for each method.

Method	Overall accuracy	Kappa coefficient
MLC(band 3, 4)	84.25	0.78
SVM(band 3, 4)	82.88	0.76
SVM (NDVI)	92.47	0.90

Using the same bands MLC algorithm provides better results when compared to the SVM. However, the use of the NDVI enhancement jointly with SVM presents better results than the two alternatives. This is a consequence of the enhancement, which neutralize the effects of the shadows and mists and allows the split of the classes. Therefore, the SVM algorithm is more effective to highlight the debris flow, since it allows classification with a unique band.

The confusion matrices are shown in Tables 2, 4 and 6. Comparing the three tables very similar results for the debris flow class are observed. It might be as consequence of the sampling of the truth points. Similar results are also seen at the commission and omission errors, shown in Tables 3, 5 and 7.

Table 2. MLC confusion matrix at the year 2000.

	Ground Truth (Percent)				
Class	**Urban and Roads**	**Water**	**Vegetation**	**Debris flow**	**TOTAL**
Unclassified	0,00%	0,00%	0,00%	0,00%	0,00%
Urban and Roads	**92,50%**	9,30%	16,33%	7,14%	34,25%
Water	0,00%	**81,40%**	0,00%	14,29%	25,34%
Vegetation	7,50%	0,00%	**81,63%**	0,00%	29,45%
Debris flow	0,00%	9,30%	2,04%	**78,57%**	10,96%
TOTAL	100,00%	100,00%	100,00%	100,00%	100,00%

Table 3. Commission and omission errors for the MLC.

	%		Pixels number	
Class	**Commission**	**Omission**	**Commission**	**Omission**
Urban	26,00%	7,50%	13/50	3/40
Water	5,41%	18,60%	2/37	8/43
vegetation	6,98%	18,37%	3/43	9/49
debris flow	31,25%	21,43%	5/16	3/14

Table 4. SVM (bands 3 and 4) confusion matrix .

	Ground Truth (Percent)				
Class	**Urban and Roads**	**Water**	**Vegetation**	**Debris flow**	**TOTAL**
Unclassified	0,00%	0,00%	0,00%	0,00%	0,00%
Urban and Roads	**70,00%**	2,33%	10,20%	0,00%	23,29%
Water	7,50%	**93,02%**	0,00%	21,43%	31,51%
Vegetation	7,50%	0,00%	**85,71%**	0,00%	30,82%
Debris flow	15,00%	4,65%	4,08%	**78,57%**	14,38%
TOTAL	100,00%	100,00%	100,00%	100,00%	100,00%

Table 5. Commission and omission errors for the SVM (bands 3 and 4).

	%		Pixels number	
Class	**Commission**	**Omission**	**Comission**	**Omission**
Urban	17,65%	30,00%	6/34	12/40
Water	13,04%	6,98%	6/46	3/43
vegetation	6,67%	14,29%	3/45	7/49
debris flow	47,62%	21,43%	10/21	3/14

Table 6. SVM (NDVI) confusion matrix.

	Ground Truth (Percent)				
Class	**Urban and Roads**	**Water**	**Vegetation**	**Debris flow**	**TOTAL**
Unclassified	0,00%	0,00%	0,00%	0,00%	0,00%
Urban and Roads	**95,00%**	4,65%	2,04%	29,63%	28,08%
Water	0,00%	**88,37%**	0,00%	21,43%	28,08%
Vegetation	0,00%	0,00%	**97,96%**	0,00%	32,88%
Debris flow	5,00%	6,98%	0,00%	**78,57%**	10,96%
TOTAL	100,00%	100,00%	100,00%	100,00%	100,00%

Table 7. Commission and omission errors for the SVM (NDVI).

	%		Pixels number	
Class	**Commission**	**Omission**	**Comission**	**Omission**
urban	7,31%	5,00%	3/41	2/40
water	7,34%	11,63%	3/41	5/43
vegetation	0,00%	2,04%	0/48	1/49
debris flow	31,25%	21,43%	5/16	3/14

Despite of the similarity of the evaluation results, only the SVM/NDVI methodology could distinguish the debris flow accident at km 42 of Anchieta road (Figures 5). This difference in the classification might be a consequence of the NDVI enhancement combined with the SVM classification. Considering that the study area has severe slopes and high humidity percentages, the classification algorithm must be used on bands or on an enhancement that minimizes the effects of the mists and shadows.

Besides, the combination of the SVM/NDVI classification with the digital terrain model allows the analysis of the areas that are most susceptible to debris flow accidents. Figure 6 shows a perspective for the overlay of the classification image and the digital terrain model. Note that the majority of pixels classified as debris flow occurs in high slope areas. Those pixels must be used to guide the mapping of debris flow occurrences allowing faster mitigation actions.

Therefore, NDVI enhancement is a good option. However, the MLC algorithm does not runs over a single band, needing at least two bands for its execution. This is the main advantage of the SVM algorithm, as it runs over a unique band allowing a classification based over and enhancement that highlights the class of interest. Moreover, SVM algorithm searches for the edges of each class, using a Kernel function, is a good algorithm to split classes which have similar configurations, meanwhile the MLC looks for the similarity of the class members.

In this sense, regarding the application of classification in order to map vulnerable areas, it is indicated, based on the results, that SVM/NDVI method should be applied coupled with hypsometric data in order to better identify areas that could be susceptible for future debris flows.

CONCLUSIONS

Comparing the two algorithms, one concludes SVM detects the class changes, creating homogenous polygons, while the MLC identifies the similar pixels creating a “salt and pepper” classification, confusing the target classes.

Nevertheless the SVM algorithm was not able to identify the debris flow areas due to shadows and mists over the study area. In order to minimize those effects SVM based on the NDVI enhancement is employed. The last method can identify and split the classes, showing the debris flow areas with higher accuracy. So it may be used with slope and land cover maps to subside the debris flow hazard areas mapping.

ACKNOWLEDGEMENTS

We thank to EPUSP/LGP for the infrastructure, to IPT for the support and theoretical material about the Anchieta Highway disaster provided. Also, we thank to FAPESP and CNPq for the scholarships of two of the authors.

REFERENCES

1. BAVARSAIKHAN, U.; BOLDGIV, B.; KIM, K.R.; PARK, K.A.; LEE, D. Change detection and classification of land cover at Hustai National Park in Mongolia. **International Journal of Applied Earth Observation and Geoinformation**, v. 11, pp. 273-280, 2009.
2. BOSER, B.; GUYON, I.; VAPNIK, V. A training algorithm for optimal margin classifiers. In: Annual Workshop on Computational Learning Theory, 5, 1992, New York. **Proceedings…** New York: ACM Press, 1992.
3. BURGES, C. J. C. A tutorial on support vector machines for pattern recognition. **Data Mining and Knowledge Discovery**, v. 3, n. 2, 1998.
4. FOODY, G.M. & MATHUR, A. The use of small sets of containing mixed pixels for accurate hard image classification: Training on mixed spectral responses for classification by SVM. **Remote Sensing of Environment**, v. 103, p. 179-189, 2006.
5. KEUCHEL, J.; NAUMANN, S.; HEILER, M.; SIEGMUND, A. Automatic land cover analysis for Tenerife by supervised classification using remotely sensed data. **Remote Sensing of the Environment**, v. 86, p. 53–541, 2003.
6. LI, D.C.; LIU, C.W. A class possibility based kernel to increaseclassification accuracy for small data sets using support vector machines, **Expert Syst. Appl.**, v. 37, p. 3104–3110, 2010.
7. LILLESAND, T.M. & KIEFER, R.W. **Remote Sensing and Image Interpretation**, 4. Ed., New Jersey: John Willey and Sons, 2002.
8. LIZARAZO, I. SVM-based segmentation and classification of remotely sensed data. **International Journal of Remote Sensing**, v. 29, n. 24, p. 7277-7283, 2008.
9. OGURA, A.T. **Relação e Previsibilidade de Eventos Meteorológicos Extremos Deflagadores de Acidentes de Movimentos de Massa na Serra do Mar**, IPT: São Paulo, 2006.
10. OGURA, A.T. **Desastres Naturais no Brasil: vulnerabilidades Sociais e Econômicas e Adaptação às Mudanças Climáticas**, IPT: São Paulo, 2009.
11. PAL, M. & MATHER, P.M. Some issues in the classification of DAIS hyperspectral data. **International Journal of Remote Sensing,** v. 27, p. 2895–2916, 2006.
12. PETROPOULOS, G.P.; KNORR, W.; SCHOLZE, M.; BOSCHETTI, L.; KARANTOUNIAS, G. Combining ASTER multispectral imagery analysis and support vector machines for rapid and cost-effective post-fire assessment: a case study from the Greek wildland fires of 2007. **Natural Hazards and Earth System Sciences**, v. 10, p. 305-317, 2010.
13. PETROPOULOS, G.P.; KONTOES, C.; KERAMITSOGLOU, I. Burnt area delineation from a uni-lateral perspective based on Landsat TM imagery classification using Support Vector Machine. **International Journal of Applied Earth Observation and Geoinformation**, v. 13, p. 70-80, 2011.
14. RAMACHANDRA, V.; DIWAKAR, P.G.; JOSHI, N.V. Spatio-temporal landscape modelling for natural hazard vulnerability analysis in select watersheds of Central Western Ghats, **ISTC/BES/TVR/205 (2007-2010)**, p. 192 – 202, 2010.
15. SOUSA, B.F.S.; TEIXEIRA, A.S.; FERREIRA, F.A.T.; ANDRADE, E.M.; BRAGA, A.P.S. Avaliação de classificadores baseados em aprendizado de máquina para a classificação do uso e cobertura da terra no bioma Caatinga. RBC. **Revista Brasileira de Cartografia (Online)**, v. 62, p. 385-399, 2010.
16. SZUSTER, B.W.; CHEN, Q.; BORGER, M. A comparison of classification techniques to support land cover and land use analysis in tropical coastal zones. **Applied Geography**, v. 31, p. 525-532, 2011.
17. VAPNIK, V. **The nature of statistical learning theory.** New York: Springer-Verlag, 1995.
18. WANA, S.; LEI, T.C. A knowledge-based decision support system to analyze the debris-flow problems. at Chen-Yu-Lan River, **Elsevier**, v. 22, p. 580–588, 2009.
19. WOLLE, C.M. **Análise dos escorregamentos translacionais numa região da Serra do Mar no contexto de uma classificação de mecanismos de instabilização de encostas**. São Paulo, 1988. Tese (Doutorado em Engenharia) – Escola Politécnica da USP.
20. YAO, X.; THAM, L.G.; DAI, F.C. Landslide susceptibility mapping based on support vector machine: a case study on natural slopes of Hong Kong, China. **Geomorphology**, v. 101, p. 572–582, 2008.
21. YILMAZ, I. Comparison of landslide susceptibility mapping methodologies for Koyulhisar, Turkey: conditional probability, logistic regression, artificial neural networks, and support vector machine. **Environmental Earth Science**, v. 61, p. 821-836, 2010.
22. ZHAO, J. & LIU, C. Monitoring dynamic change of land cover based on SVM and Satellie Images in Hanoi, Vietnam. **Wuham University Journal of Science**, v. 15, n. 4, p. 355-362, 2010.

Manuscrito recebido em: 12 de abril de 2012
Revisado e Aceito em: 04 de abril de 2014