

COMPARAÇÃO DOS CLASSIFICADORES *MAXIMUM LIKELIHOOD CLASSIFICATION* E *SUPPORT VECTOR MACHINE* PARA O MAPEAMENTO DE FRAGMENTOS FLORESTAIS NA MATA ATLÂNTICA.

Marcos Roberto Martinezes¹, Rúbia Gomes Morato², Fernando Shinji Kawakubo³, Ricardo Vicente Ferreira⁴, Marcel Fantin⁵, Rogério Hartung Toppa⁶

¹ Universidade Federal de São Carlos, UFSCAR, Rod. João Leme dos Santos, km 110 - SP-26 Bairro do Itinga, Sorocaba CEP:18052-780, mmartines@ufscar.br, ² Universidade de São Paulo, USP, Avenida Prof. Lineu Prestes, 338, Cidade Universitária, São Paulo CEP: 05508-000, rubiagm@usp.br, ³ Universidade de São Paulo, USP, Avenida Prof. Lineu Prestes, 338. Cidade Universitária, São Paulo CEP: 05508-000, fsk@usp.br, ⁴ Universidade Federal do Triângulo Mineiro, UFTM, Av. Getúlio Guaritá, 159, CEP: 38025-440, Uberaba-MG, ricardo.ferreira@uftm.edu.br, ⁵ Universidade de São Paulo, USP, Av. Trab. São Carlsense, 400, Centro, São Carlos - SP, 13566-590 mfantin@sc.usp, ⁶ Universidade Federal de São Carlos, UFSCAR, Rod. João Leme dos Santos, km 110 -SP-26 Bairro do Itinga – Sorocaba CEP 18052-780, toppa@ufscar.br

RESUMO

O objetivo desta pesquisa foi comparar a acurácia dos classificadores *Maximum Likelihood Classification* e *Support Vector Machine* para o suporte ao mapeamento de fragmentos de Mata Atlântica. Para essa aplicação foi utilizada como área de estudo o município de Salto de Pirapora (SP). Foram consideradas duas classes de análise: fragmentos de floresta nativa, denominados como classe natural, e áreas com *Eucalyptus* spp. ou *Pinus* spp., denominadas como classe exótica. As amostras de treinamento dos classificadores foram vetorizadas com o uso de imagens Sentinel 2. Para avaliar a acurácia dos classificadores, foram obtidos 80 pontos de controle com o apoio do Google Earth, e foi elaborada uma matriz de confusão e aplicado o índice Kappa. Os dois classificadores analisados possibilitaram a diferenciação das classes temáticas propostas, porém o classificador Support Vector Machine demonstrou ter maior acurácia tanto para a identificação dos padrões de Classe Natural quanto para a Classe Exótica.

Palavras-chave — sensoriamento remoto, reconhecimento de padrões, classificação supervisionada, Sentinel 2, Mata Atlântica

ABSTRACT

The objective of this research was to compare the accuracy of the Maximum Likelihood Classification and Support Vector Machine classifiers to support the mapping of Atlantic Forest fragments. For this application, the municipality of Salto de Pirapora (SP) was used as the study area. Two classes of analysis were considered: native forest fragments, denominated as natural class, and areas with Eucalyptus spp. or Pinus spp., denominated as exotic class. The training samples of the classifiers were vectored using Sentinel 2 images. To evaluate the accuracy of the

classifiers, 80 control points were obtained with the support of Google Earth, and a confusion matrix was elaborated and the Kappa index was applied. The two classifiers analyzed allowed the differentiation of the proposed thematic classes, but the Support Vector Machine classifier proved to be more accurate both for the identification of Natural Class standards and for the Exotic Class.

Key words — remote sensing, pattern recognition, supervised classification, Sentinel 2, Mata Atlântica

1. INTRODUÇÃO

O mapeamento dos remanescentes florestais da Mata Atlântica e de seus estágios de sucessão consiste em uma etapa fundamental para implantação de diversos estudos, ações de fiscalização e manejo ambiental [1]. Portanto, a identificação dos fragmentos florestais naturais do bioma da Mata Atlântica é um componente básico para uma gama ampla de estudos. Na Mata Atlântica, a maior parte dos remanescentes florestais encontra-se na forma de pequenos fragmentos, altamente perturbados, isolados, pouco conhecidos e pouco protegidos [2]. Esses fragmentos florestais naturais são pressionados por uma dinâmica de cobertura da terra em função da expansão e retração de alguns usos antrópicos. Isso modifica a paisagem formando um mosaico constituído por fragmentos de diversos tamanhos e em diversos estágios de sucessão, isolados por uma matriz antropizada [3].

A proposição de metodologias que contemplam essa temática permite avaliar quantitativa e qualitativamente os fragmentos remanescentes, bem como, a sua distribuição espacial [4], sendo os métodos voltados a identificar especialmente esses fragmentos naturais, compostos de diferentes indivíduos arbóreos formando um complexo dossel de vegetação. Dentre a composição de mosaicos de diferentes usos que circundam esses fragmentos, destacam-se de áreas de plantio com *Eucalyptus* spp. (Myrtaceae) ou *Pinus* spp. (Pinaceae), espécies arbóreas exóticas com

finalidade comercial. É fundamental realizar a distinção entre essas diferentes coberturas arbóreas a fim de estabelecer uma exata mensuração da ocorrência espacial dos fragmentos florestais naturais da Mata Atlântica.

Nesse sentido, esta pesquisa teve como objetivo aplicar dois métodos de classificação de imagem, o *Maximum Likelihood Classification* (MLC) e *Support Vector Machine* (SVM), para diferenciar os fragmentos com vegetação nativa dos fragmentos com vegetação exótica por meio da diferenciação de seus dosséis. A aplicação de classificação de imagens por sensoriamento remoto envolve um processo complexo que engloba muitos fatores, incluindo, a própria escolha do método de classificação, que requer a checagem da acurácia dos resultados. [5]

2. MATERIAIS E MÉTODOS

As seguintes etapas foram adotadas: (1) aquisição das imagens do Sentinel 2 da área de estudo e seleção das bandas espectrais; (2) reamostragem das imagens para a mesma resolução espacial; (3) seleção de amostras para treinamento e aplicação de dois classificadores para identificação das classes; (4) levantamento de amostras para avaliação da acurácia dos classificadores, *field check*.

Como área de estudo, foi selecionado o município de Salto de Pirapora (SP), localizado na região sudeste do Estado de São Paulo (Figura 1). Essa seleção se deu em função do município apresentar remanescentes de florestas pertencem ao Bioma Mata Atlântica [6], um hotspot de biodiversidade [7] e contar com uma base de dados georreferenciados elaborada pelo Núcleo de Estudos em Ecologia da Paisagem e Conservação - NEEPC.

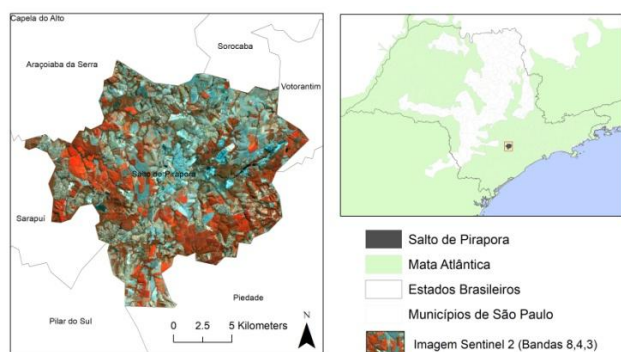


Figura 1. Localização da área de estudo, Salto de Pirapora (SP)

Foram consideradas duas classes alvos de fragmentos de floresta, Classe Natural (CN) e Classe Exótica (CE), selecionadas por meio dos pressupostos teóricos da fotointerpretação [8] delimitados através da vetorização de polígonos de treinamento sobre imagem do Sentinel 2 (Figura 2A). As classes CN e CE contemplaram estruturas florestais em diversos estágios sucessionais de vegetação nativa, *Eucalyptus* spp. e *Pinus* spp. Para o conjunto *field check* (validação), foram obtidas 40 amostras para cada

classe com auxílio das imagens de alta resolução do Google Earth (Figura 2B).

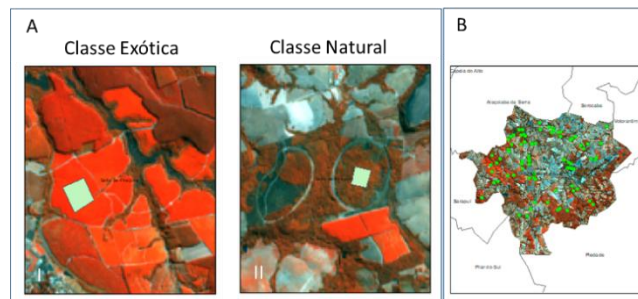


Figura 2 AI II – Exemplos de amostras para treinamento (polígono verde); 2B – Conjunto amostral para o *field check*

Para a aplicação dos dois métodos de classificação de imagem multiespectrais, foram selecionadas 10 bandas do Sentinel 2 (29/07/2018), sendo quatro bandas no visível e infravermelho (resolução espacial de 10 m) e seis bandas no “red edge” e no infravermelho de ondas curtas (resolução de 20 m) que foram reamostradas para 10 m. O primeiro classificador utilizado foi o MLC. Esse é um dos classificadores mais tradicionais descritos na literatura [9]. Esse classificador funciona calculando o vetor médio, variância e a correlação para cada classe nos dados de treinamento [10]. Um conjunto de funções de probabilidade para todas as classes é estimado com base nos dados de treinamento para descrever a probabilidade relativa de que um determinado valor de pixel pertence a cada classe [11]. Essas funções são usadas como regra de decisão para classificação de forma a atribuir um rótulo de pixel como a classe com a maior probabilidade (ou máxima verossimilhança) de associação. Esse método paramétrico pressupõe que os dados de entrada são multivariados, normalmente distribuídos, independentes e têm variâncias aproximadamente iguais [12]. Quando esses critérios são atendidos, a aplicação da técnica de MLC tem-se mostrado adequada para classificação, utilizando-se um número limitado de variáveis. [13]

O segundo classificador utilizado foi o SVM. Esse método é uma arquitetura de classificação de imagem inspirada na teoria de aprendizagem estatística e baseada no conceito de Minimização do Risco Estrutural, sendo que este método calcula o erro com relação ao conjunto de treinamento e realiza a diminuição do erro global. O objetivo de SVM consiste em obter um equilíbrio entre esses erros, diminuindo o excesso de ajustes com respeito às amostras de treinamento (*overfitting*) [14]. O problema denominado de *overfitting* consiste em o classificador memorizar os padrões de treinamento, gravando suas peculiaridades e ruídos, ao invés de extrair as características gerais que permitirão a generalização ou reconhecimento de padrões não utilizados no treinamento do classificador [15]. A partir das imagens classificadas, realizou-se o cálculo da

estimativa de área (em ha) para cada uma das classes mapeadas.

A acurácia do desempenho dos classificadores foi medida pela matriz de confusão, que é um método padrão para avaliação da classificação. A matriz de confusão fornece a base para descrever a exatidão da classificação e caracterizar tanto os acertos quanto os erros presentes em cada classe [16]. A partir dessa matriz também foi calculado o índice Kappa que consiste em uma medida que descreve o quanto a classificação está de acordo com os dados de referência [17]. A vantagem da utilização do índice Kappa é que ele utiliza todos os elementos da matriz no seu cálculo.

2. RESULTADOS E DISCUSSÃO

Os resultados das classificações utilizando cada um dos métodos estão apresentados na Figura 3.

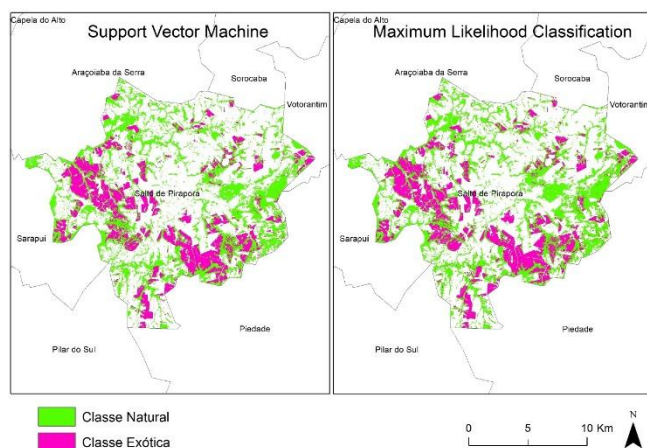


Figura 3 – Classificação temática das classes de fragmentos florestais com vegetação nativa e exótica

O método SVM para CN apresentou 6.814 ha, enquanto o MLC 4.591 ha de cobertura da vegetação (Tabela 1). Estudos realizados na área de estudo [18] sobre uso e ocupação, na escala de 1/5.000, com o uso de ortofotos com 1 metro de resolução (anos 2010-11) obteve uma cobertura de 5.996 ha para fragmentos de floresta nativa. Em comparação com a classificação realizada neste trabalho, observou-se uma diferença de área negativa para o classificador SVM (818 ha) e positiva para MLC (1.817 ha). Nessa estimativa, o classificador SVM foi o que apresentou resultados mais próximos da interpretação em escala de detalhe.

Tabela 1 – Número de pixels (N) e área em hectares (ha) para as classes temáticas por classificador (SVM - Support Vector Machine; MLC - Maximum Likelihood)

	SVM		MLC	
	N	ha	N	ha
CE	428.267	4.283	768.636	7.686
CN	681.423	6.814	459.109	4.591

A classe que apresentou a melhor exatidão individual foi a CN gerada do classificador SVM, enquanto a que apresentou a menor exatidão foi a CN utilizando o classificador MLC (Tabela 2). Tal resultado pode ser atribuído à complexidade da estrutura do dossel da vegetação nativa, com variações complexas de sombras em função da irregularidade das copas. Essa diferença de desempenho também é expressa pelo Índice Kappa (Tabela 2), que para o classificador SVM aparece com 0,93, que é considerando um desempenho excelente ($0,8 < k \leq 1,0$ Excelente), já o desempenho do MLC com 0,76 é considerado muito bom ($0,6 < k \leq 0,8$).

Tabela 2 – Matriz de Confusão e Índice Kappa Support Vector Machine

	CE	CN	Total	PC	K
CE	39	2	41	95,12	0,93
CN	1	38	39	97,44	
Total	40	40	80		

Maximum Likelihood Classification

	CE	CN	Total	PC	K
CE	37	2	39	94,87	0,76
CN	8	33	41	80,49	
Total	45	35	80		

PC: Precisão de classe individual; K: Índice Kappa; CE: Classe Exótica; CN: Classe Natural

5. CONCLUSÕES

O trabalho apresentou a diferenciação de classes temáticas de floresta, por meio de dois classificadores, para a identificação de fragmentos de Mata Atlântica em uma escala média de análise.

O desempenho dos dois classificadores analisados possibilitou a diferenciação das classes temáticas propostas, porém, o classificador SVM mostrou-se mais eficiente tanto para a identificação dos padrões de Classe Natural quanto para a Classe Exótica. Cabe destacar que em qualquer classificação supervisionada, é fundamental fornecer amostras representativas das classes mapeadas que possibilitem a diferenciação estatística entre os alvos por meio dos classificadores.

Nessa temática, é fundamental que futuros estudos contribuam para a exploração de novas possibilidades de classificação da vegetação, tanto com outros algoritmos quanto com outras imagens, assim aumentando as possibilidades de monitoramento e análise dos diferentes padrões de vegetação da Mata Atlântica.

6. REFERÊNCIAS

- [1] Ribeiro, M. C. et al. The Brazilian Atlantic forest: how much is left, and how is the remaining forest distributed? **Biological Conservation**, Liverpool, v. 142, n. 6, p. 1141- 1153, 2009.
- [2] Viana, V. M Conservação da biodiversidade de fragmentos de florestas tropicais em paisagens intensivamente cultivadas. In: **Abordagens interdisciplinares para a conservação da biodiversidade e dinâmica do uso da terra no novo mundo**.

Gainesville, Conservation International do Brasil/UFMG/ University of Florida., 1995. p. 135-154.

[3] Pires, V. R. O. et al Landscape structure analysis for the establishment of conservation strategies in Atlantic Forest patches. **Ambiência**, v. 12, p. 765, 2016.

[4] Amaral, M. V. F et al Avaliação e comparação de métodos de classificação de imagens de satélites para o mapeamento de estádios de sucessão florestal. **Revista Árvore**, 2009 33, pp. 575-582

[5] Lu, D.; Weng, Q. A survey of image classification methods and techniques for improving classification performance. **International Journal of Remote Sensing**, v. 28, n. 5, p. 823–870, 2007.

[6] IBGE. "Mapa de biomas do Brasil." 2004

[7] Myers, N. et al. Biodiversity hotspots for conservation priorities. **Nature**. v. 403. p. 853-858. 2000

[8] Marchetti, D.A. Garcia, G. J. **Princípios de fotogrametria e fotointerpretação**. Editora Nobel, 1988.

[9] Benedicsson, J. A. et al. Neural network approaches versus statistical methods in: classification of multisource remote sensing data. **IEEE**, 1990, 28, 540–551.

[10] Curran, P. Principles of Remote Sensing, London: Longman, 1985, 282 pp.

[11] Richards, A. Remote Sensing Digital Image Analysis, Berlin: Springer-Verlag, 1986, 281 pp.

[12] Tom, C. H., Miller L. D. An Automated Land-use Mapping Comparison of the Bayesian Maximum Likelihood and Linear Discriminant Analysis Algorithms, **Photogrammetric Engineering and Remote Sensing**, 1984, Vol. 50, No.2, pp. 193-207.

[13] Derek R. P. An Empirical Comparison of Evidential Reasoning, Linear Discriminant Analysis, and Maximum Likelihood Algorithms for Alpine Land Cover Classification, **Canadian Journal of Remote Sensing**. 1993

[14] Vapnik, V.N. **The Nature of Statistical Learning Theory**. USA: Springer, 2nd ed, 1998.

[15] Smola, A.J et al **Advances in large margin classifiers**. Massachusetts Institute of Technology. London, England: Ed. MIT Press, 2000.

[16] Foody, G. M. Status of land cover classification accuracy assessment. **Remote Sensing of Environment**, v. 80, p. 185– 201, 2002.

[17] Figueiredo G.C; Vieira C.A.O Estudo do comportamento dos índices de Exatidão Global, Kappa e Tau, comumente usados para avaliar a classificação de imagens do sensoriamento remoto In: Anais XIII Simpósio Brasileiro de Sensoriamento Remoto, Florianópolis, Brasil, 21-26 abril 2007, INPE, p. 5755-5762

[18] Pires, V. R. et al. Land use and occupation mapping as support to environmental planning. **Ambiência**, v. 12, p. 889, 2016.