



# Integração entre Estratégias Multiclases e diferentes Funções Kernel em Máquinas de Vetores Suporte para Classificação de Imagens de Sensoriamento Remoto

Integration of Multiclass Strategies and Different Kernel Functions into Support Vector Machines for Remote Sensing Image Classification

*Luccas Zambon Maselli*<sup>1</sup>  
*Rogério Galante Negri*<sup>1</sup>

Recebido em agosto de 2018.  
Aprovado em dezembro de 2018.

## RESUMO

Dentre diferentes métodos de classificação de imagens, Máquina de Vetores Suporte (Support Vector Machine – SVM) tem sido amplamente utilizado em diferentes aplicações em Sensoriamento Remoto. Além de sua excelente formulação matemática, a possibilidade de emprego de diferentes funções kernel e estratégias multiclases tornam o método SVM ainda mais atrativo. Enquanto as funções kernel possibilitam aumentar a capacidade de distinção entre dados não linearmente separáveis, as estratégias multiclases estendem a formulação original do método SVM a fim de lidar com problemas de classificação envolvendo além de duas classes. A escolha envolvendo uma função kernel e uma estratégia multiclases em particular implica diretamente sobre a acurácia da classificação. Este trabalho propõe duas arquiteturas para treinamento do método SVM com finalidade de diminuir o grau de liberdade que surge diante das diferentes combinações possíveis entre função kernel e estratégia multiclases. Três estudos de caso, envolvendo classificação de uso e cobertura do solo a partir de imagens adquiridas por diferentes sensores, são realizados a fim de verificar o potencial das arquiteturas formalizadas em comparação as abordagens usuais.

**PALAVRAS-CHAVE:** Estratégias multiclasse. Funções Kernel. Máquina de Vetores Suporte. Classificação de Imagens.

## ABSTRACT

Although several image classification methods have been proposed in literature, Support Vector Machine (SVM) is widely used in Remote Sensing applications. In addition to its robust mathematical formulation, the

---

<sup>1</sup> Universidade Estadual “Júlio de Mesquita Filho” - UNESP. Instituto de Ciência e Tecnologia - ICT. Departamento de Engenharia Ambiental, Brasil. E-mail: [maselliluccas@gmail.com](mailto:maselliluccas@gmail.com); [rogerio.negri@unesp.br](mailto:rogerio.negri@unesp.br).

possibility of using different kernel functions and multiclass strategies highlights the attractiveness of this method. While kernel functions make possible to enhance the classification performance face to non-linearly separable data, multiclass strategies extend the original formulation of SVM in order to cope with problems involving more than two classes. However, it worth mention that particular choice involving a kernel function and a multiclass strategy implies directly on the classification performance. Furthermore, the best choice may be not a simple task. In order to reduce the freedom degree that arises from different possible combinations between kernel function and multiclass strategy, two architectures to training SVM are proposed. Three case studies involving land use and land cover classification with images acquired by different sensors are carried in order to verify the potential of presented architectures in comparison to usual approaches.

**KEYWORDS:** Multiclass strategy. Kernel Functions. Support Vector Machines. Image Classification.

\* \* \*

## Introdução

Classificação de Imagem compreende um extenso conjunto de técnicas cuja finalidade consiste em identificar padrões/objetos homogêneos em imagens digitais. Tais técnicas recebem grande atenção diante as mais diversas aplicações em Sensoriamento Remoto, por exemplo, na investigação da distribuição espacial de determinada classe de cobertura do solo (LU, 2005), na detecção de alterações temporais (DEVRIES et al., 2015) ou mesmo em estudos climatológicos (GÓMEZ-ZOTANO et al., 2016).

Dentre diferentes métodos de classificação de imagens, destaca-se a Máquina de Vetores Suporte (Support Vector Machine – SVM), introduzida por Vapnik et al. (1979). Segundo Mountrakis et al. (2011), SVM vem sendo amplamente utilizada pela eficácia que tal método apresenta, sendo capaz de se equiparar ou até superar outros métodos de classificação.

Além da excelente formulação matemática, a possibilidade de emprego de diferentes funções kernel e estratégias multiclases tornam o método SVM ainda mais atrativo. De modo superficial, funções kernel podem ser empregadas a fim de possibilitar que o método SVM, formalizado na condição de “classificador linear”, realize classificações não lineares, as quais são

frequentes nas aplicações reais. Já as estratégias multiclases oferecem extensão à formulação original do método SVM, que permite a separação entre apenas duas classes, tenha seu uso generalizado para aplicações que independam do número de classes.

Diferentes investigações existentes na literatura lidam com o aprimoramento e/ou proposição de novas funções kernel e estratégias multiclases. Em especial, a consideração de múltiplas funções kernel é uma alternativa à seleção de uma única função. Neste contexto, a seleção de uma função kernel específica deixa de ser uma questão relevante durante o processo de treinamento, uma vez que diversas opções de kernel são empregadas simultaneamente. Dentre trabalhos relacionados ao uso de múltiplos kernels, são citados Gönen e Alpaydin (2001), Zhuang et al. (2011), Zien e Ong (2007) e Parreira (2016). Por outro lado, o desenvolvimento de novas estratégias multiclases é um tema abordado com menor frequência. Recentemente, Elkano et al. (2017) propõe uma variante da estratégia Um-Contra-Um, a qual originalmente consiste em resolver o problema multiclases a partir das análises individuais envolvendo cada um dos possíveis pares de classes. Uma investigação semelhante é conduzida em Hülsmann e Friedrich (2007).

Diante da influência exercida pela adequada seleção da função kernel e estratégia multiclases sobre a acurácia dos resultados de classificação proporcionado pelo método SVM, este estudo investiga combinações entre o uso de múltiplas funções kernel sobre diferentes estratégias multiclases. De modo superficial, são propostas duas arquiteturas de treinamento para o método SVM, as quais possibilitam a definição de variados hiperplanos de separação, segundo funções kernel distintas, as quais são posteriormente aplicados sobre os problemas binários que decompõem a estratégia multiclases adotada. As estratégias Um-Contra-Um e Um-Contra-Todos são analisadas neste estudo.

A fim de verificar o potencial das associações propostas, são realizados três estudos de caso envolvendo imagens obtidas pelos sensores IKONOS-II

OSA, ALOS PALSAR e LANDSAT-5 TM. Comparações com abordagens tradicionais, que envolvem a associação entre uma única função kernel ajustada de forma global para uma dada estratégia multiclases, são incluídas neste estudo.

## 2 Itens Teóricos.

### 2.1 Classificação de Imagens e SVM

Seja  $J$  uma imagem definida sobre um suporte  $S \in \mathbb{N}^2$  e  $\mathcal{X}$  o espaço de atributos,  $J(s) = \mathbf{x}$  denota que o pixel  $s \in S$  de  $J$  possui vetor de atributos  $\mathbf{x} \in \mathcal{X}$ . De modo genérico, um método de classificação de imagem é representado por uma função  $F: \mathcal{X} \rightarrow \Omega$ , que associa elementos do espaço de atributos  $\mathcal{X}$  a uma das classes de  $\Omega = \{\omega_1, \omega_2, \dots, \omega_c\}$ , com  $c \in \mathbb{N}^*$ .

Assim, a aplicação de  $F$  sobre  $J$  tem por objetivo a geração de uma representação simplificada que identifica diferentes classes/objetos presentes em  $J$ . Os diferentes métodos de classificação de imagem propostos na literatura podem ser entendidos como maneiras distintas de modelar  $F$  e aplicá-la em  $J$ . Para os métodos de classificação de aprendizado supervisionado, a modelagem de  $F$  faz uso de informações disponíveis em um conjunto de treinamento  $\mathcal{D} = \{(\mathbf{x}_i, \omega_j) \in \mathcal{X} \times \Omega : i = 1, \dots, m; j = 1, \dots, c\}$ , onde  $m$  é o número de dados de treinamento.

Dentre uma diversidade de propostas existentes na literatura, o método SVM tem sido frequentemente utilizado nas pesquisas em Sensoriamento Remoto (MOUNTRAKIS et al., 2011). Este método consiste em realizar a distinção entre amostras de treinamento a partir de um hiperplano que possui maior margem de separação. Tal hiperplano corresponde ao lugar geométrico onde a seguinte função é nula:

$$f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b, \quad (1)$$

sendo  $\mathbf{w}$  o vetor ortogonal ao hiperplano e  $b/||\mathbf{w}||$  a distância entre o hiperplano e a origem do espaço de atributos.  $\langle \cdot, \cdot \rangle$  denota o produto interno entre vetores. A determinação de  $\mathbf{w}$  e  $b$  é realizada mediante a resolução de um problema de otimização construído em função dos elementos em  $\mathcal{D}$  e de um parâmetro de penalidade  $C$  definido pelo analista. É importante observar que, no processo de classificação,  $f(\mathbf{x})$  atua como função discriminante. Nesta etapa, para um dado  $\mathbf{x}$ , o sinal de  $f(\mathbf{x})$  é um indicador sobre a classe ao qual o vetor analisado está associado. Maiores detalhes a respeito do processo de treinamento do método SVM podem ser encontrados em Webb e Copsey (2011).

Considerando que o método em discussão realiza a separação entre vetores de atributos a partir de um hiperplano definido no espaço de atributos original dos dados, duas suposições se desdobram: (i) um hiperplano realiza separação linear durante os processos de treinamento e classificação; (ii) um hiperplano possibilita a distinção de padrões dentre duas classes. No entanto, as suposições apontadas podem não ser válidas em aplicações práticas. As discussões conduzidas nas Seções 2.2 e 2.3 apresentam com maiores detalhes artifícios que podem auxiliar na verificação das suposições apontadas.

## 2.2 Funções Kernel

Em determinados problemas, a separação linear dos padrões pode ser dificultada devido à heterogeneidade dos dados. Em casos como este, as funções kernel podem ser usadas para facilitar tal separação, pois realizam um mapeamento implícito dos padrões para outro espaço de atributos onde a separabilidade é maior (NEGRI, 2013). Segundo Hofmann et al. (2008), as funções kernel podem ser representadas por:

$$K(\mathbf{x}, \mathbf{x}') = \langle \phi(\mathbf{x}), \phi(\mathbf{x}') \rangle; \quad \mathcal{X} \rightarrow \mathcal{H}, \quad (2)$$

onde  $\phi(\cdot)$  representa uma função responsável por remapear os padrões de  $\mathcal{X}$  para um novo espaço  $\mathcal{H}$ , denominado espaço característico de Hilbert. A

Figura 1 exemplifica um processo de remapeamento de elementos no espaço original para um novo espaço vetorial.

As funções da forma  $K(\cdot, \cdot)$  são ditas kernel desde que sejam simétricas e atendam às condições do Teorema de Mercer (SMOLA; SCHÖLKOPF, 2009). Quando tal teorema é atendido, é possível afirmar que existe um espaço vetorial onde  $K(\cdot, \cdot)$  representa um produto interno.

Apesar de serem muitas as propostas de função kernel existentes na literatura, como discutidas na obra de Christianini e Taylor (2004), apenas quatro delas são utilizadas com maior frequência nas aplicações em classificação de imagens, são elas:

$$K_{Linear}(\mathbf{x}, \mathbf{x}') = \langle \mathbf{x}, \mathbf{x}' \rangle \quad (3)$$

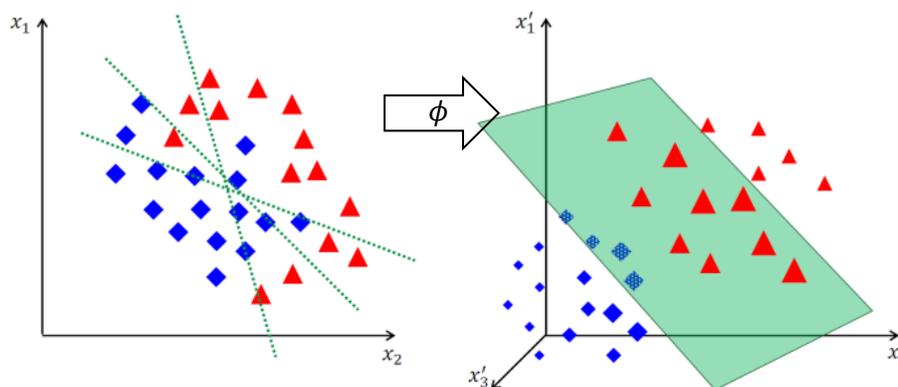
$$K_{Pol}(\mathbf{x}, \mathbf{x}') = ((\langle \mathbf{x}, \mathbf{x}' \rangle + 1))^p; p \in \mathbb{N}^* \quad (4)$$

$$K_{RBF}(\mathbf{x}, \mathbf{x}') = \exp(-\gamma \|\mathbf{x} - \mathbf{x}'\|); \gamma \in \mathbb{R}_+^* \quad (5)$$

$$K_{Sigm}(\mathbf{x}, \mathbf{x}') = \tanh(\alpha \langle \mathbf{x}, \mathbf{x}' \rangle + \beta); \alpha, \beta \in \mathbb{R} \quad (6)$$

sendo as Equações (3) a (6) denominadas por função kernel Linear, Polinomial, RBF e Sigmoid.

Figura 1 – Remapeamento do espaço de atributos original para um novo espaço cuja separação entre padrões é linear



Fonte: Negri (2013).

Para o problema envolvendo padrões não linearmente separáveis, as funções kernel podem substituir os produtos internos presentes na Equação (1) e nas equações que contemplam o treinamento do método SVM (cf. Webb e Copsey (2011)), produzindo assim um remapeamento dos padrões para um novo espaço vetorial cuja distribuição espacial possibilita a separação (linear) através de um hiperplano.

### 2.3 Estratégias Multiclasses

Como já observado, a separação entre padrões a partir de hiperplanos leva à definição do método SVM como um classificador binário. No entanto, a maioria dos problemas práticos abrange mais de duas classes de interesse, fazendo assim necessário sua extensão para problemas multiclasses.

Uma forma de lidar com problemas multiclasses é resolvê-los como um conjunto de subproblemas binários. Segundo esta abordagem, as diferentes estratégias multiclasses existentes são distintas entre si de acordo com a forma como os subproblemas binários são organizados e resolvidos. Dois exemplos são as estratégias multiclasses Um-Contra-Todos (*One-Against-All* – OAA) e Um-Contra-Um (*One-Against-One* – OAO).

Diante de um problema de classificação envolvendo  $c$  classes, sendo  $c > 2$ , a Estratégia OAA faz o uso de  $c$  classificadores binários. Para o  $l$ -ésimo classificador binário, é realizada a distinção dentre a classe  $\omega_l$  e as demais classes remanescentes. Dessa forma, um padrão  $\mathbf{x}$  é classificado pela seguinte regra de decisão:

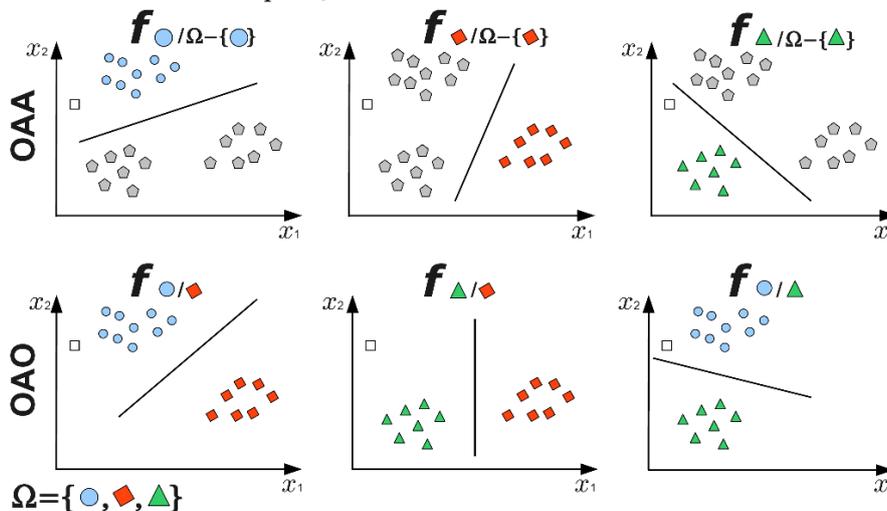
$$\mathbf{x} \in \omega_l \Leftrightarrow \arg \max_{1 \leq l \leq c} f_l(\mathbf{x}), \quad (7)$$

onde  $f_l$  é a função de classificação para o  $l$ -ésimo problema binário. Em resumo, a regra expressa em (7) estipula que o padrão  $\mathbf{x}$  pertencerá à classe cuja distância é maior em relação ao hiperplano de separação que realiza a separação dentre esta mesma classe e as demais.

Com relação a estratégia OAO, supondo novamente um problema com  $c > 2$  classes, são empregados  $\binom{c}{2}$  classificadores binários. Nesta estratégia, os padrões são classificados entre duas classes de cada vez, usando todas as combinações possíveis de pares de classes. Seguindo um esquema de “votação”, o padrão em classificação pertencerá à classe na qual foi atribuído em um número maior de vezes.

A Figura 2 ilustra o uso das estratégias OAA e OAO em um problema de classificação envolvendo 3 classes,  $\Omega = \{\circ, \diamond, \triangle\}$ . A notação  $f_{A/B}$  define que o hiperplano  $f$  realiza a separação entre os conjuntos  $A$  e  $B$ . Na estratégia OAA o padrão  $\square$  pertence à classe composta por elementos  $\circ$  devido a maior distância (perpendicular) observada diante o hiperplano  $f_{\circ/\Omega-\{\circ\}}$ . Apesar de verificada uma separação superior entre  $\square$  e o hiperplano  $f_{\diamond/\Omega-\{\diamond\}}$ , que segundo a discussão anterior faz  $\square$  pertencer à classe  $\Omega - \{\diamond\}$ , tal associação não permite determinar se este padrão pertence a classe  $\circ$  ou  $\triangle$ . Na estratégia OAO o padrão  $\square$  é classificado novamente como sendo da classe  $\circ$ , pois, dentre os três casos binários analisados ( $f_{\circ/\diamond}$ ,  $f_{\triangle/\diamond}$  e  $f_{\circ/\triangle}$ ), ele foi associado à classe  $\circ$  duas vezes, em  $f_{\circ/\diamond}$  e  $f_{\circ/\triangle}$ .

Figura 2 – Exemplo de aplicação das estratégias multiclassadas OAA e OAO em um problema de separação envolvendo três classes



Fonte: adaptado de Negri (2013).

### **3 Integração Otimizada Entre Múltiplos Kernels e Estratégias Multiclasses**

A Seção 2.3 consistiu em discutir o conceito básico de estratégias multiclasses e apresentar as estratégias OAA e OAO, usualmente empregadas nas aplicações em classificação de imagens que adotam o método SVM. Como mencionado na Seção 1, os estudos apresentados na literatura, comumente, fazem o emprego de uma combinação específica entre estratégia multiclasses e uma única função kernel. Desse modo, conforme discutido na Seção 2.3, cada subproblema binário que surge da decomposição do problema multiclasses é classificado com base em um único hiperplano de separação, obtido por uma função kernel escolhida e parametrizada a priori. Posteriormente, os resultados alcançados pelos diferentes hiperplanos são combinados para geração de uma classificação multiclasses final.

Uma alternativa a esta tendência usual consiste em definir diferentes hiperplanos de separação, segundo funções kernel distintas, para cada um dos subproblemas binários provenientes da decomposição multiclasses, utilizá-los no processo de classificação e combinar seus resultados para definição da classificação final. Nas Seções 3.1 e 3.2 são propostas duas arquiteturas de integração entre múltiplas funções kernel e as estratégias OAA e OAO.

#### **3.1 Arquitetura Híbrida**

Inicialmente, seja admitida uma estratégia multiclasses que decompõe o problema original em subproblemas binários. Uma arquitetura simples que permite o uso de múltiplos kernels em um mesmo problema de classificação consiste em determinar, para cada subproblema binário, diferentes hiperplanos de separação segundo funções kernel distintas. Em uma etapa posterior, durante o processo de classificação, cada elemento não rotulado, diante de cada um dos subproblemas binários, será avaliado pelo hiperplano que garante o maior nível de confiança de classificação. Este trabalho

emprega uma medida de avaliação de confiabilidade de classificação pelo método SVM proposta por Lin et al. (2007). É importante ressaltar que os hiperplanos considerados neste processo são determinados por parâmetros (do kernel e do SVM) ótimos.

Dessa forma, com a disponibilidade de diferentes hiperplanos, os quais são empregados na conveniência de garantir classificações binárias de maior confiança, são esperadas classificações mais acuradas. Esta característica alternável, que os hiperplanos de separação podem desempenhar durante o processo de classificação, denomina este processo por “arquitetura híbrida”.

Em termos formais, seja  $f_{\omega_i/\omega_j}$  um hiperplano responsável por separar elementos dentre as classes  $\omega_i$  e  $\omega_j$ . Analogamente, consideremos  $f_{\omega_i/\omega_j}^k$  um hiperplano que efetua o mesmo tipo de separação mencionada, porém, colocando em evidência que a função kernel  $k$  foi empregada na definição deste hiperplano. Nestas condições, a relação entre  $f_{\omega_i/\omega_j}$  e  $f_{\omega_i/\omega_j}^k$  é estabelecida por:

$$f_{\omega_i/\omega_j}(\mathbf{x}) = \arg \max_{k \in \mathcal{K}} f_{\omega_i/\omega_j}^k(\mathbf{x}), \quad (8)$$

onde  $\mathcal{K}$  representa o conjunto das funções kernel disponíveis para aplicação desta arquitetura.

Diante da formulação apresentada, verifica-se que a arquitetura híbrida é independente da estratégia multiclases adotada. Nas discussões a seguir, os acrônimos HOAA e HOAO denotam a associação entre a arquitetura híbrida e as estratégias OAA e OAO, respectivamente.

### 3.2 Combinação Linear

Em contrapartida à arquitetura híbrida, é proposta como alternativa uma combinação linear entre hiperplanos de separação obtidos a partir de diferentes funções kernel.

Nesta proposta, partindo de um conjunto de funções kernel definido *a priori* e de um problema de classificação decomposto em subproblemas binários, são obtidos hiperplanos para cada um destes subproblemas segundo as funções kernel disponíveis e supostamente parametrizadas de forma ótima. Em um segundo momento, é estabelecida uma combinação linear entre os hiperplanos obtidos anteriormente. Os coeficientes que ponderam tal combinação linear são definidos de forma que a aplicação da função resultante desta combinação em um processo de classificação proporcione resultados otimizados em termos de acurácia.

De modo análogo à formalização conduzida na Seção 3.1, seja  $f_{\omega_i/\omega_j}^k$  um hiperplano construído com uso de uma função kernel  $k$ , o qual é responsável por separar elementos dentre as classes  $\omega_i$  e  $\omega_j$ . Uma combinação linear entre hiperplanos que tratam do mesmo problema de separação (i.e., entre  $\omega_i$  e  $\omega_j$ ) segundo diferentes kernels é expresso por:

$$f_{\omega_i/\omega_j}(\mathbf{x}) = \sum_{k \in \mathcal{K}} \eta_k f_{\omega_i/\omega_j}^k(\mathbf{x}), \quad (9)$$

onde  $\eta_k$  representa o peso associado a cada função discriminante  $f_{\omega_i/\omega_j}^k(\mathbf{x})$ , de modo que  $\sum_{k=1}^{\#\mathcal{K}} \eta_k = 1$ . Novamente,  $\mathcal{K}$  representa o conjunto das funções kernel consideradas.

Cabe observar que os coeficientes que estabelecem a combinação linear expressa na Equação 9 devem ser determinados uma única vez e posteriormente são utilizados nos diferentes subproblemas binários que decompõem o problema multiclases original. Os resultados alcançados sobre cada subproblema são recombinados de acordo com a estratégia multiclases empregada inicialmente.

As denominações CLOAA e CLOAO referem-se às combinações entre as estratégias OAA e OAO e a arquitetura definida nesta seção.

## 4 Experimentos e Resultado

### 4.1 Materiais e Procedimentos metodológicos

Na Seção 3 foram propostas duas arquiteturas alternativas para o treinamento do método SVM. A fim de verificar o potencial destas propostas, foram realizados três experimentos sobre classificação de uso e cobertura do solo com emprego de imagens de Sensoriamento Remoto. Enquanto o primeiro experimento abrange uma área localizada no município de São José dos Campos, no estado de São Paulo, os outros dois experimentos têm como foco uma região no entorno da Floresta Nacional do Tapajós, localizada no estado do Pará.

A imagem apresentada na Figura 3, obtida pelo sensor OSA a bordo do satélite IKONOS-II, é referente à área de estudo do primeiro experimento mencionado. O sensor em questão é capaz de capturar imagens com 4 metros de resolução espacial, nos intervalos espectrais do azul ( $0,45-0,52\mu m$ ), verde ( $0,52-0,60\mu m$ ), vermelho ( $0,63-0,69\mu m$ ) e infravermelho próximo ( $0,76-0,90\mu m$ ), sendo que esta última não foi utilizada no presente estudo. Esta imagem possui  $600 \times 600$  pixels de dimensão.

Um extrato também de dimensões  $600 \times 600$  pixels de uma imagem adquirida pelo sensor PALSAR, a bordo do satélite ALOS, é ilustrado na Figura 4. Tal imagem possui resolução espacial de aproximadamente 20 metros nas polarizações HH, HV e VV em amplitude, e abrange uma porção da Floresta Nacional dos Tapajós.

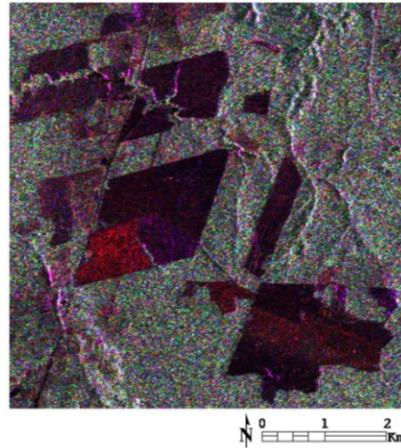
A última imagem de estudo, ilustrada na Figura 5, foi capturada pelo sensor TM a bordo do satélite LANDSAT-5. Tal imagem possui  $650 \times 650$  pixels de dimensão, 5 bandas espectrais, que compreendem o intervalo espectral do verde ao infravermelho médio, e resolução espacial de aproximadamente 30 metros. A imagem representa uma área mais abrangente no entorno da Floresta Nacional dos Tapajós, em comparação com a área ilustrada na Figura 4.

Figura 3 – Imagem IKONOS-II OSA, em composição natural, referente a uma região de São José dos Campos



Fonte: Elaborada pelos autores.

Figura 4 – Imagem ALOS PALSAR, composição R(HH)G(HV)B(VV), referente a uma região no entorno da Floresta Nacional do Tapajós



Fonte: Elaborada pelos autores.

Figura 5 – Imagem LANDSAT-5 TM, em composição natural, referente a uma região no entorno da Floresta Nacional do Tapajós

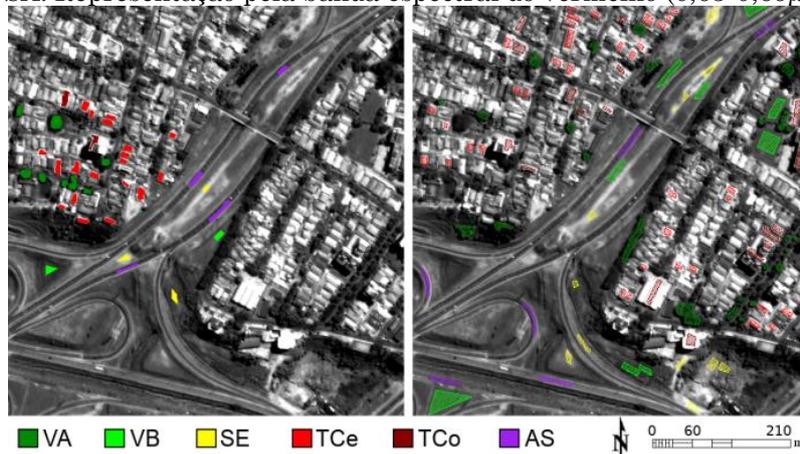


Fonte: Elaborada pelos autores

A fim de fornecer dados para treinamento do classificador e avaliação dos resultados, foram selecionadas amostras de verdades terrestres sobre as diferentes classes de cobertura do solo consideradas neste estudo. As Figuras 6, 7 e 8 apresentam as distribuições espaciais das amostras de treinamento e teste consideradas. Com relação à área de estudo referente ao município de São José dos Campos (Figura 6), foram consideradas as classes de Vegetação Alta (VA), Vegetação Baixa (VB), Telhados de Cerâmica (TCe), Telhados de Concreto (TCo), Solo Exposto (SE) e Asfalto (AS). Na área de estudo relativa

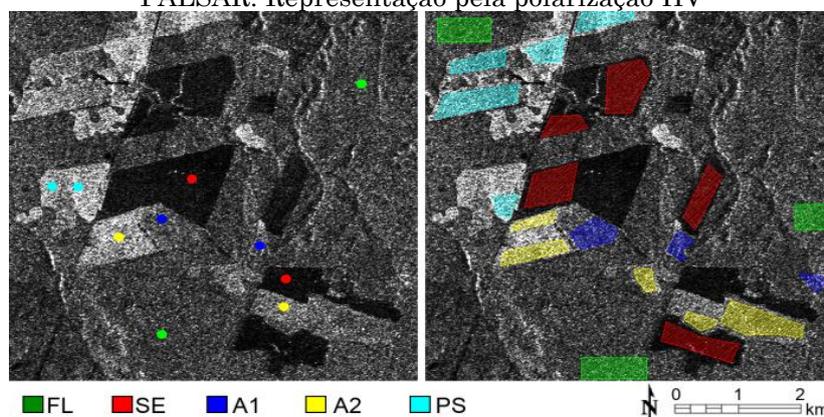
à imagem ALOS PALSAR (Figura 7), foram identificadas as classes de Floresta (FL), Solo Exposto (SE), Pastagem (PS) e dois tipos diferentes de Agricultura (A1 e A2). Por fim, na área abrangida pela imagem LANDSAT-5 (Figura 8), também referente à região da Floresta Nacional do Tapajós, foram identificadas as classes de Floresta (FL), Agricultura (AG), Pastagem (PS) e áreas de Regeneração Antiga (RA) e Jovem (RJ), onde a Regeneração Antiga corresponde à áreas de floresta que foram degradadas há mais de dez anos e a Regeneração Jovem há menos de 10 anos.

Figura 6 – Amostras de treinamento (esquerda) e teste (direita) para a imagem IKONOS-II OSA. Representação pela banda espectral do vermelho (0,63-0,69 $\mu$ m)



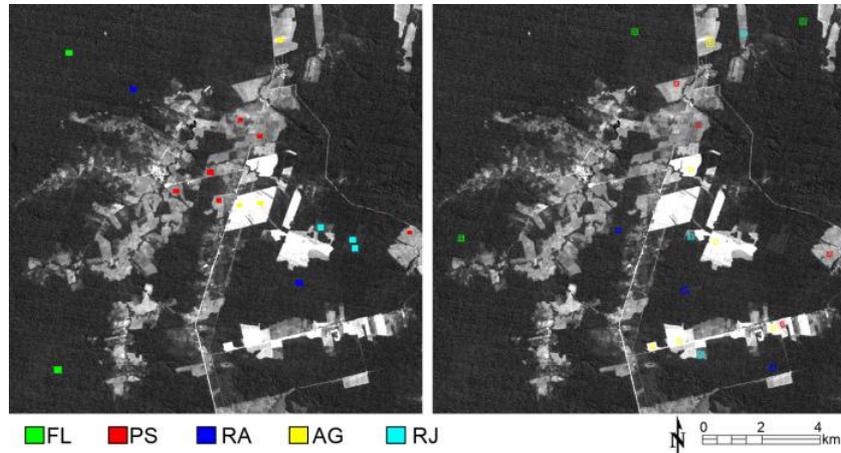
Fonte: Elaborada pelos autores.

Figura 7 – Amostras de treinamento (esquerda) e teste (direita) para a imagem ALOS PALSAR. Representação pela polarização HV



Fonte: Elaborada pelos autores.

Figura 8 – Amostras de treinamento (esquerda) e teste (direita) para a imagem LANDSAT-5 TM. Representação pela banda espectral do vermelho (0,63-0,69 $\mu$ m)



Fonte: Elaborada pelos autores.

As Tabelas 1, 2 e 3 sumarizam a quantidade de pixels que compõem as amostras de treinamento e teste com relação às imagens ilustradas nas Figuras 6, 7 e 8, respectivamente.

Tabela 1 – Quantidade de pixels utilizados para treinamento e teste na imagem IKONOS-II OSA.

Classe	Sigla	Treino	Teste
Vegetação Alta	VA	1045	3810
Vegetação Baixa	VB	255	3458
Telhado de Cerâmica	TCe	1632	2263
Telhado de Concreto	TCo	300	2741
Solo Exposto	SE	267	1323
Asfalto	AS	604	1440

Fonte: Elaborada pelos autores.

Tabela 2 – Quantidade de pixels utilizados para treinamento e teste na imagem ALOS PALSAR.

Classe	Sigla	Treino	Teste
Floresta	FL	320	8275
Solo Exposto	SE	320	14739
Agricultura 1	A1	320	3641
Agricultura 2	A2	320	8521
Pastagem	PS	320	9579

Fonte: Elaborada pelos autores.

Tabela 3 – Quantidade de pixels utilizados para treinamento e teste na imagem LANDSAT-5 TM.

<b>Classe</b>	<b>Sigla</b>	<b>Treino</b>	<b>Teste</b>
Floresta	FL	211	262
Agricultura	AG	177	405
Regeneração Antiga	RA	180	262
Pastagem	PS	356	204
Regeneração Jovem	RJ	270	216

Fonte: Elaborada pelos autores.

Primeiramente, em posse das implementações do método SVM e das diferentes estratégias multiclases e arquiteturas propostas, a citar, OAA, HOAA, CLOAA, OAO, HOAO e CLOAO, foram realizadas as classificações das imagens ilustradas nas Figuras 3, 4 e 5. Tais resultados permitem comparações simultâneas entre diferentes associações de estratégias multiclases e funções kernel. Neste trabalho foram considerados os kernels Linear, Polinomial, RBF e Sigmoides, apresentados nas Equações 3 a 6, respectivamente.

Previamente ao processo de classificação, é necessária a identificação dos parâmetros adequados para o método SVM (penalidade –  $C$ ) e para as funções kernel consideradas. Para isso, independente da estratégia ou arquitetura adotada, foi executado um procedimento de busca exaustiva (*grid search*) com validação cruzada em 10 vezes. Tal procedimento consiste em testar todas as combinações de parâmetros dentro de um espaço de busca delimitado inicialmente. Para cada configuração, é realizado o treinamento do método com 9/10 partes do conjunto original de treinamento e em seguida é calculada sua acurácia de classificação sobre 1/10 dos dados remanescentes. Este procedimento é repetido para a mesma configuração de parâmetros 10 vezes, sendo que cada repetição possui conjuntos de treinamento e testes distintos, e ao fim a média das acurácias computada representa a acurácia geral da configuração analisada. Cabe mencionar que este processo foi limitado ao uso dos dados destinados inicialmente ao treinamento com a finalidade de reservar os dados de teste ao cálculo das estimativas de

acurácias, evitando assim que o desempenho (i.e., capacidade de generalização) do método possa ser influenciado por características específicas aprendidas durante a fase de treinamento. O coeficiente de concordância kappa (CONGALTON; GREEN, 1999) foi empregado como medida de acurácia. Nos experimentos a seguir, os seguintes espaços de busca foram considerados: penalidade do método SVM –  $C \in \{1, 10, 100, 1000, 10000\}$ , grau do kernel Polinomial –  $p \in \{2, 3, 5, 7\}$  e parâmetros regularizadores dos kernels RBF e Sigmoides –  $\alpha, \gamma \in \{0,1; 0,25; 0,75; 1,00; 1,25; 1,50; 1,75; 2,00\}$ . O parâmetro  $\beta$  presente no kernel Sigmoides foi fixado em zero.

Quando as estratégias usuais OAA e OAO são consideradas, a busca pelos parâmetros adequados é realizada considerando uma mesma função kernel em todos os subproblemas binários que decompõem o problema multiclases. É importante ressaltar que, para este caso, os parâmetros escolhidos (i.e., penalidade do SVM e parâmetro de ajuste do kernel), são fixos para todos os subproblemas envolvidos.

Com relação às arquiteturas HOAA e HOAO, o ajuste dos parâmetros é conduzido com um maior grau de liberdade. Inicialmente, o problema multiclases é decomposto em subproblemas binários, conforme define a própria arquitetura em questão. Posteriormente, para cada subproblema binário, é feita a definição de um hiperplano de separação para cada uma das funções kernel disponíveis. Para isso, são selecionados os parâmetros penalidade e da função kernel segundo um processo de busca exaustiva com validação cruzada em 10 vezes, conforme discutido anteriormente. A fim de exemplificação, supondo um problema multiclases envolvendo quatro classes e uma decomposição em seis subproblemas binários, com base na estratégia OAO, e admitindo a disponibilidade de quatro funções kernel distintas, o processo que contempla a arquitetura HOAO envolve a determinação de 24 hiperplanos de separação (6 hiperplanos binários  $\times$  4 funções kernel). É importante observar que para cada função kernel será realizado o ajuste do valor de parâmetro que garante o melhor desempenho diante dos diferentes subproblemas binários. O ajuste do parâmetro de penalidade é também

definido de modo global. Por fim, durante a etapa de classificação de cada elemento (i.e., pixel) não rotulado, é feito uso da Equação 8 nos diferentes subproblemas binários e os respectivos resultados são combinados de acordo com a estratégia “base” adotada.

Nas arquiteturas CLOAA e CLOAO, a decomposição do problema multiclass e respectivo ajuste dos hiperplanos são realizados de modo similar às arquiteturas HOAA e HOAO. Nestas condições, haverá tantos hiperplanos por subproblema binário quantos forem as funções kernel disponíveis. Posteriormente, é definida uma combinação linear envolvendo os hiperplanos obtidos na etapa anterior. Os coeficientes que ponderam esta combinação linear, conforme definidos na Equação 9, são determinados por um novo processo de busca exaustiva com validação cruzada em 10 vezes sobre os dados de treinamento, visando a maximização da acurácia. Os coeficientes considerados neste processo estão limitados ao espaço de busca  $\eta_i \in \{0; 0,025; 0,05; \dots; 1\}$ . Em uma última etapa, a classificação dos elementos não rotulados é conduzida com base na Equação 9 sobre cada subproblema binário e os respectivos resultados são combinados de acordo com a estratégia “base” adotada.

Após a seleção dos parâmetros, segundo os processos descritos, o método SVM foi treinado utilizando o conjunto completo de amostras de treinamento e então aplicado na classificação das imagens de estudo. Em seguida, as amostras de teste foram consideradas no cálculo de medidas de concordância, possibilitando a análise da qualidade do resultado. O coeficiente de concordância kappa e seu desvio padrão foram empregados para esta finalidade. Ainda, foram aplicados testes de hipótese para comparação entre os diferentes valores de kappa obtidos e verificar a significância estatística entre eles. A diferença estatística entre os resultados será discutida a um nível de significância de 5%.

Para implementação do método SVM e das arquiteturas propostas, foi adotada a linguagem IDL (*Iterative Data Language*). A biblioteca LibSVM

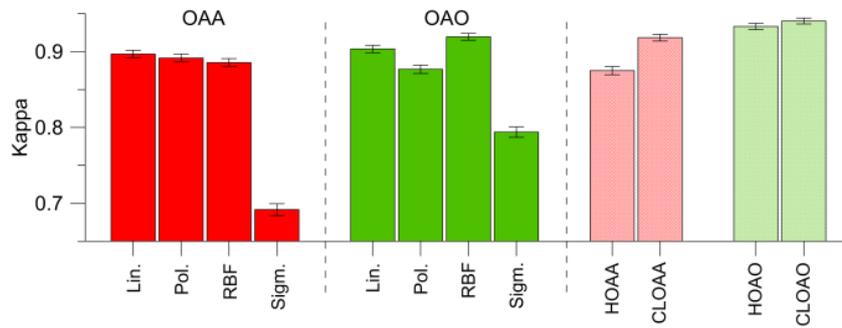
(CHANG e LIN, 2011) foi adotada para a resolução do problema de otimização associado ao treinamento do método SVM.

## 4.2 Resultados

Diante das imagens e das amostras de treinamento e teste e seguindo os procedimentos metodológicos delimitados na Seção 4.1, foram obtidos diferentes resultados de classificação. Os valores dos coeficientes kappa, que avaliam a qualidade da classificação, assim como o desvio padrão associado a cada classificação, são apresentados nos gráficos das Figuras 9, 10 e 11, os quais são referentes às imagens IKONOS-II OSA, ALOS PALSAR e LANDSAT-5 TM, respectivamente. Nas Tabelas 4, 5 e 6 são exibidos *p*-valores provenientes de testes de hipóteses realizados para comparação entre os diferentes resultados de classificação obtidos para cada uma das imagens consideradas. Valores em negrito destacam resultados equivalentes diante uma significância de 5%.

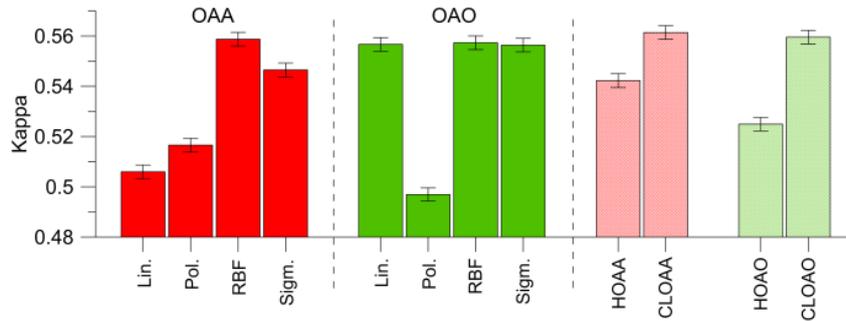
Diante os resultados obtidos sobre a imagem IKONOS-II, é possível verificar que as arquiteturas de múltiplos kernels associadas à estratégia OAO foram capazes de proporcionar resultados superiores em comparação às demais abordagens. Ainda, HOAO e CLOAO apresentam acurácias similares. No entanto, com emprego da estratégia OAA, a arquitetura baseada em combinação linear se torna equivalente ao uso do kernel RBF, assim como a arquitetura híbrida aos kernels Polinomial e RBF, como exibem os *p*-valores da Tabela 4. Cabe observar que a arquitetura híbrida não supera a proposta por combinação linear.

Figura 9 – Acurácia das classificações referentes à imagem IKONOS-II OSA



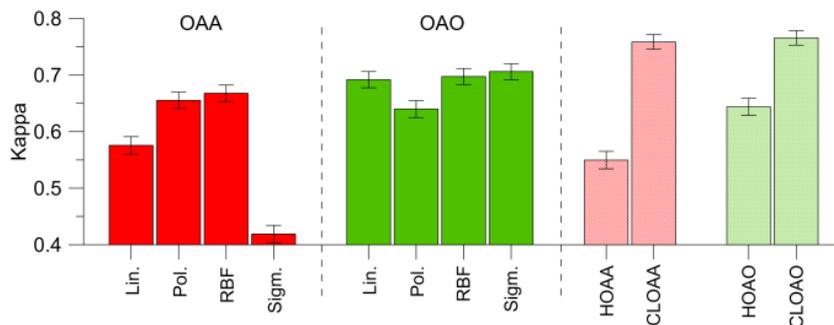
Fonte: Elaborada pelos autores.

Figura 10 – Acurácia das classificações referentes à imagem ALOS PALSAR



Fonte: Elaborada pelos autores.

Figura 11 – Acurácia das classificações referentes à imagem LANDSAT-5 TM



Fonte: Elaborada pelos autores.

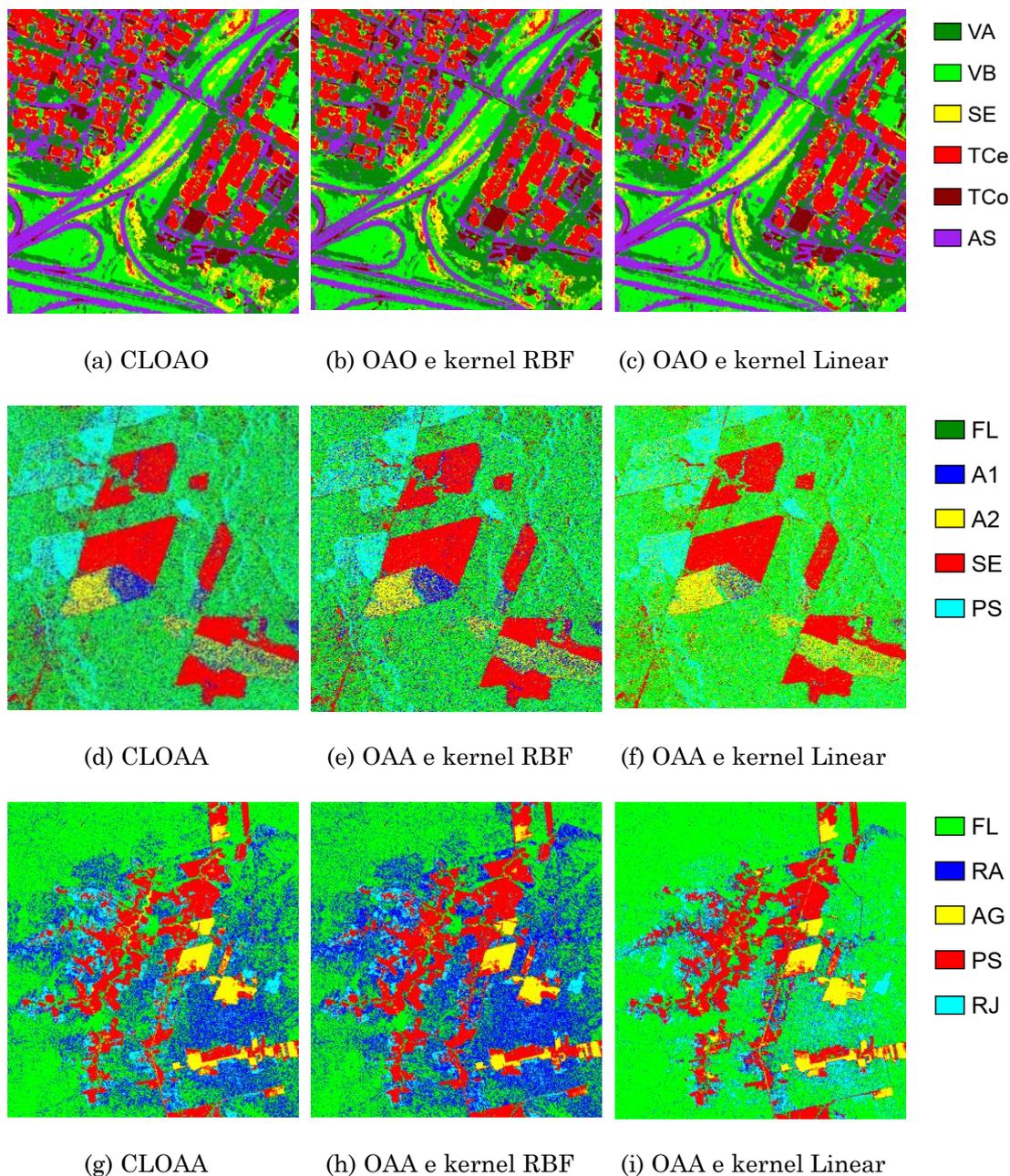
Com relação aos desempenhos obtidos sobre a imagem ALOS PALSAR, as arquiteturas de múltiplos kernels analisadas não foram superiores, de modo unânime, diante as abordagens com único kernel. Considerando apenas

o uso da estratégia OAA, o caso extremo que impediu tal unanimidade ocorreu diante da similaridade entre o uso isolado do kernel RBF e a arquitetura baseada em combinação linear. Ao considerar a estratégia OAO e com base nos  $p$ -valores mostrados na Tabela 5, é possível observar que as funções kernel Polinomial, RBF e Sigmoide são capazes de proporcionar resultados similares aos obtidos com uso da arquitetura baseada em combinação linear. Ainda, ao comparar apenas as arquiteturas propostas, é evidente a superioridade da abordagem por combinação linear, independente da estratégia multiclases adotada.

Por fim, de acordo com os resultados obtidos sobre a imagem LANDSAT-5 TM, o uso de múltiplos kernels revela vantagens novamente quando a abordagem baseada em combinação linear é considerada. A arquitetura híbrida permanece inferior ao esquema definido por combinação linear entre hiperplanos, sendo esta última insensível à estratégia multiclases adotada, como mostra a Tabela 6.

A Figura 12 apresenta, para cada imagem considerada, o melhor resultado obtido dentre as arquiteturas propostas, assim como os resultados alcançados segundo a abordagem tradicional com uso das funções kernel RBF e Linear. De acordo com os valores de kappa apresentados nos gráficos das Figuras 9 a 11, o kernel RBF possibilitou que as estratégias OAA e OAO proporcionassem resultados próximos ao uso de múltiplas funções kernel por combinação linear de hiperplanos. A importância quanto ao uso de funções kernel fica evidenciada ao comparar os resultados obtidos pelas funções RBF e Linear, uma vez que a adoção do kernel Linear pode ser interpretada como uma abordagem “sem kernel”.

Figura 12 – Resultados de classificações relevantes obtidas sobre as imagens IKONOS-II, ALOS e LANDSAT-5



Fonte: Elaborada pelos autores.

Mediante uma visão geral dos resultados obtidos, independentemente da imagem considerada, é possível verificar que a arquitetura híbrida não apresentou resultados satisfatórios, sendo frequentemente inferior às abordagens convencionais. Por outro lado, a arquitetura baseada em

combinação linear foi capaz de proporcionar resultados satisfatórios em todos os estudos de caso realizados. Ainda, tal arquitetura mostrou superioridade estatística na maioria as comparações.

Tabela 4 –  $p$ -valores ( $\times 10^{-3}$ ) referentes aos testes de hipóteses realizados para comparação entre valores de kappa obtidos a partir da classificação da imagem IKONOS-II OSA segundo as configurações consideradas. Valores destacados identificam equivalência estatística mediante significância de 5%

		OAA			OAO				Propostas			
		Pol.	RBF	Sig.	Lin.	Pol.	RBF	Sig.	HOAA	CLOAA	HOAO	CLOAO
OAA	Lin.	<b>471</b>	<b>123</b>	000	<b>339</b>	007	001	000	003	001	000	000
	Pol.	–	<b>412</b>	000	<b>094</b>	048	000	000	025	000	000	000
	RBF		–	000	012	<b>245</b>	000	000	<b>158</b>	000	000	000
	Sig.			–	000	000	000	000	000	000	000	000
OAO	Lin.				–	000	016	000	000	025	000	000
	Pol.					–	000	000	<b>812</b>	000	000	000
	RBF						–	000	000	<b>856</b>	026	000
	Sig.							–	000	000	000	000
Propo	HOAA								–	000	000	000
	CLOAA									–	016	000
	HOAO										–	<b>205</b>

Fonte: Elaborada pelos autores.

Tabela 5 –  $p$ -valores ( $\times 10^{-3}$ ) referentes aos testes de hipóteses realizados para comparação entre valores de kappa obtidos a partir da classificação da imagem ALOS PALSAR segundo as configurações consideradas. Valores destacados identificam equivalência estatística mediante significância de 5%

		OAA			OAO				Propostas			
		Pol.	RBF	Sig.	Lin.	Pol.	RBF	Sig.	HOAA	CLOAA	HOAO	CLOAO
OAA	Lin.	006	000	000	000	016	000	000	000	001	000	000
	Pol.	–	000	000	000	000	000	000	000	000	029	000
	RBF		–	002	<b>592</b>	000	<b>711</b>	<b>540</b>	000	<b>487</b>	000	<b>832</b>
	Sig.			–	008	000	005	010	<b>275</b>	000	000	001
OAO	Lin.				–	000	<b>868</b>	<b>938</b>	000	<b>218</b>	000	<b>454</b>
	Pol.					–	000	000	000	000	000	000
	RBF						–	<b>808</b>	000	<b>286</b>	000	<b>560</b>
	Sig.							–	000	<b>191</b>	000	<b>409</b>
Propo	HOAA								–	000	000	000
	CLOAA									–	000	<b>629</b>
	HOAO										–	000

Fonte: Elaborada pelos autores.

Tabela 6 –  $p$ -valores ( $\times 10^{-3}$ ) referentes aos testes de hipóteses realizados para comparação entre valores de kappa obtidos a partir da classificação da imagem LANDSAT-5 TM segundo as configurações consideradas. Valores destacados identificam equivalência estatística mediante significância de 5%

		OAA			OAO				Propostas			
		Pol.	RBF	Sig.	Lin.	Pol.	RBF	Sig.	HOAA	CLOAA	HOAO	CLOAO
OAA	Lin.	000	000	000	000	003	000	000	<b>233</b>	000	002	000
	Pol.	–	<b>543</b>	000	<b>073</b>	<b>461</b>	038	012	000	000	<b>582</b>	000
	RBF		–	000	<b>239</b>	<b>182</b>	<b>146</b>	<b>058</b>	000	000	<b>251</b>	000
	Sig.			–	000	000	000	000	000	000	000	000
OAO	Lin.				–	012	<b>782</b>	<b>476</b>	000	001	021	000
	Pol.					–	005	001	000	000	<b>855</b>	000
	RBF						–	<b>663</b>	000	002	010	000
	Sig.							–	000	006	002	002
Propo	HOAA								–	000	000	000
	CLOAA									–	000	<b>692</b>
	HOAO										–	000

Fonte: Elaborada pelos autores.

## 5 Conclusão

Neste estudo é realizada uma investigação sobre o uso de múltiplas funções kernel no método SVM na consideração de diferentes estratégias multiclases. Duas arquiteturas para lidar com o uso de múltiplas funções kernel são apresentadas. A primeira, denominada “híbrida”, é baseada no uso da função kernel que maximiza a acurácia das classificações em cada um dos subproblemas binários que decompõem o problema multiclases. Por outro lado, a segunda arquitetura consiste em uma combinação linear que pondera diferentes hiperplanos de separação definidos a partir de funções kernel distintas. A fim de verificar o potencial destas arquiteturas, foram realizados três experimentos sobre classificação de uso e cobertura do solo a partir de imagens obtidas pelos satélites IKONOS-II, ALOS e LANDSAT-5.

A partir das classificações obtidas, verifica-se a eficácia da arquitetura baseada em combinação linear, seja ela com uso das estratégias OAA ou OAO, cujas acurácias associadas atingem valores superiores às demais configurações envolvendo uma das estratégias citadas e uma única função kernel. Apesar da equivalência estatística entre a arquitetura baseada em combinação linear e com uso apenas da função RBF, tal arquitetura dispensa

o processo de seleção do kernel mais adequada, já que este processo engloba simultaneamente as funções kernel que usualmente são selecionadas e avaliadas individualmente. Não é possível identificar a superioridade de uma estratégia multiclases com relação à outra, uma vez que CLOAA e CLOAO foram estatisticamente equivalentes em dois dos três casos analisados.

Por outro lado, a arquitetura híbrida não foi capaz de proporcionar resultados relevantes. Tal arquitetura mostrou resultados competitivos apenas quando associada à estratégia OAO e aplicada na classificação da imagem IKONOS-II.

Por fim, dentre os resultados obtidos com uso de uma única função kernel, observa-se que a seleção de tal função atinge diretamente a acurácia dos resultados. Diante dos desempenhos observados, e como já observado em Mountrakis et al. (2011), a função RBF aponta como uma boa opção dentre os kernels convencionais. Ainda, não é possível apontar a superioridade dentre uma das estratégias consideradas com relação a outra.

## **Agradecimentos**

Os autores agradecem a FAPESP (Procs. 2014/14830-8 e 2016/02875-2) pelo suporte financeiro.

## **Referências**

- ALPAYDIN, E.; GÖNEN, M. Multiple kernel learning algorithms. **Journal of machine learning research**, vol. 12, 2011.
- BUSTINCE, H.; ELKANO, M.; GALAR, M.; LUCCA, G.; SANZ, J. IVOVO: A new interval-valued one-vs-one approach for multi-class classification problems. **2017 Joint 17th World Congress of International Fuzzy Systems Association and 9th International Conference on Soft Computing and Intelligent Systems (IFSA-SCIS)**, 2017. pp. 1–6,

- CHANG, C.; LIN, C. **LIBSVM : a library for support vector machines**. ACM Transactions on Intelligent Systems and Technology, 2011.
- CONGALTON, R. G.; GREEN, K. Assessing the accuracy of remotely sensed data: Principles and Practices. **CRC Press**, 2009. 183p.
- DEVRIES, B.; VERBESSELT, J.; KOOISTRA, L.; HEROLD, M. Robust monitoring of small-scale forest disturbances in a tropical montane forest using Landsat time series. **Remote Sensing of Environment**, vol.161, 2015. pp. 107-121
- FRIEDRICH, C. M.; HÜLSMANN, M. Comparison of a Novel Combined ECOC Strategy with Different Multiclass Algorithms Together with Parameter Optimization Methods. **Machine Learning and Data Mining in Pattern Recognition**, Berlin, Heidelberg: Springer Berlin Heidelberg, 2007. pp. 17–31
- GÓMEZ-ZOTANO, J.; ALCÁNTARA-MANZANARES, J.; MARTÍNEZ-IBARRA, E.; OLMEDO-COBO, J. A. Applying the technique of image classification to climate science: the case of Andalusia (Spain). **Geographical Research**, vol. 54, 2016. pp.461-470
- HOFMANN, T.; SCHÖLKOPF, B.; SMOLA, A. J. Kernel methods in machine learning. **Annals of Statistics**, vol. 36, n. 3, 2008. pp. 1171–1220
- HOI, S. C. H.; TSANG, I. W.; ZHUANG, J. Two-Layer Multiple Kernel Learning. **Proceedings of Machine Learning Research**, 2011. pp. 909–917
- IM, J.; MOUNTRAKIS, G.; OGOLE, C. Support Vector Machines in Remote Sensing: A review. **ISPRS Journal of Photogrammetry and Remote Sensing Society**, vol. 66, 2011. pp. 247-259
- LIN, C.; LIN, H.; WENG, R. C. A note on platt's probabilistic outputs for support vector machines. **Machine Learning, Kluwer Academic Publishers**, vol. 68, n. 3, 2007. pp.267–276
- LU, D. Aboveground biomass estimation using Landsat TM data in the Brazilian Amazon. **International Journal of Remote Sensing**, vol. 26, 2005. pp. 2509-2525
- NEGRI, R. G. **Máquina de Vetores de Suporte Adaptativa ao Contexto: Formalização e Aplicações em Sensoriamento Remoto**. Tese de Doutorado em Computação Aplicada, Instituto Nacional de Pesquisas Espaciais, 2013.

- ONG, C. S.; ZIEN, A. Multiclass multiple kernel learning. **Proceedings of the 24th international conference on Machine learning**, 2007. pp. 1191–1198
- PARREIRA, M. D. O. **HSMI: método de classificação hierárquico baseado em SVM multikernel com otimização meta-heurística**. Tese de Doutorado em Computação Aplicada, Instituto Nacional de Pesquisas Espaciais, 2017.
- SCHÖLKOPF, B.; SMOLA, A. J. **Learning with kernels : support vector machines, regularization, optimization, and beyond**. MIT Press, Cambridge, MA, USA, 2002.
- SHAWE-TAYLOR, J.; CRISTIANINI, N. **Kernel Methods for Pattern Analysis**. Cambridge University Press, New York, NY, USA, 2004.
- VAPNIK, V. N. **The nature of Statistical learning theory**. New York: Springer-Verlag, 1995.
- WEBB, A. R.; COPSEY, K. D. **Statistical pattern recognition**. Wiley, 2011.