

**COMBINAÇÃO DE MÚLTIPLAS ABORDAGENS  
DE CLASSIFICAÇÃO PARA INTERPRETAÇÃO  
DE IMAGENS HIPERESPECTRAIS DE  
SENSORIAMENTO REMOTO**

ANDREY BICALHO SANTOS

COMBINAÇÃO DE MÚLTIPLAS ABORDAGENS  
DE CLASSIFICAÇÃO PARA INTERPRETAÇÃO  
DE IMAGENS HIPERESPECTRAIS DE  
SENSORIAMENTO REMOTO

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais como requisito parcial para a obtenção do grau de Mestre em Ciência da Computação.

ORIENTADOR: ARNALDO DE ALBUQUERQUE ARAÚJO

COORIENTADOR: DAVID MENOTTI GOMES

Belo Horizonte

Abril de 2013

© 2013, Andrey Bicalho Santos.  
Todos os direitos reservados.

S237c Santos, Andrey Bicalho  
Combinação de Múltiplas Abordagens de  
Classificação para Interpretação de Imagens  
Hiperspectrais de Sensoriamento Remoto / Andrey  
Bicalho Santos. — Belo Horizonte, 2013  
xxi, 83 f. : il. ; 29cm

Dissertação (mestrado) — Universidade Federal de  
Minas Gerais

Orientador: Arnaldo de Albuquerque Araújo  
Coorientador: David Menotti Gomes

1. Sensoriamento Remoto. 2. Processamento de  
Imagens. 3. Múltiplos Classificadores. 4. Algoritmos  
Genéticos. 1. Computação – Teses. 2. Processamento de  
imagens – Teses. 3. Algoritmos genéticos –  
Teses. I. Orientador. II. Coorientador. III. Título.

CDU 519.6\*84



UNIVERSIDADE FEDERAL DE MINAS GERAIS  
INSTITUTO DE CIÊNCIAS EXATAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

## FOLHA DE APROVAÇÃO

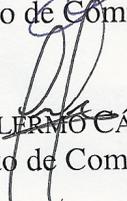
Combinação de múltiplas abordagens de classificação para interpretação de  
imagens hiperespectrais de sensoriamento remoto

**ANDREY BICALHO SANTOS**

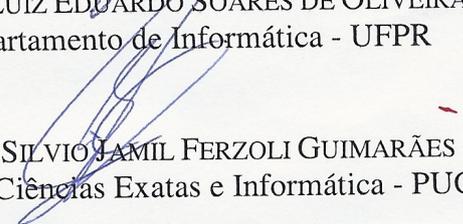
Dissertação defendida e aprovada pela banca examinadora constituída pelos Senhores:

  
PROF. ARNALDO DE ALBUQUERQUE ARAÚJO - Orientador  
Departamento de Ciência da Computação - UFMG

  
PROF. DAVID MENOTTI GOMES - Coorientador  
Departamento de Computação - UFOP

  
PROF. GUILLERMO CÁMARA CHÁVEZ  
Departamento de Computação - UFOP

  
PROF. LUIZ EDUARDO SOARES DE OLIVEIRA  
Departamento de Informática - UFPR

  
PROF. SILVIO JAMIL FERZOLI GUIMARÃES  
Instituto de Ciências Exatas e Informática - PUC/MG

Belo Horizonte, 19 de abril de 2013.

*Aos meus pais:*  
*José Anísio Tavares Santos*  
*Antônia Maria Bicalho Santos*

# Agradecimentos

Em primeiro lugar agradeço aos meus pais José Anísio Tavares Santos e Antônia Maria Bicalho Santos pelo incentivo e apoio contínuo em meus estudos.

Ao meu orientador Arnaldo de Albuquerque Araújo pelo acompanhamento e pela oportunidade concedida para fazer parte do Núcleo de Processamento Digital de Imagens (NPDI) da UFMG.

Ao meu co-orientador David Menotti Gomes pelos conhecimentos transmitidos, amizade e conselhos. Sem a sua participação e longas discussões que tivemos este trabalho não estaria no nível que se encontra.

A minha namorada Jéssica pelo carinho, paciência e constante apoio nos momentos difíceis, tornando meus dias mais felizes.

Ao meu irmão Adriano pelas diversas caronas até a UFMG.

Aos colegas do NPDI: Alberto, Bruno, Caetano, Cássio, Guilherme, Henrigue, Kleber, Marcelo, Sandra, Suellen, Tiago e Vírginia.

Aos amigos que assim como eu deixaram suas cidades e vieram para Belo Horizonte para iniciar essa jornada: Clayson, Rodolfo, Alcemir, Aécio, Evellyn, Juliana, Glauber, Fabrício.

A FAPEMIG pela bolsa concedida durante o período de mestrado.

Finalmente, obrigado a todos que diretamente ou indiretamente contribuíram para a realização deste trabalho.

*“A ausência da evidência não significa evidência da ausência.”*

(Carl Edward Sagan)

# Resumo

Recentemente, a análise de dados hiperespectrais têm sido o principal meio para interpretação de imagens de sensoriamento remoto. Uma imagem hiperespectral é composta por centenas de canais, cada um corresponde a uma determinada faixa do espectro eletromagnético. Esse tipo de dados fornece detalhada informação sobre objetos na superfície da terra, podendo nos conduzir a uma profunda investigação, e consequentemente, a uma interpretação mais precisa. A classificação da cobertura terrestre é ainda um desafio, e produzir mapas temáticos mais precisos é um objetivo comum entre os pesquisadores. Embora cada *pixel* na imagem forneça uma rica informação espectral, a união da informação espacial e espectral se faz necessária devido às relações entre *pixels* adjacentes. Entretanto, combinar esses dois tipos de informações ainda é um tópico aberto e amplamente estudado na comunidade de sensoriamento remoto. Neste trabalho aplicamos técnicas de meta-aprendizado para combinar resultados de múltiplas abordagens de classificação visando produzir mapas temáticos mais precisos. Para isso, quatro abordagens de classificação foram usadas na combinação, sendo que duas delas também fazem parte da proposta deste trabalho. Na tentativa de maximizar o ganho de informação na combinação, cada uma das abordagens de classificação possuem representações de características diferentes e algoritmos de aprendizado (classificadores) de natureza diferentes. Foram usadas duas representações baseando-se apenas na informação espectral e outras duas que se baseiam na união das informações espectral e espacial. Três diferentes tipos de algoritmo de aprendizagem foram aplicados. Os algoritmos *Support Vector Machines (SVM)* e *K-Nearest Neighbor (KNN)* foram usados para classificar os dados baseando-se nas representações da informação espectral, enquanto que *Multilayer Perceptron Neural Network (MLP)* e, novamente, SVM foram usados nas informações do tipo espectral-espacial. Dois métodos para a combinação das quatro abordagens de classificação foram propostos. Nossa primeira proposta é baseada em uma combinação linear ponderada, em que os pesos são encontrados usando Algoritmos Genéticos (*Weighted Linear Combination optimized by Genetic Algorithm (WLC-GA)*). A segunda, *Stacking-SVM*, realiza uma combinação

não linear por meio da estratégia conhecida como *Stacked Generalization*. Para isso o algoritmo de aprendizado *SVM* com *kernel Radial Basis Function (RBF)* foi utilizado. Experimentos foram realizados em duas bases de dados bem conhecidas: *Indian Pines* e *Pavia University*. Com o objetivo de avaliar o comportamento e a robustez dos métodos de combinação, experimentos em diferentes cenários e diferentes conjuntos de treinamento foram conduzidos. Foi observado que o método proposto *WLC-GA* produz resultados promissores nas duas bases de dados utilizadas, obtendo os maiores índices de acurácia dentre *Stacking-SVM* e outros métodos tradicionais como *Majority Vote (MV)* e *Average rule*. Com apenas 10% de amostras de treino, o método *WLC-GA* foi capaz de encontrar conjuntos de pesos ótimos, ou sub-ótimos, que permitem a obtenção de mapas temáticos mais precisos, superando algumas desvantagens de outros métodos de combinação.

# Abstract

In the past few decades hyperspectral data analysis has become the main source for classification of remote sensed images. A hyperspectral image is composed of hundreds of spectral channels, where each channel refers to a specific wavelength. Such a large amount of information may lead us to a deeper investigation of the materials on Earth's surface, and thus, a more precise interpretation of them. Land cover classification is still a challenge task, and producing accurate thematic map is a common goal among researchers. Although each pixel in a hyperspectral image has a detailed spectral information, the joint of both spectral and spatial information is required due to its relations among adjacent pixels. However, joining such information is a widely studied and opened topic in the remote sensing community. In this work we applied meta-learning techniques to combine multiple classification approaches aiming to produce a more accurate thematic maps. For this purpose, four classification approaches were used in the combination, in which two of them is also a part of this work. In an attempt to maximize the information gain in the combination, each one of the classification approaches has its own feature representation and also uses learning algorithms from different nature. We used two feature representation based on only spectral info and another two based on both spectral-spatial info. Three different learning algorithms were applied. The SVM and KNN classifiers were used to classify spectral based data, while MLP and, once again, SVM on spectral-spatial based data. Two methods to combine the four classification approaches were proposed. Our first proposal is based on on Weighted Linear Combination (WLC), in which weights are found by using a Algoritmos Genéticos (GA) - WLC-GA. The second one, *Stacking-SVM*, performs a nonlinear combination by means of *Stacked Generalization* strategy. In order to do it, the SVM learning algorithm with RBF kernel was used. Experiments were carried out with two well-known datasets: *Indian Pines* and *Pavia University*. In order to evaluate the robustness of the proposed combiner, experiments using different training sizes on different scenarios were conducted. It was observed that the proposed WLC-GA method achieved promising results on both datasets used, owning the highest

accuracy levels among *Stacking-SVM* and other traditional methods, such as MV and *Average* rule. With only 10% of training samples, the WLC-GA method was able to find optimal, or suboptimal, set of weights which allows an accurate combination of the classification approaches, and hence, a more precise thematic maps, overcoming some drawbacks of other combination methods.

# Lista de Figuras

1.1	Esquema ilustrativo de uma imagem hiperspectral (hipercubo). Modificado de Tarabalka (2010). . . . .	4
2.1	Exemplo esquemático da classificação espectral-espacial usando <i>MV</i> dentro das regiões segmentadas. Ilustração retirada de Tarabalka et al. (2010a). . .	14
3.1	Perfil de Decisão ( <i>Decision Profile (DP)</i> ). Colunas são valores de confiança para $c$ classes e linhas são os $L$ métodos de classificação. Adaptado de Kuncheva (2004). . . . .	24
3.2	Exemplo de informação estrutural: grande, iluminado, estrutura retangular - pequeno, circular, estrutura texturizada - comprido, fino, estrutura escura... Retirado de Fauvel (2007). . . . .	25
3.3	Exemplos de Elemento Estruturante (EE) de diferentes formas. . . . .	25
3.4	Operações típicas de cruzamento. Retirado de Weise (2008). . . . .	30
3.5	Tipos da mutação. Retirado de Weise (2008). . . . .	30
4.1	Diagrama do processo de combinação de múltiplas abordagens de classificação. Adaptado de Briem et al. (2002). . . . .	31
4.2	Representação do genoma e mapeamento das bandas espectrais, 1 significa presença e 0 ausência de determinada banda no hipercubo. Modificado de Tarabalka (2010). . . . .	34
4.3	Exemplo do <i>Extended Morphological Profiles (EMP)</i> usando duas Componentes Principais (PCs), EE circular de tamanhos 4, 6 e 8 (passo de crescimento do EE igual a 2) para a base de dados <i>Indian Pines</i> . . . . .	37
4.4	Diagrama ilustrativo da fusão da informação espectral e a espacial provida pelo <i>EMP</i> . Modificado de Fauvel et al. (2008). . . . .	38
4.5	Exemplo de uma representação do tipo <i>bit string</i> . Dez bits (isto é, genes) são usados para codificar cada peso, possibilitando $2^{10} = 1024$ diferentes valores de peso. . . . .	40

4.6	Diagrama do sistema de múltiplas abordagens de classificação proposto neste trabalho. Quatro tipo de representação de características e três tipos de algoritmos de aprendizado são usados para garantir maior diversidade.	41
4.7	Diagrama ilustrativo da abordagem proposta <i>Stacking-SVM</i> . Novas características são produzidas pela concatenação das saídas obtidas pelas abordagens de classificação base (meta-características) e então submetidas a um novo processo de aprendizado usando o classificador <i>SVM</i> .	44
5.1	<i>Airborne Visible Infra-Red Imaging Spectrometer (AVIRIS) - Indian Pines.</i>	46
5.2	<i>Reflective Optics System Imaging Spectrometer (ROSIS) - Pavia University.</i>	48
5.3	Exemplo esquemático do protocolo de treinamento e teste utilizado.	49
6.1	Comparação entre os mapas temáticos do <i>Indian Pines</i> produzidos pela classificação da informação espectral e pela combinação das quatro abordagens de classificação base (cenário 1) usando o <i>WLC-GA</i> e <i>Stacking-SVM</i> .	57
6.2	Resultados para o <i>Indian Pines</i> em diversas configurações de amostragens. O símbolo asterisco (*) indica a média do Overall Accuracy (OA) das 30 execuções e as barras verticais correspondem ao respectivo Intervalo de Confiança (CI).	61
6.3	Comparação entre os mapas temáticos do <i>Pavia University</i> produzidos pela classificação da informação espectral e pela combinação das quatro abordagens de classificação base (cenário 1) usando o <i>WLC-GA</i> e <i>Stacking-SVM</i> .	63
6.4	Resultados para o <i>Pavia University</i> em diversas configurações de amostragens. O símbolo asterisco (*) indica a média do OA das 30 execuções e as barras verticais correspondem ao respectivo CI.	64
6.5	Mapa de probabilidades, referente à classe <i>Grass/Pasture</i> do <i>Indian Pines</i> , gerado por cada uma das abordagens de classificação base.	68
6.6	Mapa de probabilidades, referente à classe <i>Grass/Pasture</i> do <i>Indian Pines</i> , obtido pela combinação usando o <i>WLC-GA</i> nos três cenários experimentados.	69
6.7	Evolução do <i>WLC-GA</i> ao longo de 100 gerações para o <i>Indian Pines</i> .	70

# Lista de Tabelas

1.1	Exemplos de Sistemas de Imageamento Hiperespectral. Modificado de Chang (2007). . . . .	4
3.1	Efeitos da erosão e dilatação em imagens em tons de cinza. . . . .	26
5.1	Número de amostras por classe para o <i>Indian Pines</i> . . . . .	47
5.2	Número de amostras por classe para o <i>Pavia University</i> . . . . .	47
6.1	Resultados do cenário 1 para o <i>Indian Pines</i> usando 10% de dados para treinamento em 30 execuções. . . . .	59
6.2	Resultados do cenário 2 para o <i>Indian Pines</i> usando 10% de dados para treinamento em 30 execuções. . . . .	59
6.3	Resultados do cenário 3 para o <i>Indian Pines</i> usando 10% de dados para treinamento em 30 execuções. . . . .	60
6.4	Resultados do cenário 1 para o <i>Pavia University</i> usando 10% de dados para treinamento em 30 execuções. . . . .	60
6.5	Resultados do cenário 2 para o <i>Pavia University</i> usando 10% de dados para treinamento em 30 execuções. . . . .	65
6.6	Resultados do cenário 3 para o <i>Pavia University</i> usando 10% de dados para treinamento em 30 execuções. . . . .	65
6.7	Tempo computacional (em segundos) para treinar (10%) e testar (90%) no <i>Indian Pines</i> . . . . .	67
6.8	Tempo computacional (em segundos) para treinar (10%) e testar (90%) no <i>Pavia University</i> . . . . .	67

# Lista de Acrônimos

<b>AA</b>	Average Accuracy
<b>a-MRF</b>	adaptative Markov Random Fields
<b>AIS</b>	Airborne Imaging Spectrometer
<b>AVIRIS</b>	Airborne Visible Infra-Red Imaging Spectrometer
<b>CA</b>	Class Accuracy
<b>CI</b>	Intervalo de Confiança
<b>CRF</b>	Conditional Random Fields
<b>DBFE</b>	Decision Boundary Feature Extraction
<b>DFC</b>	Data Fusion Contest
<b>DFTC</b>	Data Fusion Technical Committee
<b>DP</b>	Decision Profile
<b>EE</b>	Elemento Estruturante
<b>EM</b>	Expectation Maximization
<b>EMP</b>	Extended Morphological Profiles
<b>FEGA</b>	Feature Extraction by Genetic Algorithms
<b>GA</b>	Algoritmos Genéticos
<b>GDA</b>	Generalized Discriminant Analysis
<b>GRSS</b>	Geoscience and Remote Sensing Society
<b>HIAT</b>	Hyperspectral Image Analysis Toolbox

<b>HSEG</b>	Hierarchical Segmentation
<b>ICA</b>	Independent Component Analysis
<b>KFDA</b>	Kernel Fisher Discriminant Analysis
<b>KNN</b>	K-Nearest Neighbor
<b>KNWFE</b>	Kernel Nonparametric Weighted Feature Extraction
<b>KPCA</b>	Kernel Principal Component Analysis
<b>LIBSVM</b>	A Library for Support Vector Machines
<b>LDA</b>	Linear Discriminant Analysis
<b>LOP</b>	Linear Opinion Pool
<b>MCS</b>	Sistemas de Múltiplos Classificadores
<b>MV</b>	Majority Vote
<b>MRF</b>	Markov Random Fields
<b>ML</b>	Maximum Likelihood
<b>MM</b>	Morfología Matemática
<b>MP</b>	Morphological Profile
<b>MLP</b>	Multilayer Perceptron Neural Network
<b>MSSC</b>	Multiple Spectral-Spatial Classifier
<b>MSF</b>	Minimum Spanning Forest
<b>NLPCA</b>	Nonlinear Principal Component Analysis
<b>NN</b>	Neural Networks
<b>NIPALS</b>	Nonlinear Iterative Partial Least Squares
<b>NWFE</b>	Nonparametric Weighted Feature Extraction
<b>OA</b>	Overall Accuracy
<b>PC</b>	Componente Principal

<b>PCs</b>	Componentes Principais
<b>PCA</b>	Principal Component Analysis
<b>PLS</b>	Partial Least Squares
<b>PR</b>	Pós-Regularização
<b>RBF</b>	Radial Basis Function
<b>RBFNN</b>	Radial Basis Function Neural Networks
<b>Reg-AB</b>	Regularized AdaBoost
<b>Reg-RBFNN</b>	Regularized Radial Basis Function Neural Networks
<b>RGB</b>	Red-Green-Blue
<b>ROSI</b>	Reflective Optics System Imaging Spectrometer
<b>RFS</b>	Random Feature Selection
<b>RN</b>	Rede Neuronal
<b>RNA</b>	Rede Neuronal Artificial
<b>SIH</b>	Sistemas de Imageamento Hiperespectral
<b>SMC</b>	Statistical Multisource Classifier
<b>SR</b>	Sensoriamento Remoto
<b>SVM</b>	Support Vector Machines
<b>SVDSS</b>	Singular Value Decomposition Band Subset Selection
<b>WLC-GA</b>	Weighted Linear Combination optimized by Genetic Algorithm

# Sumário

Agradecimentos	vii
Resumo	ix
Abstract	xi
Lista de Figuras	xiii
Lista de Tabelas	xv
Lista de Acrônimos	xvi
<b>1 Introdução</b>	<b>3</b>
1.1 Motivação . . . . .	6
1.2 Objetivos . . . . .	7
1.3 Contribuições . . . . .	7
1.4 Organização do Documento . . . . .	8
<b>2 Revisão</b>	<b>9</b>
2.1 Abordagens de Classificação Espectrais . . . . .	9
2.2 Abordagens de Classificação Espectrais-Espaciais . . . . .	12
2.2.1 Algoritmos de segmentação para união da informação espacial . . . . .	12
2.2.2 Incorporando informação espacial no vetor de características . . . . .	15
2.3 Combinação de Múltiplas Abordagens de Classificação . . . . .	17
<b>3 Fundamentação Teórica</b>	<b>20</b>
3.1 Algoritmos de Aprendizagem de Máquina . . . . .	20
3.1.1 <i>K-Nearest Neighbor</i> . . . . .	20
3.1.2 <i>Multilayer Perceptron Neural Network</i> . . . . .	21
3.1.3 <i>Support Vector Machines</i> . . . . .	22

3.2	Métodos de Combinação de Classificadores . . . . .	22
3.2.1	Combinação Consensual . . . . .	23
3.3	Morfologia Matemática . . . . .	25
3.4	Algoritmos Genéticos . . . . .	27
3.4.1	Procedimento Geral . . . . .	27
3.4.2	Genoma . . . . .	28
3.4.3	Seleção . . . . .	29
3.4.4	Cruzamento . . . . .	29
3.4.5	Mutação . . . . .	30
<b>4</b>	<b>Metodologia</b>	<b>31</b>
4.1	Abordagens de Classificação Base . . . . .	32
4.1.1	<i>Pixelwise-SVM</i> . . . . .	32
4.1.2	<i>FEGA-KNN</i> . . . . .	33
4.1.3	<i>EMP-MLP</i> . . . . .	35
4.1.4	<i>SpecEmpPLS-SVM</i> . . . . .	37
4.2	Combinação de Múltiplas Abordagens de Classificação . . . . .	38
4.2.1	<i>WLC-GA</i> . . . . .	38
4.2.2	<i>Stacking-SVM</i> . . . . .	42
<b>5</b>	<b>Experimentos</b>	<b>45</b>
5.1	Bases de Dados Hiperespectrais . . . . .	46
5.2	Critérios de Avaliação . . . . .	48
5.2.1	Protocolo de Treinamento e Teste . . . . .	48
5.2.2	Medidas de Avaliação . . . . .	50
5.3	Configuração dos Parâmetros . . . . .	51
5.3.1	Parâmetros das Abordagens de Classificação Base . . . . .	51
5.3.2	Parâmetros dos Métodos de Combinação . . . . .	54
5.4	Cenários . . . . .	55
<b>6</b>	<b>Resultados e Discussões</b>	<b>56</b>
6.1	Análise dos Resultados . . . . .	56
6.1.1	Resultados para o <i>Indian Pines</i> . . . . .	56
6.1.2	Resultados para o <i>Pavia University</i> . . . . .	61
6.2	Análise do Tempo Computacional . . . . .	66
6.3	Discussão . . . . .	67
<b>7</b>	<b>Conclusões e Trabalhos Futuros</b>	<b>71</b>



# Publicações

A seguir é apresentado a lista de trabalhos publicados durante o período de estudo do mestrado.

## Artigos em Periódicos

- 1. Santos, A. B.; A. Araújo A. de A.; Menotti D. (2013). Combining Multiple Classification Methods for Hyperspectral Data Interpretation. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (JSTARS), vol. 6, n.3, pp. –.

## Artigos em Conferências

- 1. Santos, A. B.; A. Araújo A. de A.; Menotti D. (2012a). Combiner of Classifiers Using Genetic Algorithm for Classification of Remote Sensed Hyperspectral Images. IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 4146 -4149.
- 2. Santos, A. B.; CELES, C. S. F. S.; A. Araújo A. de A.; Menotti D. (2012c). Feature Selection for Classification of Remote Sensed Hyperspectral Images: A Filter approach using Genetic Algorithm and Cluster Validity. The 2012 International Conference on Image Processing, Computer Vision, and Pattern Recognition (ICCV'12), volume 2, pp. 675–681.

## Artigos em Workshops

- 1. Santos, A. B.; A. Araújo A. de A.; Menotti D. (2012b). Combining Multiple Approaches for Accuracy Improvement in Remote Sensed Hyperspectral Images

Classification. Workshop of Theses and Dissertations (WTD) in SIBGRAPI 2012 (XXV Conference on Graphics, Patterns and Images). pp. 54–59. <sup>1</sup>

---

<sup>1</sup>Trabalho premiado com a 2ª colocação na categoria M.Sc. no WTD-SIBGRAPI 2012: <http://www.decom.ufop.br/sibgrapi2012/index.php/call/wtd>

# Capítulo 1

## Introdução

Sensoriamento Remoto (SR) é uma área de estudo que visa coletar dados e informações de objetos através da medição de sinais (tal como a radiação solar refletida) à longas distâncias (Prasad et al., 2011). Em SR a classificação da cobertura terrestre se preocupa em identificar as diferentes áreas sobre a superfície da terra e tem sido usada para vários propósitos, tais como: planejamento e gerenciamento urbano, monitoramento ambiental, agricultura, silvicultura, *etc.* (Prasad et al., 2011; Chang, 2007). O imageamento espectroscópico, ou como é mais conhecido hoje em dia por imageamento *hiperespectral*, tem permitido muitos avanços na análise e interpretação de materiais sobre a superfície terrestre (Plaza et al., 2009). Nos anos 1980, dentro do laboratório *Jet Propulsion Laboratory* da NASA, pesquisadores começaram a desenvolver novos instrumentos. Um importante para a área de sensoriamento remoto foi o *Airborne Imaging Spectrometer (AIS)*. Posteriormente chamado de *Airborne Visible Infra-Red Imaging Spectrometer (AVIRIS)*, esse sensor cobre uma faixa de espectro eletromagnético que vai de  $0,4\mu m$  à  $2,5\mu m$  em uma resolução espectral de  $10nm$  contendo mais de duzentos canais espectrais (Green et al., 1998).

Imagens hiperespectrais são compostas de algumas dezenas até centenas de bandas espectrais (Tarabalka, 2010; Plaza et al., 2009). Cada banda espectral, ou canal, representa a reflectância de um material em um determinado comprimento de onda. Ao construir a resposta espectral para cada pixel, em cada posição, temos então a imagem hiperespectral. A Figura 1.1 ilustra a estrutura dos dados hiperespectrais, ou como também é conhecida por hipercubo (Chang, 2007). Como pode ser observado, cada pixel, ou amostra, possui várias assinaturas espectrais. A Tabela 1.1 mostra alguns exemplos de sistemas de imageamento hiperespectral para fins de pesquisa e desenvolvimento, comercial e militar, em que sensores de diferentes tipos de resoluções espectrais, espaciais e números de bandas são usualmente empregados.

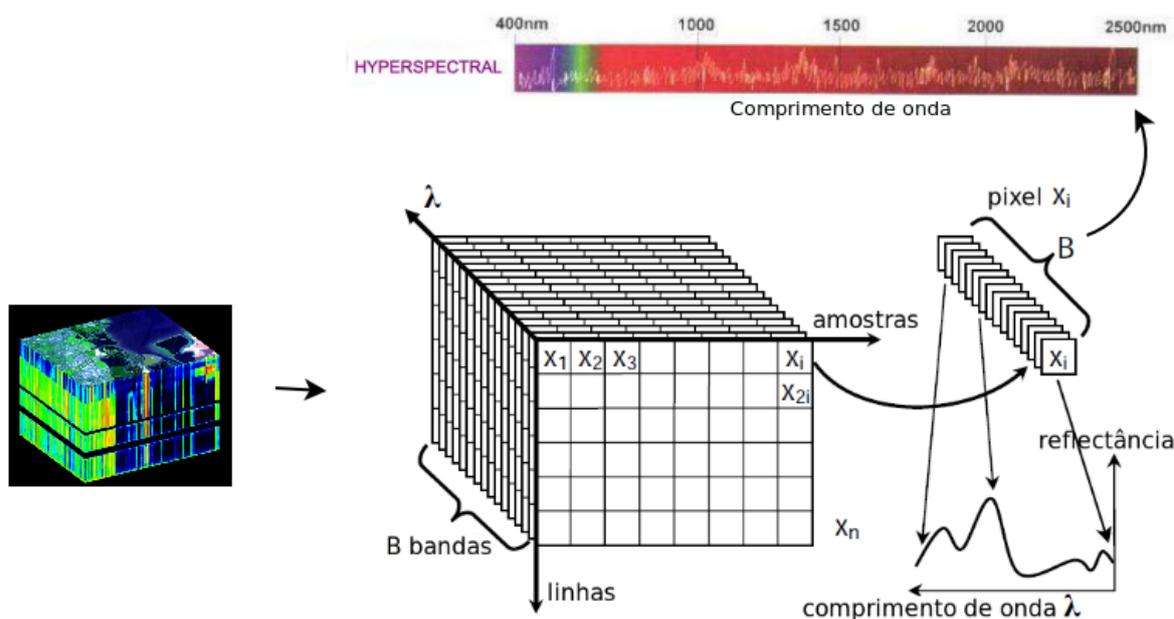


Figura 1.1: Esquema ilustrativo de uma imagem hiperspectral (hipercubo). Modificado de Tarabalka (2010).

Sensor	Objetivo	Altitude típica (km)	Faixa Espectral ( $\mu m$ )	Número de bandas	Resolução espacial (metros/pixel)
AIS (Goetz et al., 1985)	Ciência	4	1,2 a 2,4	128	8
ARCHER (Stevenson et al., 2005)	Patrulha Aérea Civil	2	0,5 a 1,1	512	5
AVIRIS (Green et al., 1998)	Ciência	20	0,4 a 2,5	224	20
DAIS (Mueller et al., 2002)	Ciência	-	0,4 a 12,3	79	5
HYDICE (Rickard et al., 1993)	Militar e ciência	6	0,4 a 2,5	210	3
HyMAP (Cocks et al., 1998)	Operações Comerciais	2	0,45 a 2,5	126	5
Hyperion (Pearlman et al., 2000)	Demonstração Espacial	705	0,4 a 2,5	200	30
ROSIS (Gege et al., 1998)	Ciência	10	0,43 a 0,86	115	1,3

Tabela 1.1: Exemplos de Sistemas de Imageamento Hiperspectral. Modificado de Chang (2007).

Essa tecnologia permite identificar materiais que não podem ser identificados por análises tradicionais de baixa resolução espectral, como por exemplo a análise Multi-espectral (Plaza et al., 2009; Chang, 2007). As imagens multispectrais, tipicamente, possuem dados adquiridos em quatro a sete canais espectrais. Portanto, podemos perceber que sistemas hiperspectrais são uma evolução natural dos multispectrais (Chang, 2007). Geralmente, as imagens são adquiridas pelos sensores em transporte aéreo ou

por satélites à pequena, média ou longa distância (Plaza et al., 2009). A análise e exploração desse tipo de dados permite grande avanço para a comunidade de sensoriamento remoto, principalmente para a interpretação (classificação) dos diversos materiais terrestres (Chang, 2007; Prasad et al., 2011). Usualmente, um mapa da área interpretada é obtido após o processo de classificação. Denominado mapa temático, essa nova imagem obtida é construída após rotular cada um dos *pixels* de acordo com as diversas classes de interesse. Assim, cada classe recebe uma cor específica que possibilite uma melhor visualização do mapa temático. Em termos de informação espacial a maioria dos sistemas possuem resoluções entre 1 e 30 metros por *pixel* (Chang, 2007). Isso representa uma grande capacidade de distinguir estruturas de objetos complexos, o que obviamente pode aumentar a acurácia na classificação. Por consequência, diversos autores visam a junção da informação espectral mais a informação espacial (espectral-espacial) na tentativa de elevar os índices de acertos na classificação (Chang, 2007; Plaza et al., 2009; Prasad et al., 2011).

Na tentativa de incluir tanto a informação espacial quanto a espectral na interpretação de imagens hiperespectrais de sensoriamento remoto, propomos a combinação de quatro abordagens de classificação. Uma maneira de se obter classificadores complementares é usando representações diferentes de uma mesma entrada (Alpaydin, 1998). Com isso em mente, propomos a combinação de abordagens de classificação que se diferenciam tanto na representação das características dos dados hiperespectrais quanto nos algoritmos de aprendizado utilizados na interpretação desses dados. Utilizamos duas abordagens que baseiam na classificação usando apenas a informação espectral, e as outras duas baseadas na informação espectral-espacial. Nesse sentido, duas abordagens de classificação (*Pixelwise-SVM* (Melgani & Bruzzone, 2004) e *EMP-MLP* (Benediktsson et al., 2005)) amplamente discutidas na literatura foram usadas. A abordagem *Pixelwise-SVM* utiliza apenas informação espectral, enquanto que a *EMP-MLP* se beneficia da informação espacial extraída através de operações morfológicas. Além disso, propomos outras duas abordagens de classificação (*FEGA-KNN* (Santos et al., 2012c) e *SpecEmpPLS-SVM*) para compor as quatro abordagens de classificação base utilizadas no processo de combinação. A abordagem *FEGA-KNN* utiliza Algoritmos Genéticos (GA) e uma medida de validação de agrupamentos chamada Índice de *Dunn* para selecionar bandas espectrais mais adequadas para o processo de classificação. A partir disso, a informação espectral gerada por esse método é caracterizada por meio do algoritmo de aprendizado *K-Nearest Neighbor (KNN)*. Na abordagem *SpecEmpPLS-SVM* utilizamos o algoritmo *Partial Least Squares (PLS)* para fundir a informação espectral e espacial de maneira mais compacta em um mesmo vetor de característica. Essa nova representação torna os dados mais adequados para uma classificação usando

o algoritmo de aprendizado *Support Vector Machines (SVM)* com *kernel Radial Basis Function (RBF)*.

Para a combinação dessas quatro abordagens de classificação, propomos dois métodos. Nossa primeira proposta se baseia em um método consensual usado para combinação de classificadores, em que uma combinação linear ponderada das probabilidades de saída de cada abordagem de classificação é realizada. Denominada *Weighted Linear Combination optimized by Genetic Algorithm (WLC-GA)* (Santos et al., 2012a,b, 2013), nossa proposta utiliza GA para encontrar um conjunto de pesos ótimos, ou sub-ótimos, que pondere cada uma das abordagens de classificação usadas na combinação. O segundo método proposto (*Stacking-SVM*) utiliza o paradigma de meta-aprendizado conhecido como *Stacked Generalization* (Wolpert, 1992) para combinar múltiplas abordagens de classificação. Devido à alta dimensionalidade imposta ao juntar as saídas de cada uma das abordagens base, e também da atenção recebida pela comunidade de SR, o algoritmo de aprendizado *SVM* foi escolhido para atuar como “função de combinação” (meta-classificador).

Experimentos foram realizados nas bases de dados *Indian Pines* e *Pavia University* capturados pelos sensores *AVIRIS* (Green et al., 1998) e *Reflective Optics System Imaging Spectrometer (ROSIS)* (Gege et al., 1998), respectivamente. Comparações com métodos tradicionais de combinação (*Majority Vote (MV)* e *Average rule*) também foram feitas.

## 1.1 Motivação

Apesar dos recentes avanços na área de interpretação de imagens hiperespectrais de sensoriamento, essa tarefa ainda se constitui um desafio visto que a utilização de técnicas tradicionais de análise de imagens multiespectrais nem sempre são adequadas devido ao elevado número de atributos (canais) presentes nos dados hiperespectrais. Por esse motivo, técnicas adequadas para imagens hiperespectrais devem ser exploradas, como por exemplo a fusão das informações espectrais e espaciais, que ainda é um problema aberto na comunidade de sensoriamento remoto. Essa estratégia tem demonstrado resultados promissores na classificação de dados hiperespectrais (Fauvel, 2007; Tarabalka, 2010; Licciardi et al., 2009; Tarabalka et al., 2010a; Prasad et al., 2011; Santos et al., 2012a,b, 2013).

Desde 2006, um evento chamado *Data Fusion Contest (DFC)*, organizado pelo comitê *Data Fusion Technical Committee (DFTC)* através da *IEEE GRSS*, tem como principal objetivo fornecer dados de referência para toda a comunidade de sensori-

amento remoto. No *DFC* de 2008 o tema abordado foi a classificação de imagens hiperespectrais de alta resolução. O objetivo não era apenas identificar a melhor abordagem de classificação, mas também, a fusão das cinco melhores abordagens afim de se conseguir mapas temáticos mais precisos quanto possíveis (Licciardi et al., 2009). O método usado na combinação foi o simples, mas democrático, *MV*. Inspirado por essa ideia, este trabalho visa realizar uma combinação mais elaborada entre diversas abordagens de classificação. Usando a combinação de múltiplas abordagens de classificação podemos acrescentar diferentes tipos de informações no processo de interpretação, portanto, é possível realizar a fusão da informação espacial e espectral de maneira indireta, e com isso obter índices de acurácia mais elevados.

## 1.2 Objetivos

Dadas as motivações acima, os principais objetivos deste trabalho são:

1. Selecionar canais, bandas, mais informativos para o processo de classificação baseada apenas na informação espectral.
2. Unir a informação espectral e espacial em um único vetor de característica.
3. Utilizar as informações espectrais e espaciais na classificação.
4. Combinar múltiplas abordagens de classificação.
5. Comparar os métodos desenvolvidos com métodos tradicionais existentes.
6. Produzir mapas temáticos mais precisos.

## 1.3 Contribuições

Ao alcançar os objetivos propostos, as principais contribuições desse trabalho de dissertação foram:

1. Algoritmo para seleção de bandas espectrais (seleção de características) baseado em GA e métricas de validação de agrupamentos.
2. Nova abordagem de classificação que se beneficia de ambas informações espectrais e espaciais (*SpecEmpPLS-SVM*) usando o algoritmo *PLS* para a modelagem da informação espectral-espacial.

3. Utilização de GA para encontrar conjuntos de pesos adequados para uma combinação ponderada de diversas abordagens de classificação aplicadas à imagens hiperespectrais de sensoriamento remoto.
4. Utilização de técnicas de meta-aprendizado no domínio das imagens hiperespectrais de sensoriamento remoto.

## 1.4 Organização do Documento

O restante desta dissertação está organizada como a seguir. O Capítulo 2 apresenta uma revisão bibliográfica sobre os principais métodos de classificação de imagens hiperespectrais de sensoriamento remoto. O Capítulo 3 fornece os fundamentos básicos para a compreensão dos métodos e algoritmos utilizados. O Capítulo 4 descreve as múltiplas abordagens de classificação base usadas na combinação, bem como os métodos de combinação utilizados neste trabalho. Posteriormente os experimentos são detalhados no Capítulo 5. No Capítulo 6 fazemos uma análise e sobre os resultados obtidos. Finalmente, uma conclusão e discussão sobre trabalhos futuros são apresentados no Capítulo 7.

# Capítulo 2

## Revisão

Nesta revisão, trataremos de trabalhos que usam a informação espectral, isto é, utilizam apenas valores de cada *pixel* para realizar a interpretação. Trabalhos que utilizam informação espectral-espacial, ou seja, a informação espacial entre *pixels* adjacentes também é levada em consideração no processo de classificação. E, finalmente, apresentamos trabalhos que utilizam sistemas de múltiplos classificadores para interpretação de imagens hiperespectrais de sensoriamento remoto.

### 2.1 Abordagens de Classificação Espectrais

De maneira geral, a classificação de imagens hiperespectrais de sensoriamento remoto consiste na atribuição de um rótulo específico para cada um dos *pixels* individuais da imagem. Abordagens que se baseiam apenas na informação espectral presente na imagem são conhecidas como *Pixelwise*. Neste tipo de abordagem cada *pixel* é, então, considerado uma amostra e seu espectro o conjunto de características dessa amostra (Tarabalka, 2010). A Figura 1.1 fornece uma ilustração desse tipo de representação. Diante da alta dimensionalidade que as imagens hiperespectrais apresentam, algoritmos de aprendizado, ou métodos de classificação, capazes de se adequarem a essa situação se fazem necessários (Plaza et al., 2009).

Ao lidar com os conhecidos problemas no domínio das imagens hiperespectrais Melgani & Bruzzone (2004); Camps-Valls & Bruzzone (2005) concluíram que algoritmos de aprendizado baseados em *kernel*, os chamados *kernel-trick methods*, são mais eficazes que os tradicionais métodos de aprendizado. Métodos baseados em *kernel* possuem a habilidade de transformar os dados originais para um novo espaço de características onde se consegue uma melhor separação entre as classes. Melgani & Bruzzone (2004), usando o algoritmo de aprendizado *SVM* aplicado as imagens hiperespectrais

revela que esse algoritmo é uma efetiva e poderosa ferramenta para a interpretação desses dados, superando tradicionais métodos como: *Radial Basis Function Neural Networks (RBFNN)* e *KNN*. Em um estudo ainda mais avançado (Camps-Valls & Bruzzone, 2005), vários algoritmos baseados em *kernel* foram explorados. Nesse estudo (Camps-Valls & Bruzzone, 2005) os algoritmos de classificação *Regularized Radial Basis Function Neural Networks (Reg-RBFNN)*, *SVM*, *Kernel Fisher Discriminant Analysis (KFDA)*, e *Regularized AdaBoost (Reg-AB)* foram analisados e comparados quanto a questão da interpretação dos dados hiperespectrais, revelando assim a superioridade do algoritmo *SVM*. Desde então o *SVM* tem sido apontado por diversos autores pelo seu bom comportamento na presença de alta dimensionalidade e poucas amostras de treino. Ainda, sua eficácia é considerada como referência em vários trabalhos da literatura (Tarabalka, 2010).

De uma maneira geral abordagens que se baseiam em aprendizado supervisionado possuem maior desempenho, em termos de acurácia, que abordagens não-supervisionadas (Alpaydin, 2010; Mitchell, 1997). No entanto, como sabemos, rotular grandes quantidades de dados é um trabalho árduo, não somente no domínio de imagens hiperespectrais como também em outras áreas, o que torna a escassez de dados com referencia um problema eminente. Por isso, técnicas de aprendizado ativo combinadas ao aprendizado semi-supervisionado (Li et al., 2010) estão sendo cada vez mais utilizadas, para interpretação desse tipo de dados, com o objetivo de minimizar o problema da escassez de dados rotulados. Por meio do aprendizado ativo, amostras que maximizam o ganho de informação na etapa de aprendizado são identificadas permitindo melhor aprendizado mesmo utilizando conjuntos de treinamento pequenos (Li et al., 2010, 2011). Ainda, no trabalho de Tuia & Camps-Valls (2011), a utilização de *textitkernels* e aprendizado semi-supervisionado demonstra uma ferramenta promissora para combinar informações de amostras rotuladas e não rotuladas, permitindo maior robustez na classificação.

Como já foi discutido anteriormente, a baixa quantidade de dados referenciados (rotulados) nas bases de dados hiperespectrais torna o problema de classificação um desafio. Dado que quanto maior a dimensionalidade, isto é, conjunto de características presentes nos dados, maior será o número de amostras de treinamento necessárias para que se haja um aprendizado desejável (Alpaydin, 2010; Mitchell, 1997). Diversos autores têm utilizado métodos de seleção e extração de características com o objetivo de manter, ou até mesmo melhorar, o poder de discriminação dos dados sem prejudicar o desempenho dos classificadores. Isso é feito de forma que uma imagem com  $B$  bandas (canais) seja representada em nova imagem contendo  $D$  bandas, em que  $D < B$ , ao mesmo tempo manter, ou até aumentar, a discriminatividade

entre as diversas classes existentes na imagem (base de dados). Recentemente, Kuo et al. (2009) comparam os principais métodos de extração de características discutidos na literatura utilizando diferentes tipos de classificadores. O método apresentado por Kuo et al. (2009) utiliza-se de propriedades dos métodos *kernel-based* para estender o tradicional *Nonparametric Weighted Feature Extraction (NWFE)* dado que este se aproveita apenas de uma transformação linear, propondo assim o *Kernel Nonparametric Weighted Feature Extraction (KNWFE)*. Em um extenso trabalho Kuo et al. (2009) compararam o *KNWFE* com diversos métodos lineares e não-lineares de extração de características presentes na literatura técnica, são eles: *Principal Component Analysis (PCA)*, *kernel-based PCA (Kernel Principal Component Analysis (KPCA))*, *Independent Component Analysis (ICA)*, *Linear Discriminant Analysis (LDA)*, *kernel-based LDA* ou *Generalized Discriminant Analysis (GDA)*, *Decision Boundary Feature Extraction (DBFE)* e o tradicional *NWFE*. Para a comparação foram utilizados os algoritmos de aprendizado: *SVM*, *Maximum Likelihood (ML)* e *KNN* em diversas configurações de amostragem de treinamento. Kuo et al. (2009) concluíram que a abordagem proposta *KNWFE* possui o melhor desempenho, em termos de acurácia, em relação aos demais métodos. No entanto, não houve um classificador em comum que obtivesse o melhor desempenho em todas as condições de testes realizados. Para os autores, encontrar a melhor combinação, extração de características + classificador, que possua a melhor acurácia em todas as bases de dados, é uma tarefa muito difícil. Embora esse problema seja considerado de difícil tratamento, existem trabalhos que buscam encontrar a melhor cadeia de processamento para se obter uma classificação mais exata possível de acordo com o tipo de base de dados utilizada (Bakos & Gamba, 2011). Nos trabalhos de Bakos & Gamba (2011) o interesse é combinar os algoritmos mais eficientes destacados na literatura em um procedimento geral que se adapte a diferentes tipos de imagens hiperespectrais, seja para classificação de áreas urbanas ou com vegetação, minerais, etc.

Outra maneira muito comum de reduzir a dimensionalidade nos dados é conhecida como seleção de características (Bishop, 2006). Diferentemente da extração, a seleção não introduz nenhum tipo de transformação nos dados originais. O processo de seleção de características é dividido em duas categorias: *filter* e *wrapper*. Contudo, as duas categorias possuem o mesmo objetivo: procurar um subconjunto menor de características que permita maior separabilidade inter-classe e conseqüentemente índices mais elevados de acurácia. Abordagens do tipo *wrapper*, como nos trabalhos de Bazi & Melgani (2006); Pal & Foody (2010); Pedernana et al. (2012), conduzem uma busca utilizando informação proveniente de um classificador (Bishop, 2006). Abordagens *filter* (Velez-Reyes & Jimenez, 1998; Velez-Reyes et al., 2000; Yu & Liu, 2003; Santos et al., 2012c)

são completamente independentes do processo de classificação. Geralmente esses métodos tentam otimizar algum tipo de função que mensura a separabilidade entre as classes, sendo assim mais rápidos que o processo de aprendizado usado nas abordagens *wrapper*. Embora algoritmos que se baseiam em abordagens *wrapper* sejam, em sua maioria, mais lentos que aqueles baseados em *filter*, estes tendem a produzir subconjuntos mais adequados para o classificador em questão (Bishop, 2006; Kohavi & John, 1997).

## 2.2 Abordagens de Classificação Espectrais-Espaciais

Basicamente há duas estratégias para incorporar informação espectral-espacial, ou informação contextual, para interpretação de imagens hiperespectrais de sensoriamento remoto (Tarabalka et al., 2010d; Zhon & Wang, 2010). A primeira estratégia modela informações contextuais/espaciais nas estruturas intrínsecas dos classificadores (Jackson & Landgrebe, 2002; Sarkar et al., 2002; Neher & Srivastava, 2005; Salzenstein & Collet, 2006; Zhon & Wang, 2010), ou ainda são consideradas após classificação tradicional, regularizando o mapa temático produzido através de um filtro da maioria (Tarabalka et al., 2009a). Dessa forma os *pixels* do mapa temático são novamente rotulados conforme sua vizinhança, em Tarabalka et al. (2009a) esse processo é conhecido como Pós-Regularização (PR). A segunda estratégia incorpora informação contextual nas características extraídas da imagem, ou seja, informação espacial e espectral são combinadas em um só vetor de características de cada *pixel*, e então, algum algoritmo de aprendizado pode ser aplicado ao novo conjunto de características (Benediktsson et al., 2005; Fauvel et al., 2008; Plaza et al., 2009; Fauvel et al., 2009; Shen & Lia, 2011; Fauvel et al., 2012).

### 2.2.1 Algoritmos de segmentação para união da informação espacial

Objetivando a interpretação das imagens de sensoriamento remoto em diversas categorias, ou classes, como água, vegetação, área urbana, tipos de cultivos, etc., algoritmos de segmentação de imagens se mostram adequadas ferramentas para esse tipo de análise. Em Tilton (1998) um método baseado em crescimento de regiões é proposto para a segmentação de imagens multiespectrais. Denominado *Hierarchical Segmentation (HSEG)*, uma segmentação hierárquica é realizada iterativamente rotulando *pixels* de acordo com

a região mais semelhante. Um critério de dissimilaridade é calculado a todos os pares de regiões espacialmente adjacentes. O menor valor de dissimilaridade é então usado como limiar, de maneira que regiões espacialmente adjacentes são agrupadas quando a dissimilaridade entre elas forem menor ou igual ao limiar. Dessa maneira é criada uma sequência hierárquica que permite flexibilidade na escolha de um apropriado nível de detalhe da segmentação. Quando dados de treino estão disponíveis, a escolha do melhor nível hierárquico é um simples processo baseado na avaliação da segmentação de um determinado nível hierárquico de acordo com os acertos e erros nesse conjunto de treino fornecido (Tarabalka et al., 2010a). Devido ao sucesso desse algoritmo nas imagens multiespectrais, o método *HSEG* foi estendido ao domínio das imagens hiperespectrais (Tilton et al., 2006) fornecendo resultados promissores ao usar um nível hierárquico apropriado. Ainda, métodos automáticos para seleção do melhor nível hierárquico têm sido propostos, tais como em Plaza & Tilton (2005) e Tarabalka et al. (2011). Embora o *HSEG* forneça informações espaciais que possam ser usadas no processo de classificação, outros métodos de segmentação podem ser investigados para realizar essa tarefa.

*Watershed* é uma poderosa técnica da morfologia matemática para segmentação de imagens (Facon, 1996; Serra, 1982). Uma extensão do algoritmo *Watershed* ao domínio das imagens hiperespectrais foi primeiramente proposta por Tarabalka et al. (2010c), onde regiões segmentadas da imagem Hiperespectral são combinadas usando *MV* (Figura 2.1), com uma classificação *Pixelwise* usando *SVM*. Usando essa abordagem, mapas temáticos com regiões mais homogêneas são obtidos (Tarabalka et al., 2008, 2010c). Utilizando esse mesmo esquema, outros algoritmos não-supervisionados com o objetivo de dividir a imagem em regiões como por exemplo os algoritmos de clusterização *ISODATA* e *Expectation Maximization (EM)*, podem ser facilmente aplicados (Tarabalka et al., 2009a). A aplicação do algoritmo *Watershed* de maneira totalmente não-supervisionado pode apresentar alguns problemas como o *oversegmentation* (Tarabalka et al., 2010c). Para contornar esse problema Tarabalka et al. (2009b) propuseram a adição de uma etapa no processo anterior. Dessa vez, a segmentação é controlada por um conjunto de *pixels* denominado sementes (Tarabalka et al., 2009b). Para fazer a seleção dos melhores candidatos à sementes, Tarabalka et al. (2009b) propuseram a utilização de *SVM* com saída probabilística (Chang & Lin, 2011). Assim, uma estimativa da probabilidade de pertencer a cada uma das classes é atribuída para cada *pixel* classificado nesta primeira etapa. Dessa forma, o *pixel* que possui probabilidade maior que um dado limiar é escolhido para fazer parte do conjunto de sementes que controlará o *Watershed*. Através dessa abordagem o problema de *oversegmentation* pode ser minimizado e índices de acurácia superiores à tradicional abordagem

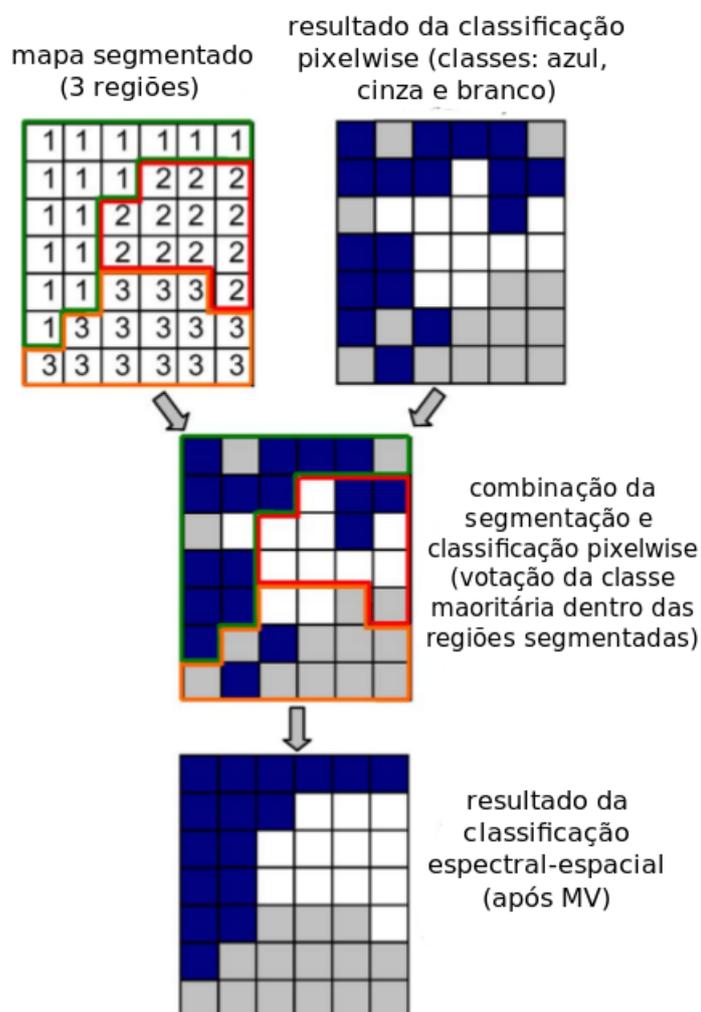


Figura 2.1: Exemplo esquemático da classificação espectral-espaical usando  $MV$  dentro das regiões segmentadas. Ilustração retirada de Tarabalka et al. (2010a).

foram reportados. A desvantagem dessa abordagem é que a escolha das sementes é muito dependente do desempenho na classificação *Pixelwise* (Fauvel et al., 2012), o que pode acontecer de sementes não tão apropriadas sejam selecionadas. Por esse motivo, Tarabalka et al. (2010a) propuseram uma abordagem baseada em Sistemas de Múltiplos Classificadores (MCS) para seleção automática dessas sementes. Essa nova abordagem, *Multiple Spectral-Spatial Classifier (MSSC)* (Tarabalka et al., 2010a), seleciona as sementes que são comuns entre os múltiplos classificadores usando o esquema de  $MV$  ilustrado na Figura 2.1, e assim aumentando a confiabilidade das sementes selecionadas. Por fim o mapa temático é obtido por meio do algoritmo *Minimum Spanning Forest (MSF)* (Tarabalka et al., 2010a).  $MSF$  tem sido bastante aplicado

para segmentação controlada (Tarabalka et al., 2010b; Bernard et al., 2011, 2012), ou supervisionada, de imagens hiperespectrais, demonstrando superioridade em relação à métodos não-supervisionados de segmentação.

Como vimos, algoritmos de segmentação têm sido bastante usados, também, para a regularização de uma classificação do tipo *Pixelwise*, no domínio das imagens hiperespectrais de sensoriamento remoto. Dessa forma, o resultado de uma classificação é combinado com um resultado de um algoritmo de segmentação, por exemplo: *HSEG* (Tilton et al., 2006), *Watershed* (Tarabalka et al., 2010c), *ISODATA* (Tarabalka et al., 2009a), *EM* (Tarabalka et al., 2009a) e *MSF* (Tarabalka et al., 2010b). Geralmente, por se tratarem de métodos não-supervisionados, (exceto *MSF* e versão do *Watershed* que utiliza um conjunto de *pixels* como sementes), a combinação da segmentação com a classificação é feita usando-se o *MV* dentro das regiões segmentadas (Tarabalka et al., 2010c, 2009a, 2010a). Como ilustrado na Figura 2.1, é feita uma contagem da classes previamente classificadas dentro da região segmentada, então, a classe majoritária é atribuída à aquela região. Contudo, esses métodos são fortemente dependentes da classificação *Pixelwise*, uma vez que uma região inteira pode ser classificada erroneamente devido a falhas nessa etapa. Uma alternativa a essa abordagem seria incluir a informação espacial de regiões conectadas no vetor de características como é a ideia básica dos *Extended Morphological Profiles (EMP)*.

## 2.2.2 Incorporando informação espacial no vetor de características

A aplicação direta de métodos de extração de características como *PCA*, *ICA* e *LDA* possui a limitação de tratar somente da informação espectral. Ferramentas matemáticas capazes de extrair correlações de sinais no espaço, no espectro, e unir estes domínios são desejáveis. Na análise de sinais, a *wavelet* de *Gabor* é uma ferramenta capaz de conseguir ótima união de tempo-frequência e espaço-frequência. Shen & Lia (2011) propuseram uma abordagem baseada na 3-D *Gabor-wavelet* para a classificação baseada em *pixel*, ou *pixelwise*, que permita unir a informação espectral e espacial. A ideia central é que cada *pixel* pode ser bem representado por uma rica informação extraída pela *wavelet* de *Gabor*. A magnitude da resposta em cada amostra  $(x, y)$  para uma determinada banda  $b$  contém informação detalhada sobre a variância do sinal naquela região (Shen & Lia, 2011). Como amostras do mesmo material (pertencentes a mesma classe) geralmente possuem variância similares, essa informação deve ser aproveitada na representação de características. Complexos conjuntos da *Gabor-wavelet* com diferentes frequências e orientações foram projetados para extrair variância do sinal hi-

perespectral nos domínios espacial e espectral, combinando assim estes dois (Shen & Lia, 2011). Elevados índices de acurácia foram reportados mesmo em condições de poucas amostras de treinamento (Shen & Lia, 2011).

Ainda neste segundo grupo, *Markov Random Fields (MRF)* é uma clássica abordagem probabilística para modelar informação contextual/espacial (Plaza et al., 2009). Diante da extensiva literatura sobre *MRF* para realização dessa tarefa (Jackson & Landgrebe, 2002; Sarkar et al., 2002; Neher & Srivastava, 2005; Salzenstein & Collet, 2006), Tarabalka et al. (2010d) propuseram a integração de *SVM* e *MRF* (*SVM-MRF*), onde em um primeiro estágio uma classificação do tipo *Pixelwise* usando *SVM* probabilístico é realizada. Em seguida, a classificação é refinada através de uma regularização usando *MRF*. Ainda, baseado no *SVM-MRF* algumas limitações do método proposto por Tarabalka et al. (2010d) foram tratadas de maneira adaptativa por Zhang et al. (2011) (*adaptive Markov Random Fields (a-MRF)*), elevando os índices de acurácia em relação ao *SVM-MRF*. Estudos ainda mais complexos desenvolvidos por Zhong & Wang (2010) usando *Conditional Random Fields (CRF)* para modelar informação espectral-espacial traz sólida alternativa aos tradicionais métodos envolvendo *MRF*.

Uma alternativa simples e eficiente para modelar e codificar informação espectral-espacial muito aplicada, também, em imagens multiespectrais e então utilizada no domínio das imagens hiperespectrais é conhecida como *EMP* (Plaza et al., 2009; Fauvel et al., 2012). *EMP* se baseia em operações morfológicas (Fauvel et al., 2012) de fechamento e abertura por reconstrução (Serra, 1982) para criar então o assim chamado *Morphological Profile (MP)* (Benediktsson et al., 2005). Essencialmente, informação espectral é mantida através da aplicação de algum método de extração de características, o que irá também, diminuir o número total de canais da imagem utilizados para construção de cada *MP* (Benediktsson et al., 2005; Plaza et al., 2009; Fauvel et al., 2012). Cada *MP* resultante de uma operação morfológica, isto é, abertura ou fechamento por reconstrução, realizada utilizando um Elemento Estruturante (EE) de vários tamanhos diferentes são concatenados juntamente com as características extraídas na primeira etapa. Dessa forma, informações espectrais-espaciais são codificadas em um único conjunto de características formando então o *EMP* (Benediktsson et al., 2005). Por fim um método de classificação é usado para aprender e inferir sobre esse novo conjunto de características. Na abordagem tradicional desenvolvida por Benediktsson et al. (2005) *Multilayer Perceptron Neural Network (MLP)* foi usada para essa tarefa. Tradicionalmente o algoritmo de extração de características *PCA* é empregado para redução do número de canais espectrais e codificação da informação espectral no *EMP* (Benediktsson et al., 2005). Em outros estudos o uso de outros algoritmos de extração de características, tais como *ICA* (Palmason et al., 2005; Dalla Mura et al.,

2011) e *KPCA* (Fauvel et al., 2009) têm melhorado significativamente os resultados da classificação. Recentemente, um estudo comparando tradicional *PCA*, *KPCA* e *PCA* não linear (*Nonlinear Principal Component Analysis (NLPCA)*) (Licciardi et al., 2012), obtida através de Redes Neurais auto-associativas, demonstra uma superioridade da *NLPCA* em termos de acurácia na interpretação de imagens hiperespectrais usando a abordagem baseada em *EMP*. Ainda, Fauvel et al. (2008) verificaram algumas redundâncias no vetor de características produzido pelo *EMP* e propuseram a aplicação de algoritmos supervisionados de extração de características, como *DBFE* e *NWFE*, sobre os dados hiperespectrais originais e também sobre o *EMP* tradicional (Benediktsson et al., 2005). Assim, novas características são concatenadas em um novo conjunto mais compacto e discriminativo (Fauvel et al., 2008).

## 2.3 Combinação de Múltiplas Abordagens de Classificação

A combinação de múltiplas abordagens de classificação vem se tornando muito comum em várias áreas do conhecimento. Os termos: *consensus theoretic methods*, *class-consicous combiners*, *meta-learning*, *decision fusion*, *mixture of experts*, *combining multiple learners* e *ensemble of classifiers* são comumente empregados para falar sobre esse assunto. Apesar da variedade de termos utilizados, todos possuem um mesmo objetivo em comum: construir um modelo mais preciso baseado em várias “opiniões” diferentes.

Em sensoriamento remoto a associação, ou fusão, de múltiplos sensores é uma técnica muito pertinente para agregar mais informações sobre uma determinada região ou objeto de interesse ao qual se deseja interpretar (Luo & Kay, 1989; Benediktsson & Swain, 1992; Benediktsson & Kanellopoulos, 1999; Briem et al., 2002; Dalponte et al., 2008; Thoonen et al., 2012).

Benediktsson & Swain (1992) avaliaram métodos de combinação consensual (*consensus theoretic methods*) (Kuncheva, 2004) para classificação de dados de sensoriamento remoto obtidos de múltiplas fontes. Nesse trabalho, Benediktsson & Swain (1992) utilizaram regras consensuais para combinar resultados de classificações usando quatro tipos diferentes de fontes de dados de sensoriamento remoto. Diferentes mecanismos estatísticos foram investigados com a finalidade de encontrar pesos mais adequados para as regras de combinações consensuais *Linear Opinion Pool (LOP)* e *Logarithmic Opinion Pool*. Contudo, a utilização de *LOP* e *Logarithmic Opinion Pool* para a tarefa de classificação desse tipo de dados se mostraram superiores quando comparados à uma

classificação usando uma única fonte de dados. Benediktsson & Swain (1992) ainda concluem que encontrar pesos mais adequados para *LOP* e *Logarithmic Opinion Pool* é uma tarefa muito difícil e deve ser profundamente investigada. Dessa maneira, Benediktsson et al. (1997) utilizaram métodos lineares estatísticos e também não-linear baseado em Rede Neuronal (RN) na otimização do conjunto de pesos para a combinação usando regras consensuais. O método proposto foi investigado quanto à aplicação da classificação de dados de sensoriamento remoto proveniente de múltiplas fontes e tem demonstrado excelentes resultados para essa tarefa (Benediktsson et al., 1997).

Além de métodos consensuais, ou seja, métodos que se baseiam em votação e combinações lineares ponderadas (*LOP*), podemos encontrar na literatura técnica (Kittler et al., 1998; Kuncheva, 2004; Alpaydin, 2010) métodos ainda mais sofisticados como *bagging* (Breiman, 1996), *boosting* (Schapire, 1989), *stacked generalization* (Wolpert, 1992) e *cascading* (Kaynak & Alpaydin, 2000). Alguns autores tem explorado algumas delas para classificação de dados multiespectrais de sensoriamento remoto obtidos por múltiplas fontes (Briem et al., 2002). Os resultados apresentados por Briem et al. (2002) demonstraram que o processo de *boosting* supera os demais métodos estudados para classificação de dados multiespectrais obtidos de múltiplas fontes.

Como foi discutido, existem vários métodos disponíveis na literatura para combinação de múltiplos classificadores (Lam & Suen, 1995; Alpaydin, 1998; Kittler et al., 1998; Kuncheva, 2004; Alpaydin, 2010), e muitos deles foram usados com sucesso para combinação de abordagens de classificação envolvendo múltiplas fontes de dados, ou múltiplos sensores (Benediktsson & Swain, 1992; Benediktsson & Kanellopoulos, 1999; Briem et al., 2002; Dalponte et al., 2008; Thoonen et al., 2012). Era de se esperar que todos esses conceitos fossem, também, introduzidos na área de imagens hiperespectrais, e desde 2006 um importante evento chamado *DFC* tem sido organizado pelo comitê *DFTC* por meio da *IEEE Geoscience and Remote Sensing Society (GRSS)*. O objetivo deste evento é avaliar as metodologias existentes para a resolução de problemas de sensoriamento remoto envolvendo dados obtidos de diferentes sensores. Dessa forma, o principal objetivo da competição é oferecer dados de referência para toda a comunidade de sensoriamento remoto (Licciardi et al., 2009). A *DFC* de 2008 foi dedicada à classificação de dados hiperespectrais de alta resolução, onde o objetivo não era apenas identificar a melhor abordagem proposta, mas também fornecer um esforço colaborativo: a fusão dos melhores algoritmos individuais objetivando a produção de mapas temáticos ainda mais precisos. Nesta última fase, as cinco melhores abordagens foram combinadas usando simples *MV* (Licciardi et al., 2009), produzindo mapas ainda mais confiáveis.

Em Lee & Ersoy (2007) RN consensuais e hierárquicas são usadas para produzir

múltiplos classificadores, de maneira que as RN são ajustadas para reduzir erros remanescentes de sucessivas aplicações. Após a fase de treinamento, as múltiplas RN são combinadas por uma regra consensual, de forma que os pesos de cada RN foram encontrados baseando-se na regra *delta* (minimização do erro quadrático) sobre um conjunto de amostras de treinamento. Os resultados demonstram que a aplicação de RN hierárquicas de maneira consensual produzem melhores resultados quando comparados à um único classificador. Ainda no domínio das imagens hiperespectrais e sistema de múltiplos classificadores, Marpu et al. (2009) avaliaram a medida de distância *Bhattacharya* para seleção das características que melhor identificam determinada classe, e então utilizaram *class-dependent Neural Networks* onde RN são construídas em duas etapas. Inicialmente, RN individuais para cada classe, usando conjuntos de características que caracterizam apenas aquela classe em questão, são treinadas de modo a produzir uma única saída de forma que atue como uma *fuzzification*. As saídas de cada RN servem então como entrada para alimentar outra RN que atuará como um “*defuzzification*” e assim produzir a probabilidade final para a classe (Marpu et al., 2009). Assim como em Marpu et al. (2009), Waske et al. (2010) investigaram como o número de características selecionado e o tamanho do conjunto de classificadores influenciam na acurácia da classificação de dados hiperespectrais. Para isso *Random Feature Selection (RFS)* foi usado, e para cada subconjunto um *SVM* foi aplicado de maneira que cada *SVM* se tornasse especialista naquele subconjunto em questão, e ao final as diversas decisões foram combinadas usando *MV* (Waske et al., 2010).

Fauvel (2007) utilizou a teoria dos conjuntos *fuzzy* (Zadeh, 1956) para a combinação dos dados hiperespectrais. Em todas as abordagens mencionadas houveram ganhos na acurácia dos mapas temáticos produzidos quando comparados a apenas sistemas de classificação que use apenas um classificador.

Contudo, devido aos promissores resultados desses métodos ao combinar dados multiespectrais obtidos por múltiplos sensores (Luo & Kay, 1989; Benediktsson & Swain, 1992; Benediktsson & Kanellopoulos, 1999; Briem et al., 2002), podemos aplicar os mesmos conceitos para combinar, de maneira indireta, as informações espectrais e espaciais presentes nas imagens hiperespectrais. Assim, neste presente trabalho, adaptamos os métodos *LOP* e *Stacked Generalization* para combinar diferentes abordagens de classificação. Cada abordagem de classificação utilizada fornece diferentes representações do mesmo conjunto de dados hiperespectrais, como também, diferentes algoritmos de aprendizado para interpretar esses dados. Dessa forma, ao combinar os resultados de diferentes abordagens de classificação, conseguimos unir indiretamente as informações espectrais e espaciais em um mesmo sistema de interpretação.

# Capítulo 3

## Fundamentação Teórica

Este Capítulo tem a finalidade de fornecer os fundamentos básicos dos métodos e algoritmos utilizados neste trabalho. Uma breve introdução sobre algoritmos de aprendizagem, métodos de combinação, morfologia matemática e GA será dada a seguir.

### 3.1 Algoritmos de Aprendizagem de Máquina

Algoritmos de aprendizagem de máquina são bem difundidos na comunidade científica e há uma vasta literatura sobre esse assunto (Mitchell, 1997; Duda et al., 2000; Scholkopf & Smola, 2001; Theodoridis & Koutroumbas, 2003; Bishop, 2006; Alpaydin, 2010). Esta seção tem o objetivo de fornecer as ideias básicas de três algoritmos de aprendizagem bem conhecidos na literatura técnica: *KNN*, *MLP* e *SVM*.

Antes de iniciar com os conceitos básicos de cada algoritmo de aprendizagem utilizado neste trabalho, iremos definir formalmente o problema de classificação de imagens hiperespectrais. Definimos  $\delta = \{1, \dots, n\}$ , um conjunto de inteiros, indexando os  $n$  *pixels* de uma imagem hiperespectral e  $\psi = \{1, \dots, C\}$  o conjunto de  $C$  classes existentes. Seja  $x = \{x_1, \dots, x_n\} \in \mathbb{R}^{d \times n}$ , os *pixels* dessa imagem, constituindo vetores de características em um espaço  $d$ -dimensional onde o tamanho  $d$  depende da representação utilizada, isto é, *Pixelwise*, *EMP*, etc. Finalmente, dado  $y = \{y_1, \dots, y_n\} \in \psi^n$  uma imagem rotulada. O objetivo da classificação é, para todo pixel  $i \in \delta$  de uma imagem, inferir um rótulo  $y_i \in \psi$  pelo vetor de características  $x_i \in \mathbb{R}^d$ .

#### 3.1.1 *K-Nearest Neighbor*

Dado que os *pixels* (amostras) das imagens hiperespectrais, estão em algum espaço de característica, o *KNN* usa uma medida de distância para definir a proximidade de

uma amostra com outra (Mitchell, 1997; Duda et al., 2000). O algoritmo *KNN* (Cover & Hart, 1967) classifica novas amostras de acordo com as  $K$  amostras mais próximas do conjunto de treinamento. Por exemplo, seja  $\psi = \{1, \dots, C\}$  o conjunto de  $C$  classes existentes,  $x = \{x_1, \dots, x_n\} \in \mathfrak{R}^{d \times n}$  um conjunto de treinamento em um espaço  $d$ -dimensional e  $y = \{y_1, \dots, y_n\} \in \psi^n$  seus respectivos rótulos (classes). Dado um conjunto de teste ainda não rotulado  $t = \{t_1, \dots, t_n\} \in \mathfrak{R}^{d \times n}$ , para classificá-lo, o algoritmo *KNN* calcula a distância de  $t_i$  para toda amostra no conjunto de treino  $x$ . A distância é determinada por uma métrica de similaridade, como por exemplo a distância Euclidiana. Então, um rótulo  $y_i \in \psi$  é atribuído à  $t_i$  considerando o mais a classe mais frequente dentre as  $K$  amostras de treino mais próximas à  $t_i$ .

Apesar de sua simplicidade, o *KNN* possui algumas desvantagens (Mitchell, 1997):

- Alto custo computacional para calcular a distância entre a amostra a ser rotulada e todas outras no conjunto de treino.
- Baixa precisão em espaços de características muito elevados.
- Encontrar o melhor parâmetro  $K$ .

### 3.1.2 *Multilayer Perceptron Neural Network*

Uma Rede Neuronal Artificial (RNA) de múltiplas camadas, ou *MLP*, é uma extensão da RNA *Perceptron* comum (Mitchell, 1997). A *MLP* é composta por um conjunto de unidades de entrada, ou neurônios, que representa a camada de entrada, pelo menos uma camada escondida, e uma camada de saída (Mitchell, 1997; Duda et al., 2000). Uma *MLP* é um poderoso algoritmo de aprendizado capaz de aprender complexos padrões nos dados (por exemplo padrões não lineares). Cada neurônio escondido realiza um mapeamento do espaço de características de entrada usando uma função de transferência. Após um mapeamento apropriado, outras camadas podem aprender essas características mapeadas como simples discriminantes lineares, assim padrões não lineares são aprendidos da mesma forma como lineares. Como resultado, regiões no espaço de características são associadas à uma classe específica e então uma nova amostra pode ser rotulada de acordo com a região em que está inserida. À medida que mais camadas são acrescentadas na *MLP*, mais complexas e sofisticadas separações podem ser feitas no espaço de características. Portanto, a *MLP* pode construir arbitrários limites de decisões, no espaço de características, determinando diferentes e complexas categorias (classes) (Duda et al., 2000). A construção de uma *MLP* possui algumas observações importantes, tais como a escolha do tipo de função de transferência, o

número de camadas escondidas e neurônios em cada camada, que devem ser escolhidos de acordo com o domínio do problema.

### 3.1.3 *Support Vector Machines*

O algoritmo de aprendizado *SVM* se baseia na separação das classes através de margens (Alpaydin, 2010). Dessa forma o treinamento do *SVM* consiste em encontrar um hiperplano separador ótimo, onde a distância de separação entre as margens de cada classe são maximizadas. As amostras que estão situadas sobre as margens são chamadas de *support vectors* e são as mais informativas para a criação do limite de decisão da classificação (Duda et al., 2000; Scholkopf & Smola, 2001; Alpaydin, 2010). O *SVM* possui uma característica interessante conhecida por *kernel trick*. Amostras são mapeadas para um espaço de características, geralmente muito maior que o original, por uma função *kernel* onde nesse espaço transformado consegue-se uma melhor separação linear dos dados (Scholkopf & Smola, 2001). Alguns *kernels* populares são: *Linear*, *Polynomial* e *RBF*. A habilidade de separar dados com distribuição não linear está relacionada com a escolha dessa função, e que deve ser analisada de acordo com o domínio do problema (Duda et al., 2000; Scholkopf & Smola, 2001; Alpaydin, 2010). Usando um mapeamento não linear apropriado, amostras de diferentes classes podem ser linearmente separáveis por um hiperplano neste novo e maior espaço de características.

## 3.2 Métodos de Combinação de Classificadores

O principal objetivo em combinar múltiplos métodos de classificação é produzir uma decisão final que seja melhor que uma única decisão (Kittler et al., 1998; L.I. Kuncheva & Duin, 2001; Kuncheva, 2004). Para Alpaydin (1998) as abordagens de combinação podem ser divididas em duas categorias dependendo da metodologia de representação de características:

- Representação única: todos os classificadores são treinados por dados usando a mesma representação.
- Múltiplas representações: diferentes classificadores são treinados usando representações diferentes de uma mesma entrada.

Ao usar um único tipo de algoritmo de aprendizado, sobre dados de representação única, para produzir diferentes modelos que sejam adequados para combinação,  $N$  conjuntos

distintos de treinamento são aplicados ao classificador, produzindo  $N$  modelos. Essa é a ideia básica das abordagens *boosting* (Schapire, 1989) e *bagging* (Breiman, 1996). Outra maneira de se produzir diferentes modelos de um mesmo algoritmo de aprendizado é através da variação de algum parâmetro desse mesmo algoritmo. Abordagens que utilizam múltiplas representações se beneficiam de informações vindas de diferentes fontes para caracterizar um mesmo conjunto de dados, o que naturalmente proporciona diferentes resultados de classificação mesmo utilizando o mesmo tipo de algoritmo de aprendizado. Em todos os casos os ganhos obtidos através da combinação de vários classificadores depende do quanto os dados são correlacionados (Alpaydin, 1998; Kuncheva, 2004), isto é, os erros e acertos produzidos por cada uma das abordagens de classificação devem ser, preferencialmente, independentes.

### 3.2.1 Combinação Consensual

Em geral, algoritmos de aprendizagem (classificadores) podem produzir saídas do tipo *hard* ou *soft*. Nas saídas do tipo *hard*, ou *crisp*, um único valor é produzido para cada uma das amostras rotulados, esse valor condiz com o número identificador da classe. Enquanto que nas saídas do tipo *soft* temos um certo grau de confiança (Kuncheva, 2004) para cada classe, produzindo mais informação para pós-processamento (Kittler et al., 1998; Alpaydin, 1998; Kuncheva, 2004). Neste trabalho estamos interessados em classificadores capazes de produzir saídas do tipo *soft*, o que possibilita o desenvolvimento de combinações mais bem elaboradas.

As saídas do tipo *soft* podem ser valores *fuzzy*, probabilidades a posteriori, graus de certeza ou confiança, possibilidades, etc., (Kuncheva, 2004), isto é, existem  $c$  valores associados à  $c$  classes. A partir desses valores, outros tipos de regras consensuais podem ser aplicadas para uma combinação mais elaborada. Para isso devemos, primeiramente, construir o chamado Perfil de Decisão (*Decision Profile (DP)*) (Kuncheva, 2004) ilustrado na Figura 3.1. Matematicamente, o *DP* para um dada amostra  $x$  pode ser definido como uma matriz  $L \times c$ :  $DP(x) = [D_1(x), D_2(x), \dots, D_L(x)]^T$ , em que  $L$  é o número de classificadores e  $c$  o número de classes. Cada predição (saída *soft*) é representada por um vetor  $D_i(x) = [d_{i,1}(x), \dots, d_{i,c}(x)]$ , em que  $d_{i,j}(x)$  é o grau de confiança do classificador  $D_i$  para a classe  $j$  (Kuncheva, 2004).

Dentre todos os métodos de combinação o voto majoritário (*MV*) é o mais simples (Kittler et al., 1998; Kuncheva, 2004) e amplamente utilizado para interpretação de imagens hiperespectrais de sensoriamento remoto (Licciardi et al., 2009; Tarabalka et al., 2010a). O *MV* atribui a classe majoritária, dentre as atribuições das  $L$  abordagens de classificação, a uma determinada amostra. Apesar de democrático, o *MV*

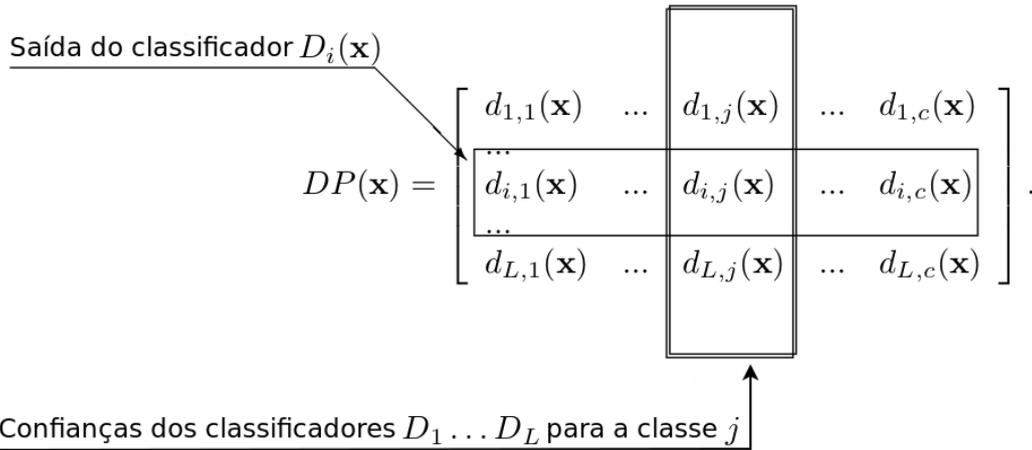


Figura 3.1: Perfil de Decisão ( $DP$ ). Colunas são valores de confiança para  $c$  classes e linhas são os  $L$  métodos de classificação. Adaptado de Kuncheva (2004).

trabalha com atribuições do tipo *hard* e não usa toda a informação disponível, isto é, as saídas do tipo *soft* produzidas pelos classificadores. Métodos de combinação que usam uma única coluna do  $DP(x)$  por vez, são conhecidos como *Class-Conscious combiners* (Kuncheva, 2004) ou *consensus theoretic methods* (Benediktsson & Swain, 1992). Abaixo são apresentados algumas regras consensuais básicas que são usadas para produzir uma combinação, entre dois ou mais classificadores, usando o esquema apresentado:

- Maximum (Max):  $\mu_j(x) = \max_i \{d_{i,j}(x)\}$
- Minimum (Min):  $\mu_j(x) = \min_i \{d_{i,j}(x)\}$
- Average (Aveg):  $\mu_j(x) = \frac{1}{L} \sum_{i=1}^L d_{i,j}(x)$
- Product (Prod):  $\mu_j(x) = \prod_{i=1}^L d_{i,j}(x)$

em que, dado uma amostra  $x$ ,  $\mu_j(x)$  é o novo grau de confiança para a classe  $j$ . Essas operações são, usualmente, chamadas de regras consensuais não-treináveis apenas pelo fato de não precisarem da estimação de nenhum parâmetro. A aplicação de alguma dessas regras sobre o esquema DP produz novos graus de confiança. Após construção adequada de novos valores de confiança para cada amostra de entrada, uma atribuição do tipo *hard* pode ser efetuada pela atribuição do maior valor de confiança do conjunto. Assim o índice do maior valor de confiança no conjunto representa a categoria ou classe assinalada para uma dada amostra.

### 3.3 Morfologia Matemática

A Morfologia Matemática (MM) visa analisar as relações espaciais entre conjuntos de *pixels*, ou seja, a forma, o tamanho e orientação que a matéria pode formar (Serra, 1982; Facon, 1996). Por exemplo, a Figura 3.2 ilustra tipos de estrutura espaciais de interesse que devem fazer parte de uma mesma região.



Figura 3.2: Exemplo de informação estrutural: grande, iluminado, estrutura retangular - pequeno, circular, estrutura texturizada - comprido, fino, estrutura escura... Retirado de Fauvel (2007).

A MM fornece conjuntos de transformações capazes de extrair e tratar essas dependências espaciais. Para isso um conhecido conjunto de formas de diferentes tamanhos, denominados EE (Figura 3.3) (Serra, 1982; Facon, 1996), é aplicado sobre a imagem com o objetivo de extrair características geométricas de interesse. Em ima-

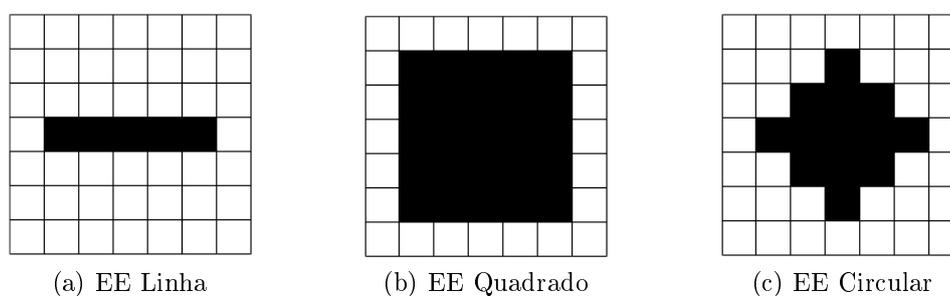


Figura 3.3: Exemplos de EE de diferentes formas.

gens binárias, isto é, apenas dois níveis de intensidades, as transformações morfológicas

podem ser consideradas como operações booleanas usando a teoria de conjuntos, aplicando, por exemplo, os operadores de União ( $\cup$ ) e Interseção ( $\cap$ ). Para imagens em tons de cinza, as transformações são vistas como funções e são realizadas por meio dos operadores *max* e *min*. Os operadores fundamentais em morfologia matemática são a erosão e a dilatação, todos os outros operadores morfológicos podem ser expressos em termos de erosão e dilatação (Serra, 1982; Facon, 1996). Em imagens binárias a erosão reduz e remove objetos, ao passo que a dilatação cresce e conecta os objetos. A Tabela 3.1 sumariza os principais efeitos da erosão e dilatação em imagens em tons de cinza. Em geral, essas operações são duais mas não inversíveis, e quando aplicadas em

Tabela 3.1: Efeitos da erosão e dilatação em imagens em tons de cinza.

<b>Erosão</b>	<b>Dilatação</b>
escurecer a imagem	clarear a imagem
alargar e engordar os vales (padrões escuros)	alargar e engordar os picos (padrões claros)
conectar vales próximos	conectar picos próximos
reduzir e as vezes eliminar picos (padrões claros)	reduzir e as vezes eliminar vales (padrões escuros)
separar picos próximos	separar vales próximos

conjunto geram novas operações conhecidas como abertura e fechamento. A abertura é definida como uma erosão seguida de uma dilatação. A ideia é recuperar a maior parte das estruturas que foram removidas pela erosão. Dessa forma, partes indesejadas são removidas sem alteração do tamanho do objeto, e em imagens em tons de cinza estruturas escuras são alargadas. Por sua vez, o fechamento é obtido por uma dilatação seguida de uma erosão (Serra, 1982; Facon, 1996), usado com o objetivo de agrupar ou preencher estruturas claras na imagem. De maneira geral, em imagens em tons de cinza, a abertura remove estruturas mais claras menores que o EE, enquanto que o fechamento remove estruturas mais escuras.

As operações de abertura e fechamento causam vários efeitos sobre a imagem, especialmente quando o EE é maior que as estruturas presentes na imagem. Por esse motivo, determinadas características geométricas de algumas estruturas podem ser distorcidas ou completamente perdidas (Serra, 1982; Facon, 1996). Obviamente esse é um efeito indesejado, e para superar esse problema operações morfológicas baseadas em reconstruções geodésicas são utilizadas quando a preservação das formas é um fator crucial (Benediktsson et al., 2005; Prasad et al., 2011). Basicamente, a completa remoção ou preservação de componentes conectados na imagem é realizada de acordo com a interação entre o EE da transformação (isto é, fechamento ou abertura) e a imagem. Se um componente da imagem é maior que o EE então ele não será afetado, caso contrário ele será agrupado à estrutura adjacente mais clara ou escura, dependendo se um fechamento ou uma abertura é respectivamente aplicado (Prasad et al., 2011;

Serra, 1982; Facon, 1996). Sendo assim, o uso da reconstrução permite uma melhor preservação das formas na imagem (Benediktsson et al., 2005). Quando aplicado, a abertura por reconstrução remove estruturas claras não conectadas, enquanto que o fechamento por reconstrução remove estruturas escuras não conectadas. Maiores informações sobre as operações morfológicas citadas podem ser encontradas em Serra (1982); Facon (1996).

## 3.4 Algoritmos Genéticos

Algoritmos evolucionários são abstrações da teoria da evolução natural adaptada a sistemas computacionais (Mitchell, 1998; Weise, 2008). Esses algoritmos se baseiam nos conceitos de seleção natural e sobrevivência dos mais adaptados discutidos por *Darwin* sobre a evolução biológica. Algoritmos Genéticos (GA) são uma subclasse dos algoritmos evolucionários. Desde sua criação, GA tem sido umas das técnicas mais difundidas para resolução de problemas de otimização, busca e também de aprendizado de máquina (Goldberg, 1989; Mitchell, 1998; Weise, 2008; Brownlee, 2011).

### 3.4.1 Procedimento Geral

Basicamente, esse processo é constituído por uma população, em que cada indivíduo da população representa uma solução em potencial para um dado problema. Usualmente, indivíduos são representados por uma sequência linear de genes denominado genoma, também frequentemente chamados de cromossomos. O genoma, ou genótipo, é então traduzido para um fenótipo, o qual representa a real solução do problema. GA provê operadores de busca que copiam os esquemas de reprodução sexuada e assexuada na natureza. Na operação de busca sexuada, os genes de dois pais são recombinados, enquanto que na assexuada, apenas a mutação muda o indivíduo. No entanto, a aplicação das duas estratégias juntas é muito comum: primeiramente é feita uma recombinação usando dois indivíduos e então esse novo indivíduo é sujeito a mutação. Cada indivíduo de uma população é avaliado por uma função denominada *fitness*, que por sua vez fornece a “qualidade” de uma solução candidata. O algoritmo aplica os operadores de busca por  $n$  gerações ou até que um determinado critério seja atingido. Ao final o melhor indivíduo da última geração é usado como solução.

GA é capaz de encontrar soluções ótimas ou sub-ótimas, em um grande espaço de busca, em tempo aceitável (Goldberg, 1989; Mitchell, 1998; Weise, 2008; Brownlee, 2011). Para uma boa convergência, devemos modelar uma representação adequada do indivíduo bem como a escolha da função *fitness*. Esses fatores devem analisados

de acordo com o domínio do problema. Um pseudocódigo do GA (Algoritmo 1) é apresentado abaixo:

---

**Algorithm 1** Pseudocódigo para o Algoritmo Genético (retirado de Brownlee (2011)).

---

**Input:**  $Population_{size}, Problem_{size}, P_{crossover}, P_{mutation}$

**Output:**  $S_{best}$

```

1:  $Population \leftarrow InitializePopulation(Population_{size}, Problem_{size})$ 
2:  $EvaluatePopulation(Population)$ 
3:  $S_{best} \leftarrow GetBestSolution(Population)$ 
4: while not( $StopCondition()$ ) do
5:    $Parents \leftarrow SelectParents(Population, Population_{size})$ 
6:    $Children \leftarrow 0$ 
7:   foreach  $Parent_1, Parent_2 \in Parents$  do
8:      $Child_1, Child_2 \leftarrow Crossover(Parent_1, Parent_2, P_{crossover})$ 
9:      $Children \leftarrow Mutate(Child_1, P_{mutation})$ 
10:     $Children \leftarrow Mutate(Child_2, P_{mutation})$ 
11:  end for
12:   $EvaluatePopulation(Children)$ 
13:   $S_{best} \leftarrow GetBestSolution(Children)$ 
14:   $Population \leftarrow Replace(Population, Children)$ 
15: end while
16: return  $S_{best}$ 

```

---

### 3.4.2 Genoma

Os indivíduos são representados por um genoma que deve ser codificado de maneira a conter a informação da solução que ele representa (Goldberg, 1989). Weise (2008) define o genoma como uma sequência linear de um certo tipo de dado. Nessa terminologia, o genoma é formado por elementos denominados genótipos que codificam as características (fenótipos) do indivíduo. O genoma é representado por um conjunto de genes, formando uma *string* que pode ter tamanho definido ou variável. Normalmente essas *strings* são vetores de inteiros (*bit string*) ou vetores de números reais (*real-encoded*). Por possuir uma estrutura linear o genoma é, também, frequentemente chamado de cromossomo (Brownlee, 2011).

### 3.4.3 Seleção

Algumas estratégias são adotadas ao longo do ciclo do GA quanto a produção de uma nova população. A estratégia mais simples substitui toda a população a cada geração, enquanto que outras, como a *steady state* e a geracional, substituem apenas uma porção da população a cada nova geração (Goldberg, 1989; Brownlee, 2011). O método de seleção tem por objetivo selecionar  $m$  indivíduos de uma população, baseado em seus respectivos valores de *fitness*, para o processo de cruzamento (Goldberg, 1989). A ideia básica dos mecanismos de seleção é que os indivíduos mais aptos tenham uma probabilidade maior de serem selecionados. De acordo com Weise (2008) os procedimentos mais comuns para seleção são:

- *Roulette Wheel*: a probabilidade de selecionar um indivíduo é proporcional à sua *fitness*.
- *Ranking*: a probabilidade de um indivíduo ser selecionado é proporcional a sua posição em um *ranking* de uma lista ordenada de todos os indivíduos de uma população.
- Torneio: nesse procedimento  $k$  indivíduos são comparados uns com os outros em um torneio, então o vencedor é selecionado para cruzamento.

Além disso, um mecanismo conhecido como elitismo é, usualmente, aplicado com o objetivo de garantir que a melhor solução será sempre passada para as gerações seguintes.

### 3.4.4 Cruzamento

A combinação dos indivíduos é feita pelo processo de cruzamento. Após selecionados dois indivíduos (pais), eles são combinados de acordo com algumas estratégias: cruzamento de um ponto, dois pontos e múltiplos pontos (Goldberg, 1989; Weise, 2008). No cruzamento de um ponto (Figura 3.4a), os dois pais são divididos em um determinado ponto de cruzamento. Subsequentemente, um filho é produzido ao juntar a segunda parte do segundo pai à primeira parte do primeiro pai. No cruzamento de dois pontos (Figura 3.4b), os pais são divididos em dois pontos. Um novo indivíduo é produzido combinando a primeira e a terceira parte do primeiro pai com a parte do meio do segundo. Esse processo pode ser adaptado para  $n$  pontos (Figura 3.4c).

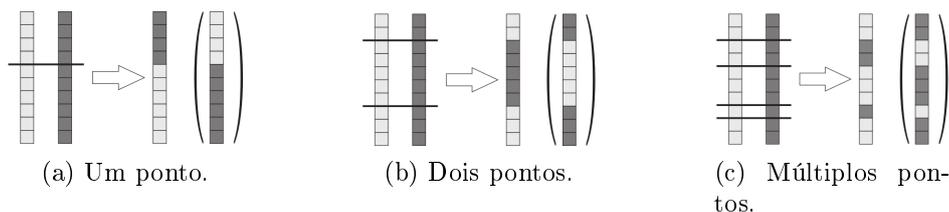


Figura 3.4: Operações típicas de cruzamento. Retirado de Weise (2008).

### 3.4.5 Mutação

A mutação é um importante método para a preservação da diversidade das soluções candidatas. Para prevenir que as soluções não caiam em um ótimo local, é feita a introdução de pequenas mudanças aleatórias nos genes dos indivíduos (Weise, 2008). Normalmente, a mutação é aplicada com uma baixa probabilidade de ocorrência. Algumas variações desse mecanismo podem ser observadas, por exemplo, em uma representação do tipo *bit string* os genes seriam invertidos de acordo com um tipo de mecanismo de mutação adotado (Figura 3.5).

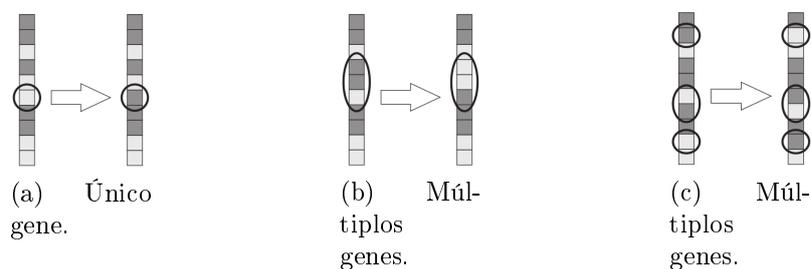


Figura 3.5: Tipos da mutação. Retirado de Weise (2008).

# Capítulo 4

## Metodologia

Baseado nas considerações feitas na Seção 3.2 do Capítulo 3, este presente trabalho visa combinar diferentes abordagens de classificação para a interpretação de dados hiperespectrais de sensoriamento remoto (Figura 4.1). Para isso diversos algoritmos de aprendizado foram treinados usando diferentes características extraídas dos dados hiperespectrais. Primeiramente serão descritos as abordagens de classificação base,

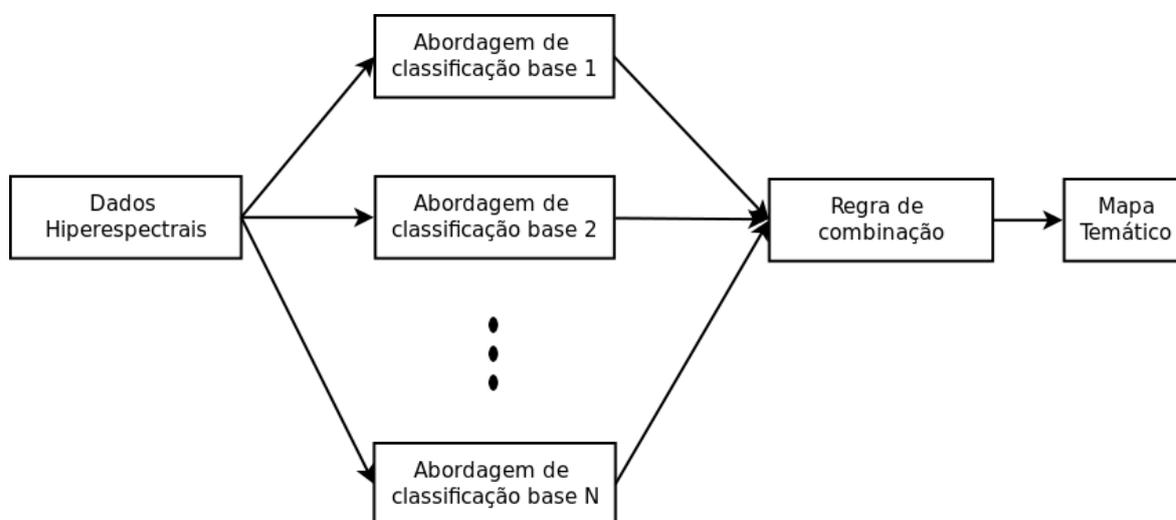


Figura 4.1: Diagrama do processo de combinação de múltiplas abordagens de classificação. Adaptado de Briem et al. (2002).

ou seja, as abordagens de classificação que serão utilizadas para a combinação. Para se obter uma maior diversidade entre as abordagens de classificação base, utilizamos diversos tipos de representações dos dados de entrada, e além disso, algoritmos de aprendizado de natureza diferentes. Dessa maneira esperamos produzir um conjunto de abordagens de classificação base que se complementam, e assim permitir maiores

chances de sucesso na combinação.

Após a descrição das abordagens de classificação base, serão apresentados os métodos de combinação propostos para o problema da classificação de imagens hiperespectrais de sensoriamento remoto. Primeiramente descrevemos a aplicação de métodos consensuais (Kuncheva, 2004; Alpaydin, 2010), tais como as regras *Average*, *Product*, *Maximum*, *Minimum* e *MV*, e então descrevemos nossa primeira proposta baseada na regra *LOP*, em que os pesos são otimizados usando GA. Chamamos essa proposta de *WLC-GA* (Santos et al., 2012a,b, 2013). Finalmente descrevemos outro método proposto para a combinação dos dados hiperespectrais de sensoriamento remoto, este se baseia em *Stacked Generalization* (Wolpert, 1992) usando o algoritmo de aprendizado *SVM* como meta-classificador.

## 4.1 Abordagens de Classificação Base

Foram utilizadas quatro abordagens de classificação base, duas que se baseiam somente na informação espectral e duas que se baseiam na informação espectral-espacial. As abordagens de classificação que se baseiam na informação espectral utilizadas neste trabalho são: a tradicional *Pixelwise* usando o classificador *SVM* (*Pixelwise-SVM*) (Melgani & Bruzzone, 2004), e outra baseada em seleção de características, que também faz parte deste trabalho de dissertação, denominada *Feature Extraction by Genetic Algorithms* (*FEGA*), utilizando o classificador *KNN* (*FEGA-KNN*) (Santos et al., 2012c). As abordagens que utilizam informação espacial-espectral são: a tradicional *EMP* utilizando o classificador *MLP* (*EMP-MLP*) (Benediktsson et al., 2005) e também, como proposta deste trabalho dissertação, uma abordagem baseada no trabalho de Fauvel et al. (2008). Nessa última, a junção da informação produzida pelo *EMP* e a informação original dos dados hiperespectrais são concatenados após redução da dimensionalidade usando *PLS* (Wold., 1985), chamamos essa abordagem de *SpecEmpPLS-SVM*.

### 4.1.1 *Pixelwise-SVM*

A representação de características chamada *Pixelwise* é a mais simples. Ela é composta por todas as bandas da imagem hiperespectral, isto é, os dados originais. Neste tipo de abordagem nenhum tipo de informação espacial é utilizada, somente a informação espectral de cada banda da imagem é levada em consideração no processo de classificação. Dessa forma, cada banda espectral se torna então uma característica presente nos dados. Diante da alta dimensionalidade imposta por esse tipo de representação, Melgani

& Bruzzone (2004) e Camps-Valls & Bruzzone (2005) concluíram que a aplicação do algoritmo de aprendizado *SVM* (Duda et al., 2000; Scholkopf & Smola, 2001; Alpaydin, 2010) é mais adequada para essa situação.

#### 4.1.2 *FEGA-KNN*

Neste trabalho de dissertação um método de seleção de características baseado em métodos *filter* mas também supervisionado é proposto. Existem várias razões para reduzir a dimensionalidade nos dados hiperespectrais. Uma delas é o fenômeno de Hughes (Hughes, 1968), ou como é mais conhecido por “maldição da alta dimensionalidade”. Esse fenômeno acontece quando o número de dimensões (características) é muito maior que o número de amostras de treinamento disponíveis, levando o classificador a produzir uma generalização imprecisa (Hughes, 1968; Duda et al., 2000; Mitchell, 1997; Theodoridis & Koutroumbas, 2003; Bishop, 2006; Alpaydin, 2010). Podemos destacar também que elevados espaços de características têm se mostrado vazios e fazer uma estimativa da densidade se torna mais difícil a medida que se aumenta a dimensionalidade dos dados (Plaza et al., 2009). Outra razão óbvia consiste na redução da complexidade computacional. Em uma RN mais características implicam em mais pesos sinápticos a serem ajustados (Bishop, 2006). Assim, a seleção de características tem como objetivo reduzir o número total de características nos dados e ao mesmo tempo manter o máximo possível de informação discriminativa (Bishop, 2006).

A seleção de  $B$  bandas espectrais (características) ótimas, isto é, com o maior poder de discriminação possível, é um problema de otimização combinatória com um espaço de busca bem elevado. Por exemplo, o número total de combinações possíveis para seleção de  $B$  bandas em um conjunto de  $N$  é  $\binom{N}{B} = \frac{N!}{B!(N-B)!}$ . Para cada combinação de características selecionadas um critério de separabilidade deve ser usado com o objetivo de encontrar o melhor subconjunto. Dado essas condições, GA se mostra uma ferramenta adequada capaz de guiar uma busca afim de otimizar um certo critério de separabilidade.

Como discutido anteriormente (Seção 3.4), cada indivíduo representa uma solução em potencial e sua adequada modelagem é um fator crucial na resolução de um problema. Nesta proposta modelamos o indivíduo (cromossomo) como a seguir: cada cromossomo é representado por um *bit string* de tamanho igual ao número de  $B$  bandas espectrais existentes na imagem hiperespectral, ou seja, a quantidade de genes é igual ao número de bandas. Cada gene representa a presença (bit 1) ou a ausência (bit 0) de uma banda específica, conforme ilustrado na Figura 4.2. Assim, a partir da composição das bandas presentes (bits 1), o subconjunto de bandas selecionadas, ou subconjunto

de características, é formado. Como dito anteriormente a qualidade de cada indivíduo é avaliada de acordo com a função *fitness* e o sucesso da utilização de GA depende da escolha de uma função *fitness* adequada.

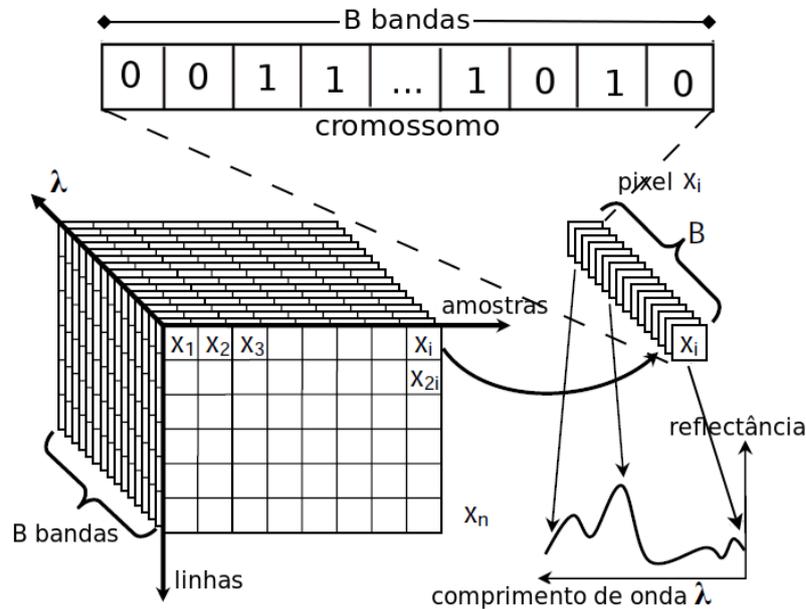


Figura 4.2: Representação do genoma e mapeamento das bandas espectrais, 1 significa presença e 0 ausência de determinada banda no hiper-cubo. Modificado de Tarabalka (2010).

Dado um conjunto de treino caracterizado por um subconjunto de bandas espectrais, estamos interessados em uma métrica baseada na homogeneidade entre amostras pertencentes a mesma partição (grupo ou classe), como também na dissimilaridade entre amostras de diferentes partições. Para isso, métricas de validação de agrupamentos (Bezdek & Pal, 1995; Halkidi et al., 2001; Bishop, 2006) podem ser utilizadas. Uma entre as mais populares métricas de validação de agrupamentos é conhecida como índice de Dunn (*Dunn's Index*) (Dunn, 1974). O índice de Dunn fornece informação sobre a qualidade do agrupamento formado, dessa forma, quanto maior o índice de Dunn, obtido ao avaliar um sistema de partições/agrupamentos caracterizado por um determinado subconjunto de características, mais densos e distantes uns dos outros estão esses agrupamentos.

Para construir os agrupamentos necessitamos de um conjunto de amostras rotuladas, chamamos de conjunto de “treinamento”, e é nesse mesmo conjunto que iremos avaliar, indiretamente, a qualidade das características ou bandas selecionadas. Seja  $U = \{C_1, \dots, C_c\}$  um sistema de agrupamentos composto por um dado subconjunto de

características. O índice de *Dunn* pode ser calculado como:

$$v_D(U) = \min_{1 \leq i \leq c} \left\{ \frac{\min_{\substack{1 \leq j \leq c \\ i \neq j}} \left\{ \frac{\delta(C_i, C_j)}{\max_{1 \leq k \leq c} \{\Phi(C_k)\}} \right\} \right\},$$

em que  $\Phi(C_k)$  é o diâmetro do agrupamento  $C_k$  e  $c$  o número de agrupamentos no sistema. Note que  $c$  é o número de classes disponíveis. Neste trabalho, adotamos  $\delta(C_i, C_j)$  como a distância entre centroides dos agrupamentos  $C_i$  e  $C_j$ , no entanto qualquer medida *inter-cluster* pode ser aplicada (Bezdek & Pal, 1995). A medida de similaridade, ou métrica de distância, usada é a distância Euclidiana:

$$\delta(x, y) = \sqrt{(x_1 - y_1)^2 + \dots + (x_n - y_n)^2}.$$

Desse modo o objetivo é minimizar distâncias entre amostras pertencentes a um mesmo agrupamento (*intra-cluster*) e maximizar distâncias entre amostras de diferentes agrupamentos (*inter-cluster*). Assim, esperamos encontrar subconjuntos de bandas espectrais que maximizem o índice de Dunn de tal forma que classificadores possam construir limites de decisão de uma maneira mais precisa.

Pelo fato de que a métrica usada para avaliação do agrupamento se baseia em distâncias *inter-cluster* e *intra-cluster*, podemos assumir que um método de classificação que também seja baseado em distância entre amostras seja mais adequado para aprender o subconjunto de características produzido pelo *FEGA*. Dessa forma optamos pela utilização algoritmo de classificação *KNN* (Cover & Hart, 1967; Mitchell, 1997; Duda et al., 2000).

Em Santos et al. (2012c), um trabalho totalmente dedicado a essa abordagem, compara-se o *FEGA* com *Singular Value Decomposition Band Subset Selection (SVDSS)* (Velez-Reyes & Jimenez, 1998), outra abordagem amplamente utilizada na área de sensoriamento remoto para seleção de subconjunto de bandas espectrais (Velez-Reyes & Jimenez, 1998; Velez-Reyes et al., 2000; Jimenez-Rodriguez et al., 2007). Utilizando diversos tipos de algoritmos de classificação, nesse trabalho o método *FEGA* se mostrou tão efetivo quanto o *SVDSS* em termos de acurácia, e em alguns casos até melhor quando utilizando o classificador *KNN*.

### 4.1.3 EMP-MLP

Como vimos na Seção 2.2 do Capítulo 2, diversos autores destacam a importância da informação contextual, ou seja, a relação de dependência entre *pixels* de uma mesma região. A abordagem de classificação *EMP-MLP* utiliza representação espectral-

espacial (Fauvel et al., 2012) das características presentes nos dados hiperespectrais através da aplicação de operações morfológicas descritas na Seção 3.3.

Em imagens hiperespectrais as transformações morfológicas são aplicadas, usando a mesma lógica daquela aplicada em imagens em tons de cinza, em cada uma das bandas separadamente. Devido ao volume de informação gerado, diversos autores (Benediktsson et al., 2005; Palmason et al., 2005; Fauvel et al., 2008, 2009; Dalla Mura et al., 2011; Licciardi et al., 2012) propõem a utilização de algum método de extração de características, tais como *PCA*, *ICA*, *DBFE*, *NWFE*, *KPCA* e *NLPCA* capaz de comprimir maior parte da informação hiperespectral em apenas alguns componentes (bandas, características). Dessa forma o volume de informação gerado pelas transformações morfológicas é consideravelmente minimizado.

Para construção do *EMP* é preciso diminuir o número de características presentes na imagem, isto é, o conjunto de bandas espectrais. A abordagem tradicional proposta por Benediktsson et al. (2005) utiliza *PCA* como extração de características. Sobre cada Componente Principal (PC) é aplicado uma série de sucessivas aberturas e fechamentos por reconstrução com EE de tamanho variável. A cada aplicação da abertura/fechamento o EE é aumentado de tamanho em um determinado passo de crescimento, isso é realizado afim de conectar, ou separar, estruturas maiores na imagem. Obviamente o tamanho inicial do EE depende da resolução espacial de cada base de dados, resoluções maiores necessitam de EE maiores para a caracterizam adequada das estruturas na imagem. Assim, o *EMP* possui um conjunto de parâmetros que deve ser otimizado de acordo com a base de dados utilizada. Os parâmetros são: número de Componentes Principais (PCs) utilizadas, número de abertura/fechamento, tamanho do EE e passo de crescimento do EE. Benediktsson et al. (2005) sugerem um conjunto “padrão” de parâmetros que tem funcionado de maneira satisfatória para as bases de dados atuais. Um número de PCs que contabilizam pelo menos 95% da variabilidade total da informação presente nos dados hiperespectrais deve ser escolhido para garantir que a maior parte da informação espectral seja mantida (Benediktsson et al., 2005).

O hipercubo é construído de forma que cada banda espectral seja composta por todas as imagens geradas pelas transformações morfológicas e também pelas  $p$  primeiras PCs. O número total de características no vetor será  $(2n + 1) \times p$ , em que  $n$  é o número de abertura/fechamento e  $p$  o número de PCs utilizadas. A Figura 4.3 ilustra a criação do *EMP* na base de dados *Indian Pines*. Essa técnica permite conectar estruturas similares nos dados hiperespectrais, possibilitando uma representação mais homogênea de objetos que pertencem a uma mesma classe (Benediktsson et al., 2005). A seguir será descrito o algoritmo de aprendizado utilizado por Benediktsson et al. (2005) para classificação utilizando a abordagem *EMP*.

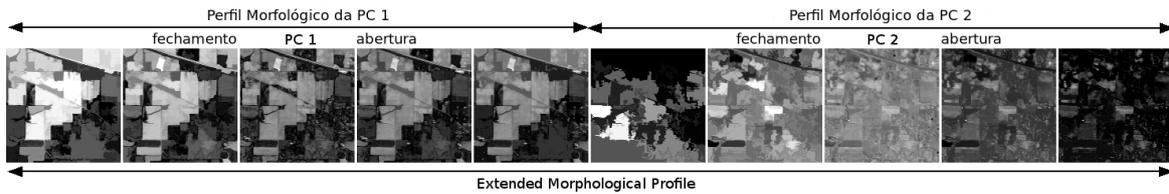


Figura 4.3: Exemplo do *EMP* usando duas PCs, EE circular de tamanhos 4, 6 e 8 (passo de crescimento do EE igual a 2) para a base de dados *Indian Pines*.

#### 4.1.4 *SpecEmpPLS-SVM*

Recentemente, a fusão dos dois tipos de informações, isto é, espectral e espacial, tem mostrado grandes avanços na classificação de imagens hiperespectrais de sensoriamento remoto (Fauvel et al., 2012). Após a introdução do *EMP* (Benediktsson et al., 2005) para a modelagem da informação espacial em imagens hiperespectrais, Fauvel et al. (2008) propuseram a utilização de algum algoritmo supervisionado de extração de características para a construção de um vetor de características mais informativo e o menos redundante possível. Primeiramente um *EMP* totalmente carregado de informação espacial é construído. Para isso Fauvel et al. (2008) recomenda usar um número de PCs que correspondem a 99% da variância total na imagem na criação do *EMP*, o que eleva consideravelmente a dimensionalidade do *EMP*. Em um segundo momento algum algoritmo de extração de características é aplicado aos dados hiperespectrais originais e ao *EMP*. Finalmente, as características extraídas tanto dos dados originais quanto do *EMP* são concatenadas em um único vetor e então classificadas pelo algoritmo de aprendizado SVM.

Baseado na abordagem proposta por Fauvel et al. (2008), esta abordagem realiza a fusão dos dados hiperespectrais originais e do *EMP* por meio das características produzidas pelo método de redução de dimensionalidade *PLS* (Wold., 1985) aplicado individualmente nos respectivos tipos de informações (espectral e espacial). O método *PLS* (Wold., 1985; Rosipal & Krämer, 2006) tem sido aplicado com sucesso na área de visão computacional para identificação de faces (Schwartz et al., 2012) e detecção de pedestres (Schwartz et al., 2011; Schwartz, 2011). Em sua forma clássica o *PLS* é baseado no algoritmo *Nonlinear Iterative Partial Least Squares (NIPALS)* (Wold., 1985), modelando relações entre as características presentes nos dados através de variáveis latentes. Basicamente, novas características são produzidas, variáveis latentes, como combinações lineares do conjunto original de características considerando suas respectivas classes. Dessa forma, o *PLS* busca combinações lineares ortogonais, das características originais, correlacionadas com a classe em questão, maximizando a vari-

ância nas novas características obtidas o máximo possível. Mais detalhes sobre o *PLS* são encontrados em Wold. (1985); Barker & Rayens (2003); Rosipal & Krämer (2006); Schwartz et al. (2012).

Após a construção da nova representação os novos conjuntos de características são submetidos ao algoritmo de aprendizado *SVM* discutido na Seção 3.1.3. A Figura 4.4 ilustra o processo de fusão utilizando *PLS* como extração de características.

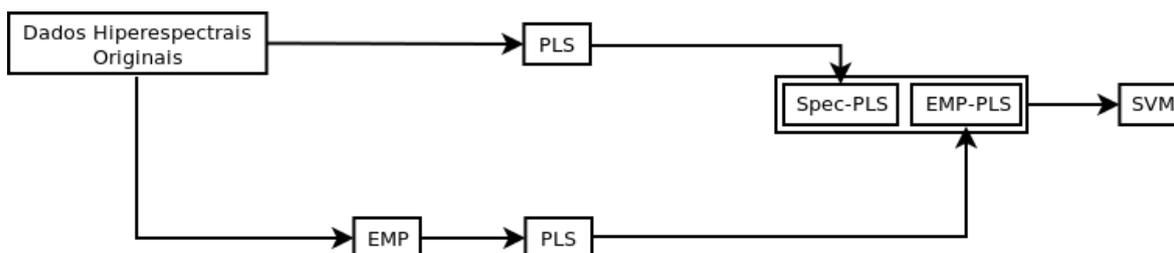


Figura 4.4: Diagrama ilustrativo da fusão da informação espectral e a espacial provida pelo *EMP*. Modificado de Fauvel et al. (2008).

## 4.2 Combinação de Múltiplas Abordagens de Classificação

Uma vez definidos os métodos de representação dos dados hiperespectrais e algoritmos de aprendizado utilizado em cada um dos diferentes tipos de características, podemos prosseguir com os métodos de combinação. Nesta seção serão descritos os dois métodos propostos para a combinação de múltiplas abordagens de classificação aplicadas aos dados hiperespectrais de sensoriamento remoto. Primeiramente é apresentado um método consensual baseado em uma combinação linear ponderada otimizada por GA. Em seguida um método usando a ideia de *Stacked Generalization* com o algoritmo SVM, atuando como regra de combinação, é apresentado.

### 4.2.1 WLC-GA

Cada abordagem de classificação converge a uma diferente solução e falha em diferentes circunstâncias, ou seja, um certo modelo pode levar a erros em determinados conjuntos de amostras e acertos em outros. Segundo Alpaydin (1998) representações diferentes produzem características explicitamente diferentes, o que permite uma melhor caracterização dos dados.

Quando classificadores possuem diferentes taxas de acertos, é sensato dar mais poder de discriminação para aqueles que possuem acurácias mais elevadas. Essa é a

ideia básica de métodos consensuais ponderados (Kuncheva, 2004), por exemplo o LOP usado por Benediktsson & Swain (1992); Benediktsson & Kanellopoulos (1999) para combinação de dados de sensoriamento remoto de múltiplas fontes. Esses métodos são chamados de treináveis, pois necessitam de encontrar o melhor conjunto de pesos para produzir os melhores conjuntos de confiança para cada amostra.

Em Kittler et al. (1998), a regra *sum* provou-se ser melhor que outras regras básicas discutidas anteriormente. Portanto, propomos a utilização da regra *sum* como base para realizar uma combinação linear ponderada de maneira consensual usando o esquema *DP*. Chamamos essa combinação de *WLC-GA* (Santos et al., 2012a,b). Assim, novos graus de confiança são produzidos de acordo como a seguir:

$$\mu_j(x) = \sum_{i=1}^L w_i \times d_{i,j}(x) \quad (4.1)$$

tal que satisfaça:

$$\forall i, w_i \geq 0, e \sum_{i=1}^L w_i = 1 \quad (4.2)$$

em que  $L$  é o número de abordagens de classificação,  $w_i$  é o peso da  $i$ -ésima abordagem de classificação,  $d_{i,j}(x)$  é o grau de confiança da  $i$ -ésima abordagem de classificação para a classe  $j$ , e  $\mu_j(x)$  é o novo grau de confiança para a classe  $j$ . Então, um rótulo é atribuído para uma dada amostra  $x$  como o índice do maior grau de confiança  $\mu_j(x)$ . Essa é a ideia básica da *LOP* descrita em Benediktsson & Swain (1992), no entanto, nesse mesmo trabalho Benediktsson & Swain (1992) ressaltam que encontrar pesos ótimos é considerado uma tarefa muito difícil e deve ser melhor investigada. Por esse motivo optamos por uma busca global usando GA.

Como sabemos, a qualidade de cada solução candidata é determinada por uma função chamada *fitness*. Cada solução, isto é, indivíduo, representa um conjunto de pesos que deverá ser otimizado com o objetivo de obter uma acurácia tão elevada quanto o possível no processo de combinação. Portanto, para avaliar e determinar a qualidade do indivíduo optamos por uma abordagem supervisionada onde o *Overall Accuracy (OA)*, em um conjunto de treino, é usado como medida de qualidade de cada um dos indivíduos. Usamos uma representação do tipo *bit string* (Figura 4.5) para codificar os pesos no genoma do indivíduo.

Nesse tipo de combinação, estamos interessados em encontrar pesos que melhor contribuem para a obtenção de uma maior acurácia, ou seja, estamos ponderando a  $i$ -ésima abordagem de classificação de acordo com sua respectiva contribuição no processo de combinação. Os pesos encontrados não alteram individualmente as probabilidades

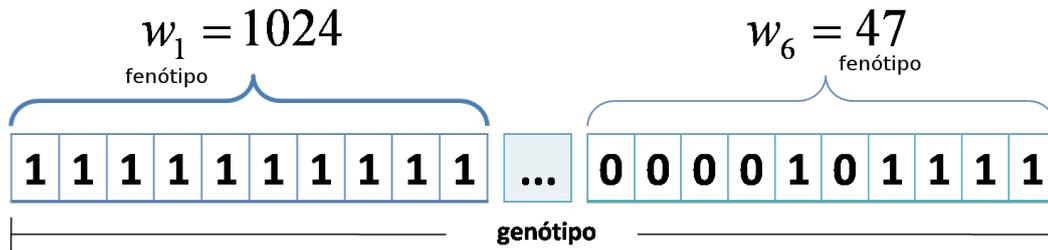


Figura 4.5: Exemplo de uma representação do tipo *bit string*. Dez bits (isto é, genes) são usados para codificar cada peso, possibilitando  $2^{10} = 1024$  diferentes valores de peso.

de cada abordagem de classificação, e por isso, a informação gerada pelos classificadores é preservada. Dessa forma, os pesos atuam de forma que um grau de influência é atribuído à cada uma das abordagens base na combinação. A Figura 4.6 ilustra o esquema de combinação utilizando *WLC-GA*.

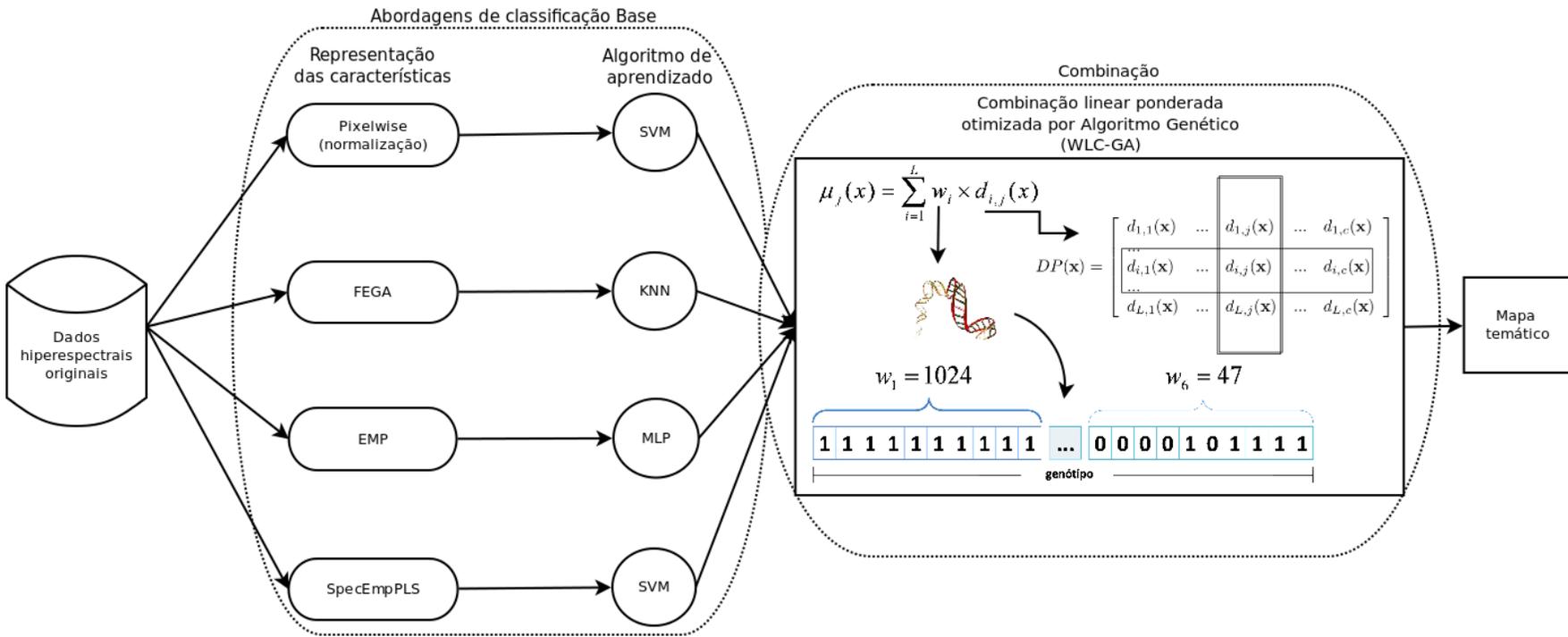


Figura 4.6: Diagrama do sistema de múltiplas abordagens de classificação proposto neste trabalho. Quatro tipo de representação de características e três tipos de algoritmos de aprendizado são usados para garantir maior diversidade.

### 4.2.2 *Stacking-SVM*

Primeiramente proposto por Wolpert (1992) a principal ideia da abordagem de combinação *Stacked Generalization* é aprender com as saídas de diferentes algoritmos de aprendizado. Cada algoritmo de aprendizado obtém diferentes modelos, seja por uma multirepresentação da entrada de dados ou por serem algoritmos de natureza diferentes. Dessa forma, erros correlacionados podem ser identificados (Alpaydin, 1998; Ting & Witten, 2011), e aprendidos, por um outro algoritmo de aprendizado (Wolpert, 1992). O primeiro passo nessa abordagem é coletar os dados previstos por cada modelo (isto é, probabilidades, confiança, etc.) e criar um novo conjunto de dados, chamado de meta-dados. Na terminologia de Wolpert (1992) são atribuídos níveis para cada estágio do processo. Os dados originais e algoritmos de aprendizado usado no primeiro estágio são chamados de dados de nível-0 e modelos de nível-0, respectivamente. Enquanto que os novos dados criados a partir dos modelos aprendidos no primeiro estágio são chamados de dados de nível-1, e o algoritmo de aprendizado neste segundo estágio de modelos de nível-1 (aqui chamado de meta-classificador).

Com o intuito de facilitar o entendimento, definimos o dados de nível-0 e modelos de nível-0 como “abordagens de classificação base”, enquanto que o método de combinação (modelos de nível-1) chamamos de meta-classificador (Figura 4.7). Dado que para toda amostra  $x$  cada abordagem de classificação base  $D_i$  produz um conjunto com os devidos graus confiança ( $d_{i,j}(x)$ ) para cada uma das  $c$  classes ( $D_i(x) = \{d_{i,1}(x), \dots, d_{i,c}(x)\}$ ). Por meio da concatenação das saídas de cada uma das abordagens de classificação base  $D_i$  é formado as meta-características  $Z(x) = [d_{1,1}(x), \dots, d_{1,j}(x), \dots, d_{1,c}(x), d_{i,1}(x), \dots, d_{i,j}(x), \dots, d_{i,c}(x), d_{L,1}(x), \dots, d_{L,j}(x), \dots, d_{L,c}(x)]$ . Nesse nível chamamos a base de dados de meta-dados. Note que esse método pode levar a um significativo aumento da dimensionalidade conforme o número de abordagens de classificação base sejam introduzidas. Logo, o número de meta-características presentes nos meta-dados será igual a  $L \times c$ , em que  $L$  é o número de abordagens de classificação base e  $c$  o número de classes existentes (saídas do tipo *soft*) no base de dados (dados de nível-0).

Devido ao sucesso do *SVM* ao lidar com alta dimensionalidade, neste trabalho investigamos o uso desse algoritmo de aprendizado atuando como meta-classificador, afim de identificar e aprender erros correlacionados, e assim, produzir uma combinação mais adequada. Denominada *Stacking-SVM*, a abordagem proposta produz novas características, (meta-características), utilizando as quatro abordagens de classificação base descritas na Seção 4.1. Propomos então a utilização do algoritmo de aprendizado *SVM* no estágio de meta-classificador, atuando como um combinador. Portanto, a

combinação é feita através do algoritmo de aprendizado *SVM* segundo a abordagem *Stacked Generalization* (Wolpert, 1992). Logo, cada *pixel* na imagem hiperespectral é novamente submetido ao processo de treinamento usando seu novo conjunto de características (isto é, meta-características) e então re-classificados. O esquema ilustrado na Figura 4.7 sumariza o método proposto.

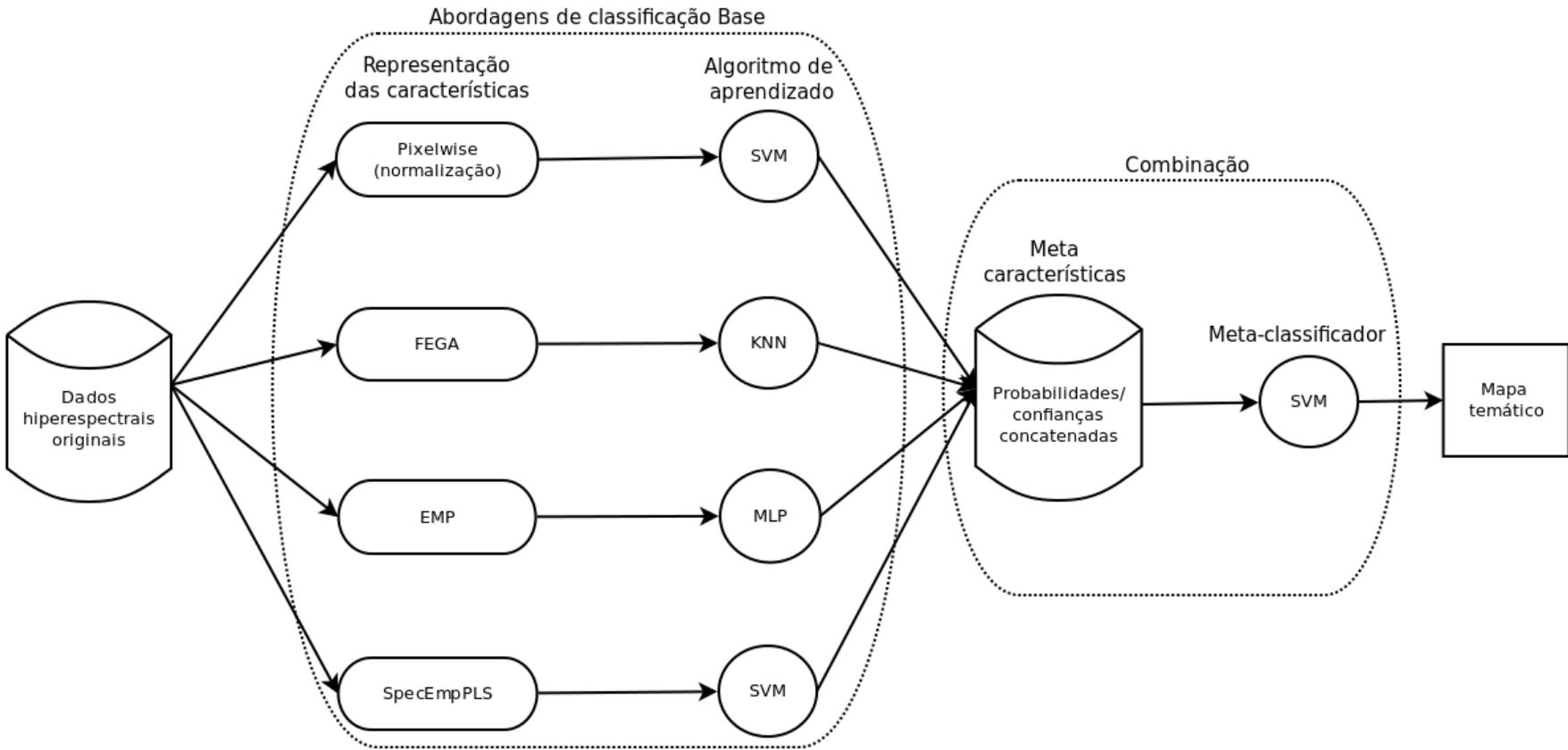


Figura 4.7: Diagrama ilustrativo da abordagem proposta *Stacking-SVM*. Novas características são produzidas pela concatenação das saídas obtidas pelas abordagens de classificação base (meta-características) e então submetidas a um novo processo de aprendizado usando o classificador *SVM*.

# Capítulo 5

## Experimentos

Este capítulo se concentra na descrição das bases de dados utilizadas bem como os critérios de avaliação, parâmetros e cenários dos experimentos. Definiremos um protocolo de experimentação baseado no procedimento padrão de treino e teste (Duda et al., 2000), de maneira que várias configurações de treino/teste são usadas para uma melhor comparação entre os métodos. Além disso, diferentes cenários, variando as abordagens de classificação base usadas na combinação, foram experimentados com o intuito de verificar o comportamento dos métodos de combinação propostos. Também serão discutidos todos os parâmetros usados em cada uma das abordagens de classificação base apresentadas no Capítulo 4.

Os experimentos foram executados em um computador pessoal com processador *Intel® Core™ i7-2630QM* e 6 GB de memória principal com sistema operacional *Ubuntu 12.04*. Foi desenvolvido um *framework* em MATLAB contendo um conjunto de operações de modo a facilitar o tratamento das imagens hiperespectrais. No entanto, algumas manipulações, como por exemplo a aplicação do método *PCA*, foram realizadas usando a ferramenta *Hyperspectral Image Analysis Toolbox (HIAT)* (Arzuaga-Cruz et al., 2004) disponível para o ambiente MATLAB (MATLAB, 2008). Para os algoritmos de aprendizado usamos a implementação do *SVM* fornecida pela biblioteca *A Library for Support Vector Machines (LIBSVM)* (Chang & Lin, 2011) e as ferramentas próprias para *RN* e *KNN* disponíveis no MATLAB (MATLAB, 2012) para a *MLP* e *KNN*, respectivamente. Tanto o *FEGA* quanto o *WLC-GA* foram escritos em linguagem C++.

## 5.1 Bases de Dados Hiperespectrais

Para uma maior confiabilidade dos resultados os experimentos foram executados em dois diferentes tipos de bases de dados hiperespectrais. Sendo que uma imagem, isto é, base de dados, é caracterizada por uma área de cultivos, enquanto que a outra representa uma área tipicamente urbana. As bases de dados *Indian Pines* e *Pavia University* foram obtidos pelo sensores *AVIRIS* (Green et al., 1998) e *ROSIS* (Gege et al., 1998), respectivamente. A seguir são apresentadas algumas informações relevantes sobre as duas bases de dados utilizadas:

1. A base de dados *Indian Pines* corresponde a uma área de vegetação e cultivos obtida na região Noroeste de *Indiana*, EUA. Possui dimensões de  $145 \times 145$  *pixels*, uma resolução espacial de  $20m$  por *pixel*, e 220 canais que cobrem uma faixa espectral de  $0,4\mu m$  a  $2,5\mu m$ . A Figura 5.1a mostra uma composição Red-Green-Blue (RGB) em falsa cor usando as bandas espectrais: 50 para o canal R, 20 para o G e 10 para o B; Usualmente, nessa base de dados, as vinte bandas mais ruidosas são removidas (canais 104-108, 150-163, e 220) (Camps-Valls et al., 2006), restando 200 canais espectrais para os experimentos. Dezesesseis classes de interesse são consideradas e a quantidade de amostras por classe é mostrado na Tabela 5.1. O *Ground-truth* pode ser visualizado na Figura 5.1b.

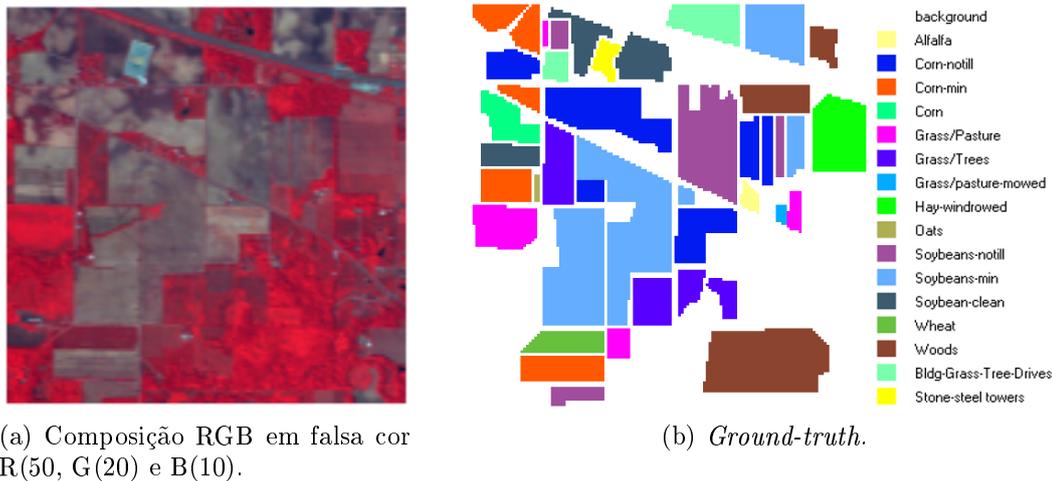


Figura 5.1: *AVIRIS - Indian Pines*.

2. A base de dados de *Pavia University* corresponde a uma área urbana capturada sobre a Universidade de Pavia na Itália. A imagem é composta de  $610 \times 340$  *pixels* com uma resolução espacial de  $1,3m/pixel$ , faixa espectral de  $0,43\mu m$  a  $0,86\mu m$

Tabela 5.1: Número de amostras por classe para o *Indian Pines*.

Classes	Número de amostras ( <i>pixels</i> )
<i>Alfalfa</i>	54
<i>Corn-notill</i>	1434
<i>Corn-min</i>	834
<i>Corn</i>	234
<i>Grass/Pasture</i>	497
<i>Grass/Trees</i>	747
<i>Grass/pasture-mowed</i>	26
<i>Hay-windrowed</i>	489
<i>Oats</i>	20
<i>Soybeans-notill</i>	968
<i>Soybeans-min</i>	2468
<i>Soybean-clean</i>	614
<i>Wheat</i>	212
<i>Woods</i>	1294
<i>Bldg-Grass-Trees-Drives</i>	380
<i>Stone-steel towers</i>	95
<i>Background</i>	10659

ao longo de seus 115 canais. Doze canais ruidosos são removidos, permanecendo 103 bandas para os experimentos<sup>1</sup>. A Figura 5.2a mostra uma composição RGB em falsa cor dessa imagem. Nove classes de interesse são identificadas. Tabela 5.2 quantifica todas as amostras de todas as nove classes. O *Ground-truth* pode ser visualizado na Figura 5.2b.

Tabela 5.2: Número de amostras por classe para o *Pavia University*.

Classes	Número de amostras ( <i>pixels</i> )
<i>Asphalt</i>	6631
<i>Meadow</i>	18649
<i>Gravel</i>	2099
<i>Trees</i>	3064
<i>Metal Sheets</i>	1345
<i>Bare Soil</i>	5029
<i>Bitumen</i>	1330
<i>Bricks</i>	3682
<i>Shadow</i>	947
<i>Background</i>	164624

Note que *background* não é uma classe, e é usada apenas para identificar as amostras não rotuladas. Sendo assim, essas amostras não são usadas para nenhum cálculo estatístico neste trabalho, apenas para a construção do mapa temático completo. Mais detalhes sobre essas bases de dados podem ser encontrados em Fauvel (2007) e Tarabalka (2010).

<sup>1</sup>Os dados disponíveis já continham apenas 103 canais, e dessa forma não foi possível identificar as bandas removidas.

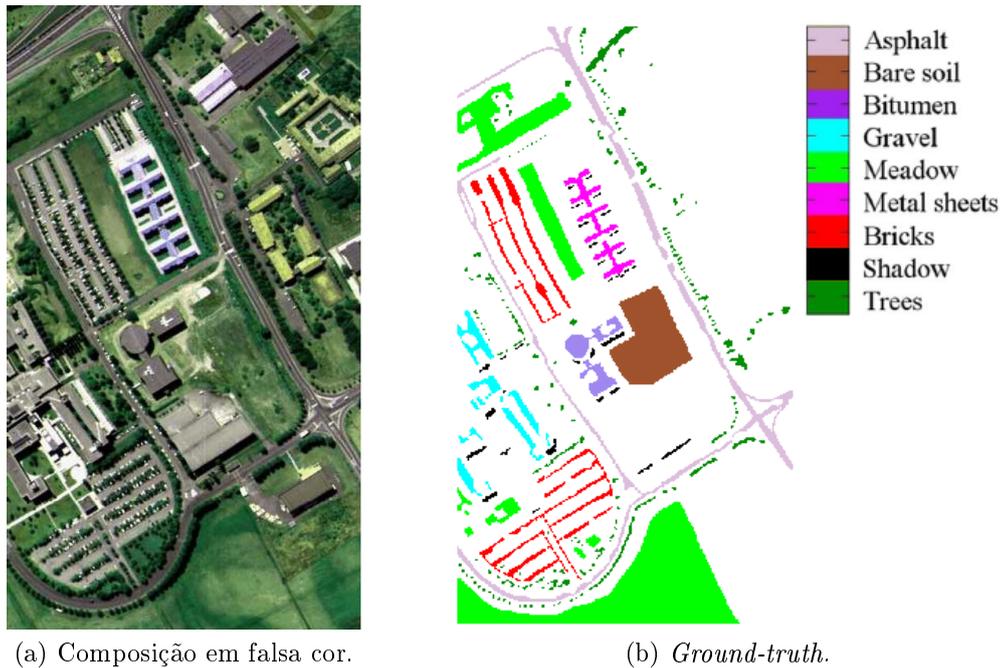


Figura 5.2: *ROSIS - Pavia University*.

## 5.2 Critérios de Avaliação

Para servir de referência, utilizamos os métodos tradicionais de combinação *MV* e *Average* descritos na Seção 3.2.1 do Capítulo 3. Como o tamanho do conjunto de treinamento é um fator importante na classificação de imagens de sensoriamento remoto, avaliamos a robustez dos métodos propostos em diferentes tamanhos de conjuntos de treinamento e diferentes cenários. Primeiramente iremos definir o protocolo de treino/teste utilizado para os experimentos, em seguida serão discutidos as medidas estatísticas usadas para avaliação e por fim descreveremos os cenários.

### 5.2.1 Protocolo de Treinamento e Teste

As abordagens de classificação base (*Pixelwise-SVM*, *FEGA-KNN*, *EMP-MLP* e *SpecEmpPLS-SVM*) foram treinadas em duas etapas. A primeira etapa consiste em treinar as abordagens de classificação base gerando modelos que irão ser usados para posterior comparação com os métodos de combinação propostos. Dessa forma, iremos identificar se algum método de combinação foi melhor, ou pior, que alguma abordagem de classificação individual. Ainda nessa etapa, serão produzidos os dados que efetivamente serão usados na combinação pelo método *WLC-GA*. Já a segunda etapa se concentra em gerar dados que irão ser usados para calibrar os pesos no *WLC-GA*,

e também, para treinar o método *Stacking-SVM*. Portanto, nessa segunda etapa, serão usados conjuntos de treino menores que aqueles usados na primeira etapa, e dessa maneira realizamos uma comparação justa.

Na primeira etapa, as abordagens de classificação base foram treinadas usando conjuntos de amostras de tamanhos 10%, 15%, 20% e 25% (“conjunto-treino-completo”), e em todos os casos as amostras foram coletadas aleatoriamente. Note que os dados gerados nessa primeira etapa não irão compor os dados usados no treinamento e calibração dos métodos de combinação, visto que já utilizamos 10%, 15%, 20% e 25% para treinamento. Utilizar conjuntos adicionais para calibrar e treinar os métodos de combinação propostos seria injusto. Além disso, não há dados hiperespectrais rotulados em abundância, e utilizar o mínimo possível de dados de treino é um objetivo comum entre os pesquisadores da área.

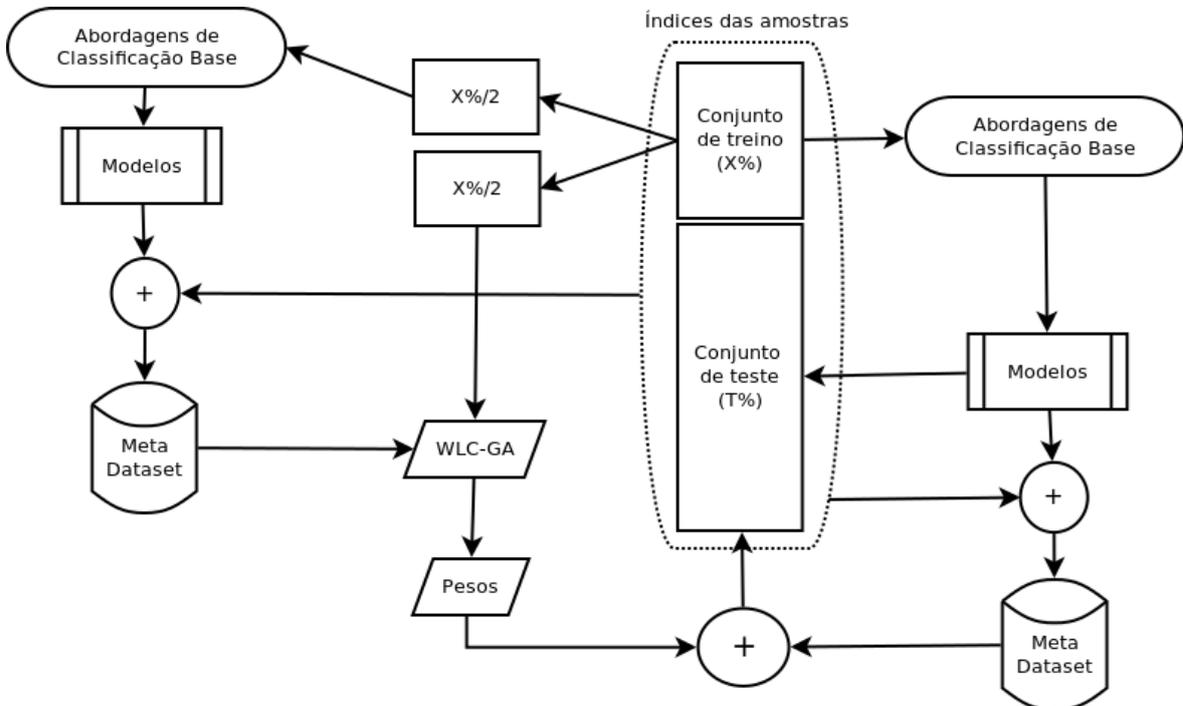


Figura 5.3: Exemplo esquemático do protocolo de treinamento e teste utilizado.

Para o treinamento das abordagens de combinação, isto é, *WLC-GA* e *Stacking-SVM*, ao invés de usarmos conjuntos de treino adicionais, dividimos cada um dos conjuntos, usados para treinamento das abordagens de classificação base na primeira etapa, em 50%. Assim, para cada conjunto-treino-completo (10%, 15%, 20% e 25%) temos dois novos conjuntos, e dessa forma obtemos configurações de treino de tamanhos iguais a 5%, 7,5%, 10% e 12,5%. Um dos conjuntos repartidos será usado para treinar novamente as abordagens de classificação base afim de produzir os dados, temporários,

para a combinação (probabilidades, graus de confiança, etc.). O outro conjunto será utilizado para treinar os métodos de combinação *WLC-GA* e *Stacking-SVM*. Por fim, utilizamos as probabilidades geradas pelas abordagens base treinadas nos conjunto-treino-completo para, efetivamente, combina-las e testar o *WLC-GA* como mostra a Figura 5.3. Dessa maneira, não utilizamos mais que 10%, 15%, 20% e 25% de dados para treinamento de todo o sistema de classificação. Os conjuntos de amostras restantes (90%, 85%, 80% e 75%) foram utilizados para teste.

## 5.2.2 Medidas de Avaliação

Para uma avaliação mais precisa executamos os experimentos 30 vezes para cada uma das configurações de treinamento/teste mencionadas anteriormente, e então medidas comuns utilizadas em classificação de padrões foram extraídas. Essas medidas são definidas como a seguir:

- **Matriz de Confusão:** é uma tabela onde cada coluna representa as amostras preditas em uma classe, enquanto que a linha representa a classe real. Cada elemento  $C_{ij}$  da matriz representa o número de amostras classificados como classe  $i$  e referenciados como classe  $j$ . Essa tabela permite uma fácil visualização do número de amostras classificadas erroneamente e nos permite extrair as medidas a seguir.
- **Acurácia global (*OA*):** é a porcentagem de amostras classificadas corretamente.

$$OA = \frac{\sum_i^K C_{ii}}{\sum_{ij} C_{ij}} \times 100$$

- **Acurácia por classe (*Class Accuracy (CA)*):** é a porcentagem de amostras classificadas corretamente para uma dada classe.

$$CA_i = \frac{C_{ii}}{\sum_j^K C_{ij}} \times 100$$

- **Acurácia média (*Average Accuracy (AA)*):** é média da *CA* para todas as classes.

$$AA = \frac{\sum_i^K CA_i}{K} \times 100$$

- ***F-score*:** muito utilizada na área de *Recuperação de Informação*, o *F-score* fornece uma melhor descrição, sobre o quanto o classificador aprendeu sobre uma

determinada classe, quando comparado a *CA*. Isso acontece pelo fato de que o *F-score* leva em consideração os erros falsos positivos (precisão) e falsos negativos (revocação), enquanto que a *CA* considera apenas os falsos positivos. Essa medida é definida como a média harmônica da precisão (*precision*) e revocação (*recall*):

$$F\text{-score} = \frac{2 \times \textit{precision} \times \textit{recall}}{\textit{precision} + \textit{recall}}$$

A precisão é o número de amostras relevantes retornadas dividido pelo número total de amostras retornadas, enquanto revocação é o número de amostras relevantes retornadas dividido pelo número total de amostras relevantes (Manning et al., 2008).

A partir dessas medidas foram calculados os intervalos de confiança (*Intervalo de Confiança (CI)*) para o *OA*, *AA* e *F-score*. Para 95% de nível de confiança o *CI* é calculado da seguinte forma:

$$CI = X \pm 1,96 \times \frac{\sigma}{\sqrt{n}}$$

em que  $X$  é a média, a constante 1,96 vem da função cumulativa da distribuição normal (Jain, 1991) e depende do nível de confiança adotado (no nosso caso corresponde a 95%).  $\sigma$  é o desvio padrão e  $n$  o número de observações realizadas.

Além disso, para verificar se há diferença estatística entre os métodos estudados, realizamos o teste de observações pareadas conhecido como *paired t-test* (Jain, 1991). Usamos um nível de confiança igual a 95%, onde as observações provenientes do *OA* foram utilizadas para o *paired t-test*. Dessa maneira podemos afirmar, com 95% de certeza, que um determinado método possui comportamento melhor que outro.

## 5.3 Configuração dos Parâmetros

### 5.3.1 Parâmetros das Abordagens de Classificação Base

Algumas considerações à respeito das representações de características utilizadas podem impactar no tamanho do vetor de características. Abaixo são listadas as representações de características utilizados e discutidos os principais parâmetros de cada uma delas:

- *Pixelwise*: Nesse tipo de representação não há nenhum parâmetro a ser ajustado, apenas os canais espectrais mais ruidosos foram removidos conforme descrito na

## Seção 5.1.

- *FEGA*: O GA foi executado 10 vezes para as duas bases de dados usadas. Ao final, as características selecionadas foram aquelas que apareciam em pelo menos 50% em todas as 10 execuções. Baseado em experimentos preliminares os parâmetros utilizados no GA foram os seguintes: número de genes iguais ao número de bandas espectrais, 100 gerações, população igual a 100 indivíduos, 80% e 0.9% para as taxas de cruzamento e probabilidade de mutação, respectivamente. Mecanismos de seleção baseados em *torneio* e *elitismo*. Pelo fato de ser um método supervisionado, 50 amostras de cada classe selecionadas aleatoriamente foram usadas no experimento. Somente para o *Indian Pines*, para aquelas classes que possuem menos de 100 amostras (*Alfalfa*, *Grass/pasture-mowed*, *Oats* e *Stone-steel towers*), utilizamos 50%. O número total de características dos subconjuntos são 95 e 34 para o *Indian Pines* e *Pavia University*, respectivamente. Maiores detalhes podem ser encontrados em Santos et al. (2012c), onde diversos tipos de classificadores foram aplicados e comparados os resultados com as abordagens *Pixelwise* e outra abordagem de seleção de características usando o método *SVDSS*.
- *EMP*: Com o objetivo de construir um *EMP* compacto, e ao mesmo tempo carregar informação espacial suficiente, que possa contribuir na melhoria da classificação, seguimos as sugestões tradicionais feita no trabalho de Benediktsson et al. (2005). Utilizamos um número de principais componentes (PCs), extraídas pelo método *PCA*, que represente pelo menos 95% da variabilidade da informação na imagem, o que corresponde a 2 PCs nas duas bases de dados usadas. Número de aberturas/fechamentos igual a 4, com EE circular de tamanho inicial igual a 4 (raio do círculo em *pixels*) e passo incremental de tamanho 2, logo, chegando a um tamanho máximo de 12 *pixels* de raio. Portanto, o número de características no *EMP* utilizando esses parâmetros é igual a 18.
- *SpecEmpPLS*: Como discutido na Seção 4.1.4, primeiramente é necessário construir um *EMP* sobrecarregado de informação espacial. Para isso utilizamos um número de PCs que some a 99% a variabilidade total da informação na imagem, correspondendo a 5 e 3 PCs nos *Indian Pines* e *Pavia University*, respectivamente. Os demais parâmetros para construção do *EMP* foram os mesmos usados anteriormente. Por se tratar de um método supervisionado, 5% das amostras foram usadas para a construção do modelo *PLS* para as duas bases de dados utilizadas. Tanto as características espectrais originais quanto as produzidas pelo

*EMP* foram expressas em termos de variáveis latentes produzidas pelo modelo *PLS*. Escolhemos um número de características que totalize pelo menos 99% da variabilidade total da informação extraída, tanto espectral quanto a espacial proveniente do *EMP*. Assim, esperamos capturar tanta informação quanto precisamos sem aumentar drasticamente o número de dimensões no vetor de características. É importante realçar que essas considerações também são recomendadas no trabalho de Fauvel et al. (2008), ao qual esta presente abordagem se baseia. No *Indian Pines* foram extraídas 16 características, à partir somente da informação espectral, e também 16 à partir da informação espacial, totalizando 32 características no vetor. Enquanto que no *Pavia University* foram 4 características obtidas através da informação espectral e 8 da espacial, totalizando 12 características.

Para que os dados estejam adequados, para os algoritmos de aprendizado, é necessário redimensionar (normalizar/padronizar) os valores das características, esse processo é conhecido como *feature scaling*. Isso se faz necessário para garantir que todas as características possuam uma mesma importância no processo de aprendizagem (Duda et al., 2000). Há várias maneiras de normalizar/padronizar as características dos dados (Jain, 1991; Duda et al., 2000). Uma maneira bastante usada consiste em deixar os dados com média igual a 0 (zero) e variância igual a 1 (um) (Jain, 1991), esse procedimento é, também, conhecido como *z-score standardization*. Em nossos experimentos, o método *z-score* foi aplicado em todas as representações de características.

Além dos parâmetros inerentes aos métodos de representação de características, existem ainda aqueles referentes aos algoritmos de aprendizado. Abaixo segue o procedimento para ajustar esses parâmetros nos diferentes algoritmos usados:

- *KNN*: O *KNN* possui apenas um único parâmetro, o valor de  $K$ . Para encontrar o melhor valor de  $K$  usamos o processo de treinamento/teste conhecido como *N-fold cross-validation* (Duda et al., 2000; Theodoridis & Koutroumbas, 2003; Bishop, 2006; Alpaydin, 2010) variando o parâmetro  $K$  dentro de uma faixa de valores preestabelecidos. A faixa de valores utilizada foi: 3, 5, 7, 9, 11, 13, 15, 17 e 19 em um esquema de *5-folds*. Esse procedimento é feito para cada uma das configurações de amostragem de treino discutidas anteriormente na Seção 5.2.1. Após encontrar o melhor parâmetro, o algoritmo é novamente treinado e avaliado utilizando o protocolo descrito na Seção 5.2.1.
- *MLP*: De acordo com Duda et al. (2000) a escolha do número de neurônios nas camadas escondidas deve-se levar em consideração a quantidade de amostras de

treino disponíveis. O número de amostras de treino deve ser maior que o número de características nos dados de entrada, caso contrário poderá levar a uma pobre generalização. Contudo, existem várias regras conhecidas como “*rules of thumb*” para determinar o número de neurônios nas camadas escondidas da *MLP*. Muitas delas sugerem definir algo entre os números de neurônios de entrada e saída, ou ainda, alguma relação entre esses dois. Pelo fato de que há amostras de treinamento suficientes, nas duas bases de dados usadas, seguimos uma regra simples: a raiz quadrada do número de características presente nos dados vezes o número de classes existentes. No casos das bases de dados usadas, achamos essa regra bem interessante pois mantém um bom comprometimento entre os números de neurônios de entrada e saída. Com relação ao número de camadas escondidas, uma *MLP* de 3 camadas, isto é, uma camada de entrada, uma escondida e uma de saída, pode construir qualquer limite de decisão (função, modelo, etc.) desde que forneça quantidade adequada de neurônios (Duda et al., 2000). Portanto, definimos o número de camadas escondidas igual a 1.

- *SVM*: O número de parâmetros do *SVM* varia conforme o *kernel* utilizado. Diante dos bons resultados, relatados na literatura técnica, obtidos pelo *SVM* ao utilizar *kernel RBF*, para classificar imagens hiperespectrais de SR, optamos por usá-lo também. Nesse algoritmo, o procedimento para ajustar  $C$  e  $\gamma$  foi, também, o *N-fold cross-validation* usando *5-folds*. Todavia, temos dois parâmetros, o que leva a um grande número de combinações possíveis se usado uma extensa faixa de valores para  $C$  e  $\gamma$ . Por isso, os parâmetros são, geralmente, incrementados em potências de 2, avaliando uma faixa de intervalos maiores em um tempo aceitável (Chang & Lin, 2011). Além disso, um algoritmo “guloso” para buscar esses parâmetros está disponível na biblioteca *LIBSVM* (Chang & Lin, 2011).

Essas considerações foram feitas em ambas as bases de dados utilizadas.

### 5.3.2 Parâmetros dos Métodos de Combinação

Como era de se esperar, os métodos de combinação também possuem seus respectivos parâmetros. No caso do *WLC-GA* os parâmetros dizem respeito aos aqueles usado em GA, enquanto que no *Stacking-SVM* os parâmetros dependem do *kernel* usado no *SVM*.

- *WLC-GA*: Foram utilizados 20 bits para a representação de cada peso, totalizando 80 genes no genoma de cada indivíduo. Número de gerações iguais a

100, população de 30 indivíduos, probabilidades de cruzamento e mutação iguais 80% e 0.9%, respectivamente. Seleção por torneio e elitismo.

- *Stacking-SVM*: Utilizamos o *kernel RBF* pelo fato desse *kernel* ter se mostrado superior em relação a outros bastante usados para a classificação de dados hiperespectrais (Camps-Valls & Bruzzone, 2005). Assim como nas abordagens de classificação base que usam o *SVM*, os parâmetros  $C$  e  $\gamma$  foram ajustados usando o procedimento de *N-fold cross-validation* com 5 *folds*. É importante observar que esses parâmetros são estimados no conjunto de treino.

## 5.4 Cenários

Com o objetivo de investigar mais profundamente o comportamento dos métodos de combinação, executamos os testes em três cenários diferentes. Cada cenário corresponde a um grupo de abordagens de classificação base que irão compor o conjunto de entrada para os métodos de combinação descritos na Seção 4.2. Abaixo são descritos as abordagens de classificação base usadas em cada um dos cenários:

1. Cenário 1: Neste cenário todas as quatro abordagens de classificação base descritas na Seção 4.1 foram utilizadas.
2. Cenário 2: Neste segundo cenário, retiramos a abordagem de classificação base que produz a maior taxa de acurácia. Apenas as abordagens de classificação *base Pixelwise-SVM*, *FEGA-KNN* e *EMP-MLP* foram usadas na combinação.
3. Cenário 3: Neste último cenário, retiramos a abordagem de classificação base que possui os menores índices de acurácia, restando então, *Pixelwise-SVM*, *EMP-MLP* e *SpecEmpPLS-SVM* para efetuar a combinação.

Avaliar os métodos de combinação nesses três cenários pode nos dar algumas intuições a respeito de quando usar tais métodos, ou seja, em que situações métodos de combinação podem trazer melhorias significativas na classificação. Por fim, definido o processo de experimentação e avaliação utilizados neste trabalho, podemos seguir com a análise dos resultados no capítulo seguinte.

# Capítulo 6

## Resultados e Discussões

Neste capítulo apresentamos e discutimos os resultados obtidos por cada uma das abordagens de classificação base apresentadas no Capítulo 4, bem como os resultados obtidos pelos métodos de combinação *WLC-GA* e *Stacking-SVM* propostos. Além disso, comparamos os métodos propostos com outros métodos consensuais não-treináveis (*Average rule* e *MV*). Para que este presente documento fique mais claro, e menos repetitivo, iremos omitir os mapas temáticos produzidos nos cenários 2 e 3, e também, as tabelas com os resultados para as amostragens de 15% 20% e 25% dos cenários 1, 2 e 3. Todavia, os resultados utilizando essas amostragens poderão ser visualizados nos gráficos.

### 6.1 Análise dos Resultados

Nas Tabelas a seguir mostramos o *OA*, *AA* e *F-score* obtidos pelas abordagens de classificação base e pelos métodos de combinação.

#### 6.1.1 Resultados para o *Indian Pines*

Nas Tabelas 6.1, 6.2 e 6.3, podemos perceber que o método proposto *WLC-GA* detém os maiores valores de acurácia em todos os cenários experimentados. Para uma melhor validação, apresentamos os mapas temáticos produzidos pelos métodos *WLC-GA* e *Stacking-SVM* nas Figuras 6.1d e 6.1c, respectivamente. Para efeito de comparação, o mapa temático produzido pela classificação usando apenas a informação espectral (*Pixelwise-SVM*) é mostrado na Figura 6.1b. O respectivo *Ground-truth* pode ser visualizado na Figura 6.1a. Esses mapas foram gerados pela execução do experimento

que mais se aproximou da média (entre as 30 execuções) reportada na Tabela 6.1, referente ao cenário 1.

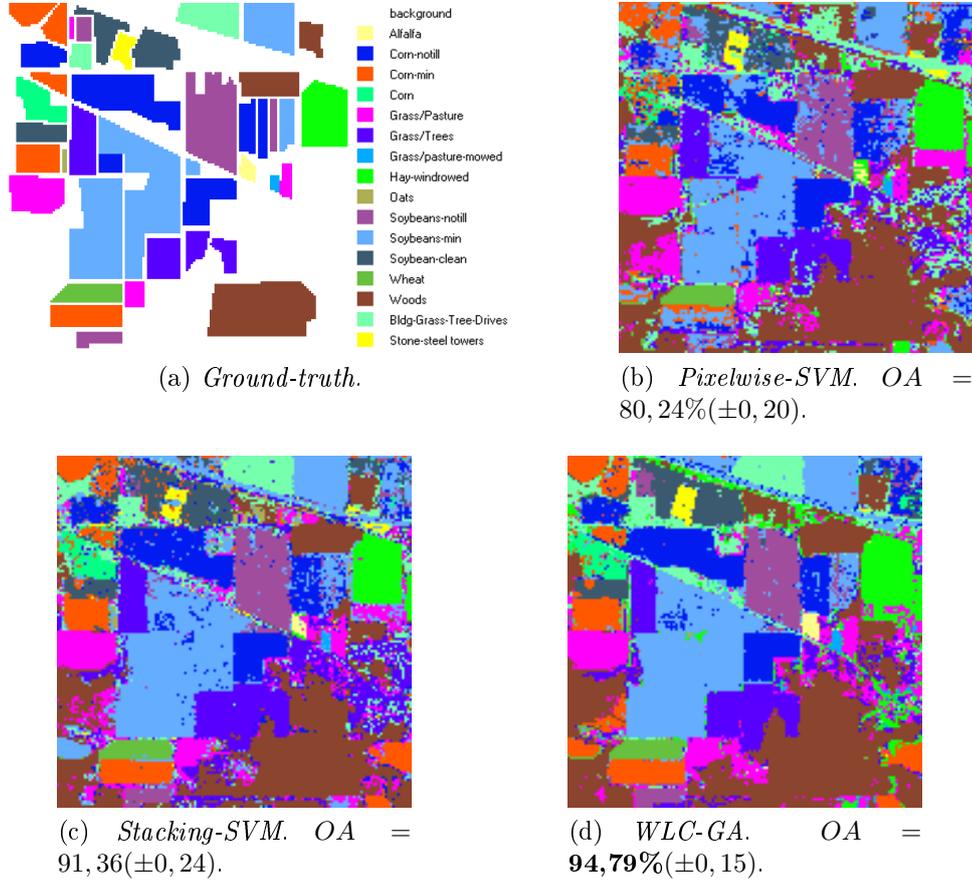


Figura 6.1: Comparação entre os mapas temáticos do *Indian Pines* produzidos pela classificação da informação espectral e pela combinação das quatro abordagens de classificação base (cenário 1) usando o *WLC-GA* e *Stacking-SVM*.

Além da comparação entre os métodos de combinação *MV*, *Average*, *WLC-GA* e *Stacking-SVM*, utilizamos também, como comparação, a melhor abordagem de classificação base usada no processo de combinação. Esses resultados podem ser visualizados nos Gráficos 6.2a, 6.2b e 6.2c. O símbolo asterisco (\*) indica a média do *OA* das 30 execuções e as barras verticais correspondem ao respectivo *CI*. Note que avaliamos os métodos *MV* e *Average* nos mesmos conjuntos de teste usados para o *WLC-GA* e *Stacking-SVM*.

No cenário 1, o *test-t pareado* indica que os métodos *Stacking-SVM* e *Average* são estatisticamente diferentes nos experimentos com conjuntos de treinamento iguais a 10% e 15%, sendo que no restante não há diferença estatística para um nível de confiança igual a 95%. Portanto, nesse cenário, o *Average* possui um melhor desem-

## 6. RESULTADOS E DISCUSSÕES

penho, em termos de acurácia, que o *Stacking-SVM* usando 10% e 15% de amostras de treino. A abordagem de classificação base *SpecEmpPLS-SVM* se saiu melhor que o *MV*, *Stacking-SVM* e *Average* em todas as configurações de treinamento, todavia, conseguimos melhorar ainda mais a classificação por meio do método *WLC-GA*. Apesar dos resultados entre o *SpecEmpPLS-SVM* e *WLC-GA* estarem bem próximos nesse cenário, o *test-t pareado* indica diferença estatística, em todas as configurações de treinamento, para um nível de confiança de 95%. Essa diferença estatística também pode ser observada pelos intervalos de confiança (*CI*) que não se sobrepõem.

No cenário 2 removemos a abordagem *SpecEmpPLS-SVM*, restando apenas três abordagens para a combinação. Nesse cenário, investigamos o comportamento dos métodos de combinação ao remover a melhor abordagem de classificação base. O gráfico da Figura 6.2b mostra que nesse cenário os índices de acurácias são menores que no cenário 1. Porém, é importante realçar que os métodos de combinação proporcionaram um melhor ganho na acurácia quando comparados às abordagens de classificação base isoladamente. Apesar dos ganhos obtidos pelo *Stacking-SVM* nesse cenário, o método *WLC-GA* ainda continua com os melhores índices de acurácia. Novamente, o *MV* foi método com os menores índices de acurácia dentre os métodos de combinação. Ainda, como pode ser observado pela Tabela 6.2 e do gráfico da Figura 6.2b, o *MV* não garante ser mais preciso que a melhor abordagem de classificação base usada na combinação. Esse comportamento do *MV* pode ser verificado também nos outros cenários (Tabelas 6.2a e 6.2c). Observe que a abordagem *EMP-MLP* obteve acurácia máxima usando um conjunto de treino de 20% e caiu em 25%. Isso foi observado também na segunda base de dados utilizada.

Ao remover a abordagem de classificação base com os mais baixos valores de *OA* (*FEGA-KNN*), os resultados dos métodos de combinação se mostram similares a aqueles obtidos no cenário 1. Isso indica que, nesse cenário, a abordagem *FEGA-KNN* não contribui de maneira significativa na combinação pelo *WLC-GA*. No entanto, comparando o cenário 1 e 3 ao utilizar o método *Stacking-SVM* na combinação, concluímos que o *Stacking-SVM* consegue capturar mais informações fornecidas pelo *FEGA-KNN* que as outras abordagens de combinação, pois ao remover o *FEGA-KNN* da combinação o *Stacking-SVM* sofre uma redução nos índices de acurácia. Observamos ainda que novamente o *WLC-GA* obteve os melhores índices de acurácia dentre os métodos de combinação utilizados.

Tabela 6.1: Resultados do cenário 1 para o *Indian Pines* usando 10% de dados para treinamento em 30 execuções.

	num. de amostras	amostras de treino (%)	Abordagens de Classificação Base				Métodos de Combinação			
			<i>Pixelwise SVM</i> (200 features)	<i>FEGA KNN</i> (95 features)	<i>EMP MLP</i> (18 features)	<i>SpecEmpPLS SVM</i> (16+16 features)	<i>MV</i>	<i>Aveg.</i>	<i>WLC-GA</i>	<i>Stacking SVM</i> (64 meta-features)
<b>OA (%)</b>			80, 24(±0, 20)	61, 24(±0, 36)	84, 60(±0, 78)	93, 95(±0, 17)	88, 77(±0, 23)	91, 82(±0, 22)	<b>94,79(±0, 15)</b>	91, 36(±0, 24)
<b>AA (%)</b>			67, 56(±0, 72)	50, 03(±0, 87)	80, 42(±1, 00)	85, 82(±0, 90)	79, 42(±0, 70)	83, 40(±0, 69)	<b>89,62(±0, 61)</b>	78, 52(±0, 81)
<b>Classes</b>			<b>F-score</b>							
<b>Alfalfa</b>	54	10	0, 22(±0, 08)	0, 21(±0, 06)	0, 72(±0, 08)	<b>0,93(±0, 01)</b>	0, 81(±0, 06)	0, 75(±0, 06)	0, 92(±0, 01)	0, 06(±0, 07)
<b>Corn-notill</b>	1434	10	0, 77(±0, 00)	0, 50(±0, 01)	0, 78(±0, 03)	0, 90(±0, 00)	0, 84(±0, 01)	0, 89(±0, 01)	<b>0,92(±0, 00)</b>	0, 88(±0, 01)
<b>Corn-min</b>	834	10	0, 68(±0, 01)	0, 41(±0, 01)	0, 77(±0, 01)	0, 93(±0, 01)	0, 87(±0, 01)	0, 91(±0, 01)	<b>0,94(±0, 00)</b>	0, 89(±0, 01)
<b>Corn</b>	234	10	0, 53(±0, 04)	0, 40(±0, 03)	0, 53(±0, 05)	0, 86(±0, 01)	0, 69(±0, 03)	0, 80(±0, 02)	<b>0,88(±0, 01)</b>	0, 78(±0, 02)
<b>Grass/Pasture</b>	497	10	0, 89(±0, 01)	0, 65(±0, 02)	0, 83(±0, 02)	0, 95(±0, 00)	0, 93(±0, 00)	0, 93(±0, 00)	<b>0,96(±0, 00)</b>	0, 90(±0, 01)
<b>Grass/Trees</b>	747	10	0, 93(±0, 00)	0, 67(±0, 01)	0, 90(±0, 01)	<b>0,98(±0, 00)</b>	0, 95(±0, 00)	0, 95(±0, 00)	<b>0,98(±0, 00)</b>	0, 96(±0, 01)
<b>Grass/pasture-mowed</b>	26	10	0, 54(±0, 08)	0, 17(±0, 06)	0, 79(±0, 06)	0, 67(±0, 12)	0, 72(±0, 08)	0, 74(±0, 05)	<b>0,89(±0, 03)</b>	0, 63(±0, 09)
<b>Hay-windrowed</b>	489	10	0, 93(±0, 00)	0, 91(±0, 00)	0, 91(±0, 02)	<b>1,00(±0, 00)</b>	0, 97(±0, 00)	0, 96(±0, 00)	0, 99(±0, 00)	0, 98(±0, 01)
<b>Oats</b>	20	10	0, 02(±0, 02)	0, 01(±0, 01)	0, 38(±0, 06)	0, 12(±0, 06)	0, 12(±0, 04)	0, 28(±0, 06)	<b>0,45(±0, 07)</b>	0, 32(±0, 07)
<b>Soybeans-notill</b>	968	10	0, 73(±0, 01)	0, 53(±0, 01)	0, 77(±0, 01)	0, 89(±0, 00)	0, 85(±0, 01)	0, 87(±0, 01)	<b>0,90(±0, 00)</b>	0, 85(±0, 01)
<b>Soybeans-min</b>	2468	10	0, 78(±0, 00)	0, 64(±0, 00)	0, 87(±0, 01)	0, 94(±0, 00)	0, 89(±0, 00)	0, 92(±0, 00)	<b>0,95(±0, 00)</b>	0, 92(±0, 00)
<b>Soybean-clean</b>	614	10	0, 73(±0, 01)	0, 31(±0, 01)	0, 73(±0, 04)	0, 91(±0, 01)	0, 83(±0, 01)	0, 89(±0, 01)	<b>0,93(±0, 01)</b>	0, 88(±0, 01)
<b>Wheat</b>	212	10	0, 96(±0, 01)	0, 81(±0, 01)	0, 99(±0, 00)	<b>1,00(±0, 00)</b>	0, 98(±0, 00)	0, 99(±0, 00)	<b>1,00(±0, 00)</b>	0, 99(±0, 00)
<b>Woods</b>	1294	10	0, 94(±0, 00)	0, 82(±0, 00)	0, 98(±0, 00)	<b>0,99(±0, 00)</b>	0, 96(±0, 00)	0, 98(±0, 00)	<b>0,99(±0, 00)</b>	<b>0,99(±0, 00)</b>
<b>Bldg-Grass-Trees-Drives</b>	380	10	0, 67(±0, 01)	0, 24(±0, 02)	0, 94(±0, 01)	0, 95(±0, 01)	0, 79(±0, 01)	0, 93(±0, 01)	<b>0,96(±0, 01)</b>	0, 94(±0, 01)
<b>Stone-steel towers</b>	95	10	0, 83(±0, 02)	0, 83(±0, 01)	0, 91(±0, 03)	0, 92(±0, 02)	0, 90(±0, 01)	0, 92(±0, 01)	<b>0,93(±0, 01)</b>	0, 88(±0, 03)

Tabela 6.2: Resultados do cenário 2 para o *Indian Pines* usando 10% de dados para treinamento em 30 execuções.

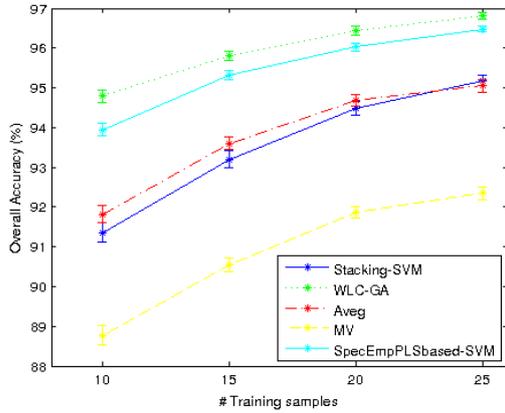
	num. de amostras	amostras de treino (%)	Abordagens de Classificação Base			Métodos de Combinação			
			<i>Pixelwise SVM</i> (200 features)	<i>FEGA KNN</i> (95 features)	<i>EMP MLP</i> (18 features)	<i>MV</i>	<i>Aveg.</i>	<i>WLC-GA</i>	<i>Stacking SVM</i> (48 meta-features)
<b>OA (%)</b>			80, 24(±0, 20)	61, 24(±0, 36)	84, 60(±0, 78)	80, 82(±0, 35)	85, 16(±0, 30)	<b>88,90(±0, 41)</b>	86, 03(±0, 45)
<b>AA (%)</b>			67, 56(±0, 72)	50, 03(±0, 87)	80, 42(±1, 00)	68, 60(±0, 88)	76, 10(±0, 67)	<b>82,85(±0, 86)</b>	73, 38(±1, 04)
<b>Classes</b>			<b>F-score</b>						
<b>Alfalfa</b>	54	10	0, 22(±0, 08)	0, 21(±0, 06)	0, 72(±0, 08)	0, 34(±0, 07)	0, 59(±0, 07)	<b>0,73(±0, 07)</b>	0, 06(±0, 07)
<b>Corn-notill</b>	1434	10	0, 77(±0, 00)	0, 50(±0, 01)	0, 78(±0, 03)	0, 72(±0, 01)	0, 80(±0, 01)	<b>0,85(±0, 01)</b>	0, 82(±0, 01)
<b>Corn-min</b>	834	10	0, 68(±0, 01)	0, 41(±0, 01)	0, 77(±0, 01)	0, 69(±0, 01)	0, 79(±0, 01)	<b>0,84(±0, 01)</b>	0, 80(±0, 01)
<b>Corn</b>	234	10	0, 53(±0, 04)	0, 40(±0, 03)	0, 53(±0, 05)	0, 51(±0, 04)	0, 65(±0, 03)	<b>0,67(±0, 04)</b>	0, 57(±0, 04)
<b>Grass/Pasture</b>	497	10	0, 89(±0, 01)	0, 65(±0, 02)	0, 83(±0, 02)	0, 87(±0, 01)	0, 89(±0, 01)	<b>0,91(±0, 01)</b>	0, 86(±0, 01)
<b>Grass/Trees</b>	747	10	0, 93(±0, 00)	0, 67(±0, 01)	0, 90(±0, 01)	0, 89(±0, 01)	0, 92(±0, 01)	<b>0,94(±0, 00)</b>	0, 93(±0, 01)
<b>Grass/pasture-mowed</b>	26	10	0, 54(±0, 08)	0, 17(±0, 06)	0, 79(±0, 06)	0, 52(±0, 08)	0, 71(±0, 06)	<b>0,82(±0, 04)</b>	0, 62(±0, 09)
<b>Hay-windrowed</b>	489	10	0, 93(±0, 00)	0, 91(±0, 00)	0, 91(±0, 02)	0, 93(±0, 00)	0, 95(±0, 00)	<b>0,96(±0, 00)</b>	0, 95(±0, 01)
<b>Oats</b>	20	10	0, 02(±0, 02)	0, 01(±0, 01)	0, 38(±0, 06)	0, 15(±0, 05)	0, 26(±0, 06)	<b>0,39(±0, 06)</b>	0, 30(±0, 07)
<b>Soybeans-notill</b>	968	10	0, 73(±0, 01)	0, 53(±0, 01)	0, 77(±0, 01)	0, 74(±0, 01)	0, 79(±0, 01)	<b>0,82(±0, 01)</b>	0, 78(±0, 01)
<b>Soybeans-min</b>	2468	10	0, 78(±0, 00)	0, 64(±0, 00)	0, 87(±0, 01)	0, 82(±0, 00)	0, 84(±0, 00)	<b>0,89(±0, 00)</b>	0, 87(±0, 01)
<b>Soybean-clean</b>	614	10	0, 73(±0, 01)	0, 31(±0, 01)	0, 73(±0, 04)	0, 74(±0, 02)	0, 79(±0, 02)	<b>0,84(±0, 02)</b>	0, 77(±0, 02)
<b>Wheat</b>	212	10	0, 96(±0, 01)	0, 81(±0, 01)	<b>0,99(±0, 00)</b>	0, 97(±0, 00)	0, 98(±0, 00)	<b>0,99(±0, 00)</b>	<b>0,99(±0, 00)</b>
<b>Woods</b>	1294	10	0, 94(±0, 00)	0, 82(±0, 00)	<b>0,98(±0, 00)</b>	0, 95(±0, 00)	0, 95(±0, 00)	<b>0,98(±0, 00)</b>	<b>0,98(±0, 00)</b>
<b>Bldg-Grass-Trees-Drives</b>	380	10	0, 67(±0, 01)	0, 24(±0, 02)	0, 94(±0, 01)	0, 74(±0, 01)	0, 84(±0, 01)	0, 93(±0, 01)	0, 92(±0, 01)
<b>Stone-steel towers</b>	95	10	0, 83(±0, 02)	0, 83(±0, 01)	0, 91(±0, 03)	0, 87(±0, 01)	0, 90(±0, 01)	<b>0,93(±0, 02)</b>	0, 87(±0, 03)

Tabela 6.3: Resultados do cenário 3 para o *Indian Pines* usando 10% de dados para treinamento em 30 execuções.

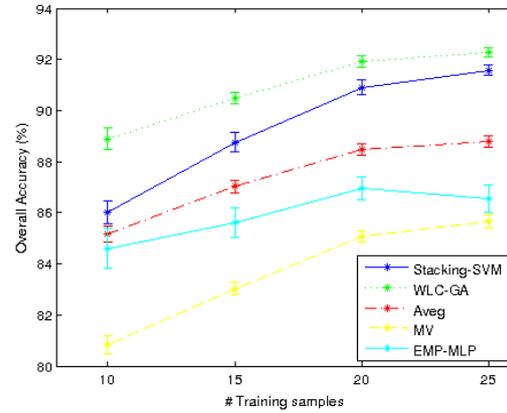
	num. de amostras	amostras de treino (%)	Abordagens de Classificação Base			Métodos de Combinação			
			<i>Pixelwise SVM</i> (200 features)	<i>EMP MLP</i> (18 features)	<i>SpecEmpPLS SVM</i> (16+16 features)	<i>MV</i>	<i>Aveg.</i>	<i>WLC-GA</i>	<i>Stacking SVM</i> (48 meta-features)
<b>OA (%)</b>			80,24(±0,20)	84,60(±0,78)	93,95(±0,17)	92,09(±0,22)	94,11(±0,17)	<b>94,82(±0,16)</b>	91,70(±0,29)
<b>AA (%)</b>			67,56(±0,72)	80,42(±1,00)	85,82(±0,90)	83,80(±0,80)	88,58(±0,74)	<b>89,68(±0,57)</b>	79,08(±1,20)
<b>Classes</b>			<b>F-score</b>						
<b>Alfalfa</b>	54	10	0,22(±0,08)	0,72(±0,08)	<b>0,93(±0,01)</b>	0,84(±0,05)	0,84(±0,06)	<b>0,93(±0,01)</b>	0,12(±0,10)
<b>Corn-notill</b>	1434	10	0,77(±0,00)	0,78(±0,03)	0,90(±0,00)	0,87(±0,00)	<b>0,92(±0,00)</b>	<b>0,92(±0,00)</b>	0,89(±0,01)
<b>Corn-min</b>	834	10	0,68(±0,01)	0,77(±0,01)	0,93(±0,01)	0,88(±0,01)	0,93(±0,01)	<b>0,94(±0,00)</b>	0,90(±0,01)
<b>Corn</b>	234	10	0,53(±0,04)	0,53(±0,05)	0,86(±0,01)	0,71(±0,03)	0,85(±0,01)	<b>0,88(±0,01)</b>	0,80(±0,02)
<b>Grass/Pasture</b>	497	10	0,89(±0,01)	0,83(±0,02)	0,95(±0,00)	0,93(±0,00)	0,95(±0,00)	<b>0,96(±0,00)</b>	0,91(±0,01)
<b>Grass/Trees</b>	747	10	0,93(±0,00)	0,90(±0,01)	<b>0,98(±0,00)</b>	0,97(±0,00)	0,97(±0,00)	<b>0,98(±0,00)</b>	0,97(±0,01)
<b>Grass/pasture-mowed</b>	26	10	0,54(±0,08)	0,79(±0,06)	0,67(±0,12)	0,74(±0,08)	0,87(±0,03)	<b>0,89(±0,03)</b>	0,57(±0,11)
<b>Hay-windrowed</b>	489	10	0,93(±0,00)	0,91(±0,02)	<b>1,00(±0,00)</b>	0,98(±0,00)	0,98(±0,00)	0,99(±0,00)	0,98(±0,01)
<b>Oats</b>	20	10	0,02(±0,02)	0,38(±0,06)	0,12(±0,06)	0,16(±0,05)	0,44(±0,06)	<b>0,45(±0,06)</b>	0,35(±0,08)
<b>Soybeans-notill</b>	968	10	0,73(±0,01)	0,77(±0,01)	0,89(±0,00)	0,87(±0,01)	<b>0,90(±0,01)</b>	<b>0,90(±0,01)</b>	0,86(±0,01)
<b>Soybeans-min</b>	2468	10	0,78(±0,00)	0,87(±0,01)	0,94(±0,00)	0,93(±0,00)	0,94(±0,00)	<b>0,95(±0,00)</b>	0,92(±0,01)
<b>Soybean-clean</b>	614	10	0,73(±0,01)	0,73(±0,04)	0,91(±0,01)	0,91(±0,01)	0,92(±0,00)	<b>0,93(±0,01)</b>	0,89(±0,01)
<b>Wheat</b>	212	10	0,96(±0,01)	0,99(±0,00)	<b>1,00(±0,00)</b>	<b>1,00(±0,00)</b>	<b>1,00(±0,00)</b>	<b>1,00(±0,00)</b>	0,99(±0,00)
<b>Woods</b>	1294	10	0,94(±0,00)	0,98(±0,00)	<b>0,99(±0,00)</b>	<b>0,99(±0,00)</b>	<b>0,99(±0,00)</b>	<b>0,99(±0,00)</b>	<b>0,99(±0,00)</b>
<b>Bldg-Grass-Trees-Drives</b>	380	10	0,67(±0,01)	0,94(±0,01)	0,95(±0,01)	<b>0,96(±0,00)</b>	<b>0,96(±0,01)</b>	<b>0,96(±0,01)</b>	0,93(±0,01)
<b>Stone-steel towers</b>	95	10	0,83(±0,02)	0,91(±0,03)	0,92(±0,02)	0,92(±0,01)	<b>0,94(±0,01)</b>	0,93(±0,01)	0,89(±0,03)

Tabela 6.4: Resultados do cenário 1 para o *Pavia University* usando 10% de dados para treinamento em 30 execuções.

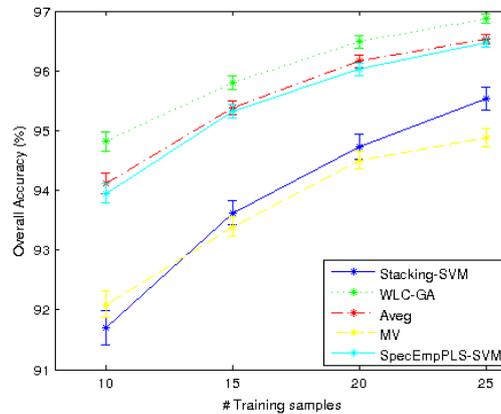
	num. de amostras	amostras de treino (%)	Abordagens de Classificação Base				Métodos de Combinação			
			<i>Pixelwise SVM</i> (103 features)	<i>FEGA KNN</i> (34 features)	<i>EMP MLP</i> (18 features)	<i>SpecEmpPLS SVM</i> (4+8 features)	<i>MV</i>	<i>Aveg.</i>	<i>WLC-GA</i>	<i>Stacking SVM</i> (36 meta-features)
<b>OA (%)</b>			93,12(±0,07)	88,32(±0,08)	95,73(±0,25)	98,94(±0,03)	96,67(±0,07)	98,33(±0,04)	<b>99,03(±0,03)</b>	98,69(±0,06)
<b>AA (%)</b>			90,54(±0,12)	85,53(±0,13)	93,67(±0,39)	98,18(±0,06)	94,85(±0,10)	97,01(±0,06)	<b>98,22(±0,07)</b>	97,63(±0,12)
<b>Classes</b>			<b>F-score</b>							
<b>Asphalt</b>	6631	10	0,94(±0,00)	0,89(±0,00)	0,97(±0,00)	<b>0,99(±0,00)</b>	0,98(±0,00)	<b>0,99(±0,00)</b>	<b>0,99(±0,00)</b>	0,98(±0,00)
<b>Meadow</b>	18649	10	0,97(±0,00)	0,94(±0,00)	0,98(±0,00)	<b>1,00(±0,00)</b>	0,98(±0,00)	<b>1,00(±0,00)</b>	<b>1,00(±0,00)</b>	<b>1,00(±0,00)</b>
<b>Gravel</b>	2099	10	0,78(±0,00)	0,69(±0,00)	0,80(±0,02)	<b>0,95(±0,00)</b>	0,89(±0,00)	0,91(±0,00)	<b>0,95(±0,00)</b>	<b>0,95(±0,00)</b>
<b>Trees</b>	3064	10	0,94(±0,00)	0,88(±0,00)	0,97(±0,00)	<b>0,99(±0,00)</b>	0,96(±0,00)	<b>0,99(±0,00)</b>	<b>0,99(±0,00)</b>	<b>0,99(±0,00)</b>
<b>Metal Sheets</b>	1345	10	0,99(±0,00)	0,99(±0,00)	0,98(±0,01)	<b>1,00(±0,00)</b>	<b>1,00(±0,00)</b>	<b>1,00(±0,00)</b>	<b>1,00(±0,00)</b>	<b>1,00(±0,00)</b>
<b>Bare Soil</b>	5029	10	0,90(±0,00)	0,79(±0,00)	0,96(±0,01)	<b>1,00(±0,00)</b>	0,95(±0,00)	0,99(±0,00)	<b>1,00(±0,00)</b>	0,99(±0,00)
<b>Bitumen</b>	1330	10	0,85(±0,00)	0,77(±0,00)	0,86(±0,01)	<b>0,96(±0,00)</b>	0,94(±0,00)	0,95(±0,00)	<b>0,96(±0,00)</b>	0,94(±0,00)
<b>Bricks</b>	3682	10	0,85(±0,00)	0,78(±0,00)	0,90(±0,01)	<b>0,97(±0,00)</b>	0,93(±0,00)	0,95(±0,00)	<b>0,97(±0,00)</b>	<b>0,97(±0,00)</b>
<b>Shadow</b>	947	10	0,99(±0,00)	0,99(±0,00)	0,98(±0,01)	<b>1,00(±0,00)</b>	<b>1,00(±0,00)</b>	<b>1,00(±0,00)</b>	<b>1,00(±0,00)</b>	<b>1,00(±0,00)</b>



(a) Resultados para o cenário 1.



(b) Resultados para o cenário 2.



(c) Resultados para o cenário 3.

Figura 6.2: Resultados para o *Indian Pines* em diversas configurações de amostragens. O símbolo asterisco (\*) indica a média do OA das 30 execuções e as barras verticais correspondem ao respectivo CI.

### 6.1.2 Resultados para o *Pavia University*

Observando o gráfico referente ao cenário 1 (Figura 6.4a) e a Tabela 6.4, podemos perceber que a abordagem *SpecEmpPLS-SVM* possui resultados similares ao *WLC-GA* para o *Pavia University*. No entanto, o *test-t pareado* revela diferença estatística para um nível de confiança igual a 95%, logo, podemos concluir que o *WLC-GA* possui um desempenho superior ao *SpecEmpPLS-SVM*. O *Stacking-SVM* demonstrou ser melhor que *Average* e *MV*. Os mapas temáticos podem ser visualizados na Figura 6.3.

Assim como na base de dados anterior, ao remover o *FEGA-KNN* no cenário 2, os ganhos obtidos pelos métodos de combinação se tornam mais significantes em relação à melhor abordagem base usada na combinação (Tabela 6.5). Nesse cenário os métodos *WLC-GA* e *Stacking-SVM* não são estatisticamente diferentes em quase todas as configurações de treino/teste. Apenas em 25% há a diferença, e portanto, nesse cenário o *Stacking-SVM* se mostra ligeiramente superior ao *WLC-GA* em termos

de acurácia. É importante realçar que ao analisar o cenário 3 (Tabela 6.6), percebemos novamente como o *Stacking-SVM* sofre com a retirada do *FEGA-KNN* na combinação, ficando abaixo dos resultados obtidos pelo *Average* nas amostragens de 10% e 15%. Em todos os cenários o método *MV* obteve os menores índices de acurácia dentre os métodos de combinação experimentados, reportando inclusive acurácias inferiores à melhor abordagem de classificação base (*SpecEmpPLS-SVM*).

O *Pavia University* possui grande quantidade de amostras para cada uma das 9 classes existentes e não sofre tanto com o problema de desbalanceamento de classes como o *Indian Pines*. Por isso, podemos ver que a maioria dos métodos possuem *OA* superiores a 90% para essa base de dados, apenas o *FEGA-KNN* possui valores um pouco abaixo ( $OA = 88,32(\pm 0,08)$ ).

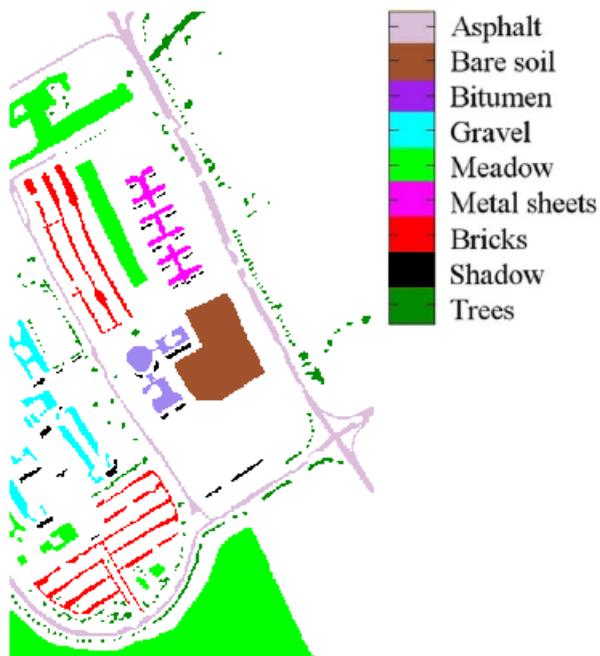
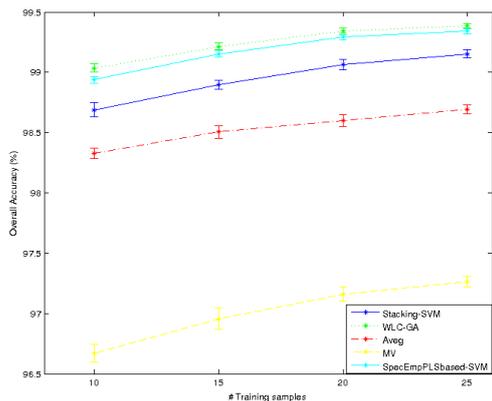
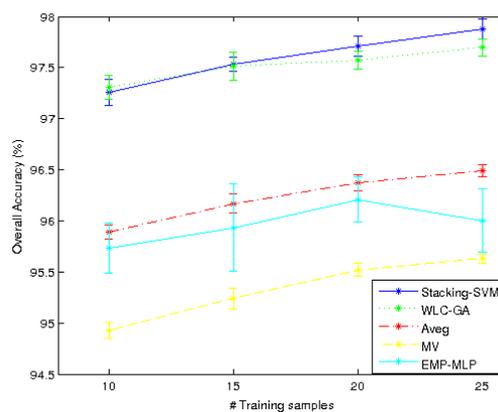
(a) *Ground-truth.*(b) Pixelwise-SVM.  $OA = 93,12\%(\pm 0,07)$ .(c) Stacking-SVM.  $OA = 98,69(\pm 0,06)$ .(d) WLC-GA.  $OA = 99,03\%(\pm 0,03)$ .

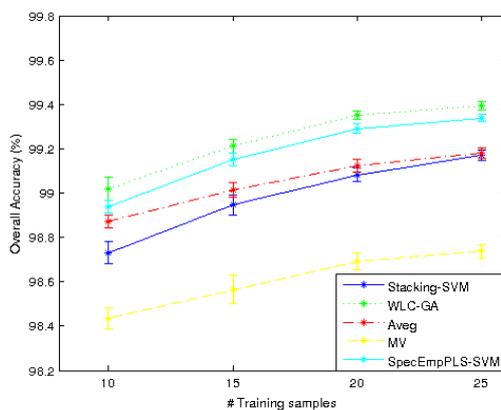
Figura 6.3: Comparação entre os mapas temáticos do *Pavia University* produzidos pela classificação da informação espectral e pela combinação das quatro abordagens de classificação base (cenário 1) usando o *WLC-GA* e *Stacking-SVM*.



(a) Resultados para o cenário 1.



(b) Resultados para o cenário 2.



(c) Resultados para o cenário 3.

Figura 6.4: Resultados para o *Pavia University* em diversas configurações de amostragens. O símbolo asterisco (\*) indica a média do OA das 30 execuções e as barras verticais correspondem ao respectivo CI.

Tabela 6.5: Resultados do cenário 2 para o *Pavia University* usando 10% de dados para treinamento em 30 execuções.

	num. de amostras	amostras de treino (%)	Abordagens de Classificação Base			Métodos de Combinação			
			<i>Pixelwise SVM</i> (103 features)	<i>FEGA KNN</i> (34 features)	<i>EMP MLP</i> (4+8 features)	<i>MV</i>	<i>Aveg.</i>	<i>WLC-GA</i>	<i>Stacking SVM</i> (36 meta-features)
<b>OA (%)</b>			93, 12(±0, 07)	88, 32(±0, 08)	95, 73(±0, 25)	94, 93(±0, 08)	95, 89(±0, 07)	<b>97, 31(±0, 12)</b>	97, 26(±0, 13)
<b>AA (%)</b>			90, 54(±0, 12)	85, 53(±0, 13)	93, 67(±0, 39)	92, 63(±0, 11)	93, 86(±0, 09)	95, 62(±0, 17)	<b>95, 67(±0, 21)</b>
<b>Classes</b>			<b>F-score</b>						
<b>Asphalt</b>	6631	10	0, 94(±0, 00)	0, 89(±0, 00)	0, 97(±0, 00)	0, 95(±0, 00)	0, 97(±0, 00)	<b>0, 98(±0, 00)</b>	<b>0, 98(±0, 00)</b>
<b>Meadow</b>	18649	10	0, 97(±0, 00)	0, 94(±0, 00)	0, 98(±0, 00)	0, 98(±0, 00)	0, 98(±0, 00)	<b>0, 99(±0, 00)</b>	<b>0, 99(±0, 00)</b>
<b>Gravel</b>	2099	10	0, 78(±0, 00)	0, 69(±0, 00)	0, 80(±0, 02)	0, 82(±0, 00)	0, 83(±0, 00)	0, 86(±0, 01)	<b>0, 89(±0, 01)</b>
<b>Trees</b>	3064	10	0, 94(±0, 00)	0, 88(±0, 00)	0, 97(±0, 00)	0, 96(±0, 00)	0, 97(±0, 00)	<b>0, 99(±0, 00)</b>	0, 98(±0, 00)
<b>Metal Sheets</b>	1345	10	0, 99(±0, 00)	0, 99(±0, 00)	0, 98(±0, 01)	0, 99(±0, 00)	0, 99(±0, 00)	<b>1, 00(±0, 00)</b>	<b>1, 00(±0, 00)</b>
<b>Bare Soil</b>	5029	10	0, 90(±0, 00)	0, 79(±0, 00)	0, 96(±0, 01)	0, 93(±0, 00)	0, 94(±0, 00)	<b>0, 97(±0, 00)</b>	<b>0, 97(±0, 00)</b>
<b>Bitumen</b>	1330	10	0, 85(±0, 00)	0, 77(±0, 00)	0, 86(±0, 01)	0, 91(±0, 00)	<b>0, 92(±0, 00)</b>	<b>0, 92(±0, 00)</b>	<b>0, 92(±0, 01)</b>
<b>Bricks</b>	3682	10	0, 85(±0, 00)	0, 78(±0, 00)	0, 90(±0, 01)	0, 87(±0, 00)	0, 89(±0, 00)	0, 92(±0, 00)	<b>0, 93(±0, 00)</b>
<b>Shadow</b>	947	10	0, 99(±0, 00)	0, 99(±0, 00)	0, 98(±0, 01)	<b>1, 00(±0, 00)</b>	0, 99(±0, 00)	<b>1, 00(±0, 00)</b>	<b>1, 00(±0, 00)</b>

Tabela 6.6: Resultados do cenário 3 para o *Pavia University* usando 10% de dados para treinamento em 30 execuções.

	num. total de amostras	amostras de treino (%)	Abordagens de Classificação Base			Métodos de Combinação			
			<i>Pixelwise SVM</i> (103 features)	<i>EMP MLP</i> (18 features)	<i>SpecEmpPLS SVM</i> (4+8 features)	<i>MV</i>	<i>Aveg.</i>	<i>WLC-GA</i>	<i>Stacking SVM</i> (36 meta-features)
<b>OA (%)</b>			93, 12(±0, 07)	95, 73(±0, 25)	98, 94(±0, 03)	98, 43(±0, 05)	98, 87(±0, 03)	<b>99, 02(±0, 05)</b>	98, 73(±0, 05)
<b>AA (%)</b>			90, 54(±0, 12)	93, 67(±0, 39)	98, 18(±0, 06)	97, 13(±0, 09)	97, 93(±0, 05)	<b>98, 23(±0, 07)</b>	97, 70(±0, 10)
<b>Classes</b>			<b>F-score</b>						
<b>Asphalt</b>	6631	10	0, 94(±0, 00)	0, 97(±0, 00)	<b>0, 99(±0, 00)</b>	0, 98(±0, 00)	<b>0, 99(±0, 00)</b>	<b>0, 99(±0, 00)</b>	0, 98(±0, 00)
<b>Meadow</b>	18649	10	0, 97(±0, 00)	0, 98(±0, 00)	<b>1, 00(±0, 00)</b>	<b>1, 00(±0, 00)</b>	<b>1, 00(±0, 00)</b>	<b>1, 00(±0, 00)</b>	<b>1, 00(±0, 00)</b>
<b>Gravel</b>	2099	10	0, 78(±0, 00)	0, 80(±0, 02)	<b>0, 95(±0, 00)</b>	0, 92(±0, 00)	0, 94(±0, 00)	<b>0, 95(±0, 00)</b>	<b>0, 95(±0, 00)</b>
<b>Trees</b>	3064	10	0, 94(±0, 00)	0, 97(±0, 00)	<b>0, 99(±0, 00)</b>	<b>0, 99(±0, 00)</b>	<b>0, 99(±0, 00)</b>	<b>0, 99(±0, 00)</b>	<b>0, 99(±0, 00)</b>
<b>Metal Sheets</b>	1345	10	0, 99(±0, 00)	0, 98(±0, 01)	<b>1, 00(±0, 00)</b>	<b>1, 00(±0, 00)</b>	<b>1, 00(±0, 00)</b>	<b>1, 00(±0, 00)</b>	<b>1, 00(±0, 00)</b>
<b>Bare Soil</b>	5029	10	0, 90(±0, 00)	0, 96(±0, 01)	<b>1, 00(±0, 00)</b>	0, 99(±0, 00)	0, 99(±0, 00)	<b>1, 00(±0, 00)</b>	0, 99(±0, 00)
<b>Bitumen</b>	1330	10	0, 85(±0, 00)	0, 86(±0, 01)	<b>0, 96(±0, 00)</b>	0, 95(±0, 00)	<b>0, 96(±0, 00)</b>	<b>0, 96(±0, 00)</b>	0, 94(±0, 00)
<b>Bricks</b>	3682	10	0, 85(±0, 00)	0, 90(±0, 01)	<b>0, 97(±0, 00)</b>	0, 95(±0, 00)	<b>0, 97(±0, 00)</b>	<b>0, 97(±0, 00)</b>	<b>0, 97(±0, 00)</b>
<b>Shadow</b>	947	10	0, 99(±0, 00)	0, 98(±0, 01)	<b>1, 00(±0, 00)</b>	<b>1, 00(±0, 00)</b>	<b>1, 00(±0, 00)</b>	<b>1, 00(±0, 00)</b>	<b>1, 00(±0, 00)</b>

## 6.2 Análise do Tempo Computacional

Mensuramos o tempo computacional gasto para realizar as fases de treinamento e teste de cada abordagem de classificação base e métodos de combinação. Nas Tabelas 6.7 e 6.8, os tempos foram discriminados individualmente, e então o tempo médio sobre 30 execuções foi reportado. Abaixo segue algumas considerações sobre os tempos mensurados:

1. O tempo para criação das características (isto é, para seleção de características usando a abordagem *FEGA*, para construção do *EMP* e para a construção do modelo *PLS*) não são considerados.
2. Na fase de treinamento do *WLC-GA* o tempo considerado (descrito nas Tabelas 6.7 e 6.8) para encontrar os pesos diz respeito ao treinamento usando 5% das amostras, pois conforme descrito na Seção 5.2.1, a estimação dos pesos acontece usando conjuntos de treinos de tamanho igual a 50% do tamanho do conjunto-treino-completo. Além disso, o tempo total requerido para o treino *WLC-GA* inclui também o tempo gasto nas abordagens de classificação base.
3. Na fase de teste, o tempo do *WLC-GA* considera apenas a aplicação da combinação linear, dado que os pesos já foram encontrados, portanto o tempo total para a classificação usando o *WLC-GA* também deve incluir o tempo gasto pela classificação através das abordagens base.
4. Os métodos *MV* e *Average* não requerem treinamento.
5. Os tempos de treino do *FEGA-KNN* diz respeito apenas a atribuição dos 10% de dados ao modelo *KNN* usado no MATLAB, esse algoritmo de aprendizado não requer treinamento explícito.
6. Embora o método *WLC-GA* possui o maior custo computacional na fase de treinamento, podemos minimizar esse tempo ajustando o número de gerações do GA. Mais detalhes sobre isso serão fornecidos adiante.
7. O tempo para treinamento da abordagem de classificação base *EMP-MLP* também pode ser minimizado ao ajustar o número de épocas na *MLP*.

Tabela 6.7: Tempo computacional (em segundos) para treinar (10%) e testar (90%) no *Indian Pines*.

	<i>Pixelwise SVM</i>	<i>FEGA KNN</i>	<i>EMP MLP</i>	<i>SpecEmpPLS SVM</i>	<i>MV</i>	<i>Aveg.</i>	<i>WLC-GA</i>	<i>Stacking SVM</i>
Treino	1,3421	0,0119	14,5725	0,6829	-	-	19,80	0,2204
Teste	3,2136	1,1824	0,0196	1,0348	0,3773	3,3904	0,0140	0,7165

Tabela 6.8: Tempo computacional (em segundos) para treinar (10%) e testar (90%) no *Pavia University*.

	<i>Pixelwise SVM</i>	<i>FEGA KNN</i>	<i>EMP MLP</i>	<i>SpecEmpPLS SVM</i>	<i>MV</i>	<i>Aveg.</i>	<i>WLC-GA</i>	<i>Stacking SVM</i>
Treino	4,2077	0,0115	31,8357	1,3930	-	-	49,533	0,3269
Teste	10,4173	6,7825	0,0539	1,6248	1,0218	7,6338	0,0389	1,0984

### 6.3 Discussão

Cada uma das abordagens de classificação base são treinadas em diferentes circunstâncias, isto é, diversos tipos de representações de características e algoritmos de aprendizado de natureza diferentes. Dessa maneira, cada uma delas aprende determinadas regiões do mapa de maneiras diferentes, levando a erros e acertos em várias situações.

Ao analisar a distribuição de probabilidades (confiança), obtidas por cada uma das abordagens de classificação base, podemos perceber como o processo de combinação pode melhorar a acurácia para a interpretação dos dados hiperespectrais. Para exemplificar, vamos tomar como referência a classe *Grass/Pasture* do *Indian Pines*. Em todos os cenários estudados houveram ganhos na medida *F-score* para essa classe. Podemos notar também que cada uma das abordagens de classificação base obtiveram diferentes valores de *F-score* nessa classe. Isso indica, também, como as abordagens de classificação base se diferenciam umas das outras em termos de aprendizado. As repostas obtidas, isto é, probabilidade, confiança, etc., por cada uma delas podem ser vistas nos mapas de cores mostrados na Figura 6.5. No mapa de cor, as probabilidades são visualizadas usando cores, onde o azul representa baixas probabilidades, e o vermelho altas probabilidades. Podemos perceber que, apesar de concordarem em determinadas regiões, ainda há regiões no mapa que uma abordagem de classificação base “aprendeu” melhor que outra. Por esse motivo, ao efetuar a combinação obtivemos ganhos na medida *F-score* melhorando a interpretação desses dados e maximizando os índices de acurácia.

Observando os resultados dos cenários 1, 2 e 3 para o *Indian Pines* (Tabelas 6.1, 6.2 e 6.3), os ganhos obtidos pelo *WLC-GA* em relação à melhor abordagem de classificação base usada na combinação foram de 0,84%, 4,3% e 0,87%, respectivamente. Note que os ganhos mais expressivos, quando comparados o *WLC-GA* à melhor abordagem

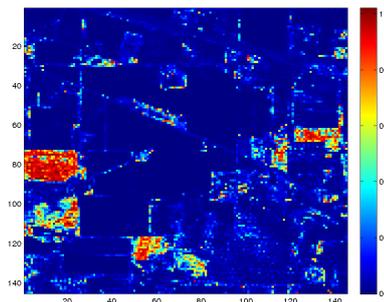
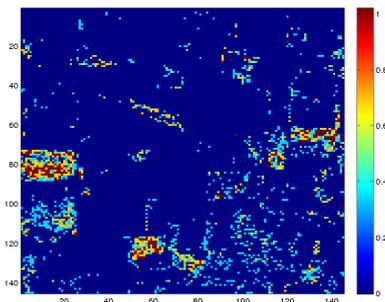
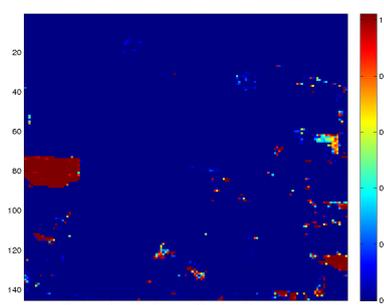
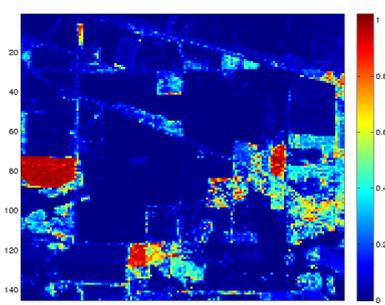
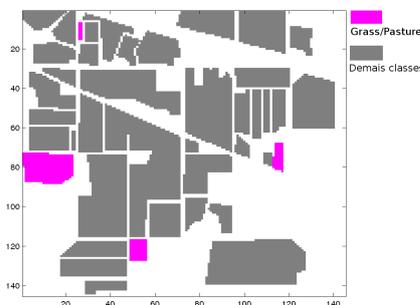
(a) *Pixelwise-SVM*.(b) *FEGA-KNN*.(c) *EMP-MLP*.(d) *SpecEmpPLS-SVM*.(e) Localização da classe *Grass/Pasture* no *Ground-truth*.

Figura 6.5: Mapa de probabilidades, referente à classe *Grass/Pasture* do *Indian Pines*, gerado por cada uma das abordagens de classificação base.

de classificação base usada na combinação, foram obtidos no cenário 2, onde a abordagem *SpecEmpPLS-SVM* não foi utilizada na combinação. Devido aos altos índices de acurácia obtidos pela abordagem de classificação *SpecEmpPLS-SVM*, é possível que as outras abordagens de classificação base não acrescentem muita informação no processo de combinação. Isso acontece quando uma ou mais abordagens concordam entre si em praticamente todas as amostras classificadas. Por esse motivo, ao utilizar a abordagem

*SpecEmpPLS-SVM*, (cenários 1 e 3), não foi possível obter ganhos expressivos usando os métodos de combinação propostos.

Os mapas de cores da Figura 6.6 mostram como o processo de combinação, usando *WLC-GA*, interpreta as diferentes regiões referentes à classe *Grass/Pasture*. Nas Fi-

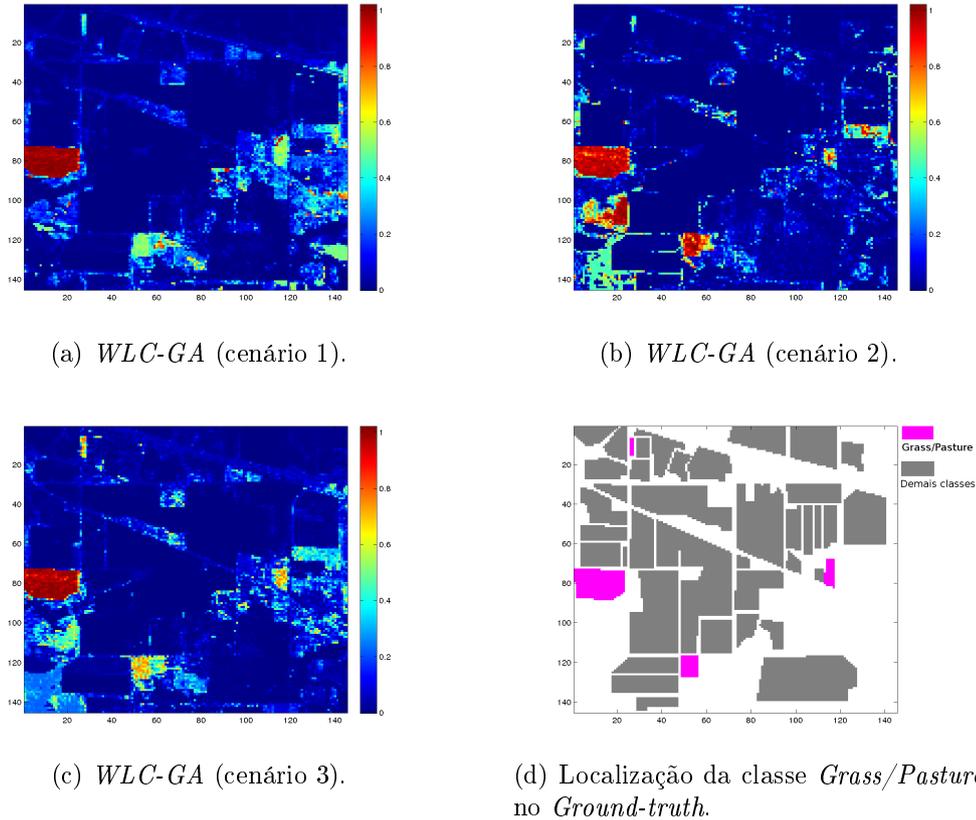


Figura 6.6: Mapa de probabilidades, referente à classe *Grass/Pasture* do *Indian Pines*, obtido pela combinação usando o *WLC-GA* nos três cenários experimentados.

Figuras 6.6a, 6.6b e 6.6c, podemos perceber que apesar de uma leve atenuação das probabilidades/confianças de algumas amostras em determinadas regiões, ainda sim, foi possível o aprendizado de praticamente todas as regiões para essa classe. Por esse motivo, conseguimos melhorar o *F-score* para essa classe em todos os cenários experimentados. Essa mesma suposição é aplicável a todas as classes existentes, e as Tabelas 6.1, 6.2 e 6.3, confirmam o aumento da medida *F-score* para praticamente todas as classes. Experimentalmente verificamos que esses conceitos e afirmações são transferíveis à segunda base de dados testada, o *Pavia University*.

Ao verificar o comportamento evolutivo do *WLC-GA*, para uma amostragem de treino de 5%, na Figura 6.7, observamos que a melhor solução é atingida em torno de 30 gerações. Isso demonstra que esse problema é facilmente solucionado usando

GA. O mesmo comportamento foi observado para o *Pavia University* e para todas as outras amostragens de treino experimentadas, no entanto, os gráficos desses resultados foram omitidos. É importante observar que a medida que se insere mais abordagens de classificação base, para efetuar a combinação, o espaço de busca se torna cada vez maior e a utilização de GA pode fornecer soluções promissoras em tempo aceitável.

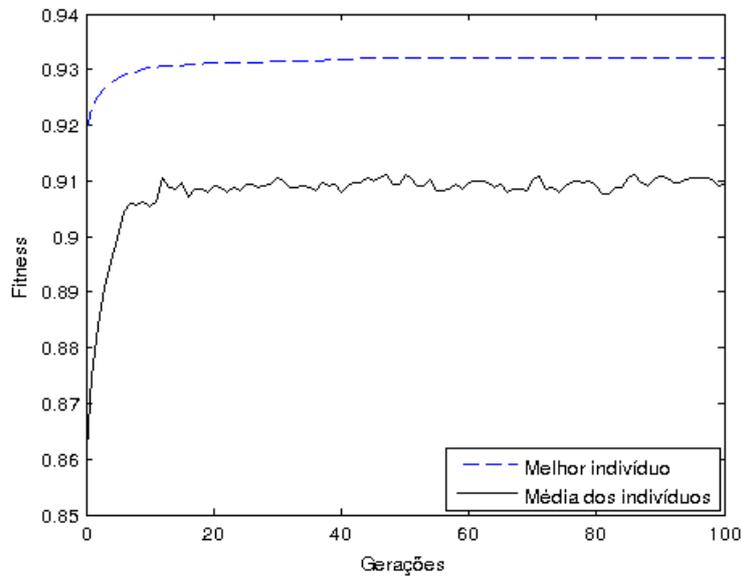


Figura 6.7: Evolução do *WLC-GA* ao longo de 100 gerações para o *Indian Pines*.

# Capítulo 7

## Conclusões e Trabalhos Futuros

Neste trabalho, propomos, implementamos e testamos diferentes abordagens de classificação, e mecanismos para combinação de múltiplos classificadores focados na interpretação de dados hiperespectrais de sensoriamento remoto. Usamos quatro tipos diferentes de abordagens de classificação para efetuar a combinação, duas que se baseiam somente na informação espectral e outras duas que se baseiam na informação espectral-espacial. Cada uma dessas abordagens de classificação base diferem tanto na representação das características quanto nos algoritmos de aprendizado utilizados. Portanto, cada uma converge e falha em diferentes circunstâncias. Assim, essa variedade de informações permite uma melhor caracterização dos dados, tornando possível um melhor desempenho dos métodos de combinação, em termos de acurácia.

Dois métodos de combinação foram propostos. O primeiro baseado em uma combinação linear ponderada o *Weighted Linear Combination optimized by Genetic Algorithm (WLC-GA)* consiste em uma combinação consensual em que pesos são encontrados usando Algoritmos Genéticos (GA). O segundo, *Stacking-SVM*, se baseia em uma combinação não linear conhecida como *Stacked Generalization*, nessa proposta usamos o algoritmo de aprendizado *Support Vector Machines (SVM)* para atuar como o combinador (meta-classificador). Além disso, comparamos os métodos propostos com outros métodos amplamente utilizados na literatura (*Majority Vote (MV)* e *Average Rule*).

Diferentes cenários foram elaborados afim de investigar o comportamento dos métodos de combinação. Cada cenário diz respeito a quais métodos fazem parte do processo de combinação. Três cenários foram usados: um cenário contendo as quatro abordagens de classificação base implementadas, e dois outros cenários usando apenas três, um sem aquela que obtém os melhores índices de acurácia e outra sem aquela que obtém os mais baixos.

Diversas configurações de treino e teste foram experimentadas, e com isso demonstramos que o método proposto WLC-GA é capaz de melhorar a acurácia com apenas 10% de amostras de treinamento. As principais conclusões e considerações a respeito dos resultados experimentais foram:

- Dentre as abordagens de classificação base, a abordagem que obteve os melhores resultados, para as duas bases de dados, foi a *SpecEmpPLS-SVM*. A partir disso, destacamos a importância de se ter uma abordagem de classificação baseada na informação espacial-espectral em nosso conjunto de classificadores.
- O método de combinação *WLC-GA* conseguiu melhorar a exatidão dos mapas temáticos nas duas bases de dados utilizadas. Apesar de sua simplicidade, o WLC-GA se mostrou superior quando comparado ao *MV*, *Average*, *Stacking-SVM* e *SpecEmpPLS-SVM*. O método também se mostrou adequado quando há conjuntos de treino de tamanho reduzido.
- Por se tratar de um novo aprendizado, envolvendo um outro classificador, o método de combinação *Stacking-SVM* necessita de uma maior quantidade de amostras de treino para se ter desempenho aproximado ao WLC-GA.
- O método MV não garante ser mais preciso que a melhor abordagem de classificação base usada na combinação.
- Ao utilizar abordagens de classificação base muito precisas, ou seja, com altos índices de acurácia, os métodos de combinação tendem a produzir resultados similares às essas. Isso pode ser explicado pelo fato de que as abordagens de classificação base podem concordar entre si na maior parte das amostras classificadas, não acrescentando informação para os métodos de combinação.

Com relação aos trabalhos futuros podemos relacioná-los em duas categorias: os trabalhos relacionados às abordagens de classificação base e aqueles pertinentes aos métodos de combinação e técnicas mais avançadas de meta-aprendizado. Na literatura técnica, é notório a investigação de novas alternativas para representações dos dados hiperespectrais de modo a melhorar a interpretação pelos algoritmos de aprendizado. Com isso, os trabalhos futuros relacionados às abordagens de classificação base propostas neste trabalho podem se concentrar nos seguintes aspectos:

- Seleção de subconjuntos de bandas espectrais. Por exemplo, na abordagem proposta denominada *FEGA-KNN*, outras medidas para validação de agrupamento podem ser utilizadas, além disso diferentes tipos de métodos de otimização podem ser testados.

- Ponderação de bandas espectrais usando métodos de *feature re-weighting*.
- União da informação espacial e espectral através de algoritmos de segmentação de imagens.
- Utilização de transformadas *wavelets* para codificar e representar a informação espectral-espacial em um único vetor de características.
- Aplicação de técnicas de *Deep learning* e *Unsupervised feature learning* para uma melhor representação dos dados espectrais, ou até mesmo para uma nova codificação das características produzidas pelo *Extended Morphological Profiles (EMP)*.

Como trabalhos futuros relacionados aos métodos de combinação, podemos considerar uma extensão do método WLC-GA na qual, além da ponderação, as probabilidades em cada abordagem de classificação são corrigidas de maneira a maximizar os acertos. Isso pode ser feito atribuindo pesos para cada uma das classes em cada uma das abordagens de classificação base, diferentemente do método tradicional onde usamos apenas um único peso para cada abordagem de classificação base. Ainda, uma função *fitness* mais adequada pode ser elaborada, dado que a função atual não trata o *overfitting*. Em relação ao método *Stacking-SVM*, experimentos focando a análise de diferentes tipos de *kernels*, ou até mesmo diferentes algoritmos de aprendizado atuando como meta-classificador, podem ser realizados. Para capturar determinados padrões de erros correlacionados, técnicas de *deep learning* e *unsupervised feature learning* aliadas ao meta-aprendizado podem fornecer uma boa alternativa aos métodos convencionais.

Embora as técnicas de combinação empregadas neste trabalho terem se mostrado adequadas para a interpretação de imagens hiperespectrais de sensoriamento, técnicas ainda mais avançadas podem ser exploradas e avaliadas. Contudo, nossos próximos passos nessa área envolvem temas como *feature re-weighting*, *deep learning* e *unsupervised feature learning*.

# Referências Bibliográficas

- Alpaydin, E. (1998). Techniques for combining multiple learners. Em *Engineering of Intelligent Systems*, pp. 6--12.
- Alpaydin, E. (2010). *Introduction to Machine Learning*. The MIT Press.
- Arzuaga-Cruz, E.; Jimenez-Rodriguez, L.; Velez-Reyes, M.; Kaeli, D.; Rodriguez-Diaz, E.; Velazquez-Santana, H.; Castrodad-Carrau, A.; Santos-Campis, L. & Santiago, C. (2004). A matlab toolbox for hyperspectral image analysis. Em *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, volume 7, pp. 4839--4842.
- Bakos, K. & Gamba, P. (2011). Hierarchical hybrid decision tree fusion of multiple hyperspectral data processing chains. *IEEE Trans. on Geoscience and Remote Sensing (TGARS)*, 49(1):388--394.
- Barker, M. & Rayens, W. (2003). Partial least squares for discrimination. *Journal of Chemometrics*, 17(3):166--173.
- Bazi, Y. & Melgani, F. (2006). Toward an optimal svm classification system for hyperspectral remote sensing images. *IEEE Trans. on Geoscience and Remote Sensing (TGARS)*, 44(11):3374--3385.
- Benediktsson, J. & Kanellopoulos, I. (1999). Classification of multisource and hyperspectral data based on decision fusion. *Geoscience and Remote Sensing, IEEE Transactions on*, 37(3):1367--1377.
- Benediktsson, J.; Palmason, J. & Sveinsson, J. (2005). Classification of hyperspectral data from urban areas based on extended morphological profiles. *IEEE Trans. on Geoscience and Remote Sensing (TGARS)*, 43(3):480--491.
- Benediktsson, J.; Sveinsson, J. & Swain, P. (1997). Hybrid consensus theoretic classification. *Geoscience and Remote Sensing, IEEE Transactions on*, 35(4):833--843.

- Benediktsson, J. & Swain, P. (1992). Consensus theoretic classification methods. *Systems, Man and Cybernetics, IEEE Transactions on*, 22(4):688 –704.
- Bernard, K.; Tarabalka, Y.; Angulo, J.; Chanussot, J. & Benediktsson, J. (2011). A stochastic minimum spanning forest approach for spectral-spatial classification of hyperspectral images. Em *Image Processing (ICIP), 2011 18th IEEE International Conference on*, pp. 1265 –1268.
- Bernard, K.; Tarabalka, Y.; Angulo, J.; Chanussot, J. & Benediktsson, J. (2012). Spectral-spatial classification of hyperspectral data based on a stochastic minimum spanning forest approach. *Image Processing, IEEE Transactions on*, 21(4):2008 – 2021.
- Bezdek, J. & Pal, N. (1995). Cluster validation with generalized dunn’s indices. Em *International Two-Stream Conference on Artificial Neural Networks and Expert Systems*, pp. 190--193.
- Bishop, C. (2006). *Pattern recognition and machine learning*, volume 4. springer New York.
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2):123--140.
- Briem, G.; Benediktsson, J. & Sveinsson, J. (2002). Multiple classifiers applied to multisource remote sensing data. *Geoscience and Remote Sensing, IEEE Transactions on*, 40(10):2291 – 2299.
- Brownlee, J. (2011). *Clever Algorithms: Nature-Inspired Programming Recipes*. Lulu.com.
- Camps-Valls, G. & Bruzzone, L. (2005). Kernel-based methods for hyperspectral image classification. *IEEE Trans. on Geoscience and Remote Sensing (TGARS)*, 43(6):1351--1362.
- Camps-Valls, G.; Gomez-Chova, L.; Munoz-Mari, J.; Vila-Francés, J. & Calpe-Maravilla, J. (2006). Composite kernels for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters (GRSL)*, 3(1):93--97.
- Chang, C. (2007). *Hyperspectral data exploitation: theory and applications*. Wiley-Blackwell.
- Chang, C.-C. & Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Trans. on Intelligent Systems and Technology*, 2(3):1--27. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.

- Cocks, T.; Jenssen, R.; Stewart, A.; Wilson, I. & Shields, T. (1998). The hmaptm airborne hyperspectral sensor: the system, calibration and performance. Em *Proceedings of the 1st EARSeL workshop on Imaging Spectroscopy*, pp. 37--42.
- Cover, T. & Hart, P. (1967). Nearest neighbor pattern classification. *Information Theory, IEEE Transactions on*, 13(1):21 -27.
- Dalla Mura, M.; Villa, A.; Benediktsson, J.; Chanussot, J. & Bruzzone, L. (2011). Classification of hyperspectral images by using extended morphological attribute profiles and independent component analysis. *IEEE Geoscience and Remote Sensing Letters (GRSL)*, 8(3):542--546.
- Dalponte, M.; Bruzzone, L. & Gianelle, D. (2008). Fusion of hyperspectral and lidar remote sensing data for classification of complex forest areas. *Geoscience and Remote Sensing, IEEE Transactions on*, 46(5):1416 -1427.
- Duda, R.; Hart, P. & Stork, D. (2000). *Pattern Classification and Scene Analysis*. John Wiley & Sons, 2 edição.
- Dunn, J. (1974). Well separated clusters and optimal fuzzy partitions. Em *Journal of Cybernetics*, volume 4, pp. 95--104.
- Facon, J. (1996). *Morfologia Matemática: Teoria e Exemplos*.
- Fauvel, M. (2007). *Spectral and Spatial Methods for the Classification of Urban Remote Sensing Data*. Tese de doutorado, University of Iceland and Grenoble Institute of Technology.
- Fauvel, M.; Chanussot, J.; Benediktsson, J. & Sveinsson, J. (2008). Spectral and spatial classification of hyperspectral data using svms and morphological profiles. *IEEE Trans. on Geoscience and Remote Sensing (TGARS)*, 46(11):3804--3814.
- Fauvel, M.; Chanussot, J. & Benediktsson, J. A. (2009). Kernel principal component analysis for the classification of hyperspectral remote sensing data over urban areas. *EURASIP J. Adv. Signal Process*, 2009(11):1--14.
- Fauvel, M.; Tarabalka, Y.; Benediktsson, J. A.; Chanussot, J. & Tilton, J. C. (2012). Advances in spectral-spatial classification of lyperspectral images. *Proceedings of the IEEE*, PP(99):1--24. accepted for publication.
- Gege, P.; Beran, D.; Mooshuber, W.; Schulz, J. & Van Der Piepen, H. (1998). System analysis and performance of the new version of the imaging spectrometer rosis. Em *The First EARSeL Workshop on Imaging Spectroscopy*, pp. 29--35.

- Goetz, A.; Vane, G.; Solomon, J. & Rock, B. (1985). Imaging spectrometry for earth remote sensing. *Science*, 228(4704):1147.
- Goldberg, D. (1989). *Genetic algorithms in search, optimization, and machine learning*. Addison-wesley.
- Green, R.; Eastwood, M.; Sarture, C.; Chrien, T.; Aronsson, M.; Chippendale, B.; Faust, J.; Pavri, B.; Chovit, C.; Solis, M. et al. (1998). Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (aviris). *Remote Sensing of Environment*, 65(3):227--248.
- Halkidi, M.; Batistakis, Y. & Vazirgiannis, M. (2001). On clustering validation techniques. *Journal of Intelligent Information Systems*, 17:107--145.
- Hughes, G. (1968). On the mean accuracy of statistical pattern recognizers. *IEEE Trans. on Information Theory*, 14(1):55--63.
- Jackson, Q. & Landgrebe, D. (2002). Adaptive bayesian contextual classification based on markov random fields. *Geoscience and Remote Sensing, IEEE Transactions on*, 40(11):2454 - 2463.
- Jain, R. (1991). *Art of Computer Systems Performance Analysis Techniques For Experimental Design Measurements Simulation And Modeling*. Wiley Computer Publishing, John Wiley and Sons, Inc.
- Jimenez-Rodriguez, L.; Arzuaga-Cruz, E. & Vélez-Reyes, M. (2007). Unsupervised linear feature-extraction methods and their effects in the classification of high-dimensional data. *IEEE Trans. on Geoscience and Remote Sensing (TGARS)*, 45(2):469--483.
- Kaynak, C. & Alpaydin, E. (2000). Multistage cascading of multiple classifiers: One man's noise is another man's data. Em *Proceedings of the Seventeenth International Conference on Machine Learning, ICML '00*, pp. 455--462, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.
- Kittler, J.; Hatef, M.; Duin, R. P. W. & Matas, J. (1998). On combining classifiers. *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, 20(3):226--239.
- Kohavi, R. & John, G. (1997). Wrappers for feature subset selection. *Artificial intelligence*, 97(1-2):273--324.

- Kuncheva, L. (2004). *Combining Pattern Classifiers: Methods and Algorithms. 2004.* Wiley-Interscience.
- Kuo, B.-C.; Li, C.-H. & Yang, J.-M. (2009). Kernel nonparametric weighted feature extraction for hyperspectral image classification. *IEEE Trans. on Geoscience and Remote Sensing (TGARS)*, 47(4):1139--1155.
- Lam, L. & Suen, C. (1995). Optimal combinations of pattern classifiers. *Pattern Recognition Letters*, 16(9):945--954.
- Lee, J. & Ersoy, O. (2007). Consensual and hierarchical classification of remotely sensed multispectral images. *IEEE Trans. on Geoscience and Remote Sensing (TGARS)*, 45(9):2953--2963.
- Li, J.; Bioucas-Dias, J. & Plaza, A. (2010). Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning. *IEEE Trans. on Geoscience and Remote Sensing (TGARS)*, 48(11):4085--4098.
- Li, J.; Bioucas-Dias, J. M. & Plaza, A. (2011). Hyperspectral image segmentation using a new bayesian approach with active learning. *IEEE Trans. on Geoscience and Remote Sensing (TGARS)*, PP(99):1139--1155.
- L.I. Kuncheva, J. B. & Duin, R. (2001). Decision templates for multiple classifier fusion: an experimental comparison. *Pattern Recognition*, 34:299--314.
- Licciardi, G.; Marpu, P.; Chanussot, J. & Benediktsson, J. (2012). Linear versus nonlinear pca for the classification of hyperspectral data based on the extended morphological profiles. *Geoscience and Remote Sensing Letters, IEEE*, 9(3):447 - 451.
- Licciardi, G.; Pacifici, F.; Tuia, D.; Prasad, S.; West, T.; Giacco, F.; Thiel, C.; Inglada, J.; Christophe, E.; Chanussot, J. & Gamba, P. (2009). Decision fusion for the classification of hyperspectral data: Outcome of the 2008 GRS-S data fusion contest. *IEEE Trans. on Geoscience and Remote Sensing (TGARS)*, 47(11):3857--3865.
- Luo, R. & Kay, M. (1989). Multisensor integration and fusion in intelligent systems. *Systems, Man and Cybernetics, IEEE Transactions on*, 19(5):901 - 931.
- Manning, C. D.; Raghavan, P. & Schtze, H. (2008). *Introduction to Information Retrieval.* Cambridge University Press. Online version available at <http://nlp.stanford.edu/IR-book/>.

- Marpu, P.; Gamba, P. & Niemeyer, I. (2009). Hyperspectral data classification using an ensemble of class-dependent neural networks. *IEEE GRSS WHISPERS*, n/a(n/a):1-4.
- MATLAB (2008). *version 7.7.0.471 (R2008b)*. The MathWorks Inc., Natick, Massachusetts.
- MATLAB (2012). *version 7.14.0.739 (R2012a)*. The MathWorks Inc., Natick, Massachusetts.
- Melgani, F. & Bruzzone, L. (2004). Classification of hyperspectral remote sensing images with support vector machines. *IEEE Trans. on Geoscience and Remote Sensing (TGARS)*, 42(8):1778--1790.
- Mitchell, M. (1998). *An Introduction to Genetic Algorithms*. MIT Press, Cambridge, MA, USA.
- Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill Science/Engineering/Math.
- Mueller, A.; Hausold, A. & Strobl, P. (2002). Hysens-dais/rosis imaging spectrometers at dlr. Em *Proceedings of SPIE*, volume 4545, p. 225.
- Neher, R. & Srivastava, A. (2005). A bayesian mrf framework for labeling terrain using hyperspectral imaging. *Geoscience and Remote Sensing, IEEE Transactions on*, 43(6):1363 – 1374.
- Pal, M. & Foody, G. (2010). Feature selection for classification of hyperspectral data by svm. *Geoscience and Remote Sensing, IEEE Transactions on*, 48(5):2297 –2307.
- Palmason, J.; Benediktsson, J.; Sveinsson, J. & Chanussot, J. (2005). Classification of hyperspectral data from urban areas using morphological preprocessing and independent component analysis. Em *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, volume 1, pp. 4--pp.
- Pearlman, J.; Segal, C.; Liao, L.; Carman, S.; Folkman, M.; Browne, B.; Ong, L. & Ungar, S. (2000). Development and operations of the eo-1 hyperion imaging spectrometer. Em *SPIE Int. Soc. Opt. Eng.*, volume 4135, pp. 243--253.
- Pedernana, M.; Marpu, P.; Mura, M.; Benediktsson, J. & Bruzzone, L. (2012). A novel supervised feature selection technique based on genetic algorithms. Em *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*, pp. 60 –63.

- Plaza, A.; Benediktsson, J.; Boardman, J.; Brazile, J.; Bruzzone, L.; Camps-Valls, G.; Chanussot, J.; Fauvel, M.; Gamba, P.; Gualtieri, A.; Marconcini, M.; Tilton, J. & Trianni, G. (2009). Recent advances in techniques for hyperspectral image processing. *Remote Sensing Environmet*, 113(1):110--122.
- Plaza, A. & Tilton, J. (2005). Automated selection of results in hierarchical segmentations of remotely sensed hyperspectral images. Em *Geoscience and Remote Sensing Symposium, 2005. IGARSS '05. Proceedings. 2005 IEEE International*, volume 7, pp. 4946 – 4949.
- Prasad, S.; Bruce, L. M. & Chanussot, J. (2011). *Optical Remote Sensing: Advances in Signal Processing and Exploitation Techniques*, volume 3. Springer.
- Rickard, L.; Basedow, R.; Zalewski, E.; Silverglate, P. & Landers, M. (1993). Hydice: An airborne system for hyperspectral imaging. Em *Proceedings of SPIE*, volume 1937, pp. 173--179.
- Rosipal, R. & Krämer, N. (2006). Overview and recent advances in partial least squares. Em *Proceedings of the 2005 international conference on Subspace, Latent Structure and Feature Selection, SLSFS'05*, pp. 34--51, Berlin, Heidelberg. Springer-Verlag.
- Salzenstein, F. & Collet, C. (2006). Fuzzy markov random fields versus chains for multispectral image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(11):1753 –1767.
- Santos, A. B.; de A. Araújo, A. & Menotti, D. (2012a). Combiner of classifiers using genetic algorithm for classification of remote sensed hyperspectral images. Em *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 4146 – 4149.
- Santos, A. B.; de A. Araújo, A. & Menotti, D. (2012b). Combining multiple approaches for accuracy improvement in remote sensed hyperspectral images classification. Em Lopes, H. & Hirata, N., editores, *Workshop of Theses and Dissertations (WTD) in SIBGRAPI 2012 (XXV Conference on Graphics, Patterns and Images)*, pp. 54--59, Ouro Preto, MG, Brazil.
- Santos, A. B.; de A. Araújo, A. & Menotti, D. (2013). Combining multiple classification methods for hyperspectral data interpretation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, PP(99):1–10.

- Santos, A. B.; de S. Celes, C. S. F.; de A. Araújo, A. & Menotti, D. (2012c). Feature selection for classification of remote sensed hyperspectral images: A filter approach using genetic algorithm and cluster validity. Em *The 2012 International Conference on Image Processing, Computer Vision, and Pattern Recognition (ICCV'12)*, volume 2, pp. 675–681.
- Sarkar, A.; Biswas, M.; Kartikeyan, B.; Kumar, V.; Majumder, K. & Pal, D. (2002). A mrf model-based segmentation approach to classification for multispectral imagery. *Geoscience and Remote Sensing, IEEE Transactions on*, 40(5):1102–1113.
- Schapire, R. (1989). The strength of weak learnability. Em *Foundations of Computer Science, 1989., 30th Annual Symposium on*, pp. 28–33.
- Scholkopf, B. & Smola, A. J. (2001). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge, MA, USA.
- Schwartz, W. R. (2011). Human Detection Based on Large Feature Sets Using Graphics Processing Units. *Informatica*, 35(4):473–479.
- Schwartz, W. R.; Davis, L. S. & Pedrini, H. (2011). Local Response Context Applied to Pedestrian Detection. Em *Iberoamerican Congress on Pattern Recognition*, pp. 181–188.
- Schwartz, W. R.; Guo, H.; Choi, J. & Davis, L. S. (2012). Face Identification Using Large Feature Sets. *IEEE Transactions on Image Processing*, 21(4):2245–2255.
- Serra, J. (1982). *Image Analysis and Mathematical Morphology*. Academic Press.
- Shen, L. & Lia, S. (2011). Three-dimensional gabor wavelets for pixel-based hyperspectral imagery classification. *IEEE Trans. on Geoscience and Remote Sensing (TGARS)*, 49(12):5039–5046.
- Stevenson, B.; O'Connor, R.; Kendall, W.; Stocker, A.; Schaff, W.; Holasek, R.; Even, D.; Alexa, D.; Salvador, J.; Eismann, M. et al. (2005). The civil air patrol archer hyperspectral sensor system. Em *Proceedings of SPIE*, volume 5787, p. 17.
- Tarabalka, Y. (2010). *Classification of Hyperspectral Data Using Spectral-Spatial Approaches*. Tese de doutorado, University of Iceland and Grenoble Institute of Technology.
- Tarabalka, Y.; Benediktsson, J. & Chanussot, J. (2009a). Spectral-spatial classification of hyperspectral imagery based on partitional clustering techniques. *IEEE Trans. on Geoscience and Remote Sensing (TGARS)*, 47(8):2973–2987.

- Tarabalka, Y.; Benediktsson, J.; Chanussot, J. & Tilton, J. (2010a). Multiple spectral-spatial classification approach for hyperspectral data. *IEEE Trans. on Geoscience and Remote Sensing (TGARS)*, 48(11):4122--4132.
- Tarabalka, Y.; Chanussot, J. & Benediktsson, J. (2009b). Classification based marker selection for watershed transform of hyperspectral images. *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, n/a(n/a):105--108.
- Tarabalka, Y.; Chanussot, J. & Benediktsson, J. (2010b). Segmentation and classification of hyperspectral images using minimum spanning forest grown from automatically selected markers. *IEEE Trans. on Systems, Man and Cybernetics. Part B.*, 40(5):1267--1279.
- Tarabalka, Y.; Chanussot, J. & Benediktsson, J. (2010c). Segmentation and classification of hyperspectral images using watershed transformation. *Pattern Recognition*, 43(7):2367--2379.
- Tarabalka, Y.; Chanussot, J.; Benediktsson, J.; Angulo, J. & Fauvel, M. (2008). Segmentation and classification of hyperspectral data using watershed. *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, n/a(n/a):652--655.
- Tarabalka, Y.; Fauvel, M.; Chanussot, J. & Benediktsson, J. (2010d). Svm- and mrf-based method for accurate classification of hyperspectral images. *IEEE Geoscience and Remote Sensing Letters (GRSL)*, 7(4):736--740.
- Tarabalka, Y.; Tilton, J. C.; Benediktsson, J. A. & Chanussot, J. (2011). Marker-based hierarchical segmentation and classification approach for hyperspectral imagery. pp. 1089--1092.
- Theodoridis, S. & Koutroumbas, K. (2003). *Pattern Recognition*. Academic Press An imprint of Elsevier Science.
- Thoonen, G.; Mahmood, Z.; Peeters, S. & Scheunders, P. (2012). Multisource classification of color and hyperspectral images using color attribute profiles and composite decision fusion. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, 5(2):510 -521.
- Tilton, J. (1998). Image segmentation by region growing and spectral clustering with a natural convergence criterion. Em *Geoscience and Remote Sensing Symposium Proceedings, 1998. IGARSS '98. 1998 IEEE International*, volume 4, pp. 1766 -1768 vol.4.

- Tilton, J.; Lawrence, W. T. & Plaza, A. (2006). Utilizing hierarchical segmentation to generate water and snow masks to facilitate monitoring change with remotely sensed image data. *GIScience and Remote Sensing*, 43(1):39 -- 66.
- Ting, K. M. & Witten, I. H. (2011). Issues in stacked generalization. *CoRR*, abs/1105.5466.
- Tuia, D. & Camps-Valls, G. (2011). Urban image classification with semisupervised multiscale cluster kernels. *IEEE Trans. on Geoscience and Remote Sensing (TGARS)*, 4(1):65--74.
- Velez-Reyes, M. & Jimenez, L. (1998). Subset selection analysis for the reduction of hyperspectral imagery. Em *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, volume 3, pp. 1577--1581.
- Velez-Reyes, M.; Jimenez, L.; Linares, D. & Velazquez, H. (2000). Comparison of matrix factorization algorithms for band selection in hyperspectral imagery. Em *SPIE proceedings series*, pp. 288--297.
- Waske, B.; Linden, S. V. D.; Benediktsson, J.; Rabe, A. & Hostert, P. (2010). Sensitivity of support vector machines to random feature selection in classification of hyperspectral data. *IEEE Trans. on Geoscience and Remote Sensing (TGARS)*, 48(7):2880--2889.
- Weise, T. (2008). Global optimization algorithms - theory and application.
- Wold., H. (1985). Partial least squares. *Encyclopedia of Statistical Sciences*, 6:581--591.
- Wolpert, D. (1992). Stacked generalization. *Neural Networks*, 5(2):241--259.
- Yu, L. & Liu, H. (2003). Feature selection for high-dimensional data: A fast correlation-based filter solution. Em *International Conference on Machine Learning*, volume 20, pp. 856--863.
- Zadeh, L. A. (1956). Fuzzy sets. *Information and Control*, 8:338--353.
- Zhang, B.; Li, S.; Jia, X.; Gao, L. & Peng, M. (2011). Adaptive markov random field approach for classification of hyperspectral imagery. *IEEE Geoscience and Remote Sensing Letters (GRSL)*, 8(5):973--977.
- Zhou, P. & Wang, R. (2010). Learning conditional random fields for classification of hyperspectral images. *IEEE Trans. on Image Processing (TIP)*, 19(7):1890--1907.