

UNIVERSIDADE FEDERAL DE SÃO CARLOS

CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

**UM NOVO MÉTODO *WRAPPER*
MULTIOBJETIVO PARA SELEÇÃO DE
BANDAS DE IMAGENS HIPERESPECTRAIS**

DIEGO SAQUI

ORIENTADOR: PROF. DR. JOSÉ HIROKI SAITO

São Carlos – SP

Fevereiro/2020

UNIVERSIDADE FEDERAL DE SÃO CARLOS

CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

**UM NOVO MÉTODO *WRAPPER*
MULTIOBJETIVO PARA SELEÇÃO DE
BANDAS DE IMAGENS HIPERESPECTRAIS**

DIEGO SAQUI

Tese apresentada ao Programa de Pós-Graduação em
Ciência da Computação da Universidade Federal de
São Carlos, como parte dos requisitos para a obten-
ção do título de Doutor em Ciência da Computação,
área de concentração: Processamento de Imagens e
Sinais

Orientador: Prof. Dr. José Hiroki Saito

O presente trabalho foi realizado com apoio
da Coordenação de Aperfeiçoamento de
Pessoal de Nível Superior – Brasil (CAPES)
- Código de Financiamento 001.

São Carlos – SP

Fevereiro/2020



Folha de Aprovação

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Tese de Doutorado do candidato Diego Saqui, realizada em 19/02/2020:

Prof. Dr. José Hiroki Saito
UFSCar

Prof. Dr. Alexandre Luis Magalhães Levada
UFSCar

Prof. Dr. Edilson Reis Rodrigues Kato
UFSCar

Prof. Dr. Denis Henrique Pinheiro Salvadeo
UNESP

Prof. Dr. Moacir Antonelli Ponti
ICMC/USP

RESUMO

Imagens hiperspectrais (IHs) são caracterizadas pela resolução espectral maior que outros tipos de imagens, tendo aplicações em áreas como medicina, mineração e, principalmente, na agricultura. Essas imagens em conjunto com sensoriamento remoto tornaram-se uma ferramenta útil para agricultura de precisão, possibilitando tarefas de identificação e análise de condições de saúde de áreas agrícolas. Para essa identificação se faz necessária a segmentação de imagens, que pode ser obtida por meio de classificação. Um problema intrínseco às IHs é o volume de dados que pode representar um desafio em termos de transmissão, armazenamento, processamento e, também, no desempenho de algoritmos de classificação (causados pela maldição da dimensionalidade). Técnicas de redução da dimensionalidade de IHs são promissoras, mas muitas delas foram projetadas para lidar com um único objetivo e não podem assegurar um equilíbrio entre objetivos conflitantes. Exemplos de objetivos conflitantes podem ser baseados na melhora da classificação de pixels e redução da quantidade de bandas de IHs simultaneamente, sendo este último relacionado a dimensionalidade dessas imagens. Para tentar lidar com soluções com múltiplos objetivos conflitantes podem ser aplicados algoritmos multiobjetivos que são projetados para essa finalidade. Métodos de seleção de bandas baseados em algoritmos multiobjetivos têm sido propostos recentemente na literatura, porém muitas estratégias ainda não foram exploradas ou combinadas adequadamente. Baseado em diferentes abordagens da literatura, nesta pesquisa foi elaborado um método de seleção de bandas multiobjetivo chamado *Wrapper Multiobjective Evolutionary Band Selection* (WMOEBS) composto por estratégias que foram testadas experimentalmente. O WMOEBS é baseado na estratégia *Wrapper* incorporando o classificador *Support Vector Machine* (SVM), que utiliza informação espacial e espectral, e realiza uma seleção inicial para diminuir as bandas correlacionadas, consistindo num algoritmo multiobjetivo para lidar com resultados da classificação e quantidade de bandas simultaneamente e um tomador de decisão para retornar uma única solução final. O WMOEBS foi comparado com os métodos estado-da-arte da literatura em critérios de melhoria de classificação em diferentes métricas e capacidade de redução da quantidade de bandas. Experimentos demonstraram que o WMOEBS apresenta resultados superiores à maioria dos casos testados para métricas de classificação, incluindo testes estatísticos, além de também ser vantajoso na redução da quantidade de bandas.

Palavras-chave: Seleção de bandas, Algoritmos Multiobjetivos, Imagens Hiperespectrais, Sensoriamento Remoto, estratégia *Wrapper*, *Support Vector Machine*

ABSTRACT

Hyperspectral images (HIs) are characterized by higher spectral resolution than other kind of images, having applications in areas such as medicine, mining, and especially in agriculture. These images in conjunction with remote sensing have become a useful tool for precision farming, enabling identification and analysis of health conditions in agricultural areas. For this identification it is necessary to segment the images that can be obtained through classification. An intrinsic problem with HIs is the volume of data that can pose a challenge in terms of transmission, storage, processing and also the performance of classification algorithms (caused by the curse of dimensionality). Techniques that reduce dimensionality are promising for HIs, but many of them are designed to deal with a single objective and cannot assure a balance between conflicting objectives. Examples of conflicting objectives can be based on improving the pixel classification and reducing the number of HIs bands simultaneously, the latter being related to the dimensionality of these images. To try to deal with solutions for conflicting objectives, it can be applied multiobjective algorithms that are designed for this purpose. Band selection methods based on multiobjective algorithms have recently been proposed in the literature, but many strategies have not yet been explored or properly combined. Based on different approaches from the literature, in this research it was developed a multiobjective band selection method called Wrapper Multiobjective Evolutionary Band Selection (WMOEBS) composed of strategies that were experimentally tested. WMOEBS is based on the Wrapper strategy incorporating the Support Vector Machine (SVM) classifier, using spatial and spectral information as input, it makes an initial selection to narrow down correlated bands, being a multiobjective algorithm dealing with classification results and number of bands simultaneously and a decision maker to return a single final solution. WMOEBS has been compared with state-of-the-art methods in the classification improvement criteria for different metrics and bandwidth reduction capability. Experiments have shown that WMOEBS presents superior results to most cases tested for classification metrics, including when applying statistical tests, and it is also advantageous in reducing the number of bands.

Keywords: Band Selection, Multiobjective Algorithms, Hyperspectral Images, Remote Sensing, Wrapper Strategy, *Support Vector Machine*

LISTA DE FIGURAS

1.1	Exemplo de mapa de classificação da imagem Indian Pines. Adaptado de Xie et al. (2018).	14
1.2	Aplicação prática do método proposto.	22
2.1	Sensoriamento remoto (adaptado de Landgrebe (2003)).	25
2.2	Componentes (bandas) de uma imagem RGB (adaptado de Gonzalez e Woods (2018)).	26
2.3	Imageamento multiespectral (A) e hiperespectral (B) (adaptado de Habermann (2018)).	26
2.4	Cubo hiperespectral e assinatura espectral (adaptado de Habermann (2018)). . .	27
2.5	Etapas básicas da classificação de pixels de imagem com múltiplas bandas (adaptado de Lillesand, Kiefer e Chipman (2004)).	30
2.6	Exemplo das distribuições de probabilidades de classes de imagens nas bandas 3 e 4 (adaptado de Lillesand, Kiefer e Chipman (2004)).	33
2.7	Representação do SVM com margem rígida.	34
2.8	Ilustração da aplicação de um filtro de média e filtragem espacial linear.	40
3.1	Exemplos de bandas da imagem Indian Pines. Bandas com baixos valores de entropia: 200, 199, 145, 2 e 104. Bandas com elevados valores de entropia: 27, 112, 29, 30 e 31 (adaptado de Gong, Zhang e Yuan (2016)).	49
3.2	Ilustração de elementos de um algoritmo genético com a função de <i>fitness</i> baseada na distância de Bhattacharyya. Adaptado de Zhang, Sun e Li (2009). . . .	59
3.3	Representação do MOEA/D onde um problema multiobjetivo foi decomposto em 9 subproblemas e a vizinhança do subproblema 3 é composta pelos subproblemas 1, 2, 3, 4 e 5 (adaptado de Zhou, Zhang e Zhang (2012)).	65

4.1	Imagem Hiperespectral Indian Pines, GT e legenda com classes.	74
4.2	Imagem Hiperespectral Salinas, GT e legenda com classes.	75
4.3	Imagem Hiperespectral Pavia University, GT e legenda com classes.	75
4.4	Grupos de pixels usados como amostras para experimentos usando <i>kernel</i> 3×3	78
4.5	Configuração das amostras para os experimentos com <i>5-fold Cross-Validation</i> .	79
5.1	<i>Framework</i> para seleção de bandas multiobjetivo baseado na estratégia <i>Wrapper</i> .	84
5.2	Representação de uma solução por meio de um vetor binário.	84
5.3	Representação da região balanceada (KIMOVSKI et al., 2018)	97
5.4	<i>Wrapper multiobjective evolutionary band selection</i> (WMoEBS)	114
6.1	Evolução da precisão e quantidade de bandas para os métodos WMoEBS, WMoEBS2, IRMOBS, GA-SVM e IDMMoBS. (a) (c) e (e) Precisão (OA). (b), (d) e (f) Quantidade de bandas.	125

LISTA DE TABELAS

5.1	Configurações do MOEA/D utilizado nos testes dos componentes avaliados. . .	101
5.2	Resultados da quantidade de bandas e classificação dos experimentos usando ou não a informação espacial, f_1 baseado em diferentes índices e com bandas selecionadas usando o MOEA/D.	104
5.3	Resultados da quantidade de bandas e classificação dos experimentos considerando a utilização da informação espacial obtida por diferentes filtros e com bandas selecionadas usando o MOEA/D.	108
5.4	Resultados da quantidade de bandas e classificação dos experimentos considerando diferentes DMs e com bandas selecionadas usando o MOEA/D.	111
5.5	Resultados dos p -values do teste T-pareado para o índice OA para estratégia proposta para os diferentes DMs analisados.	111
5.6	Resultados dos p -values do teste T-pareado para o índice AA para estratégia proposta para os diferentes DMs analisados.	112
5.7	Comparação das vitórias considerando o p -value dos DM baseados em diferentes estratégias para todas imagens analisadas.	113
5.8	Comparação das vitórias considerando a quantidade de bandas proporcionada pelos DMs baseado em centroide ou seleção direta.	113
6.1	Configurações gerais dos experimentos e do WMoEBS.	118
6.2	Resultados da quantidade de bandas e índices de classificação dos experimentos replicados com métodos supervisionados, WMoEBS e WMoEBS2.	120
6.3	Resultados da quantidade de bandas e índices de avaliação de classificação dos experimentos replicados com métodos não supervisionados, WMoEBS e WMoEBS2.	122

6.4	Resultados da quantidade de bandas e índices de avaliação de classificação dos experimentos com métodos supervisionados implementados, WMoEBS e WMoEBS2.	127
6.5	Resultados dos <i>p-values</i> do teste T-pareado para o índice OA em métodos supervisionados, WMoEBS e WMoEBS2.	128
6.6	Resultados dos <i>p-values</i> do teste T-pareado para o índice AA em métodos supervisionados, WMoEBS e WMoEBS2.	129
6.7	Resultados da quantidade de bandas e índices de classificação dos experimentos com métodos não supervisionados implementados, WMoEBS e WMoEBS2, . . .	131
6.8	Resultados dos <i>p-values</i> do teste T-pareado para o índice OA em métodos não supervisionados, WMoEBS e WMoEBS2,	131
6.9	Resultados dos <i>p-values</i> do teste T-pareado para o índice AA em métodos não supervisionados, WMoEBS e WMoEBS2,	132
6.10	Resultados da quantidade de bandas e índices de classificação dos experimentos com a utilização de todas as bandas, WMoEBS e WMoEBS2.	133
6.11	Resultados dos <i>p-values</i> do teste T-pareado para o índice OA com a utilização de todas as bandas, WMoEBS e WMoEBS2.	134
6.12	Resultados dos <i>p-values</i> do teste T-pareado para o índice AA com a utilização de todas as bandas, WMoEBS e WMoEBS2.	134

SUMÁRIO

CAPÍTULO 1 – INTRODUÇÃO	13
1.1 Problemas e motivação	15
1.2 Objetivos, hipóteses e principais contribuições	20
1.2.1 Objetivos específicos	20
1.2.2 Hipóteses	21
1.3 Principais contribuições e delimitação de escopo	21
1.4 Organização do trabalho	23
CAPÍTULO 2 – CONCEITOS DE IMAGENS HIPERESPECTRAIS	24
2.1 Sensoriamento Remoto e Imagens Hiperespectrais	24
2.2 Características sobre a alta dimensionalidade de dados	28
2.3 Classificação e segmentação em Imagens Hiperespectrais	29
2.3.1 Exemplos de classificadores para imagens hiperespectrais	31
2.3.2 <i>Gaussian Maximum Likelihood Classifier</i> (GMLC)	32
2.3.3 <i>Support Vector Machines</i> (SVM)	34
2.3.4 Informação espacial para classificação e segmentação	36
2.3.5 Obtenção de informação espacial por meio de filtragem	38
2.4 Considerações Finais do Capítulo	41
CAPÍTULO 3 – SELEÇÃO DE BANDAS EM IMAGENS HIPERESPECTRAIS	43
3.1 Seleção de características/bandas	43

3.1.1	Métodos baseados em ranqueamento	47
3.1.2	Métodos baseados em <i>clustering</i>	51
3.1.3	Métodos baseados em esparsidade	53
3.1.4	Métodos baseados em aprendizado embarcado	54
3.1.5	Métodos baseados em estratégias de busca	56
3.1.6	Métodos baseados em algoritmos de otimização multiobjetivo	62
3.1.7	Métodos baseados em esquemas híbridos	68
3.2	Considerações Finais do Capítulo	71
CAPÍTULO 4 – MATERIAIS E MÉTODOS		73
4.1	Amostras / Imagens Hiperespectrais (IHs)	73
4.2	Ferramentas utilizadas	76
4.3	Configurações dos experimentos e critérios de avaliação	76
4.3.1	Estratégia de captura de amostras das IHs sem sobreposição	77
4.3.2	Configuração de amostras de treino, validação e testes	78
4.3.3	Índices avaliados e exibição de resultados	80
4.4	Considerações Finais do Capítulo	81
CAPÍTULO 5 – WMOEBS - WRAPPER MULTIOBJECTIVE EVOLUTIONARY BAND SELECTION		82
5.1	<i>Framework</i> para seleção de bandas multiobjetivo baseado na estratégia <i>Wrapper</i>	83
5.2	Componentes analisados para estruturação do WMoEBS	86
5.2.1	Classificação de pixels e Funções de <i>Fitness</i>	86
5.2.1.1	Funções de <i>Fitness</i>	86
5.2.1.2	Organização das amostras no processo de seleção de bandas .	88
5.2.1.3	Classificação de pixels	88
5.2.2	Método de otimização multiobjetivo	89
5.2.3	Inicialização do conjunto de soluções (população)	92

5.2.4	Heurística de reparo	94
5.2.5	Utilização da informação espacial obtida por meio de filtros	96
5.2.6	<i>Decision-Maker</i> (DM)	97
5.3	Experimentos para as escolhas dos componentes do WMoEBS	99
5.3.1	Grupo de experimentos 1: Classificadores, uso da informação espacial e função de <i>fitness</i> f_1	102
5.3.1.1	Análise dos Resultados do Grupo de Experimentos 1	103
5.3.2	Grupo de experimentos 2: Avaliação de diferentes filtros para obtenção de informações espaciais	106
5.3.2.1	Análise dos Resultados do Grupo de Experimentos 2	107
5.3.3	Grupo de experimentos 3: Avaliação dos tomadores de decisão (<i>Decision Maker</i> - DM)	109
5.3.3.1	Análise dos Resultados do Grupo de Experimentos 3	110
5.4	WMoEBS	114
5.5	Considerações Finais do Capítulo	115
CAPÍTULO 6 – EXPERIMENTOS COMPARATIVOS E RESULTADOS		116
6.1	Configurações de amostras e experimentos comparativos	116
6.2	Grupo de experimentos 1: Experimentos com comparação direta de resultados	118
6.2.1	Supervisionados	120
6.2.1.1	Análise dos Resultados do Grupo de Experimentos 1: Supervisionados	120
6.2.2	Não Supervisionados	121
6.2.2.1	Análise dos Resultados do Grupo de Experimentos 1: Não Supervisionados	121
6.3	Grupo de experimentos 2: Experimentos com métodos implementados	123
6.3.1	Evolução dos resultados durante a execução de métodos de busca	124
6.3.2	Avaliações das amostras de testes para os métodos analisados	126

6.3.2.1	Supervisionados	126
6.3.2.2	Análise dos Resultados do Grupo de Experimentos 2: Supervisionados	128
6.3.2.3	Não Supervisionados	130
6.3.2.4	Análise dos Resultados do Grupo de Experimentos 2: Não Supervisionados	130
6.4	Grupo de experimentos 3: IHs com todas as bandas e bandas selecionadas pelo WMoEBS	133
6.4.1	Análise dos Resultados do Grupo de Experimentos 3	133
6.5	Considerações Finais do Capítulo	135
CAPÍTULO 7 – CONSIDERAÇÕES FINAIS, CONCLUSÕES E TRABALHOS FUTUROS		136
7.1	Elaboração do WMoEBS	137
7.2	Seleção de bandas com o WMoEBS e classificação de imagens hiperespectrais	138
7.3	Limitações do WMoEBS e dificuldades encontradas	140
7.4	Publicações geradas desta pesquisa	141
7.5	Trabalhos Futuros	142
REFERÊNCIAS		143
GLOSSÁRIO		155
APÊNDICE A – ARTIGOS PRODUZIDOS DURANTE ESTA PESQUISA		158

Capítulo 1

INTRODUÇÃO

Um dos desafios da agricultura é a realização adequada do monitoramento de condições relacionadas ao meio ambiente, de recursos naturais, do solo, práticas agrícolas e plantações (JORGE; INAMASU, 2014). Nessa área, obter as informações necessárias para gerenciar recursos por métodos manuais, ou seja, uma pessoa caminhando na lavoura, é um processo demorado e custoso (LANDGREBE, 2003). Uma alternativa a esse processo é a adoção de uma estratégia de monitoramento chamada Sensoriamento Remoto (SR), que pode ser aplicada com o uso de plataformas como satélites, aviões ou drones. Em SR, essas plataformas possuem sensores acoplados que permitem obter informações para análise e, posteriormente, o manejo adequado das áreas monitoradas.

Se comparado com uma pessoa caminhando, SR apresenta a vantagem de coleta de dados de grandes áreas de maneira mais rápida que, quando aplicado em grandes escalas, podem reduzir o custo dessa tarefa se tornando economicamente viável. Por esse motivo, essa ferramenta é adotada na Agricultura de Precisão que, por sua vez, pode ser entendida como um conjunto de técnicas e metodologias que visam otimizar o manejo de cultivos e a utilização dos insumos agrícolas, proporcionando máxima eficiência econômica (MAPA, 2009). Aplicações de SR se concentram no monitoramento da cobertura terrestre de áreas do meio ambiente, urbanas e agrícolas, não se limitando a áreas agrícolas, mas também em diversos outros contextos. Porém, principalmente na agricultura, drones têm surgido como uma estratégia para obtenção de informações importantes por meio de um custo reduzido (JORGE; INAMASU, 2014). Segundo FAO (2018), a importância de SR com drones é por causa da possibilidade de auxiliarem no aumento da produção agrícola e, conseqüentemente, na crescente demanda global por alimentos.

Imagens são uma das principais formas de representação de dados obtidos por SR e são geradas a partir de processos específicos aplicados na energia eletromagnética capturada por sensores (LILLESAND; KIEFER; CHIPMAN, 2015). Essas imagens permitem obter informa-

ções importantes sobre as áreas analisadas e são constituídas por bandas espectrais, que, por sua vez, representam diferentes regiões do espectro eletromagnético. Conforme a quantidade e tipo de bandas das imagens, elas podem ser caracterizadas como *red-green-blue* (RGB), multiespectral e, em especial, hiperespectral. Imagens hiperespectrais (IHs) têm diversas bandas adicionais em relação aos outros tipos, caracterizando um aspecto de continuidade e maior abrangência no espectro eletromagnético. Essas bandas adicionais podem ter maior capacidade discriminativa e, portanto, facilitar e melhorar a precisão na classificação de objetos presentes nas imagens (LANDGREBE, 2003; ATTAS et al., 2003; YAO; QIAN, 2009; KHAN et al., 2018). Especificamente na agricultura, IHs proporcionam aplicações como análises do tipo de vegetação (LILLESAND; KIEFER; CHIPMAN, 2004), condições de saúde (ATTAS et al., 2003), assim como outros fenômenos bióticos e abióticos (KHAN et al., 2018).

A análise de informações de IHs obtidas por SR pode ser facilitada com o uso de mapas de classificação (ou mapas temáticos). Um exemplo de uma IH da região de Indian Pines é apresentada na Figura 1.1, onde na sequência, da esquerda para a direita, é possível visualizar a imagem original, o mapa de classificação e uma legenda com os elementos do mapa. A imagem original apresentada é gerada por bandas RGB e o mapa de classificação é apresentado logo na fase de classificação e, portanto, apresenta alguns ruídos na imagem que posteriormente podem ser eliminados por técnicas de processamento de imagens.

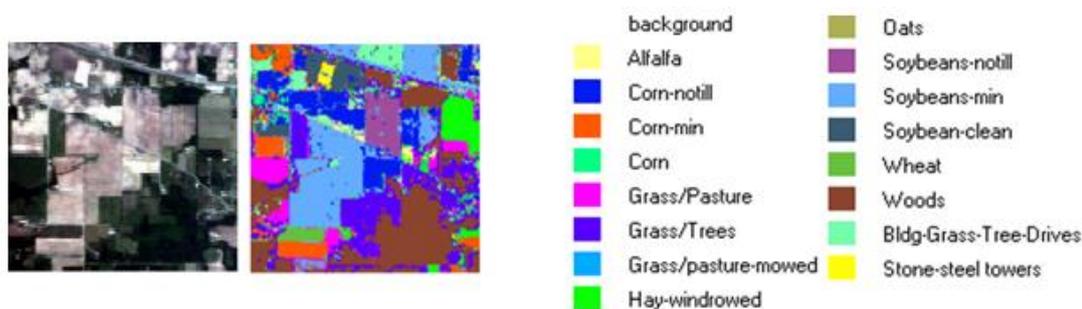


Figura 1.1. Exemplo de mapa de classificação da imagem Indian Pines. Adaptado de Xie et al. (2018).

Como pode ser observado no exemplo, mapas de classificação representam os elementos com propriedades diferentes (ou classes) das imagens por meio de cores distintas, sendo uma poderosa ferramenta para uma análise eficiente de áreas agrícolas ou regiões específicas em outros contextos. Mapas de classificação podem ser gerados por um sistema de reconhecimento de padrões (SRP), que conforme Duda, Hart e Stork (2001) é composto por cinco etapas:

1. Detecção de dados: Pode ser realizada por uma câmera. A dificuldade está nas limitações proporcionadas pelo ambiente que, por exemplo, podem ocasionar ruídos nos dados;

2. Segmentação: Todas as análises são realizadas sobre padrões¹ distintos, portanto, deve-se aplicar uma técnica de segmentação para permitir a distinção de diferentes padrões. No contexto de imagens, a segmentação é o processo de divisão da imagem em regiões constituintes (GONZALEZ; WOODS, 2018);
3. Extração e seleção de características: O objetivo dessa etapa é encontrar características que proporcionem o reconhecimento de objetos de uma mesma classe por valores de variáveis similares e permitir que objetos de classes distintas tenham valores diferentes;
4. Classificação: Consiste em utilizar as características extraídas ou selecionadas na etapa anterior para classificar um novo padrão;
5. Pós-Processamento: Considera o contexto do problema. Por exemplo, em mapas de classificação pode analisar a vizinhança de um pixel rotulado e alterar o seu valor com base na maioria dos valores evitando ruídos na imagem (HABERMANN, 2018).

Segundo Gonzalez e Woods (2018), se a classificação é realizada pixel por pixel, permite formar segmentos entre pixels de mesma classe conectados. Com a aplicação de técnicas de pós-processamento, ruídos internos a esses segmentos formados podem ser eliminados e regiões de mapas de classificação podem ser estabelecidas. Isso mostra que embora as etapas de SRP sejam normalmente apresentadas de forma sequencial, na prática, algumas etapas como segmentação e classificação podem ser complementares ou serem executadas concomitantemente.

1.1 Problemas e motivação

Embora a segmentação e IHS proporcionem as vantagens descritas anteriormente, existem desafios relacionados à alta dimensionalidade dos dados:

1. Sohaib, Haq e Mushtaq (2013) citam a alta demanda de recursos computacionais, como armazenamento, transmissão e capacidade de processamento. A captura de imagens com drones, podem gerar até 140 GB por Km², porém esses próprios equipamentos podem ter limitações de *hardware*. Essa limitação pode representar uma ameaça para futuras missões em tempo real com esses equipamentos (FAO, 2018);
2. Aumento da complexidade dos algoritmos de classificação relacionados com o número de dimensões (KUMAR, 2004);

¹Objetos que podem ser imagens, sinais ou outra forma de dados mensuráveis que precisam ser classificados (THEODORIDIS; KOUTROUMBAS, 2003).

3. Muitas bandas, principalmente adjacentes, podem ser altamente correlacionadas resultando em informações redundantes que afetam o desempenho de algoritmos (MARTINEZ-USO et al., 2007; DATTA; GHOSH; A., 2012; GONG; ZHANG; YUAN, 2016);
4. A maldição da dimensionalidade que pode afetar a acurácia de classificadores, sendo causada por limitações no número de amostras de treinamento para elevadas dimensões (FUKUNAGA, 1990; JIMENEZ; LANDGREBE, 1998; JAIN; DUIN; MAO, 2000; LANDGREBE, 2003);
5. Alto custo da produção de sensores e dispositivos para processar as IHS, impossibilitando o acesso para pequenos e médio empreendedores. Com bandas selecionadas pode-se projetar sensores adequados para aplicações específicas (BRIS et al., 2014).

Devido a esses problemas, uma das etapas mais importantes para a geração de mapas de classificação por meio de um SRP é a extração/seleção de características, principalmente em casos de alta dimensionalidade como em IHS. Quando o espaço de características contém somente as características relevantes, o classificador será mais rápido e ocupará menos memória. Além disso, um espaço de características pequeno com um conjunto de amostras de treinamento limitado, pode evitar a maldição da dimensionalidade proporcionando erros menores ao classificador (JAIN; DUIN; MAO, 2000).

Métodos de extração de características utilizam procedimentos matemáticos para realizar uma modificação no espaço de características para uma dimensão menor. Essa nova dimensão pode ser constituída por uma combinação das características do conjunto original (YAO; QIAN, 2009; PATRA; MODI; BRUZZONE, 2015). Um método comum para essa finalidade é a Análise de Componentes Principais (*Principal Component Analysis - PCA*), sendo utilizado em IHS nos trabalhos de Rodarmel e Shan (2002) e Agarwal et al. (2007), e que realiza uma transformação das bandas para um novo conjunto de características (componentes) não correlacionadas. Um problema que pode ocorrer nos métodos baseados na extração de características é que no processo de transformação para um novo espaço, dados intrínsecos podem ser distorcidos e a informação física ser perdida em caso de serem selecionadas somente alguns elementos do novo espaço (ZHANG; GONG; CHAN, 2018).

Em contrapartida à extração de características, a seleção de características (ou seleção de bandas para IHS) tenta encontrar um subconjunto a partir do conjunto original (GONG; ZHANG; YUAN, 2016). A seleção deve preservar as bandas que contêm informações importantes das IHS, de forma a eliminar somente aquelas que apresentem informações redundantes ou inúteis para uma aplicação específica. A vantagem da seleção de bandas é que a represen-

tação do espaço original não é alterada e, portanto, é preferível quando o significado original dos dados é necessário (MASAELI; FUNG; DY, 2010). Para IHS, há o interesse em manter o significado original das bandas para construção de sensores específicos. Na agricultura, por exemplo, bandas específicas podem também auxiliar em pesquisas para a exploração de novos índices de vegetação (JORGE; INAMASU, 2014).

Uma abordagem direta para tarefa de seleção de bandas é examinar todas as combinações possíveis selecionando o subconjunto que melhor otimiza o critério estabelecido. Porém, isso faz com que a quantidade de subconjuntos possíveis constituídos por diferentes combinações seja exponencial com relação ao número de bandas inviabilizando a busca exaustiva (GRECO et al., 2006). Devido a isso, uma série de métodos de seleção de bandas são explorados na literatura e uma forma de organizá-los é por meio da categorização em estratégias do tipo *Filter*, *Wrapper* ou *Embedded* (MONTEIRO; MURPHY, 2011).

Ainda segundo Monteiro e Murphy (2011), métodos baseados na estratégia *Embedded* analisam a importância de cada uma das bandas (variáveis) com algum índice como, por exemplo, o ganho de informação e, então, estabelecem o modelo de classificação. Durante o processo, esses métodos podem ou não descartar as bandas que menos contribuem no modelo. Algoritmos baseados em árvores de decisão são um exemplo de método baseado na estratégia *Embedded*.

Métodos baseados na estratégia *Filter* utilizam métricas para a avaliação de variáveis ou subconjunto de variáveis antes da construção do modelo. Após a seleção, constroem o modelo (TUV et al., 2009). A maioria dos métodos de seleção de bandas encontrados nesta pesquisa pertencem a essa categoria e alguns exemplos são baseados em algoritmos evolucionários (AE) ou inteligência de enxame (GONG; ZHANG; YUAN, 2016; ZHANG; MA; GONG, 2017).

Métodos baseados em AE, como os Algoritmos Genéticos (AGs), podem ser aplicados na forma da estratégia *Wrapper* conforme mostrado em Zhuo et al. (2008) e Saqui et al. (2016). Por meio dessa estratégia, os métodos podem incorporar um modelo classificador qualquer ao processo de seleção de bandas, e utilizam um índice preditivo como a acurácia para que iterativamente ajustem o modelo e avaliem os subconjuntos de bandas. A diferença entre as estratégias *Wrapper* e *Embedded* é que na primeira o classificador não depende das variáveis escolhidas para ser estabelecido podendo usar todas elas, enquanto que na segunda estratégia o modelo de classificação depende exatamente da escolha de variáveis para ser construído. Por incorporarem o classificador ao processo de seleção de bandas, métodos baseados na estratégia *Wrapper* normalmente apresentam desempenho de classificação superior em relação aos métodos da estratégia *Filter*, pois direcionam as bandas selecionadas diretamente para um determinado classificador.

Além das estratégias *Filter*, *Wrapper* ou *Embedded*, métodos de seleção de bandas podem ser organizados em supervisionados (também semi-supervisionados) ou não supervisionados, ou até mesmo com base na natureza do algoritmo em que o método foi proposto. Juntamente a essas estratégias, alguns recursos/componentes são comuns em métodos da literatura que promoveram bons resultados de classificação.

Independentemente da estratégia adotada, a maioria dos métodos de seleção tenta encontrar um conjunto reduzido que melhora um determinado índice relacionado de forma exclusiva com a informação espectral. Porém, a informação espacial tem sido incorporada ao processo de seleção de bandas e classificação de IHS conforme mostrado nos trabalhos de Cao et al. (2016) e Yuan, Lin e Wang (2016). Essa informação pode diminuir a necessidade de uma quantidade elevada de bandas na classificação proporcionando melhores resultados quando comparados com a não utilização. Liang et al. (2017) argumentam que um problema encontrado em estudos como, por exemplo, dos métodos baseados na obtenção de informação espacial por meio de estratégias de filtragem, têm sido avaliados experimentalmente de forma inadequada. Isso ocorre porque selecionam amostras aleatoriamente para treinamento e testes, sendo que essas podem ser vizinhas e, a informação espacial de cada uma das amostras ter influência sobre a outra. Portanto, a utilização dessa informação espacial pode ser considerada no processo de seleção de bandas e classificação de IHS, porém experimentos adequados são necessários.

Embora promissores para a melhoria da classificação, uma característica de muitos AEs para seleção de bandas é que eles conduzem o processo utilizando a melhoria de um único índice ou métrica como, por exemplo, uma função baseada na precisão² da classificação dos pixels. De outra forma, podem tentar combinar esse índice com a redução do número de bandas, utilizando a técnica de soma ponderada, conforme demonstrado em Zhuo et al. (2008). Um problema dos algoritmos com uma função com um único valor (chamada de função de *fitness* ou função objetivo), é que eles não conseguem operar adequadamente para o melhoramento simultâneo de mais de um índice como, por exemplo, o aumento da qualidade de segmentação e redução da quantidade de bandas. Esses dois objetivos ainda podem ser conflitantes pois, se bandas que contribuem para segmentação forem erroneamente eliminadas, o desempenho de classificação será afetado. Portanto, faz-se necessária a exploração de estratégias que permitam a busca por soluções com objetivos balanceados, chamadas de soluções de *tradeoff*, ou de compromisso. Essas soluções podem ser pensadas com a característica de equilíbrio nos índices que são considerados para estabelecer valores de *fitness*. Para essa finalidade podem ser aplicados métodos baseados na seleção de bandas por otimização multiobjetivo (*Multiobjective*

²Também referenciada como "Exatidão Global" ou "Acurácia" $\frac{1}{n} \sum_{i=1}^n (1 \times (p_i = \hat{p}_i))$, onde p_i é a classe estimada para o pixel i , \hat{p}_i é a classe real e n é a quantidade de amostras/pixels.

optimization band selection - MOBS).

MOBS buscam por soluções de *tradeoff* e têm sido aplicados para lidar com objetivos conflitantes em seleção de bandas de IHS (GONG; ZHANG; YUAN, 2016). Em comparação com a abordagem *Embedded*, os MOBS, se modelados como *Filters* ou *Wrappers*, permitem controlar o melhoramento de mais de um índice mesmo quando conflitantes. Técnicas como o C4.5 para árvores de decisão (*Embedded*) podem eliminar algumas variáveis irrelevantes, porém não buscam o equilíbrio entre a redução de variáveis e o desempenho do classificador. Em relação à estratégia *Wrapper*, os métodos baseados na estratégia *Embedded* ainda tornam o processo de seleção de características dependente de um modelo de classificação específico.

Apesar da elevada demanda computacional para execução de métodos baseados na estratégia *Wrapper*, após encontrar as bandas e construir o modelo classificador, o processo de seleção não é repetido. As bandas selecionadas e o modelo construído podem ser utilizados para produção de mapas de classificação. Além disso, resultados melhores durante a geração desses mapas podem ser obtidos quando considerada a informação espacial. Baseado nos problemas e métodos previamente apresentados, nesta tese é proposto e investigado experimentalmente, um método para seleção de bandas supervisionado por otimização multiobjetivo que incorpora um modelo classificador, ou seja, o método é baseado na estratégia *Wrapper*.

Uma característica comum nas estratégias baseadas na fronteira de Pareto (FP), como as previamente discutidas como MOBS, é que elas podem retornar mais de uma solução. Portanto, podem ser necessários métodos adequados ou a participação de um especialista para a seleção de uma solução dessa fronteira. As principais propostas para escolher uma solução da FP são caracterizadas por um tomador de decisões (*Decision Maker* - DM) automático. Esse DM pode ser baseado no centroide, como proposto por Kimovski et al. (2018) ou também na correlação entre soluções da fronteira e pontos ideais, como em Xu, Shi e Pan (2017), além de outras estratégias. Portanto, a utilização de um DM também é proposta e avaliada.

O método proposto inova ao combinar a seleção de bandas multiobjetivo junto a estratégia *Wrapper* com o propósito de simultaneamente melhorar a classificação enquanto faz a redução da quantidade de bandas. Além disso, outra inovação é que o método proposto combina outras estratégias importantes, tais como:

- a inicialização do conjunto de soluções;
- uso de informações da vizinhança de pixels de amostras em cada banda;
- DM automático para seleção de de uma única solução da FP.

Essa combinação não foi explorada na literatura, porém quando exploradas individualmente tendem a apresentar resultados positivos. Portanto, surge o questionamento se:

- Se o método proposto que é baseado nessa combinação de técnicas pode apresentar bons resultados para seleção e redução da quantidade de bandas de IHS e melhora de qualidade na classificação de pixels?

Diante dessa discussão, na sequência são apresentados os objetivos desta tese.

1.2 Objetivos, hipóteses e principais contribuições

Nesta tese é formulado um método chamado de *Wrapper multiobjective evolutionary band selection* (WMOEBS) para seleção de bandas por otimização multiobjetivo baseado na estratégia *Wrapper* e que incorpora um DM. O WMOEBS tem o objetivo de encontrar um conjunto de soluções com a dimensionalidade dos dados de IHS reduzida contribuindo para a classificação de pixels. Especialmente também é considerada a utilização da informação espacial em conjunto à espectral durante o processo de seleção e avaliado seu impacto na classificação com bandas selecionadas. Portanto, esta tese teve como objetivo principal:

- Demonstrar que o método WMOEBS (*Wrapper Multiobjective Evolutionary Band Selection*) consegue selecionar um conjunto otimizado de bandas de IHS, que promova alto índice de classificação usando informação espacial e espectral.

Considera-se boa qualidade do conjunto de bandas otimizado para resultados que são similares ou melhores aos proporcionados por métodos avançados ou atuais disponíveis na literatura.

1.2.1 Objetivos específicos

- Constituir um novo *framework* baseado na estratégia *Wrapper* e otimização multiobjetivo para seleção de bandas de IHS que funcione como base estrutural para o método proposto;
- Elaborar estratégias de seleção de solução da FP em problemas multiobjetivos;
- Adequar estratégias da literatura como possíveis componentes e elaborar o WMOEBS;
- Demonstrar a influência do uso da informação espacial na seleção de bandas; e
- Proporcionar a produção de artigos científicos com avaliações e testes comparativos de diferentes algoritmos para seleção de bandas e classificação de pixels de IHS.

1.2.2 Hipóteses

Considerando a necessidade de seleção de bandas de IHS e a possibilidade de melhoramento nos resultados de classificação dos pixels das imagens geradas com bandas selecionadas, este estudo investiga as seguintes hipóteses:

- A utilização da informação espacial em conjunto com o processo de seleção de bandas pode promover melhores resultados que a utilização somente da informação espectral na tarefa de classificação dos pixels com bandas selecionadas;
- O WMoEBS permite reduzir a dimensionalidade dos dados de IHS e as bandas selecionadas melhoram o desempenho da classificação de pixels de IHS quando comparado com as imagens originais; e
- As bandas selecionadas pelo WMoEBS melhoram o desempenho da classificação de pixels de IHS quando comparado com bandas selecionadas por outros métodos de seleção que estão no estado da arte na literatura e mantém uma quantidade reduzida de bandas.

1.3 Principais contribuições e delimitação de escopo

As principais contribuições desta tese são listadas a seguir:

- Criação de um método de seleção de bandas multiobjetivo baseado na estratégia *Wrapper*.
- Criação de um *framework* geral para seleção de bandas por otimização multiobjetivo baseado na estratégia *Wrapper*.
- Avaliação da utilização de forma adequada da informação espacial na seleção de bandas e na classificação dos pixels de IHS.
- Métodos e avaliações de possíveis tomadores de decisão (*DM*) para selecionar uma solução da FP.
- Elaboração de um método de inicialização do conjunto de bandas com diferentes combinações que reduz a correlação e custo computacional de busca.
- Comparações com diferentes métodos da literatura e com IHS com todas as bandas.

Uma ilustração do contexto prático onde a pesquisa desta tese pode ser aplicada é ilustrada na Figura 1.2, onde é representado um sistema de monitoramento por SR organizado em duas etapas que são definidas a seguir:

- Uma etapa (Figura 1.2a - destacado em amarelo) para selecionar bandas e gerar o modelo de classificação constituídos de bandas relevantes para determinadas aplicações. A etapa de selecionar bandas é onde essa pesquisa pode auxiliar com o método proposto;
- e uma segunda etapa para fazer a classificação de IHS com bandas selecionadas (Figura 1.2b - em amarelo), onde essa pesquisa tem o objetivo de contribuir com o melhoramento da classificação por meio das bandas selecionadas e informação espacial.

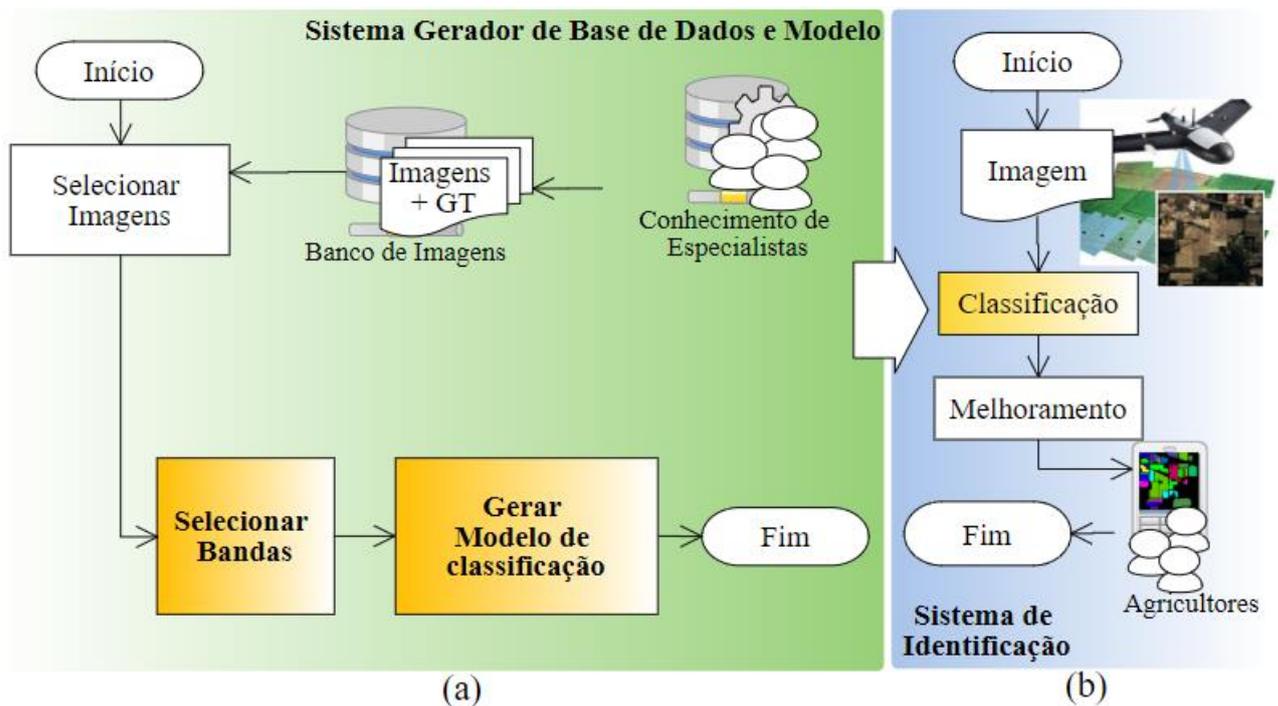


Figura 1.2. Aplicação prática do método proposto.

Uma importante observação é que embora a agricultura seja uma área onde existe uma recente expansão para o uso de IHS o método proposto não se limita a essas área, mas pode ser explorado em qualquer área de aplicações de IHS ou até mesmo, quando adaptado, em qualquer tipo de dados de alta dimensionalidade.

A delimitação do escopo envolve principalmente os seguintes itens:

- Não é considerada a instalação do método elaborado em um dispositivo acoplado em drones, pois a geração do modelo de classificação e validação do mesmo é suficiente para os experimentos demonstrarem a qualidade do método proposto;
- Os problemas do custo computacional relacionados ao tempo de classificação de IHS e a maldição da dimensionalidade são conhecidos na literatura, portanto também não são explorados neste estudo;

- A validação do método proposto considera a quantidade de bandas e a qualidade da classificação dos pixels de IHS. O tempo de processamento do processo de busca não é avaliado, pois entende-se que mesmo que possa ter um elevado custo durante a busca, após as bandas serem encontradas, elas podem ser utilizadas diretamente em um modelo de classificação treinado.
- A contribuição e avaliação individual de cada banda também não é considerada pois não é abordado uma aplicação específica, mas sim, avaliado se o método consegue selecionar um o conjunto inteiro das bandas otimizado que posteriormente proporcione boa classificação.

1.4 Organização do trabalho

Neste capítulo foram apresentados o contexto desta pesquisa, a motivação, a justificativa e os objetivos deste trabalho. O restante do trabalho está organizado da seguinte forma:

- No Capítulo 2, são apresentados os fundamentos de IHS, assim como técnicas que permitem sua utilização como classificação e segmentação. Esse e o capítulo seguinte são complementares às teorias apresentadas na Introdução;
- No Capítulo 3, é apresentado o problema de seleção de bandas em IHS, assim como os aspectos teóricos da redução da dimensionalidade, além dos métodos relacionados ao tema da pesquisa (estado da arte);
- No Capítulo 4, é apresentada a metodologia de pesquisa utilizada, considerando as amostras utilizadas nos experimentos, procedimentos específicos da pesquisa e explicações sobre os experimentos realizados;
- No Capítulo 5, o *framework* que foi utilizado como base para estruturar o WMoEBS e os principais componentes avaliados para compor esse método são apresentados. Também são mostrados os experimentos iniciais onde possíveis métodos foram avaliados para compor o WMoEBS, assim como também é apresentado o WMoEBS em si;
- No Capítulo 6, são apresentados os experimentos comparativos com outros métodos da literatura realizados, assim como a discussão dos resultados obtidos;
- No Capítulo 7, são realizadas as considerações finais da tese, apresentadas as limitações encontradas e discutidos os trabalhos futuros.

Capítulo 2

CONCEITOS DE IMAGENS HIPERESPECTRAIS

A contribuição desta pesquisa ocorre principalmente na etapa de seleção de características que, para o contexto desta tese, são bandas de imagens hiperespectrais (IHs). Além disso, informações espaciais e espectrais das bandas selecionadas que devem promover bom desempenho nas tarefas de classificação e segmentação também são consideradas.

Para o entendimento do comportamento das informações espaciais e das bandas espectrais na tarefa de seleção são necessários conhecimentos que vão além de sistemas de reconhecimento de padrões (SRP). Portanto, neste capítulo, são apresentados aspectos teóricos sobre sensoriamento remoto (SR) e IHs que fazem parte das etapas de aquisição de dados em SRP e são importantes para a contextualização do tema abordado. Além disso, são apresentados métodos de classificação e segmentação de IHs sendo alguns específicos para essas imagens.

2.1 Sensoriamento Remoto e Imagens Hiperespectrais

SR é a ciência de obtenção de informações sobre objetos, áreas, ou fenômenos por meio de dados adquiridos por dispositivos (sensores) que não estão em contato com os objetos ou áreas que estão sendo observados (LILLESAND; KIEFER; CHIPMAN, 2004). Na Figura 2.1 é representado um processo de SR, onde pode-se observar etapas semelhantes às existentes em sistemas de SRP apresentadas na introdução. Nessa Figura, um sensor em um drone é responsável pela coleta de dados da superfície terrestre, que depois são processados para que possam ser analisados. O objetivo da etapa de análise é o entendimento das informações geradas por meio dos dados capturados, e pode ser composta por métodos automatizados com ou sem a participação de um especialista. No final, as informações são disponibilizadas para uso (LANDGREBE, 2003).

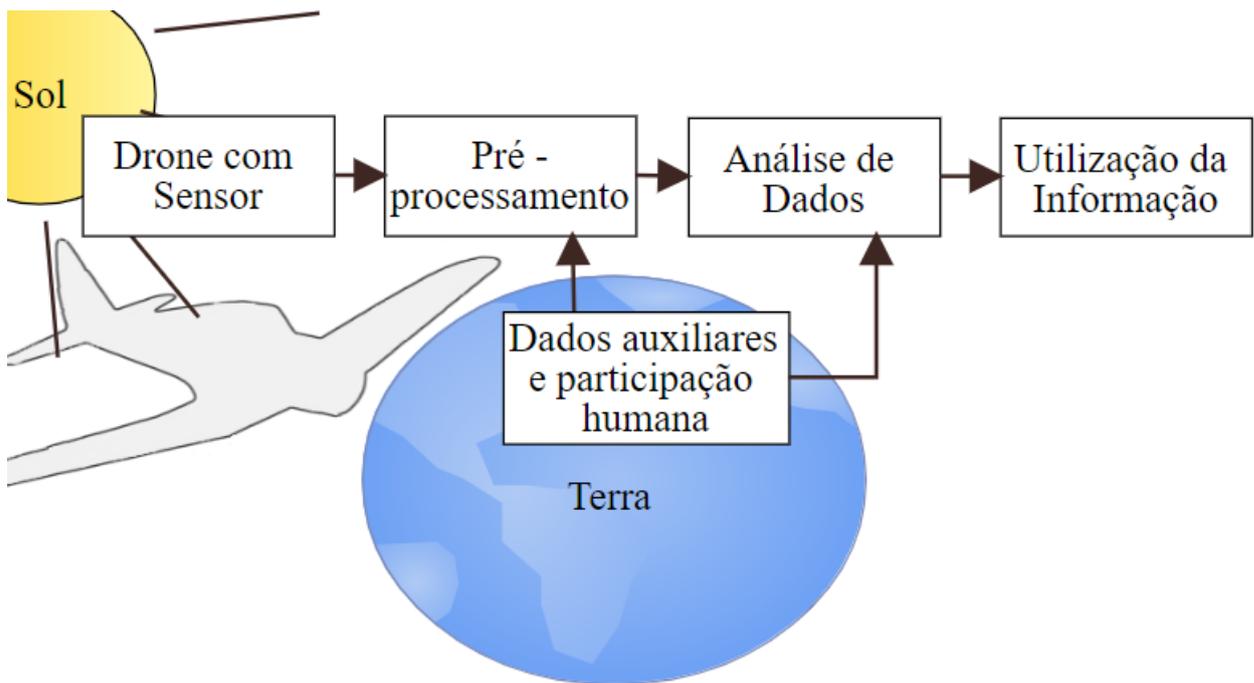


Figura 2.1. Sensoriamento remoto (adaptado de Landgrebe (2003)).

SR tem sido aplicado em diferentes áreas como geologia, agricultura, planejamento urbano, entre outros (LILLESAND; KIEFER; CHIPMAN, 2015). A expansão de dispositivos modernos como drones tem possibilitado a aquisição de imagens de maior resolução e frequência que dispositivos tradicionais como satélites (KULBACKI et al., 2018). Isso tem promovido um rápido crescimento de aplicações de SR em diferentes áreas (TERWILLIGER et al., 2017). Drones podem ter acoplados sensores que capturam imagens digitais coloridas em RGB (*red-Vermelho, Green-Verde, Blue-Azul*), multiespectrais ou hiperespectrais.

Gonzalez e Woods (2018) definem uma imagem digital como uma função bidimensional $f(x,y)$ onde x e y são coordenadas espaciais e a amplitude de f é o nível de cinza da imagem naquele ponto. As variáveis x , y e a função f são finitas para quantidades discretas. A função $f(x,y)$ também pode ser representada como uma superfície no espaço (x,y,z) , onde z é a altura na posição (x,y) .

Imagens coloridas no espaço RGB, conforme representado na Figura 2.2, são formadas pelo uso dos componentes vermelho (z_1), verde (z_2) e azul (z_3) que podem ser organizados na forma de um vetor $z = [z_1, z_2, z_3]^T$. O caso geral de imagens multiespectrais envolve d componentes, ou seja, um vetor resultante $z = [z_1, z_2, z_3, \dots, z_d]^T$ com d dimensões.

A diferença entre IHS e imagens multiespectrais, está na quantidade de bandas e o espaçamento da faixa espectral em que as mesmas são obtidas. Enquanto imagens multiespectrais normalmente têm entre 5 e 12 bandas com espaçamento maior no espectro eletromagnético, as

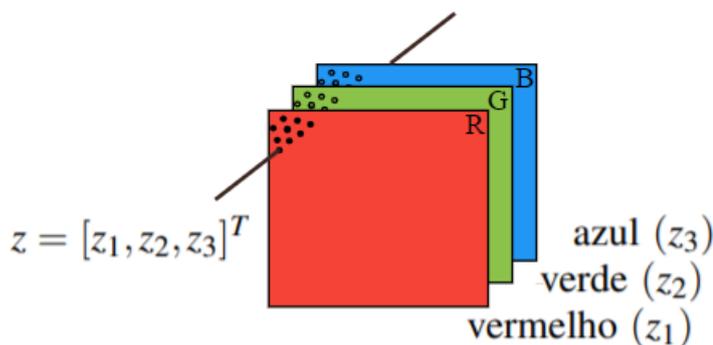


Figura 2.2. Componentes (bandas) de uma imagem RGB (adaptado de Gonzalez e Woods (2018)).

IHs são constituídas de centenas ou milhares de bandas espaçadas de forma mais estreita (5-20 nm). A Figura 2.3 representa a diferença entre o imageamento multiespectral (A) e hiperespectral (B), onde a luz visível abrange apenas uma pequena fração do espectro (entre 400-700 nm). Com a pequena fração do espectro visível, é notável que grande parte das informações no espectro não são exploradas por imagens com somente bandas na faixa visível.

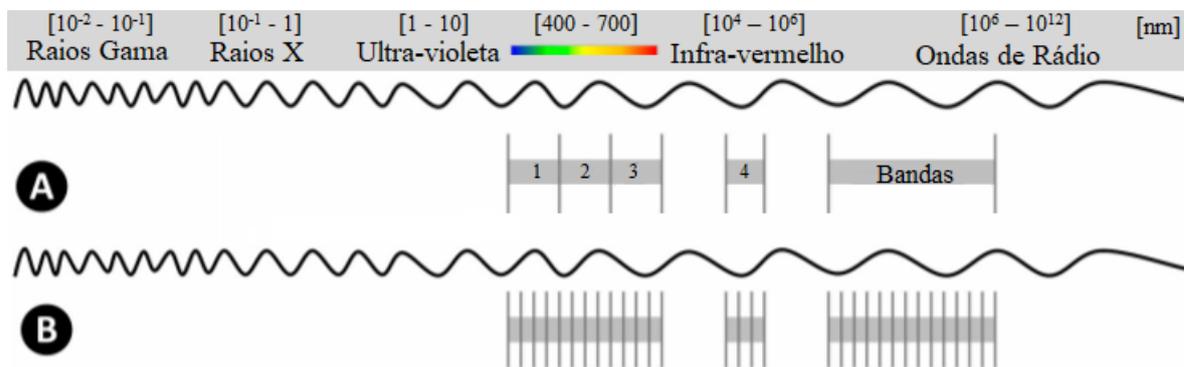


Figura 2.3. Imageamento multiespectral (A) e hiperespectral (B) (adaptado de Habermann (2018)).

Devido à maior resolução espectral nas IHs, ou seja, espaçamento de bandas mais estreito, informações com mais detalhes em cada pixel estão disponíveis. As bandas de cada pixel constituem a assinatura espectral (THENKABAIL et al., 2014), que pode ser visualizada na parte inferior da Figura 2.4, representada por uma curva que caracteriza o comportamento espectral da área monitorada e possibilita a discriminação de objetos com precisão (MANOLAKIS; MARDEN; SHAW, 2003). Na mesma figura também pode-se observar, no canto superior direito, a representação do cubo espectral constituído por um conjunto de bandas e cada pixel que é um vetor que caracteriza a assinatura espectral. Como os dados se manifestam em intervalos maiores sobre o espectro em imagens multiespectrais, elas não atingir o mesmo nível de detalhe.

As informações adicionais das IHs elevam a capacidade de distinção correta de alvos e

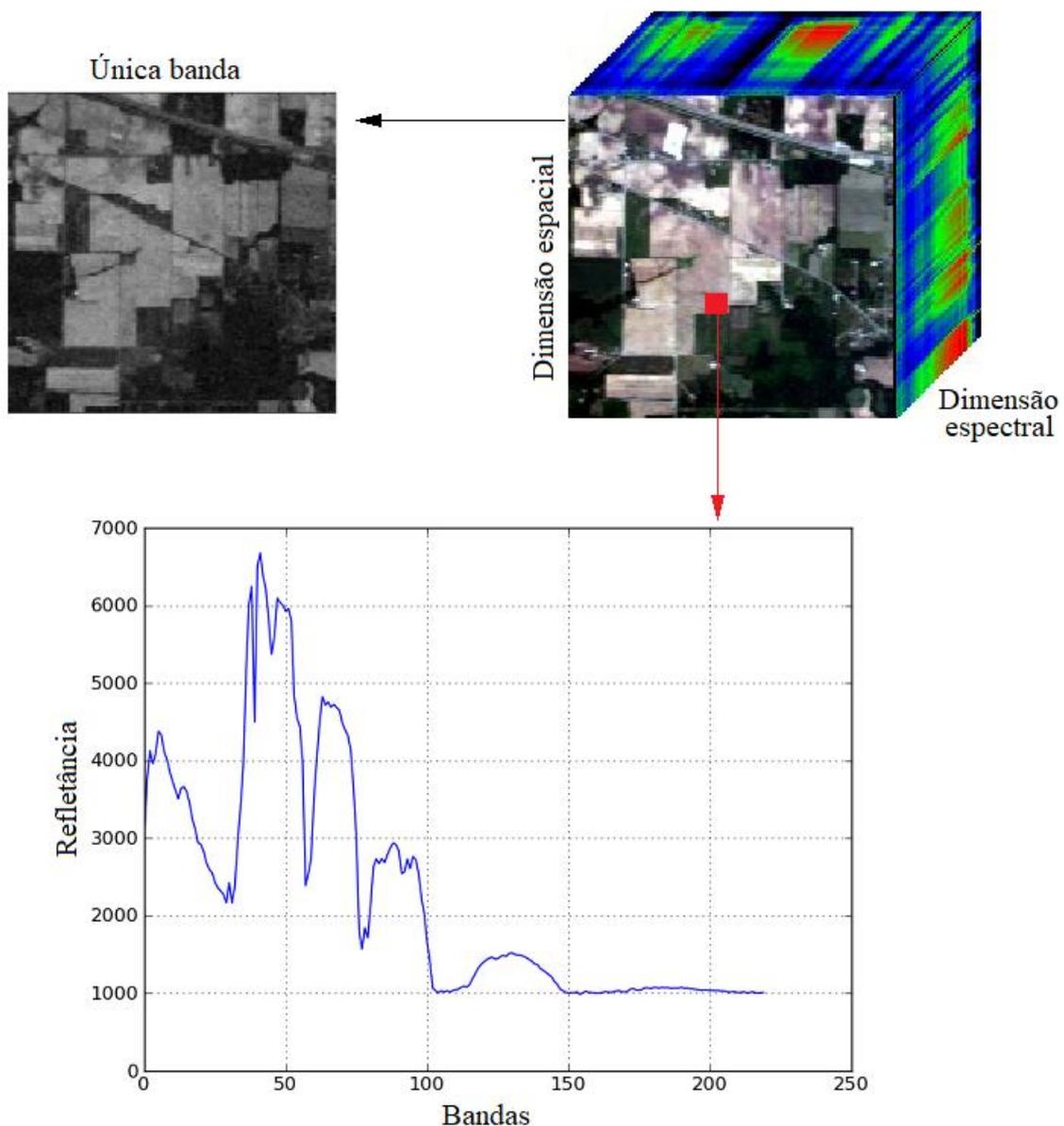


Figura 2.4. Cubo hiperespectral e assinatura espectral (adaptado de Habermann (2018)).

possibilitam muitas aplicações em diferentes áreas. Exemplos dessas áreas são a Mineralogia (KLIMA et al., 2014), Agricultura de precisão (QIN et al., 2013; MURUGAN et al., 2016), Geologia (MURPHY; MONTEIRO; SCHNEIDER, 2012), Ecologia (LU; HU; GUO, 2009), Hidrologia (KLEMAS, 2014), Aplicações militares (HUA; GUO; LIU, 2015; WANG et al., 2018), Diagnósticos médicos (CHANG; LI; LI, 2018), entre outras.

Junto com a melhoria da resolução espectral das IHS, seu processamento também tem a complexidade aumentada, já que essas imagens têm muitas bandas que podem ser difíceis de serem manipuladas em tempo real (ADAO et al., 2017). Essa complexidade e quantidade de bandas levam a duas questões importantes:

- Características da dimensionalidade de dados de IHS que convergem para os problemas que foram elencados no capítulo de introdução e são relacionados à correlação, e ao fenômeno de Hughes, entre outros; e
- Algoritmos para classificação e segmentação adequados que são utilizados para aplicações específicas com IHS e SR.

Detalhes sobre essas questões são apresentados na sequência.

2.2 Características sobre a alta dimensionalidade de dados

Devido à alta dimensionalidade das IHS, características importantes devem ser consideradas para aplicações que necessitam da tarefa de classificação e segmentação automática. Uma dessas características descreve a necessidade de uma elevada quantidade de pixels rotulados para amostras de treinamento, que normalmente são apresentados por meio de mapas de *Ground Truths* (GTs). Se a redução da dimensionalidade não é realizada, esses classificadores, quando aplicados em IHS, podem ter seu desempenho degradado. Esse fenômeno pode ser explicado pela maldição da dimensionalidade (também chamada de fenômeno de Hughes ou fenômeno de pico) descrita como segue:

- "Se um número finito de amostras de treinamento é usado, o erro de classificação tende a aumentar a partir de um determinado número de dimensões"(HUGHES, 1968; FUKUNAGA, 1990; JAIN; DUIN; MAO, 2000; THEODORIDIS; KOUTROUMBAS, 2003).

Outros estudos sobre propriedades geométricas, estatísticas e assintóticas da distribuição dos dados de alta dimensionalidade são agrupados no trabalho de Jimenez e Landgrebe (1998). Um desses estudos, realizado por Scott (1992), mostra que conforme a dimensionalidade dos dados aumenta, o volume do hipercubo se concentra nos cantos e o volume de uma hiperesfera (ou hiperelipsoide) se concentra na camada externa (borda). Baseado na distribuição desses volumes, algumas observações foram realizadas por Jimenez e Landgrebe (1998) e Benediktsson e Ghamisi (2015):

- Um espaço de alta dimensão é esparsos, o que implica que os dados multivariados geralmente estão em uma estrutura dimensional inferior, ou seja, dados de alta dimensão podem ser projetados em um subespaço inferior sem perdas de informações em termos de separabilidade entre diferentes classes estatísticas; e

- Dados com distribuições Gaussianas têm tendência de se concentrar nas caudas e dados uniformemente distribuídos têm uma tendência de se concentrar nos cantos, o que dificulta a estimativa de densidade de dados de alta dimensão para ambas as distribuições.

Fukunaga (1990) mostra que existe uma relação entre a quantidade necessária de amostras de treinamento e o número de dimensões para diferentes tipos de classificadores. Esse número é linearmente relacionado à dimensionalidade dos classificadores lineares e ao quadrado da dimensionalidade dos classificadores quadráticos. Para classificadores não paramétricos, Scott (1992) e Hwang, Lay e Lippman (1994) mostram que conforme a dimensionalidade aumenta, a quantidade de amostras deve aumentar exponencialmente para se ter uma estimativa efetiva de densidades multivariadas.

Não somente essas características, mas outras também podem ser encontradas em Jimenez e Landgrebe (1998), que, em conjunto com os problemas apresentados na introdução desta tese, motivam a aplicação de métodos de redução de dimensionalidade e classificadores adequados para IHS.

2.3 Classificação e segmentação em Imagens Hiperespectrais

A segmentação consiste na subdivisão da imagem em suas regiões homogêneas que contêm pixels com propriedades em comum (GONZALEZ; WOODS, 2018). Essa técnica é um dos importantes recursos para análise automática de imagens, pois possibilita a identificação de regiões que representem objetos de interesse (RUSS, 1998). Além disso, é também uma das etapas preliminares para diversas tarefas de sistemas de visão computacional (VC), tais como classificação, contagem, reconhecimento e outras tarefas. Gonzalez e Woods (2018) ainda descrevem que a segmentação pode ser baseada em duas propriedades básicas:

1. *descontinuidade*: particiona a imagem em regiões com base nas variações bruscas de intensidade, sendo que essas variações podem caracterizar elementos como bordas;
2. *similaridade*: baseada no particionamento de uma imagem em suas regiões que são similares quanto ao conjunto de critérios predefinidos.

O reconhecimento de cada objeto de uma imagem segmentada é proporcionado por rótulos que podem ser caracterizados por diferentes cores falsas ou outras formas de representações. Pixels ou regiões que pertencem a objetos diferentes têm seus respectivos rótulos distintos.

Conforme previamente apresentado, uma imagem ou conjunto de dados hiperespectrais pode ser visto como uma pilha de vetores de pixels, onde cada pixel é denotado por $x = (x_1, \dots, x_d)^T$, sendo d a quantidade de bandas e também o tamanho do vetor. Cada característica ou banda espectral é uma dimensão do espaço de características d -dimensional. Um método de classificação tenta definir um pixel não conhecido a uma das y classes que são definidas sobre o conjunto de treinamento. Cada classe é estabelecida com base na similaridade em relação a outras classes, estabelecendo fronteiras de decisão no espaço de características.

Em SR, classificadores podem ser utilizados onde é considerada a existência de pixels rotulados, que são estruturados em um GT. Em um exemplo de IHS geográficos, o GT é constituído por diversas coberturas do solo (KUMAR, 2004), e sua elaboração é normalmente realizada com o auxílio de especialistas que conhecem a área que será monitorada. Em um primeiro momento, pode ser um pouco trabalhoso a construção do GT, mas é importante para promover a construção de modelos de classificação, que por sua vez, permitem a classificação automática de regiões.

Etapas comuns para classificadores aplicados em imagens de SR são apresentadas na Figura 2.5, onde uma imagem com 5 bandas é considerada. Nessa imagem, a primeira etapa consiste em definir o conjunto de coberturas do solo da imagem que serão classificadas, como no exemplo: água, terra e vegetação. Nessa etapa também devem ser escolhidos elementos representativos como dados de treinamento. Na fase de treinamento os dados são usados para construir um modelo classificador. O modelo construído pode então ser utilizado na fase de classificação de novos dados.

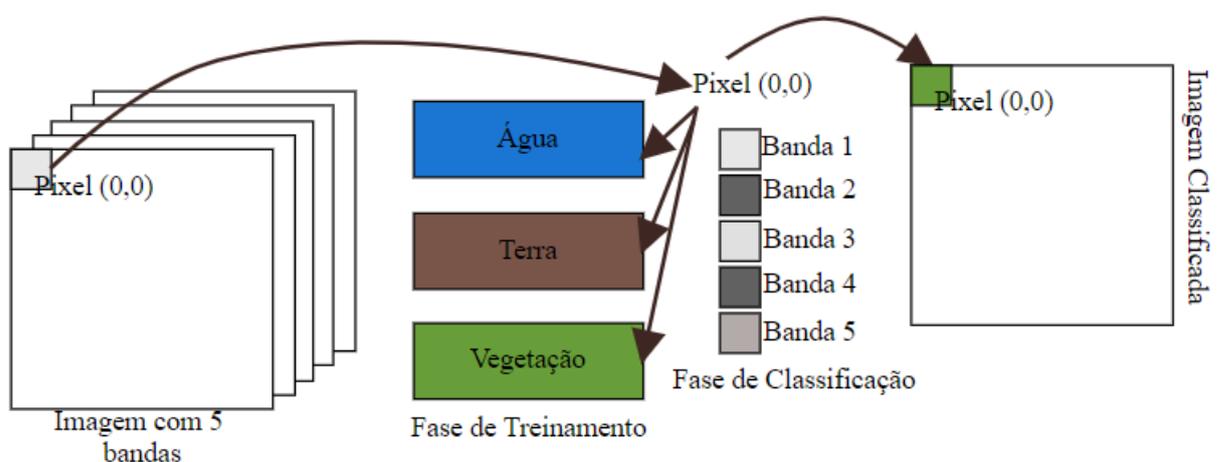


Figura 2.5. Etapas básicas da classificação de pixels de imagem com múltiplas bandas (adaptado de Lillesand, Kiefer e Chipman (2004)).

No exemplo da figura anterior, é representada a classificação do pixel na posição (0,0), onde os resultados são utilizados para gerar mapas de classificação. O pixel na posição (0,0) foi

classificado como "Vegetação"(cor verde). Por fim, o resultado obtido pode ser avaliado por meio de alguma estratégia (LILLESAND; KIEFER; CHIPMAN, 2004).

No Capítulo 1 - Introdução, Figura 1.1 foi mostrado uma parte da imagem como sendo o mapa de classificação da IH Indian Pines que foi segmentada com base na similaridade. O mapa é uma imagem composta por cores falsas que representam os diferentes segmentos. Para geração desse mapa foi aplicada uma estratégia chamada de classificação baseada em pixels (ou pixel por pixel), que é uma das abordagens mais utilizadas em análise de dados de SR. Esse tipo de classificação pode promover uma segmentação na imagem que particiona seus elementos em diferentes tipos de regiões (GONZALEZ; WOODS, 2018). Os segmentos são definidos por pixels conectados pertencentes a uma mesma classe. Nesse tipo de classificação, pixels podem ser analisados por sua informação espectral e/ou espacial (GHAMISI et al., 2017).

2.3.1 Exemplos de classificadores para imagens hiperespectrais

Em Ghamisi et al. (2017), uma revisão detalhada sobre classificadores que são mais utilizados em IHS é apresentada. Os autores fizeram um levantamento de trabalhos da literatura e destacam que os mais utilizados até a presente data são as Redes Neurais Artificiais (RNA), *Random Forests* (RF) e o *Support Vector Machines* (SVM).

O RF, além de promover a classificação também opera como um método de redução da dimensionalidade (baseado na estratégia *Embedded*) (SAQUI et al., 2019b). Estudos com métodos de seleção de bandas realizados por Gong, Zhang e Yuan (2016) e Xu, Shi e Pan (2017) utilizam outros classificadores, como os algoritmos *K-Nearest Neighbors* (KNN), *Extreme Learning Machine* (ELM) e SVM para avaliar o resultado da classificação promovido com as bandas selecionadas. Nesses trabalhos, os autores mostraram o bom desempenho de classificação do SVM, que foi superior ao KNN e equivalente ao ELM. O SVM mostra um desempenho notável quando comparado com outros métodos em termos de qualidade de classificação devido à sua propriedade robusta à dimensão espectral das IHS (CAO et al., 2017). Além disso, o SVM também é destacado em Ghamisi et al. (2017) onde foram realizados testes em IHS, onde esse método com *kernel Radial basis function* (RBF) e as RNAs-*multilayer perceptron* (com uma camada oculta) obtiveram precisão semelhantes entre si e ambas melhores que as RFs (com 300 árvores).

Ghamisi et al. (2017) também descrevem que algoritmos de classificação podem ser categorizados como paramétricos ou não-paramétricos. Os métodos não-paramétricos, como o SVM, não fazem nenhuma suposição sobre a distribuição dos dados utilizados como entrada e os métodos paramétricos assumem distribuições específicas nos dados de entrada.

Um exemplo de algoritmo classificador paramétrico que é muito utilizado para imagens de SR é o *Gaussian Maximum Likelihood Classifier* (GMLC), que é baseado na máxima verossimilhança. O GMLC também é utilizado em conjunto com métodos de seleção de bandas como mostrado em Cui et al. (2011) e Saqui et al. (2016), onde proporcionou bom desempenho após a redução do número de bandas. Em IHS, dados de amostras podem ser limitados e, portanto, métodos não paramétricos podem ser mais adequados no critério de classificação em relação a outra categoria, ou seja, métodos paramétricos (GHAMISI et al., 2017). Já suposições realizadas por Richards e Jia (2013) sugerem que distribuições gaussianas em dados de imagens de SR são bastante aceitáveis. Portanto, nesta tese foi considerado um exemplar de cada uma dessas categorias para a classificação de IHS e constituição do método explorado, que são o GMLC (paramétrico) e o SVM (não paramétrico) que são discutidos na sequência. Como em métodos da estratégia *Wrapper* o classificador é incorporado ao processo de busca e então modelado, é importante utilizar o mesmo classificador após a seleção de características/bandas para obter bons resultados, por isso é necessário uma escolha adequada.

2.3.2 *Gaussian Maximum Likelihood Classifier* (GMLC)

GMLC é um algoritmo paramétrico baseado na máxima verossimilhança e assume que os dados/amostras de treinamento de cada classe podem ser caracterizados por distribuições de probabilidade Gaussianas (distribuições Normais). Na Figura 2.6 é apresentado um gráfico de exemplo das distribuições de probabilidade para duas bandas (3 e 4) de uma imagem que contém 6 classes, sendo que cada classe é representada por uma distribuição de probabilidade. Nesse gráfico, o eixo vertical está associado à probabilidade de um valor de pixel pertencer a uma das classes, onde é perceptível que quanto mais próximo um pixel está do centro de uma determinada distribuição Gaussiana (ou seja, de uma classe), maior a chance dele pertencer a essa classe.

Como no GMLC é realizada a suposição que os dados têm distribuições Gaussianas, são considerados que as classes de um determinado conjunto de dados (como pixels de imagens) podem ser definidas por vetores de média e matrizes de covariância (KUMAR, 2004). Junto com os vetores de média e matrizes de covariância, no GMLC é utilizada uma função de probabilidade de densidade para classificar novos pixels (KUMAR, 2004; LILLESAND; KIEFER; CHIPMAN, 2004). A suposição de distribuições Gaussianas para classes em IHS é aceitável, pois espera-se que a maioria dos pixels de uma classe, ou seja, a maior densidade dos pixels de uma classe em uma imagem, se concentre em direção ao centro e conforme mais afastado desse centro, menor a chance dos pixels serem dessa classe (RICHARDS; JIA, 2013). Essa

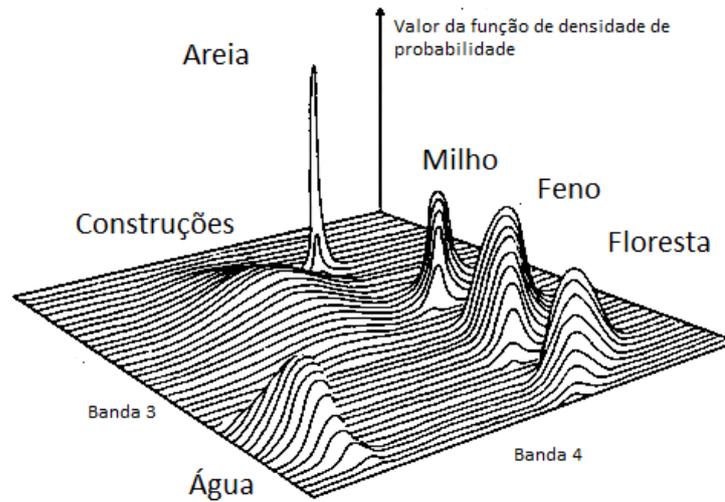


Figura 2.6. Exemplo das distribuições de probabilidades de classes de imagens nas bandas 3 e 4 (adaptado de Lillesand, Kiefer e Chipman (2004)).

característica colabora para a adequação do modelo Gaussiano nos pixels de imagens de SR.

Para classificação de IHS, dado um conjunto de pixels deve-se estabelecer o vetor de média μ_i e a matriz de covariância C_i , onde os valores são estabelecidos conforme os pixels de uma determinada classe ω_i , para cada $i = 1 \dots M$ onde M é o número de classes. Os valores obtidos são utilizados no treinamento de um classificador GMLC, onde essas métricas são utilizadas na função discriminante $g_i(x)$ do algoritmo representada na equação 2.1:

$$g_i(x) = -\ln|C_i| - (x - \mu_i)^T C_i^{-1} (x - \mu_i), \quad (2.1)$$

onde x é um pixel a ser classificado e \ln é o logaritmo natural.

A função $g_i(x)$ é utilizada na regra de decisão descrita na equação 2.2:

$$\forall (j \neq i), \quad x \in \omega_i \quad \text{se} \quad g_i(x) > g_j(x), \quad (2.2)$$

onde ω_i e ω_j são classes do conjunto considerado. A regra na equação 2.2 é utilizada para classificar o pixel x , por meio da verificação da maior probabilidade desse pixel pertencer a cada uma das classes consideradas.

Assim como outros classificadores paramétricos, o GMLC pode não classificar dados de IHS com boa acurácia se a quantidade de amostras é muito pequena em relação à quantidade de características. Isso acontece, principalmente, quando no processo de treinamento não é possível estimar corretamente estatísticas de primeira e segunda ordem (ou seja, média e covariância) fundamentais para operacionalização do algoritmo. Portanto, para tentar melhorar o

desempenho dessa categoria de algoritmos, métodos de redução da dimensionalidade podem ser necessários. Uma revisão completa sobre esses problemas relacionado a classificadores pode ser encontrada em Jimenez e Landgrebe (1998) e Fukunaga (1990).

2.3.3 Support Vector Machines (SVM)

Support Vector Machines (SVM) é um algoritmo supervisionado baseado na teoria do aprendizado estatístico (teoria Vapnik-Chervonenkis) projetado para tarefas de classificação (VAPNIK, 1995). O SVM busca um hiperplano ótimo como uma função de decisão em um espaço de características que pode ter muitas dimensões (BOSER; GUYON; VAPNIK, 1992; CRISTIANINI; SHAW-TAYLOR, 2000). Para essa otimização, o SVM introduz uma minimização do risco estrutural, do inglês *structural risk minimization* (SRM), considerando o melhor separador, aquele que minimiza o erro de generalização e tentando evitar problemas de *overfitting* (GUO et al., 2019).

O caso mais simples, envolve problemas com duas classes distintas onde os dados podem ser linearmente separáveis e que estão em um espaço de características bidimensional. Dado um número infinito de possibilidades para obter uma função de classificação, o SVM seleciona aquela que maximiza a margem que separa os chamados vetores de suporte que são dados representantes de cada uma das classes e presentes no conjunto de treinamento. Um exemplo de um hiperplano separador, margem e vetores de suporte é ilustrado na Figura 2.7.

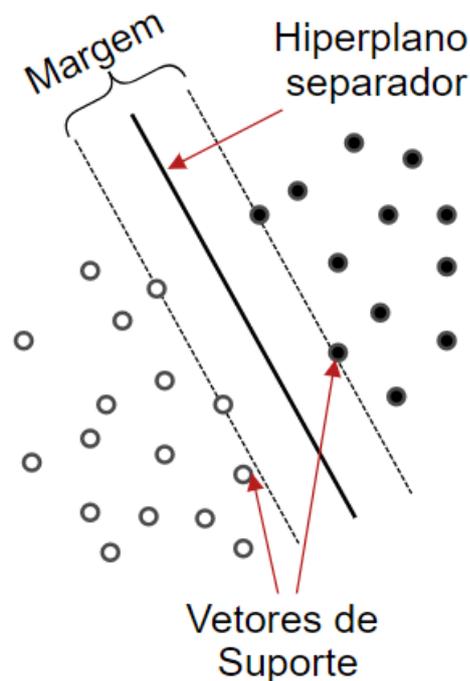


Figura 2.7. Representação do SVM com margem rígida.

A ilustração da Figura 2.7 representa um exemplo de SVM com margem rígida, onde são definidas fronteiras lineares a partir de dados linearmente separáveis. Porém, em problemas reais *outliers* são comuns e as SVM de margens suaves são mais adequadas de forma que introduzem um relaxamento. Esse recurso suaviza restrições impostas na determinação do hiperplano ótimo e admite a ocorrência de algumas amostras com classes possivelmente erradas (*outliers*). O relaxamento é controlado por um parâmetro C que determina como a margem é construída, operando como um ponto de equilíbrio entre a maximização da margem e minimização do erro de classificação, podendo ser mais ou menos tolerante a erros.

SVMs lineares são eficazes na classificação de conjuntos de dados com uma distribuição linear ou "quase"linear e a versão de margens suaves tolera a presença de *outliers*. Porém, existem situações onde os dados não são separáveis linearmente e portanto não se pode dividir corretamente o conjunto de dados de treinamento com um hiperplano. SVMs tratam problemas não lineares utilizando um truque de *kernel*, mapeando o conjunto de treinamento de seu espaço original para um novo espaço Euclidiano de maior dimensão (GHAMISI et al., 2017). Para isso, a função de *kernel*, caracterizada por uma função matemática, transforma o espaço de entrada em um espaço de dimensão maior, porém que permite que os dados possam ser separados linearmente (LORENA; CARVALHO, 2007). O truque do *kernel* é um dos principais recursos do SVM e em tarefas de classificação em IHS o *kernel radial basis function* (RBF) é o mais utilizado (NAGASUBRAMANIAN et al., 2017; GUO et al., 2019). O RBF permite estabelecer separações circulares ou em formato de curvas nas fronteiras de decisão. O controle dessas curvas de separações é estabelecido por um parâmetro γ onde valores maiores fazem com que a curva seja mais circular em torno de cada amostra e valores menores fazem separações com curvas mais suaves e abrangentes. Os parâmetros C (tolerância a erros) e γ (característica da curva) são normalmente definidos experimentalmente, onde a técnica de validação cruzada é utilizada e a combinação de valores que possibilita melhores resultados de classificação são escolhidas.

Além da separação em duas classes, podem acontecer situações onde existem múltiplas classes para serem consideradas e isso pode ser resolvido utilizando as estratégias um-contra-todos (*one-against-all*) ou um-contra-um (*one-against-one*) (VAPNIK, 1995).

O classificador SVM apresenta bons resultados em IHS, porém pode ter um custo computacional que cresce quadraticamente com o tamanho dos dados e métodos de seleção de características, ou bandas, o que deve ser levado em consideração (ARCHIBALD; FANN, 2007). Argumenta-se que a dimensionalidade do espaço de entrada não tem importância em relação à complexidade do SVM, visto que o algoritmo depende apenas dos vetores de suporte (GUAL-

TIERI, 2009). Esta hipótese foi apoiada por uma série de estudos com SVM, como aqueles que empregam o *kernel* RBF para aplicações de classificação de SR com coberturas de solo (MELGANI; BRUZZONE, 2004).

Estudos mostraram que a qualidade da classificação SVM pode ser aumentada pela redução da dimensionalidade do conjunto de dados (ANTHONY; RUTHER, 2007). Particularmente, Weston et al. (2000) mostram que o SVM pode ter um desempenho ruim em situações onde existem muitas características irrelevantes, um problema que é corrigido justamente usando abordagens de seleção de características.

2.3.4 Informação espacial para classificação e segmentação

Pesquisas que utilizam os classificadores como o SVM e GMLC são relacionadas à seleção de bandas que é o problema principal desta tese. Muitos desses métodos consideram apenas a informação espectral contida nas IHS. Porém, estudos sobre IHS, como os realizados por Tu et al. (2018) e Cao et al. (2017), mostram que a utilização da informação espacial integrada com a espectral pode melhorar muito o resultado da qualidade da classificação em IHS. Essa combinação é cada vez mais comum em técnicas de classificação recentes, porém, é importante observar que é pouco explorada em conjunto com métodos de seleção de bandas.

A melhoria da qualidade da classificação de IH com o uso da informação espacial é influenciada pela resolução espacial dessas imagens. Essa mesma resolução espacial, quando alta, pode reduzir significativamente problemas com pixels mistos (pixel único que contém informações de objetos distintos) e detectar mais detalhes nas cenas (GHAMISI et al., 2017). É devido a essas características que ocorre a melhoria de classificação dos pixels de IHS. Tanto a resolução espectral quanto a resolução espacial de IHS têm sido melhoradas com a evolução de tecnologias de imageamento espectral. A tendência é que cada vez mais, os dispositivos tenham a capacidade aumentada motivando o uso de informações espaciais em conjunto às espectrais.

Muitas estratégias são capazes de considerar as informações espaciais, tais como métodos baseados nos Campos Aleatórios de Markov, do inglês *Markov Random Field*, técnicas de filtragem, *extended morphological profiles* e, também, alguns métodos diretamente baseados em segmentação por similaridade e proximidade de pixels (CAO et al., 2017). Para a categoria de estratégia de segmentação existe um conjunto de métodos que consideram a informação espacial por meio do uso de superpixels.

Superpixels caracterizam uma região de forma adaptativa e de acordo com a estrutura espacial dos pixels (ZHU et al., 2017). Gonzalez e Woods (2018) descrevem que o propósito

dos superpixels é substituir a grade de pixels padrão por agrupamento que caracterizam regiões primitivas, que são mais perceptivamente significativas do que pixels individuais. A utilização de superpixels diminuem a carga computacional e melhoram o desempenho de algoritmos de segmentação por meio da utilização da informação espacial e redução de detalhes irrelevantes (ZHANG et al., 2017).

No contexto de IHS, a segmentação baseada em superpixels tem sido considerada em métodos como o *joint sparse representation classification* (JSRC) proposto por Chen, Nasrabadi e Tran (2011). Esses autores propõem o uso de um classificador baseado na representação de esparsidade, onde pixels de uma mesma classe podem compartilhar as mesmas características. Por consequência, partem da premissa que novos pixels a serem classificados podem ser relacionados por uma combinação linear de algumas amostras de treinamento contidas em um dicionário estruturado com representantes de todas as classes. O vetor esparso é recuperado resolvendo um problema de otimização com restrição de esparsidade e pode determinar diretamente o rótulo da classe da amostra (pixel) de teste.

No JSRC, os autores utilizaram a representação dos pixels de diferentes classes por uma janela fixa que considera informação espacial e espectral. Já Feng et al. (2014) propõem o *multiscale adaptive sparse representation* (MASR), que considera a representação esparsa ajustando o tamanho e forma das regiões adaptativamente formando diferentes escalas. O MASR foi comparado com outros métodos tradicionais incluindo o JSRC, e demonstrou melhores resultados por causa de suas variações de escala.

Os dois exemplos apresentados anteriormente utilizam uma estratégia própria para agrupar os pixels e demonstram que uma tarefa difícil é encontrar o tamanho adequado dos grupos de pixels. Zhang et al. (2017) propõem o *Multiscale Superpixel-Based Sparse Representation* (MSSR) que utiliza algoritmos clássicos para a geração de superpixels. Nesse método, foram explorados os algoritmos Felzenszwalb-Huttenlocher (FELZENSZWALB; HUTTENLOCHER, 2004), o *simple linear iterative clustering* (SLIC) (ACHANTA et al., 2012) e o *entropy rate superpixel* (LIU et al., 2011), que são propostos para estabelecer superpixels conforme a similaridade e proximidade dos pixels nas imagens. O MSSR é comparado com os dois métodos apresentados anteriormente demonstrando melhores resultados.

Por considerar a informação espacial e agrupamentos de pixels, os superpixels podem ter uma presença menor de artefatos (ruídos) em seus resultados e essa característica é uma das principais vantagens em relação à classificação direta pixel a pixel. Porém, técnicas baseadas em superpixels são menos comuns em pesquisas com IHS e, provavelmente, isso ocorre devido ao baixo número de imagens que estão disponibilizadas gratuitamente. Em testes com métodos

que consideram superpixels, uma amostra é uma região inteira ao invés de um único pixel como nos métodos de classificação pixel a pixel. De outra forma, a informação espacial de um único pixel obtida por meio de seus vizinhos pode dificultar a seleção de amostras de forma adequada para o treinamento e testes de classificadores. Isso pode limitar a execução de experimentos onde quantidades razoáveis de pixels de IHS são necessárias.

Outra questão é que o planejamento da captura de amostras para avaliar métodos de classificação deve ser realizado de forma a evitar a ocorrência de avaliações injustas e tendenciosas, problemas comuns em alguns estudos que envolvem o uso de informação espacial. Por exemplo, em processos de amostragem aleatória com pixels que passaram por algum processo que consideram a informação da vizinhança, dois pixels vizinhos podem ser selecionados respectivamente para treinamento e teste, influenciando na avaliação do classificador (LIANG et al., 2017).

Outra possível solução para o uso da informação espacial, em casos onde existem poucas amostras, é obtê-la por meio de filtros espaciais. O tópico na sequência apresenta uma breve revisão sobre métodos dessa categoria.

2.3.5 Obtenção de informação espacial por meio de filtragem

Em processamento de imagens, filtragem se refere a alguma operação (passagem, modificações ou rejeição) em valores de componentes como pixels. A filtragem espacial modifica uma imagem substituindo o valor de cada pixel por uma função que pode considerar seu próprio valor e os valores de pixels vizinhos. Um simples filtro de média suaviza variações locais em uma imagem e o ruído é reduzido como resultado do desfoque (GONZALEZ; WOODS, 2018). Tanto esse quanto outros filtros podem ser aplicados em conjunto à seleção de bandas e classificação de IHS.

Segundo Cao et al. (2017), a filtragem espacial tem sido introduzida para melhorar a qualidade de classificação de IHS e algoritmos dessa categoria quando comparados com outros que consideram a informação espacial e espectral, são mais fáceis de implementar e podem apresentar resultados satisfatórios. Ainda segundo os mesmos autores, duas estratégias diferentes e independentes podem ser adotadas em conjunto com a classificação de IHS:

1. De forma similar a um pré-processamento, cada banda das IHS é suavizada separadamente por um determinado filtro espacial. Posteriormente, as IHS com bandas filtradas são usadas como entrada para um classificador; e
2. A aplicação do filtro acontece após a obtenção dos resultados de classificação para cada

uma das bandas, ou seja, algoritmos de classificação são usados individualmente em cada banda. Posteriormente, a classe final de cada pixel é determinada pela frequência máxima das classes das diferentes bandas filtradas.

Além da forma de como ocorre a aplicação dos filtros espaciais nas bandas espectrais, eles ainda podem ser organizados em duas categorias conforme a forma de operar:

1. Linear: executa uma operação de soma de produtos entre uma imagem f e um *kernel* filtro, Ω . O *kernel* (também chamado de máscara ou janela) é uma matriz cujo tamanho define a vizinhança da operação e os coeficientes determinam a natureza do filtro; e
2. Não-linear: Diferentemente do caso anterior, a resposta é baseada na ordenação/classificação dos pixels contidos na região abrangida pelo filtro (GONZALEZ; WOODS, 2018).

Alguns exemplos de filtros dessas duas categorias que foram encontrados em aplicações com segmentação de IHS são apresentadas na sequência:

- Filtragem espacial linear:

1. Filtro de média: Nesse filtro, o valor de intensidade do pixel filtrado é obtido por meio do cálculo do valor médio da intensidade dos pixels vizinhos, sendo a vizinhança determinada por um *kernel* (GONZALEZ; WOODS, 2018). Um filtro de média suaviza variações locais em uma imagem e o ruído é reduzido como resultado do desfoque de cada banda ou imagem.

A equação 2.3 e a Figura 2.8 representam a aplicação de um filtro de média e também exemplifica uma filtragem espacial linear:

$$f'(i) = \frac{1}{mn} \sum_{j \in \Omega_i} f(j), \quad (2.3)$$

onde i é o pixel que está sendo filtrado que possui coordenadas referentes à sua posição nas linhas e colunas e cada j -ésimo pixel faz parte da vizinhança de i representada por um *kernel* Ω_i , que contém os pixels j . Esse *kernel* é de tamanho $m \times n$, que normalmente é de tamanho ímpar e está centralizado no pixel i , e todos os coeficientes dos pixels de Ω_i têm valor $\frac{1}{mn}$. O filtro de média calcula o valor médio para um pixel i de uma imagem f com ruído na área definida por Ω_i , que considera a vizinhança do pixel em questão, gerando uma nova imagem restaurada $f'(i)$ onde o novo pixel i tem o valor de média calculado;

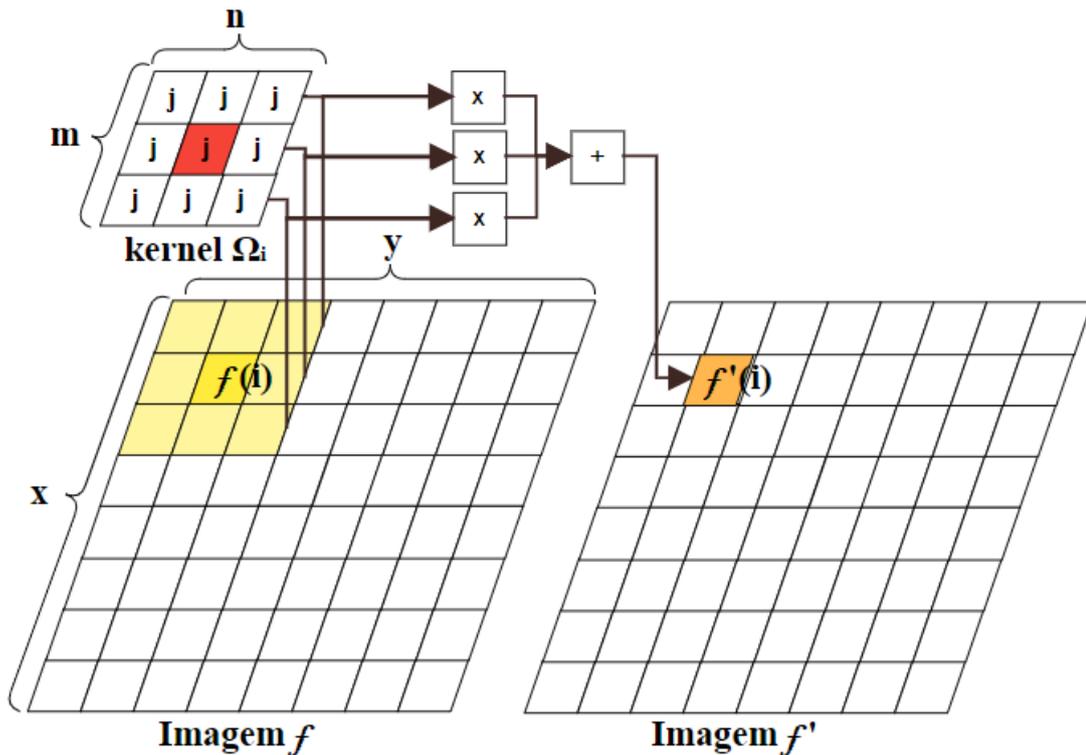


Figura 2.8. Ilustração da aplicação de um filtro de média e filtragem espacial linear.

2. Filtro Gaussiano: O filtro Gaussiano é uma técnica de remoção de ruído que utiliza uma função Gaussiana (AKYUREK; KOCER, 2017; CAO et al., 2017). O resultado da aplicação do filtro Gaussiano é uma suavização mais tênue que a proporcionada pelo filtro de média. *Kernels* Gaussianos possuem dois parâmetros importantes, o primeiro é o tamanho do *kernel* já apresentado previamente e o segundo é um parâmetro σ (desvio padrão) que determina o nível de desfoque da imagem resultante (BAGGIO, 2015). σ com valores maiores tendem a desfocar mais a imagem proporcionando resultados mais próximos do filtro de Média. O *kernel* do filtro de média atribui o mesmo peso a todos os pixels, os valores dos coeficientes do *kernel* Gaussiano diminuem em função da distância do centro do *kernel* (GONZALEZ; WOODS, 2018). A operação é semelhante à demonstrada na Figura 2.8 e esse filtro é calculado por uma média ponderada, ou seja, os coeficientes têm valores diferentes entre si seguindo o funcionamento da função Gaussiana (CAO et al., 2017).

- Filtragem não linear:

1. Filtro da mediana: O filtro da mediana é o mais conhecido da filtragem não linear, realizando uma suavização que preserva alguns detalhes das bordas, enquanto remove o ruído. A principal ideia desse filtro é substituir o valor da intensidade do pixel central pelo valor da mediana de seus vizinhos considerando, uma *kernel* de

tamanho específico (CAO et al., 2017). Esse filtro é efetivo para remoção de ruídos do tipo sal e pimenta gerando menos desfoque que filtros lineares de tamanhos semelhantes (GONZALEZ; WOODS, 2018);

2. Filtro Bilateral: Um filtro não-linear que realiza a remoção de ruídos com a capacidade de preservar bordas. O filtro Bilateral combina o domínio espacial (pode ser Gaussiano) e o espaço de cores/eixo de intensidade (NAZ et al., 2018).

O filtro Gaussiano tradicional considera os pixels vizinhos, mas não considera se esses pixels têm valores de intensidade próximos e nem se o pixel está em uma borda ou não. O efeito resultante indesejado é que o filtro Gaussiano pode desfocar as bordas. Assim, no filtro Bilateral, a função Gaussiana do espaço garante que apenas os pixels que são vizinhos espaciais são considerados para filtragem, enquanto o componente aplicado no domínio da intensidade (também pode ser Gaussiano) garante que apenas os pixels com intensidades semelhantes ao pixel central sejam incluídos para calcular o resultado da filtragem. Dessa forma, esse método preserva as bordas, pois para pixels próximos às bordas exibem grandes variações de intensidade quando comparados ao pixel central, não influenciando tanto o desfoque. No filtro Bilateral, dois parâmetros são importantes, σ_s que representa o desvio padrão espacial e σ_r que representa o desvio padrão do eixo de intensidade. σ_r com valores maiores, faz com que o resultado do filtro Bilateral se aproximam dos promovidos pelo Filtro Gaussiano.

Outros dois exemplos de filtros que tem sido explorados recentemente em IHS são o *Non-Local Means* (TU et al., 2018) e o *Guided Filter* (CAO et al., 2017; CHANG; LI; LI, 2018). Esses métodos tem particularidades na forma de operar ocasionando especificidades na condução da organização de amostras para experimentos quando aplicados. Devido a limitações da quantidade de amostras rotuladas nas IHS, esses dois filtros não foram considerados nesta tese pois requerem uma elevada quantidade de pixels de amostras vizinhos para operar (*kernel* de tamanho grande).

2.4 Considerações Finais do Capítulo

Neste capítulo, foram apresentados os conceitos e aspectos teóricos de IHS, assim como as diferentes aplicações relacionadas e proporcionadas por essas imagens. Um ponto importante é a composição dessas imagens caracterizada por centenas de bandas espectrais, ou seja, dados de alta dimensionalidade. As bandas das IHS proporcionam capacidades e aplicações adicionais

em relação a outros tipos de imagens.

Características da alta dimensionalidade também foram apresentadas, assim como os problemas relacionados, onde destacam-se a maldição da dimensionalidade e os elevados recursos computacionais necessários para processar essas imagens. Adicionalmente, para o conhecimento relacionado à capacidade proporcionada a aplicações práticas com essas imagens, foram discutidos algoritmos de classificação, segmentação e estratégias relacionadas à possibilidade da utilização da informação espacial nesses processos. Esses conceitos são importantes para estruturação do método proposto nesta tese.

O capítulo posterior apresenta uma discussão de métodos que visam lidar com a alta dimensionalidade dos dados, mais precisamente métodos de seleção de bandas de IHS, que tentam melhorar ou manter uma boa capacidade de classificação.

Capítulo 3

SELEÇÃO DE BANDAS EM IMAGENS HIPERESPECTRAIS

Considerando os problemas relacionados à alta dimensionalidade apresentados anteriormente, assim como aspectos que podem promover melhor classificação de imagens hiperespectrais (IHs), no presente capítulo serão discutidos estratégias e conceitos de seleção de bandas para essas imagens. Métodos de seleção de bandas são baseados em diversas estratégias e objetivam promover bom desempenho de classificação à medida que reduzem a dimensionalidade de dados (quantidade de bandas), portanto, são interessantes em diversas aplicações com IHs.

3.1 Seleção de características/bandas

Greco et al. (2006) e Xie et al. (2018) descrevem que a seleção de bandas é uma técnica de redução de dimensionalidade, onde deve-se encontrar uma solução X que caracteriza um subconjunto de bandas de uma IH, que otimiza um determinado critério. Essa afirmação é também colocada por Sun e Du (2019), que categorizam a seleção de bandas como um problema de otimização, principalmente quando envolve métodos de busca, porém pode-se estender esse conceito para métodos de diferentes categorias. Nesse problema, normalmente, tenta-se remover a redundância entre bandas mantendo uma boa capacidade de classificação, diminuindo a quantidade de bandas e otimizando alguma métrica relacionada à classificação (XIE et al., 2018).

A partir da perspectiva que a seleção de bandas pode ser caracterizada como um problema de otimização, o subconjunto de bandas selecionadas deve ser menor que um conjunto original Z de forma que $x \subseteq Z$ e, por exemplo, um critério a ser melhorado pode ser o resultado de uma função f que avalia a qualidade de uma classificação realizada por um algoritmo (JAIN;

DUIN; MAO, 2000). Tratando-se a seleção de bandas como problema de otimização, ela pode ser representada conforme a equação 3.1:

$$\begin{aligned} & \text{otimizar } f(x) \\ & \text{sujeito a } x \in \Omega, \end{aligned} \tag{3.1}$$

onde x representa uma possível solução, ou seja, no contexto de seleção pode ser um conjunto com uma combinação específica de bandas, que é buscada em Ω (espaço de busca) de forma que $f(x)$ deve ser ótima.

Diferente de métodos de extração de características como, por exemplo, a Análise de Componentes Principais, do inglês *Principal Component Analysis* (PCA), a seleção de bandas (ou características) não faz transformações no espaço dos atributos dos dados e, dessa forma, é indicada para quando a manutenção da informação física sobre os mesmos é necessária (LI et al., 2011; JIA; QIAN; JI, 2009).

Patra, Modi e Bruzzone (2015) destacam algumas vantagens da seleção de bandas de IHS:

- Possibilidade de uma validação das bandas selecionadas e, portanto, da informação usada pelo classificador. Métricas como o Ganho de Informação, do inglês *Information Gain* (IG), podem ser utilizadas para mensurar o quanto uma determinada banda pode contribuir em um processo de classificação usando determinados algoritmos;
- Os resultados da seleção podem ser usados como uma ferramenta de mineração de dados para inferir uma informação física (isto é, bandas espectrais) com base nas classes em que são discriminadas. Pode-se utilizar desse recurso para saber quais bandas espectrais que melhor contribuem para classificação de um determinado objeto como, por exemplo, uma determinada cultura com presença de pragas na área de agricultura;
- Possibilidade de definir sistemas que bandas irrelevantes ou com alta correlação não sejam adquiridas e armazenadas. Isso pode simplificar o processo de aquisição, reduzir o tempo de processamento e os requisitos de armazenamento de dados.

As bandas que devem ser selecionadas são as que proporcionam uma boa separação de classes, melhorando o desempenho de classificadores, não sendo redundantes e contendo informações relevantes para uma determinada tarefa (MOLINA; BELANCHE; NEBOT, 2002; MONTEIRO; MURPHY, 2011; DUDA; HART; STORK, 2001). Portanto, quando se tem o

interesse em conhecer as bandas que são utilizadas em tarefas de classificação e segmentação em um determinado contexto, a seleção de bandas deve ser utilizada.

Diversos métodos de seleção de bandas podem ser encontrados na literatura e com isso categorizações com base em diferentes aspectos de estratégias também podem ser encontradas. Um exemplo tradicional é categorizar os métodos entre supervisionados (semi-supervisionados) ou não supervisionados, onde esses métodos estão relacionados à existência de amostras rotuladas para treinamento ou não.

Uma revisão de literatura sobre seleção de bandas é apresentada em Sun e Du (2019), que organizam os métodos em seis tipos:

1. métodos baseados em ranqueamento;
2. métodos baseados em agrupamento (*clustering*);
3. métodos baseados em esparsidade;
4. métodos baseados no aprendizado embarcado;
5. métodos baseados em busca; e
6. métodos baseados em esquemas híbridos.

Essa categorização é adotada como um dos critérios de organização nos trabalhos relacionados apresentados nesta tese, pois permite observar a variedade de algoritmos que dão origem a métodos de seleção de bandas. Uma observação importante é que os métodos das seis categorias anteriores foram categorizados em problemas de otimização de uma única métrica ou função. Portanto, uma nova categoria, de métodos baseados em busca multiobjetivo, é adicionada para se adequar a uma das principais estratégias de seleção de bandas atuais e exploradas nesta pesquisa:

7. métodos baseados em algoritmos de otimização multiobjetivo.

Além disso, os métodos apresentados estão sub-categorizados conforme o uso de rótulos proporcionados por especialistas ou não, proporcionando as categorias supervisionados (incluído também os semi-supervisionados, pois também fazem uso de rótulos determinados por especialistas) ou não supervisionados. Os trabalhos comparados nos experimentos também são posteriormente organizados em função dessas subcategorias.

Outra categorização bastante tradicional utilizada em Habermann (2018) e Molina, Belanche e Nebot (2002) é classificar os métodos de seleção de características em *Filter*, *Wrapper* e *Embedded*, conforme as estratégias seguintes:

1. *Filter*: os métodos dessa categoria estabelecem métricas de avaliação e então eliminam algumas bandas antes da construção do modelo de classificação (TUV et al., 2009). Dessa forma, o objetivo da seleção de bandas é que primeiro ela seja realizada, para que posteriormente sejam aplicadas as etapas necessárias para o treinamento em algoritmos de classificação. Quase todos os métodos da sub-categoria não supervisionada encontrados durante essa pesquisa têm a característica comum de serem baseados na estratégia *Filter*. Isso está em conformidade com o princípio de que métodos baseados na estratégia *Filter* não incorporam um classificador no processo de busca e, portanto, não necessitam de dados rotulados (classes) (HABERMANN, 2018). Porém, os métodos baseados na estratégia *Filter* não se limitam a serem não supervisionados, pois pode-se fazer o uso de amostras rotuladas para serem estabelecidas métricas para as bandas sem que um classificador seja utilizado, do contrário, caracterizariam os métodos como baseados na estratégia *Wrapper*;
2. *Wrapper*: Métodos baseados nessa estratégia utilizam um classificador e avaliam diretamente o valor de cada característica (TUV et al., 2009). Nessa estratégia, a seleção ocorre durante o treinamento do classificador. Para cada banda ou conjunto de bandas adicionado ou removido das possíveis combinações, o classificador deve ser treinado novamente, gerando um modelo que é avaliado (HABERMANN, 2018).

Uma desvantagem dessa estratégia em relação a *Filter* é sua lentidão devido aos repetitivos processos de treinamento e, além disso, podem ser necessários conjuntos com grande quantidades de amostras para serem executadas, principalmente quando utilizados classificadores paramétricos. Normalmente, pesquisadores que utilizam a seleção de bandas supervisionada têm que conceber métodos capazes de lidar com poucos dados de treinamento. Isso pode ser uma questão desafiadora para métodos *Filters* baseados na classificação e medidas de separação de classes. Porém, para métodos baseados na estratégia *Wrapper*, a escassez de dados de treinamento pode piorar o problema, uma vez que os métodos contam exclusivamente com classificadores para gerar resultados. Em compensação a essa desvantagem, devido aos métodos da estratégia *Wrapper* utilizarem o classificador incorporado ao processo de seleção de bandas, as bandas selecionadas podem auxiliar a promover um modelo classificador com desempenho de acerto muito superior em relação a métodos baseados na estratégia *Filter* (MOLINA; BELANCHE;

NEBOT, 2002; SHAHANA; PREEJA, 2016).

As bandas selecionadas por métodos dessa categoria podem operar bem somente com o classificador incorporado ao processo de seleção de bandas. Portanto, recomenda-se utilizar a seleção com o mesmo classificador que será utilizado na aplicação para a qual as bandas foram selecionadas (KOHAVI; JOHN, 1997);

3. *Embedded*: Essa estratégia tenta combinar as qualidades das duas estratégias anteriores, usando algoritmos de classificação que têm os seus próprios métodos de seleção de características. Métodos baseados na estratégia *Embedded* usam todas as variáveis para gerar um modelo de classificação e, então, analisam esse modelo inferindo a importância de suas variáveis (TUV et al., 2009). Os representantes mais conhecidos dessa estratégia são baseados em árvores de decisão ou redes neurais artificiais (MITCHELL, 1982).

Todos os tipos de categorizações apresentados podem ser utilizados em conjunto e constituir diferentes níveis de organização. Por exemplo, é comum existirem métodos baseados na estratégia *Filter* que são supervisionados e utilizam recursos de ranqueamento. Portanto, nesta tese, sempre que pertinente, além de discutir a categoria principal e subcategoria que o método está organizado, também serão discutidos se ele pertence à estratégia *Filter*, *Wrapper* ou *Embedded*. Mais do que a categorização e descrição dos trabalhos relacionados, nos tópicos a seguir são apresentados os principais elementos e características dos métodos que contribuem para seleção de bandas.

3.1.1 Métodos baseados em ranqueamento

Métodos baseados em ranqueamento quantificam a importância das bandas conforme uma métrica de priorização baseada em estratégias supervisionadas ou não supervisionadas. Após essa quantificação, esses métodos selecionam as bandas melhores ranqueadas (SUN; DU, 2019).

- **Não Supervisionados:**

Existem diferentes métodos não supervisionados baseados na estratégia *Filter*, sendo o *Maximum variance principal component analysis* (MVPCA) proposto por Chang et al. (1999) um dos mais tradicionais e ainda muito utilizado em comparações. Esse método utiliza uma estratégia de ranqueamento por meio de fatores de carga que são baseados em procedimentos que envolvem autovalores e autovetores. Os autores também propõem uma métrica chamada de "poder das bandas" que permite calcular a quantidade de bandas necessárias para obter uma taxa

de variância desejada. Essa métrica é estabelecida conforme os fatores de carga de uma quantidade m de bandas desejadas. Depois, uma segunda métrica baseada na razão entre o poder das m bandas e o total de variância presente no conjunto de todas as bandas é estabelecida. Além das métricas obtidas, os autores discutem que as bandas selecionadas pelo MVPCA podem conter alta correlação e, então, propõem uma adequação para evitar essa correlação no processo de ranqueamento, gerando um método chamado *maximum-signal-to-noise ratio* (SNR)-PCA.

Uma métrica bem conhecida chamada de entropia da informação, também é explorada para a criação de métodos de ranqueamento e frequentemente utilizada em diferentes métodos de seleção de bandas. Um exemplo de método baseado nessa métrica é mostrado em Groves e Bajcsy (2003), onde a entropia é usada para mensurar cada banda individualmente, podendo ser definida conforme a equação 3.2:

$$H(z_i) = - \sum_{h_k \in \Omega} p(h_k) \log p(h_k), \quad (3.2)$$

onde $H(z_i)$ é a entropia da banda z_i , Ω é o conjunto com distribuições de probabilidades (intensidades em escala de cinza obtidos a partir dos valores de refletância) e $p(h_k)$ representa a probabilidade do evento $h_k \in \Omega$ ocorrer. No caso de seleção de bandas, $p(h_k)$ é dado pela razão entre a intensidade k do histograma da banda z_i e a quantidade total de valores de intensidade presentes nos pixels dessa banda (GONG; ZHANG; YUAN, 2016).

Groves e Bajcsy (2003) descrevem que em seu método as bandas são ordenadas crescentemente, da banda com o menor valor (pequena quantidade de informações) para a banda com o maior valor de entropia (grande quantidade de informações). A importância dessa ordenação pode ser observada pela imagem da Figura 3.1 que foi adaptada do estudo de Gong, Zhang e Yuan (2016).

Gong, Zhang e Yuan (2016) ainda citam, como mostrado na Figura 3.1, cinco bandas da imagem Indian Pines com baixa entropia: 200, 199, 145, 2, e 104, e as cinco bandas com entropia elevada: 27, 112, 29, 30, e 31. É perceptível que as cinco bandas com elevada entropia são mais distintas do que as cinco bandas com baixa entropia. Então pode-se observar que bandas com mais entropia são mais informativas e cruciais para tarefas de classificação, enquanto as bandas com menos entropia são mais ruidosas e indistintas.

Uma métrica derivada da entropia, chamada de informação mútua (IM), também é utilizada em estratégias de seleção de bandas, como em Guo et al. (2006), onde a IM é utilizada para ranquear as bandas. Além disso, essa métrica também é utilizada em métodos de outras categorias como, por exemplo, agrupamento (*clustering*). A IM mede a dependência estatística entre duas

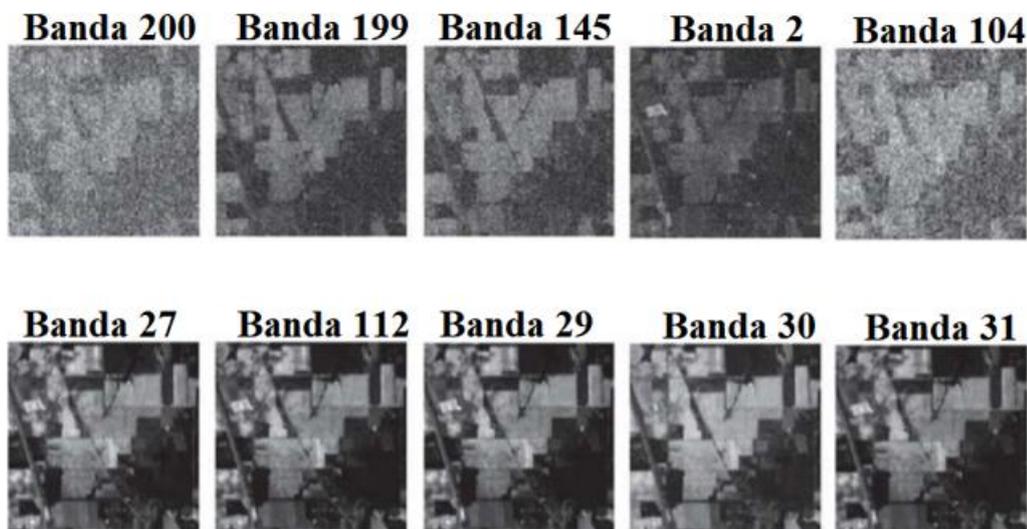


Figura 3.1. Exemplos de bandas da imagem Indian Pines. Bandas com baixos valores de entropia: 200, 199, 145, 2 e 104. Bandas com elevados valores de entropia: 27, 112, 29, 30 e 31 (adaptado de Gong, Zhang e Yuan (2016)).

variáveis aleatórias e é definida conforme a equação 3.3 (SUN; DU, 2019):

$$I(z_i, z_j) = H(z_i) + H(z_j) - H(z_i, z_j), \quad (3.3)$$

onde $H(z_i)$ e $H(z_j)$ são as entropias das bandas z_i e z_j , respectivamente, e $H(z_i, z_j)$ é a entropia conjunta. Em Groves e Bajcsy (2003) é aplicada uma estratégia para estimar um GT evitando a participação de um especialista. Neste caso, a IM é utilizada para avaliar a utilidade de cada banda espectral para classificação em relação ao GT estimado. Portanto, o método proposto tem uma característica particular em relação à maioria dos métodos apresentados, pois não incorpora um classificador e, portanto, não é baseado na estratégia *Wrapper* e utiliza um GT estimado a partir de conhecimento *a priori* durante seus processamentos. Então essas bandas podem ser ranqueadas conforme a IM e um número determinado de bandas pode ser selecionado.

Outro critério que pode ser usado para seleção de bandas é a análise de informação de dados de bandas considerando um equilíbrio entre as métricas integradas de precisão (*overall accuracy*) e redundância, como demonstrado no método proposto *joint maximum overall accuracy and minimum redundancy-based* (MAMR) em Sui et al. (2015). Os autores formulam uma equação baseada na precisão estimada e redundância que utiliza um parâmetro adaptativo para controlar o equilíbrio dessas métricas por meio de ponderação. A estratégia empregada faz com que o parâmetro seja ajustado automaticamente, evitando a necessidade de definição do usuário. O método de estimativa da precisão funciona de forma não supervisionada e foi adotado para evitar a necessidade de amostras de treinamento. A precisão estimada é usada no início do

processo de seleção de bandas para ranqueá-las e, em conjunto com um parâmetro estabelecido, é utilizada para descartar bandas com menor valor, ou seja, é realizada uma seleção inicial de bandas. Posteriormente, a métrica que integra a precisão e redundância utilizando o parâmetro de equilíbrio adaptativo é empregada para ranquear as bandas restantes e, conforme uma quantidade determinada pelo usuário, um conjunto de bandas é selecionado. Outra característica importante observada nesse estudo é a utilização de uma estratégia para selecionar um conjunto inicial de bandas que são exploradas. O método é comparado com outros métodos como *Ward's Linkage strategy Using Divergence* (WaLuDi) e MVPCA, onde os autores consideraram o método proposto como o mais robusto em termos de precisão e redundância.

- **Supervisionados:**

Embora os estudos apresentados anteriormente fizeram uso da IM como métrica para elaboração de métodos não supervisionados, ela também pode ser utilizada para estabelecer métodos supervisionados. Um exemplo de uma nova estratégia baseada na IM considera a correlação entre três variáveis, rótulo de classes (portanto, supervisionado) e duas bandas espectrais, sendo esse estudo apresentado em Feng et al. (2014).

Além da entropia e IM, outras métricas que não são baseadas nas tradicionais também são exploradas, por exemplo, Patra, Modi e Bruzzone (2015) apresentam um método de ranqueamento de bandas utilizando a teoria de *Rough Set*. Segundo os mesmos autores, *Rough Set* é um paradigma para lidar com imprecisão, incompletude e incerteza de dados, podendo ser utilizado para estruturar métodos de seleção de bandas. O método proposto pelos autores utiliza diretamente amostras rotuladas (GTs). Por meio da teoria de *Rough Set* são propostos critérios para estabelecer a relevância e significância de cada uma das bandas onde elas são ordenadas e aquelas de maior valor são selecionadas. A métrica de relevância permite medir o quão bem uma banda pode contribuir para classificação e a significância é o critério que mensura o quão importante uma banda é para o conjunto de bandas selecionado, permitindo descartar bandas com alta correlação. Bandas que não são importantes (com baixa significância) não são úteis e podem ser descartadas. A quantidade de bandas a serem selecionadas deve ser previamente estabelecida pelo usuário.

Feng et al. (2017) propõem uma estratégia supervisionada de ranqueamento para seleção de bandas por meio de um método chamado *Non-homogeneous Hidden Markov Chains* (NHHMC) que utiliza a transformada de *Wavelets* para reduzir a redundância enquanto preserva a informação útil. Esse método é composto por dois módulos: 1) um módulo de treinamento de características baseado no NHHMC e 2) um módulo específico para seleção de bandas. O

primeiro módulo baseado no NHHMC coleta aleatoriamente um conjunto de pixels de treinamento, computa os coeficientes da transformada de *Wavelets* para cada espectro, e utiliza os resultados como entrada para uma unidade de treinamento do NHHMC. O NHHMC emite como saída parâmetros do modelo de Markov e rótulos para cada um dos espectros de treinamento a serem usados como características para posterior ranqueamento. O segundo módulo determina valores baseados na métrica estabelecida anteriormente e na correlação entre bandas, permitindo a ordenação das mesmas e, conseqüentemente, a seleção do subconjunto final. Embora a estratégia proposta leve em consideração o uso de um GT, ela é baseada na estratégia *Filter*, porque a seleção de bandas é feita antes da classificação realizada pelo SVM, não incorporando o classificador ao processo de seleção de bandas. A estratégia também é um algoritmo de busca sequencial operando como um *sequential forward selection* (SFS).

3.1.2 Métodos baseados em *clustering*

Métodos baseados em *clustering* agrupam as bandas do conjunto original e, então, selecionam as mais representativas de cada grupo (SUN; DU, 2019).

- **Não Supervisionados:**

Exemplos tradicionais de algoritmos baseados em *clustering* para seleção de bandas são o *Ward's linkage* com a métrica de divergência de Kullback–Leibler (*Ward's Linkage strategy Using Divergence* - WaLuDi) e *Ward's linkage* com a IM (*Ward's Linkage strategy Using Mutual Information* - WaLuMi), propostos por Martinez-Uso et al. (2007), que inspiram diversos outros métodos de seleção de bandas e são comumente comparados com outros métodos. Outros exemplos de algoritmos de *clustering* usados em métodos de seleção de bandas são:

- K-Means (SOHAIB; HAQ; MUSHTAQ, 2013; YANG et al., 2017b);
- *Spectral Clustering* (SC) (SUN et al., 2015; ZHU et al., 2017);
- *Affinity propagation* (AP) (JIA; QIAN; JI, 2009; YAO; QIAN, 2009; YANG et al., 2017a);
e
- *Density-based spatial clustering of applications with noise* (DBSCAN) (DATTA; GHOSH; A., 2012).

As métricas que compõem o WaLuDi e o WaLuMi são utilizadas em diversos outros estudos, sendo a divergência uma métrica de discrepância entre duas distribuições de probabilidade.

A IM, previamente apresentada, é uma métrica para o grau de independência entre variáveis aleatórias e, também, quanto uma variável pode explicar a informação contida na outra. Além da IM ser utilizada da forma supervisionada previamente apresentada, onde foi comparada com o GT (GROVES; BAJCSY, 2003), ela pode ser utilizada para mensurar o quanto uma banda se sobrepõe a uma outra sem considerar o GT e, portanto, não supervisionada. Dessa forma, a IM permite obter uma métrica semelhante à correlação que é comumente utilizada para agrupamento.

O WaLuDi ou WaLuMi constroem uma matriz de similaridades que, posteriormente, é utilizada para gerar m (definido pelo usuário) grupos de bandas realizado pelo *Ward's linkage*. Com os grupos formados, uma métrica para avaliar a qualidade das bandas é utilizada para ranqueá-las conforme seus grupos. Finalmente, as bandas de melhor qualidade de cada grupo são selecionadas para serem utilizadas na formação de novas IHS. Os métodos são capazes de selecionar bandas que promovem um desempenho melhor ou competitivo no critério de classificação de pixels em relação aos outros métodos comparados como o MVPCA já apresentado.

As técnicas de *clustering* até aqui descritas são tradicionais, porém métodos modernos também são estruturados, como o *information-assisted density peak index* (IaDPI) proposto por Luo, Xue e Yin (2017) e o *dual-clustering-based band selection by context analysis* (DCCA) proposto por Yuan, Lin e Wang (2016). O IaDPI é baseado na entropia de cada banda e em uma estratégia de *clustering* por pico de densidade - *density peak* proposta por Rodriguez e Laio (2014). O IaDPI considera a entropia entre bandas de uma densidade local e a distância entre *clusters* para garantir centros com alta qualidade. Já o DCCA considera o agrupamento da informação espacial (bandas e pixels vizinhos) junto à informação espectral para seleção de bandas. Uma importante observação nesse método é a utilização dos pixels vizinhos, ou seja, informação espacial no processo de seleção de bandas.

Em Cao et al. (2016), um método constituído de duas etapas é proposto. Primeiro, a estrutura espacial é explorada para determinar a capacidade discriminante de cada banda, onde as com pouca capacidade são descartadas. Então, um novo algoritmo de *clustering* que faz uso de um classificador (*K-nearest neighbors* (KNN), *support vector machine* (SVM) ou outros) é utilizado para definir o número de grupos e, conseqüentemente, a quantidade de bandas a serem selecionadas. Esse classificador não utiliza amostras rotuladas reais no treinamento, mas amostras com rótulos gerados por meio de uma estratégia não supervisionada. Após o estabelecimento de grupos, uma banda representativa de cada um é selecionada.

Métodos para gerar grupos de bandas também podem ser obtidos por meio de teoria dos grafos e Yuan, Zheng e Lu (2017) propõem o método chamado *multigraph determinantal point*

process (MDPP) que tem o objetivo de descobrir a relação entre bandas, formando grupos de assinaturas espectrais, e encontrar o subconjunto de bandas ótimo. Para isso, múltiplos grafos são projetados para capturar o relacionamento intrínseco entre as bandas. O MDPP incorpora um método chamado *Mixture determinantal point process* que representa múltiplas dependências nos grafos gerados e busca por uma ótima combinação de bandas. Assim o MDPP consegue modelar as múltiplas dependências em grafos, providenciando uma estratégia de busca eficiente para o processo de seleção de bandas.

- **Supervisionados:**

Além de métodos de seleção baseados em agrupamento não supervisionados, outras estratégias consideram métodos semi-supervisionados. Jiao et al. (2015) propõem uma métrica chamada *normalized trivariable mutual information* (NTMI) para mensurar três tipos diferentes de correlação entre bandas para classificação considerando pixels rotulados. Utilizando o NTMI e também a IM (para bandas correlacionadas e sem considerar os rótulos), uma medida geral de correlação é estruturada. O método de *clustering* AP é adequado com as métricas em questão e então é capaz de selecionar bandas informativas e com capacidade discriminatória. As bandas selecionadas são utilizadas para formar novas IHS, que têm seus pixels classificados pelo algoritmo SVM com função *radial basis function* (RBF) e quando comparadas com outros métodos da literatura como aplicação direta do AP ou uso de todas as bandas espectrais (sem seleção), elas apresentaram bons resultados.

3.1.3 Métodos baseados em esparsidade

Métodos baseados na esparsidade formulam a seleção de bandas como um problema de otimização usando matrizes esparsas (SUN; DU, 2019). De acordo com a teoria da esparsidade aplicada para IHS, cada banda espectral pode ser esparsamente representada usando apenas alguns coeficientes diferentes de zero de uma matriz associados a um dicionário. Dessa forma, um problema de seleção de bandas pode ser modelado por meio da equação 3.4:

$$Y = DX, \quad (3.4)$$

onde $Y = [Y_1, \dots, Y_{l \times c}]$ é uma matriz estruturada a partir da IH original, onde cada Y_i representa um dos $l \times c$ pixels que contém os valores para todas as Z bandas, D é a matriz que caracteriza um dicionário incompleto com tamanho de $Z \times Z$ e X é a matriz de coeficientes de representação de Y (LI; QI, 2011) com o tamanho idêntico a Y . Os métodos de seleção de bandas baseados

na esparsidade utilizam essa representação para revelar estruturas subjacentes nos dados IH. As bandas representativas podem ser encontradas resolvendo o problema de otimização demonstrado na equação 3.4, onde X caracteriza coeficientes que permitem ranquear as bandas.

Para essa categoria, foram exemplificados apenas métodos não supervisionados, que são apresentados a seguir.

- **Não Supervisionados:**

O *Sparsity based band selection* (SpaBS) proposto por Li e Qi (2011) utiliza o algoritmo K-SVD, que funciona de uma forma similar ao K-Means sendo um tipo de generalização desse algoritmo. O K-SVD é capaz de decompor os dados hiperespectrais representados pela matriz Y da equação 3.4 na multiplicação do dicionário D pela matriz de coeficientes $X = [X_1, \dots, X_{l \times c}]$. X possui a propriedade de esparsidade e tem a capacidade de revelar o quanto cada banda contribui na formação dos dados hiperespectrais. Ao calcular o histograma da matriz de coeficientes X , é possível obter as m bandas (definido a partir de um coeficiente de esparsidade) que aparecem com mais frequência para atender à necessidade de redução de dimensionalidade e, ao mesmo tempo, preservar o significado físico das bandas selecionadas.

Outro exemplo de método baseado em esparsidade chamado de *dissimilarity-weighted sparse self-representation* (DWSSR) foi apresentado em Sun et al. (2016) para selecionar um subconjunto de bandas apropriado para a classificação de IHS. Assim como o caso anterior, o DWSSR supõe que o conjunto de todas as bandas pode ser representado por subconjunto de bandas menor utilizando a teoria da esparsidade. O problema é formulado por meio de uma representação esparsa de todas as bandas em um modelo chamado de *sparse self-representation* (SSR), que considera a inexistência do dicionário D da equação 3.4. O DWSSR faz uma série de ajustes à proposta de seleção de bandas por representação esparsa, construindo um modelo SSR para considerar ruídos gaussianos que incorpora uma matriz ponderada de dissimilaridade entre bandas e utiliza uma estratégia específica para gerar a matriz de coeficientes. Assim como o caso anterior, a partir da matriz de coeficientes, as m bandas melhor ranqueadas são selecionadas.

3.1.4 Métodos baseados em aprendizado embarcado

Métodos baseados em aprendizado embarcado incorporam a seleção de banda na otimização de modelos para aplicações específicas como, por exemplo, classificação, detecção de alvos e separação espectral (SUN; DU, 2019). Diversos métodos encontrados nesta pesquisa podem ser facilmente agrupados na categoria de aprendizado embarcado, principalmente os que fazem

uso de rótulos e conduzem o processo de seleção de bandas por meio desses rótulos como, por exemplo, os métodos baseados nas estratégias *Wrapper* ou *Embedded*. Como ambos os casos são para tarefa de classificação (ou até reconhecimento) de elementos específicos da imagem, pode-se considerar muitos métodos como aprendizado embarcado e aqueles que são exemplificados na sequência são todos supervisionados.

- **Supervisionados:**

Um exemplo de método que pode ser categorizado como aprendizado embarcado e também baseado na estratégia *Embedded* pode ser observado no trabalho de Gomez-Chova et al. (2003) em que os autores propõem a utilização do método *Classification And Regression Trees* (CART) para seleção de bandas de IH para classificação de plantações. Esse algoritmo utiliza uma métrica para avaliar a classificação chamada de índice de Gini, baseada na entropia para construção da árvore de decisão. O CART permite selecionar as bandas mais discriminantes.

Monteiro e Murphy (2011) propõem um *framework* de seleção de bandas para IHS usando árvores de decisão chamadas de *boosting decision trees*. A ideia de *boosting* é treinar muitos classificadores fracos, *weak learners*, que são classificadores melhores que estratégias aleatórias sobre várias distribuições de dados de entrada e, então, combinar seus resultados. Nesse método algumas árvores de decisão binárias com um nó único são construídas para formar um *ensemble*. Após isso, cada nó de decisão (variável ou banda) é avaliado com uma métrica de importância baseada na média de todas as árvores e todas as classes, ou seja, no *ensemble*. No processo, os nós mais importantes são selecionados. Outro algoritmo comum, chamado de *Random Forest* (RF), que é uma composição de árvores de decisão, tem sido proposto para classificação de IHS por diferentes autores como em Dinç e Aygün (2013), Amini et al. (2018), Xie et al. (2018) e Sun et al. (2018). Esse algoritmo também foi aplicado em conjunto a métodos híbridos, portanto será discutido posteriormente.

Devido a amostras de treinamento desequilibradas, o classificador SVM é um método muito popular usado na análise de imagens hiperespectrais. A eliminação recursiva de características, do inglês *recursive-feature elimination* (RFE), foi integrado ao SVM em alguns trabalhos da literatura, formando o RFE-SVM (SUN; DU, 2019). O RFE-SVM pode fornecer desempenho superior na seleção de bandas hiperespectrais e, devido a problemas de velocidade, em Archibald e Fann (2007), modificações nessa estratégia são realizadas para melhorar a velocidade de convergência e eliminar automaticamente as bandas redundantes sem ter conhecimento prévio do número de bandas adequado. Outra proposta de melhoria desse algoritmo pode ser encontrada em Zhang e Ma (2009) onde um novo critério no processo de treinamento foi proposto

com o propósito de melhorar a classificação e reduzir o tempo de processamento médio. Além disso, em Kuo et al. (2014), o RFE-SVM foi modificado para operar como um algoritmo de seleção de bandas baseado em *Kernel*, que usa a magnitude dos coeficientes do SVM como um critério de ranqueamento para selecionar bandas importantes.

Em Sharma et al. (2016), as Redes Neurais Convolucionais, do inglês *convolutional neural network* (CNN), e a AdaBoost com SVM foram usadas para explorar características profundas das bandas em IHS, porém em um contexto diferente do SR, que no caso é reconhecimento facial. Basicamente, um esquema foi proposto para treinar a CNN para classificação de IHS e então uma nova abordagem para explorar as características da CNN para buscar por bandas espectrais discriminativas foi aplicado usando o AdaBoost com SVM. Outro trabalho semelhante a este é apresentado em Lorenzo et al. (2018), onde as CNNs são utilizadas especificamente para a seleção de bandas na forma de estratégia *Embedded* e baseada no aprendizado embarcado, onde exploraram as regiões mais informativas do espectro de IHS e, assim, puderam identificar e selecionar bandas relevantes. Um *framework* baseado nas CNNs e em uma métrica chamada densidade de distância, do inglês *distance density* (DD) é proposto em Zhan et al. (2017). Nesse caso, a DD é usada, ao invés de uma busca aleatória, para seleção de bandas candidatas que são usadas para estruturar o classificador CNN. Experimentos mostram que a seleção de bandas baseada na DD é mais rápida que a versão de busca aleatória e a CNN em conjunto com a DD proporcionaram melhores resultados que os métodos comparados.

3.1.5 Métodos baseados em estratégias de busca

Métodos baseados em estratégia de busca convertem a seleção de características em um problema de otimização para encontrar uma combinação de bandas que melhora uma determinada função de *fitness* (função objetivo). A equação 3.1, previamente apresentada, representa esse problema de otimização. Nos métodos baseados em estratégia de busca, duas questões são importantes:

1. A função de *fitness* que dever ser melhorada durante a busca das bandas adequadas; e
2. A estratégia de busca que é responsável por determinar a melhor maneira de encontrar uma solução ótima ou sub-ótima.

Métodos baseados em estratégias de busca podem ser supervisionados, semi-supervisionados ou não supervisionados.

- **Não Supervisionados:**

Em vez de testar exaustivamente todas as combinações, os métodos baseados em estratégia de busca incremental adicionam sequencialmente bandas em um conjunto visando otimizar uma função de *fitness* até atingir uma quantidade desejada e um exemplo é de método baseado em estratégia de busca incremental é o *sequential forwarding selection* (SFS) (SUN; DU, 2019). Como alternativa ao SFS existe o *sequential backward selection* (SBS), que começa com um conjunto com todas as bandas e iterativamente remove a banda que proporciona pior *fitness*.

Um exemplo do uso do SBS é mostrado em Sun et al. (2014), onde foi proposto um método não supervisionado baseado na estratégia *Filter* chamado de *minimum noise band selection* (MNBS), que combina as ideias das técnicas chamadas *noise-adjusted principal component* (NAPC) e *maximum determinant of covariance matrix* (MDCM) em uma métrica para avaliar a qualidade das IHS. NAPC é uma métrica para maximizar o *signal-to-noise ratio* (SNR), que é um índice capaz de medir a qualidade de imagens (LEE; WOODYATT; BERMAN, 1990). Já o MDCM é um método para mensurar uma métrica de quantidade de informação e baixa correlação ao mesmo tempo. Baseado na combinação desses métodos é estruturada uma nova métrica e o MNBS com o SBS é responsável pela seleção de bandas.

Métodos mais sofisticados são baseados em algoritmos evolucionários (AE) e têm sido amplamente explorados para seleção de bandas (SUN; DU, 2019). Um exemplo, com o algoritmo de seleção clone, do inglês *clonal selection algorithm* (CSA), que é uma teoria baseada em sistemas imunológicos representante de AEs e, em Feng et al. (2016), o mesmo é utilizado para selecionar bandas com uma função derivada da IM e entropia chamada de *Maximum information and minimum redundancy* (MIMR). O método tenta selecionar um conjunto de bandas otimizando a função MIMR.

Algoritmos meméticos, que podem ser considerados como uma proposta nova baseada nos Algoritmos Genéticos (AG), foram usados recentemente para seleção de bandas. Em Zhang et al. (2017), uma função de *fitness* é projetada combinando uma métrica de desvio padrão dos pixels de cada banda (como métrica de informação baseada no contraste) e a métrica da IM entre bandas, sendo esse método baseado na estratégia *Filter*. O algoritmo memético tem seu operador de cruzamento e sua estratégia de busca local projetados para IHS e, então, é utilizado no processo de seleção.

- **Supervisionados:**

Métodos incrementais supervisionados usam conhecimento a priori para seleção de bandas na tentativa de maximizar a separabilidade de classes (SUN; DU, 2019). Fauvel et al. (2015) propõem um método que seleciona bandas que são avaliadas pelo resultado do classificador

Gaussian Mixture Model (GMM) em validação cruzada. A seleção é realizada por um método chamado *nonlinear parsimonious feature selection* (NPFS) que incorpora o GMM. O NPFS, usa uma busca por meio do SFS, selecionando iterativamente uma banda a partir do conjunto original e a adicionando ao conjunto de bandas selecionadas. Esse conjunto é estabelecido conforme o GMM, onde bandas são selecionadas de acordo com a qualidade de classificação.

Bai et al. (2015), exploraram um método baseado em informações espaciais e espectrais, onde primeiramente é construído um modelo de hipergrafo usando todas as bandas para medir a similaridade entre os pixels no domínio espectral (por meio da IM) e espacial. Um método semi-supervisionado para propagação de rótulos é usado para estabelecer classes de amostras não rotuladas. Após isso, um modelo de regressão linear baseado em uma matriz de dispersão é construído e o método LASSO (algoritmo de otimização) é aplicado para selecionar um conjunto de bandas. Por fim, pixels com bandas selecionadas são utilizados para treinar o algoritmo SVM que, posteriormente, é testado e comparado aos resultados de classificação de IHS com bandas selecionadas por outros métodos.

AEs são uma das categorias de métodos mais comuns baseados na estratégia *Filter* e em Kavzoglu e Mather (2002), os AGs foram utilizados para selecionar as bandas que são utilizadas como entradas de uma rede neural artificial *Multilayer Perceptron* (RNA-MLP). Nesse estudo, os autores se preocupam em testar diferentes métricas como funções de *fitness*, tais como as de separação: divergência, transformada da divergência, distância de Bhattacharyya e a distância de Jeffries-Matusita; e, também, os métodos estatísticos: Wilks' Λ e Hotelling's T^2 . Os autores também investigam os AGs baseados na estratégia *Wrapper* onde duas métricas foram avaliadas, a acurácia média e a precisão, ambas obtidas por meio do classificador baseado na distância de Mahalanobis, do inglês *Mahalanobis distance classifier* (MDC) incorporado ao método proposto. Cada uma dessas métricas foi aplicada como critério de função de *fitness* junto aos AGs e, também, com o método SFS. Embora as bandas selecionadas pelo MDC não tenham proporcionado os melhores resultados com o RNA-MLP, durante o processo de seleção, o MDC apresentou melhor precisão do que as bandas selecionadas que foram utilizadas na RNA-MLP. Isso pode ser explicado devido a diferenças no viés indutivo de cada classificador, sugerindo que os métodos baseados na estratégia *Wrapper* funcionam melhor com a aplicação do próprio classificador. Em sua discussão, os autores também destacam que os métodos implementados com AGs foram melhores que os métodos implementados com SFS. As bandas selecionadas pelos AGs com *fitness* usando as métricas de Transformada da Divergência e Jeffries-Matusita foram as que apresentaram melhor precisão em conjunto com a RNA-MLP.

Em Zhang, Sun e Li (2009), um método baseado na estratégia *Filter* é modelado com al-

goritmos genéticos (AG) que mescla a representação de soluções (ou indivíduos) binária e a métrica de distância de Bhattacharyya para avaliar a separação de classes. Uma ilustração dos AGs com essa estratégia pode ser vista na Figura 3.2, onde observa-se o conjunto de soluções, ou se seja, 4 indivíduos representados por um vetor binário que constituem a população. Cada posição do vetor representa uma banda, onde o conteúdo com valor 1 representa sua existência naquela solução, e o valor 0, a ausência. A operação de cruzamento guia o processo de seleção de bandas, permitindo misturar soluções selecionadas (pais) para gerar soluções novas (filhos). O critério de seleção normalmente dá a preferência para as soluções com bons valores de *fitness* e duas estratégias tradicionais são chamadas de torneio e roleta. Após a operação de cruzamento, as novas soluções ainda podem passar por uma alteração em seus valores com a operação chamada de mutação, que tem o propósito de evitar que o algoritmo de busca fique preso em ótimos locais. No trabalho de Zhang, Sun e Li (2009), o uso da representação binária permitiu a variação na quantidade de bandas selecionadas, não necessitando restringir a um número fixo de bandas.

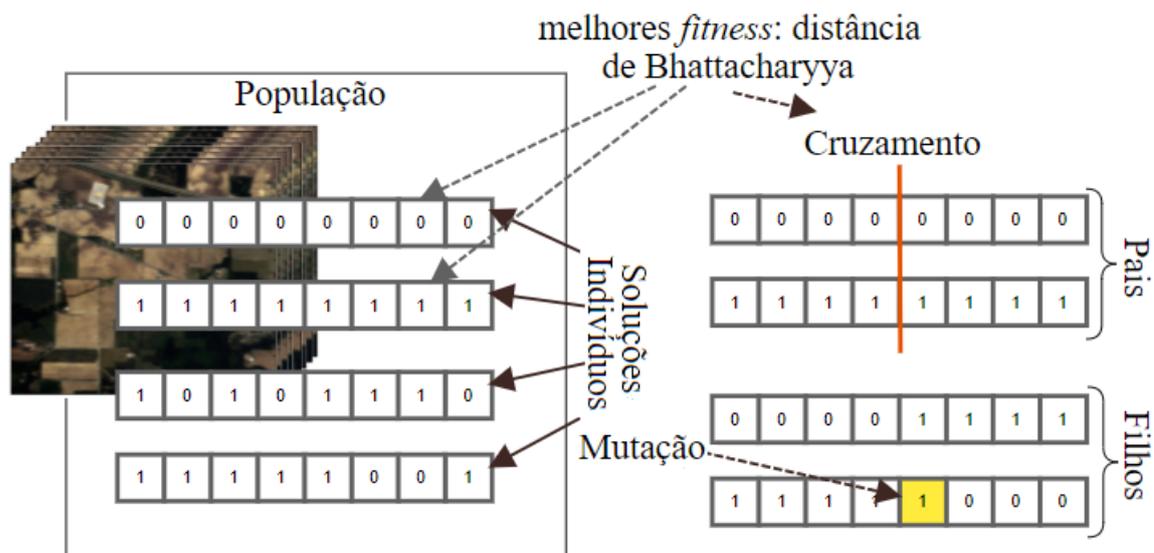


Figura 3.2. Ilustração de elementos de um algoritmo genético com a função de *fitness* baseada na distância de Bhattacharyya. Adaptado de Zhang, Sun e Li (2009).

Em Cui et al. (2011), um método baseado na estratégia *Filter* combinando a seleção de bandas por meio do AG e a extração de características por Análise Linear Discriminante - *Linear Discriminant Analysis* (LDA) lidando com a alta dimensionalidade e resultando em um espaço de características útil. Duas funções de *fitness* são propostas e avaliadas separadamente compondo dois métodos baseados na combinação proposta. Um dos métodos utiliza uma função de *fitness* estruturada a partir da distância de Bhattacharyya obtida por meio da separação entre classes que deve ser otimizada. O outro método é baseado na função de *fitness* estruturada a partir da relação de Fisher. O AG é responsável pela seleção de bandas inicial e o LDA

é aplicado em sequência para gerar um espaço de características otimizado para classificação. Bandas selecionadas por cada um dos métodos são comparadas com bandas selecionadas pela aplicação direta do LDA e uma técnica chamada SLDA. O GMLC foi aplicado para avaliar as bandas selecionadas em cada um dos casos e métodos.

Um dos primeiros trabalhos para seleção de características de imagens baseado na estratégia *Wrapper* usando AGs foi proposto por Siedlecki e Sklansky (1989), onde esse algoritmo foi usado para selecionar características de imagens no infravermelho e os indivíduos foram representados por vetores binários com 30 posições (atributos). A classificação das informações foi realizada com base no algoritmo *K-Nearest Neighbors* (KNN) e o propósito era melhorar o erro de classificação entre duas classes junto com a redução do número de características. Para isso os autores estabeleceram uma função de *fitness* específica que combinava em uma equação uma métrica para o erro de classificação e para o número de características. O método incorpora essa função dentro do processo de busca de AGs, portanto, apresenta características da estratégia *Wrapper*. O procedimento é executado até um número de iterações (gerações) ser alcançado e, em cada geração, os melhores conjuntos de características são mantidos, podendo ser modificados para próxima geração com as operações genéticas. No final, o conjunto de características de melhor *fitness* é retornado com as bandas selecionadas.

Em Bazi e Melgani (2006), o AG é utilizado em conjunto com o SVM com *Kernel Radial Basis Function* (SVM-RBF) para seleção de bandas. Os autores propõem além da seleção de bandas, uma maneira automática de determinar os parâmetros do SVM, sendo esse método baseado na estratégia *Wrapper*. Para isso, utilizam as bandas em indivíduo sendo representadas por posições de um vetor binário, porém adicionam dois valores finais para os parâmetros de regularização e *kernel* do SVM. Testes com duas diferentes funções de *fitness* chamadas de *radius margin bound* e a contagem por vetores de suporte (VS) foram realizados. Nos experimentos realizados, o método baseado em contagem por VS apresentou os melhores resultados entre todos os comparados chegando a alcançar 90,6% de precisão considerando-se 9 classes bem definidas da imagem Indian Pines e 43 características.

Um método semelhante ao de Bazi e Melgani (2006) é apresentado em Zhuo et al. (2008), onde os AGs foram utilizados para seleção de bandas em conjunto com a otimização dos parâmetros do SVM formando um método baseado na estratégia *Wrapper*. O método utiliza média ponderada em uma função de *fitness* que combina a precisão da classificação e a grandeza inversamente proporcional à quantidade de bandas conforme mostrado na equação 3.5:

$$f(x) = w_a \times *precisao + \frac{w_b}{\sum_{i=1}^n g_i}, \quad (3.5)$$

onde w_a é o valor do peso para precisão da classificação, w_b representa o peso para a quantidade de bandas, n é a quantidade de bandas. g_i é o conteúdo da i -ésima posição do vetor, tal que 1 representa que a banda i é selecionada e 0 representa que a banda i não é selecionada. Nesse método é utilizada a estratégia de média ponderada, comum para tentar lidar com mais de um objetivo simultaneamente. Os autores comparam o conjunto de bandas selecionado em relação ao uso de todas as bandas. Com os resultados é mostrado que com um número reduzido de bandas e parâmetros adequados, pode-se melhorar o desempenho de classificação do SVM. Essa combinação de GA-SVM também é estudada em trabalhos recentes, conforme proposto em Nagasubramanian et al. (2017), onde é utilizado para selecionar bandas que auxiliam na identificação precoce de doenças de soja.

Outros algoritmos baseados nos AGs para estratégia *Wrapper* são apresentados em:

- Saqui et al. (2016), em conjunto com o algoritmo GMLC;
- Bris et al. (2014), com o *Sequential Forward Floating Search* (SFFS) e SVM linear; e
- Vaiphasa et al. (2007), com a métrica *Spectral Angle Mapper* (SAM).

Outros métodos baseados em AEs têm sido recentemente explorados para seleção de bandas de forma supervisionada, por exemplo: o método proposto por Su, Yong e Du (2016), que utiliza o algoritmo de otimização Firefly e a distância de Jeffreys-Matusita para otimização da separação de classes. Em Sun et al. (2017) um método de seleção de bandas baseado na estratégia *Wrapper* usando o classificador *extreme learning machine* (ELM) e o algoritmo de otimização Firefly foi proposto. O algoritmo Firefly é usado em momentos diferentes no método proposto. No passo inicial, por meio do Firefly, é realizada uma pré-seleção de bandas, onde busca-se otimizar a distância JM e essas bandas são utilizadas para otimizar o número de neurônios ocultos no ELM. Depois, novamente, a seleção de bandas é realizada pelo Firefly, porém agora, otimiza-se uma função de *fitness* baseada na precisão. Novamente, com o algoritmo Firefly, o número de neurônios da camada oculta é otimizado. Por fim, o algoritmo Firefly otimiza os parâmetros do ELM com as bandas selecionadas.

O CSA foi utilizado em Feng et al. (2014) para estabelecer os métodos *Trivariate mutual information-clonal selection algorithm* (TMI-CSA) e *Semi-supervised band selection approach*

based on TMI and graph regulation (STMIGR), ambos baseados na estratégia *Filter*. Esses métodos tentam otimizar uma métrica baseada na correlação entre três atributos: uma classe e duas bandas, sendo o primeiro método totalmente supervisionado e o segundo semi-supervisionado.

Outros métodos comuns aplicados no contexto de seleção de bandas são o algoritmo de Otimização por Enxame de Partículas, do inglês *Particle Swarm Optimization* (PSO) (XU; DU; YOUNAN, 2017) e o algoritmo de Otimização da Colônia de Formigas, do inglês *Ant Colony Optimization* (ACO) (GAO et al., 2014).

3.1.6 Métodos baseados em algoritmos de otimização multiobjetivo

Muitos métodos de seleção de bandas baseados em AEs operam com uma função de *fitness* baseada em uma única métrica e, por isso, podem não ser adequados para lidar com problemas com múltiplos objetivos. Mesmo quando duas ou mais métricas são combinadas em uma única função, onde normalmente a estratégia de média ponderada é aplicada, os métodos podem não conseguir operar com métricas conflitantes, além de ser difícil definir as proporções da ponderação. Exemplos de métodos que buscam essa ponderação foram previamente apresentados no MAMR proposto por Sui et al. (2015) e no AG-SVM proposto por Zhuo et al. (2008).

Uma alternativa à estratégia da média ponderada com uma única função de *fitness*, é definir uma quantidade fixa de bandas e tentar encontrar a melhor combinação para a função de *fitness* com base nessa quantidade. Porém, isso não permite explorar muitas das possíveis soluções, que têm conjunto de bandas com combinações diferentes que poderiam levar a bons resultados de classificação. Além disso, essa estratégia evita que o algoritmo possa ter simultaneamente uma tendência na redução da quantidade de bandas à medida que melhora a precisão da classificação. Outro problema que pode ocorrer em algoritmos que melhoram uma única métrica de precisão é que podem ignorar a necessidade de redução da quantidade de bandas e o processo de busca ficar "preso" em soluções que apresentam boa classificação, mas que tem elevada quantidade de bandas. Essa situação pode ser indesejável visto que um dos propósitos da seleção é justamente a redução da quantidade de bandas para proporcionar as diversas melhorias previamente citadas. Baseada nessas condições e problemas, a seleção de bandas pode ser redefinida como um problema de otimização multiobjetivo (GONG; ZHANG; YUAN, 2016).

Tratar a seleção de bandas como um problema de otimização multiobjetivo é bastante viável, visto que métodos dessa categoria podem simultaneamente obter resultados que apresentam boa quantidade de bandas e métrica de separabilidade de classes. Essas métricas podem ser conflitantes, pois se durante a tentativa da redução da quantidade de bandas forem removidas bandas importantes para classificação pode-se perder a qualidade de separação das classes. Por-

tanto, o tratamento natural proporcionado por métodos multiobjetivos pode ser adequado para a seleção de bandas. Baseado nessas considerações, a expressão *multiobjective optimization band selection* (MOBS) ou seleção de bandas por otimização multiobjetivo é encontrada em Gong, Zhang e Yuan (2016) para se referenciar a essa categoria de métodos.

Um problema de otimização multiobjetivo é caracterizado por otimizar várias funções de *fitness* simultaneamente e pode ser descrito conforme demonstrado na equação 3.6:

$$\begin{aligned} \text{otimizar } & F(x) = (f_1(x), \dots, f_m(x))^T \\ \text{sujeito a } & x \in \Omega, \end{aligned} \tag{3.6}$$

onde x é uma solução para um problema de otimização; R^m é o espaço de objetivos, e $F : \Omega \rightarrow \mathbb{R}^m$ é o mapeamento das soluções do espaço de busca em valores das m funções de *fitness*.

No contexto da seleção de bandas, x é um conjunto de bandas, $f_1(x)$ pode ser uma função relacionada à quantidade de bandas, $f_2(x)$ uma função relacionada à entropia, $f_3(x)$ relacionada a alguma outra métrica de segmentação e assim por diante. Gong, Zhang e Yuan (2016) defenderam o problema de seleção de bandas como multiobjetivo com base no fato de que diferentes funções de *fitness* para este problema podem ser conflitantes. A melhoria de uma métrica em um determinado instante pode prejudicar a outra, e métodos multiobjetivo são adequados para o tratamento desse conflito buscando por soluções balanceadas. Diferentemente dos métodos com um único objetivo, nos problemas multiobjetivo não existe apenas uma única solução ótima possível, mas um conjunto de soluções que são buscadas pelos métodos de otimização, que por sua vez, são baseados no conceito chamado de Fronteira de Pareto (FP).

A FP é um conjunto constituído por soluções onde não se pode determinar se uma solução é melhor que a outra. Logo, as soluções da FP não podem ser ordenadas considerando todos os objetivos. Sobre essas soluções da FP é dito que elas não são dominadas por nenhuma outra solução, porém são melhores que muitas outras soluções do espaço de busca (COLLETTE; SIARRY, 2003). Para se verificar as relações de dominância entre as soluções é dito que uma solução x domina uma solução y , ou seja, $x \succ y$ se:

- x é igual ou melhor que y para todos objetivos e;
- x é melhor que y em ao menos um objetivo,

x é ótimo global no senso de Pareto, se não existe nenhuma outra solução no espaço de busca que a domine e assim é uma solução da FP.

- **Não Supervisionados:**

Embora o termo MOBS tenha sido usado por Gong, Zhang e Yuan (2016), o primeiro estudo encontrado durante esta pesquisa que utilizou um algoritmo multiobjetivo para o propósito de seleção de bandas foi em Kumar (2004), onde foi utilizado o *Nondominated Sorting Genetic Algorithm 2* (NSGA2) para seleção de bandas. O NSGA2 originalmente proposto por Deb et al. (2002) é um dos algoritmos para problemas multiobjetivo mais tradicionais e utilizados, principalmente em aplicações do mundo real. Na formulação original, o NSGA2 superava outros métodos clássicos da época em termos de complexidade computacional e em encontrar soluções adequadas. Uma de suas principais características é sua capacidade de selecionar soluções para as operações genéticas de cruzamento e mutação considerando um método de ordenação baseado na FP e na relação de dominância chamado de *Fast nondominated-sorting*, que estabelece um ranqueamento conforme a não dominância de soluções (DEB et al., 2002).

O método de seleção de bandas proposto por Kumar (2004) é baseado na estratégia *Filter* e é constituído por duas etapas apresentadas na sequência:

- A primeira etapa consiste em selecionar um subconjunto de bandas espectrais utilizando o NSGA2 com os três objetivos:
 1. maximizar a Informação Associada com as bandas espectrais;
 2. maximizar a Informação Associada com a entropia e ;
 3. reduzir o número de bandas.

A Informação Associada é uma métrica baseada no valor absoluto de correlação (chamado de *factor loadings*) de cada banda, obtida a partir de sua importância em cada um dos componentes principais calculados pelo PCA. Essa métrica deve ser maximizada, portanto, bandas com baixo valor de Informação Associada devem ser descartadas. A partir do NSGA2 com os respectivos objetivos estabelecidos, um subconjunto de bandas pode ser selecionado conforme a FP.

- Na segunda etapa, esse subconjunto de bandas é utilizado no processo que consiste no uso de um algoritmo de otimização de função única que visa maximizar a separação das classes. O objetivo desse algoritmo é reduzir a função de *fitness Brightness Value Overlap Index*, que permite quantificar a separação das classes.

Além de métodos tradicionais como o NSGA2 e o *Strength Pareto Evolutionary Algorithm 2* (SPEA2), que são diretamente baseados na atribuição de um *fitness* para não dominância (SAQUI et al., 2019b), outras estratégias podem ser exploradas, como a decomposição de soluções em problemas de otimização escalar.

Uma solução ótima da FP para um problema multiobjetivo pode ser aproximada por x^1, \dots, x^n soluções ótimas de $1, \dots, n$ subproblemas de otimização escalar. Neste caso, a função de *fitness* de uma solução é uma agregação de todas as funções de *fitness* f_i individuais (ZHANG; LI, 2007; GONG; ZHANG; YUAN, 2016). O *Multiobjective Evolutionary Algorithm based on Decomposition* (MOEA/D) proposto por Zhang e Li (2007) é uma estratégia multiobjetivo baseada na decomposição de problemas e uma ilustração pode ser visualizada na Fig. 3.3, onde F^1, \dots, F^n são vetores de funções de *fitness* para possíveis soluções x^1, \dots, x^n .

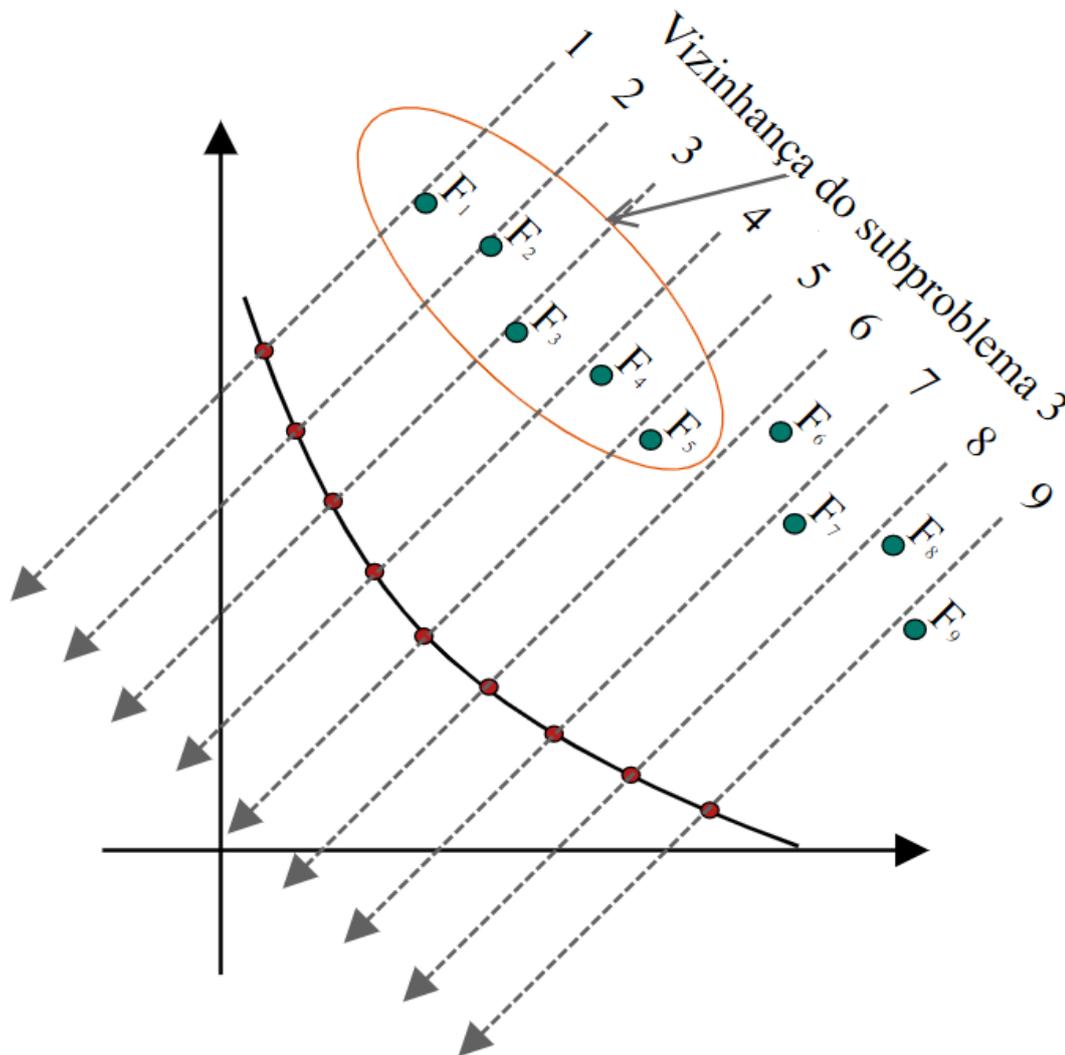


Figura 3.3. Representação do MOEA/D onde um problema multiobjetivo foi decomposto em 9 subproblemas e a vizinhança do subproblema 3 é composta pelos subproblemas 1, 2, 3, 4 e 5 (adaptado de Zhou, Zhang e Zhang (2012)).

Existem muitas abordagens para converter o problema de aproximação da PF em subproblemas de otimização escalar, porém a mais explorada é a estratégia de decomposição de Tchebycheff, do inglês *Tchebycheff Decomposition Strategy* (TDS), explorado em Gong, Zhang e Yuan (2016) como uma estratégia *Filter* para seleção de bandas de forma não supervisionada. O TDS é usado para agregar os f_1, \dots, f_m objetivos em uma única função de agregação, onde os valores

são obtidos por meio da diferença dos valores de *fitness* de uma solução x e um ponto de referência ideal $z^* = [z_1^*, \dots, z_m^*]^T$. A decomposição usando TDS para o i -ésimo subproblema com a solução x é mostrada na eq. 3.7 (ZHOU; ZHANG; ZHANG, 2012):

$$\min g^i(x) = \min g(x|\lambda^i, z^*) = \max_{1 \leq j \leq m} \lambda_j^i |f_j(x) - z_j^*|, \quad (3.7)$$

onde $\lambda^i = (\lambda_1^i, \dots, \lambda_m^i)^T$ é o vetor de pesos para os diferentes m objetivos.

No processo de evolução dos subproblemas, dada uma determinada solução, a função de agregação das soluções vizinhas e novas soluções geradas por cruzamento e mutação são calculadas para posterior definição das soluções que podem ser atualizadas. Soluções que têm altos valores de g têm maiores chances de serem substituídas por aquelas outras que estão mais próximas da solução ideal (XU; SHI; PAN, 2017).

Cada subproblema é otimizado pelo uso da informação obtida a partir de subproblemas vizinhos (como mostrado para o subproblema 3 da Fig. 3.3) e não do conjunto todo. A estratégia de otimizar a partir da vizinhança faz com que o MOEA/D tenha um custo computacional menor que os algoritmos clássicos. Como soluções ótimas para dois subproblemas vizinhos podem ser muito similares (no espaço de decisão e/ou no domínio de objetivos), considerar somente informações desses subproblemas pode também promover um melhor refinamento do espaço de decisão (LOPEZ, 2017). Em sua proposta original, o MOEA/D foi comparado com outros algoritmos da literatura além do NSGA2, como o *multiobjective genetic local search* (MOGLS), onde apresentou melhores resultados de complexidade computacional e melhores resultados para problemas de otimização (ZHANG; LI, 2007).

No estudo de Gong, Zhang e Yuan (2016), a seleção de bandas é modelada como um problema multiobjetivo e duas funções com relacionamento conflitante são projetadas. Uma função é baseada na entropia enquanto a outra visa reduzir a quantidade de bandas. Conforme já apresentado, a entropia está relacionada ao montante de informação presente em cada banda e conforme os autores, maiores valores significam uma maior quantidade de detalhes nas bandas espectrais. Os autores discutem que as bandas com maior entropia da informação apresentam a possibilidade de maior distinção nos objetos presentes na imagem e, portanto, são mais informativas. As IHS testadas são segmentadas pelos algoritmos KNN, SVM e ELM. As bandas selecionadas pelo algoritmo proposto foram comparadas com bandas selecionadas por uma série de outros métodos clássicos. Alguns exemplos de métodos comparativos foram o WaLuMi, WaLuDi e MVPCA. Nos testes com três imagens clássicas da literatura, Indian Pines, Pavia University e Salinas, o método proposto foi capaz de obter subconjuntos de bandas que promoveram um desempenho de classificação competitivo e estável.

Em Xu, Shi e Pan (2017) foi proposto um método chamado *Incorporated Rank Based Multi-Objective Band Selection Algorithm* (IRMoBS) baseado no MOEA/D e utilizando três funções de *fitness* que devem ser minimizadas: uma baseada na entropia (eq. 3.8), outra na variância (eq. 3.9) e outra no número de bandas (eq. 3.10):

$$f_1(b) = \frac{1}{|b|} \sum_{i=1}^{|b|} \frac{1}{H(x_b^i)}, \quad (3.8)$$

$$f_2(b) = \frac{1}{|b|} \sum_{i=1}^{|b|} \frac{1}{S(x_b^i)}, \quad (3.9)$$

$$f_3(b) = ||b| - k|, \quad (3.10)$$

onde b é a representação de um vetor binário de bandas, X_b denota os dados das bandas selecionadas, $H(x_b^i)$ e $S(x_b^i)$ são, respectivamente, a entropia de informação e a variância da i -ésima banda. $|b|$ é a cardinalidade de b e k é a quantidade de bandas desejada. Particularmente, nesse estudo, um tomador de decisões, do inglês *decision maker* (DM), incorporado ao processo de seleção de bandas foi proposto, o que permite escolher automaticamente uma solução da FP.

A seleção de uma solução da FP é uma tarefa difícil. Além da estratégia proposta no IRMoBS, outros métodos podem ser empregados. Um exemplo é mostrado em Kimovski et al. (2018), onde um DM foi proposto para selecionar uma única solução da FP. Para isso, regiões são estabelecidas conforme o valor de centroide obtido a partir das soluções da FP. Esse centroide funciona como um limiar, criando regiões com soluções balanceadas e outras mais tendenciosas para uma das funções de *fitness*. A partir dessas regiões podem ser estabelecidas regras que permitem selecionar uma determinada solução. Esse método é explorado nesta tese e alguns detalhes são demonstrados no capítulo posterior.

- **Supervisionados:**

Além do NSGA2 ter sido explorado como um método não supervisionado baseado na estratégia *Filter*, recentemente em Saqui et al. (2019), ele foi explorado como método supervisionado. Nesse estudo, o NSGA2 foi utilizado diretamente para melhorar a acurácia balanceada (*recall* médio) de classificação gerada pelo GMLC enquanto reduz a quantidade de bandas das IHS. Uma característica importante e moderna em relação a outros métodos da literatura, é que esse método incorpora o DM proposto por Kimovski et al. (2018) ao processo de seleção de bandas, o que permite selecionar diretamente uma única solução da FP. As bandas selecionadas

pelo método proposto foram comparados com bandas selecionadas por uma estratégia baseada em GA e pelo IRMOBS, sendo que para o classificador GMLC proporcionaram os melhores resultados para as imagens consideradas.

3.1.7 Métodos baseados em esquemas híbridos

Métodos dessa categoria implementam vários esquemas dos métodos previamente apresentados conjuntamente com o propósito de selecionar bandas apropriadas para uma determinada tarefa. Alguns métodos com elementos híbridos já foram apresentados, porém nesta subseção novos métodos com características híbridas bem definidas são apresentados (SUN; DU, 2019).

- **Não Supervisionados:**

Métodos híbridos de seleção de bandas que utilizam *clustering* e foram encontrados durante esta pesquisa podem ser organizados da seguinte forma:

1. métodos de otimização que são utilizados para otimizar parâmetros de algoritmos de *clustering*, como demonstrado em Zhang, Ma e Gong (2017), onde o *Particle Swarm Optimization* atenua desvantagens do *Fuzzy-C-Means* como a escolha dos pontos iniciais e evita que o algoritmo fique preso em ótimos locais; ou então
2. métodos de *clustering* que são utilizados no processo de seleção em conjunto com outros métodos. Um exemplo é mostrado em Wang et al. (2017), onde foi proposto o método *trace Lasso-based band selection*, composto pelo método de otimização *Least Absolute Shrinkage and Selection Operator* (LASSO), representações esparsas e *spectral clustering* (SC) para seleção de bandas com base na correlação.

Outro exemplo da segunda categoria apresentada anteriormente é mostrado em Wang, Lin e Yuan (2016), onde foi proposto um algoritmo chamado de *Salient Band Selection via Manifold Ranking* (MR). MR é uma técnica de ranqueamento baseada em teoria dos grafos e utiliza métricas de similaridade para estabelecer relacionamentos entre os elementos que estão sendo considerados, para que posteriormente possam ser estabelecidos critérios de ranqueamento. O *Salient Band Selection via MR* recebe como parâmetros de entrada um conjunto de bandas espectrais e a quantidade de bandas desejadas. Durante sua execução, primeiramente, são gerados grupos com subconjuntos de bandas com base na similaridade entre bandas. Esses grupos iniciais podem ser gerados por técnicas de agrupamento como o K-Means. Para esse agrupamento, os diferentes pixels das imagens são utilizados como dimensões e cada banda espectral está

inserida em um espaço Euclidiano com essas dimensões. Assim, pode-se determinar a similaridade entre diferentes bandas espectrais e formar os respectivos grupos. Posteriormente, uma banda representativa de cada grupo é selecionada por meio de um algoritmo de seleção clonal, baseado no paradigma de sistemas imunológicos artificiais. Esse método de seleção é aplicado para minimizar a distância Euclidiana entre elementos de uma mesma classe, enquanto maximiza a distância de elementos de classes diferentes. Após isso, se a quantidade de bandas selecionada é menor que a quantidade definida pelo usuário, as bandas já escolhidas são usadas como as entradas do algoritmo MR e usadas como base para ordenar as demais bandas, ou seja, aquelas que não foram selecionadas como representativas. O MR ordena as bandas remanescentes utilizando um ranqueamento baseado em uma matriz de afinidade que, por sua vez, utiliza a mesma métrica de distância utilizada no processo de agrupamento. A banda mais dissimilar é adicionada para o grupo de bandas representativas. Esse passo é repetido até uma quantidade desejada de bandas, pré-definida pelo usuário, ser alcançada.

- **Supervisionados:**

Tuv et al. (2009) e Sun et al. (2018) discutem que abordagens híbridas entre estratégias *Embedded* e *Wrapper* podem promover melhores resultados de classificação que os métodos tais como àqueles baseados em árvores de decisão. Um exemplo de trabalho dessa categoria é explorado por Stein et al. (2005), onde os AGs são utilizados para seleção de características que são utilizadas em árvores de decisão. As árvores de decisão geradas com as características selecionadas pelo método apresentam melhores resultados em relação à técnica de árvores de decisão puramente baseadas na estratégia *Embedded*. Outro trabalho semelhante, porém, operando com MOBS baseado na estratégia *Wrapper* foi apresentado por Saqui et al. (2019b), onde o algoritmo *Strength Pareto Evolutionary Algorithm* (SPEA2) com os objetivos de melhorar o desempenho de classificação e reduzir o número de bandas, que é utilizado para formar árvores de decisão em um classificador RF. SPEA2 é um algoritmo de otimização multiobjetivo que tem a característica importante de utilizar uma população externa onde são guardadas as soluções não dominadas (ZITZLER; LAUMANN; THIELE, 2001). O trabalho mostra a evolução ao longo das iterações do SPEA2 dos resultados da precisão da classificação do RF e a redução do número de bandas. Com as bandas selecionadas, um outro algoritmo classificador, o GMLC, também foi aplicado e os resultados são melhores quando comparados com a utilização de todas as bandas ou com bandas selecionadas por um método baseado em AG.

Um método híbrido foi apresentado em Li et al. (2011), que conseguiram resultados melhores que o método proposto por Bazi e Melgani (2006), previamente apresentado, adicionando

um agrupamento de bandas com alta correlação antes da execução do AG. Essa é uma boa estratégia para reduzir o custo computacional e alta correlação durante o processo de busca. Os autores utilizam a IM para mensurar a correlação entre bandas e agrupam as bandas de alta correlação até uma determinada quantidade. Dessa forma, é possível selecionar apenas uma banda representante de cada grupo evitando redundâncias. Depois, o AG com SVM-RBF foram aplicados conjuntamente, constituindo um método baseado na estratégia *Wrapper* para buscar uma combinação de bandas. O AG seleciona apenas uma banda de cada grupo, reduzindo o espaço de busca em relação aos casos gerais. A função de *fitness* utilizada representa a precisão da classificação. Por último, o algoritmo *branch and bound* foi utilizado para remover os grupos de bandas irrelevantes do resultado.

Outro método importante chamado *Information Gain - Gray Wolf Optimizer (IG-GWO)*, que combinou as estratégias *Filter* e *Wrapper*, foi apresentado em Xie et al. (2018). O método é baseado no ganho de informação como um critério de ranqueamento de bandas e no GWO, sendo esse último um algoritmo de otimização evolucionário capaz de oferecer resultados competitivos com outros métodos meta-heurísticos da literatura. Bandas com maior ganho de informação proporcionam melhores resultados de classificação, porém, o uso simples dessa métrica para seleção pode ocasionar a presença de bandas altamente correlacionadas causando perda no desempenho da classificação. No IG-GWO, inicialmente as bandas são agrupadas conforme o ganho de informação e na curva espectral. Posteriormente, k bandas são selecionadas aleatoriamente de cada grupo para constituir o conjunto que é manipulado pelo GWO. Finalmente, o GWO é executado e garante que a combinação de bandas selecionadas proporciona uma boa qualidade de classificação e evita bandas redundantes.

Em Habermann, Fremont e Shiguemori (2019), um método supervisionado baseado na estratégia *Filter* e redes neurais artificiais foi proposto. Para cada classe, um classificador baseado em redes neurais de uma única camada realiza a classificação entre a classe e os dados remanescentes. Nesse esquema, o modelo um contra-todos é adotado. Depois disso, as bandas relacionadas com os maiores e menores pesos são selecionadas, portanto, o método é orientado pelas classes. Ao final do processo, se o número de bandas for maior que o desejado pelos usuários, uma técnica de *clustering* K-Means é aplicada para gerar um número de grupos igual ao número de bandas desejado pelo usuário.

Por fim e elaborado também durante esta pesquisa, em Saqui et al. (2019c) um método chamado *Incorporated Decision-Maker-Based Multiobjective Band Selection (IDMMoBS)*, que pode ser considerado uma versão prévia do método estruturado nesta tese foi apresentado. Esse método combina elementos do algoritmo WaLuMi para inicialização de soluções, MOEA/D

para seleção de bandas multiobjetivo e SVM em uma estratégia *Wrapper* e um DM para seleção de uma solução da fronteira de Pareto. Detalhes desse método são apresentados conjuntamente ao método proposto nos capítulos seguintes.

3.2 Considerações Finais do Capítulo

Neste capítulo, foram apresentados os principais métodos de seleção de bandas encontrados durante esta pesquisa. Os métodos foram categorizados conforme os tipos de algoritmos ou estratégias nos quais foram baseados, demonstrando quando eram supervisionados/semi-supervisionados ou não supervisionados e, também, se eram baseados nas estratégias *Filter*, *Wrapper* ou *Embedded*. Alguns desses métodos foram escolhidos para comparações com o resultado de bandas selecionadas pelo método proposto, onde buscou-se diversificar as características entre os métodos escolhidos. Os métodos escolhidos são:

- Métodos multiobjetivos, que foram escolhidos por serem alguns dos poucos exemplares dessa categoria assim como o próprio método proposto neste tese:
 - *Incorporated rank based multi-objective band selection algorithm (IRMOBS)* (XU; SHI; PAN, 2017): Método baseado na estratégia *Filter*, não supervisionado e com os objetivos baseados na entropia, variância e número de bandas;
 - *Incorporated Decision-Marker-based multiobjective band selection (IDMMoBS)* (SAQUI et al., 2019c): Método baseado na estratégia *Wrapper* e supervisionado desenvolvido durante essa pesquisa e similar ao WMoEBS;
- Métodos recentes com único objetivo, que foram escolhidos por apresentarem os componentes mais atuais explorados na literatura:
 - *Trivariate mutual information-clonal selection algorithm (TMI-CSA)* (FENG et al., 2014): Baseado em uma adequação da IM que considera a correlação entre três variáveis, o rótulo da classe e duas bandas. Utiliza o CSA sendo um método de busca por otimização baseado na estratégia *Filter* e aprendizado supervisionado;
 - *Semi-supervised band selection approach based on TMI and graph regulation (ST-MIGR)* (FENG et al., 2014): Semelhante ao TMI-CSA, porém utiliza uma estratégia de aprendizado semi-supervisionado que permite a propagação de rótulos;
 - *Maximum information and minimum redundancy-clonal selection algorithm (MIMR-CSA)* (FENG et al., 2016): Utiliza o CSA para otimizar uma função derivada da IM

- e entropia chamada de *Maximum information and minimum redundancy* (MIMR). Esse método opera de forma não supervisionada, com base na estratégia *Filter* e utiliza busca por otimização;
- *Information gain - gray wolf optimizer* (IG-GWO) (XIE et al., 2018): Opera como um método supervisionado com uma função de *fitness* baseada no ganho de informação das bandas, porém baseado na estratégia *Filter*, pois não incorpora um método de classificação ao processo de seleção de bandas;
 - *Single-layer neural network* (SLN) (HABERMANN; FREMONT; SHIGUEMORI, 2019): O método considerado *Filter* porque usa a SLN para selecionar bandas e outros algoritmos para classificação;
- Métodos que são comumente comparados ou apresentaram bons resultados na literatura:
 - *Ward's linkage and Mutual Information* (WaLuMi) (MARTINEZ-USO et al., 2007): Método baseado em agrupamento hierárquico na estratégia *Filter* e não supervisionado que utiliza a informação mútua entre bandas;
 - *Classification and regression tree* (CART) (GOMEZ-CHOVA et al., 2003): Um método supervisionado baseado em árvores de decisão e na estratégia *Embedded*;
 - *Genetic Algorithm with Support Vector Machines* (GA-SVM) (NAGASUBRAMANIAN et al., 2017; ZHUO et al., 2008): Um método baseado na estratégia *Wrapper*, supervisionado e utilizando uma função de *fitness* com média ponderada entre OA da classificação e quantidade de bandas. Essa combinação é tradicional na literatura, sendo explorada por meio de diferentes formas conforme apresentado em Nagasubramanian et al. (2017), Zhuo et al. (2008). Aqui é utilizada a função de *fitness* proposta por Zhuo et al. (2008) e previamente apresentada na equação 3.5.

Também foram discutidas as características mais comuns entre os métodos encontrados e quais deles eram mais relevantes. Um exemplo de uma característica desses métodos que pode ser importante é a IM, usada em muitos dos métodos encontrados para reduzir a correlação entre bandas. Métodos baseados em otimização multiobjetivo também demonstram uma capacidade interessante que é a de lidar com objetivos conflitantes, permitindo buscar por soluções que permitem uma boa capacidade de classificação com uma quantidade reduzida de bandas. A informação espacial também se demonstra um recurso importante para qualidade da classificação. Assim, o estudo dessas características permitiu elaborar o *framework* proposto como base para o método proposto nesta tese, além de selecionar quais delas que foram os componentes avaliados para composição do proposto.

Capítulo 4

MATERIAIS E MÉTODOS

O Capítulo 1 desta tese apresentou a motivação, problemas e objetivos que foram explorados nesta pesquisa. O Capítulo 2 descreveu as importantes teorias envolvidas com seleção de bandas em imagens hiperespectrais (IHs), alta dimensionalidade, técnicas de classificação e segmentação que são aplicadas em IHs e o Capítulo 3 apresentou alguns dos trabalhos da literatura relacionados com a redução da dimensionalidade. O capítulo atual apresenta a metodologia aplicada durante o desenvolvimento desta pesquisa, que junto a ações conduzidas e apresentadas nos outros capítulos desta tese, possibilitaram o planejamento e execução de atividades importantes. Dessa forma, ele está organizado nas seguintes etapas:

1. Apresentação das IHs (*datasets*), as quais foram realizados os experimentos para elaboração do *Wrapper multiobjective evolutionary band selection* (WMoEBS) e experimentos comparativos com outros métodos da literatura;
2. Apresentação das ferramentas utilizadas no desenvolvimento desta pesquisa; e
3. Apresentação dos critérios e formas de análises de resultados consideradas.

4.1 Amostras / Imagens Hiperespectrais (IHs)

Existe uma limitação em IHs com GT disponíveis gratuitamente e de fácil acesso. Porém, essas IHs possuem uma grande quantidade de pixels que podem ser utilizados como amostras nos experimentos realizados. Algumas dessas imagens não são exclusivamente de áreas agrícolas, mas permitem aplicações similares e, portanto, podem ser utilizadas para seleção de bandas. Isso possibilita a avaliação de métodos supervisionados baseados na estratégia *Wrapper*, uma vez que precisam dos grupos de amostras de pixels para treinamentos e testes. Portanto, para os

experimentos realizados nesta tese, foram consideradas três IHs clássicas da literatura ¹.

Na sequência, essas imagens estão organizadas junto com uma descrição, número de bandas, sua representação em RGB, e seus respectivos GTs com as classes presentes. A região na cor preta presente nos GTs representa uma área que não foi fornecida por nenhum especialista, ou seja, não representam classes úteis para o presente estudo. Portanto, para executar os procedimentos necessários nessas imagens são consideradas apenas as regiões coloridas dos GTs e as suas respectivas áreas nas IHs originais ou com bandas selecionadas. São elas:

- Indian Pines (Figura 4.1): A imagem foi capturada pelo sensor de dados hiperespectrais *Airborne Visible / Infrared Imaging Spectrometer (AVIRIS)* no noroeste do estado de Indiana no Estados Unidos (EUA). Essa imagem tem 220 bandas, sendo de tamanho de 145 X 145 pixels, ou seja, um total de 21025 pixels, porém 10249 pixels possuem representação no GT. As bandas espectrais estão na faixa do comprimento de onda de 400 - 2500 nanômetros (nm). Vinte bandas que sofrem influência da atmosfera pela presença de vapor d'água e que não contém informações espectrais com qualidade para caracterização dos alvos foram eliminadas para realização dos experimentos. Portanto restaram 200 bandas para utilização. O GT, ou seja, o mapa com culturas e outros objetos presentes na imagem contém 16 classes que representam alfafa, tipos de diferentes plantios de milho, grama, semeados com pastagem, feno, aveia, soja, trigo, vegetação natural e construções.

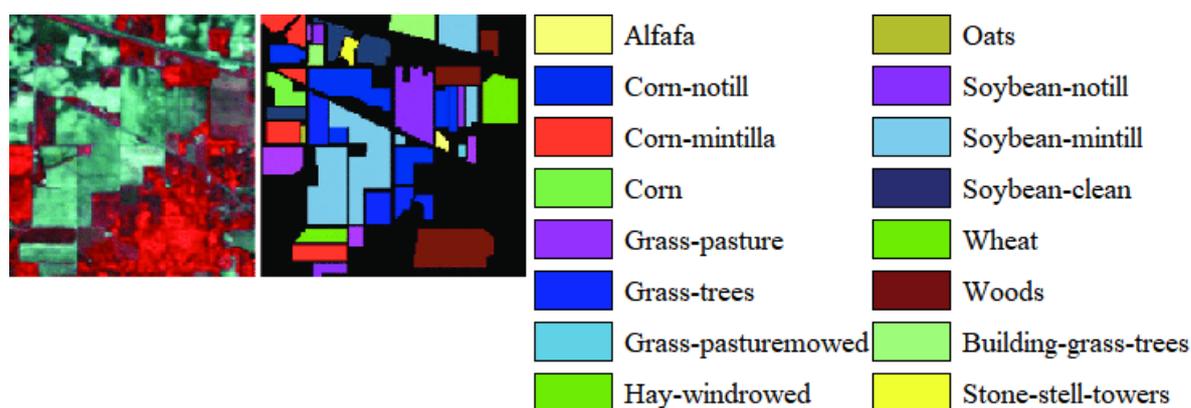


Figura 4.1. Imagem Hiperespectral Indian Pines, GT e legenda com classes.

- Salinas (Figura 4.2): A imagem de 512 x 217 pixels (total de 111104 pixels) também coletada pelo sensor AVIRIS, sobre a região do Salinas Valley (vale em Salinas) no estado da Califórnia nos EUA. Nessa imagem 54129 pixels possuem GT e estão adequados

¹Obtidas em: <http://www.ehu.es/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes>. Acesso em: 05 de Setembro de 2019

para avaliação de classificadores. Essa imagem tem resolução espacial de 3,7 metros por pixel e contém 224 bandas, sendo utilizadas 204 bandas, onde, assim como para imagem Indian Pines, 20 bandas de absorção de vapor d'água foram descartadas. O GT (Figura 4.2) dessa imagem contém 16 classes que representam culturas de brócolis, pousio (terra em repouso sem semeadura), restolho (palha que fica no campo após a colheita), salsaão, uvas em diferentes estágios de desenvolvimento, milho, alface com diferentes semanas de desenvolvimento.

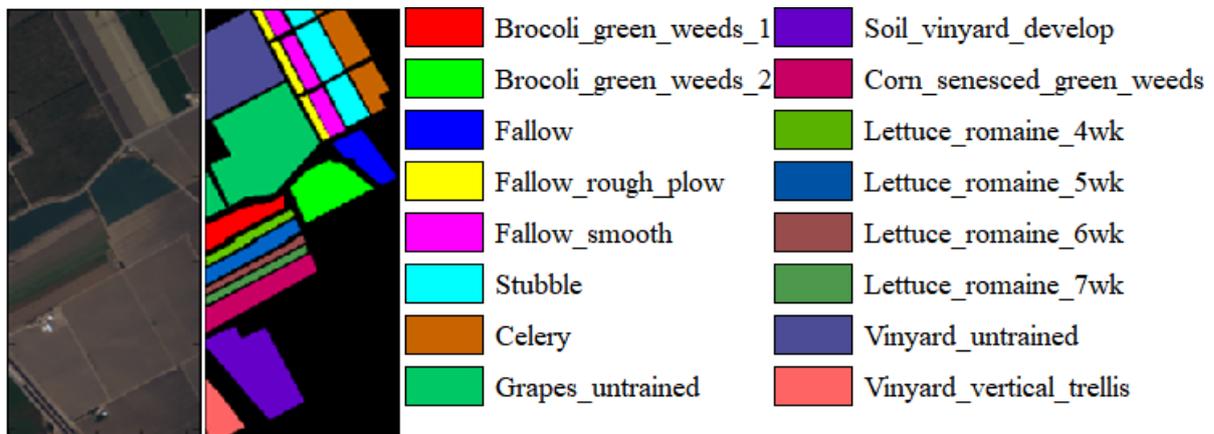


Figura 4.2. Imagem Hiperespectral Salinas, GT e legenda com classes.

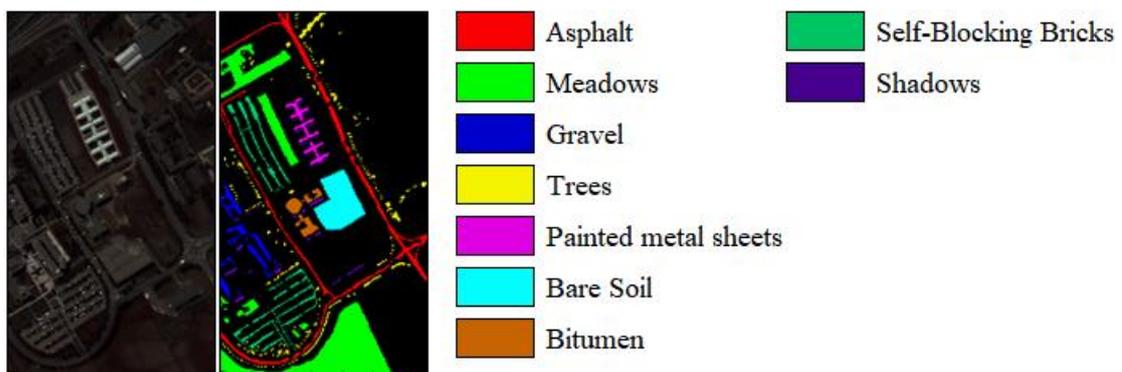


Figura 4.3. Imagem Hiperespectral Pavia University, GT e legenda com classes.

- Pavia University (Figura 4.3): Imagem de 610 x 340 pixels (total de 372100 pixels, onde 42776 com GT) e 103 bandas obtida pelo sensor *Reflective Optics System Imaging Spectrometer* (ROSIS) da universidade de Pavia na Itália. A IH apresenta nove classes, sendo asfalto, pastagem, cascalho, árvores, construções metálicas, solo limpo, construções, sombras, entre outras.

4.2 Ferramentas utilizadas

Para utilização das imagens selecionadas como conjuntos de dados, elas foram convertidas para o formato *Comma-separated values* (.csv), com atributos/bandas separados por ponto e vírgula (;) e cada linha representando um pixel de amostra. A vantagem do uso desse tipo de arquivo, é que pode ser facilmente manipulado na linguagem Python com o auxílio das bibliotecas Numpy e Pandas ². Conseqüentemente, essas estruturas podem facilmente serem manipuladas em bibliotecas específicas de Aprendizado de Máquina como o Scikit-learn ³.

Devido à facilidade proporcionada por esses recursos, nesta tese foram utilizadas as seguintes ferramentas:

- Linguagem de programação Python;
- Bibliotecas tradicionais Numpy, Pandas e Scikit-learn;
- Biblioteca *Distributed Evolutionary algorithms in Python* (DEAP) ⁴, que auxilia no desenvolvimento de algoritmos evolutivos; e
- Biblioteca Spectral Python (SPy) ⁵, que auxilia na manipulação de IHS.

4.3 Configurações dos experimentos e critérios de avaliação

Nesta tese, dois grupos de experimentos foram necessários e realizados:

- Grupo 1: Esse grupo teve o objetivo de avaliar experimentalmente os componentes que estruturaram o WMoEBS e junto com a metodologia apresentada neste capítulo, faz parte dos procedimentos específicos utilizados para construção do método proposto.
- Grupo 2: Esse grupo de experimentos teve o propósito de comparar o WMoEBS já estruturado com outros métodos da literatura e com a utilização de todas as bandas (IHS original), para demonstrar sua eficiência em relação aos critérios avaliados e atingir um dos objetivos desta tese.

Para avaliação desses grupos, foram propostas estratégias específicas para cada caso e sub-grupos para cada caso foram organizados. Uma das estratégias é discutida na sequência e trata

²Disponível em: <<https://pandas.pydata.org>>. Acesso em: 05 de Setembro de 2019

³Disponível em: <<https://scikit-learn.org/stable/>>. Acesso em: 05 de Setembro de 2019

⁴Disponível em: <<https://deap.readthedocs.io/en/master/>>. Acesso em: 05 de Setembro de 2019

⁵Disponível em: <<http://www.spectralpython.net>>. Acesso em: 05 de Setembro de 2019

das condições para permitir uma avaliação justa quando a informação espacial obtida por meio de filtros é utilizada.

4.3.1 Estratégia de captura de amostras das IHS sem sobreposição

Conforme apresentado por Liang et al. (2017) e observado durante esta pesquisa, na literatura, existem muitos métodos que tentam melhorar a classificação, porém poucos estudos se preocupam com o projeto e organização dos experimentos e testes. Ainda segundo Liang et al. (2017), muitas pesquisas utilizam de maneira inadequada as técnicas de processamento espectral-espacial para classificação, causando avaliações de desempenho injustas e tendenciosas devido a violações da suposição de independência dos dados.

Quando utilizada a informação espacial, avaliações injustas podem ocorrer quando amostras de treinamento e testes são obtidas de forma aleatória. Nesse caso, a dependência causada por sobreposição entre amostras de treinamento e teste vizinhas pode ser ocasionada por métodos que permitem obter informação espacial, como os próprios filtros espaciais. A utilização de amostragem aleatória dificulta saber se a melhoria da qualidade da classificação é causada pela incorporação de informações espaciais ou pelo aumento da sobreposição entre treinamento e amostras de teste. Portanto, a estratégia de amostragem aleatória pode não ser adequada para avaliar algoritmos de classificação em imagens quando esses utilizam informação espacial. Para evitar o problema de sobreposição entre amostras de treinamento e testes, nos experimentos realizados nesta tese uma estratégia específica foi proposta.

A estratégia proposta considera que para cada imagem sejam estabelecidos grupos de pixels organizados por *kernels* de tamanhos, 2×2 ou 3×3 que são adequados conforme o tipo do filtro aplicado e sempre são sem sobreposição. Esses *kernels* permitem controlar a influência dos pixels vizinhos a uma amostra, permitindo reduzir ruídos comuns em IHS e melhorando os resultados da classificação. Com os *kernels* são constituídos grupos de 4 ou 9 pixels vizinhos.

A Figura 4.4 mostra um exemplo dos grupos de pixels estabelecidos com *kernel* 3×3 (visivelmente agrupados por linhas azuis na Figura), onde os pixels do mesmo grupo estão representados pelo mesmo número. Sempre o pixel central de cada grupo, que nessa Figura está preenchido na cor preta, é o pixel a ser selecionado como amostra para os experimentos, evitando sobreposição de amostras vizinhas indesejadas. Quando o *kernel* 2×2 é aplicado, o pixel inferior no canto direito é selecionado também evitando sobreposição.

Essa estratégia é aplicada nos métodos que envolvem a utilização da informação espacial e são comparados entre si, ou mesmo quando algum método que faz uso da informação espacial

1	1	1	2	2	2	3	3	3
1	1	1	2	2	2	3	3	3
1	1	1	2	2	2	3	3	3
4	4	4	5	5	5	6	6	6
4	4	4	5	5	5	6	6	6
4	4	4	5	5	5	6	6	6

Figura 4.4. Grupos de pixels usados como amostras para experimentos usando $kernel\ 3 \times 3$

é comparado com outros que não a consideram. A estratégia evita que os pixels selecionados nos experimentos tenham influência uns em relação aos outros e, assim, as avaliações realizadas podem ser justas e evitar qualquer tipo de tendência.

Nas imagens consideradas, mesmo a máscara de tamanho 3×3 , que pode não ser adequada para alguns dos filtros apresentados, permite a aplicação de todos os casos apresentados no capítulo anterior e por isso foi utilizada.

4.3.2 Configuração de amostras de treino, validação e testes

Baseado na estratégia sem sobreposição e, também, quando a informação espacial não foi necessária, dois critérios foram considerados para obtenção e configurações de amostras:

1. Critério com uso da informação espacial: Foi utilizada a estratégia sem sobreposição com um dos *kernels* apresentados anteriormente, evitando avaliações injustas, porém limitando a quantidade de pixels usados. Para seleção dos pixels foi utilizada uma abordagem de amostragem estratificada e sistemática explicada posteriormente; e
2. Critério sem uso da informação espacial: Também foi utilizada uma abordagem de amostragem estratificada e sistemática por classes, porém utilizando todos os pixels rotulados.

Para todos os métodos e componentes avaliados nesta tese foram utilizadas estratégias de validação cruzada com 5 grupos (*5-fold cross-validation*). Uma representação da estratégia *5-fold cross-validation* pode ser visualizada na Figura 4.5, onde para cada *fold* (grupo de pixels), os pixels de amostras separados para treinamento formaram um montante de 80%, ou seja, $\frac{4}{5}$ da quantidade total de pixels. Esses pixels são representados pelos blocos azuis da Figura 4.5 e

foram utilizados durante cada um dos processos de seleção de bandas, sendo divididos em dois subgrupos, sendo um para treinamento e outro para validação.

	<i>Fold 1</i>	<i>Fold 2</i>	<i>Fold 3</i>	<i>Fold 4</i>	<i>Fold 5</i>
Grupo A	Treino	Treino	Treino	Treino	Teste
Grupo B	Treino	Treino	Treino	Teste	Treino
Grupo C	Treino	Treino	Teste	Treino	Treino
Grupo D	Treino	Teste	Treino	Treino	Treino
Grupo E	Teste	Treino	Treino	Treino	Treino

} *Cross-Validation*
5-fold

Figura 4.5. Configuração das amostras para os experimentos com *5-fold Cross-Validation*

Os pixels de teste estão representados pela cor laranja na Figura 4.5 e uma característica importante, foi a forma como os 20%, ou seja, $\frac{1}{5}$ dos pixels de testes de cada um dos *folds* foram organizados. Para essa organização, uma estratégia de separação foi aplicada por meio de amostragem estratificada e sistemática para formar cada um dos *folds*, sendo essa estratégia descrita na sequência:

- Primeiramente, os pixels (amostras) foram organizados em classes obtidas a partir do GT;
- Então, em cada classe, os pixels foram ordenados conforme as posições originais na imagem, observando primeiro as colunas e depois as linhas;
- Por fim, para cada uma das classes, os pixels eram selecionados alternadamente para compor cada um dos 5 *folds*.

Esse procedimento foi repetido até que cada *fold* contivesse 20% dos pixels de testes, ou seja, por meio dessa organização cada *fold* era totalmente diferente dos demais. Quando necessária e aplicada a utilização da informação espacial, a estratégia sem sobreposição apresentada anteriormente foi aplicada.

Para cada *fold*, após a seleção de bandas, os 80% dos pixels (amostras de treinamento) são novamente utilizados para o treinamento de um modelo classificador. Nesse momento é utilizado o mesmo classificador do processo de seleção de bandas, porém agora treinado com as bandas selecionadas e usando de uma única vez todos os 80% dos pixels. Após essa realização,

os outros 20% do total dos pixels (amostras de testes) são utilizados para os testes do conjunto de bandas selecionados e avaliação do desempenho de classificação para cada um dos métodos e casos considerados.

Essas configurações de distribuições dos pixels para cada um dos *folds* apresentadas neste subcapítulo foram obtidas uma única vez e utilizadas nos diferentes casos de seleção de bandas aplicados nos experimentos iniciais proporcionando experimentos justos. Após a realização dos testes em cada conjunto de bandas selecionado, para cada caso, a média dos diferentes índices obtidos pelo valor dos 5 grupos foi apresentado nos resultados.

4.3.3 Índices avaliados e exibição de resultados

Assim como no estudos de Xie et al. (2018) e Saqui et al. (2019c), para as avaliações de todos os casos testados foram considerados os seguintes índices (métricas):

- Precisão ou Acurácia Geral (*Overall Accuracy - OA*): refere-se à razão entre a quantidade de pixels corretamente classificados e total de pixels. Fornece a proporção de acerto dos pixels classificados independentemente da classe. Esse é o índice mais tradicional em trabalhos da literatura para avaliar a qualidade da classificação dos pixels de IHS com bandas selecionadas;
- Acurácia Média (*Average Accuracy - AA*): obtido a partir do valor médio da OA da classificação de cada uma das classes. Esse índice, também é comumente encontrado na literatura. As equações desse índice e do anterior são mostradas no capítulo posterior, pois foram utilizados como base para guiar o processo de seleção de bandas;
- Coeficiente Kappa de Cohen (*Cohen's Kappa - KC*): uma medida estatística de consistência que tem sido utilizada na literatura para realizar comparações entre o GT e o resultado proporcionado por um classificador. Esse índice é mostrado na equação 4.1:

$$k = \frac{p_0 - p_e}{1 - p_e}, \quad (4.1)$$

onde p_0 é a precisão observada que caracteriza a probabilidade de concordância no rótulo atribuído às amostras, p_e é a concordância esperada caracterizada pela proporção esperada de uma quantidade de vezes que o classificador e GT concordam, essa métrica é estimada por uma equação. Os valores apresentados na sequência indicam os níveis de concordância indicados Landis e Koch (1977) e comumente utilizados na literatura como referência para comparar dois observadores.

- Menor que zero → insignificante.
- Entre 0 e 0,2 → fraca.
- Entre 0,21 e 0,4 → razoável.
- Entre 0,41 e 0,6 → moderada.
- Entre 0,61 e 0,8 → forte.
- Entre 0,81 e 1 → quase perfeita.

Dessa forma, resultados promovidos por um classificador e o GT superiores a 0,8 são considerados quase perfeitos.

- Quantidade de bandas.

Os resultados proporcionados por esses índices são apresentados nos capítulos seguintes em tabelas e o valor médio e desvios padrões obtidos por meio do *5-fold cross-validation* é apresentado. Também são apresentados gráficos para exibição dos resultados obtidos pelos processos de seleção de bandas iterativos, demonstrando a evolução dos resultados. Além disso, sempre que possível e necessário são apresentados alguns valores de testes estatísticos.

Especificamente, quando foram realizadas validações por meio de testes estatísticos, o chamado teste T-pareado foi aplicado. Esse teste pode ser aplicado para comparar os resultados proporcionados por dois modelos classificadores na validação cruzada. No sentido prático, o teste T-pareado avalia se a diferença da média de resultados de dois classificadores é zero (hipótese nula). A hipótese nula ocorre quando o *p-value* é maior que um dado limiar, por exemplo, 0,05. Do contrário da hipótese nula, ou seja, a hipótese alternativa, indica que os dois resultados proporcionados pelos modelos classificadores são diferentes, sugerindo que um classificador pode ter proporcionado resultados melhores que o outro.

4.4 Considerações Finais do Capítulo

Neste capítulo, foram apresentadas informações relacionadas a metodologia aplicada nesta pesquisa. Foram apresentadas as amostras utilizadas para validar o WMoEBS e métodos comparados. Por fim, foram apresentadas a estrutura e configurações dos experimentos realizados e critérios de avaliação. No próximo capítulo são apresentados as etapas específicas e os experimentos que levaram a constituição do método WMoEBS.

Capítulo 5

WMOEBS - *Wrapper Multiobjective Evolutionary Band Selection*

Para criação do método proposto, chamado de *Wrapper Multiobjective Evolutionary Band Selection* (WMOEBS), importantes etapas desta pesquisa foram a estruturação de um *framework* geral para seleção de bandas e a escolha dos componentes que proporcionaram a composição do método proposto. Esses componentes são novos, ou foram modificados e necessitaram serem ajustados e avaliados a fim de obter conclusões que conduzissem à escolha dos mais qualificados para compor o WMOEBS. Portanto, inicialmente são apresentados o *framework*, os componentes e, posteriormente, os experimentos e resultados para escolha desses componentes. Além do objetivo de escolher os componentes adequados, os experimentos também visaram demonstrar possíveis estratégias que podem ser adequadas como opções para métodos futuros. O WMOEBS, com a composição dos componentes escolhidos, é apresentado no final deste capítulo.

As etapas que levaram à constituição do WMOEBS foram também utilizadas na elaboração do *Incorporated Decision-Maker-Based Multiobjective Band Selection* (IDMMoBS), que foi publicado em Saqui et al. (2019c) anteriormente à conclusão desta tese, sendo o WMOEBS uma evolução do IDMMoBS. Os dois métodos se diferenciam principalmente devido à chamada heurística de reparo discutidas nas etapas seguintes. A pesquisa conduzida nesta tese também se diferencia do trabalho publicado em Saqui et al. (2019c) porque contém experimentos mais específicos com testes estatísticos.

5.1 **Framework para seleção de bandas multiobjetivo baseado na estratégia Wrapper**

Na Figura 5.1 é representado um novo *framework* para seleção de bandas onde não são estabelecidos métodos específicos em algumas das suas etapas, permitindo assim que sejam aplicados diferentes algoritmos para composição de métodos. A diferença em relação às *frameworks* baseados na estratégia *Wrapper* é que o atual foi pensado especificamente para seleção de bandas multiobjetivo. A ideia de constituir cada etapa do *framework* com diferentes algoritmos é inspirada em outros trabalhos com métodos híbridos como Li et al. (2011), Sun et al. (2017), Saqui et al. (2019b) e Saqui et al. (2019c), onde sua estrutura foi pensada a partir de partes em comuns de muitos desses trabalhos. Esse *framework* foi utilizado como base estrutural para elaboração do WMoEBS.

No *framework*, inicialmente são definidas as amostras, ou seja, pixels que serão utilizados para o processo de seleção de bandas e as classes desses pixels por meio do *Ground Truth* (GT), contendo amostras de classes como exemplos para treinamento. A presença e posterior utilização do GT, caracteriza o *framework* como próprio para métodos de seleção de bandas supervisionados, porém ainda não é suficiente para caracterizá-lo como *Wrapper*. A justificativa para utilização do GT é a possibilidade da utilização de funções baseadas em classes reais, ou seja, supervisionada, e assim conduzir a melhoria de índices de classificação. Outras definições nessa fase são os parâmetros dos métodos que devem ser utilizados nas etapas seguintes como, por exemplo, no caso da utilização de algoritmos evolucionários (AE), o número de gerações (iterações de busca) e tamanho da população (do conjunto de soluções).

A **Etapa 1** consiste na geração do conjunto inicial de soluções que são avaliadas e utilizadas como base para constituição de novas soluções das gerações seguintes e conseqüentemente explorar o espaço de busca. Alguns dos métodos apresentados anteriormente, como em Li et al. (2011) e Sui et al. (2015), utilizam uma pré-seleção de bandas antes da aplicação do método principal, gerando um conjunto intermediário de bandas que apresentam características específicas em relação a formas tradicionais que geram conjuntos aleatoriamente. A justificativa para utilização desses métodos de pré-seleção, é que muitas vezes eles eliminam a alta correlação de bandas e, conseqüentemente, devido a essa redução da quantidade de bandas, evitam uma necessidade elevada por recursos computacionais durante o processo de busca. Portanto, a **Etapa 1** do *framework* proposto é baseada nessa perspectiva, onde pode ser aplicado um método específico para criação do conjunto de soluções (população) inicial.

As bandas utilizadas no *framework* proposto são representadas por um vetor binário, mesma

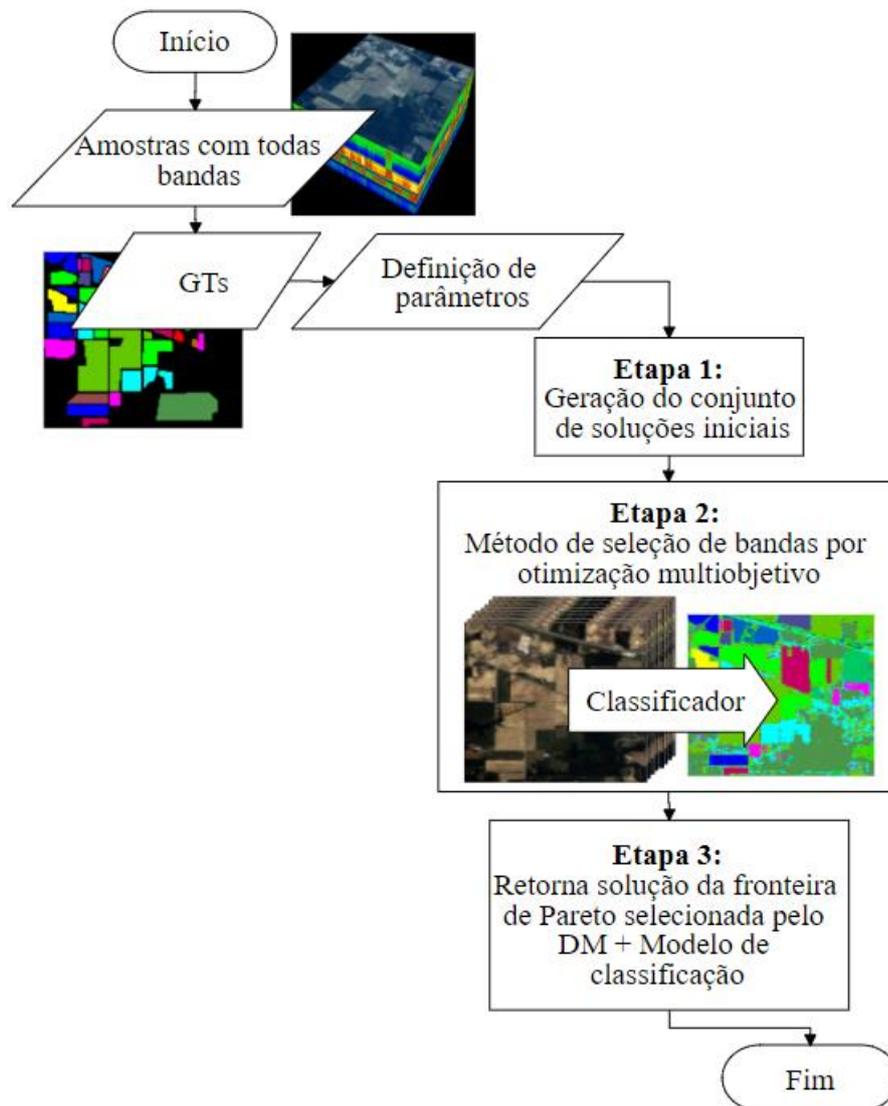


Figura 5.1. Framework para seleção de bandas multiobjetivo baseado na estratégia Wrapper.

estrutura utilizada em Xu, Shi e Pan (2017) e outros trabalhos da literatura. Um exemplo da representação de uma solução é mostrado na Figura 5.2. O tamanho do vetor representa o número total original de bandas, o valor 0 (zero) representa que a banda referente àquela posição não foi selecionada e o valor 1 (um) representa a presença da banda daquela posição.

1	0	1	1	0	1	0	1
---	---	---	---	---	---	---	---

Figura 5.2. Representação de uma solução por meio de um vetor binário.

Adequações podem ser realizadas no *framework* proposto para outros tipos de representação, porém a representação binária é adequada e de fácil manipulação para bandas espectrais.

Na **Etapa 2**, é aplicado o método de seleção de bandas supervisionado baseado na otimização multiobjetivo e na estratégia *Wrapper*, portanto deve-se definir algoritmos dessas categorias. A justificativa da utilização de um método de otimização multiobjetivo é devido ao *framework* ter sido pensado para lidar com a seleção de bandas onde existem soluções com objetivos conflitantes. Por exemplo, os objetivos explorados nesta tese estão relacionados à redução da quantidade de bandas e melhoria da qualidade da classificação, podendo estes ser conflitantes, visto que a melhoria de um pode afetar negativamente o desempenho do outro. Métodos multiobjetivo são adequados para esse tipo de problema. Além da definição do algoritmo de otimização, na **Etapa 2**, também é necessário definir o algoritmo classificador (ex.: *Gaussian Maximum Likelihood Classifier* (GMLC), *Support Vector Machine* (SVM), etc), o que permite caracterizar o *framework* como estratégia *Wrapper*. A justificativa de se utilizar a estratégia *Wrapper* é que ela pode apresentar melhores resultados em relação a métodos baseados em outros tipos de estratégias como, por exemplo, *Filter*. Esse comportamento acontece principalmente porque a estratégia *Wrapper* é caracterizada por incorporar o modelo de classificação ao processo de seleção. As funções de *fitness* que serão otimizadas (ex.: acurácia, número de bandas, etc) e a forma de avaliação aplicada em cada solução (ex.: *holdout*, validação cruzada, etc) também devem ser definidas para serem utilizadas nessa etapa. Para que o método com base nesse *framework* seja funcional, a forma de avaliação em cada solução é necessária ser pelo menos *holdout*, mas pode ser estruturada em validação cruzada com muitos *folds*, separando os pixels em amostras de treinamento e testes, tentando evitar *overfitting* durante a execução do processo.

Todos algoritmos de otimização multiobjetivo encontrados nesta pesquisa são baseados na fronteira de pareto (FP), que pode conter mais de uma solução ótima no senso de Pareto. Portanto, uma proposta ideal é pensar e propor uma estratégia para selecionar uma única solução a partir da FP. Para essa seleção, deve ser definido um tomador de decisão (*Decision Maker-DM*). A justificativa da utilização do DM é que, na **Etapa 3**, é possível utilizá-lo para selecionar uma solução da FP e retorná-la. Junto com a solução selecionada pelo DM, também é retornado o modelo classificador treinado. Portanto, o modelo classificador treinado e as bandas selecionadas pelo DM são as saídas do *framework* de seleção de bandas proposto.

Na sequência, são discutidos os componentes usados para estruturar o WMoEBS, sendo estes caracterizados por algoritmos, métodos e etapas que permitem conduzir a seleção de bandas e classificação de pixels das Imagens Hiperspectrais (IHs) com bandas selecionadas.

5.2 Componentes analisados para estruturação do WMoEBS

Os seguintes componentes foram considerados como possíveis para compor o WMoEBS:

- Classificador e Funções de *Fitness*;
- Método de otimização multiobjetivo;
- Inicialização do conjunto de soluções;
- Heurística de Reparo;
- Utilização da informação espacial; e
- *Decision Maker* (DM).

5.2.1 Classificação de pixels e Funções de *Fitness*

5.2.1.1 Funções de *Fitness*

Os primeiros componentes estruturados são duas funções de *fitness*, $f_1(x)$ e $f_2(x)$ que foram utilizadas no método de seleção de bandas com o propósito de maximização conforme a eq. 5.1:

$$\max f(x) = |f_1(x), f_2(x)|. \quad (5.1)$$

Inspirado em Fauvel et al. (2015), a função $f_1(x)$ é baseada na classificação dos pixels. Diferente de muitos outros métodos que utilizaram índices de classificação, para tentar evitar a ocorrência de *overfitting*, nesta tese $f_1(x)$ é estruturada a partir de validação cruzada. A outra função, $f_2(x)$, é estruturada conforme a quantidade de bandas e baseada em trabalhos com otimização multiobjetivo como em Gong, Zhang e Yuan (2016) e Xu, Shi e Pan (2017).

Duas equações foram consideradas e avaliadas em momentos distintos para serem utilizadas como possíveis funções de *fitness* para f_1 :

1. A primeira equação 5.2 caracteriza o resultado médio de uma validação cruzada com *k-folds*, onde a acurácia média (*Average Accuracy* - AA) é aplicada:

$$f_1(x) = \frac{1}{k} \sum_{j=1}^k \left(\frac{1}{|Cl|} \sum_{i=1}^{Cl} (A_{ij}) \right), \quad (5.2)$$

onde i é cada classe do conjunto de classes Cl e j é cada um dos k *folds*. Essa função

é obtida a partir do valor médio da acurácia de cada classe de cada j -ésimo *fold*. A_{ij} mostrada na equação 5.3 é a acurácia (BRODERSEN et al., 2010; GUYON et al., 2015):

$$A_{ij} = \frac{tp_{ij}}{tp_{ij} + fn_{ij}}, \quad (5.3)$$

onde tp_{ij} é o total de pixels corretamente classificados para uma determinada classe i e fn_{ij} é o total de pixels erroneamente classificados nessa mesma classe i .

2. A segunda equação 5.4, caracteriza o resultado médio de uma validação cruzada com k -*folds* onde o índice de precisão (*Overall Accuracy-OA*) é utilizado:

$$f_1(x) = \frac{1}{k} \sum_{j=1}^k OA_j = \frac{1}{k} \sum_{j=1}^k \frac{1}{n} \sum_{i=1}^n (1 \times (p_{ij} = \hat{p}_{ij})), \quad (5.4)$$

onde p_{ij} é a classe do pixel real obtida pelo GT, e \hat{p}_{ij} é a classe estimada para o pixel i .

A OA é o índice mais comum de ser avaliado em resultados de classificação de pixels de IHS com bandas selecionadas, porém ela pode não ser adequada para problemas com classes não balanceadas, onde existem classes com poucas amostras. A AA, mostrada na equação 5.2, é indicada para problemas multi-classes, especialmente onde os conjuntos de dados das classes têm diferentes quantidades de amostras (classes não balanceadas). Como a OA é o índice mais frequentemente avaliado e comparado em estudos com classificação de IHS enquanto o índice AA é mais adequado para problemas com classes desbalanceadas, os dois índices foram avaliados e analisados como possíveis componentes f_1 do WMoEBS.

A outra função utilizada no problema de seleção de bandas multiobjetivo, f_2 , é representada na equação 5.5 e está relacionada a redução da quantidade de bandas:

$$f_2(x) = 1 - \frac{|t(x) - Q|}{len(x)}, \quad (5.5)$$

onde $len(x)$ é a quantidade de bandas totais possíveis nas IHS, ou seja, a quantidade original. $t(x)$ é o total de bandas presentes na solução x conforme mostrado na equação 5.6:

$$t(x) = \sum_{z=0}^{len(x)} x_z, \quad (5.6)$$

onde z representa uma banda na solução x . Q é um parâmetro que caracteriza a quantidade de bandas desejada. Q é importante por evitar que o algoritmo tente reduzir drasticamente a

quantidade de bandas, o que pode causar grande perda de desempenho de classificação e pode ser definido pelo usuário.

Por meio das funções f_1 e f_2 , os métodos de seleção devem buscar por soluções balanceadas, onde a redução da quantidade de bandas selecionadas e o índice escolhido em f_1 da classificação consigam ser melhorados conjuntamente.

Uma observação importante se diz ao uso da validação cruzada nos dois índices utilizados como possível f_1 , onde os pixels são separados para treinamento e testes. Além disso, os pixels de treinamento são subdivididos em um subgrupo de treinamento e validação para guiar o processo de busca. Isso porque eles guiam a estruturação do modelo classificador ao mesmo tempo em que as bandas são selecionadas e, as bandas selecionadas ao fim do processo são avaliadas por meio dos pixels de testes.

5.2.1.2 Organização das amostras no processo de seleção de bandas

Para cada solução avaliada durante o processo de seleção nesta tese, ou seja, no classificador incorporado ao método de seleção de bandas, os 80% dos pixels de treinamento que foram obtidos do processo de *5-fold cross-validation* usado nos experimentos, eram configurados em *3-fold cross-validation* para diminuir riscos de *overfitting*. Dessa forma eram constituídos dois *folds* para treinamento e um *fold* para validação para cada solução avaliada na seleção.

5.2.1.3 Classificação de pixels

Dois classificadores apresentados previamente foram utilizados, o SVM e o GMLC. O SVM foi escolhido porque mostrou bons resultados em diversos experimentos com classificação de IHS, além disso, ele opera bem quando a quantidade de amostras de treinamento é limitada. O GMLC foi escolhido porque, além de apresentar bons resultados em trabalhos da literatura, existe a suposição feita por Richards e Jia (2013) de que distribuições gaussianas para classes em imagens de Sensoriamento Remoto (SR) são aceitáveis, pois espera-se que a maioria dos pixels de uma classe, ou seja, a maior densidade dos pixels de uma classe em uma imagem, se concentre em direção ao centro e conforme mais afastado desse centro, menor a chance de os pixels serem dessa classe.

As particularidades utilizadas em cada classificador são descritas na sequência:

- SVM: Junto com esse algoritmo foi utilizado o *kernel radial basis function* (RBF), apontado por Ghamisi et al. (2017) como adequado para lidar com distribuições de classes não

lineares complexas. Os parâmetros C (controle do relaxamento da margem de separação) e γ (controle da curva de separação do RBF), foram estabelecidos em um processo de validação cruzada com *5-folds* considerando os valores mais frequentes para diferentes combinações de bandas.

Para este método também foi utilizada a estratégia um contra todos (*one-against-all*) que, dado $|Cl|$ classes, o SVM treina um modelo classificador para cada classe i sendo que $i \in Cl$. Cada um desses modelos classificadores é binário e classifica uma nova instância entre uma determinada classe i ou em uma classe genérica que agrupa todas outras $|Cl| - 1$ classes. A classe final da nova instância é a mais frequente por votação entre todos modelos construídos;

- GMLC: Esse algoritmo pode ter problemas para estabelecer a matriz de covariância em suas equações para casos com poucas amostras em relação a elevados número de características. Para tratar esse problema uma estratégia de estimativa de covariância, chamada em inglês de *pooled covariance*, foi aplicada quando não era possível o cálculo da alternativa tradicional.

5.2.2 Método de otimização multiobjetivo

Diferentes algoritmos de otimização multiobjetivo podem ser utilizados para seleção de bandas, tais como:

- *Non-dominated sorting genetic algorithm II* (NSGA2) (KUMAR, 2004);
- *Strength Pareto Evolutionary Algorithm* (SPEA2) (SAQUI et al., 2019b);
- *Multiobjective evolutionary algorithm based on decomposition* (MOEA/D) (GONG; ZHANG; YUAN, 2016).

Para elaboração do método proposto, foram utilizados o MOEA/D e a *Tchebycheff Decomposition Strategy* (TDS). A motivação para a escolha do uso do MOEA/D + TDS foi que, além de ser um dos métodos multiobjetivos mais recentemente explorados em seleção de bandas, em comparações com outros métodos da literatura, como o NSGA2, ele tem apresentado bons resultados no processo de otimização multiobjetivo. O algoritmo MOEA/D tem mecanismos que se adequam ao problema de seleção de bandas multiobjetivo, como por exemplo, a chamada "estratégia de reparo" que permite delimitar o espaço de busca e a utilização de vizinhos de cada solução no processo de busca, o que reduz o custo computacional.

O TDS é usado no WMoEBS para agregar os m objetivos (f_1 e f_2 neste caso) em uma única função de agregação. Valores dessa função de agregação são obtidos por meio da diferença dos valores de *fitness* de f_1 e f_2 de uma determinada solução x e um ponto ideal $z^* = [z_1^*, z_2^*]^T$. A função com pior valor entre as diferenças dessas funções de *fitness* e os valores do ponto ideal é usada para representar o valor da agregação que, por sua vez, é utilizado para comparar soluções. Isso possibilita ao MOEA/D tentar melhorar sempre o pior valor das diferentes funções *fitness*, causando uma tendência de buscar por aquelas que estão mais próximas da FP.

Uma representação do MOEA/D adequada para estabelecer o WMoEBS é mostrada no Algoritmo 1, onde inicialmente são definidos os parâmetros do algoritmo, os pixels *OrigHI* de amostra da IH com todas as bandas e o *GT* dos pixels de amostra que é utilizado no processo de treinamento. Os parâmetros utilizados são: *NGEN*, que representa o número de iterações (gerações) que o algoritmo será executado; *N*, que representa a quantidade de soluções (indivíduos) na população; *T*, que representa os vizinhos da população e, *mr*, a taxa de mutação.

No **Passo 1**, é inicializado o conjunto de soluções, ou seja, a população P . Para essa inicialização, métodos específicos podem ser aplicados e uma proposta é descrita posteriormente. Internamente a esse passo, uma população externa adicional (**Passo 1.1**) é usada. Cada solução é estruturada conforme a representação binária previamente apresentada.

No **Passo 1.2**, são calculados os valores das funções de *fitness* f_1 e f_2 para cada solução x do conjunto inicial P . As funções de *fitness* utilizadas foram baseadas na redução da quantidade de bandas e melhora do valor da média obtida a partir da validação cruzada de índices de acurácia média ou precisão da classificação.

No **Passo 1.3**, são estabelecidos os valores de ponderação $\lambda^1 \dots \lambda^N$, sendo $\lambda^i = [\lambda_1^i, \dots, \lambda_m^i]$ para m -funções de *fitness* e, por definição, a soma dos pesos de cada objetivo de cada λ^i deve ter valor 1. Também são estabelecidas a distância Euclidiana entre todos os vetores de ponderação no espaço de objetivos e a vizinhança $NG(i) = i^1, \dots, i^{NS}$, onde $\lambda^{i^1}, \dots, \lambda^{i^{NS}}$ são os NS vetores de ponderação mais próximos a λ^i . Posteriormente, no **Passo 1.4**, é iniciado o valor ideal de z^* . Uma abordagem comum e utilizada nesta tese, é utilizar os melhores valores de cada *fitness* da população corrente para z^* .

O **Passo 2**, é caracterizado por um número de iterações onde as atualizações são executadas e também a análise de todos os N subproblemas. No **Passo 2.1** ocorre a seleção das soluções para reprodução, onde duas soluções para i^a e i^b , que fazem parte da vizinhança $NG(i)$ da solução corrente para i são selecionadas aleatoriamente e, então, é executada a aplicação dos operadores genéticos de cruzamento e mutação para geração de novas soluções NI .

Algoritmo 1: MOEA/D para seleção de bandas

Entrada: $NGEN$ (número de gerações/ critério de parada),
 N (tamanho de população/número de subproblemas),
 NS (tamanho da vizinhança),
 GT , $OrigHI$ (IH com todas as bandas),
e mr (taxa de mutação).

Saída : Solução selecionada pelo DM da população externa (EP).

Passo 1 - Inicialização:

Passo 1.1 - Cria uma população P gerando soluções com combinações de bandas a partir de $OrigHI$. Cria uma população externa $EP = \emptyset$

Passo 1.2 - Calcula funções de *fitness* f_1 e f_2

Passo 1.3 - Define o conjunto de vetores $\lambda^1 \dots \lambda^N$

Computa a distância Euclidiana entre os vetores de pesos e então opera nos NS vetores vizinhos mais próximas de cada vetor de pesos.

Para cada $i = 1 \dots N$, define um conjunto de vizinhos $NG(i) = i^1, \dots, i^{NS}$, onde $\lambda^{i^1}, \dots, \lambda^{i^{NS}}$ são os NS vetores mais próximos de λ^i .

Passo 1.4 - Inicialização z^* .

Passo 2 - Atualização:

Para $ngen = 1 \dots NGEN$ faça (critério de parada),

Para $i = 1 \dots N$, faça,

Passo 2.1 - Reprodução:

Seleciona aleatoriamente duas soluções, i^a e i^b , da vizinhança de i em $NG(i)$, e então gera novas soluções NI a partir de i^a e i^b pelo uso de operadores genéticos.

Passo 2.2 - Reparo:

Repara as soluções NI geradas usando uma heurística particular.

* Estabelece os valores de *fitness* de NI .

Passo 2.3 - Atualiza z^* :

Para cada $y' \in NI$,

Para cada $j = 1, \dots, m(\text{objetivos})$,

se $f_j(y') > z_j^*$, então

define $z_j^* = f_j(y')$

Passo 2.4 - Atualiza as soluções vizinhas:

Para cada índice $k \in NG(i)$,

se $g(y' | \lambda^i, z^*) \leq g(x^k | \lambda^i, z^*)$ então

define $x^k = y'$, $f_1(x^k) = f_1(y')$, e $f_2(x^k) = f_2(y')$.

Passo 2.5 - Atualiza EP :

Remove todas soluções dominadas por y' a partir EP .

Adiciona y' para EP se não há soluções em EP que dominam em y' .

Passo 3 - Fim:

DM seleciona e exibe a solução de EP .

No **Passo 2.2** ocorre a chamada operação de reparo, que pode ajustar uma determinada solução conforme uma heurística determinada para respeitar restrições do espaço de busca. Essa heurística é uma das principais vantagens do MOEA/D, permitindo controlar a busca de soluções em um espaço factível e não foi explorada em algoritmos de seleção de bandas como pode-se observar em Gong, Zhang e Yuan (2016) e Xu, Shi e Pan (2017). Portanto, é uma das principais inovações do método proposto no contexto de seleção de bandas. Após essa etapa são calculados os valores de *fitness* das novas soluções.

No **Passo 2.3**, com as novas soluções $y' \in NI$, o ponto de referência z^* é atualizado e no **Passo 2.4**, é atualizada a vizinhança de i corrente com base nas novas soluções e TDS.

No **Passo 2.5**, são removidas de EP todas soluções dominadas por y' e adiciona y' para EP se não há soluções em EP que dominem y' .

No **Passo 3**, um DM seleciona uma solução da FP e, então, as bandas selecionadas e o modelo do classificador são retornadas.

5.2.3 Inicialização do conjunto de soluções (população)

A inicialização do conjunto de soluções/população para AE é importante pois promove uma seleção inicial de bandas nas soluções que serão base da evolução de AE com base em alguns critérios. Conforme o estudo realizado em Li et al. (2011), uma técnica de *clustering* pode ser utilizada para gerar um conjunto de soluções iniciais, reduzindo a quantidade de bandas, ao mesmo tempo que melhora a qualidade da classificação. O índice utilizado pelos autores é baseado na informação mútua que permite agrupar bandas com alta correlação.

Nesta tese, para composição do WMoEBS, um método inspirado no algoritmo *Ward's Linkage strategy Using Mutual Information* (WaLuMi) desenvolvido por Martinez-Uso et al. (2007) foi considerado e uma estratégia de inicialização é apresentada no Algoritmo 2.

No Algoritmo 2, é recebido como entrada o conjunto de pixels de amostras de treinamento representados por *OrigHI*. Esses pixels são representantes de diferentes amostras que também são utilizadas posteriormente para o processo de seleção de bandas e, inicialmente, contém todas as bandas da imagem original. Os valores denominados *quantidadeMin* e *quantidadeMax* são utilizados para estabelecer um intervalo na quantidade de bandas a serem selecionadas inicialmente e N representa o total de soluções a serem geradas. *OrigHI* é uma matriz onde cada linha representa um pixel e cada coluna as bandas existentes.

No **Passo 1**, a transposta de *OrigHI*, denominada *MatT*, é utilizada para se estabelecer uma métrica de distância entre cada par de bandas. Essa métrica é baseada entre a diferença do valor

1 e a informação mútua normalizada entre cada par de bandas i e j , cuja matriz de similaridade obtida a partir desses valores é representada por *MatrizDist*.

Algoritmo 2: Inicialização de soluções baseada no WaLuMi

Entrada: *OrigHI* (Pixels de amostras de treinamento com todas as bandas)

quantidadeMin (quantidade mínima de bandas)

quantidadeMax (quantidade máxima de bandas)

N (tamanho da população número de subproblemas)

Saída : População inicial P .

Passo 1 - Cálculo da métrica de distância baseada na informação mútua (IM):

$MatT = \text{transposta}(OrigHI)$:

Para cada banda i em $MatT$:

Para cada banda j em $MatT$:

$MatrizDist[i, j] = 1 - \text{IM}$ (normalizada) entre i e j

Passo 2 - Formação de grupos:

Aplica-se o método *Ward's linkage clustering* na *MatrizDist* para formar um agrupamento

hierárquico (dendograma) entre as bandas

Passo 3 - Geração de soluções:

Para cada k até N :

Aleatoriamente escolhe o valor p entre a *quantidadeMin* até *quantidadeMax*

Cria p grupos com base na hierarquia construída no Passo 2.

Aleatoriamente escolhe uma banda de cada grupo e cria uma nova solução.

Adiciona a solução criada à população inicial P .

Passo 4 - Finalização:

Retorna o conjunto de soluções inicial P .

No **Passo 2**, o *Ward's linkage clustering* é utilizado para estabelecer uma hierarquia, na forma de um dendograma, utilizando a matriz de similaridade *MatrizDist* criada no **Passo 1**.

No **Passo 3**, baseado na hierarquia (dendograma) estabelecida no **Passo 2**, diferentes grupos de bandas com tamanhos estabelecidos aleatoriamente e dentro da faixa de *quantidadeMin* e *quantidadeMax* são criados. Então, uma banda de cada grupo é selecionada aleatoriamente para formar uma solução. Essa nova solução é inserida na população P . O procedimento de geração é repetido até uma quantidade de soluções N estarem inseridas no conjunto de soluções P e finalmente, no **Passo 4**, esse conjunto de soluções é retornado.

Esse procedimento se diferente do WaLuMi tradicional, pois as bandas são selecionadas aleatoriamente a partir dos grupos formados, sendo outra inovação desta pesquisa. No WaLuMi, uma única combinação de bandas é selecionada por meio de uma equação específica.

5.2.4 Heurística de reparo

A heurística de reparo proposta nesta tese e utilizada em conjunto ao MOEA/D tem o propósito de restringir a quantidade máxima de bandas. Essa heurística é uma das principais inovações desta tese e incluídas no MOEA/D para o contexto de seleção de bandas. Nenhuma heurística de reparo foi explorada nos métodos de seleção multiobjetivo apresentados em Gong, Zhang e Yuan (2016) e Xu, Shi e Pan (2017) que foram estudados nesta pesquisa. A restrição proporcionada por essa heurística não descaracteriza o problema de seleção de bandas como multiobjetivo, visto que o método continua lidando com soluções com objetivos conflitantes, porém, com um limiar para a quantidade máxima de bandas.

A heurística proposta é baseada na razão do ganho de informação, do inglês *Information Gain ratio* (IGR), inspirada no método de seleção de bandas apresentado em Xie et al. (2018). Conforme esses autores, o *Information Gain* (IG) pode mensurar quanto cada característica pode contribuir para a classificação.

O IDMMoBS (SAQUI et al., 2019c) elaborado durante o desenvolvimento desta pesquisa fez uso de uma estratégia de reparo pensada previamente à estratégia usada no WMoEBS. Essa estratégia de reparo foi aplicada somente quando a quantidade de bandas era superior ao limiar p . As características da estratégia são mostradas no Algoritmo 3 e elencadas na sequência.

Nos **Passos 1, 2 e 3** do Algoritmo 3, para uma determinada solução x (com combinação específica de bandas), o método *Ward's linkage clustering* é utilizado para agrupar as bandas conforme a informação mútua. Especificamente no **Passo 1**, ocorre a redefinição dos pixels da *OrigHI* conforme as bandas existentes em x . Além disso, nessas etapas, uma quantidade p (limiar) de grupos são formados.

No **Passo 4**, em cada grupo formado as bandas são ranqueadas conforme a IGR. Posteriormente, no **Passo 5**, as bandas com maior IGR de cada grupo são selecionadas constituindo um novo conjunto. Por fim, no **Passo 6**, a solução x em questão é retornada com a nova combinação de bandas encontrada.

Durante a execução do IDMMoBS foram apresentados alguns problemas relacionados a estratégia de reparo, sendo destacado que:

- embora o cálculo da informação mútua entre bandas pudesse ser realizado uma única vez no início do processo de seleção, o *Ward's linkage clustering* tinha que ser aplicado novamente para cada solução com uma combinação diferente de bandas em cada geração (**Passo 2 e 3**), ocasionando um elevado custo computacional (processamento, memória e tempo) ao processo de seleção; e

- o limiar fixo p causou uma tendência ao processo de busca, fazendo com que o método encontrasse muitas soluções com uma quantidade exata de p bandas, evitando em quase todo o processo soluções com uma quantidade de bandas menor que p .

Algoritmo 3: Estratégia de Reparo do IDMMoBS

Entrada: p (limiar máximo de bandas)

x (solução com combinação específica de bandas)

Saída : x (solução com nova combinação de bandas)

Passo 1 - Cálculo da métrica de distância baseada na informação mútua (IM):

$SelecHI = \text{seleciona}(OrigHI, x)$

$MatT = \text{transposta}(SelecHI)$:

Para cada banda i em $MatT$:

Para cada banda j em $MatT$:

$MatrizDist[i, j] = 1 - IM$ (normalizada) entre i e j

Passo 2 - Formação de grupos:

Aplica o método *Ward's linkage clustering* na $MatrizDist$ somente nas bandas de x para formar um agrupamento hierárquico (dendograma) entre as bandas

Passo 3 - Geração de grupos:

Cria p grupos com base na hierarquia construída no Passo 2.

Passo 4 - Ranqueamento:

As bandas são ordenadas conforme o IGR em cada um dos p grupos.

Passo 5 - Geração do novo conjunto de bandas:

Seleciona a banda melhor ranqueada de cada grupo e constrói um novo conjunto de bandas.

Passo 6 - Finalização:

Redefine x como a nova combinação de bandas e o retorna.

Pensando em resolver os problemas da estratégia de reparo usada no IDMMoBS, no WMoEBS, algumas soluções foram estabelecidas, sendo essas apresentadas no Algoritmo 4.

Como **Entrada** do Algoritmo 4, o $conjIGR$ são os valores do IGR para cada uma das bandas espectrais, que pode ser obtido uma única vez, inicialmente na execução do método de seleção, uma vez que não existe a necessidade de ele ser recalculado ao longo das gerações.

No **Passo 1**, as bandas da solução x são ranqueadas conforme o $conjIGR$. No **Passo 2**, as m bandas da solução x com maior IGR são selecionadas constituindo um novo conjunto. Por fim, no **Passo 3**, a solução x em questão é retornada com a nova combinação de bandas encontrada.

Uma condição importante é que m é reajustado em cada iteração/geração do algoritmo de seleção de bandas e dado aleatoriamente entre:

1. a quantidade de bandas de uma solução da FP selecionada aleatoriamente; ou
2. uma quantidade estabelecida aleatoriamente igual ou inferior ao limiar p (definido pelo usuário no início do método de seleção).

Algoritmo 4: Estratégia de Reparo do WMoEBS

Entrada: x (solução com combinação específica de bandas)

$conjIGR$ (conjunto com valores específicos da razão do ganho de informação)

Saída : x (solução com nova combinação de bandas)

Passo 1 - Ranqueamento:

As bandas presentes em x são ordenadas conforme o IGR em $conjIGR$

Passo 2 - Geração do novo conjunto de bandas:

Seleciona as m bandas melhor ranqueadas e constrói um novo conjunto de bandas.

Passo 3 - Finalização:

Redefine x como a nova combinação de bandas e o retorna.

A estratégia de definir um m automaticamente ao longo das gerações e variável evita que o método fique tendencioso somente em soluções com uma determinada quantidade de bandas. Retirar a aplicação do *Ward's linkage clustering* para cada solução melhorou o desempenho computacional e o IGR de cada banda pode ser calculado já no início do algoritmo, evitando que seja calculado a cada iteração.

Resultados entre comparações entre o WMoEBS e IDMMoBS são apresentados nos experimentos comparativos entre métodos, onde pode-se observar a diferença nos resultados proporcionados por cada um dos métodos.

5.2.5 Utilização da informação espacial obtida por meio de filtros

Uma importante consideração realizada nesta tese para compor o método proposto envolve a utilização ou não da informação da vizinhança de cada pixel durante o processo de seleção de bandas e, também, na classificação de IH com as bandas selecionadas. Essa informação pode ser obtida por meio do uso de filtros espaciais que podem melhorar o desempenho na classificação de IHS, além de serem facilmente implementados em relação a outros métodos para mesma finalidade (CAO et al., 2017). Adicionalmente, os filtros espaciais fazem parte da etapa de pré-processamento removendo ruídos das bandas espectrais.

Para composição do WMoEBS, os seguintes filtros foram experimentalmente avaliados:

1. Filtro de média;

2. Filtro de mediana;
3. Filtro bilateral; e
4. Filtro Gaussiano.

Esses filtros têm o propósito de redução de ruídos e alguns deles também são direcionados para preservar bordas nas imagens. Além disso, eles permitem controlar a área na qual os pixels são considerados no processo de filtragem por meio do *kernel*. O controle dessa área é importante para organização dos experimentos que foram realizados com métodos que utilizam a informação espacial, pois deve-se ter a precaução da influência entre as amostras de treinamento e testes para não se realizar avaliações injustas.

5.2.6 Decision-Maker (DM)

Como algoritmos multiobjetivos estão sujeitos a encontrar um conjunto de soluções de *tradeoff* que não são dominadas por nenhuma outra e constituem a FP, uma estratégia com um tomador de decisão, do inglês *Decision-Maker* (DM), pode ser aplicada para selecionar uma das soluções a partir dessa fronteira. Essa estratégia é adotada no método proposto nesta tese e algumas possíveis constituições desse DM são consideradas.

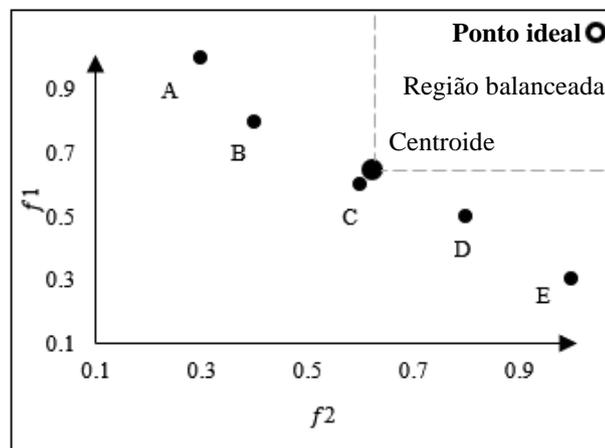


Figura 5.3. Representação da região balanceada (KIMOVSKI et al., 2018)

Algumas das possíveis constituições são baseadas no DM proposto no trabalho de Kimovski et al. (2018), e um exemplo que auxilia em seu entendimento é apresentado na Figura 5.3, onde duas funções que devem ser maximizadas são apresentadas. Na Figura 5.3, o centroide é calculado a partir dos valores das funções f_1 e f_2 das soluções pertencentes a FP. Esse centroide funciona como um limiar para criar uma região de soluções balanceadas e pode ser útil para

escrever regras de tomadas de decisão. A região balanceada pode ser visualizada na área tracejada da Figura 5.3, onde o centroide está representado como o círculo maior no canto inferior esquerdo da região balanceada e o ponto ideal é o círculo sem preenchimento. A-E são soluções do espaço de decisão que constituem a FP. Com o centroide calculado, podem ser estabelecidas regras para tomar a decisão e escolher uma solução de interesse.

Outra constituição de DM considera a seleção direta de uma solução da FP com base em uma das funções de *fitness* e pode ser relacionada a restrições das soluções factíveis do espaço de busca. Dessa forma, é possível selecionar a solução da FP com melhor valor para função de *fitness* relacionada a classificação dentro de uma quantidade máxima de bandas.

Além desses DMs, foi considerado também a utilização do *Sequential Backward Floating Selection* (SBFS) para selecionar uma solução da FP com uma quantidade de bandas ainda mais reduzida. O SBFS é uma evolução do *sequential backward selection* (SBS), porém testa a reinclusão de características removidas em passos anteriores a cada nova remoção de uma característica (PUDIL; NOVOVICOVA; KITTLER, 1994).

Na seleção de bandas, o SBFS iterativamente remove a banda que proporciona melhor valor de uma função de *fitness* do conjunto de bandas após sua remoção. Depois de cada iteração, o SBFS testa individualmente a reinserção de cada uma das bandas removidas nas iterações anteriores no conjunto atual. Se esses testes não promoverem a melhoria da função avaliada, o conjunto atual é mantido, do contrário, a banda reinserida que promoveu o melhor valor de *fitness* é mantida. Posteriormente a essas etapas, o SBFS passa para próxima iteração e avalia a próxima banda para remoção. Nos testes do DM desta tese, a quantidade de bandas que promoveu o melhor valor de *fitness* usando o SBFS é usada.

Com essas características, nesta tese foram avaliados DMs com as seguintes regras:

1. DM baseado no centroide:

- Se há soluções na região balanceada, então a solução daquela região mais próxima ao ponto ideal é selecionada;
- Se não há soluções na região balanceada, então a melhor solução em f_2 (baseada na quantidade de bandas) da FP com valor superior a um dado limiar definido pelo centroide para função f_1 (que mensura a qualidade da classificação) é selecionada;
- De outra maneira, entre todas as possíveis soluções na FP, aquela mais próxima do ponto ideal é escolhida.

Através dessa estratégia, é possível selecionar somente uma solução da FP privilegiando uma das funções na decisão final. Isso pode ser importante para IHS porque embora

objetiva-se melhorar o número de bandas e qualidade da classificação simultaneamente, em casos conflitantes deve-se privilegiar primeiro essa última característica;

2. DM baseado no centroide + SBFS:

- baseado na regra anterior, porém após a aplicação da regra, no final da seleção de bandas, é aplicado o SBFS para tentar reduzir um pouco mais a quantidade de bandas sem perder a qualidade da classificação.

3. DM baseado na seleção direta de uma solução da FP:

- Dado a FP, a solução de maior valor de *fitness* em f_1 , ou seja, para avaliar a classificação dos pixels é selecionada, visto que as restrições estabelecidas pela heurística de reparo no espaço de busca podem limitar a quantidade de bandas.

4. DM baseado na seleção direta de uma solução da FP + SBFS:

- baseado na regra anterior, ou seja, na seleção direta de uma solução da FP, porém após a aplicação da regra é aplicado o SBFS para tentar reduzir um pouco mais o número de bandas sem perder a qualidade da classificação.

5.3 Experimentos para as escolhas dos componentes do WMoEBS

Os experimentos realizados foram organizados em grupos que consideraram avaliações de componentes individualmente ou em conjuntos. Os componentes avaliados são:

- Possível algoritmo de classificação utilizado no WMoEBS, onde foram considerados o algoritmo paramétrico *Gaussian Maximum Likelihood Classifier* (GMLC) e o não paramétrico *Support Vector Machines* com *kernel radial basis function* (SVM-RBF);
- Possível função de *fitness* f_1 utilizada para avaliar a classificação durante a seleção de bandas, podendo ser baseada na precisão (Acurácia Geral, ou do inglês *Overall Accuracy-OA*) ou acurácia média (*Average Accuracy-AA*). Lembrando-se que a função de *fitness* denotado como f_2 é baseada na quantidade de bandas;
- Possibilidade da utilização ou não das informações espaciais considerando os pixels vizinhos de cada amostra das imagens hiperspectrais (IHs) tornando o WMoEBS um método espectral-espacial, sendo que essas informações foram estabelecidas por diferentes filtros da área de processamento de imagens; e

- Possível tomador de decisão, ou seja, *Decision Maker-* (DM), e o impacto de sua utilização durante a seleção de bandas.

Após a seleção dos componentes que foram candidatos para compor o WMoEBS, foram realizadas escolhas e experimentos para avaliação de cada um. O propósito foi encontrar os mais adequados para operar com a seleção de bandas multiobjetivo baseada na estratégia *Wrapper*.

Um componente que não foi avaliado, porém escolhido para ser utilizado, foi a estratégia de inicialização das soluções apresentada anteriormente e baseada no *Ward's Linkage strategy using Mutual Information* (WaLuMi). Esse componente é aplicado antes do procedimento de seleção de bandas para iniciar um conjunto de soluções de base para busca (e evolução) de possíveis novas soluções.

Outro componente escolhido, foi a própria estrutura do MOEA/D utilizada para guiar o processo de seleção de bandas por meio de uma estratégia multiobjetivo. A opção por utilizar o MOEA/D é porque ele é um algoritmo explorado recentemente para seleção de bandas, porém, por meio da estratégia *Filter*, que apresentou bons resultados e vantagens em relação a outros métodos tradicionais. Esse algoritmo possui elementos que podem melhorar o processo de seleção de bandas, como a possibilidade do uso da "estratégia de reparo". Além desse componente e do método baseado no WaLuMi, o *framework* geral que foi proposto, foi utilizado como base para estruturar o WMoEBS.

Nesta seção são apresentados os experimentos e discussões realizados com os componentes que foram avaliados e resultados obtidos. Esses experimentos foram organizados em grupos que consideraram as avaliações individuais ou em conjunto com combinações de componentes. Os grupos considerados são:

- Grupo de experimentos 1: Avaliação dos classificadores (SVM ou GMLC), uso ou não da informação espacial por meio de filtros e índice utilizado como base para a função de *fitness* f_1 (acurácia média ou precisão);
- Grupo de experimentos 2: Avaliação de diferentes possíveis filtros espaciais, aplicados com *kernel* de tamanho 3×3 ;
- Grupo de experimentos 3: Avaliação dos diferentes e possíveis tomadores de decisão (*Decision Maker-DM*).

Conforme cada um dos grupos de experimentos foi realizado, sempre que possível, os componentes já eram selecionados para a composição do WMoEBS, eliminando-se as outras possíveis alternativas a serem avaliadas nos grupos seguintes. Isso possibilitou reduzir bastante a quantidade de experimentos necessários posteriormente.

Para a avaliação dos componentes, os três conjuntos de dados/imagens apresentados anteriormente foram utilizados para os experimentos iniciais. Nesses experimentos foi considerada a utilização do *kernel* sem sobreposição com tamanho 3×3 , logo as seguintes quantidades de pixels/amostras estavam passíveis para utilização:

- Indian Pines: 1.134 amostras;
- Salinas: 6.037 amostras; e
- Pavia University: 4.681 amostras.

Essas três imagens têm uma quantidade considerável de pixels e classes para realização dos experimentos.

Ainda, para a realização dos experimentos, o processo de seleção de bandas foi conduzido pelo mesmo MOEA/D, que é posteriormente utilizado no WMoEBS. Os parâmetros utilizados na execução desse algoritmo em todos os experimentos realizados e apresentados nesse capítulo são mostrados na Tabela 5.1.

Tabela 5.1. Configurações do MOEA/D utilizado nos testes dos componentes avaliados.

Configurações do MOEA/D utilizado nos testes dos componentes	
Parâmetro/Algoritmo	Valor
Iterações (gerações)	500
Número de Soluções	100
Cruzamento	Cruzamento de Dois Pontos
Mutação	<i>Troca de bits (bit-flip)</i>
Taxa de Mutação	5%
Quantidade de vizinhos	20
f_1	OA ou AA (<i>3-fold cross-validation</i>)
Classificador para f_1	SVM ou GMLC
Limiar para f_2	30
Estratégia de reparo	Limiar fixo
Limiar de reparo	30

Especificamente, para avaliação dos possíveis componentes do WMoEBS, a estratégia de reparo do IDMMoBS que limita a quantidade máxima de bandas foi aplicada. Com isso um importante parâmetro foi o limiar máximo da estratégia de reparo, onde foi utilizada a quantidade de 30 bandas. O mesmo valor também foi utilizado como quantidade desejada na função de *fitness* f_2 . Esse valor foi observado em Xia et al. (2018), onde é tomado como um limiar superior para quantidade de bandas desejadas.

Outras importantes definições são sobre o classificador SVM incorporado ao processo de seleção de bandas, no qual foi utilizado com *kernel Radial Basis Function* (RBF). Os parâmetros

do SVM γ e C foram obtidos previamente ao processo de seleção de bandas, através de uma estratégia de busca em grade por meio de validação cruzada com multiplicidade 5-(5 *folds*). Para essa busca, para cada uma das IHS foram geradas 100 soluções de forma aleatória com 30 bandas e esse procedimento permitiu encontrar os parâmetros mais frequentes que otimizavam o SVM e que, posteriormente, foram utilizados durante o processo de seleção.

5.3.1 Grupo de experimentos 1: Classificadores, uso da informação espacial e função de *fitness* f_1

Os experimentos do primeiro grupo foram baseados nos seguintes componentes:

- O classificador incorporado no processo de seleção de bandas, ou seja, o GMLC (paramétrico) ou o SVM (não-paramétrico). Essa avaliação teve o objetivo de auxiliar na escolha do classificador que é incorporado no WMoEBS;
- Utilização ou não da informação espacial obtida por meio de estratégias de filtragem. Essa avaliação visou auxiliar na análise da possibilidade do uso de informação espacial integrada à informação espectral no processo de seleção de bandas. Para isso foram mostrados os resultados e discutidas as possíveis vantagens do uso dessa informação; e
- Função de *fitness* f_1 , ou seja, baseada em OA ou AA. Em cada um dos casos analisados nesse grupo de experimentos, um desses dois índices foi sempre utilizado em conjunto com f_2 apresentada na equação 5.5.

Para selecionar uma única solução da fronteira de Pareto, o DM que captura a solução com melhor valor de f_1 foi utilizado. Para permitir o uso da informação espacial, o filtro de média foi inicialmente escolhido, por ser suficiente para avaliar o impacto da informação espacial na seleção de bandas e classificação, conforme já demonstrado em outros trabalhos da literatura, como em Liang et al. (2017). Posteriormente, no segundo grupo de experimentos a utilização da informação espacial proporcionada por diferentes filtros foi avaliada de forma integrada ao processo de seleção de bandas na busca da opção mais adequada para compor o WMoEBS. Dessa forma determinou-se as seguintes configurações para os experimentos:

- MOEA/D para o processo de seleção de bandas com parâmetros mostrados na Tabela 5.1;
- 5-fold *cross-validation* para testar as bandas selecionadas em operações de classificação;
- 3-fold *cross-validation* para composição da função de *fitness* baseada na OA ou AA durante os processos de seleção de bandas;

- Filtros com *kernels* de tamanho 3×3 ; e
- DM que captura a solução com melhor valor de f_1 da fronteira de Pareto.

Foram realizados experimentos usando cada um dos possíveis índices para se estabelecer a função de *fitness* f_1 , ou seja, OA ou AA. Assim, as seguintes combinações de componentes foram testadas para cada um desses índices:

1. **SVM**: Classificador SVM-RBF sem o uso da informação espacial, ou seja, considerando apenas a informação espectral;
2. **SVM+informação espacial**: Classificador SVM-RBF com o uso da informação espacial obtida por filtro de média considerando os pixels da vizinhança;
3. **GMLC**: Classificador GMLC sem o uso da informação espacial, ou seja, considerando apenas a informação espectral; e
4. **GMLC+informação espacial**: Classificador GMLC com o uso da informação espacial obtida por filtro de média considerando os pixels da vizinhança.

A Tabela 5.2 apresenta os valores obtidos para cada uma das combinações de componentes nas três imagens comparadas. Nessa tabela, na primeira coluna é apresentado o índice utilizado como base para f_1 , a segunda coluna representa as combinações de componentes avaliadas nos experimentos e, também, a quantidade média de bandas obtida por meio do *5-fold cross-validation*. As colunas restantes representam os resultados para um dos três índices avaliados (OA, AA e Coeficiente Kappa (KC)) e seus respectivos desvios padrões σ .

5.3.1.1 Análise dos Resultados do Grupo de Experimentos 1

- **Índice KC**: Conforme previamente apresentado, os valores adequados do índice KC para concordância quase perfeito entre classificador e GT ocorrem quando superam 0,8. Para esse critério, somente o SVM-RBF proporcionou resultados adequados quando considerados os dados de todas as imagens avaliadas com as bandas selecionadas. Resultados proporcionados por esse classificador indicaram concordância quase perfeita com o GT.

O GMLC conseguiu promover um desempenho bom para imagens Salinas e Pavia University, porém, para imagem Indian Pines, o desempenho foi ruim mesmo quando utilizada a informação espacial promovida pelo filtro de média.

Quando utilizada a informação espacial em conjunto ao SVM-RBF ou GMLC verifica-se que os valores foram maiores em ambos os casos. Porém, mesmo quando a informação

Tabela 5.2. Resultados da quantidade de bandas e classificação dos experimentos usando ou não a informação espacial, f_1 baseado em diferentes índices e com bandas selecionadas usando o MOEA/D.

Indian Pines							
f_1	Método	Resultados					
		OA %	σ	AA %	σ	KC	σ
OA	SVM (30 bandas)	83,7	1,8	78,6	3,2	0,81	0,02
	SVM+inf. espacial (30 bandas)	89,8	3,3	86,7	5,1	0,88	0,04
	GMLC (30 bandas)	61,0	5,2	59,3	4,3	0,56	0,06
	GMLC+inf. espacial (30 bandas)	70,3	3,6	65,5	3,6	0,66	0,04
AA	SVM (30 bandas)	82,8	3,4	74,7	5,0	0,81	0,04
	SVM+inf. espacial (30 bandas)	86,2	2,1	81,3	5,1	0,84	0,02
	GMLC (30 bandas)	58,9	4,0	55,3	2,5	0,53	0,04
	GMLC+inf. espacial (30 bandas)	70,3	3,9	65,4	3,7	0,66	0,04
Salinas							
f_1	Método	Resultados					
		OA %	σ	AA %	σ	KC	σ
OA	SVM (30 bandas)	92,8	0,6	96,6	0,3	0,92	0,01
	SVM+inf. espacial (30 bandas)	95,5	0,5	97,8	0,3	0,95	0,01
	GMLC (30 bandas)	90,2	1,2	96,0	0,6	0,89	0,01
	GMLC+inf. espacial (30 bandas)	90,5	0,6	96,0	0,5	0,89	0,01
AA	SVM (30 bandas)	92,3	0,8	96,5	0,4	0,91	0,01
	SVM+inf. espacial (27 bandas)	95,2	0,8	97,8	0,4	0,95	0,01
	GMLC (30 bandas)	90,2	1,0	95,8	0,3	0,89	0,01
	GMLC+inf. espacial (30 bandas)	94,9	0,7	97,8	0,3	0,94	0,01
Pavia University							
f_1	Método	Resultados					
		OA %	σ	AA %	σ	KC	σ
OA	SVM (27 bandas)	93,2	0,3	91,4	1,9	0,91	0,01
	SVM+inf. espacial (29 bandas)	97,1	0,6	96,1	1,0	0,96	0,01
	GMLC (23 bandas)	91,3	0,9	88,7	1,7	0,88	0,01
	GMLC+inf. espacial (24 bandas)	96,7	0,25	94,3	1,1	0,96	0,00
AA	SVM (29 bandas)	93,0	0,9	91,5	1,7	0,90	0,01
	SVM+inf. espacial (29 bandas)	96,7	0,3	95,9	0,5	0,95	0,00
	GMLC (20 bandas)	90,8	1,1	90,3	1,5	0,88	0,01
	GMLC+inf. espacial (21 bandas)	95,8	1,3	91,1	6,5	0,94	0,02

espacial foi utilizada para o GMLC, esse algoritmo não conseguiu atingir o índice de 0,8 para imagem Indian Pines.

- **Índices OA e AA:** O bom desempenho proporcionado pelo SVM-RBF também é apresentado para os índices OA e AA, e a diferença dos resultados entre esse classificador e o GMLC é apresentada em todos os casos, porém, é maior para a imagem Indian Pines. Quando utilizada a informação espacial, a melhoria de desempenho acontece para ambos os classificadores e para todas os índices avaliados. A vantagem do SVM-RBF em relação

ao GMLC ocorre tanto para o uso da informação espacial proporcionada pela aplicação do filtro de média, quanto sem a utilização dessa informação.

Por exemplo, para a imagem Indian Pines, o índice OA com o uso da informação espacial, o SVM-RBF apresentou 89,8 (destacado em negrito na Tabela 5.2) contra 70,3 de valor médio de OA quando o GMLC foi aplicado. Já para a mesma utilização da informação espacial e f_1 baseado no índice de OA o valor obtido pelo SVM-RBF foi bem maior que o valor médio de 83,7 apresentado quando a informação espacial não foi utilizada. Ainda para imagens Indian Pines, e quando o índice de AA é usado como base para f_1 , o SVM apresentou o valor médio de 86,2 para OA com o uso da informação espacial e 82,8 sem o uso da informação espacial. Para as imagens Salinas e Pavia University, o GMLC promoveu resultados bastante próximos aos proporcionados pelo SVM-RBF, principalmente quando AA foi utilizada como base para f_1 .

A utilização da informação espacial proporcionada pelo uso de filtragem espacial pode ter auxiliado na redução de ruídos presentes nas bandas espectrais. Isso pode ter ocorrido porque elas são provenientes de Sensoriamento Remoto e podem ser afetadas por interferências entre os sensores e objetos monitorados.

Uma possível justificativa para o desempenho inferior do GMLC é devido à necessidade desse algoritmo ter uma quantidade maior de amostras de treinamento para poder possibilitar a obtenção adequada dos vetores de média e matrizes de covariância, ou seja, para que as classes possam ser adequadas à suposição de terem distribuições Gaussianas.

O algoritmo SVM-RBF, como é um classificador não paramétrico, não necessita fazer suposições sobre a distribuição dos dados, o que provavelmente permitiu um melhor desempenho de classificação. Além disso, esse algoritmo é frequentemente apontado como uma boa estratégia para quando poucas amostras estão disponíveis, sendo uma boa alternativa para IHS, o que fortalece a escolha dele como opção para compor o WMoEBS.

Quando comparadas aos dois índices possíveis para serem utilizadas no estabelecimento de f_1 e assim conduzir o processo de seleção de bandas em conjunto a f_2 , observa-se que a OA, permitiu evoluir os três índices (OA, AA e KC) avaliados conjuntamente nos três casos analisados. Uma importante observação é que mesmo para o índice de AA a seleção de bandas conduzida por f_1 baseada no OA, conseguiu apresentar bons resultados. Isso mostra que houve pouca influência em amostras com classes desbalanceadas e, portanto, motivou a escolha de OA para composição do WMoEBS.

Para todos os casos, a utilização da informação espacial em conjunto ao classificador SVM-RBF, OA como base para f_1 e bandas selecionadas, foi a combinação que apre-

sentou os melhores resultados de classificação e isso foi um primeiro indicativo para sua escolha na composição do WMoEBS.

- **Quantidade de bandas:** Para as imagens Indian Pines e Salinas pode-se observar que a quantidade média de bandas obtida para maioria dos casos foi de 30, o que pode sugerir que o limiar utilizado na estratégia de reparo para essa imagem está muito baixo e valores maiores poderiam ser considerados. Quantidades menores de bandas podem ser observadas principalmente para a imagem Pavia University, onde, para esse critério, o GMLC foi melhor que o SVM com ou sem informação espacial. Análises sobre a quantidade de bandas são detalhadas no terceiro grupo de experimentos.

Síntese dos Resultados do Grupo de Experimentos 1: Baseado nas avaliações realizadas no primeiro grupo de experimentos, foram possíveis realizar escolhas de alguns dos componentes para compor o WMoEBS. Os primeiros componentes escolhidos são listados a seguir:

- Classificador: SVM-RBF, principalmente por operar bem com poucas amostras e apresentar bons resultados nos experimentos;
- Utilização da informação espacial por meio de filtros: pois junto ao componente classificador SVM-RBF, permitiu bons resultados de classificação sendo muitas vezes bem melhores que a não utilização; e
- OA para compor f_1 : pois apresentou a maioria dos melhores resultados com os três índices analisados, ou seja, OA, AA e KC.

Essas definições possibilitaram reduzir os números de testes necessários nos experimentos posteriores. O filtro de média utilizado anteriormente, poderia não ser o mais adequado para considerar as informações espaciais e, portanto, outros filtros considerados.

5.3.2 Grupo de experimentos 2: Avaliação de diferentes filtros para obtenção de informações espaciais

No segundo grupo de experimentos, foram consideradas as avaliações dos seguintes filtros com os respectivos parâmetros:

- (1) Filtro de média;
- (2) Filtro de mediana;
- (3) Filtro Gaussiano com $\sigma = 1$ (desvio padrão espacial); e

- (4) Filtro bilateral com $\sigma_s = 75$ e com $\sigma_r = 75$ (desvio padrão do eixo de intensidade).

O parâmetro do filtro Gaussiano foi escolhido para evitar que os resultados fossem próximos ao filtro de Média, por isso o valor pequeno. Já os parâmetros do filtro bilateral, seguiu-se recomendações da documentação do OPENCV (BRADSKI, 2000). Esses filtros proporcionaram a obtenção e consideração da informação espacial integrada ao processo de seleção de bandas. Portanto, nesse grupo de experimentos, as seguintes combinações foram avaliadas:

1. **SVM+inf. espacial c/ filtro de média:** Classificador SVM-RBF em conjunto com a informação espacial obtida pelo filtro de média aplicado à vizinhança dos pixels com *kernel* de tamanho 3×3 e sem sobreposição;
2. **SVM+inf. espacial c/ filtro de mediana:** Classificador SVM-RBF em conjunto com a informação espacial obtida pelo filtro de mediana aplicado à vizinhança com *kernel* de tamanho 3×3 e sem sobreposição;
3. **SVM+inf. espacial c/ filtro Gaussiano:** Classificador SVM-RBF em conjunto com a informação espacial obtida pelo filtro Gaussiano aplicado à vizinhança com *kernel* de tamanho 3×3 e sem sobreposição; e
4. **SVM+inf. Espacial c/ filtro bilateral:** Classificador SVM-RBF em conjunto com a informação espacial obtida pelo filtro bilateral com *kernel* de tamanho 3×3 e sem sobreposição para ambos os domínios (espacial e faixa de intensidades).

Com a utilização dos filtros espaciais, as configurações dos experimentos para esse grupo foram as mesmas utilizadas no primeiro caso.

Os resultados da classificação dos pixels de cada uma das três IHS com bandas selecionadas pela estratégia proposta, considerando a aplicação dos diferentes filtros espaciais são apresentados na Tabela 5.3. A primeira coluna dessas tabelas apresenta os diferentes filtros que foram utilizados no processo de seleção de bandas e posteriormente na classificação e as colunas seguintes apresentam os resultados obtidos para as diferentes índices considerados, ou seja, OA, AA e KC e os respectivos desvios padrões. Assim como nos casos do primeiro grupo de experimentos, os melhores valores incluindo empates são destacados em negrito.

5.3.2.1 Análise dos Resultados do Grupo de Experimentos 2

- **Índice KC:** A informação espacial quando obtida a partir dos filtros propostos, possibilitou bons resultados para todos os casos, atingindo um valor superior a 0,8. Isso confirma a vantagem de sua utilização para tarefas de classificação e segmentação.

Tabela 5.3. Resultados da quantidade de bandas e classificação dos experimentos considerando a utilização da informação espacial obtida por diferentes filtros e com bandas selecionadas usando o MOEA/D.

Indian Pines						
Filtro aplicado	Resultados					
	OA %	σ	AA %	σ	KC	σ
Média (30 bandas)	89,8	3,3	86,7	5,1	0,88	0,04
Mediana (30 bandas)	84,3	1,9	80,8	4,6	0,82	0,02
Gaussiano (30 bandas)	88,1	1,9	84,4	5,9	0,86	0,02
Bilateral (30 bandas)	83,6	3,1	79,3	6,7	0,81	0,04
Salinas						
Filtro aplicado	Resultados					
	OA %	σ	AA %	σ	KC	σ
Média (30 bandas)	95,5	0,5	97,8	0,3	0,95	0,01
Mediana (30 bandas)	93,6	0,7	96,9	0,4	0,93	0,01
Gaussiano (30 bandas)	94,9	0,3	97,6	0,1	0,94	0,00
Bilateral (30 bandas)	93,5	0,4	96,9	0,3	0,93	0,01
Pavia University						
Filtro aplicado	Resultados					
	OA %	σ	AA %	σ	KC	σ
Media (29 bandas)	97,1	0,6	96,1	1,0	0,96	0,01
Mediana (30 bandas)	95,5	0,5	94,5	1,0	0,94	0,01
Gaussiano (28 bandas)	97,1	0,8	95,8	1,2	0,96	0,01
Bilateral (28 bandas)	94,3	0,2	92,6	1,6	0,92	0,00

- **Índices OA e AA:** O filtro de média se destacou em relação aos outros, apresentando o valor de 89,8 de OA para classificação para imagem Indian Pines, maior que os valores de 88,1, 84,3 e 83,6 respectivamente para os filtros Gaussiano, de mediana e bilateral. De forma semelhante o índice OA, o filtro de média também proporcionou os melhores resultados do índice AA. Os mesmos comportamentos para o critério de classificação apresentados na imagem Indian Pines são repetidos para as imagens Salinas e Pavia University, onde o filtro de média proporcionou a maioria dos melhores resultados. Especificamente para imagem Pavia University, observa-se que o filtro Gaussiano apresentou os mesmos valores de OA que o filtro de média, porém foi inferior para o índice AA.

Uma provável explicação para o filtro bilateral ter apresentado resultados inferiores aos outros métodos é devido à difícil estimativa dos parâmetros para múltiplas bandas espectrais das IHS e um estudo adequado sobre a estimativa desses parâmetros podem ser realizados no futuro. O filtro de mediana pode não ter operado tão bem como o filtro de média, por causa das grandes variações de valores espectrais que podem ocorrer em cada uma das bandas das IHS, causando grandes modificações nos valores dos pixels por causa de sua vizinhança. Para o filtro Gaussiano que proporcionou os resultados que

melhor se aproximam dos resultados do filtro de média, *kernels* maiores que o tamanho testado podem ser necessários para melhorar o desempenho. Uma provável explicação para os bons resultados proporcionados pelo filtro de média é que ele suaviza somente as bandas que apresentam alguma interferência de ruídos e não interferem em casos onde a vizinhança dos pixels apresenta um padrão, além disso não dependem de um *kernel* de tamanho muito grande para operar corretamente.

- **Quantidade de bandas:** Uma importante observação, é que quando qualquer um dos filtros foi aplicado em conjunto com a classificação do SVM-RBF, a quantidade de bandas que proporcionou os melhores resultados sempre foi de 30, sugerindo novamente que o limiar de 30 bandas pode ser um valor baixo para imagem Indian Pines. Para imagem Pavia University, foram apresentados os valores de 28 bandas para os filtros Gaussianos e bilateral e 29 bandas para o filtro de Média, sugerindo-se que para essa imagem, o limiar de 30 bandas funcionou melhor do que quando usado para imagem Indian Pines.

Síntese dos Resultados do Grupo de Experimentos 2: Os resultados obtidos nesse grupo de experimentos demonstraram que para a maioria dos casos o filtro de média proporcionou os melhores resultados para o uso da informação espacial em conjunto à seleção de bandas e classificação de IHS. Além disso, para os casos onde outros filtros proporcionaram uma quantidade menor de bandas, essa quantidade foi bastante pequena. Portanto, o filtro de média foi escolhido para fazer parte do WMoEBS e utilizado nos experimentos seguintes.

5.3.3 Grupo de experimentos 3: Avaliação dos tomadores de decisão (*Decision Maker - DM*)

Neste subcapítulo com diferentes tomadores de decisão (DMs) foram consideradas as mesmas configurações dos dois primeiros grupos, porém aqui, em conjunto com os diferentes DMs, foram utilizados apenas o filtro de média para obtenção da informação espacial. Para isso, as seguintes combinações foram analisadas:

1. Centróide: DM baseado no centróide;
2. Centróide + *Sequential Backward Floating Selection* (SBFS): DM baseado no centróide + SBFS;
3. Seleção direta: DM baseado na seleção direta de uma solução da FP considerando um limiar; e

4. Seleção direta + SBFS: DM baseado na seleção direta de uma solução da FP considerando um limiar + SBFS.

Assim como nos casos anteriores, os resultados da classificação dos pixels de cada uma das três IHS com bandas selecionadas pela estratégia proposta, considerando a aplicação dos diferentes DMs são apresentados na Tabela 5.4. A primeira coluna dessa tabela apresenta o nome do DM cujos resultados são analisados e as colunas seguintes apresentam os diferentes resultados obtidos.

Como os resultados apresentados na aplicação dos diferentes DMs foram bastante próximos, para esse grupo de experimentos também se realizou o teste estatístico T-pareado para os índices de OA e AA e os resultados são mostrados nas Tabelas 5.5 e 5.6. Conforme já apresentado, esse teste indica que a hipótese nula não deve ser rejeitada quando o *p-value* apresenta valores maiores que um determinado limiar, aqui sendo considerado 0,05, indicando que a diferença da média dos resultados proporcionados por dois classificadores é zero. O índice KC foi omitido do teste T-pareado porque seus resultados são altamente correlacionados com o índice OA, variando conjuntamente e, portanto, apresentando praticamente os mesmos *p-values*. Nessas tabelas, o valor destacado em negrito, indica que o método da linha apresentou resultados melhores de OA ou AA que o método da coluna, do contrário os métodos tiveram resultados médios idênticos ou o método da linha é inferior.

5.3.3.1 Análise dos Resultados do Grupo de Experimentos 3

- **Índice KC:** Todos os DMs proporcionaram bons resultados, sugerindo novamente que a utilização de algum desses métodos como componente do WMoEBS é viável.
- **Índices OA e AA:** Para esses índices, os maiores valores relacionados à classificação foram proporcionados para o DM baseado no centroide e para o DM baseado na seleção direta, que diferentemente dos outros dois métodos, empataram em todos os casos nos testes estatísticos (*p-value* maior que 0,05).

Quando o método SBFS foi incorporado em qualquer um dos dois DMs (Centroide ou Seleção direta), ele conseguiu reduzir a quantidade de bandas, porém ocasionou um pouco de redução no desempenho da classificação para todos os índices considerados e isso é mostrado inclusive pelos testes estatísticos. Isso foi a indicação que motivou o descarte do SBFS para composição do WMoEBS.

- **Quantidade de bandas:** Com relação à quantidade de bandas, foi mostrado que os DMs que incorporam o SBFS conseguiram a menor quantidade de bandas e essa diferença foi

Tabela 5.4. Resultados da quantidade de bandas e classificação dos experimentos considerando diferentes DMs e com bandas selecionadas usando o MOEA/D.

Indian Pines						
DMs	Resultados					
	OA %	σ	AA %	σ	KC	σ
Centroide (29 bandas)	89,8	3,5	86,2	6,2	0,88	0,04
Centroide + SBFS (20 bandas)	88,8	2,4	84,2	6,1	0,87	0,03
Seleção direta (30 bandas)	89,8	3,3	86,7	5,1	0,88	0,04
Seleção direta + SBFS (22 bandas)	87,8	3,3	82,9	6,3	0,86	0,04
Salinas						
DMs	Resultados					
	OA %	σ	AA %	σ	KC	σ
Centroide (29 bandas)	95,2	0,6	97,6	0,3	0,95	0,01
Centroide + SBFS (21 bandas)	94,9	0,6	97,6	0,4	0,94	0,01
Seleção direta (30 bandas)	95,5	0,5	97,8	0,3	0,95	0,01
Seleção direta + SBFS (19 bandas)	94,7	0,6	97,5	0,4	0,94	0,01
Pavia University						
DMs	Resultados					
	OA %	σ	AA %	σ	KC	σ
Centroide (28 bandas)	97,2	0,5	96,1	1,1	0,96	0,01
Centroide + SBFS (20 bandas)	96,6	0,6	95,6	1,2	0,96	0,01
Seleção direta (29 bandas)	97,1	0,6	96,1	1,0	0,96	0,01
Seleção direta + SBFS (20 bandas)	96,6	0,9	95,8	1,3	0,95	0,01

Tabela 5.5. Resultados dos *p-values* do teste T-pareado para o índice OA para estratégia proposta para os diferentes DMs analisados.

Indian Pines				
DM	Centroide	Centroide+SBFS	Seleção direta	Seleção direta+SBFS
Centroide	-	0,22	0,99	0,30
Centroide+SBFS	0,22	-	0,22	0,46
Seleção direta	0,99	0,22	-	0,12
Seleção direta+SBFS	0,30	0,46	0,12	-
Salinas				
DM	Centroide	Centroide+SBFS	Seleção direta	Seleção direta+SBFS
Centroide	-	0,05	0,44	0,28
Centroide+SBFS	0,05	-	0,08	0,24
Seleção direta	0,44	0,08	-	0,05
Seleção direta+SBFS	0,28	0,24	0,05	-
Pavia University				
DM	Centroide	Centroide+SBFS	Seleção direta	Seleção direta+SBFS
Centroide	-	0,04	0,59	0,19
Centroide+SBFS	0,04	-	0,06	0,89
Seleção direta	0,59	0,06	-	0,15
Seleção direta+SBFS	0,19	0,89	0,15	-

Tabela 5.6. Resultados dos *p-values* do teste T-pareado para o índice AA para estratégia proposta para os diferentes DMs analisados.

Indian Pines				
DM	Centroide	Centroide+SBFS	Seleção direta	Seleção direta+SBFS
Centroide	-	0,08	0,62	0,19
Centroide+SBFS	0,08	-	0,17	0,60
Seleção direta	0,62	0,17	-	0,05
Seleção direta+SBFS	0,19	0,60	0,05	-
Salinas				
DM	Centroide	Centroide+SBFS	Seleção direta	Seleção direta+SBFS
Centroide	-	0,07	0,40	0,92
Centroide+SBFS	0,07	-	0,14	0,28
Seleção direta	0,40	0,14	-	0,09
Seleção direta+SBFS	0,92	0,28	0,09	-
Pavia University				
DM	Centroide	Centroide+SBFS	Seleção direta	Seleção direta+SBFS
Centroide	-	0,19	0,97	0,63
Centroide+SBFS	0,19	-	0,04	0,65
Seleção direta	0,97	0,04	-	0,42
Seleção direta+SBFS	0,63	0,65	0,42	-

alta em relação a sua não utilização, porém foi considerado que não é uma justificativa viável a incorporação dessa estratégia no WMoEBS visto que os métodos promoveram uma diminuição na qualidade de classificação.

Entre a utilização do DM baseado no centroide e do DM baseado na seleção direta, ambos os que apresentaram melhores resultados de classificação, o primeiro caso foi discretamente melhor em relação a essa quantidade.

Síntese dos Resultados do Grupo de Experimentos 3: Considerando os resultados dos índices OA e AA e a análise dos *p-values* para todas as imagens e todos os DMs que foram testados, na Tabela 5.7, são mostrados um resumo dos resultados com a contagem de empates, vitórias e derrotas de cada comparação. Respectivamente, nessa tabela, o primeiro símbolo nas colunas é referente à imagem Indian Pines, o segundo sinal referente à imagem Salinas e o terceiro referente à imagem Pavia University.

O símbolo + indica que o método da linha apresentou resultados melhores que o método da coluna, o símbolo de igual indica empate e o símbolo de - que o método apresentou resultados inferiores. Sempre que os resultados de um método foram melhores ou iguais a todos os outros comparados, esse método está destacado em negrito para facilitar a visualização.

Como pode-se observar, os DMs que proporcionaram melhores resultados em relação aos outros no critério de classificação foram o DM baseado em centroide e o DM baseado na sele-

Tabela 5.7. Comparação das vitórias considerando o *p-value* dos DM baseados em diferentes estratégias para todas imagens analisadas.

Precisão - Overall Accuracy (OA)				
DM	Centroide	Centroide+SBFS	Seleção direta	Seleção direta+SBFS
Centroide	-	= + +	= = =	= = =
Centroide+SBFS	= - -	-	= = =	= = =
Seleção direta	= = =	= = =	-	= = +
Seleção direta+SBFS	= = =	= = =	= = -	-
Acurácia Média - Average Accuracy (AA)				
DM	Centroide	Centroide+SBFS	Seleção direta	Seleção direta+SBFS
Centroide	-	= = =	= = =	= = =
Centroide+SBFS	= = =	-	= = -	= = =
Seleção direta	= = =	= = +	-	+ = =
Seleção direta+SBFS	= = =	= = =	- = =	-

ção direta, portanto a escolha foi em função desses dois DMs descartando-se o uso do SBFS. Observa-se que o DM baseado na seleção direta, conseguiu ser ainda superior em três casos para os dois índices, enquanto o DM baseado em centroide foi superior em dois dos casos.

Tabela 5.8. Comparação das vitórias considerando a quantidade de bandas proporcionada pelos DMs baseado em centroide ou seleção direta.

DM	Indian Pines	Salinas	Pavia University
Centroide	29	29	28
Seleção direta	30	30	29

Mesmo com o fato do DM baseado na seleção direta ter sido superior em um caso a mais quando comparado ao outro, antes da decisão de qual dos dois casos escolher para compor o WMoEBS, observou-se o número de bandas. A Tabela 5.8 mostra o resumo dos resultados em relação ao número de bandas para esses dois DMs, onde observa-se que o DM baseado no centroide foi superior em relação ao outro para todos os casos comparados.

Considerando que os DM baseado no centroide:

- foi melhor que a maioria dos outros casos comparados e que apresentou estatisticamente a mesma média que o DM baseado na seleção direta para o critério de classificação; e
- ter apresentado melhores resultados em relação à quantidade de bandas que o DM baseado na seleção direta, além de não ser projetado somente para capturar a melhor seleção, mas sim uma que tenta ponderar os resultados dos diferentes *fitness*;

nesta tese optou-se pela escolha desse método para compor o WMoEBS.

5.4 WMoEBS

A partir dos experimentos realizados, o WMoEBS foi estruturado com base nos componentes selecionados. A estrutura do WMoEBS é apresentada no esquema da Figura 5.4 e um resumo de suas etapas são listadas na sequência.

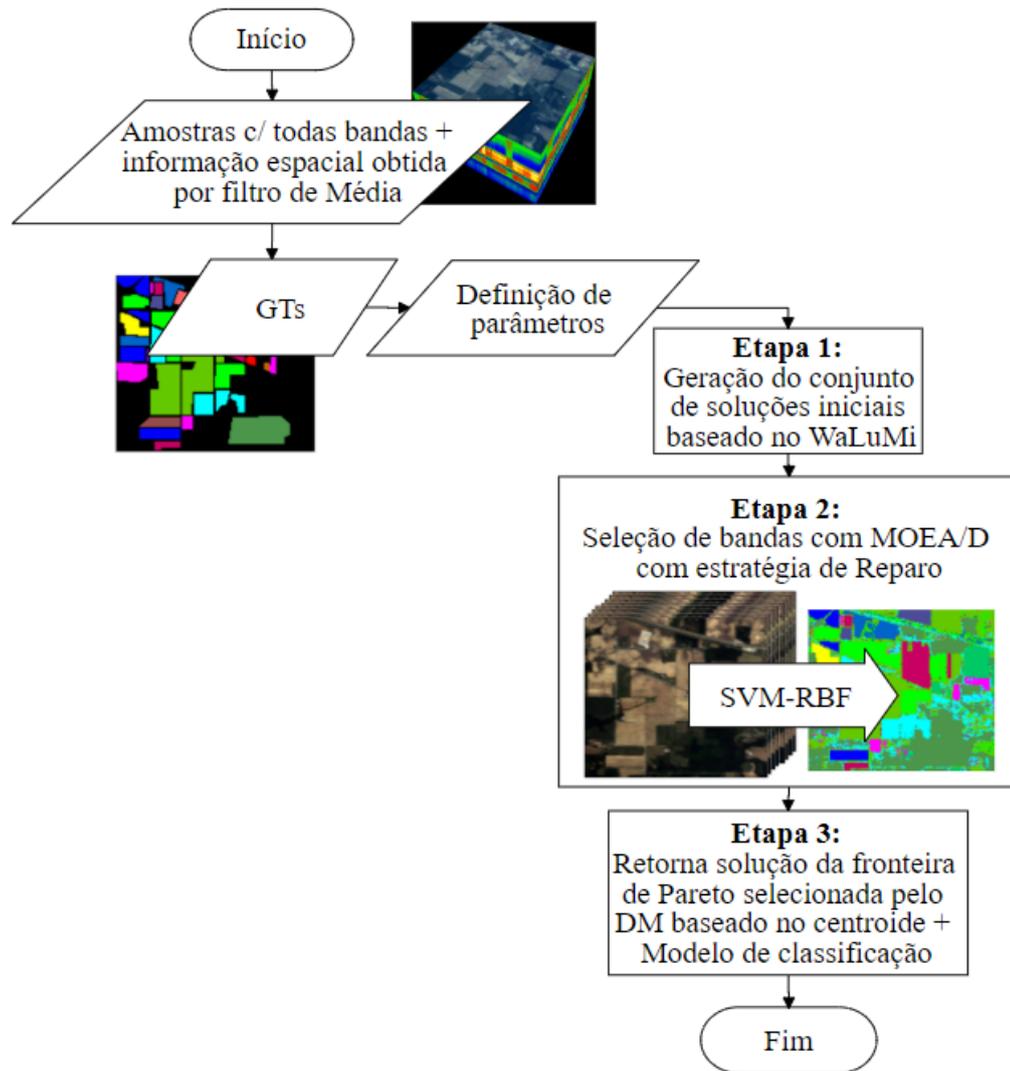


Figura 5.4. Wrapper multiobjective evolutionary band selection (WMoEBS)

- **Definição de parâmetros:** Inicia-se o WMoEBS com os pixels de amostras com todas as bandas, GT, parâmetros do SVM-RBF e considerando a informação espacial provinda de um filtro de média considerando o *kernel* de tamanho 2×2 e estratégia previamente apresentados. Como foi aplicado um filtro de média, não foi necessário um *kernel* com muitos pixels permitindo utilizar o tamanho de 2×2 ;
- **Etapa 1:** É gerado o conjunto de soluções iniciais, onde essas soluções representam

bandas que são derivadas de um agrupamento baseado no WaLuMi que é aplicado de forma a evitar correlação entre as bandas.

- **Etapa 2:** É aplicado o processo de seleção de bandas, onde o procedimento é baseado no MOEA/D e tem a particularidade de aplicar uma estratégia de reparo para estabelecer restrições em relação à quantidade de bandas. Nesse procedimento, o SVM-RBF foi utilizado para classificar os pixels para estabelecer as funções f_1 (baseada na OA e 3-*folds cross-validation*) e f_2 (baseada na quantidade de bandas).
- **Etapa 3:** Baseado na fronteira de Pareto obtida, um processo de tomada de decisão, ou seja, o DM é utilizado para escolher uma única solução dessa fronteira. A solução selecionada representa o conjunto de bandas identificado e é retornada em conjunto com o modelo classificador treinado para posteriormente serem usadas em novas classificações. O DM utilizado nessa etapa é o baseado em centroide inspirado a partir dos estudos de Kimovski et al. (2018).

Com o WMoEBS aplicado, a outra etapa deste estudo consistiu em avaliá-lo por meio de comparações com outros métodos da literatura. Os experimentos que permitiram essas avaliações são apresentados no capítulo seguinte.

5.5 Considerações Finais do Capítulo

Este capítulo apresentou todas etapas e elementos para constituição do método proposto que é avaliado nesta tese. Esses elementos incluíram a constituição de um *framework*, componentes (algoritmos, estratégias, etc) e experimentos que levaram a constituição do WMoEBS. Os componentes foram avaliados experimentalmente e os que apresentaram melhores resultados utilizados na composição do método proposto. No fim do capítulo, o WMoEBS foi apresentado e é comparado com outros métodos da literatura no próximo capítulo.

Capítulo 6

EXPERIMENTOS COMPARATIVOS E RESULTADOS

Neste capítulo são apresentados os experimentos realizados no que se referem a comparações do *Wrapper Multiobjective Evolutionary Band Selection* (WMoEBS) com outros métodos de seleção de bandas da literatura, incluindo o *Incorporated Decision-Marker-Based Multiobjective Band Selection* (IDMMoBS) produzido durante esta pesquisa. Conforme previamente apresentado, o WMoEBS considera a utilização da informação espacial, porém, para demonstrar a efetividade da seleção de bandas também foi considerada a utilização do método sem o uso da informação espacial. Nesse caso, a notação WMoEBS2 é utilizada.

Os métodos utilizados nas comparações estão categorizados e organizados da mesma forma dos grupos apresentados no Capítulo 3. Portanto, adicionalmente à avaliação do WMoEBS, permitiram observar o comportamento de um método multiobjetivo baseado na estratégia *Wrapper*, comparando-o com métodos dessa mesma estratégia e das estratégias *Filter* e *Embedded* que utilizam conceitos supervisionados e não supervisionados. Uma importante observação, assim como os casos anteriores, o que são avaliados são a quantidade de bandas selecionadas das imagens hiperespectrais (IHs) e os resultados de classificação dos pixels com essas bandas proporcionados pelos métodos analisados. Para realização desses experimentos, foram consideradas as IHs apresentadas e utilizadas nos capítulos anteriores.

6.1 Configurações de amostras e experimentos comparativos

Similarmente aos experimentos realizados para a estruturação do WMoEBS, quando foi utilizada a informação espacial obtida por meio do filtro de média em cada banda espectral no WMoEBS, a estratégia de "*kernel* sem sobreposição" foi aplicada. Como foi aplicado um filtro de média, não foi necessário um *kernel* com muitos pixels portanto, sendo possível utilizar o de tamanho 2×2 . Isso possibilitou considerar um maior número de amostras nas seleções, treinamentos e testes dos classificadores proporcionando uma confiabilidade ainda maior nos

resultados. Um importante destaque, é que esse *kernel* foi aplicado somente quando o objetivo foi comparar o WMoEBS com a informação espacial, e em casos onde não foi considerada essa informação, foi possível a utilização de todos os pixels como amostras.

Para todos os casos avaliados, a estratégia *5-fold cross-validation* foi aplicada considerando as particularidades de cada grupo de experimentos e métodos. O *Support Vector Machines* (SVM) com *kernel RBF* e usando exatamente as mesmas estratégias aplicadas durante a composição do WMoEBS foi aplicado em todos os casos de comparação. A estratégia de *5-fold cross-validation* também foi aplicada para estimar os parâmetros do SVM em cada caso.

Os índices de avaliação de classificação analisados foram a precisão (*Overall Accuracy - OA*), acurácia média (*Average Accuracy - AA*), coeficiente de Kappa (*Cohen's Kappa - KC*), além da quantidade de banda selecionadas. Sempre que necessário e possível, foram demonstrados graficamente os processos de evolução de busca de alguns algoritmos e também realizados os testes estatísticos (testes T-Pareado com validação cruzada) nos resultados de classificação.

Existem diversos experimentos com seleção de bandas que podem ter seus resultados comparados diretamente, sem a necessidade de implementação dos métodos que estão sendo comparados ao proposto. Outros experimentos requerem mais esforço com a implementação de métodos, porém possibilitam uma análise mais detalhada e completa dos métodos. Ainda, é possível avaliar diretamente o desempenho de classificação considerando bandas selecionadas por determinados métodos em relação às bandas selecionadas pelo método proposto. Para a avaliação do WMoEBS, três grupos de experimentos foram realizados:

- Grupo de experimentos 1: Comparação direta de resultados, em que são considerados os experimentos que possibilitaram a comparação direta de resultados apresentados na literatura. Esses experimentos estão bem detalhados nos trabalhos da literatura, o que permitiu replicá-los para gerar resultados com o método proposto e, posteriormente, compará-los com os resultados de outros métodos. A vantagem da replicação de experimentos é que possibilita demonstrar a capacidade do WMoEBS2 (sem a informação espacial), quando comparado com diversos outros métodos da literatura, sem ter que implementá-los, ou seja, comparando-se diretamente os resultados. Uma desvantagem é que testes mais detalhados, como testes estatísticos que permitem avaliações mais confiáveis, são de difícil aplicação e por isso o segundo grupo de experimentos foi elaborado;
- Grupo de experimentos 2: Experimentos com métodos comparativos implementados, em que são considerados métodos da literatura que foram re-implementados durante esta pesquisa. A vantagem desse grupo é que os experimentos foram adequados para permitirem testes estatísticos. Portanto, resultados e análises mais detalhadas foram possíveis.

- Grupo de experimentos 3: IHS com todas as bandas e bandas selecionadas pelo WMoEBS, em que são considerados experimentos com classificação e a utilização de todas as bandas espectrais de cada imagem em relação ao método proposto com e sem informação espacial. De forma similar ao grupo anterior, testes estatísticos foram possíveis.

As configurações gerais de todos os experimentos e WMoEBS são mostradas na Tabela 6.1. Com exceção das configurações de informação espacial, esses mesmos parâmetros também são utilizados no WMoEBS2.

Tabela 6.1. Configurações gerais dos experimentos e do WMoEBS.

Parâmetro/Algoritmo	Valor
Iterações (gerações)	500
Número de Soluções	100
Geração da população inicial	Método baseado no <i>WaLuMi</i>
Cruzamento	Cruzamento de Dois Pontos
Mutação	<i>Troca de bits (bit-flip)</i>
Taxa de Mutação	5%
Quantidade de vizinhos	20
f_1	OA (3-fold cross-validation)
Classificador para f_1	SVM-RBF
Limiar para f_2	30 ou 20
Estratégia de reparo	Limiar adaptativo
Limiar de reparo	30 ou 20
filtragem para informação espacial	Filtro de Média
<i>kernel</i>	2×2
<i>Índices avaliados</i>	OA, AA, KC e Quantidade de bandas

6.2 Grupo de experimentos 1: Experimentos com comparação direta de resultados

Nesse grupo foram replicados alguns dos experimentos realizados em Xie et al. (2018) permitindo a comparação direta dos resultados de diferentes métodos avaliados nesse estudo com os resultados proporcionados pelo WMoEBS e WMoEBS2. Os métodos utilizados já foram justificados e discutidos no Capítulo 3, porém suas particularidades são descritas quando necessário e também estão organizados em métodos de aprendizado supervisionado (ou semi-supervisionado) e não supervisionado. Essa organização foi utilizada na comparação, pois o uso do *Ground Truth* (GT) pode influenciar na qualidade da classificação.

- Supervisionados:

- *Trivariate mutual information-clonal selection algorithm* (TMI-CSA) (FENG et al., 2014);
- *Semi-supervised band selection approach based on TMI and graph regulation* (ST-MIGR) (FENG et al., 2014);
- *Information gain - gray wolf optimizer* (IG-GWO) (XIE et al., 2018);
- Não Supervisionados:
 - *Ward's linkage and Mutual Information* (WaLuMi) (MARTINEZ-USO et al., 2007);
e
 - *Maximum information and minimum redundancy-clonal selection algorithm* (MIMR-CSA) (FENG et al., 2016).

Em Xie et al. (2018) as imagens consideradas foram as mesmas desta pesquisa, que são a Indian Pines, Salinas e Pavia University. O SVM com *5-fold cross-validation* com *radial basis function* (RBF) também foi utilizado para classificar os pixels das imagens com bandas selecionadas por cada um dos métodos estudados, ou seja, o mesmo classificador que foi considerado nas avaliações do WMoEBS e WMoEBS2. A mesma estratégia de estimativa de parâmetros do SVM também foi aplicada. A abordagem de classificação utilizada no SVM também é um contra todos. Para métodos que necessitam de uma quantidade pré-determinada de bandas, foram definidas 30 bandas para imagens Indian Pines e Salinas enquanto para imagens Pavia University foram definidas 20 bandas. Essas mesmas quantidades foram utilizadas como limiar máximo da estratégia de reparo para o WMoEBS e WMoEBS2.

Uma diferença entre o WMoEBS e os métodos comparados é a utilização da informação espacial, o que permite observar a influência do uso dessa informação frente a outros métodos da literatura. Além disso, como o filtro espacial com *kernel* 2×2 sem sobreposição foi aplicado, o WMoEBS teve uma quantidade inferior de pixels de treinamento em relação aos demais métodos comparados, o que pode influenciar negativamente em seus resultados. Justamente por isso e para uma avaliação mais adequada do processo de seleção de bandas em si, sem considerar a informação espacial, o WMoEBS2 também foi aplicado e comparado com outros métodos da literatura. Particularmente, para o método WMoEBS2 foi possível utilizar exatamente a mesma configuração do Xie et al. (2018), ou seja, todas as amostras possíveis dos pixels puderam ser distribuídas aleatoriamente entre as multiplicidades (*folds*) consideradas. Ainda, conforme foi apresentado em Xie et al. (2018), o desvio padrão das quantidades de bandas foi omitido neste primeiro grupo de experimentos.

Os resultados de classificação e quantidade de bandas proporcionadas pelos diferentes métodos para as imagens Indian Pines, Salinas e Pavia University estão organizados nos grupos

supervisionado (ou semi-supervisionados) e não supervisionados e são apresentados a seguir.

6.2.1 Supervisionados

Os resultados de classificação e quantidade de bandas usados para comparar os métodos supervisionados com o método proposto são mostrados na Tabela 6.2.

Tabela 6.2. Resultados da quantidade de bandas e índices de classificação dos experimentos replicados com métodos supervisionados, WMoEBS e WMoEBS2.

Imagem	Método	OA %	σ	AA %	σ	KC	σ
Indian Pines	WMoEBS (28 bandas)	89,6	0,8	84,5	4,6	0,88	0,01
	WMoEBS2 (26 bandas)	89,3	0,4	89,9	1,4	0,88	0,01
	TMI-CSA (30 bandas)	72,9	2,0	55,5	2,3	0,69	0,02
	STMIGR (30 bandas)	74,4	1,0	57,4	1,2	0,70	0,01
	IG-GWO (30 bandas)	85,2	0,8	82,6	6,6	0,83	0,01
Salinas	WMoEBS (26 bandas)	94,9	0,3	97,8	0,1	0,94	0,00
	WMoEBS2 (28 bandas)	94,4	0,1	97,5	0,1	0,94	0,00
	TMI-CSA (30 bandas)	90,8	0,2	94,3	0,2	0,90	0,00
	STMIGR (30 bandas)	91,3	0,3	95,3	0,3	0,90	0,00
	IG-GWO (28 bandas)	93,9	0,2	97,0	0,7	0,93	0,00
Pavia University	WMoEBS (17 bandas)	96,6	0,2	95,8	0,7	0,95	0,00
	WMoEBS2 (20 bandas)	94,8	0,3	94,3	0,4	0,93	0,00
	TMI-CSA (20 bandas)	86,2	1,6	79,4	2,7	0,81	0,02
	STMIGR (20 bandas)	87,2	1,5	82,8	1,6	0,83	0,02
	IG-GWO (18 bandas)	94,2	0,2	92,3	1,0	0,92	0,00

6.2.1.1 Análise dos Resultados do Grupo de Experimentos 1: Supervisionados

Comparando-se os métodos WMoEBS e o WMoEBS2 com os métodos TMI-CSA, STMIGR e IG-GWO em cada uma dos índices são realizadas as seguintes observações:

- **Índice KC:** Considerando os três métodos supervisionados, ou seja, TMI-CSA, STMIGR e IG-GWO, com o WMoEBS e, também, o WMoEBS2, observa-se que somente as bandas selecionadas pelo IG-GWO, o WMoEBS e também o WMoEBS2 promoveram bons resultados para o KC, superiores a 0,8, para IH Indian Pines. Porém, para as imagens Salinas e Pavia University, todos os métodos tiveram o KC superior a 0,8, ou seja, resultados desejados. Esse primeiro índice é um indicativo que o IG-GWO e o método proposto podem proporcionar bons resultados mesmo sem a utilização da informação espacial.
- **Índices OA e AA:** Para OA e AA, o WMoEBS proporcionou os melhores desempenhos em relação aos outros algoritmos, seguido pelo próprio WMoEBS2, o que novamente sugere que o método proposto é superior aos métodos supervisionados modernos mesmo

quando a informação espacial não é utilizada. Apenas para imagem Indian Pines, o WMoEBS2 proporcionou resultados melhores para os índices AA do que quando utilizada a informação espacial no WMoEBS. Uma explicação para o melhor desempenho é que para o WMoEBS2 foi considerada uma maior quantidade de amostras durante a seleção de bandas, treinamento e testes permitindo uma elaboração de modelo e testes mais adequadas. Quando comparados com os outros métodos testados, observa-se que os resultados proporcionados pelo IG-GWO (*Filter*) foram os que mais se aproximaram dos resultados proporcionados pelo WMoEBS e WMoEBS2, porém ainda inferior.

- **Quantidade de bandas:** O WMoEBS e o WMoEBS2 também apresentaram bons resultados de quantidade de bandas quando comparados com os outros métodos. Além dos métodos propostos, o IG-GWO conseguiu reduzir a quantidade de bandas de forma próxima a cada um desses métodos para as imagens Salinas e Pavia University. Os outros métodos tiveram a quantidade de bandas pré-determinada.

Síntese dos Resultados do Grupo de Experimentos 1 – Supervisionados: Na observação direta da classificação com os índices de avaliação, o WMoEBS e WMoEBS2, ou seja, com ou sem a informação espacial, apresentaram melhores resultados em todos os casos analisados onde foram comparados com outros métodos supervisionados. O bom desempenho também é repetido para a quantidade de bandas. O método IG-GWO foi o que mais se aproximou dos valores do WMoEBS e WMoEBS para classificação e quantidade de bandas. Porém, mesmo quando a informação espacial não foi utilizada no método proposto (WMoEBS2), o resultado de classificação do IG-GWO foi inferior. Os outros métodos comparados tiveram uma quantidade pré-definida de bandas desejadas, porém mesmo com essa maior quantidade em relação aos outros, apresentaram resultados inferiores de classificação. Assim como mostrado previamente, o bom desempenho do WMoEBS é mais destacado quando a utilização da informação espacial foi utilizada, pois isso foi demonstrado para maioria dos casos.

6.2.2 Não Supervisionados

De forma similar ao caso anterior, os resultados usados para comparar o método proposto com métodos com aprendizado não supervisionados da literatura são mostrados na Tabela 6.3.

6.2.2.1 Análise dos Resultados do Grupo de Experimentos 1: Não Supervisionados

- **Índice KC:** Dos dois métodos não supervisionados analisados, o único que não apresentou um valor de KC maior que 0,8 para todos os casos foi o WaLuMi. O MIMR-CSA teve bom desempenho nesse caso, assim como o WMoEBS e o WMoEBS2.

Tabela 6.3. Resultados da quantidade de bandas e índices de avaliação de classificação dos experimentos replicados com métodos não supervisionados, WMoEBS e WMoEBS2.

Imagem	Método	OA %	σ	AA %	σ	KC	σ
Indian Pines	WMoEBS (28 bandas)	89,6	0,8	84,5	4,6	0,88	0,01
	WMoEBS2 (26 bandas)	89,3	0,4	89,9	1,4	0,88	0,01
	WaLuMi (30 bandas)	80,6	0,8	68,2	2,0	0,78	0,01
	MIMR-CSA (30 bandas)	84,5	0,9	79,4	1,8	0,82	0,01
Salinas	WMoEBS (26 bandas)	94,9	0,3	97,8	0,1	0,94	0,00
	WMoEBS2 (28 bandas)	94,4	0,1	97,5	0,1	0,94	0,00
	WaLuMi (30 bandas)	93,0	0,2	96,4	0,2	0,92	0,00
	MIMR-CSA (30 bandas)	93,5	0,2	96,8	0,1	0,93	0,00
Pavia University	WMoEBS (17 bandas)	96,6	0,2	95,8	0,7	0,95	0,00
	WMoEBS2 (20 bandas)	94,8	0,3	94,3	0,4	0,93	0,00
	WaLuMi (20 bandas)	90,1	1,0	86,7	1,2	0,87	0,01
	MIMR-CSA (20 bandas)	92,5	0,3	89,1	0,6	0,90	0,00

- **Índices OA e AA:** Assim como no KC, o método não supervisionado MIMR-CSA foi o que proporcionou melhores resultados para os índices OA e AA nos casos testados, porém com valores não muito próximos ao WMoEBS e o WMoEBS2, sendo esses últimos ainda melhores. O outro método não supervisionado, WaLuMi, não teve um desempenho bom de AA principalmente para imagem Indian Pines.
- **Quantidade de bandas:** Para esse índice, ambos os métodos não supervisionados comparados utilizam uma quantidade de bandas pré-definida, portanto apresentaram quantidades de bandas maiores que o WMoEBS e WMoEBS2, visto que o propósito é a redução dessas quantidades.

Síntese dos Resultados do Grupo de Experimentos 1 – Não Supervisionados: Os métodos não supervisionados, quando comparados com o WMoEBS e WMoEBS2 proporcionaram um desempenho inferior de classificação para todas as imagens analisadas. Nem todos os métodos não supervisionados conseguem um bom desempenho do índice KC para a imagem Indian Pines, ao contrário do método proposto que, com ou sem a informação espacial, sempre proporcionou bons resultados. De todos os métodos não supervisionados, o MIMR-CSA foi o que apresentou resultados mais próximos aos propostos, portanto no futuro pode ser explorado como alternativa para quando dados rotulados não estão disponíveis.

Os experimentos apresentados por Xie et al. (2018) neste subcapítulo foram reproduzidos para possibilitarem comparações de diferentes métodos em relação ao WMoEBS e WMoEBS2. O WMoEBS, com ou sem informação espacial, apresentou a maioria dos melhores resultados quando comparado diretamente com outros métodos da literatura para os índices avaliados em relação aos resultados de classificação ou quantidade de bandas. Para possibilitar uma expan-

são na análise de resultados e comprovar o bom funcionamento do WMoEBS, na sequência é apresentado o segundo grupo de experimentos onde diversos métodos comparativos foram reproduzidos e testes estatísticos também foram realizados.

6.3 Grupo de experimentos 2: Experimentos com métodos implementados

Nesta seção, as bandas selecionadas pelo WMoEBS, WMoEBS2 e outros métodos supervisionados e não supervisionados da literatura são comparadas. Assim como no caso anterior, o WMoEBS2 indica o método proposto sem considerar a informação espacial. Com exceção do *Incorporated Decision-Marker-based multiobjective band selection* (IDMMoBS) e do WMoEBS, o WMoEBS2 e outros métodos comparados não consideram a informação espacial e foram utilizados conforme as propostas originais. Nos experimentos deste grupo, para todos os casos foram utilizados exatamente os mesmos pixels selecionados para o WMoEBS (com informação espacial), sendo executado esse método novamente após os experimentos anteriores. Portanto, pode-se observar uma variação nos resultados do WMoEBS2 em relação aos experimentos do grupo anterior, que nos experimentos implementados uma menor quantidade de amostras foi utilizada.

Nesse grupo de experimentos, foi possível observar os valores médios da evolução do índice OA obtidos a partir do *3-fold cross-validation* e da quantidade de bandas de alguns dos métodos durante o processo de seleção de bandas. Além disso e como nos casos anteriores, outras avaliações consideraram os valores médios dos índices OA, AA e KC obtidos a partir da classificação com *5-fold cross-validation* das bandas selecionadas por diferentes métodos. Por fim, com os métodos comparativos implementados foram possíveis avaliações mais detalhadas com testes estatísticos e desvios padrões das quantidades de bandas para métodos que variam essa métrica.

Os métodos considerados e suas particularidades são listados na sequência e posteriormente são apresentadas as avaliações de cada caso:

- Supervisionados:
 - *Genetic Algorithm with Support Vector Machines* (GA-SVM) (NAGASUBRAMANIAN et al., 2017; ZHUO et al., 2008): Um método baseado na estratégia *Wrapper*, supervisionado e utilizando uma função de *fitness* com média ponderada entre OA da classificação e quantidade de bandas. Aqui é utilizada a função de *fitness* proposta por Zhuo et al. (2008) e previamente apresentada no Capítulo 3 na equação

3.5. Os pesos utilizados nessa função foram 0,8 para acurácia e 0,2 para a quantidade de bandas;

- *Single-layer neural network* (SLN) (HABERMANN; FREMONT; SHIGUEMORI, 2019);
 - *Classification and regression tree* (CART) (GOMEZ-CHOVA et al., 2003);
 - *Incorporated Decision-Marker-based multiobjective band selection* (IDMMoBS) (SAQUI et al., 2019c);
- Não Supervisionados:
 - *Incorporated rank based multi-objective band selection algorithm* (IRMOBS) (XU; SHI; PAN, 2017); e
 - WaLuMi.
 - Conjunto de bandas original: Pixels de cada imagem com todas as bandas.

6.3.1 Evolução dos resultados durante a execução de métodos de busca

Como métodos baseados em diferentes Algoritmos Evolutivos (AE) foram analisados, tais como o WMoEBS, WMoEBS2, IRMOBS, GA-SVM e IDMMoBS, pode-se observar a evolução da qualidade de classificação e da quantidade de bandas proporcionadas ao longo das gerações. Essa observação não tem significado estatístico, porém auxilia no entendimento do funcionamento desses métodos, permitindo observar a característica evolutiva e da otimização dos múltiplos objetivos.

Os gráficos das Figuras 6.1 apresentam a evolução dos resultados de OA e quantidade de bandas proporcionados pelos métodos para as diferentes imagens consideradas, onde os resultados foram obtidos a partir do primeiro conjunto (*folds*) considerado nos experimentos. O resultado desejado tem o comportamento de que quando a quantidade de bandas diminui, ao mesmo tempo, aumenta o índice OA.

Os valores mostrados em cada par de gráficos (OA \times quantidade de bandas) foram obtidos a partir da mesma solução selecionada por algum critério, em cada geração. Para o GA-SVM o critério utilizado para a solução foi aquele que apresentou melhor valor de *fitness*. Para o IRMOBS foi utilizada a própria estratégia proposta no algoritmo que permite selecionar uma solução da fronteira de Pareto considerando a proximidade da solução de um ponto ideal e, também, a correlação da solução ideal e melhores soluções para cada *fitness*. Para o WMoEBS, WMoEBS2 e o IDMMoBS foram utilizados o DM proposto. Os gráficos 6.1a, 6.1c e 6.1e repre-

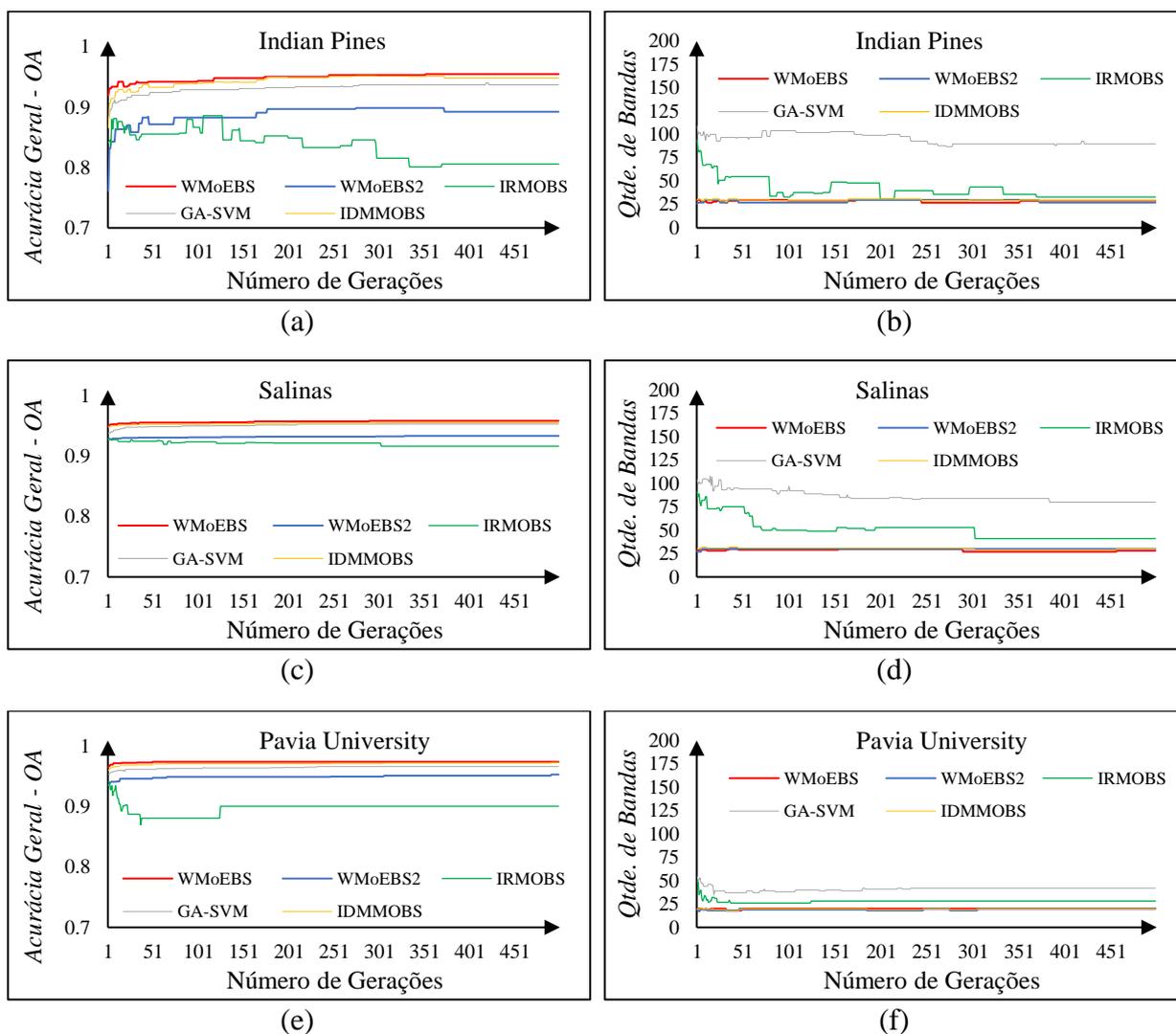


Figura 6.1. Evolução da precisão e quantidade de bandas para os métodos WMOEBS, WMOEBS2, IRMOBS, GA-SVM e IDMMOBS. (a) (c) e (e) Precisão (OA). (b), (d) e (f) Quantidade de bandas.

sentam a evolução da OA e os gráficos 6.1b, 6.1d e 6.1f representam a evolução da quantidade de bandas para diferentes imagens analisadas.

Conforme pode ser observado, os métodos WMOEBS, WMOEBS2, IDMMOBS e GA-SVM, todos utilizando a estratégia *Wrapper* e aprendizado supervisionado, apresentam uma evolução quase constante da OA enquanto mantém ou fazem a redução da quantidade de bandas. As evoluções para esses métodos são mais perceptíveis até próximo da geração 150 e principalmente para imagem Indian Pines, depois, têm um comportamento quase constante. O método não supervisionado acabou perdendo um pouco do desempenho conforme índice OA durante as gerações para imagens Indian Pines (6.1a) e Salinas (6.1c). Já para imagem Pavia University (6.1e) oscilou no início das gerações e depois se manteve constante. O desempenho do IRMOBS pode ser explicado por suas funções de *fitness* não são baseadas em rótulos supervisionados, mas sim baseadas na entropia média e variância entre bandas.

Os métodos WMoEBS e IDMMoBS que utilizam a informação espacial, apresentam o melhor desempenho de evolução, seguido pelo GA-SVM (sem informação espacial). O bom desempenho do GA-SVM pode ser explicado por causa da tendência dos pesos de sua função de *fitness*, direcionados em maior parte para melhorar o índice de OA. Em contrapartida, o método GA-SVM não apresentou uma boa redução na quantidade de bandas quando comparado com os outros métodos, mesmo quando considerado o IRMOBS. Para a redução da quantidade de bandas, o WMoEBS e WMoEBS2 proporcionaram os melhores valores ao longo das gerações.

Com base nessas evoluções, os métodos que apresentaram o comportamento mais desejado foram o WMoEBS e IDMMoBS, pois eles conseguem ter um bom desempenho de classificação quando utilizada a informação espacial ao mesmo tempo em que conseguem reduzir a quantidade de bandas. O mesmo comportamento também foi exibido pelo WMoEBS2, que mesmo tendo desempenho inferior quando ao índice OA em relação ao GA-SVM, proporcionou um bom equilíbrio entre os dois índices analisados.

Uma característica observada na evolução do IDMMoBS, foi que ele teve uma tendência a encontrar soluções com quantidades de bandas muito próximas ao limiar da estratégia de reparo. Nas soluções testadas no processo de busca, esse algoritmo acabava encontrando uma boa solução, porém a maioria das soluções tinha quantidade de bandas semelhante ao limiar, sugerindo que ele estava tendo uma tendência a buscar por soluções com essa quantidade de bandas. A estratégia de reparo do WMoEBS foi proposta para diminuir esse problema.

Para uma análise dos resultados proporcionados pelas bandas selecionadas, a técnica de validação cruzada e testes estatísticos foram aplicados e os resultados são demonstrados na sequência, onde também foi possível a comparação com outros métodos além dos baseados em AE já observados nos gráficos anteriores.

6.3.2 Avaliações das amostras de testes para os métodos analisados

Com o propósito de organizar e facilitar o entendimento das comparações realizadas com o WMoEBS (com informação espacial), WMoEBS2 (sem informação espacial) e outros métodos da literatura, assim como os casos anteriores os métodos estão organizados em aprendizados supervisionados (junto com semi-supervisionados), não supervisionados e imagem original com o uso de todas as bandas.

6.3.2.1 Supervisionados

A Tabela 6.4 apresenta os valores médios de resultados de classificação obtidos a partir do *5-fold cross validation* para diferentes índices e da quantidade de bandas selecionadas propor-

cionados pelos métodos supervisionados comparados com o WMoEBS e WMoEBS2.

Tabela 6.4. Resultados da quantidade de bandas e índices de avaliação de classificação dos experimentos com métodos supervisionados implementados, WMoEBS e WMoEBS2.

Imagem	Método	OA %	σ	AA %	σ	KC	σ
Indian Pines	WMoEBS (28 bandas, $\sigma=1,7$)	89,6	0,8	84,5	4,6	0,88	0,01
	WMoEBS2 (26 bandas , $\sigma=2,8$)	85,9	2,4	82,0	2,6	0,84	0,03
	GA-SVM (93 bandas, $\sigma=9,2$)	86,2	1,4	82,1	2,7	0,84	0,02
	SLN (30 bandas)	77,3	2,1	71,2	5,8	0,74	0,02
	CART (134 bandas)	66,9	1,9	61,8	4,8	0,62	0,02
	IDMMoBS (29 bandas, $\sigma=0,8$)	91,4	1,1	91,4	3,9	0,90	0,13
Salinas	WMoEBS (26 bandas , $\sigma=0,49$)	94,9	0,3	97,8	0,1	0,94	0,00
	WMoEBS2 (27 bandas, $\sigma=1,8$)	91,5	4,0	95,5	3,0	0,91	0,04
	GA-SVM (92 bandas, $\sigma=4,6$)	92,2	3,3	95,9	2,3	0,91	0,04
	SLN (30 bandas)	92,5	0,8	96,2	0,4	0,92	0,01
	CART (173 bandas)	86,4	1,0	90,4	2,0	0,85	0,01
	IDMMoBS (29 bandas, $\sigma=0,75$)	94,4	0,4	97,3	0,1	0,94	0,04
Pavia U.	WMoEBS (17 bandas, $\sigma=1,2$)	96,6	0,2	95,8	0,7	0,95	0,00
	WMoEBS2 (20 bandas, $\sigma=0,4$)	94,4	0,4	93,3	0,9	0,93	0,01
	GA-SVM (48 bandas, $\sigma=2,1$)	94,7	0,4	93,5	0,5	0,93	0,01
	SLN (20 bandas)	88,7	1,8	87,4	2,3	0,85	0,02
	CART (99 bandas)	86,9	0,5	84,3	1,2	0,82	0,01
	IDMMoBS (15 bandas , $\sigma=2,33$)	96,2	0,3	95,5	0,5	0,95	0,04

O WMoEBS e IDMMoBS que utilizam a informação espacial proporcionaram os melhores resultados quanto aos índices OA, AA e KC em relação a todos os outros métodos. O WMoEBS ainda foi superior ao IDMMoBS em dois casos, sendo inferior em um terceiro. Porém, para comprovações, uma análise do desempenho da classificação proporcionada pelas bandas selecionadas por cada método, foram analisados os valores chamados "*p-values*" obtidos pelo teste T-pareado calculado com base no *5-fold cross-validation* mostrados nas Tabelas 6.5 e 6.6. Os valores apresentados nessas tabelas são respectivamente para os índices OA e AA e *p-values* não foram obtidos para o índice KC, pois esse índice é altamente correlacionada com a OA.

Nestes experimentos, para o teste T-pareado foi considerado o limiar de 0,05, portanto, quando o *p-value* for superior a esse valor deve-se aceitar a hipótese nula, onde considera-se que os valores obtidos têm médias iguais, ou do contrário, significa que os valores médios são diferentes um ao outro. Os valores em negrito das Tabelas 6.5 e 6.6 indicam que o método da linha não proporcionou média igual ao método da coluna e foi maior no resultado do respectivo índice mostrado na Tabela 6.4.

Tabela 6.5. Resultados dos *p-values* do teste T-pareado para o índice OA em métodos supervisionados, WMoEBS e WMoEBS2.

Indian Pines						
Método	WMoEBS	WMoEBS2	GA-SVM	SLN	CART	IDMMoBS
WMoEBS	-	0,04	0,0	0,0	0,0	0,06
WMoEBS2	0,04	-	0,87	0,0	0,0	0,03
GA-SVM	0,0	0,87	-	0,0	0,0	0,01
SLN	0,0	0,0	0,0	-	0,0	0,0
CART	0,0	0,0	0,0	0,0	-	0,0
IDMMoBS	0,06	0,03	0,01	0,0	0,0	-
Salinas						
Método	WMoEBS	WMoEBS2	GA-SVM	SLN	CART	IDMMoBS
WMoEBS	-	0,16	0,16	0,0	0,0	0,20
WMoEBS2	0,16	-	0,21	0,57	0,03	0,25
GA-SVM	0,16	0,21	-	0,77	0,01	0,27
SLN	0,0	0,57	0,77	-	0,0	0,02
CART	0,0	0,03	0,01	0,0	-	0,0
IDMMoBS	0,20	0,25	0,27	0,02	0,0	-
Pavia University						
Método	WMoEBS	WMoEBS2	GA-SVM	SLN	CART	IDMMoBS
WMoEBS	-	0,0	0,0	0,0	0,0	0,14
WMoEBS2	0,0	-	0,26	0,0	0,0	0,01
GA-SVM	0,0	0,26	-	0,0	0,0	0,01
SLN	0,0	0,0	0,0	-	0,14	0,0
CART	0,0	0,0	0,0	0,14	-	0,0
IDMMoBS	0,14	0,01	0,01	0,0	0,0	-

6.3.2.2 Análise dos Resultados do Grupo de Experimentos 2: Supervisionados

Comparando-se os métodos WMoEBS e o WMoEBS2 com os métodos supervisionados GA-SVM, SLN, CART e IDMMoBS em cada um dos índices foi observado que:

- **Índice KC:** Os WMoEBS, WMoEBS2, GA-SVM, SLN, CART e IDMMoBS proporcionaram bom desempenho para imagens Salinas e Pavia University, porém essa característica não se repete para o SLN e o CART, quando considerado a imagem Indian Pines.
- **Índices OA e AA:** O WMoEBS proporcionou um desempenho superior que os algoritmos SLN e CART para todos os casos em ambos os índices. Para imagem Salinas, o WMoEBS, o GA-SVM, IDMMoBS e o WMoEBS2 proporcionaram um desempenho equivalente para ambos os índices, pois observa-se que o *p-value* apresentado nestes casos, foi maior que 0,05, indicando que os resultados médios de classificação são estatisticamente iguais. O WMoEBS e o IDMMoBS apresentaram médias iguais para quase todos os casos e somente para a imagem Salinas e para o índice AA, o WMoEBS foi um

Tabela 6.6. Resultados dos *p-values* do teste T-pareado para o índice AA em métodos supervisionados, WMoEBS e WMoEBS2.

Indian Pines						
Método	WMoEBS	WMoEBS2	GA-SVM	SLN	CART	IDMMoBS
WMoEBS	-	0,25	0,38	0,0	0,0	0,06
WMoEBS2	0,25	-	0,96	0,01	0,0	0,03
GA-SVM	0,38	0,96	-	0,01	0,0	0,0
SLN	0,0	0,01	0,01	-	0,0	0,0
CART	0,0	0,0	0,0	0,0	-	0,0
IDMMoBS	0,06	0,03	0,0	0,0	0,0	-
Salinas						
Método	WMoEBS	WMoEBS2	GA-SVM	SLN	CART	IDMMoBS
WMoEBS	-	0,20	0,19	0,0	0,0	0,0
WMoEBS2	0,20	-	0,26	0,63	0,0	0,28
GA-SVM	0,19	0,26	-	0,84	0,0	0,29
SLN	0,0	0,63	0,84	-	0,0	0,0
CART	0,0	0,0	0,0	0,0	-	0,0
IDMMoBS	0,0	0,28	0,29	0,0	0,0	-
Pavia University						
Método	WMoEBS	WMoEBS2	GA-SVM	SLN	CART	IDMMoBS
WMoEBS	-	0,0	0,0	0,0	0,0	0,63
WMoEBS2	0,0	-	0,65	0,01	0,0	0,04
GA-SVM	0,0	0,65	-	0,0	0,0	0,02
SLN	0,0	0,01	0,0	-	0,0	0,0
CART	0,0	0,0	0,0	0,0	-	0,0
IDMMoBS	0,63	0,04	0,02	0,0	0,0	-

pouco melhor. Esses métodos também foram superiores em relação a todos outros para as imagens Indian Pines e Salinas para ambas os índices, incluindo o WMoEBS2 (sem informação espacial).

- **Quantidade de bandas:** O bom potencial do WMoEBS pode também ser observado quando considerada a quantidade de bandas. Nesse índice quando comparados ao WMoEBS, o WMoEBS2 mostrou um desempenho superior para imagem Indian Pines e o IDMMoBS para imagens Pavia University, porém os resultados proporcionados pelo método proposto foram bastante próximos a ambos os casos. O SLN teve uma quantidade de bandas pré-definida e o CART buscou livremente por uma quantidade adequada. Para ambos os casos a quantidade média de bandas obtidas e utilizadas foi superior em relação ao WMoEBS e WMoEBS2.

Síntese dos Resultados do Grupo de Experimentos 2 – Supervisionados: O GA-SVM apresentou os resultados de classificação mais próximos ao método proposto com informação espacial e também o IDMMoBS, sendo que em alguns casos proporcionou um desempenho igual a

ambos. O SLN e o CART não proporcionaram bons desempenhos de classificação e o CART também não para a quantidade de bandas. Considerando o WMoEBS2 e o GA-SVM, observa-se que para os casos testados, a não utilização da informação espacial no método proposto em comparação com um método direcionado quase que totalmente para melhoria da precisão ocasionou mesmo desempenho médio de classificação. Isso mostra o bom desempenho do método proposto nos critérios de classificação, mesmo quando a informação espacial não é utilizada. Além disso, com relação ao GA-SVM, este método tem o problema de não conseguir realizar adequadamente uma redução da quantidade de bandas, o que é perceptível tanto quando comparado com o WMoEBS, WMoEBS2 ou demais métodos comparados. A justificativa para o desempenho ruim do GA-SVM no critério de redução da quantidade de bandas ocorre por causa da natureza da função de *fitness* com média ponderada e os pesos utilizados em cada índice. Esses pesos podem causar uma tendência no método em buscar por uma elevada OA na classificação ignorando uma propriedade de redução da quantidade de bandas. Essa característica demonstra a dificuldade dos métodos de média ponderada em lidar com problemas onde múltiplos objetivos devem ser melhorados, situação que não ocorre no caso do WMoEBS ou IDMMoBS e destacando a boa capacidade do método proposto. O WMoEBS quando comparado com sua versão prévia IDMMoBS apresentou resultados idênticos para maioria dos casos, sendo vitorioso em apenas um caso, porém em relação a "estratégia de reparo" utilizada no WMoEBS possui um custo computacional menor, o que pode ser uma vantagem no processo de seleção de bandas.

6.3.2.3 Não Supervisionados

Na Tabela 6.7 são apresentados os valores médios da quantidade de bandas selecionadas e dos índices de classificação nas mesmas condições anteriores, porém para os métodos não supervisionados comparados com o WMoEBS e WMoEBS2. Também são mostrados os resultados dos *p-values* para OA e AA respectivamente nas Tabelas 6.8 e 6.9.

6.3.2.4 Análise dos Resultados do Grupo de Experimentos 2: Não Supervisionados

- **Índice KC:** Nesta comparação o método não supervisionado IRMOBS conseguiu proporcionar um bom desempenho de classificação para o índice KC para todas imagens, assim como os já apresentados WMoEBS e WMoEBS2. Já o método WaLuMi não conseguiu proporcionar um bom resultado para esse índice quando aplicado na imagem Indian Pines.
- **Índices OA e AA:** Para o índice OA e assim como no caso anterior, o WMoEBS propor-

Tabela 6.7. Resultados da quantidade de bandas e índices de classificação dos experimentos com métodos não supervisionados implementados, WMoEBS e WMoEBS2,

Imagem	Método	OA %	σ	AA %	σ	KC	σ
Indian Pines	WMoEBS (28 bandas, $\sigma=1,7$)	89,6	0,8	84,5	4,6	0,88	0,01
	WMoEBS2 (26 bandas, $\sigma=2,8$)	85,9	2,4	82,0	2,6	0,84	0,03
	IRMOBS (37 bandas, $\sigma=2,6$)	82,9	1,4	78,5	1,2	0,81	0,02
	WaLuMi (30 bandas)	69,8	2,2	62,7	4,2	0,65	0,03
Salinas	WMoEBS (26 bandas, $\sigma=0,49$)	94,9	0,3	97,8	0,1	0,94	0,00
	WMoEBS2 (27 bandas, $\sigma=1,8$)	91,5	4,0	95,5	3,0	0,91	0,04
	IRMOBS (43 bandas, $\sigma=9,6$)	93,1	0,4	96,7	0,3	0,92	0,01
	WaLuMi (30 bandas)	87,4	0,5	89,1	0,4	0,86	0,01
Pavia U.	WMoEBS (17 bandas, $\sigma=1,2$)	96,6	0,2	95,8	0,7	0,95	0,00
	WMoEBS2 (20 bandas, $\sigma=0,4$)	94,4	0,4	93,3	0,9	0,93	0,01
	IRMOBS (27 bandas, $\sigma=2,4$)	90,5	1,1	89,1	1,5	0,87	0,02
	WaLuMi (20 bandas)	91,2	0,9	90,5	1,1	0,88	0,01

Tabela 6.8. Resultados dos *p-values* do teste T-pareado para o índice OA em métodos não supervisionados, WMoEBS e WMoEBS2,

Indian Pines				
Método	WMoEBS	WMoEBS2	IRMOBS	WaLuMi
WMoEBS	-	0,04	0,0	0,0
WMoEBS2	0,04	-	0,01	0,0
IRMOBS	0,0	0,01	-	0,0
WaLuMi	0,0	0,0	0,0	-
Salinas				
Método	WMoEBS	WMoEBS2	IRMOBS	WaLuMi
WMoEBS	-	0,16	0,0	0,06
WMoEBS2	0,16	-	0,51	0,09
IRMOBS	0,0	0,51	-	0,0
WaLuMi	0,06	0,09	0,0	-
Pavia University				
Método	WMoEBS	WMoEBS2	IRMOBS	WaLuMi
WMoEBS	-	0,0	0,0	0,0
WMoEBS2	0,0	-	0,0	0,0
IRMOBS	0,0	0,0	-	0,51
WaLuMi	0,0	0,0	0,51	-

cionou os melhores resultados de classificação, seguido pelo WMoEBS2 que foi melhor que os métodos não supervisionados em duas imagens. O WMoEBS2 teve resultados com médias iguais ao WMoEBS apenas para imagem Salinas, sendo inferior nas outras duas imagens. Para o índice AA, o WMoEBS e WMoEBS2 proporcionaram resultados médios iguais em dois casos, imagens Indian Pines e Salinas e na outra imagem o método com informação espacial foi melhor. Para imagem Salinas, o método não supervisionado IRMOBS e o WMoEBS2 proporcionaram resultados com médias iguais de OA e AA e,

Tabela 6.9. Resultados dos *p-values* do teste T-pareado para o índice AA em métodos não supervisionados, WMoEBS e WMoEBS2,

Indian Pines				
Método	WMoEBS	WMoEBS2	IRMOBS	WaLuMi
WMoEBS	-	0,25	0,07	0,0
WMoEBS2	0,25	-	0,0	0,06
IRMOBS	0,07	0,0	-	0,07
WaLuMi	0,0	0,06	0,07	-
Salinas				
Método	WMoEBS	WMoEBS2	IRMOBS	WaLuMi
WMoEBS	-	0,20	0,0	0,0
WMoEBS2	0,20	-	0,47	0,01
IRMOBS	0,0	0,47	-	0,0
WaLuMi	0,0	0,01	0,0	-
Pavia University				
Método	WMoEBS	WMoEBS2	IRMOBS	WaLuMi
WMoEBS	-	0,0	0,0	0,0
WMoEBS2	0,0	-	0,02	0,0
IRMOBS	0,0	0,02	-	0,25
WaLuMi	0,0	0,0	0,25	-

portanto, o IRMOBS pode ser uma boa alternativa para casos onde amostras rotuladas não estão disponíveis. O WaLuMi não conseguiu um bom desempenho para maioria dos casos, e apenas apresenta médias iguais ao IRMOBS para imagem Pavia University.

- **Quantidade de bandas:** Quando considerada a quantidade de bandas, o WMoEBS e WMoEBS2 proporcionaram os melhores resultados. Sendo o WMoEBS2, que não considera o uso da informação espacial, ainda melhor que o WMoEBS. Ambos os métodos conseguem reduzir a quantidade de bandas de forma inferior ao limiar estabelecido e manter uma boa qualidade de classificação.

Síntese dos Resultados do Grupo de Experimentos 2 – Não Supervisionados: Quando os métodos WMoEBS e WMoEBS2 foram comparados com os métodos não supervisionados considerados, apresentaram simultaneamente melhores resultados de classificação para a maioria dos casos e uma quantidade reduzida de bandas. Esse comportamento é esperado para os métodos que utilizam estratégia *Wrapper* quando comparado com métodos baseados na estratégia *Filter*. O IRMOBS, mesmo sendo *Filter* e não supervisionado, apresentou bons resultados de classificação podendo ser uma alternativa para quando as amostras rotuladas não estão disponíveis. Uma deficiência observada no IRMOBS é a falta de um mecanismo para evitar bandas redundantes e isso foi explorado na geração da população inicial do WMoEBS e WMoEBS2. Como o WaLuMi apresentou um desempenho inferior ao WMoEBS e WMoEBS2, fica caracte-

rístico a capacidade do WMoEBS e WMoEBS2 de melhorar a seleção de bandas após a seleção de outro método, pois o WaLuMi foi uma estratégia base para iniciar as soluções exploradas pelo método proposto.

6.4 Grupo de experimentos 3: IHs com todas as bandas e bandas selecionadas pelo WMoEBS

Na Tabela 6.10 são apresentados os resultados de classificação proporcionados pelo WMoEBS (e WMoEBS2) e IHs com todas as bandas sem informação espacial. Essas comparações têm o propósito de demonstrar a qualidade da classificação proporcionada pelas bandas selecionadas pelo WMoEBS quanto às IHs com todas as bandas. Para as IHs com todas as bandas, o mesmo classificador foi utilizado, ou seja, o SVM-RBF.

Tabela 6.10. Resultados da quantidade de bandas e índices de classificação dos experimentos com a utilização de todas as bandas, WMoEBS e WMoEBS2.

Imagem	Método	OA %	σ	AA %	σ	KC	σ
Indian Pines	WMoEBS (28 bandas, $\sigma=1,7$)	89,6	0,8	84,5	4,6	0,88	0,01
	WMoEBS2 (26 bandas, $\sigma=2,8$)	85,9	2,4	82,0	2,6	0,84	0,03
	Original (200 bandas)	86,4	1,4	80,2	3,3	0,84	0,02
Salinas	WMoEBS (26 bandas, $\sigma=0,49$)	94,9	0,3	97,8	0,1	0,94	0,00
	WMoEBS2 (27 bandas, $\sigma=1,8$)	91,5	4,0	95,5	3,0	0,91	0,04
	Original (204 bandas)	91,4	2,8	94,1	2,0	0,90	0,03
Pavia U.	WMoEBS (17 bandas, $\sigma=1,2$)	96,6	0,2	95,8	0,7	0,95	0,00
	WMoEBS2 (20 bandas, $\sigma=0,4$)	94,4	0,4	93,3	0,9	0,93	0,01
	Original (103 bandas)	95,1	0,5	93,6	0,9	0,93	0,01

6.4.1 Análise dos Resultados do Grupo de Experimentos 3

- **Índice KC:** Com relação ao índice KC, todos os casos proporcionaram bons resultados.
- **Índices OA e AA:** Para OA, o WMoEBS proporcionou a maioria dos melhores resultados em relação ao uso da IH, e em casos onde não foi melhor, o teste T-Pareado indicou empate entre os dois casos. Em relação ao WMoEBS2 e a utilização de todas as bandas teve desempenho de classificação médio iguais para imagens Indian Pines e Salinas. O WMoEBS2 teve desempenho inferior da OA apenas para imagem Pavia University, sugerindo que o método proposto proporcionou um bom desempenho de classificação para esse índice mesmo quando a informação espacial não foi utilizada. Para métrica AA e a imagem Indian Pines ocorreram empates entre todos os casos e para imagens Salinas e Pavia University, o WMoEBS foi melhor que o uso de todas as bandas. Os resultados indicam

Tabela 6.11. Resultados dos *p-values* do teste T-pareado para o índice OA com a utilização de todas as bandas, WMoEBS e WMoEBS2.

Indian Pines			
Método	WMoEBS	WMoEBS2	Original
WMoEBS	-	0,04	0,01
WMoEBS2	0,04	-	0,59
Original	0,01	0,59	-
Salinas			
Método	WMoEBS	WMoEBS2	Original
WMoEBS	-	0,16	0,06
WMoEBS2	0,16	-	0,86
Original	0,06	0,86	-
Pavia University			
Método	WMoEBS	WMoEBS2	Original
WMoEBS	-	0,0	0,0
WMoEBS2	0,0	-	0,0
Original	0,0	0,0	-

Tabela 6.12. Resultados dos *p-values* do teste T-pareado para o índice AA com a utilização de todas as bandas, WMoEBS e WMoEBS2.

Indian Pines			
Método	WMoEBS	WMoEBS2	Original
WMoEBS	-	0,25	0,07
WMoEBS2	0,25	-	0,06
Original	0,07	0,06	-
Salinas			
Método	WMoEBS	WMoEBS2	Original
WMoEBS	-	0,20	0,02
WMoEBS2	0,20	-	0,18
Original	0,02	0,18	-
Pavia University			
Método	WMoEBS	WMoEBS2	Original
WMoEBS	-	0,0	0,0
WMoEBS2	0,0	-	0,39
Original	0,0	0,39	-

que a utilização da informação espacial no WMoEBS pode proporcionar resultados melhores que a utilização de todas as bandas da imagem original e quando essa informação não é considerada, o WMoEBS2 apresentou resultados competitivos com uma quantidade reduzida de bandas.

- **Quantidade de bandas:** Com relação a quantidade de bandas, a redução realizada pelo WMoEBS, foi de 86% para imagem Indian Pines, 87,25% para imagem Salinas e 83,5%

para imagem Pavia University. Já redução realizada pelo WMoEBS2, foi de 87% para imagem Indian Pines, 86,8% para imagem Salinas e 80,6% para imagem Pavia University. Conforme previamente discutido, esses resultados de redução foram melhores que os métodos comparados o que demonstra uma boa capacidade para o critério de redução de bandas em relação a quantidade original de bandas.

6.5 Considerações Finais do Capítulo

Neste capítulo foram apresentadas diversas comparações de métodos da literatura com o WMoEBS e WMoEBS2 (sem informação espacial) propostos nesta tese. Especialmente, o método IDMMoBS também foi comparado com os métodos propostos, sendo ele uma versão anterior ao WMoEBS. Diferentes grupos de experimentos foram organizados, sendo alguns desses replicados e outros tiveram algoritmos implementados. Os diferentes métodos comparados utilizam estratégias de aprendizados supervisionados (ou semi-supervisionados) ou não supervisionados, além de serem baseados em diferentes estratégias (*Wrapper*, *Filter* ou *Embedded*). Com os resultados e discussões apresentados, observa-se que o WMoEBS foi superior nos critérios de classificação para maioria dos casos comparados, considerando os métodos da literatura, uso de todas as bandas e até mesmo o IDMMoBS. O método também foi capaz de reduzir a quantidade de bandas, escolhendo aquelas que foram mais adequadas para promover uma boa classificação.

Mesmo quando a informação espacial não foi utilizada, como mostrado no WMoEBS2, o desempenho proporcionado pela classificação na maioria dos casos foi superior ou idêntico em relação a muitos dos outros métodos. Para alguns casos, o WMoEBS2 quando comparado com o tradicional método GA-SVM com uma função de *fitness* que considera média ponderada, proporcionou desempenhos de classificação médio idênticos, porém vantagens em relação à quantidade de bandas. As características do WMoEBS2 demonstram que o método pode também ser utilizado em qualquer contexto de seleção de características e não somente para bandas espectrais.

Baseado nesses resultados, ficou evidente que o WMoEBS é adequado para seleção de bandas reduzindo a quantidade e promovendo boa classificação, principalmente quando considera a utilização da informação espacial. Portanto, pode ser aplicado no contexto de seleção de bandas de IHS não somente de áreas agrícolas, mas também de outras, auxiliando na automatização de processos de monitoramento e reconhecimento de objetos.

Capítulo 7

CONSIDERAÇÕES FINAIS, CONCLUSÕES E TRABALHOS FUTUROS

Nesta tese foram contextualizadas as Imagens hiperespectrais (IHs), que contém diversas e específicas informações sobre cenas que são capturadas por sensoriamento remoto (SR) que, posteriormente, são analisadas. Essas IHs proporcionam novas aplicações em diferentes áreas e, em especial, na agricultura podem trazer contribuições que melhoram o manejo de culturas, economizando recursos e auxiliando na produção da demanda global de alimentos.

A composição de IHs é resultado de suas numerosas bandas espectrais que abrangem uma longa e contínua faixa do espectro eletromagnético que vai além de ondas (cores) visíveis. Materiais diferentes têm valores de refletância distintos ao longo dessa faixa, caracterizando as chamadas assinaturas espectrais. Objetos com características semelhantes em termos de cores e formas podem também ser distintos no domínio espectral utilizando o comportamento da assinatura espectral.

Conforme previamente apresentado, em um sistema de reconhecimento de padrões (SRP), a enorme quantidade de dados contidos nas IHs pode causar diversos problemas, tais como: dificuldade em armazenamento e transmissão, baixo desempenho na qualidade de classificação e segmentação, entre outros. Uma maneira de diminuir a ocorrência ou impacto desses problemas é por meio da redução da dimensionalidade, onde no contexto de IHs pode ser realizada por estratégias de extração de características ou seleção de bandas. De interesse desta tese, a seleção de bandas tenta reduzir a quantidade de dados mantendo as informações úteis e em formato original.

Em métodos que utilizam o aprendizado supervisionado, é possível realizar a seleção das bandas levando em consideração a disponibilidade da informação de classe para os dados de amostra. Com essa característica, por exemplo, medidas de similaridade em amostras de classes conhecidas podem ser utilizadas para avaliar a qualidade de bandas espectrais permitindo

estruturar métodos de seleção supervisionados. Os próprios classificadores podem ser incorporados no processo de seleção e ajustados durante a busca do conjunto de bandas mais adequado formando métodos baseados na estratégia *Wrapper*. Outra categoria de métodos que podem ser explorados na tarefa de seleção de bandas, são baseados em algoritmos multiobjetivos, que permitem conduzir a busca por soluções que apresentam um equilíbrio entre diferentes funções de *fitness* de interesse, que pode ser baseados em índices como a quantidade de bandas e a qualidade de classificação.

Para a tarefa de seleção de bandas multiobjetivo decorrente da estratégia *Wrapper*, nesta tese, foi elaborado um novo método que incorpora a informação espacial ao processo de seleção de bandas e que, posteriormente, também pode ser usada na classificação de IHS com as bandas selecionadas. O método proposto é denominado *Wrapper Multiobjective Evolutionary Band Selection* (WMoEBS) que combina características específicas de métodos da literatura que apresentaram resultados satisfatórios em seus estudos na tarefa de seleção de bandas. A composição do método, junto a diferentes comparações e avaliações, foram realizadas nesta tese e, na sequência, são apresentadas as principais conclusões.

7.1 **Elaboração do WMoEBS**

A constituição do WMoEBS foi realizada por meio de diversos componentes (que são algoritmos, métodos e configurações), sendo alguns explorados previamente por outros autores e outros criados exclusivamente para serem utilizados no WMoEBS. Esses componentes foram selecionados com base em resultados e justificativas já apresentadas na literatura além de também avaliações realizadas nesta tese. Com base nas estruturas estudadas, foi pensado e proposto um *framework* para seleção de bandas multiobjetivo apresentado no capítulo de materiais e métodos. Para selecionar os componentes que foram utilizados no *framework* proposto e compor o WMoEBS, diversos testes e avaliações foram realizados. Os melhores componentes avaliados foram os escolhidos para compor o WMoEBS.

De característica multiobjetivo, o WMoEBS é baseado no *Multiobjective Evolutionary Algorithm Based on Decomposition* (MOEA/D) e seleciona conjuntos de bandas com base em duas funções de *fitness*. Uma dessas funções é baseada no valor médio do *3-fold cross-validation* do índice de precisão/*Overall Accuracy* (OA) e a outra função é baseada em um índice de quantidade de bandas. Por ser multiobjetivo, o WMoEBS consegue lidar com essas duas funções de *fitness* simultaneamente, mesmo em situações onde elas podem ser conflitantes, ou seja, quando uma afeta negativamente à outra. Uma estratégia baseada em *Decision-Maker* (DM) foi incorporada ao WMoEBS permitindo que uma única solução da fronteira de Pareto

fosse escolhida. O algoritmo *Support Vector Machine* (SVM) com o *kernel radial basis function* (RBF) foi o classificador incorporado ao processo de seleção de bandas caracterizando o método proposto como estratégia *Wrapper*. Uma estratégia de inicialização do conjunto de soluções iniciais, que são utilizadas na condução do processo de busca, foi baseada no algoritmo *Ward's Linkage strategy Using Mutual Information* (WaLuMi). Esse processo evita uma elevada quantidade de bandas correlacionadas no início da busca do WMoEBS.

Durante a escolha desses componentes do WMoEBS e, também, nos experimentos comparativos com outros métodos considerados, foi mostrado que a utilização da informação espacial, obtida por um simples filtro de média, proporcionou melhores resultados do que a sua não utilização. Com essas avaliações foi possível atingir uma das hipóteses desta tese:

- A utilização da informação espacial em conjunto com o processo de seleção de bandas, pode promover melhores resultados que a utilização somente da informação espectral na tarefa de classificação dos pixels com bandas selecionadas.

Além dessa hipótese, a própria elaboração do *framework* e do WMoEBS permitiram alcançar os alguns dos objetivos específicos elencados no capítulo inicial.

7.2 Seleção de bandas com o WMoEBS e classificação de imagens hiperespectrais

Avaliações da quantidade de bandas e qualidade da classificação usando os índices OA, Acurácia Média/*Average Accuracy* (AA) e coeficiente de Kappa/*Kappa Coefficient* (KC) foram realizadas com base em um conjunto de dados de testes gerado a partir de IHS. Para evitar problemas de *overfitting*, esses dados de testes foram separados do conjunto de treinamento usado na seleção de bandas e construção do modelo. Resultados dos diferentes índices foram proporcionados por bandas selecionadas pelo WMoEBS, outros métodos distintos da literatura e a utilização de todas as bandas (quantidade de bandas original). Amostras com as bandas selecionadas pelos diferentes métodos foram comparadas usando a classificação pelo SVM em três grupos de experimentos que foram organizados.

O primeiro grupo de experimentos desta tese foi a replicação dos estudos de Xie et al. (2018), permitindo comparações diretas de resultados de diferentes métodos da literatura com o WMoEBS (com ou sem a informação espacial). Os métodos comparados com o WMoEBS foram da categoria de aprendizados supervisionados/semi-supervisionados ou não supervisionados. Nessas comparações, três índices de classificação diferentes e a quantidade de bandas foram analisadas. Para todos os casos no critério de classificação, mesmo quando a informa-

ção espacial não foi utilizada, o WMoEBS proporcionou os melhores resultados. O método estado-da-arte *Information Gain-Gray Wolf Optimizer* (IG-GWO) foi o que mais se aproximou dos resultados de classificação proporcionados pelo WMoEBS, porém não foram iguais nem superiores. Com relação a quantidade de bandas, o IG-GWO também teve resultados bastante próximos aos proporcionados pelo WMoEBS.

O segundo grupo de experimentos considerou a implementação de métodos da literatura, permitindo a aplicação de estratégias de avaliações estatísticas e, para alguns métodos, observar a evolução do índice de OA e quantidade de bandas ao longo de iterações. A capacidade de lidar simultaneamente com mais de um objetivo foi perceptível nos gráficos das funções de *fitness*, resultados apresentados da OA e quantidade de bandas demonstrados no capítulo anterior e quando comparados com outros métodos baseados em algoritmos evolucionários. Para esses casos, quando observado a evolução da OA, o WMoEBS com a utilização da informação espacial apresentou valores maiores e com uma pequena evolução ao longo de todas as gerações quando comparado com os outros casos. Os resultados foram similares a outro método gerado durante a elaboração desta tese, o chamado *Incorporated Decision-Marker-based multiobjective band selection* (IDMMoBS) (SAQUI et al., 2019c). Para OA, o método que apresentou os melhores resultados sem a utilização da informação espacial é estruturado com base em algoritmos genéticos e SVM (GA-SVM), porém com o índice de quantidade de bandas sua evolução não foi tão eficaz.

Em relação às comparações realizadas com todos os métodos do segundo grupo de experimentos (além dos métodos baseados em algoritmos evolucionários), o índice KC e testes estatísticos para OA e AA demonstraram que o WMoEBS apresentou os melhores resultados de classificação para maioria dos casos considerados. O WMoEBS sem a informação espacial, em comparação ao GA-SVM, também apresentou bons resultados, observando-se que ambos promoveram resultados médios de classificação para os índices OA e AA idênticos. O bom desempenho do WMoEBS é também apresentado para quantidade de bandas que, em relação ao GA-SVM, é muito superior. Os resultados dos dois grupos de experimentos atenderam uma das hipóteses desta tese:

- As bandas selecionadas pelo WMoEBS melhoram o desempenho da classificação de pixels de IHS quando comparado com bandas selecionadas por outros métodos de seleção que são estado-da-arte na literatura e mantém uma quantidade reduzida de bandas.

Sobre a melhoria dos resultados de classificação com a utilização de uma quantidade de bandas reduzida em relação ao total original de bandas da IHS sem a informação espacial, em um terceiro grupo de experimentos, observou-se que o WMoEBS, sem a utilização da informação

espacial, proporcionou resultados médios para OA e AA similares ao uso de todas as bandas quando classificado pelo SVM e, em ambos os casos, um bom valor de KC foi observado. Nas comparações realizadas e considerando análises estatísticas, somente para um caso com o índice OA o uso de todas as bandas promoveu resultados melhores que quando as bandas foram selecionadas pelo WMoEBS e, em contrapartida, para um caso com o índice AA, o WMoEBS foi aquele que proporcionou melhores resultados. Portanto, para o WMoEBS sem a utilização da informação espacial, ocorreu um empate nos resultados de classificação, o que demonstra a boa capacidade do WMoEBS mesmo quando a informação espacial não é utilizada. Já quando a informação espacial é utilizada, o WMoEBS apresentou melhores resultados para todos os índices e IHS comparadas quando elas utilizavam todas as bandas e sem informação espacial. Esses resultados atendem parcialmente a uma das hipóteses desta tese, pois o desempenho do WMoEBS é melhor quando a informação espacial é utilizada. Essa hipótese afirma que:

- O WMoEBS permite reduzir a dimensionalidade dos dados de IHS e as bandas selecionadas melhoram o desempenho da classificação de pixels de IHS quando comparado com as imagens originais.

Com base na discussão da seleção de bandas com o WMoEBS pode-se observar, que para todos objetivos, tanto gerais como específicos, hipóteses e contribuições apresentadas no capítulo de introdução foram alcançados durante o desenvolvimento até a conclusão desta tese. Isso indica que o método de seleção de bandas pode ser aplicado no contexto de seleção de IHS para agricultura e também outras áreas. Outra característica importante observada é que o WMoEBS sem a utilização da informação espacial, chamado de WMoEBS2 nos experimentos, pode operar como uma estratégia de seleção de características em qualquer contexto, não se limitando a IHS.

7.3 Limitações do WMoEBS e dificuldades encontradas

A principal limitação observada no WMoEBS é o tempo necessário para selecionar as bandas uma vez que o método de seleção incorpora o classificador ao processo de busca, necessitando das etapas de treinamento e classificação, as quais o tempo e qualidade são influenciados pela quantidade de pixels e bandas. Nos testes realizados, o método proposto chegou a utilizar uma semana para encontrar combinações de bandas em um computador pessoal com 8GB de memória e processador Intel i5, porém avaliações específicas e formais necessitam ser realizadas para o critério de tempo computacional. Mesmo com essa limitação, após as bandas serem selecionadas, o classificador desejado pode ser aplicado diretamente, com seu tempo de processamento reduzido em relação à utilização de todas as bandas. Portanto, se o tempo de seleção

de bandas for um fator relevante o método pode não ser indicado. Porém, se o importante for o resultado de quais bandas foram selecionadas para posterior utilização das mesmas, o método se demonstrou eficiente.

A maior dificuldade encontrada durante esta pesquisa foi encontrar conjuntos de dados/IHs com *Ground Truth* (GT) (rótulos/classes), pois bases gratuitas com esses GTs são bastante limitadas. Outra dificuldade, foi categorizar e compreender os diferentes métodos de seleção de bandas existentes.

7.4 Publicações geradas desta pesquisa

Durante esta tese, quatro publicações principais foram realizadas e são enumeradas a seguir:

1. *Methodology for Band Selection of Hyperspectral Images Using Genetic Algorithms and Gaussian Maximum Likelihood Classifier*, publicado no evento *International Conference on Computational Science and Computational Intelligence (CSCI)* em 2016, sem Qualis área de Ciência da Computação da avaliação da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) apresentada no documento Capes QUALIS CONFERÊNCIAS - CIÊNCIA DA COMPUTAÇÃO - 2016 vigente até conclusão deste texto.
2. *Approach based on SPEA2-band selection and Random Forest classifier to generate thematic maps from hyperspectral images*, publicado no evento *The Americas Conference on Information Systems (AMCIS)* em 2019, com Qualis A2 na área de Ciência da Computação da avaliação do documento da Capes QUALIS CONFERÊNCIAS - CIÊNCIA DA COMPUTAÇÃO - 2016.
3. *NSGA2-Based Method for Band Selection for Supervised Segmentation in Hyperspectral Imaging*, publicado no evento *Systems, Man, and Cybernetics Society (SMC)* em 2019, com Qualis A2 na área de Ciência da Computação da avaliação do documento da Capes QUALIS CONFERÊNCIAS - CIÊNCIA DA COMPUTAÇÃO - 2016.
4. *Incorporated Decision-Maker-based multiobjective band selection for pixel classification of hyperspectral images* publicado na revista *Advances in Electrical and Computer Engineering* em 2019, com Qualis B1 na área de Ciência da Computação do documento da AVALIAÇÃO DE PERIÓDICOS QUADRIÊNIO 2013-2016.

O último desses trabalhos propôs um método que é uma versão prévia ao do WMoEBS se diferenciando na estratégia de reparo. Os métodos dos outros artigos foram explorados e testados durante o desenvolvimento desta tese, que motivaram os critérios de avaliação e as

escolhas dos componentes que foram avaliados para composição do método proposto. Cópias desses artigos estão disponíveis nos Apêndices.

7.5 **Trabalhos Futuros**

Em trabalhos futuros pretende-se realizar novos experimentos com a:

- adição de novas funções de *fitness* ao processo de busca das bandas como, por exemplo, algum índice baseado na informação mútua para diminuir a existência de bandas correlacionadas em soluções durante a própria seleção;
- utilização de técnicas para estimar a OA, gerando um método não supervisionado e *Wrapper*;
- utilização de outras estratégias de DM;
- exploração de outros classificadores como estratégias de *deep learning*;
- explorar o WMoEBS e adequações em classes específicas de culturas da área agrícola; e
- a aplicação de outros filtros para utilização da informação de vizinhança dos pixels.

Adicionalmente, pretende-se explorar novas estratégias multiobjetivo e, também, construir uma base online de imagens hiperespectrais e GT.

Segundo Dietterich T. (1998) o teste T-pareado tem uma probabilidade alta de apresentar o erro estatístico do Tipo I (rejeitar a hipótese nula quando ela é verdadeira). Para o método proposto nesta tese isso não é problema, pois o erro do Tipo I indicaria que o método teria um desempenho igual aos demais onde o erro ocorreu, mas ainda apresentando bom desempenho e vantagens na quantidade de bandas. De qualquer forma, novas adequações nos experimentos como a utilização de um teste chamado de 5x2cv que está menos sujeito a esse tipo de erro também serão realizadas no futuro.

REFERÊNCIAS

- ACHANTA, R.; SHAJI, A.; SMITH, K.; LUCCHI, A.; FUA, P.; SUSSTRUNK, S. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE transactions on pattern analysis and machine intelligence*, v. 34, 05 2012.
- ADAO, T.; HRUSKA, J.; PáDUA, L.; BESSA, J.; PERES, E.; MORAIS, R.; SOUSA, J. Hyperspectral imaging: A review on uav-based sensors, data processing and applications for agriculture and forestry. *Remote Sensing*, v. 2017, p. 1110, 10 2017.
- AGARWAL, A.; EL-GHAZAWI, T.; EL-ASKARY, H.; LE-MOIGNE, J. Efficient hierarchical-pca dimension reduction for hyperspectral imagery. In: *2007 IEEE International Symposium on Signal Processing and Information Technology*. [S.l.: s.n.], 2007. p. 353–356. ISSN 2162-7843.
- AKYUREK, H. A.; KOCER, B. Three-dimensional spatial-spectral filtering based feature extraction for hyperspectral image classification. *Advances in Electrical and Computer Engineering*, n. 17 (2), p. 95–102, 2017. Disponível em: <<http://dx.doi.org/10.4316/AECE.2017.02013>>.
- AMINI, S.; HOMAYOUNI, S.; SAFARI, A.; DARVISHSEFAT, A. A. Object-based classification of hyperspectral data using random forest algorithm. *Geo-spatial Information Science*, Taylor Francis, v. 21, n. 2, p. 127–138, 2018.
- ANTHONY, G.; RUTHER, H. Comparison of feature selection techniques for svm classification. In: . [S.l.: s.n.], 2007. p. 1–6.
- ARCHIBALD, R.; FANN, G. Feature selection and classification of hyperspectral images with support vector machines. *IEEE Geoscience and Remote Sensing Letters*, v. 4, n. 4, p. 674–677, Oct 2007.
- ATTAS, M.; CLOUTIS, E.; COLLINS, C.; GOLTZ, D.; MAJZELS, C.; MANSFIELD, J. R.; MANTSCH, H. H. Near-infrared spectroscopic imaging in art conservation: investigation of drawing constituents. *Journal of Cultural Heritage*, v. 4, n. 2, p. 127 – 136, 2003. ISSN 1296-2074. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1296207403000244>>.
- BAGGIO, D. *OpenCV 3.0 Computer Vision with Java*. Packt Publishing, 2015. ISBN 9781783283989. Disponível em: <<https://books.google.com.br/books?id=LFtICgAAQBAJ>>.
- BAI, J.; XIANG, S.; SHI, L.; PAN, C. Semisupervised pair-wise band selection for hyperspectral images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, v. 8, n. 6, p. 2798–2813, June 2015. ISSN 1939-1404.

- BAZI, Y.; MELGANI, F. Toward an optimal svm classification system for hyperspectral remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, v. 44, n. 11, p. 3374–3385, Nov 2006. ISSN 0196-2892.
- BENEDIKTSSON, J. A.; GHAMISI, P. *Spectral-Spatial Classification of Hyperspectral Remote Sensing Images*. Norwood, MA, USA: Artech House, Inc., 2015. ISBN 1608078124, 9781608078127.
- BOSER, B. E.; GUYON, I. M.; VAPNIK, V. N. A training algorithm for optimal margin classifiers. In: *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*. New York, NY, USA: ACM, 1992. (COLT '92), p. 144–152. ISBN 0-89791-497-X. Disponível em: <<http://doi.acm.org/10.1145/130385.130401>>.
- BRADSKI, G. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.
- BRIS, A. L.; CHEHATA, N.; BRIOTTET, X.; PAPANODITIS, N. Use intermediate results of wrapper band selection methods: A first step toward the optimization of spectral configuration for land cover classifications. In: *2014 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*. [S.l.: s.n.], 2014. p. 1–4. ISSN 2158-6276.
- BRODERSEN, K. H.; ONG, C. S.; STEPHAN, K. E.; BUHMANN, J. M. The balanced accuracy and its posterior distribution. In: *2010 20th International Conference on Pattern Recognition*. [S.l.: s.n.], 2010. p. 3121–3124. ISSN 1051-4651.
- CAO, X.; JI, B.; JI, Y.; WANG, L.; JIAO, L. Hyperspectral image classification based on filtering: A comparative study. *Journal of Applied Remote Sensing*, v. 11, p. 035007, 08 2017.
- CAO, X.; WU, B.; TAO, D.; JIAO, L. Automatic band selection using spatial-structure information and classifier-based clustering. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, v. 9, n. 9, p. 4352–4360, Sep. 2016. ISSN 1939-1404.
- CHANG, C.-I.; DU, Q.; SUN, T.-L.; ALTHOUSE, M. L. G. A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, v. 37, n. 6, p. 2631–2641, Nov 1999. ISSN 0196-2892.
- CHANG, L. S.; LI, W.; LI, Q. Guided filter-based medical hyperspectral image restoration and cell classification. *Journal of Medical Imaging and Health Informatics*, v. 8, p. 826–835, 05 2018.
- CHEN, Y.; NASRABADI, N. M.; TRAN, T. D. Hyperspectral image classification using dictionary-based sparse representation. *IEEE Transactions on Geoscience and Remote Sensing*, v. 49, n. 10, p. 3973–3985, Oct 2011. ISSN 0196-2892.
- COLLETTE, Y.; SIARRY, P. *Multiobjective Optimization. Principles and Case Studies*. [S.l.: s.n.], 2003. ISBN 3-540-40182-2.
- CRISTIANINI, N.; SHAW-TAYLOR, J. Support vector machines. In: _____. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. [S.l.]: Cambridge University Press, 2000. p. 93–124.

CUI, M.; PRASAD, S.; MAHROOGHY, M.; BRUCE, L. M.; AANSTOOS, J. Genetic algorithms and linear discriminant analysis based dimensionality reduction for remotely sensed image analysis. In: *2011 IEEE International Geoscience and Remote Sensing Symposium*. [S.l.: s.n.], 2011. p. 2373–2376. ISSN 2153-7003.

DATTA, A. K.; GHOSH, S.; A., G. Clustering based band selection for hyperspectral images. *2012 International Conference on Communications, Devices and Intelligent Systems (CODIS)*, p. 101–104, 2012.

DEB, K.; PRATAP, A.; AGARWAL, S.; MEYARIVAN, T. A. M. T. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *Evolutionary Computation, IEEE Transactions on*, v. 6, p. 182 – 197, 05 2002.

DIETTERICH T., G. Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation*, v. 10, p. 1895–1923, 1998.

DINÇ, S.; AYGÜN, R. S. Evaluation of hyperspectral image classification using random forest and fukunaga-koontz transform. In: PERNER, P. (Ed.). *Machine Learning and Data Mining in Pattern Recognition*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013. p. 234–245. ISBN 978-3-642-39712-7.

DUDA, R. O.; HART, P. E.; STORK, D. G. *Pattern Classification*. 2. ed. New York: Wiley, 2001. ISBN 978-0-471-05669-0.

FAO. *e-Agriculture in action: Drones in agriculture*. [S.l.]: Food and Agriculture Organization of the United Nations and International Telecommunication Union, 2018.

FAUVEL, M.; DECHESNE, C.; ZULLO, A.; FERRATY, F. Fast forward feature selection for the nonlinear classification of hyperspectral images. *CoRR*, abs/1501.00857, 2015. Disponível em: <<http://arxiv.org/abs/1501.00857>>.

FELZENSZWALB, P. F.; HUTTENLOCHER, D. P. Efficient graph-based image segmentation. *International Journal of Computer Vision*, v. 59, n. 2, p. 167–181, Sep 2004. ISSN 1573-1405. Disponível em: <<https://doi.org/10.1023/B:VISI.0000022288.19776.77>>.

FENG, J.; JIAO, L.; LIU, F.; SUN, T.; ZHANG, X. Unsupervised feature selection based on maximum information and minimum redundancy for hyperspectral images. *Pattern Recognition*, v. 51, p. 295 – 309, 2016. ISSN 0031-3203. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0031320315003064>>.

FENG, J.; JIAO, L. C.; ZHANG, X.; SUN, T. Hyperspectral band selection based on trivariate mutual information and clonal selection. *IEEE Transactions on Geoscience and Remote Sensing*, v. 52, n. 7, p. 4092–4105, July 2014.

FENG, S.; ITOH, Y.; PARENTE, M.; DUARTE, M. F. Hyperspectral band selection from statistical wavelet models. *IEEE Transactions on Geoscience and Remote Sensing*, v. 55, n. 4, p. 2111–2123, April 2017. ISSN 0196-2892.

FUKUNAGA, K. *Introduction to Statistical Pattern Recognition*. San Diego, CA, USA: Academic Press Professional, Inc., 1990. ISBN 0122698517.

- GAO, J.; DU, Q.; GAO, L.; SUN, X. Ant colony optimization-based supervised and unsupervised band selections for hyperspectral urban data classification. *Journal of Applied Remote Sensing*, v. 8, p. 085094, 08 2014.
- GHAMISI, P.; PLAZA, J.; CHEN, Y.; LI, J.; PLAZA, A. Advanced spectral classifiers for hyperspectral images: A review. *IEEE Geoscience and Remote Sensing Magazine*, v. 5, n. 1, p. 8–32, March 2017. ISSN 2168-6831.
- GOMEZ-CHOVA, L.; CALPE, J.; OLIVAS, E.; CAMPS-VALLS, G.; MARTÍN-GUERRERO, J.; MORENO, J. Cart-based feature selection of hyperspectral images for crop cover classification. In: . [S.l.: s.n.], 2003. v. 3, p. 589–592.
- GONG, M.; ZHANG, M.; YUAN, Y. Unsupervised band selection based on evolutionary multiobjective optimization for hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing*, v. 54, n. 1, p. 544–557, Jan 2016. ISSN 0196-2892.
- GONZALEZ, R. C.; WOODS, R. E. *Digital Image Processing*. 4nd. ed. [S.l.]: Pearson, 2018.
- GRECO, M.; ACITO, N.; CORSINI, G.; DIANI, M. Band selection for spectral signature based target detection in hyperspectral data. In: *2006 IEEE International Symposium on Geoscience and Remote Sensing*. [S.l.: s.n.], 2006. p. 3899–3902. ISSN 2153-6996.
- GROVES, P.; BAJCSY, P. Methodology for hyperspectral band and classification model selection. In: *IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data, 2003*. [S.l.: s.n.], 2003. p. 120–128.
- GUALTIERI, J. The support vector machine (svm) algorithm for supervised classification of hyperspectral remote sensing data. In: _____. [S.l.: s.n.], 2009. p. 49 – 83. ISBN 9780470748992.
- GUO, B.; GUNN, S. R.; DAMPER, R. I.; NELSON, J. D. B. Band selection for hyperspectral image classification using mutual information. *IEEE Geoscience and Remote Sensing Letters*, v. 3, n. 4, p. 522–526, Oct 2006. ISSN 1545-598X.
- GUO, Y.; YIN, X.; ZHAO, X.; YANG, D.; BAI, Y. Hyperspectral image classification with svm and guided filter. *EURASIP Journal on Wireless Communications and Networking*, v. 2019, n. 1, p. 56, Mar 2019. ISSN 1687-1499. Disponível em: <<https://doi.org/10.1186/s13638-019-1346-z>>.
- GUYON, I.; BENNETT, K.; CAWLEY, G.; ESCALANTE, H. J.; ESCALERA, S.; HO, T. K.; MACIA, N.; RAY, B.; SAEED, M.; STATNIKOV, A.; VIEGAS, E. Design of the 2015 chlearn automl challenge. In: *2015 International Joint Conference on Neural Networks (IJCNN)*. [S.l.: s.n.], 2015. p. 1–8. ISSN 2161-4407.
- HABERMANN, M. *Band selection in hyperspectral images using artificial neural networks*. Tese (Doutorado) — Université de Technologie de Compiègne, 2018.
- HABERMANN, M.; FREMONT, V.; SHIGUEMORI, E. H. Supervised band selection in hyperspectral images using single-layer neural networks. *International Journal of Remote Sensing*, Taylor Francis, v. 40, n. 10, p. 3900–3926, 2019.
- HUA, W.; GUO, T.; LIU, X. Camouflage target reconnaissance based on hyperspectral imaging technology. v. 9622, 08 2015.

- HUGHES, G. On the mean accuracy of statistical pattern recognizers. *IEEE Transactions on Information Theory*, v. 14, n. 1, p. 55–63, January 1968. ISSN 0018-9448.
- HWANG, J.-N.; LAY, S.-R.; LIPPMAN, A. Nonparametric multivariate density estimation: a comparative study. *IEEE Transactions on Signal Processing*, v. 42, n. 10, p. 2795–2810, Oct 1994. ISSN 1053-587X.
- JAIN, A. K.; DUIN, R. P. W.; MAO, J. Statistical pattern recognition: A review. 2000.
- JIA, S.; QIAN, Y.; JI, Z. Band selection for hyperspectral imagery using affinity propagation. *IET Computer Vision*, v. 3, n. 4, p. 213–222, December 2009. ISSN 1751-9632.
- JIAO, L.; FENG, J.; LIU, F.; SUN, T.; ZHANG, X. Semisupervised affinity propagation based on normalized trivariable mutual information for hyperspectral band selection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, v. 8, n. 6, p. 2760–2773, June 2015. ISSN 1939-1404.
- JIMENEZ, L. O.; LANDGREBE, D. Supervised classification in high-dimensional space: geometrical, statistical, and asymptotical properties of multivariate data. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, v. 28, n. 1, p. 39–54, Feb 1998. ISSN 1094-6977.
- JORGE, L. A.; INAMASU, R. Y. Uso de veículos aéreos não tripulados (vants) em agricultura de precisão. In: BERNARDI, A. C. C.; NAIME, J. M.; RESENDE, A. V.; BASSOI, L. H.; INAMASU, R. Y. (Ed.). *Agricultura de Precisão: Resultados de um Novo Olhar*. Brasília, DF: Embrapa, 2014. cap. 8, p. 109–134.
- KAVZOGLU, T.; MATHER, P. M. The role of feature selection in artificial neural network applications. *International Journal of Remote Sensing*, Taylor Francis, v. 23, n. 15, p. 2919–2937, 2002. Disponível em: <<https://doi.org/10.1080/01431160110107743>>.
- KHAN, M. J.; KHAN, H. S.; YOUSAF, A.; KHURSHID, K.; ABBAS, A. Modern trends in hyperspectral image analysis: A review. *IEEE Access*, v. 6, p. 14118–14129, 2018. ISSN 2169-3536.
- KIMOVSKI, D.; RISTOV, S.; MATHÁ, R.; PRODAN, R. Multi-objective service oriented network provisioning in ultra-scale systems. In: HERAS, D. B.; BOUGÉ, L.; MENCAGLI, G.; JEANNOT, E.; SAKELLARIOU, R.; BADIA, R. M.; BARBOSA, J. G.; RICCI, L.; SCOTT, S. L.; LANKES, S.; WEIDENDORFER, J. (Ed.). *Euro-Par 2017: Parallel Processing Workshops*. Cham: Springer International Publishing, 2018. p. 529–540. ISBN 978-3-319-75178-8.
- KLEMAS, V. V. Advances in coastal wetland remote sensing. In: *2014 IEEE/OES Baltic International Symposium (BALTIC)*. [S.l.: s.n.], 2014. p. 1–16. ISSN 2150-6027.
- KLIMA, R. L.; IZENBERG, N. R.; HOLSCLAW, G. M.; HELBERT, J.; D'AMORE, M.; MCCLINTOCK, W. E.; SOLOMON, S. C. Visible to near-infrared hyperspectral measurements of mercury: Challenges for deciphering surface mineralogy. In: *2014 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*. [S.l.: s.n.], 2014. p. 1–4. ISSN 2158-6276.

- KOHAVI, R.; JOHN, G. H. Wrappers for feature subset selection. *Artificial Intelligence*, v. 97, n. 1, p. 273 – 324, 1997. ISSN 0004-3702. Relevance. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S000437029700043X>>.
- KULBACKI, M.; SEGEN, J.; KNIEĆ, W.; KLEMPPOUS, R.; KLUWAK, K.; NIKODEM, J.; KULBACKA, J.; SERESTER, A. Survey of drones for agriculture automation from planting to harvest. In: *2018 IEEE 22nd International Conference on Intelligent Engineering Systems (INES)*. [S.l.: s.n.], 2018. p. 000353–000358. ISSN 1543-9259.
- KUMAR, M. Feature selection for classification of hyperspectral remotely sensed data using nsga-ii. *Water Resources Seminar*, 2004.
- KUO, B.; HO, H.; LI, C.; HUNG, C.; TAUR, J. A kernel-based feature selection method for svm with rbf kernel for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, v. 7, n. 1, p. 317–326, Jan 2014.
- LANDGREBE, D. A. *Signal Theory Methods in Multispectral Remote Sensing*. [S.l.]: Wiley, 2003.
- LANDIS, J. R.; KOCH, G. G. The measurement of observer agreement for categorical data. *Biometrics*, International Biometric Society, v. 33, n. 1, 1977.
- LEE, J. B.; WOODYATT, A. S.; BERMAN, M. Enhancement of high spectral resolution remote-sensing data by a noise-adjusted principal components transform. *IEEE Transactions on Geoscience and Remote Sensing*, v. 28, n. 3, p. 295–304, May 1990. ISSN 0196-2892.
- LI, S.; QI, H. Sparse representation based band selection for hyperspectral images. In: *2011 18th IEEE International Conference on Image Processing*. [S.l.: s.n.], 2011. p. 2693–2696.
- LI, S.; WU, H.; WAN, D.; ZHU, J. An effective feature selection method for hyperspectral image classification based on genetic algorithm and support vector machine. *Knowledge-Based Systems*, v. 24, n. 1, p. 40 – 48, 2011. ISSN 0950-7051. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0950705110001097>>.
- LIANG, J.; ZHOU, J.; QIAN, Y.; WEN, L.; BAI, X.; GAO, Y. On the sampling strategy for evaluation of spectral-spatial methods in hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, v. 55, n. 2, p. 862–880, Feb 2017.
- LILLESAND, T.; KIEFER, R.; CHIPMAN, J. *Remote sensing and image interpretation*. Wiley, 2004. ISBN 9780471152279. Disponível em: <<https://books.google.com.br/books?id=XiydAAAAMAAJ>>.
- LILLESAND, T.; KIEFER, R.; CHIPMAN, J. *Remote sensing and image interpretation*. [S.l.]: Wiley, 2015. ISBN 978-1-118-34328-9.
- LIU, M.-Y.; TUZEL, O.; RAMALINGAM, S.; CHELLAPPA, R. Entropy rate superpixel segmentation. In: . [S.l.: s.n.], 2011. p. 2097–2104.
- LOPEZ, L. S. *Investigação de técnicas eficientes para algoritmos evolutivos multiobjetivo baseados em decomposição*. Tese (Doutorado) — Universidade Federal de Minas Gerais, 2017.

- LORENA, A.; CARVALHO, A. de. Uma introdução às support vector machines. *RITA*, v. 14, p. 43–67, 01 2007.
- LORENZO, P. R.; TULCZYJEW, ; MARCINKIEWICZ, M.; NALEPA, J. *Band Selection from Hyperspectral Images Using Attention-based Convolutional Neural Networks*. 2018.
- LU, X.; HU, Z.; GUO, S. The quantitative estimation of periurban vegetation ecology using hyperspectral remote sensing. In: *2009 Joint Urban Remote Sensing Event*. [S.l.: s.n.], 2009. p. 1–6. ISSN 2334-0932.
- LUO, X.; XUE, R.; YIN, J. Information-assisted density peak index for hyperspectral band selection. *IEEE Geoscience and Remote Sensing Letters*, v. 14, n. 10, p. 1870–1874, Oct 2017. ISSN 1545-598X.
- MANOLAKIS, D.; MARDEN, D.; SHAW, G. Hyperspectral image processing for automatic target detection applications. *Lincoln Lab J*, v. 14, 01 2003.
- MAPA. *Ministério da Agricultura, Pecuária e Abastecimento. Agricultura de Precisão - Boletim Técnico*. Brasília, DF: [s.n.], 2009.
- MARTINEZ-USO, A.; F., P.; M., S. J.; P., G.-S. Clustering-based hyperspectral band selection using information measures. *IEEE Transactions on Geoscience and Remote Sensing*, v. 45, n. 12, p. 4158–4171, Dec 2007. ISSN 0196-2892.
- MASAELI, M.; FUNG, G.; DY, J. G. From transformation-based dimensionality reduction to feature selection. In: *Proceedings of the 27th International Conference on International Conference on Machine Learning*. USA: Omnipress, 2010. (ICML'10), p. 751–758. ISBN 978-1-60558-907-7. Disponível em: <<http://dl.acm.org/citation.cfm?id=3104322.3104418>>.
- MELGANI, F.; BRUZZONE, L. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, v. 42, n. 8, p. 1778–1790, Aug 2004.
- MITCHELL, T. M. Generalization as search. *Artificial Intelligence*, v. 18, n. 2, p. 203 – 226, 1982. ISSN 0004-3702. Disponível em: <<http://www.sciencedirect.com/science/article/pii/0004370282900406>>.
- MOLINA, L. C.; BELANCHE, L.; NEBOT, A. Feature selection algorithms: a survey and experimental evaluation. In: *2002 IEEE International Conference on Data Mining, 2002. Proceedings*. [S.l.: s.n.], 2002. p. 306–313.
- MONTEIRO, S. T.; MURPHY, R. J. Embedded feature selection of hyperspectral bands with boosted decision trees. In: *2011 IEEE International Geoscience and Remote Sensing Symposium*. [S.l.: s.n.], 2011. p. 2361–2364. ISSN 2153-7003.
- MURPHY, R. J.; MONTEIRO, S. T.; SCHNEIDER, S. Evaluating classification techniques for mapping vertical geology using field-based hyperspectral sensors. *IEEE Transactions on Geoscience and Remote Sensing*, v. 50, n. 8, p. 3066–3080, Aug 2012. ISSN 0196-2892.
- MURUGAN, D.; GARG, A.; AHMED, T.; SINGH, D. Fusion of drone and satellite data for precision agriculture monitoring. In: *2016 11th International Conference on Industrial and Information Systems (ICIIS)*. [S.l.: s.n.], 2016. p. 910–914.

- NAGASUBRAMANIAN, K.; JONES, S.; SARKAR, S.; SINGH, A. K.; SINGH, A.; GANAPATHYSUBRAMANIAN, B. Hyperspectral band selection using genetic algorithm and support vector machines for early identification of charcoal rot disease in soybean stems. *Plant Methods*, v. 14, n. 86, 2017. ISSN 1746-4811.
- NAZ, B.; XIAO, L.; JAWAID, M.; SOOMRO, S.; JAFFARI, N. A bilateral filter based post-processing approach for supervised spectral-spatial hyperspectral image classification. *Mehran University Research Journal of Engineering and Technology*, v. 37, p. 615–630, 07 2018.
- PATRA, S.; MODI, P.; BRUZZONE, L. Hyperspectral band selection based on rough set. *IEEE Transactions on Geoscience and Remote Sensing*, v. 53, n. 10, p. 5495–5503, Oct 2015. ISSN 0196-2892.
- PUDIL, P.; NOVOVICOVA, J.; KITTLER, J. Floating search methods in feature selection. *Pattern Recognition Letters*, v. 15, n. 11, p. 1119 – 1125, 1994. ISSN 0167-8655. Disponível em: <<http://www.sciencedirect.com/science/article/pii/0167865594901279>>.
- QIN, J.; CHAO, K.; KIM, M. S.; LU, R.; BURKS, T. F. Hyperspectral and multispectral imaging for evaluating food safety and quality. *Journal of Food Engineering*, v. 118, n. 2, p. 157 – 171, 2013. ISSN 0260-8774. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0260877413001659>>.
- RICHARDS, J. A.; JIA, X. *Remote Sensing Digital Image Analysis. An Introduction*. 5. ed. [S.l.]: Springer, 2013.
- RODARMEL, C.; SHAN, J. Principal component analysis for hyperspectral image classification. *Surv Land inf Syst*, v. 62, 01 2002.
- RODRIGUEZ, A.; LAIO, A. Clustering by fast search and find of density peaks. *Science*, American Association for the Advancement of Science, v. 344, n. 6191, p. 1492–1496, 2014. ISSN 0036-8075. Disponível em: <<https://science.sciencemag.org/content/344/6191/1492>>.
- RUSS, J. C. *The Image Processing Handbook, Third Edition*. CRC-Press, 1998. (Image Processing Handbook). ISBN 9780849325328. Disponível em: <<https://books.google.com.br/books?id=ovpRAAAAMAAJ>>.
- SAQUI, D.; LIMA, D. Caio de; SAITO, J. H.; JORGE, L.; M, S. A. T. An approach based on spea2-band selection and random forest classifier to generate thematic maps from hyperspectral images. In: . [S.l.: s.n.], 2019b.
- SAQUI, D.; SAITO, J. H.; JORGE, L. A. D. C.; FERREIRA, E. J.; LIMA, D. C.; P., H. J. Methodology for band selection of hyperspectral images using genetic algorithms and gaussian maximum likelihood classifier. p. 733–738, Dec 2016.
- SAQUI, D.; SAITO, J. H.; JORGE, L. A. D. C.; FERREIRA, E. J.; LIMA, D. C.; ATAKY T. M, S.; FAMBRINI, F. Nsga2-based method for band selection for supervised segmentation in hyperspectral imaging. Oct 2019.
- SAQUI, D.; SAITO, J. H.; LIMA, D. C. D.; CURA, L. M. D. V.; ATAKY, S. T. M. Incorporated decision-maker-based multiobjective band selection for pixel classification of hyperspectral images. *Advances in Electrical and Computer Engineering (AECE)*, v. 19, n. 4, p. 21–28, Dez 2019c. ISSN 1582-7445.

- SCOTT, D. *Multivariate Density Estimation: Theory, Practice, and Visualization*. Wiley, 1992. (A Wiley-interscience publication). ISBN 9780471547709. Disponível em: <<https://books.google.com.br/books?id=7crCUS\F2ocC>>.
- SHAHANA, A. H.; PREEJA, V. Survey on feature subset selection for high dimensional data. In: *2016 International Conference on Circuit, Power and Computing Technologies (ICCPCT)*. [S.l.: s.n.], 2016. p. 1–4.
- SHARMA, V.; DIBA, A.; TUYTELAARS, T.; GOOL LUC, V. Hyperspectral cnn for image classification band selection, with application to face recognition. In: . [S.l.: s.n.], 2016.
- SIEDLECKI, W.; SKLANSKY, J. A note on genetic algorithms for large-scale feature selection. *Pattern Recognition Letters*, v. 10, n. 5, p. 335 – 347, 1989. ISSN 0167-8655. Disponível em: <<http://www.sciencedirect.com/science/article/pii/0167865589900378>>.
- SOHAIB, M.; HAQ, I. U.; MUSHTAQ, Q. Dimensional reduction of hyperspectral image data using band clustering and selection based on statistical characteristics of band images. *International Journal of Computer and Communication Engineering*, v. 2, p. 101–105, 03 2013.
- STEIN, G.; CHEN, B.; WU, A. S.; HUA, K. A. Decision tree classifier for network intrusion detection with ga-based feature selection. In: *Proceedings of the 43rd Annual Southeast Regional Conference - Volume 2*. New York, NY, USA: ACM, 2005. (ACM-SE 43), p. 136–141. ISBN 1-59593-059-0. Disponível em: <<http://doi.acm.org/10.1145/1167253.1167288>>.
- SU, H.; YONG, B.; DU, Q. Hyperspectral band selection using improved firefly algorithm. *IEEE Geoscience and Remote Sensing Letters*, v. 13, n. 1, p. 68–72, Jan 2016. ISSN 1545-598X.
- SUI, C.; TIAN, Y.; XU, Y.; XIE, Y. Unsupervised band selection by integrating the overall accuracy and redundancy. *IEEE Geoscience and Remote Sensing Letters*, v. 12, n. 1, p. 185–189, Jan 2015. ISSN 1545-598X.
- SUN, K.; GENG, X.; JI, L.; LU, Y. A new band selection method for hyperspectral image based on data quality. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, v. 7, n. 6, p. 2697–2703, June 2014. ISSN 1939-1404.
- SUN, M.; CHUNYANG, W.; WANG, S.; ZHAO, Z.; LI, X. A new semisupervised-entropy framework of hyperspectral image classification based on random forest. *Advances in Multimedia*, v. 2018, p. 1–27, 09 2018.
- SUN, W.; DU, Q. Hyperspectral band selection: A review. *IEEE Geoscience and Remote Sensing Magazine*, v. 7, n. 2, p. 118–139, June 2019. ISSN 2168-6831.
- SUN, W.; TIAN, L.; XU, Y.; ZHANG, D.; DU, Q. Fast and robust self-representation method for hyperspectral band selection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, v. 10, n. 11, p. 5087–5098, Nov 2017. ISSN 1939-1404.
- SUN, W.; ZHANG, L.; DU, B.; LI, W.; LAI, M. Y. Band selection using improved sparse subspace clustering for hyperspectral imagery classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, v. 8, n. 6, p. 2784–2797, June 2015. ISSN 1939-1404.

- SUN, W.; ZHANG, L.; ZHANG, L.; LAI, Y. M. A dissimilarity-weighted sparse self-representation method for band selection in hyperspectral imagery classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, v. 9, n. 9, p. 4374–4388, Sep. 2016.
- TERWILLIGER, B.; ISON, D. C.; ROBBINS, J.; VINCENZI, D. *Small Unmanned Aircraft Systems Guide: Exploring Designs, Operations, Regulations, and Economics*. [S.l.]: Aviation Supplies & Academics, Inc., 2017. ISBN 161954394X, 9781619543942.
- THENKABAIL, P.; GUMMA, M.; TELUGUNTLA, P.; AHMED, M. I. Hyperspectral remote sensing of vegetation and agricultural crops. *Photogrammetric Engineering and Remote Sensing*, v. 80, 08 2014.
- THEODORIDIS, S.; KOUTROUMBAS, K. *Pattern Recognition*. [S.l.]: Elsevier Science, 2003. ISBN 9780080513621.
- TU, B.; ZHANG, X.; WANG, J.; ZHANG, G.; OU, X. Spectral–spatial hyperspectral image classification via non-local means filtering feature extraction. *Sensing and Imaging*, v. 19, 12 2018.
- TUV, E.; BORISOV, A.; RUNGER, G.; TORKKOLA, K. Feature selection with ensembles, artificial variables, and redundancy elimination. *J. Mach. Learn. Res.*, JMLR.org, v. 10, p. 1341–1366, dez. 2009. ISSN 1532-4435. Disponível em: <<http://dl.acm.org/citation.cfm?id=1577069.1755828>>.
- VAIPHASA, C.; SKIDMORE, A. K.; BOER, W. F. d.; VAIPHASA, T. A hyperspectral band selector for plant species discrimination. *ISPRS Journal of Photogrammetry and Remote Sensing*, v. 62, n. 3, p. 225 – 235, 2007. ISSN 0924-2716. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0924271607000512>>.
- VAPNIK, V. N. *The Nature of Statistical Learning Theory*. Berlin, Heidelberg: Springer-Verlag, 1995. ISBN 0-387-94559-8.
- WANG, J.; ZHANG, K.; WANG, P.; MADANI, K.; SABOURIN, C. Unsupervised band selection using block-diagonal sparsity for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, v. 14, n. 11, p. 2062–2066, Nov 2017. ISSN 1545-598X.
- WANG, Q.; LIN, J.; YUAN, Y. Salient band selection for hyperspectral image classification via manifold ranking. *IEEE Transactions on Neural Networks and Learning Systems*, v. 27, n. 6, p. 1279–1289, June 2016. ISSN 2162-237X.
- WANG, T.; ZHANG, H.; LIN, H.; JIA, X. A sparse representation method for a priori target signature optimization in hyperspectral target detection. *IEEE Access*, v. 6, p. 3408–3424, 2018. ISSN 2169-3536.
- WESTON, J.; MUKHERJEE, S.; CHAPELLE, O.; PONTIL, M.; POGGIO, T.; VAPNIK, V. Feature selection for svms. In: . [S.l.: s.n.], 2000. v. 13, p. 668–674.
- XIA, J.; GHAMISI, P.; YOKOYA, N.; IWASAKI, A. Random forest ensembles and extended multiextinction profiles for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, v. 56, n. 1, p. 202–216, Jan 2018. ISSN 0196-2892.

- XIE, F.; LI, F.; LEI, C.; KE, L. Representative band selection for hyperspectral image classification. *ISPRS International Journal of Geo-Information*, v. 7, n. 9, 2018. ISSN 2220-9964. Disponível em: <<https://www.mdpi.com/2220-9964/7/9/338>>.
- XU, X.; SHI, Z.; PAN, B. A new unsupervised hyperspectral band selection method based on multiobjective optimization. *IEEE Geoscience and Remote Sensing Letters*, v. 14, n. 11, p. 2112–2116, Nov 2017. ISSN 1545-598X.
- XU, Y.; DU, Q.; YOUNAN, N. H. Particle swarm optimization-based band selection for hyperspectral target detection. *IEEE Geoscience and Remote Sensing Letters*, v. 14, n. 4, p. 554–558, April 2017.
- YANG, C.; TAN, Y.; BRUZZONE, L.; LU, L.; GUAN, R. Discriminative feature metric learning in the affinity propagation model for band selection in hyperspectral images. *Remote Sensing*, v. 9, p. 782, 2017.
- YANG, R.; SU, L.; ZHAO, X.; WAN, H.; SUN, J. Representative band selection for hyperspectral image classification. *Journal of Visual Communication and Image Representation*, v. 48, p. 396 – 403, 2017. ISSN 1047-3203. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1047320317300391>>.
- YAO, F.; QIAN, Y. Band selection based gaussian processes for hyperspectral remote sensing images classification. p. 2845–2848, Nov 2009. ISSN 1522-4880.
- YUAN, Y.; LIN, J.; WANG, Q. Dual-clustering-based hyperspectral band selection by contextual analysis. *IEEE Transactions on Geoscience and Remote Sensing*, v. 54, n. 3, p. 1431–1445, March 2016. ISSN 0196-2892.
- YUAN, Y.; ZHENG, X.; LU, X. Discovering diverse subset for unsupervised hyperspectral band selection. *IEEE Transactions on Image Processing*, v. 26, n. 1, p. 51–64, Jan 2017. ISSN 1057-7149.
- ZHAN, Y.; HU, D.; XING, H.; YU, X. Hyperspectral band selection based on deep convolutional neural network and distance density. *IEEE Geoscience and Remote Sensing Letters*, v. 14, n. 12, p. 2365–2369, Dec 2017. ISSN 1545-598X.
- ZHANG, M.; GONG, M.; CHAN, Y. Hyperspectral band selection based on multi-objective optimization with high information and low redundancy. *Applied Soft Computing*, v. 70, p. 604 – 621, 2018. ISSN 1568-4946. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1568494618303326>>.
- ZHANG, M.; MA, J.; GONG, M. Unsupervised hyperspectral band selection by fuzzy clustering with particle swarm optimization. *IEEE Geoscience and Remote Sensing Letters*, v. 14, n. 5, p. 773–777, May 2017. ISSN 1545-598X.
- ZHANG, M.; MA, J.; GONG, M.; LI, H.; LIU, J. Memetic algorithm based feature selection for hyperspectral images classification. In: *2017 IEEE Congress on Evolutionary Computation (CEC)*. [S.l.: s.n.], 2017. p. 495–502.
- ZHANG, Q.; LI, H. Moea/d: A multiobjective evolutionary algorithm based on decomposition. *IEEE Transactions on Evolutionary Computation*, v. 11, n. 6, p. 712–731, Dec 2007. ISSN 1089-778X.

ZHANG, R.; MA, J. Feature selection for hyperspectral data based on recursive support vector machines. *International Journal of Remote Sensing*, Taylor Francis, v. 30, n. 14, p. 3669–3677, 2009. Disponível em: <<https://doi.org/10.1080/01431160802609718>>.

ZHANG, S.; LI, S.; FU, W.; FANG, L. Multiscale superpixel-based sparse representation for hyperspectral image classification. *Remote Sensing*, v. 9, n. 2, 2017. ISSN 2072-4292. Disponível em: <<https://www.mdpi.com/2072-4292/9/2/139>>.

ZHANG, X.; SUN, Q.; LI, J. Optimal band selection for high dimensional remote sensing data using genetic algorithm. In: . [s.n.], 2009. v. 7471. Disponível em: <<https://doi.org/10.1117/12.847907>>.

ZHOU, A.; ZHANG, Q.; ZHANG, G. A multiobjective evolutionary algorithm based on decomposition and probability model. In: *2012 IEEE Congress on Evolutionary Computation*. [S.l.: s.n.], 2012. p. 1–8. ISSN 1941-0026.

ZHU, G.; HUANG, Y.; LI, S.; TANG, J.; LIANG, D. Hyperspectral band selection via rank minimization. *IEEE Geoscience and Remote Sensing Letters*, v. 14, p. 2320–2324, 2017.

ZHUO, L.; ZHENG, J.; LI, X.; WANG, F.; AI, B.; QIAN, J. A genetic algorithm based wrapper feature selection method for classification of hyperspectral images using support vector machine. v. 7147, 2008. Disponível em: <<https://doi.org/10.1117/12.813256>>.

ZITZLER, E.; LAUMANN, M.; THIELE, L. Spea2: Improving the strength pareto evolutionary algorithm. *TIK-Report*, v. 103, 07 2001.

GLOSSÁRIO

AE – *Algoritmos Evolutivos/Evolucionários*

AG – *Algoritmos Genéticos*

AP – *Affinity propagation*

CSA – *clonal selection algorithm*

DBSCAN – *Density-based spatial clustering of applications with noise*

DCCA – *Dual-Clustering-based band selection by context analysis*

DM – *Decision-Maker*

DP – *Density Peak*

DWSSR – *dissimilarity-weighted sparse self-representation*

ELM – *Extreme Learning Machine*

FAO – *Food and Agriculture Organization of the United Nations*

FP – *Fronteira de Pareto*

GB – *Gigabytes*

GMLC – *Gaussian Maximum Likelihood Classifier*

GMM – *Gaussian Mixture Model*

GT – *Ground Truth*

IDMMoBS – *Incorporated Decision-Maker-Based Multiobjective Band Selection*

IG-GWO – *Information Gain-Grey Wolf Optimization*

IGR – *Information Gain Ratio*

IG – *Information Gain*

IHs – *Imagens hiperespectrais*

IM – *informação mútua*

IRMoBS – *incorporated rank based multi-objective band selection algorithm*

IaDPI – *information-assisted density peak index*

- JSRC** – *Joint Sparse Representation Classification*
- KM** – *Kilômetro*
- KNN** – *K-Nearest Neighbors*
- LASSO** – *Least Absolute Shrinkage and Selection Operator*
- LDA** – *Linear Discriminant Analysis*
- MAMR** – *joint maximum overall accuracy and minimum redundancy-based*
- MASR** – *multiscale adaptive sparse representation*
- MDCM** – *maximum determinant of covariance matrix*
- MDC** – *Mahalanobis distance classifier*
- MDPP** – *multigraph determinantal point process*
- MIMR** – *Maximum information and minimum redundancy*
- MLP** – *Multilayer Perceptron*
- MNBS** – *minimum noise band selection*
- MOBS** – *Multiobjective optimization band selection*
- MOEA/D** – *Multiobjective Evolutionary Algorithm based on Decomposition*
- MOGLS** – *Multiobjective genetic local search*
- MRF** – *Markov Random Field*
- MSSR** – *Multiscale Superpixel-based Sparse Representation*
- MVPCA** – *Maximum variance principal component analysis*
- NAPC** – *noise-adjusted principal component*
- NHHMC** – *Non-homogeneous Hidden Markov Chains*
- NPFS** – *nonlinear parsimonious feature selection*
- NSGA2** – *Nondominated sorting genetic algorithm II*
- NTMI** – *normalized trivariable mutual information*
- PCA** – *Principal Component Analysis*
- POM** – *problema de otimização multiobjetivo*
- PSO** – *particle swarm optimization*
- RBF** – *Radial Basis Function*
- RFE** – *Recursive-Feature Elimination*
- RF** – *Random Forest*
- RGB** – *Red-Green-Blue*

- RNA** – *Redes Neurais Artificiais*
- RP** – *Reconhecimento de padrões*
- SAM** – *Spectral Angle Mapper*
- SBFS** – *Sequential Backward Floating Selection*
- SBS** – *sequential backward selection*
- SC** – *Spectral Clustering*
- SBFS** – *Sequential Forward Floating Search*
- SFS** – *sequential forwarding selection*
- SLIC** – *Simple linear iterative clustering*
- SNR** – *signal-to-noise ratio*
- SPEA2** – *Strength Pareto Evolutionary Algorithm*
- SRP** – *Sistema de reconhecimento de padrões*
- SR** – *Sensoriamento Remoto*
- SSR** – *sparse self-representation*
- SVM** – *Support Vector Machines*
- SpaBS** – *Sparsity based band selection*
- TDS** – *Tchebycheff Decomposition Strategy*
- VS** – *Vetores de Suporte*
- WMoEBS** – *Wrapper Multiobjective Evolutionary Band Selection*
- WaLuDi** – *Ward's Linkage strategy Using Divergence*
- WaLuMi** – *Ward's Linkage strategy Using Mutual Information*
- nm** – *Nanômetro*

Apendice A

ARTIGOS PRODUZIDOS DURANTE ESTA PESQUISA

Durante esta tese, quatro publicações principais foram realizadas na literatura e os respectivos artigos são:

1. *Methodology for Band Selection of Hyperspectral Images Using Genetic Algorithms and Gaussian Maximum Likelihood Classifier*, publicado no evento *International Conference on Computational Science and Computational Intelligence (CSCI)* em 2016, sem Qualis área de Ciência da Computação da avaliação do documento da Capes QUALIS CONFERÊNCIAS - CIÊNCIA DA COMPUTAÇÃO - 2016 vigente até conclusão deste texto.
2. *Approach based on SPEA2-band selection and Random Forest classifier to generate thematic maps from hyperspectral images*, publicado no evento *The Americas Conference on Information Systems (AMCIS)* em 2019, com Qualis A2 na área de Ciência da Computação da avaliação do documento da Capes QUALIS CONFERÊNCIAS - CIÊNCIA DA COMPUTAÇÃO - 2016 vigente até conclusão deste texto.
3. *NSGA2-Based Method for Band Selection for Supervised Segmentation in Hyperspectral Imaging*, publicado no evento *Systems, Man, and Cybernetics Society (SMC)* em 2019, com Qualis A2 na área de Ciência da Computação da avaliação do documento da Capes QUALIS CONFERÊNCIAS - CIÊNCIA DA COMPUTAÇÃO - 2016 vigente até conclusão deste texto.
4. *Incorporated Decision-Maker-based multiobjective band selection for pixel classification of hyperspectral images* publicado na revista *Advances in Electrical and Computer Engineering* em 2019, com Qualis B1 na área de Ciência da Computação do documento da AVALIAÇÃO DE PERIÓDICOS QUADRIÊNIO 2013-2016 vigente até conclusão deste texto.

Methodology for band selection of Hyperspectral Images using Genetic Algorithms and Gaussian Maximum Likelihood Classifier

Diego Saqui, José H. Saito*

Computing Department
UFSCAR - Federal University of São Carlos
São Carlos, SP - Brazil

* Faculty Campo Limpo Paulista – FACCAMP

Lucio A. de C. Jorge, Ednaldo J. Ferreira, Daniel C.

Lima, João P. Herrera.
Embrapa Instrumentation
São Carlos, SP - Brazil

Abstract— Remote Sensing is a technique to obtain data from objects without physical contact using sensors. If it is performed by hyperspectral sensors, then is possible to have a large number of bands that allows many applications. In this case, a large amount of data needs to be processed and analyzed, and therefore the practical use of these images have challenges with storage, transport and processing time. Another issue is that many of these bands contain redundant information due to high correlation between them that can cause instability in the convergence of computing processes. Because of these problems there is a need to select important bands to be processed. In many monitoring applications through remote sensing prior knowledge of areas can be used to improve the processing. Algorithms that use this information are known as supervised methods. This study introduces a methodology for band selection based on Genetic Algorithms and Gaussian Maximum Likelihood Classifier. The results were significant and demonstrate that the proposed method can reduce about 26,56% of the number of bands in hyperspectral images maintaining good quality for interpretation. The result of the similarity of the segmentation of images generated with fewer bands and more bands in the proposed methodology and the other algorithm were also compared. With the results expected to apply the proposed methodology in practical work and further study.

Keywords—Band Selection, Genetic Algorithm, Gaussian Maximum Likelihood, Hyperspectral Images, Remote Sensing.

I. INTRODUCTION

Remote Sensing (RS) is a technique to obtain data of the earth's surface objects without a physical contact between the sensor and the object [1]. Through the data obtained by sensors may be generated images that allow analyzes and interpretations [2]. Images obtained by RS, usually have high resolution and can be categorized as RGB (red-green-blue), multispectral, hyperspectral or others. RS and Hyperspectral Images (HI) are used in different applications, such as precision agriculture, for detection of health conditions of crops [3]. HIs have the characteristic of a high volume of data that require advanced algorithms for processing and analysis [4].

HI increase sensitivity beyond the visible region because there are more bands than RGBs images [3]. Each pixel contains a spectral signature composed by a large number of

bands with different characteristics enabling distinction of the materials [5]. HI challenges are generated due to the high number of bands which lead to volumetric expansion data and the increasing complexity of the analysis algorithms [1][6]. Because of this, the practical use and research on HI has challenges with storage, transmission and processing time [7], information redundancy caused by high correlation, difficulty in the determination of training samples for classifiers [8].

Based on these problems, a research interest in HI is to select bands which are relevant for certain applications, i.e., reduce the data volume. By reducing the number of bands, another solution held is related to the manufacturing cost of hyperspectral sensors. If the selected bands are close to each other in the spectral signature, multispectral sensors (which are cheaper than hyperspectral sensors) can be used [9]. In Brazil, hyperspectral sensors have been used drones to research the mapping of agricultural areas and identification of health conditions of crops. Due to the high cost of hyperspectral sensors, this technology can take a long time to be accessible to agricultural producers of small- and medium-size, so cameras affordable (multispectral) may be an alternative.

Methods that reduce/identifies characteristics of the data set are described as dimensionality reduction and can be divided into feature extraction and feature selection. In feature extraction, during processing it is common to perform a transformation of the original feature space of the data causing loss of the original data representation. If the data have been restored to its original representation, then some information may be distorted. Feature selection is more viable for dimensionality reduction in HI because it preserves the original data representation. This characteristic is very important to reduce the data in HI because during the process is necessary to determine the specific bands [10]. For any of these strategies, in an image segmentation process with a large number of bands, may be used prior knowledge provided by experts (supervised methods). Experts may indicate an area of interest, allowing the definition of training samples to an algorithm. For example, in agriculture, an expert may indicate areas of plantations in the image with different health conditions and this information can be used in a training process for a subsequent classification.

In this study we propose a supervised methodology for band selection based on Genetic Algorithm (GA) and Gaussian Maximum Likelihood Classifier (GMLC). The GA has the function of trying to find the best possible solution in time computationally possible. GA tries to evaluate the fitness function by comparing the segmented image with all bands and the segmented image with reduced bands. The GMLC is used to segment images in the process. GMLC is a supervised algorithm and thus require knowledge of areas of interest and this knowledge is represented by a segmented image with all the bands. The Structural Similarity Index (SSIM) considering the spatial information (the pixels) is used as fitness function. By adapting these techniques, the proposed methodology is able to achieve significant results.

The remainder of this paper is organized in the following sections: II. Related Work; III. Proposed Methodology; IV. Experiments and Results; and V. Conclusion.

II. RELATED WORK

An example of feature extraction is Principal Component Analysis (PCA). PCA and Independent Component Analysis (ICA) can be used for dimensionality reduction in Hyperspectral Images (HI) but can change the data representation [11]. The PCA is based on the analysis of the correlation matrix and drops the lowest variance components. The ICA is similar to PCA, but considers the independence of the statistical features. The application of these techniques results in components minimized by separating the classes [12]. Image fusion and recursive filtering is a technique based on the combination of adjacent bands by statistical methods and applying specific filters on these subsets. The result shows a smaller number of features in relation to the total number of original bands. These features can be applied to segmentation processes. This methodology has the advantage of considering spectral and spatial information to extract characteristics [13].

The three methodologies described as feature extraction are unsupervised, and do not use prior knowledge. This information could be useful to determine which bands are really significant for an analysis. Another issue is that feature extraction methodology does not preserve the original data representation. The original representation of the data can be important for construction of sensors that capture a specific range of the spectrum. Features selection methodologies have the advantage of preserving the original representation of data and there are many such methodologies that are supervised, i.e., consider the knowledge of experts for its implementation.

Features selection methodologies can be based on different approaches. Approaches based on clustering such as K-Means [7], Fuzzy C-Means [9] and hierarchical clustering [14] [15] are interesting to find hidden patterns in data sets. Many of these methodologies use techniques to estimate the number of clusters to be defined. In HI these methodologies can be useful for identifying patterns introduced by segmentation of the areas of interest. These methodologies are unsupervised, i.e., they are useful to get hidden patterns for understanding of HI, on the other hand they do not consider the knowledge of experts, and this information can be very useful for many applications.

Others methodologies for feature selection as the ICA-based band selection instead of transforming the original data representation of HI evaluate a matrix of weights to determine how each band contributes in relation to others. Bands that contain more information are selected to compose new images [16]. Search Based Methods can be structured based on a priori knowledge of experts, statistical separation between classes using correlation applied to spatial information, statistical separation between classes using correlation applied to spectral information, and others [17]. There are also methodologies which use unsupervised and supervised strategies to try to estimate the best set of bands of a HI to a classification [18].

Finally, there are methodologies with optimization algorithms directed to search for different combinations of a set of data indicating the best combination. Many algorithms are bio inspired, such as the Genetic Algorithms (GA) [8] and Particle Swarm Optimization (PSO) [19]. The GA is a technique inspired by the concept of evolutionary theory of Darwin [8] and the PSO is inspired by the behavior and social interaction of birds and fish [19]. Both algorithms have the characteristic of searching the best possible combination in a given time of different elements (individuals) in a data set (population) based on the evaluation of an objective function (fitness function). The search for each algorithm (evolution) is determined based on the best of populations in each iteration (generation). These algorithms have been explored in the literature to evaluate different compositions of HI bands presenting good results.

III. PROPOSED METHODOLOGY

The proposed methodology is shown in the flowchart of Figure 1. For the execution of the proposed methodology is expected as input parameter of one or a set of hyperspectral images of a common area and a training sample of interest areas of the images using a priori knowledge. This training sample must be selected from a ground truth of the image.

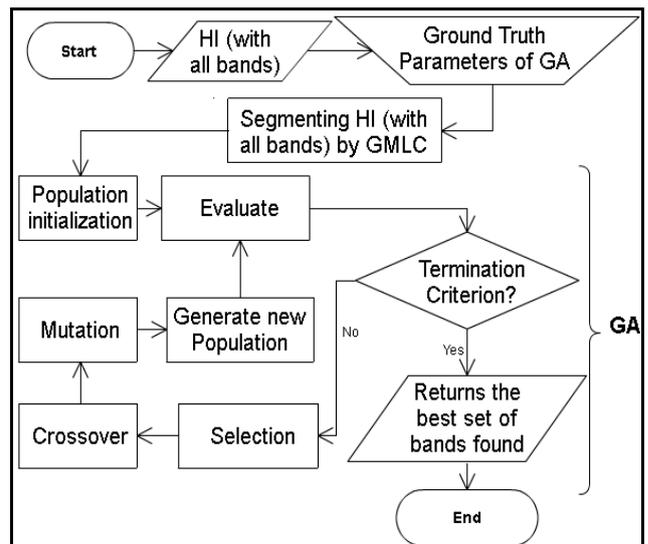


Fig. 1. Flowchart of proposed methodology.

A. *Gaussian Maximum Likelihood Classifier (GMLC) and Structural Similarity Index (SSIM)*

GMLC is an algorithm that evaluates, quantitatively, the variance and covariance of the spectral class responsible for patterns when it is classifying an unknown pixel. The probability density function is used to classify pixels by computing the probability of value belonging to each class. The pixel is assigned to the most likely class or can be set as unknown if the probability values are all below a threshold set by the analyst [24]. GMLC is applied to the original images which are composed of all bands to classify and segment areas. Based on these images, for the classification is considered a training stage with areas identified by experts through the ground truth. Some sampled pixels are selected and used subsequently for image segmentation.

During the execution of the Genetic Algorithm (GA) in the “Population Initialization” and “Generate New Population” are generated images with different combinations of bands (fewer bands than the original). In this context, the GMLC is also used to segment areas of interest in these images. The segmentation rules of GMLC applied to images with all bands and with fewer bands are the same. For example, the class of each image with more pixels always gets the same grayscale value in the segmentation, the same happens for the class with the second largest number of pixels and so on. This characteristic is important because the comparison is subsequently performed on the segmented images.

The SSIM is a method used for measuring the similarity between two images. This method is based on the assumption that the human visual system is adapted for extracting structural information from a scene. The SSIM function considers luminance, contrast and structure on the comparison. SSIM is applied to windows of different images (with predetermined size) and considers the dependency between pixels that are next to each other defining a relevant information about the structure of objects in the scene. The resultant SSIM index is a decimal value where 1 is only reachable in the case of two identical sets of data [22]. To establish the similarity between two full of images can be calculated the average SSIM values. The SSIM function is used as fitness function for image comparison. In the context of this study, the SSIM operates with the objective of evaluating the pixels between the segmented images with all bands and segmented images with a smaller number of bands.

B. *Genetic algorithms (GA)*

GA are inspired by Darwin’s theory of evolution and natural selection for adaptive heuristic search processes. GA explores historical information, searching the region of better performance within the search space. Each individual in a population can be represented by a vector (such as binary vector) characterizing a possible solution to a given problem. The variables that compose the chromosome of individuals are analogous to genes. The fitness function is responsible for evaluating each individual. This evaluation is used to select individuals that will be used in the crossover operation. The algorithm is run during a number of generations for the purpose of finding the best solution (the best individual) in viable time (determined by the number of generations). In the GA, three

types of genetic operators are common: selection (selects chromosomes with better fitness for crossover); crossover (through the selection process chooses two chromosomes of individuals and mix their genes to create two new chromosomes); and mutation (flips some of the values of genes in a chromosome). In this study, GA has the main function for the process of band selection. The main features of the GA are described below. The representation of the chromosome in the GA is structured as the binary vector shown in Figure 2.

0	1	0	0	1	1	0	...	number of bands
---	---	---	---	---	---	---	-----	-----------------

Fig. 2. Chromosome

This vector has a number of positions equal to the total number of bands of each image. A value of 1 indicates the presence of a band and 0 the absence. The fitness function considers the quality of segmentation and the smallest number of bands that can achieve. Probabilities and rates of mutation and crossover do not have any restrictions or defined patterns. The number of generations is related to the algorithm execution time. Since comparison and segmentation of images have a long processing time, it is recommended a small number of generations. For evaluation step in GA, it is first generated an image with bands indicated for a determined chromosome. Then this image is segmented with the GMLC and the previously indicated ground truth. Finally, the generated image is compared with the original through the fitness function (SSIM).

IV. EXPERIMENTS AND RESULTS

A. *Experiments*

The algorithms used in the proposed methodology were developed in Python 2.7 with frameworks Spectral Python (SPY) [22] and Distributed Evolutionary Algorithms in Python (DEAP) [25]. In this study was used 7 remote sensing hyperspectral images (HI) of different types and areas of coverage for the experiments. The HI shown in Figure 3 was used for the tests and the Figure 3a is an example of segmented areas of interest.

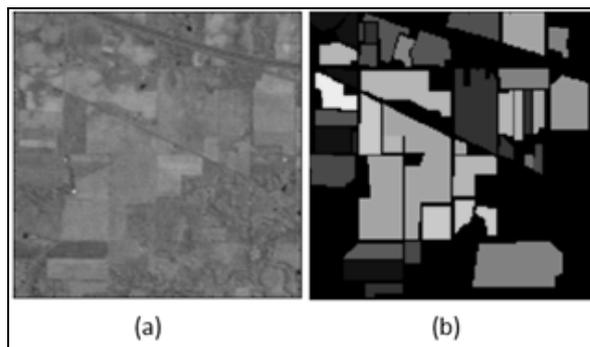


Fig. 3. (a) Image of Indian Pines. (b) Ground truth.

This image is a hyperspectral image (HI) (represented with bands 2, 97, 220) with ERDAS/Lan file format which represents an area in the Indian Pines in North-western Indiana with buildings, different crops and other types of natural

vegetation. This image can be found in the literature with the name '92AV3C.spc' with size 145 x 145 pixels, and has 220 spectral bands and 17 classes [23]. In Figure 3b is shown the ground truth of Figure 3a. The ground truth is defined by experts who know some areas that should be segmented. The Gaussian Maximum Likelihood Classifier (GMLC) was applied to the image contents all 220 bands and the result of the segmentation is shown in Figure 4.



Fig. 4. Segmented image with all bands.

Other classical images obtained from an image database have also been considered in this study. Some of these images are shown in Figure 5, Figure 5a represents an image collected by the 224-band AVIRIS sensor over Salinas Valley (Salinas Valley I), California with 512 x 217 pixels, and 16 class with vegetables, bare soils, and vineyard fields. Figure 5b represents another image obtained of the Salinas Valley (Salinas Valley II), California with 86 x 83 pixels, and 6 class with different types of vegetables. Figure 5c and Figure 5d are two scenes acquired by the ROSIS sensor over Pavia, Italy. The number of spectral bands is 102 for Pavia Centre (Figure 5c) and 103 for Pavia University (Figure 5d). Pavia Centre is a 1096*1096 pixels' image and Pavia University is 610*610 pixels' image. Both images have 9 different types of classes (water, vegetation, asphalt, etc.).

Two other images not illustrated were also used for this experiment. One of the images was acquired by the NASA AVIRIS and represents date of the Kennedy Space Center (KSC), Florida, USA. This image has 224 bands and 13 classes (vegetation and other kinds of coverings). The other image was acquired by the satellite NASA EO-1 with Hyperion sensor on the Okavango Delta, Botswana, Africa. The image has 145 bands and 14 classes represented in seasonal swamps, occasional swamps, and drier woodlands located in the distal portion of the Delta [26].

For each image the following process was performed:

- The genetic algorithm (GA) was performed for 50 generations with populations with 50 individuals (generated randomly for first population), i.e., 50 images with different combinations of bands;

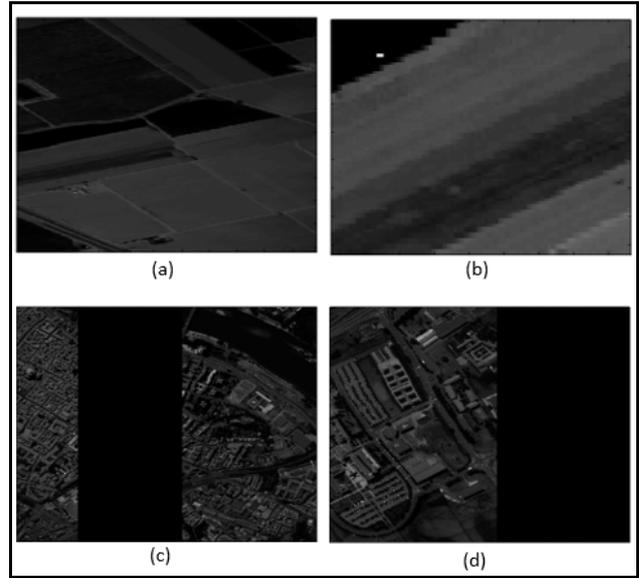


Fig. 5. (a) Salinas Valley I – 224 bands. (b) Salinas Valley II – 224 bands. (c) Pavia Centre – 102 bands. (d) Pavia University – 103 bands.

- Using all bands, a training process was performed by selecting some pixels of each class (defined by the ground truth);
- After training, the GMLC is applied to the image with all the bands and generated a new segmented image. The segmented image is used as a reference in the fitness function of the GA;
- At each generation the above process is repeated for images with fewer bands (1- presence of bands and 0-absence on chromosome);
- Each image (chromosome) of each population (fewer bands) is compared to the reference image (all bands) by the fitness function (Structural Similarity Index - SSIM);
- The segmented image with fewer bands more like the segmented image with all bands have their bands indicated at the end of the process.

The probabilities of occurring genetic operators of crossover and mutation are respectively 0.5 and 0.2. In crossover, two points method was used and the mutation can occur in multiple genes at once, with a probability of 0.05 per gene. The method used for selection is the tournament.

B. Tests

The graph in Figure 6 shows the results of the best individual of each generation of the GA for each image, i.e., the evolution of GA.

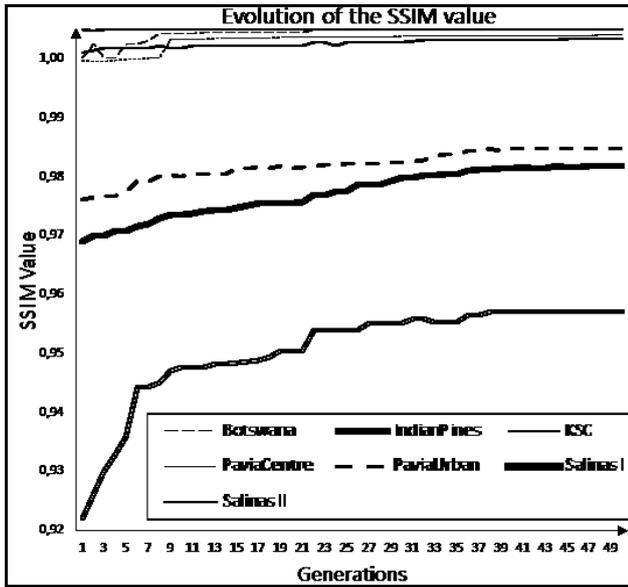


Fig. 6. Best individual of each generation of the GA for each image.

The fitness function (SSIM) shows an improvement over the generations for all images. Salinas Valley II image registered the best fitness in relation to other images and in many cases reached the maximum similarity in SSIM function. The methodology proposed to achieve good results because the image has a small size, well-defined classes and few classes. When the methodology was applied to image Indian Pines was presented the worst fitness value, but the best evolution during the application.

TABLE I. VALUES OF THE BEST FITNESS (SSIM) AND NUMBER OF SELECTED BANDS

Image	SSIM	Original bands	Reduced Bands	Reduction
Botswana	0.999	145	113	22,07%
Indian Pines	0.952	220	159	27,73%
KSC	0.998	224	134	40,18%
Pavia Centre	0.998	102	88	13,73%
Pavia University	0.979	103	85	17,48%
Salinas Valley I	0.976	224	173	22,78%
Salinas Valley II	1.0	224	130	41,96%

In the Table 1 are shown the values of the best individual (second column) for each image (first column) of all generations. Third column shows the total number of original image bands and the fourth column the number of bands after the implementation of the proposed methodology. The

proposed methodology reduces on average 26,56% of the number of image bands which shows a considerable decrease.

An Indian Pines image generated with the selected bands (shown in Table 1) is shown in Figure 7a. In Figure 7b is possible to visualize an image generated by another methodology (K-Means and NWHFC, see [7]). Although the K-Means algorithm and NWHFC present a greatly reduced number of bands (17 bands in this picture) you can view a large number of artifacts in the segmented image. Already the image obtained from the proposed methodology can have greater similarity to the original image.

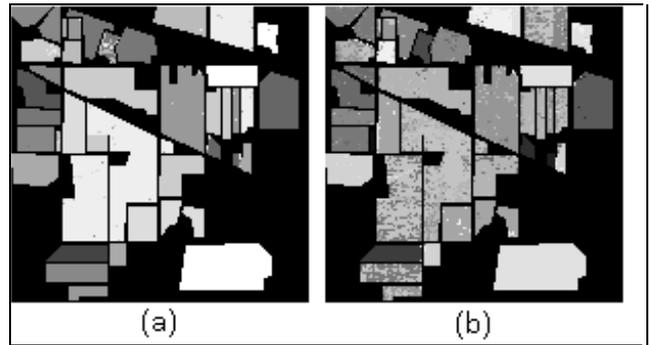


Fig. 7. (a) Image generated by proposed methodology. (b) Image generated by methodology using the K-Means and NWHFC.

V. CONCLUSION

This paper introduces a methodology based on Genetic Algorithm and Gaussian Maximum Likelihood Classifier (GMLC) to reduce the number of bands in hyperspectral images (HI). The methodology is supervised and thus requires knowledge of areas of interest in the images. During the experiments, the proposed methodology has been compared with other methodology of literature (based on K-Means and NWHFC) and showed good results in the category of quality of segmented image. The proposed methodology showed a reduction in the number of bands of about 26,56% and therefore the number of image bands can still be considered high. This is a limitation that will be explored in future studies.

In future studies, it will be conducted experiments with other images, considering other algorithms of segmentation, try a parameter for reducing the number of bands and exploring the possibility of treating the problem presented as multi-objective one. In future studies, it will be conducted experiments with other images, considering other algorithms of segmentation/classification, try a parameter for reducing the number of bands and exploring the possibility of treating the problem presented as multi-objective one.

ACKNOWLEDGMENT

The authors would like to thank Embrapa Instrumentation and CNPq with the project number 310310/2013-0 by providing resources for the development of this research.

REFERENCES

- [1] P. R. Meneses, T. de Almeida, A. N. C. S. Rosa, E. E. Sano, E. B. de Souza, G. M. M. Baptista and R. S. Brites, "Introdução ao processamento de imagens de sensoriamento remoto," 2012. [Online]. Available: <http://www.cnpq.br/documents/10157/56b578c4-0fd5-4b9fb82a-e9693e4f69d8>. [Accessed on April 2016].
- [2] G. Polder, and V. D. H. W. A. M. Gerie, "Visualization of spectral images," in SPIE Conf. Visualization and Optimization Techniques, 2001.
- [3] L. A. C. Jorge, and R. Y. Inamasu, "Uso de veículos aéreos não tripulados (VANTs) em Agricultura de Precisão," in Agricultura de Precisão. Resultados de um Novo Olhar., Embrapa, 2014.
- [4] S. Aksoy, N. H. Younan and L. Bruzzone, "Pattern recognition in remote sensing," Pattern Recognition Letters, vol. 31, n. 10, p. 1069 – 1070, 2010.
- [5] C. I. Chang, Hyperspectral Imaging: Techniques for spectral Detection, New York: Plenum Publishers, 2003.
- [6] F. Baret, G. Guyot and D. J. Major, "TSAVI: a vegetation index which minimizes soil brightness effects on LAI and APAR estimation," International Geoscience and Remote Sensing Symposium (IGARSS'89). Canadian Symposium on Remote Sensing, IEEE, pp. 1355-1358, 1989.
- [7] M. Sohaib, I. Ulhaq and Q. Mushtaq, "Dimensional reduction of hyperspectral image data using band clustering and selection based on statistical characteristics of band images," International Journal of Computer and Communication Engineering, vol. 2, 2013.
- [8] X. Zhang, Q. Sun and J. Li, "Optimal band selection for high dimensional remote sensing data using genetic algorithm," in SPIE 7471, Second International Conference on Earth Observation for Global Changes, 2009.
- [9] D. Saqui, J. H. Saito, J. R. Campos and L. A. C. Jorge, "Approach based on fuzzy c-means for band selection in hyperspectral images," International Journal of Computer, Electrical, Automation, Control and Information Engineering, vol. 10, n. 5, p. 738 – 744, 2016.
- [10] C. I. Chang, and S. Wang, "Constrained band selection for hyperspectral imagery," in IEEE Transactions on Geoscience and Remote Sensing, 2006.
- [11] X. Kang, S. Li and J. A. Benediktsson, "Feature extraction of hyperspectral images with image fusion and recursive filtering," IEEE Transactions on Geoscience and Remote Sensing, vol. 52, n. 6, p. 3742–3752, 2014.
- [12] S. A. Robila, and P. K. Varshney, "Feature Extraction from Hyperspectral Data Using ICA," in Advanced Image Processing Techniques for Remotely Sensed Hyperspectral Data, Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, p. 99–216.
- [13] P. H. Hsu, "Feature extraction of hyperspectral images using wavelet and matching pursuit," Journal of Photogrammetry and Remote Sensing, vol. 62, n. 2, p. 78 – 92, 2007.
- [14] H. Su, P. Du and Q. Du, "Hierarchical band clustering for hyperspectral image analysis," Pattern Recognition in Remote Sensing (PRRS), pp. 1-4, 2012.
- [15] A. Martínez-Uso, F. Pla, J. M. Sotoca and P. García-Sevilla, "Clustering-based hyperspectral band selection using information measures," IEEE Transactions on Geoscience and Remote Sensing, vol. 45, n. 12, p. 4158–4171, 2007.
- [16] H. Du, H. Qi, X. Wang, R. Ramanath and W. E. Snyder, "Band selection using independent component analysis for hyperspectral image processing," Applied Imagery Pattern Recognition Workshop, p. 93–98, 2003.
- [17] G. M. Petrie, P. G. Heasler and T. Warner, "Optimal band selection strategies for hyperspectral data sets," Geoscience and Remote Sensing Symposium Proceedings, vol. 3, p. 1582–1584, 1998.
- [18] P. Groves, and P. Bajcsy, "Methodology for hyperspectral band and classification model selection," Advances in Techniques for Analysis of Remotely Sensed Data, p. 120–128, 2003.
- [19] X. Zhang, W. Wang, Y. Li and L. C. Jiao, "Pso-based automatic relevance determination and feature selection system for hyperspectral image classification," Electronics Letters, vol. 48, n. 20, p. 1263–1265, 2012.
- [20] M. Mitchell, An Introduction to Genetic Algorithms, Cambridge, MA, USA: MIT Press, 1998.
- [21] T. Boggs, "Spectral Python (SPy)," 2014. [Online]. Available: <http://www.spectralpython.net/>. [Accessed on Junem 2016].
- [22] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," in IEEE Transactions on Image Processing, 2004.
- [23] D. A. Landgrebe, "Multispectral data analysis: A signal theory perspective," 1998. [Online]. Available: https://engineering.purdue.edu/~biehl/MultiSpec/Signal_Theory.pdf. [Accessed on March 24, 2016].
- [24] T. M. Lillesand, R. W. Kiefer and J. W. Chipman, Remote Sensing and Image Interpretation, John Wiley & Sons, 2002.
- [25] F. A. Fortin, F. M. D. Rainville, M. A. Gardner, M. Parizeau and C. Gagné, "DEAP: Evolutionary algorithms made easy," Journal of Machine Learning Research, vol. 13, p. 2171–2175, [Accessed on July 20, 2016].
- [26] "Hyperspectral Remote Sensing Scenes," Grupo de Inteligencia Computacional de la Universidad del País Vasco (UPV/EHU), [Online]. Available: http://www.ehu.es/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes. [Accessed on July 29, 2016].

Approach Based on SPEA2-Band Selection and Random Forest Classifier to Generate Thematic Maps from Hyperspectral Images

Completed Research

Diego Saqui

Federal University of São Carlos (UFSCar),
Federal Institute of Mato Grosso do Sul
(IFMS)

diego.saqui@ufscar.br

Daniel C. Lima

Federal University of São Carlos (UFSCar)
daniel.lima@dc.ufscar.br

José H. Saito

Federal University of São Carlos (UFSCar),
University Center of Campo Limpo
Paulista (UNIFACCAMP)

saito@dc.ufscar.br

Lucio A. de C. Jorge

Embrapa Instrumentation
lucio.jorge@embrapa.br

Steve T. M. Ataky

Université du Québec à Montréal (UQAM)
steve.ataky@nca.ufma.br

Abstract

Hyperspectral images (HIs) and segmentation have become a promising solution for different applications such as the production of thematic maps (TMs) of agricultural areas. However, problems such as the Hughes phenomenon and high demand for computational resources are related to the high number of bands of HIs. This study proposes a hybrid approach of Strength Pareto Evolutionary Algorithm 2 (SPEA2) and Random Forest classifier for producing TMs, aiming at band selection and improvement of average recall of segmentation. In experiments, the proposed approach reduced the number of bands on average from 220 to 30 in the Indian Pines image and from 224 to 42 in the Salinas image. The proposed approach was statistically whether identical or better than other approaches regarding the average recall of segmentation. Therefore, the proposed approach is promising as regards band selection and competitive in segmentation being a potential tool for generating TMs.

Keywords

Thematic Maps, Band Selection, Hyperspectral images, Remote Sensing.

Introduction

Hyperspectral images (HIs) contain spectral information of materials beyond the visible range of the spectrum. Each pixel can be represented by a spectral signature composed of a large number of bands that vary according to the kind of the object, allowing tasks of identification and analysis (Chang 2003). Among other types of images, HIs have the advantage of having unique bands that enable discrimination of different types of materials more easily (Attas et al. 2003; Khan 2018). Because of this capability, they are useful for medical imaging, remote sensing (RS), agricultural monitoring, among other applications (Gong, Zhang and Yuan 2015).

Applications with images and RS can occur by thematic maps (TMs). TMs show the spatial distribution of identifiable earth surface features; it provides an informational description over a given area, rather than a data description. The themes can range, for example, from categories such as soil, vegetation, and surface water in a general description of a rural area, to different types of soil, vegetation, and water depth or clarity for a more detailed description (Schowengerdt 2007). In particular, in agricultural monitoring, TMs can

help identify crop problems, such as the presence of pests, and prevent pesticides from being applied in excess, impacting the economy and preserving the environment. TMs facilitate the understanding of the observed area and the communication between experts from different specialties. Figure 1 shows an example of TM for Indian Pines HI, the ground truth (GT) of this image that is provided by a specialist and represents the actual region of the captured scene and their respective classes.

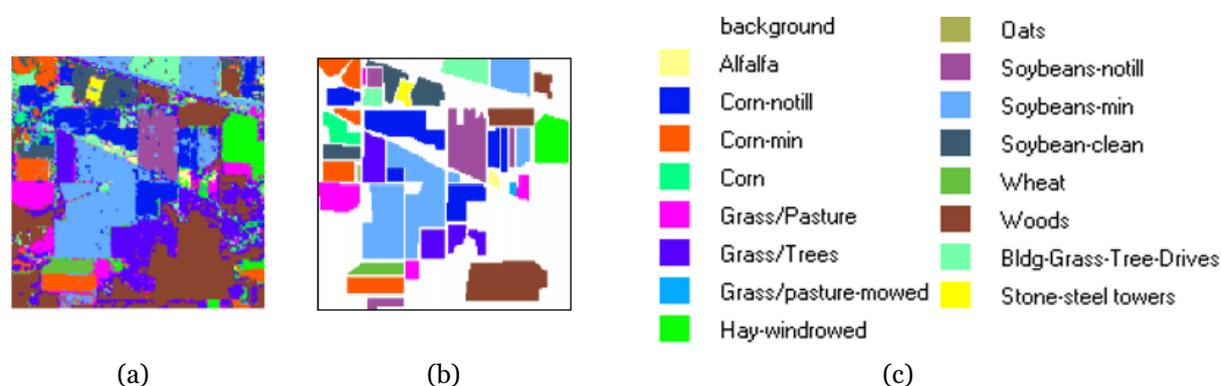


Figure 1. (a) Thematic Map. (b) Ground Truth. (c) Classes of Indian Pines HI. Adapted from Tinoco et al. (2013).

The Indian Pines HI in Figure 1 was captured by RS and was obtained from an agricultural area. This HI is used for large-scale monitoring and this monitoring is more efficient than manual methods, especially with the expansion of RS devices as drones. Therefore, applications that are capable of generating TMs are promising for agriculture and other areas that need monitoring.

For creating TMs, algorithms with peculiar statistical properties are required for classification and segmentation. The classification categorizes a pixel or a set of the pixel of an image in its corresponding area. Some examples of classifiers are Support Vector Machines (SVM), Gaussian Maximum Likelihood Classifier (GMLC) (Lillesand, Kiefer and Chipman 2014), and Random Forest (RF) (Dinç and Aygün 2013; Amini et al. 2018; Xia et al. 2018; Sun et al. 2018). The pixels of the same category when connected form regions called segments, that is, it characterizes the segmentation.

Although TM generated from HIs have many advantages, some challenges are related to the data volume: (a) information redundancy due to the high correlation of neighboring bands; (b) the curse of dimensionality; (c) high demand for computational resources, such as storage, processing and data transmission; and (d) the high cost of sensors to capture HIs. Because of these problems, the selection of relevant bands for further classification and thematic mapping process is required.

Methods that reduce features of the data are described as dimensionality reduction and can be categorized into feature extraction and feature selection (Kumar 2004). Feature selection is more feasible for dimensionality reduction in HIs because it preserves the physical information of the original data (Gong, Zhang and Yuan 2015) and for this purpose, it is called band selection. Another categorization divides feature selection into Filter, Wrapper or Embedded approaches.

Filter approaches score variables and then eliminate some before constructing a model and the wrapper approaches use a predictive measure of the model, such as accuracy, and directly evaluate the value of the feature set (Tuv et al. 2009). Owing to the fact that wrappers approaches are directly related to the model, they can produce better results than filters, on the other hand, they demand for more computing resources.

Some examples of these approaches are based on Genetic Algorithms (GAs) (Zhang, Sun and Li 2009; Saqui et al. 2016; Vaiphasa et al. 2017; Zhuo et al. 2008) and Particle Swarm (Zhang et al. 2012; Zang, Ma and Gong 2017), that can return satisfactory results during a given number of iterations. These algorithms operate with a single objective function which is usually a metric that leads to a good segmentation such as pixel classification (for wrappers) or region separation (for filters). They can also combine one of these metrics with reducing the number of bands as shown in Zhuo et al. (2008), but the problem is that they are

subject to unbalanced solutions because of the existence of conflicting objectives such as reducing the number of bands and improving the classification.

Multiobjective optimization band selection (MOBS) search for balanced solutions that improve segmentation and reduce the number of bands simultaneously (Gong, Zhang and Yuan 2015). Different filters MOBS can be found in the literature such as Non-dominated Sorting Genetic Algorithm 2 (NSGA2) (Kumar 2004) and Multiobjective Evolutionary Algorithm Based on Decomposition (MOEA/D) with Tchebycheff Decomposition (Gong, Zhang and Yuan 2015; Xu, Shi and Pan 2017). Another algorithm that can be applied as filter-based feature selection is the Strength Pareto Evolutionary Algorithm 2 (SPEA2) (Xue et al., 2013; Zitzler, Laumanns and Thiele, 2001). Studies have shown that SPEA2 can operate better than the traditional NSGA2 for a small number of iterations, small population, and noisy data as obtained by sensors, that is, HIs (Bui et al. 2005; Chlodowicz and Orlowski 2017).

Embedded approaches use all the variables to generate and analyze a model to infer the importance of the variables (Tuv et al. 2009). Examples of embedded approaches are decision trees (DTs) and RF. RF is a supervised algorithm which can produce promising results in HI segmentation (Dinç and Aygün 2013; Amini et al. 2018). A hybrid approach of DTs and feature selection with GA showed superior results about the use of the DT approach without prior band selection technique. The authors suggest that this is because the hybrid approach can focus on relevant features (Stein et al. 2005). This performance improvement may have occurred due to feature selection before the application of RF (Tuv et al. 2009; Sun et al. 2018).

In this paper, we propose a hybrid approach of SPEA2-based band selection (wrapper) and RF classifier to generate TMs. The proposed approach is modeled as a supervised MOBS dealing with the high dimensionality of HIs by band selection and maintaining the average recall performance of segmentation. Therefore, the main purposes of this study are:

- Select relevant bands of HIs with the proposed approach;
- Produce TMs with the selected bands;
- Structure a supervised MOBS to search tradeoff solutions;
- Evaluate the proposed approach by comparing it with GA-based supervised approach that combines two metrics (Zhuo et al. 2008) and the direct application of RF in HIs with all bands.

The remainder of this paper is organized as follows: The second section presents a detailed description of this approach. The experiments are conducted and results are reported in the third section. At last, the main contribution of this paper is summarized and future work direction is pointed.

Approach based on SPEA2-band selection and Random Forest

Multiobjective Optimization Band Selection (MOBS)

GA is an adaptive optimization search methodology based on Darwin's theory of natural selection and genetics in biological systems. It is an alternative to conventional heuristic methods.

GAs can represent solutions (individuals) to a problem by a vector, and an example of representation for HIs and bands is to use the number 1 for presence and 0 for the absence of the band equivalent to the position of these values in the vector. With this representation, the GA performs a search process trying different combinations of the bands in the solutions in order to improve the value obtained from the so-called fitness function.

GA can be started with a random set of solutions and try to optimize them by some iterations (called generations) or until meet a requirement. The optimization happens through the selection of the solutions with better values obtained from the fitness function and the application of crossover and mutation operators. The crossover is the action of combining parts of solutions, generating new solutions. Mutation consists of changing the values of solutions (such as 0 to 1) and can be applied so that the search process does not get stuck in local optima (Zhuo et al. 2008).

Usually, the fitness function used in GA can improve a single metric, such as overall accuracy, entropy or other metrics related to segmentation. This function can also be used to combine two metrics as shown in Zhuo et al. (2008). One problem with this single objective functions is that they may not be able to operate with conflicting metrics. For example, a reduction in the number of bands cannot occur if only a fitness

function for segmentation improvement is applied. This happens because, during the search process, bands that improve and not improve the segmentation be present together in the solutions and the latter are never eliminated since the fitness function is not designed for this. MOBS were proposed to try to solve these problems since they are suitable to operate with conflicting objectives searching for tradeoff solutions.

The term MOBS was the first time introduced by Gong, Zhang and Yuan (2015) to refer to multiobjective algorithms of feature selection applied in the band selection of HIs but a similar study had already been developed by Kumar (2005). The main algorithms found in this category were the MOEA/D and the NSGA2. All these approaches operated with metrics that characterize them as filters, that is, they did not incorporate a supervised classification algorithm during their execution. Another common feature among these studies is the use of the Pareto frontier (PF) theory.

The best solution set of the search space composes the PF and solutions of this set cannot be said better among each other. For the establishment of the PF, it is necessary to compare all solutions in the search set and for this, algorithms like the NSGA2 and SPEA2 use the non-dominance relation given in (1) that means:

- The solution x dominates the solution y because x is better than y in at least one objective and equal or better in all other objectives

$$x \succ y \quad (1)$$

Solutions not dominated by any other are from the PF.

SPEA2, which is of interest in this study, is also a multiobjective optimization algorithm that can be modeled as MOBS. Reports in the literature, describe that this algorithm works well for noisy data, as in the case of images acquired by sensors and therefore is explored in this study.

For the establishment of a ranking of the solutions of the PF and other solutions the SPEA2 uses a specific function, where, first for each solution x , the raw fitness $R(x)$ in (2) is calculated.

$$R(x) = \sum_{y \succ x} S(y) \quad (2)$$

$S(y)$ is strength value, given by the total number of solutions that y dominates, established using (1), in the population and the archive. Then, SPEA2 uses a density estimation technique represented in (3).

$$D(x) = 1/(\sigma_x^k + 2) \quad (3)$$

where σ_x^k is the Euclidean distance of x concerning to the k -th solution of a list of distances. This list is in ascending order and contains the distances of all solutions in archive and population relative to x . k is given by the square root of the sum of the population size and archive size.

Finally, the value of each solution used in the selection operation is calculated by the function shown in (4) (Zitzler, Laumanns and Thiele, 2001).

$$F(x) = R(x) + D(x) \quad (4)$$

An important observation is that (1) is established with a set of fitness functions that must be previously established, such as the function to evaluate the segmentation and another to evaluate the number of bands.

One of the functions used in this study is based on the average recall that to be calculated it is necessary to apply an algorithm for classification where the RF is used.

Random Forest (RF)

RF is an algorithm that creates an ensemble of DTs to obtain a more accurate and more stable prediction (Breiman 2001). Supervised algorithms such as C4.5 can model DTs, but it is required a set of training samples with defined classes provided by a specialist. The modeling usually happens by evaluating how important a feature (or band) is for classification in those samples, and metrics used for this evaluation are

the Gini impurity and Information Gain Entropy. In addition to the classification, the RF also performs a feature selection while structuring each leaf node of the tree and therefore it is explored in this study.

Proposed approach

In this approach, the SPEA2 conducts the band selection by some iterations (or generations). In each generation, it establishes a population and an external archive that contains solutions that are analyzed in the search process. A binary vector characterizes each solution/HI.

The following fitness functions are applied to each solution:

- **Fitness 1:** Increasing the average recall of the pixel classification performed by RF in images generated by each solution as shown in (5),

$$\text{Average Recall} = \frac{1}{n} \sum_{i=1}^n tp_i / (tp_i + fn_i) \quad (5)$$

where n is the number of classes, i is the class, tp_i is the total of pixels correctly classified, and fn_i is the total of pixels erroneously classified.

The recall is the rate of correctly classified pixels in each class and if this value is bad for some class then the average recall of all classes can be penalized. Therefore, improving the average recall of the classes implies increasing the segmentation performance, and so this metric was chosen as fitness.

- **Fitness 2:** Reducing the number of bands of HIs where it is considered the smallest possible value.

The process of the proposed approach is adapted from the original SPEA2 and shown in Algorithm 1.

Algorithm 1 SPEA2-based approach for band selection

Input: N (population size), \bar{N} (archive size), T (number of generations), GT , $OrigHI$ (HI with all bands), M (mutation rate), and R (recombination rate).

Output: Solution with best average recall of the last generation.

Step 1 - Initialization: Creates an initial population P_0 and empty external archive $\bar{P}_0 = \emptyset$. Set $t = 0$. For P_0 are generated solutions with random combinations of bands from $OrigHI$.

while $t \leq T$ **do:**

Step 2* - Classification: Train and classify the pixels of HIs generated from new solutions using RF.

Step 3 - Fitness assignment: Establish fitness values of solutions in P_t and \bar{P}_t , calculating the average recalls in (5), the number of bands, and the fitness values in (4).

Step 4 - Environmental selection: Copy all nondominated solutions in P_t and \bar{P}_t to \bar{P}_{t+1} . If size of \bar{P}_{t+1} exceeds \bar{N} then reduce \bar{P}_{t+1} by means of the truncation operator, otherwise if size of \bar{P}_{t+1} is less than \bar{N} then fill \bar{P}_{t+1} with dominated solutions in P_t and \bar{P}_t .

Step 5 - Mating selection: Performs binary tournament selection with replacement on \bar{P}_{t+1} to fill the mating pool.

Step 6- Variation: Apply recombination and mutation to the mating pool and set P_{t+1} to the resulting population. Increment the counter ($t = t + 1$).

Step 7- Select one solution: Selects the solution with best average recall from \bar{P}_t .

In algorithm 1, initially, population and archive sizes, HI with all bands ($OrigHI$), GT , and mutation and recombination rates are defined. In step 1, the population of solutions is created with a random combination of bands generated from HI with all bands.

Then, in step 2, for each solution, 3-fold cross-validation is applied for training and validation, and the average value of the average recall obtained is used as fitness. The samples (pixels) are obtained by systematic sampling from the GT through an iterative process by selecting pixels alternately for each fold. The training samples are used to establish a model based on the RF, which is later validated with the validation samples. *The use of the RF in step 2 is the main modification about the other MOBS proposed in the literature characterizing the proposed approach as the wrapper.

In step 3, fitness values and non-dominance relationships between all solutions are established.

In step 4 the environmental selection is performed, where all non-dominated solutions of the population and archive are copied to the new archive of the next generation. If this archive is smaller than the predetermined size, the dominated solutions are sorting according to (4) and then copied to fill up to the archive size. If the next generation archive has the number of non-dominated solutions that exceeds the predefined size, then the solutions to be removed are those that have less distance σ_x^k than others. If there are solutions with the same distance, SPEA2 considers the second minimum distance, and so on.

In step 5 the binary tournament selection for mating is performed and in step 6 the recombination and mutation operations are applied to compose the population of the next generation. A predetermined rate determines the occurrence of recombination between solutions, and in this paper, the one-point crossover is used. The mutation operation occurs in new solutions, where we apply the flip bit that allows changing the value of each vector position (from 1 to 0 or 0 to 1) based on a predetermined rate. In step 7 the solution that generated the image with the best average recall of the PF of the current population is selected.

After applying these steps, the selected bands can be used to construct an RF model that can later be applied to segmentation and generation of TM.

Experiments and results

Experiments

There is a limitation of free datasets of HIs with GT available, so for this study, two traditional HIs of agricultural areas were used, Indian Pines and Salinas. These HIs were obtained through the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) and correspond mainly to land covers with crops. Table I shows information about these datasets.

Name	# of pixels	# of classes	# of bands
Indian Pines	21025	16	220
Salinas	7138	6	224

Table 1. Information about datasets.

The proposed approach was compared with the GA-based approach that searches the best combination of bands with the average recall fitness as shown in equation (6).

$$f = w_a * average\ recall + (w_n / number\ of\ bands) \quad (6)$$

where w_a and w_n are the weights for the average recall and number of bands and values of 0.8 and 0.2 were used, respectively. This was the only wrapper approach of band selection found combining two metrics and therefore was chosen for comparison.

Both approaches were configured to run for 500 generations and with a population of 100 solutions. The recombination/crossover and mutation rates were 50% and 5% respectively. The recombination/crossover and mutation operation of the two approaches are the same, and the binary tournament selection was used.

For RF classifier, ten DTs generated from the Gini indices (measures the degree of heterogeneity of each node) and bootstrap samples were used. In the RF application, the rules previously presented of training and validation were used, as well as the criteria for selecting samples.

We also compared the proposed approach with the results of segmentation of GA-based approach and the original HIs with all bands. In this comparison, we used the last generation solution of each approach and evaluated the average recall and the accuracy of the segmentation.

The process of band selection (training and validation) and tests was organized in a 10-fold cross-validation for each image. We use 80% of the pixels for the band selection and 20% for final tests. The 20% pixels of each test fold were selected to be different from each other, and so was the 80% pixel of each fold used in

band selection. Based on this structure, the different band selection algorithms were applied and after the band selection, statistics were obtained from the test data.

The algorithms used were developed using the Python 3.5 language and OpenCV, Scikit-Learn, Spectral Python and Distributed Evolutionary Algorithms in Python (DEAP) packages.

Results

The following graphs show the average evolution of the solution with the best average recall in each generation obtained from the ten executions for the proposed approach (with SPEA2 and RF) and the GA-based approach. These graphs help analyze the behavior of each approach in the band selection process.

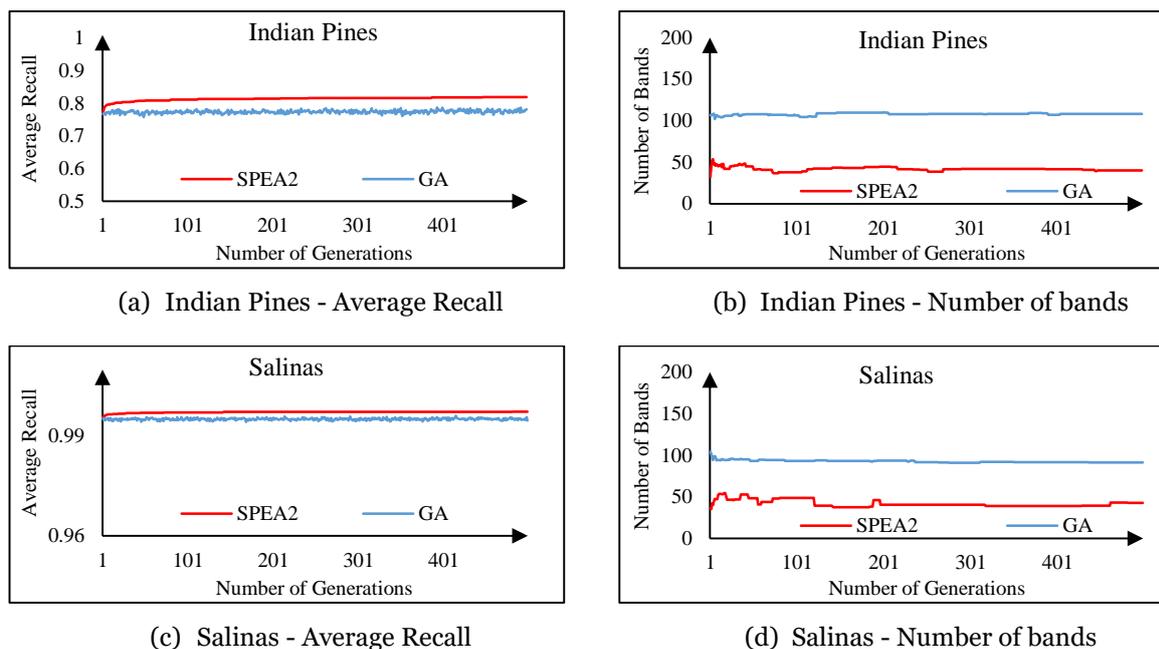


Figure 2. Graphs of the evolution of the average recall and number of bands for Indian Pines and Salinas HIs with band selection performed by proposed approach and GA.

In the graph of Figure 2a, which refers to Indian Pines, it is shown that the proposed approach had an average evolution of the average recall better than the GA-based approach. The graph in Figure 2b shows the average evolution of the number of bands under the same conditions of the graph of Figure 2a, and when analyzed together, we can see that the proposed approach reduces the number of bands whilst practically maintaining the average recall, whereas the GA-based approach cannot reduce the number of bands. The improvement of the average recall and reduction of the number of bands performed by the proposed approach is noticeable mainly at the beginning of the generations (between the first and the one hundred fiftieth generation). This characterizes the behavior of MOBS in the proposed approach where tradeoff solutions were found.

Figures 2c and 2d show the average evolution of the average recall and the number of bands respectively for Salinas image. In these graphs the behavior of both algorithms for Salinas is very similar to the results of the Indian Pines, maintaining a high average recall value, and the proposed approach reduces the number of bands more than the GA-based approach.

An observation in Figures 2b and 2d where the number of bands is shown, for the proposed approach, the function is not monotonically decreasing because the best average recall solution is selected from the Pareto set of each generation.

The characteristics shown in the average evolution of these graphs (Figure 2) suggest an improvement in the treatment of the dimensionality and the removal of redundant bands without degrading the average recall. Under these conditions, the proposed approach was better than the GA-based approach.

In Figure 3 are shown the results of the average recall and overall accuracy for each HI with bands selected by the methods compared, obtained from the classification result by the algorithm GMLC. The values shown above the columns are the averages obtained from the 10-Fold cross-validation with their respective standard deviations.

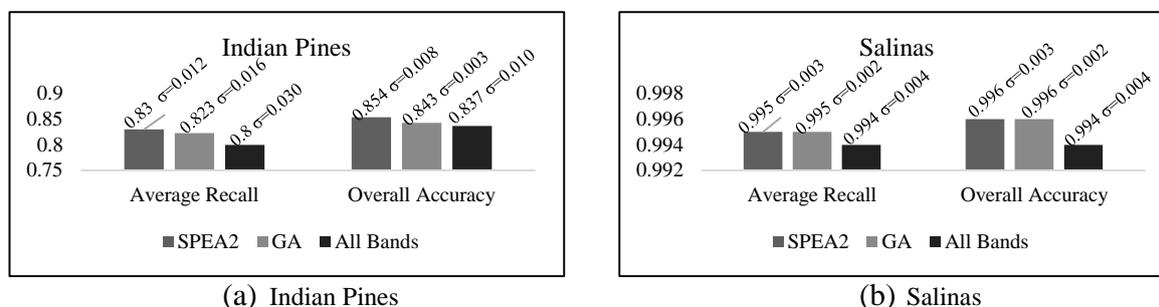


Figure 3. Average values of average recall and overall accuracy obtained from the classification result by the GMLC.

A paired t-test with the significance of 0.05 was applied in the average recall and overall accuracy, and, statistically, the proposed approach had the same mean as the other cases compared, thus equal performance for these metrics. Just for Indian Pines image, the proposed approach was statistically better than the use of all bands for overall accuracy.

Name	Indian Pines	Salinas
Proposed Approach	39	42
GA	108	91
HI with all bands	220	224

Table 2. Average number of bands selected at the end of the executions of each approach and total of original bands.

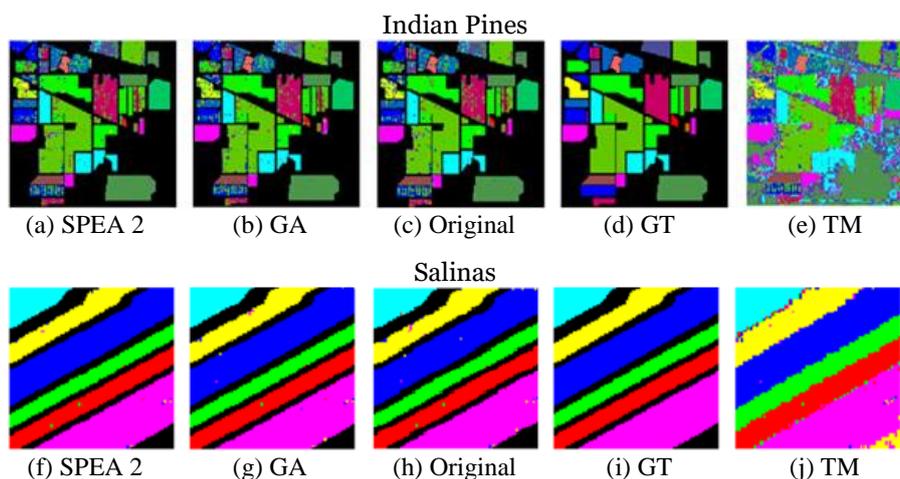


Figure 4. HIs segmented: (a) Indian Pines with bands selected by SPEA2. (b) Indian Pines with bands selected by GA (c) Indian Pines with all bands. (d) GT Indian Pines. (e) TM for Indian Pines. (f) Salinas with bands selected by SPEA2. (g) Salinas with bands selected by GA (h) Salinas with all bands. (i) GT Salinas. (j) TM for Salinas

The main advantage of the proposed approach was in the reduction of bands and the final average number of bands selected by each approach is shown in Table II. Together with the results obtained in the pixel classification, it shows that the proposed approach can be applied to generate thematic maps from HIs.

Figure 4 shows examples of TM obtained from the band selection of the first execution of each approach. In Figures 4a-4c and 4f-4h the segmented images of Indian Pines and Salinas are respectively shown. The areas not present in the GT were removed. Figures 4d and 4i, respectively show the GT of the Indian Pines and Salinas. These regions are important because they are areas of interest in thematic maps shown in Figures 4e and 4j. The result of this segmentation is obtained from the training of RFs with 30% of the pixels of each class, which were replaced in the HIs for later classification.

The results shown in this section corroborate that the proposed approach can improve the selection of bands performed by the RF, contributing to a good classification result and allowing the generation of TMs.

Conclusion

A novel supervised MOBS using SPEA2 and RF for the production of thematic maps from HIs has been described herein. Given the relationship between quality of segmentation and number of bands, the idea of the approach is to search for tradeoff solutions with the objectives of improving the average recall of the pixel classification and reduction of the number of bands. Therefore, the feature space used by RF to generate their DTs can be optimized reducing the number of bands for classification without loss of quality in the segmentation.

The proposed approach was compared with the direct application of RF as well as a GA-based approach. Experiments on the Indian Pines and Salinas HIs showed that the proposed approach was competitive in the average recall and outperformed in the reduction of the number of bands. Therefore, the approach can be applied to generate thematic maps from HIs to reduce the computational cost of processing these images without loss of quality.

As the approach has always selected the best average recall solution of the Pareto Frontier, we believe that a decision-making algorithm can improve the performance of reducing the number of bands and will be exploited in future works. Other subjects for future works are implementing a strategy to correct misclassified pixels in the segments, testing other fitness functions with metrics such as accuracy, precision, recall, F1-score and Area Under the Curve (AUC) for the Receiver Operating Characteristic (ROC) and making comparisons with other approaches in the literature.

REFERENCES

- Amini, S., Homayouni, S., Safari, A., Darvishsefat, A. A. 2018. "Object-based classification of hyperspectral data using Random Forest algorithm". In: *Geo-spatial Information Science*, vol. 21, pp. 127-138.
- Attas, M., Cloutis, E., Collins, C., Goltz, D., Majzels, C., Mansfield, J. R., Mantsch, H. H. 2003. "Near-infrared spectroscopic imaging in art conservation: Investigation of drawing constituents", *Journal of Cultural Heritage*, vol. 4, no. 2, pp. 127-136.
- Breiman, L. 2001. "Random forests". In: *Machine Learning*, vol. 45, no. 1, pp. 5-32.
- Bui, L. T., Essam, D., Abbas, H. A., Green D. 2005. "Performance analysis of evolutionary multiobjective optimization methods in noisy environments". In: *Proceedings of The 8th Asia Pacific Symposium on Intelligent and Evolutionary Systems*, pp. 29-39.
- Chang, C. I. 2003. *Hyperspectral Imaging: Techniques for spectral Detection and Classification*. Springer Science & Business Media.
- Cholodowicz, E., Orłowski, P. 2017. "Comparison of SPEA2 and NSGA-II Applied to Automatic Inventory Control System Using Hypervolume Indicator". *Studies in Informatics and Control*, vol. 26, pp. 67-74.
- Dinç, S., Aygün, R. S. 2013. "Evaluation of Hyperspectral Image Classification Using Random Forest and Fukunaga-Koontz Transform". In: *International Workshop on Machine Learning and Data Mining in Pattern Recognition. MLDM 2013: Machine Learning and Data Mining in Pattern Recognition. Lecture Notes in Computer Science*, vol 7988. Springer, pp. 234-245.
- Gong, M., Zhang, M., Yuan, Y. 2015. "Unsupervised Band Selection Based Evolutionary Multiobjective Optimization for Hyperspectral Images". In: *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, n. 1, pp. 544- 577.

- Khan, M. J., Khan, H. S., Yousaf, A., Khurshid, K., Abbas, A. 2018. "Modern Trends in Hyperspectral Image Analysis: A Review". In: IEEE Access, vol. 6, pp. 14118-14129.
- Kumar, M. 2004. "Feature Selection for Classification of Hyperspectral Remotely Sensed data using NSGA-II". In: Water Resources Seminar, Citeseer.
- Lillesand, T. M., Kiefer, R. W., Chipman, J. W. 2004. "Remote Sensing and Image Interpretation", 5th ed., Ed. John Wiley & Sons.
- Saqui, D., Saito, J.H., Jorge, L.A.C, Ferreira, E.J., Lima, D.C., Herrera, J.P. 2016. "Methodology for band selection of Hyperspectral Images using Genetic Algorithms and Gaussian Maximum Likelihood Classifier". In: International Conference on Computational Science and Computational Intelligence, Las Vegas, EUA, pp. 733-738.
- Schowengerdt, R. A. 2007. *Thematic Classification*, Remote Sensing (Third Edition), Academic Press.
- Stein, G., Chen, B., Wu, A. S., Hua, K. A. 2005. "Decision tree classifier for network intrusion detection with GA-based feature selection". In: Proceedings of the 43rd Annual Southeast Regional Conference on - ACM-SE 43.
- Sun, M., Wang, C., Wang, S., Zhao, Z., Li, X. 2018. "A New Semisupervised-Entropy Framework of Hyperspectral Image Classification Based on Random Forest". *Advances in Multimedia*, vol. 2018.
- Tinoco, S., Santos, A., Santos, H., Menotti, D., Santos, J. A. 2013. "Ensemble of classifiers for remote sensed hyperspectral land cover analysis: An approach based on linear programming and weighted linear combination", *Proc. IGARSS*, pp. 4082-4085.
- Tuv, E., Borisov, A., Runger, G. C., Torkkola, K. 2009. "Feature selection with ensembles, artificial variables, and redundancy elimination". *Journal of Machine Learning Research*, vol. 10, pp. 1341-1366.
- Vaiphasa, C., Skidmore, A. K., Boer W. F., Vaiphasa, T. 2007. "A hyperspectral band selector for plant species discrimination". *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 62, n. 3, p. 225-235.
- Xia, J., Ghamisi, P., Yokoya, N., Iwasaki, A. 2018. "Random Forest Ensembles and Extended Multi-Extinction Profiles for Hyperspectral Image Classification". In: *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, n. 1.
- Xu, X., Shi, Z., Pan, B. 2017. "A New Unsupervised Hyperspectral Band Selection Method Based on Multiobjective Optimization". *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 11, pp. 2112-2116.
- Xue, B., Cervante, L., Shang, L., Browne, W., N., Zhang, M. 2013. "Multi-objective evolutionary algorithms for filter based feature selection in classification". *International Journal on Artificial Intelligence Tools*, vol. 22.
- Zhang, M., Ma, J., Gong, M. 2017. "Unsupervised Hyperspectral Band Selection by Fuzzy Clustering with Particle Swarm Optimization". *IEEE Geoscience and Remote Sensing Letters*, vol. 14, n. 5, pp. 773-777.
- Zhang, X., Sun, Q., Li, J. 2009. "Optimal band selection for high dimensional remote sensing data using genetic algorithm". In: *Proc. SPIE 7471, Second International Conference on Earth Observation for Global Changes*.
- Zhang, X., Wang, W., Li, Y., Jiao, L. C. 2012. "Pso-based automatic relevance determination and feature selection system for hyperspectral image classification". *Electronics Letters*, vol. 48, n. 20, pp. 1263-1265.
- Zhuo, L., Zheng, J., Wang, F., Li, X., Ai, B., Qian, J. 2008. "A Genetic Algorithm Based Wrapper Feature Selection Method for Classification of Hyperspectral Images Using Support Vector Machine". In: *Proceedings of SPIE 7147, Geoinformatics 2008 and Joint Conference on GIS and Built Environment: Classification*.
- Zitzler, E., Laumanns, M., Thiele, L. 2001. "SPEA2: Improving the strength Pareto evolutionary algorithm". In: *Evolutionary methods for design optimization and control with applications to industrial problems*, Athens, Greece, pp. 95-100.

NSGA2-based method for band selection for supervised segmentation in hyperspectral imaging

Diego Saqui^{1,2}, José H. Saito^{1,3}, Daniel C. de Lima¹, Lucio A. de C. Jorge⁴,
Ednaldo J. Ferreira⁴, Steve T. M. Ataky⁵ and Francisco Fambrini¹

Abstract— The identification of crop areas marked by biotic and abiotic factors is a challenge in precision agriculture (PA). Remote sensing and supervised segmentation of hyperspectral images (HIs) have become a fast and precise solution of PA for the identification of problems in crops. However, well-known problems for knowledge extraction in HIs are related to data volumes and capacity for modeling each phenomenon due to the so-called curse of dimensionality. Some techniques for dealing with dimensionality, such as genetic algorithms (GAs), are promising, but they cannot assure a compromise between the segmentation effectiveness and reduction of dimensionality. In this paper, a multi-objective method based on Nondominated Sorting Genetic Algorithm 2 and Gaussian Maximum Likelihood Classifier is proposed for supervised segmentation of HIs dealing with the problem of dimensionality. Experiments with three datasets and cross-validation showed that the proposed method reduced the average number of bands between 75.4% and 88.2% for each evaluation. In addition to reducing the number of bands, statistical tests showed that our method had a pixel classification performance equivalent or more significant than other state-of-the-art methods. Based on the results, the method is promising for band selection in HIs and can contribute to segmentation.

I. INTRODUCTION

Nowadays, manual agricultural monitoring of crop conditions is a difficult task, but it can be improved with precision agriculture (PA) techniques. A useful tool in PA is the monitoring by images and sensors, called by Remote Sensing (RS). RS is "the science of obtaining information about an object, surface area, or phenomenon, through the analysis of data acquired by a sensor device that is not in contact with what is being observed" [1].

The evolution of technologies such as drones, sensors, and analytical software have encouraged research and large-scale application of RS in agriculture. RS tools can lead to increase in productivity and reduction in the price of agricultural products [2]. This technology is useful for obtaining hyperspectral images (HIs) that contain bands that are not available in

other images (multispectral and RGB). Some of these bands allow the discrimination of materials more efficiently [3] [4]. Therefore, the application of algorithms for discrimination of materials using HIs can promote superior performance than other types of images. This feature enables, for example, the identification and analysis of vegetation type [1], crop health [3], soil and others biotic and abiotic changes [4].

A technique for analyzing crops in HIs is the segmentation, which makes specific characteristics distinguishable. Segmentation is the process of subdivision of the image into its constituent regions [5]. In this sense, segmentation of aerial HIs can be modeled as part of a supervised process based on previous ground truth (GT) maps. Supervised segmentation implies classifying and generalizing intra-segment responses aiming to provide a map with cohesive areas. Thus, supervised segmentation can be obtained from a direct classification where connected pixels of the same class defines a new segment. Some algorithms that can be used for the pixel classification and subsequent segmentation of HIs are Support Vector Machines, Artificial Neural Networks and the Gaussian Maximum Likelihood Classifier (GMLC) [1].

Due to the high dimensionality of HIs some problems are common such as, Hughes's phenomenon [6]; difficulty in being processed in computational devices with limited resources; information redundancy caused by the high correlation [7] and the high cost of sensors [8]. Therefore, segmentation and classification algorithms, when applied in HIs, are often preceded by dimensionality reduction strategies.

Dimensionality reduction methods are categorized in feature selection and feature extraction [9]. Feature selection is the problem of selecting a subset of the original features, resulting in a smaller number of features. In contrast to feature extraction that performs a feature space modification to a new space; in feature selection, the original representation of the variables is not changed. Feature selection is preferred when the meaning of the features is demanded [10]. For HIs, there is an interest in maintaining the meaning of the information because it can allow the construction of specific sensors and identify bands that characterize certain conditions of the crops; thus, the feature/band selection is preferred.

Genetic Algorithms (GA) and Particle Swarm are examples of methods used for band selection that establish a single objective function that usually consists of some measure of segmentation that is conducted individually or combined with the reduction of the number of bands [12] [13] [14]. This kind of methods makes difficult finding solutions for classifying pixels and reducing the number of bands so

*This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001 and São Paulo Research Foundation (FAPESP)- Project code: #2017/50270-5.

¹Diego Saqui, José H. Saito, Daniel C. de Lima and Francisco Fambrini are with the Federal University of So Carlos (UFSCar), São Carlos - SP, Brazil. diego.saqi@ufscar.br, saito@dc.ufscar.br

²Diego Saqui is also with the Federal Institute of Mato Grosso do Sul (IFMS), Corumbá - MS, Brazil.

³José H. Saito is also with the University Center of Campo Limpo Paulista (UNIFACCAMP), Campo Limpo Paulista - SP, Brazil

⁴Lucio A. de C. Jorge and Ednaldo J. Ferreira are with Embrapa Instrumentation, São Carlos - SP, Brazil.

⁵Steve T. M. Ataky is with École de Technologie Supérieure (ÉTS), Montréal, Quebec, Canada.

that the optimization of an objective does not damage the other. The need to reduce the number of bands and improve segmentation performance makes the band selection problem to be characterized as multiobjective. For this multiobjective optimization band selection (MOBS) can be applied. These methods try to find balanced solutions that have a better performance for all objectives and usually use the Pareto Frontier (PF) concept where cannot be considered better between them but are better than others in the search space.

Different unsupervised MOBS, which can also be categorized as filters, can be found in the literature. Non-dominated Sorting Genetic Algorithm 2 (NSGA2) was proposed to optimize cumulative texture and spectral information of a band combination using a minimum number of bands [9]. Another common method that has been explored is the Tchebycheff Decomposition Strategy (TDS) for the multiobjective evolutionary algorithm based on decomposition. This method was explored to optimize the number of bands and entropy contained in the selected band subsets [15].

A state-of-the-art method called incorporated rank-based multiobjective band selection (IRMoBS) using TDS was proposed to optimize three objectives, band number, variance and information entropy [16]. Three traditional band selection methods, improved sparse subspace clustering-based, dissimilarity-weighted sparse self-representation-based, and dominant set extraction-based, were compared with IRMoBS which presented advantages for most cases tested.

NSGA2 used in some of these studies is an algorithm for multiobjective problems. This algorithm tries to find a set of solutions that optimize more than one objective simultaneously [9]. The set found contains candidate solutions to belong to PF and cannot be considered better about each other [17]. Therefore, it is often necessary to use a decision maker (DM) to select a solution from this set.

If the GT is available, supervised/wrapper methods can be used to fit a classification model based on a priori knowledge, and with this, it is expected to obtain better results during the selection of bands. Wrapper methods usually require a high computational cost but present better results to filter methods [18]. This type of method together with multiobjective algorithms may be an alternative to the band selection.

In this paper, the NSGA2 is modeled as a new wrapper method for MOBS of HIs of agricultural areas. This method deals with dimensionality problems, searching for balanced solutions and assisting in segmentation. The main difference compared to others in the literature is that our proposal is a supervised MOBS using a GT and a classifier. The method also consists of a DM that automatically selects a solution of the PF. The main contributions of this paper are:

- 1) A new supervised MOBS, based on NSGA2, GMLC and DM, since supervised MOBS are little explored.
- 2) Use of DM that is rarely discussed in MOBS.
- 3) Comparisons with other supervised (based on GA) and unsupervised (IRMoBS) literature methods.

II. NSGA2-BASED BAND SELECTION METHOD FOR SEGMENTATION

In this paper, the NSGA2 guides the band selection for some generations. In each generation, this method maintains a population with a set of solutions that are analyzed in the search for candidates to belong to the PF.

A vector characterizes each solution representing a HI with a specific combination of bands. This vector has the number of positions equal to the total of bands of original HIs, the number 1 indicates the presence of the band equivalent to that position and 0 otherwise.

For each solution two objectives are used and calculated:

- 1) Increase the average recall (also called balanced accuracy) of the pixel classification performed by GMLC;
- 2) Reducing the number of bands of the HIs.

Equation (1) represents the average recall that evaluates the correctly classified pixels in each class of the HIs segmented in comparison to the GT.

$$AverageRecall = \frac{1}{n} * \sum_{i=1}^n \left(\frac{tp_i}{tp_i + fn_i} \right) \quad (1)$$

where n is the number of classes, i is the class, tp_i is the total of pixels correctly classified, and fn_i is the total of pixels erroneously classified. The recall is the rate of correctly classified pixels in a class, and improving the average recall of the classes implies in increasing the classification performance. The average recall is useful for imbalanced datasets. For the reduction of bands, it is considered the least possible number of bands.

A comparison between the solutions of the current population is performed to find the set that can be part of the PF. To compare the solutions the NSGA2 uses a non-dominance relation. A solution x dominates a solution y if x is better than y in at least one objective and equal or better in the all others objectives. This relationship is shown in (2).

$$x \succ y \quad (2)$$

Solutions that are not dominated by any other are Pareto-optimal solutions, that is, they are present on the PF and frontier solutions cannot be ordered directly.

In NSGA2 the Fast Nondominated Sorting (FNS) algorithm establishes a rank of the solutions of the current population according to the level of non-dominance. First, the FNS determines the candidate solutions to belong to the PF; then it ranks the other solutions according to non-dominance. Solutions that are dominated by a smaller number of others are closer to the top of the ranking [17] [18].

For each solution, a value called crowding distance is calculated to estimate the density (how isolated) a solution is in your ranking and this value is used in the selection operation to apply cross and mutation later. In this study, as in the original NSGA2 proposal, the binary tournament selection operator based on the crowded-comparison operator is used. Crowded-comparison operator works in sorting solutions with the following rule:

- The first criterion is by PF and other rankings, and

- the second criterion is by the crowding distance.

Besides the crowded-comparison operator being used in the binary tournament selection operator, it is also used before establishing the population of the new generation, guaranteeing elitism.

A predetermined rate determines the occurrence of the crossover between solutions and in this paper is used the one-point. The mutation operation occurs in new solutions, where we apply the flip bit based on a predetermined rate.

A. GMLC

The GMLC algorithm used in the classification assumes that the pixels have Gaussian distribution and thus each class can be represented by mean and covariance. In the classification process, this algorithm uses a probability distribution function to verify the chance of a pixel belonging to each class. The class with the highest value probability will be determined for the new pixel [3].

The training and classification steps are applied to each solution of each generation of NSGA2. These procedures are necessary because each new solution is a HI with a different combination of bands and therefore the GMLC must be trained and to classify new data.

The pixels selected for training are used to calculate the mean and covariance of each class where these results are subsequently used to classify new pixels. The function $g_i(x)$ showed in eq. (3), is used to verify the probability of a pixel (represented by a vector x) belonging to the class w_i

$$g_i(x) = -\frac{1}{2} \ln \left| \sum_i i \right| - \frac{1}{2} (x - m_i)^t * \sum_i i * (x - m_i) \quad (3)$$

where i is the number of the class, \ln is the natural logarithm, $\sum_i i$ and m_i are the covariance matrix and mean vector of the data in class w_i . The following rule is used by the GMLC:

$$x \in w_i, \text{ if } g_i(x) > g_j(x) \text{ for all } j \neq i$$

Thus, the class with the highest value in eq. (3) will determine the class of each pixel.

B. Decision-Maker (DM)

Since NSGA2 can return a set of non-dominated solutions, it is necessary a DM to choose a solution automatically.

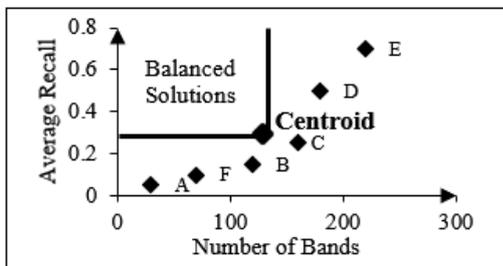


Fig. 1. Automated decision making. Adapted from [19].

In the proposed method we incorporate a strategy based on the centroid that is proposed in [19]. The centroid is calculated from the number of bands, and average recall of

the Pareto-optimal solutions acts as a threshold to create a region of balanced solutions.

Fig. 1 shows an example of the region of balanced solutions based on the centroid and that the solution C, which is outside the balanced region, but it is the choice of DM because it is the closest to the centroid.

The following rule is used to decide the chosen solution:

- If there are solutions in the balanced region then the solution in that region closest to the centroid is selected;
- otherwise, the solution of all Pareto-optimal solutions found closest to the centroid of the region is selected.

C. Process summary

The process summary of the proposed method is shown in the flowchart in Fig. 2 and described later.

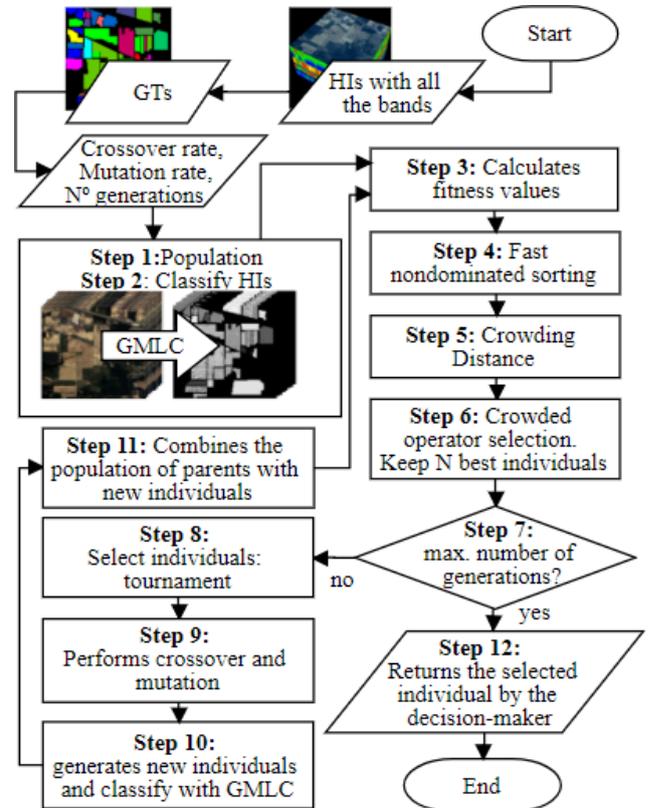


Fig. 2. Flowchart of the proposed method

Initially, the method receives the original HIs (with all bands), with the GT, mutation rates and crossover, and the number of generations. Subsequently, the population is initialized with N individuals (solutions) (Step 1), and the pixels are classified by the GMLC (Step 2). Then the fitness values of average recall and number of bands are calculated (Step 3). After this, in Step 4, the FNS categorizes PF solutions and other rankings. In Step 5, the crowding distance is calculated to estimate how isolated a solution is. Solutions with more dispersed neighbors are considered more important because they characterize more genetic diversity. In Step 6, the solutions are sorted according to the crowded

operator where the first sorting criterion is the Pareto-optimal solution and the second is crowding distance. Crowding distance is used in the comparison of solutions that are on the same rank or at the Pareto-optimal solution set. After this, the N best individuals are kept in the population (elitism).

If the maximum number of generations is reached (**Step 7**) the method ends and go to **Step 12**, otherwise the method will generate new individuals (**Step 8**); In **Step 8**, the method performs a selection by tournament for genetic operations of crossover and mutation (**Step 9**). Then, in **Step 10**, new individuals are generated and classified by the GMLC. After this, the method combines the population of parents with new individuals composing the population of the new generation (**Step 11**). Start a new generation by returning to **Step 3**. In **Step 12**, the individual selected by DM is returned.

III. EXPERIMENTS AND RESULTS

A. Experimental setting

Indian Pines, Salinas, and Pavia University datasets/HIs were used to evaluate the proposed method. The first two HIs were collected by the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) sensor, and the classes of these images represent mainly crops, that is, they are images of agricultural areas. Both images have 16 classes representing Alfalfa, Corn, Grass, Oats, Soybean, Woods, Corn, Fallow, Lettuce, Vineyard, among others. The last image, was obtained by the Reflective Optics System Imaging Spectrometer (ROSIS), it has 9 classes represent urban areas, and some of them being Water, Trees, Asphalt, among others. Table I shows information about these HIs.

TABLE I.
Information about datasets.

Name	Resolution	N. of classes	N. of bands
Indian Pines	145x145	16	220
Salinas	86x83	16	224
Pavia University	640x340	9	103

HIs with bands selected by different methods were compared with the proposed method. The following cases were used for band selection:

- 1) Wrapper method based on GA [12] that maximizes a fitness that combines two metrics based on the number of bands and accuracy. The solution with the best fitness was selected in each generation.
- 2) The state-of-the-art IRMoBS [16], with the objective of optimizing the variance, information entropy and number of bands. As in the NSGA2-based method the DM was used to select a solution in each generation.
- 3) HIs with all bands were also compared.

During the band selection performed by GA and for displaying the IRMoBS data in the following subsection, the GMLC was applied with 3-Fold Cross-Validation, with the same rules used by NSGA2-based method. For performance evaluation of the band selection obtained by each method, a new classification was performed by the GMLC with unused pixels in the band selection process.

The band selection experiments and subsequent classification of the HIs with the selected bands were organized with 90% of the pixels being used for band selection and model construction and 10% of the pixels for validation of the classification. This procedure was performed 5 times for each image, always separating different pixels in each validation set, that is, a 5-Fold cross validation was performed. After the execution of the algorithm in each fold, the mean and variance were obtained and a paired T-Test was applied.

All methods of band selection have been set to run for 500 generations with 100 solutions in the population. These values were empirically selected based on the runtime of each generation on the tested devices. The crossover rate and mutation were 50% and 5% respectively for all methods.

B. Experimental results

In Fig. 3a, Fig. 3b, and Fig. 3c the evolution of the average recall for Indian Pines, Salinas and Pavia U. where one of the five folds for each test of each image is shown.

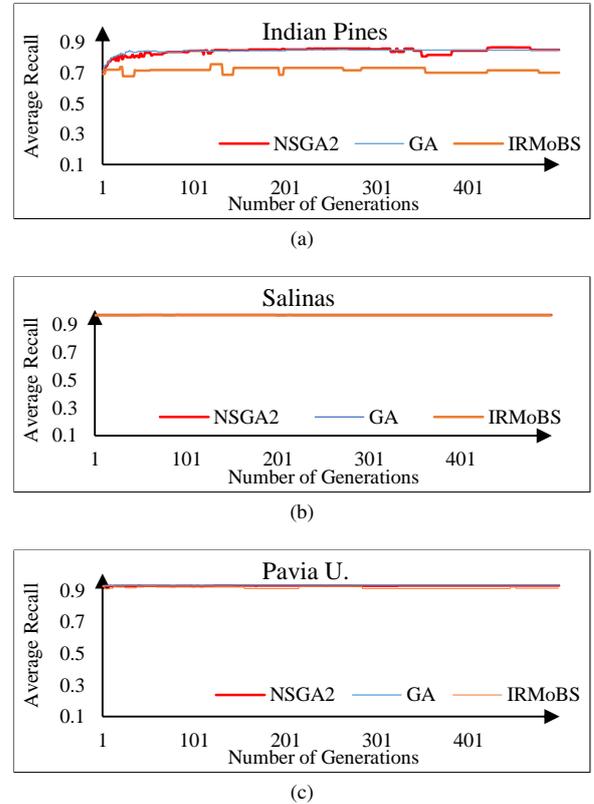


Fig. 3. Evolution of average recall

In the graphs of Fig. 3, for all HIs, the average recall performance achieved is similar using the NSGA2-based method or the GA-based method. However, the IRMoBS (filter method) showed lower performance for Indian Pines image than the other methods, and this difference suggests the advantage of using the wrappers methods.

The graphs in Fig. 4 show the evolution of the number of bands where for each HI the same solution of the respective graph of Fig. 3 was plotted. Through a simultaneous analysis

of the graphs of Fig. 3 and Fig. 4, it is observed that all methods can reduce the number of bands compared to the original image while practically maintaining or improving the average recall. The main advantage of the NSGA2-based method about the other two is the reduction of the number of bands. For Indian Pines HI, it is the most noticeable improvement in the average recall as the number of bands is reduced for methods based on NSGA2 and GA. This behavior may suggest a reduction in the Hughes phenomenon/ curse of dimensionality problem.

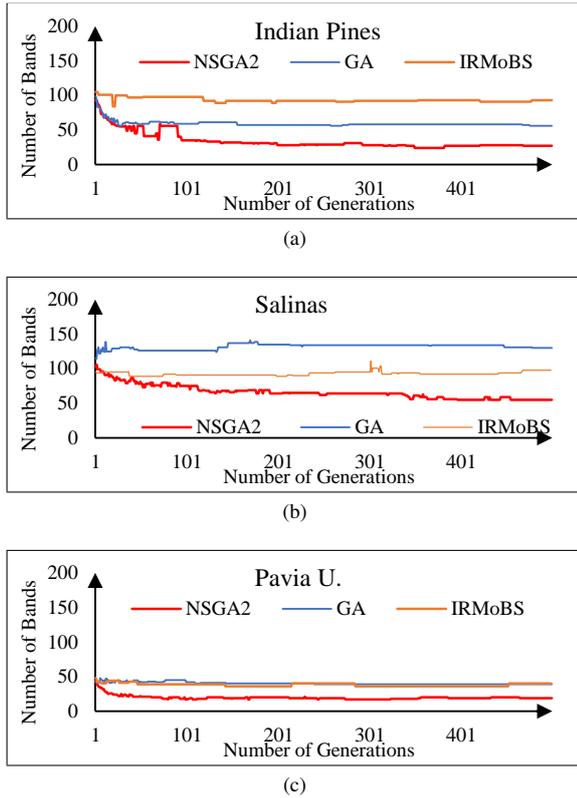


Fig. 4. Evolution of number of bands

Experiments were also performed with average recall (AvgRec), overall accuracy (OA) and paired T-Test applied to the test pixels with the bands selected by each method.

TABLE II.
Mean of Average Recall (AvgRec) and Overall Accuracy (OA) for 5-fold cross validation

Image	Method	Results			
		AvgRec	std	OA	std
Indian P.	All Bands	0.698	0.024	0.692	0.016
	IRMoBS	0.826	0.023	0.839	0.020
	GA	0.870	0.018	0.831	0.009
	NSGA2	0.928	0.018	0.862	0.013
Salinas	All Bands	0.452	0.366	0.410	0.414
	IRMoBS	0.997	0.003	0.998	0.001
	GA	0.996	0.001	0.997	0.001
	NSGA2	0.997	0.004	0.998	0.002
Pavia U.	All Bands	0.912	0.009	0.925	0.006
	IRMoBS	0.962	0.004	0.960	0.006
	GA	0.933	0.005	0.931	0.005
	NSGA2	0.964	0.006	0.961	0.005

The AvgRec was used in fitness function being a good metric for imbalanced data set. The OA informs a percentage of correct pixel classification, being a common metric for evaluating the segmented HIs. The paired T-test is widely used to evaluate cross-validation results where two average results are obtained by two different classifier models. The paired T-Test indicates whether these results have the same (null hypothesis) or different mean values.

Table II shows the mean of the AvgRec and the mean of the OA obtained from 5-fold cross-validation and the respective standard deviations. This table show that the Indian Pines and Pavia University HIs with all bands have an AvgRec and OA much lower when compared to HIs with selected bands. GMLC is a parametric algorithm that assumes Gaussian distributions in the data and it needs enough samples to operate well with many bands, otherwise it may perform poorly due to Hughes's phenomenon. This phenomenon may have occurred in these cases, which can explain the inferior performance for these HIs with all bands. For Pavia University image, even when the classification is performed with all bands, the performance of the classifier is not so poor. A possible explanation is that the amount of pixels in this image is larger, which allows a greater amount of training samples and consequently the method is less subject to the Hughes's phenomenon. The images with bands selected by the IRMOBS method, even this method being unsupervised, showed good results and often even higher than the images with bands selected by the wrapper method based on GA. In these tests, for most cases, the HIs with bands selected by the method based on NSGA2, presented the best results, however for this confirmation the paired T-Test test was applied and the results shown in Table III.

TABLE III.
p-values between proposed method and others.

Average Recall (AvgRec)			
Name	Indian P.	Salinas	Pavia U.
GA	0.000	0.517	0.000
IRMoBS	0.000	0.374	0.364
All Bands	0.000	0.041	0.000
Overall Accuracy (OA)			
Name	Indian P.	Salinas	Pavia U.
GA	0.014	0.099	0.000
IRMoBS	0.023	0.374	0.731
All Bands	0.000	0.047	0.000

In Table III, for p-values greater than 0.05 (5% significance) means that the value of the means obtained in the classification of the pixels between HIs with bands selected by the indicated method and the NSGA2-based method is the same. Otherwise, this indicates that one of the algorithms is better than the other and therefore the one with the highest AvgRec value or the OA value is considered. Analyzing Tables II and III together, the values highlighted in bold indicate when the HIs with bands selected by NSGA2 proposed in this study had better pixel classification performance or identical HIs with bands selected by other methods. Therefore, the pixel classification of HIs with bands selected by the NSGA2-based method, showed the results

better than the other methods for most cases tested. The GA and IRMoBS also performed well, especially the latter, since it tied with NSGA2-based method in some cases.

The average number of selected bands and percentage of reduction from HIs with all bands of the last solution found by each method is shown in Table IV. In this Table, it is possible to verify that all methods reduce the number of bands. The NSGA2-based method was able to reduce the number of bands more effectively.

TABLE IV. Results of number of bands to the last selected solution and percentage reduction for each method

Name	All Bands	IRMoBS	GA	NSGA2
Indian P.	220	94 (57.3%)	53 (75.9%)	26 (88.2%)
Salinas	224	96 (57.1%)	128 (42.8%)	55 (75.4%)
Pavia U.	103	40 (61.1%)	38 (63.1%)	21 (79.6%)

The improvement of the AvgRec and OA is more prominent for solutions with bands selected by the NSGA2-based method than the others case, where it can also be observed in Table IV a high reduction in the number of bands. This shows that for the cases tested the NSGA2-based method can be an interesting alternative for band selection.

IV. CONCLUSION

In this study, a supervised MOBS method based on NSGA, GMLC, and a DM is proposed. This method tries to find tradeoff solutions, that is, solutions with good results of the average recall and the number of bands. The method is aimed at reducing the number of bands allowing to deal with the problem of the curse of dimensionality improving the pixel classification performance of HIs.

Experiments were conducted on images of agricultural areas and urban area. The method was compared to pixel classification of original HIs with all bands, HIs with bands selected by a GA-based method, and the state-of-the-art IRMoBS. The NSGA2-based method, showed superior performance for the number of bands and the average recall and overall accuracy for most of the cases tested.

Based on these results, the NSGA2-based method appear as a good alternative MOBS of HIs. Thus, the improvement of the pixels classification of these HIs is expected, making possible an evolution in this AP tool.

A negative point in the method proposed is that it takes a long time to run. This is because of the training and classification steps that are influenced by the number of samples and executed for each solution in each generation. However, this is not a problem for practical applications with images after selected bands, but studies can be conducted to optimize this time.

In future works, it will be evaluated the method with other classification algorithms as decision trees, SVM, among others. We also intend to test the algorithm with different parameters of the algorithms, such as the number of generations, crossing rate, mutation, and new fitness functions. Other algorithms such as multiobjective PSO will also be tested as wrappers.

REFERENCES

- [1] T. M. Lillesand, R. W. Kiefer and J. W. Chipman, *Remote Sensing and Image Interpretation*, 5th ed., Ed. John Wiley & Sons, 2004.
- [2] FAO (Food and Agriculture Organization of the United Nations and International Telecommunication Union), *E-Agriculture in action: Drones for Agriculture*, Edited by Gerard Sylvester, 2018.
- [3] M. Attas E. Cloutis, C. Collins, D. Goltz, C. Majzels, J. R. Mansfield, H. H. Mantsch. "Near-infrared spectroscopic imaging in art conservation: Investigation of drawing constituents", *J. Cultural Heritage*, vol. 4, no. 2, pp. 127-136, 2003.
- [4] M. J. Khan, H. S. Khan, A. Yousaf, K. Khurshid, A. Abbas, *Modern Trends in Hyperspectral Image Analysis: A Review*. IEEE Access, vol. 6, pp. 14118-14129, 2018.
- [5] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd Edition. Ed. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2008.
- [6] G. Hughes, On the mean accuracy of statistical pattern recognizers. *IEEE Trans. Inf. Theory*, vol. 14, n. 1, pp. 5563, Jan. 1968.
- [7] Zhang, X., Sun, Q., Li, J.: Optimal band selection for high dimensional remote sensing data using genetic algorithm, *Proc. SPIE 7471*, Second International Conference on Earth Observation for Global Changes, October 2009; doi:10.1117/12.847907
- [8] D. Saqui, J. H. Saito, L. A. C. Jorge, E. J. Ferreira, D. C. Lima, J. P. Herrera, *Methodology for band selection of Hyperspectral Images using Genetic Algorithms and Gaussian Maximum Likelihood Classifier*, International Conference on Computational Science and Computational Intelligence, Las Vegas, EUA, pp. 733-738, 2016.
- [9] M. Kumar, *Feature Selection for Classification of Hyperspectral Remotely Sensed data using NSGA-II*, Water Resources Seminar, Citeseer, 2004.
- [10] M. Masaeli, G. Fung, J. G. Dy, From transformation-based dimensionality reduction to feature selection. *ICML'10 Proceedings of the 27th International Conference on International Conference on Machine Learning*, pp. 751-758, 2010.
- [11] C. Vaiphasa, A. K. Skidmore, W. F. Boer and T. Vaiphasa. A hyperspectral band selector for plant species discrimination. *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 62, n. 3, 2007. p. 225-235.
- [12] L. Zhuo, J. Zheng, F. Wang, X. Li, B. Ai, and J. Qian, A Genetic Algorithm Based Wrapper Feature Selection Method for Classification of Hyperspectral Images Using Support Vector Machine. *Proceedings of SPIE 7147*, Geoinformatics 2008 and Joint Conference on GIS and Built Environment: Classification of Remote Sensing Images, vol. 7147, pp. 397402 2008.
- [13] X. Zhang, W. Wang, Y. Li and L. C. Jiao, Pso-based automatic relevance determination and feature selection system for hyperspectral image classification, *Electronics Letters*, vol. 48, n. 20, pp. 12631265, 2012.
- [14] M. Zhang and J. Ma and M. Gong. Unsupervised Hyperspectral Band Selection by Fuzzy Clustering with Particle Swarm Optimization. *IEEE Geoscience and Remote Sensing Letters*, vol. 14, n5, pp. 773-777. 2017.
- [15] M. Gong, M. Zhang and Y. Yuan, Unsupervised Band Selection Based Evolutionary Multiobjective Optimization for Hyperspectral Images. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, n. 1, pp. 544-577, 2015. 10.1109/TGRS.2015.2461653.
- [16] X. Xu, Z. Shi and B. Pan, "A New Unsupervised Hyperspectral Band Selection Method Based on Multiobjective Optimization," in *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 11, pp. 2112-2116, Nov. 2017.
- [17] K. Deb; S. Agrawal, A. Pratap and T. A. Meyaruvan, Fast Elitist Non-Dominated Sorting Genetic Algorithm for Multi-Objective Optimization: NSGA-II. *Proceedings of the 6th International Conference on Parallel Problem Solving from Nature*, 849-858, 2000.
- [18] K. Deb; A. Pratap; S. Agrawal and T. A. Meyaruvan, A Fast and Elitist MultiObjective Genetic Algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, 182-197, 2002.
- [19] D. Kimovski, R. Matha, S. Ristoy, and R. Prodan, Multiobjective service oriented network provisioning in ultra-scale systems. *European Conference on Parallel Processing*, pp. 529-540, 2017.

Incorporated Decision-maker-based Multiobjective Band Selection for Pixel Classification of Hyperspectral Images

Diego SAQUI^{1,2}, José Hiroki SAITO^{1,3}, Daniel Caio De LIMA¹, Luis Mariano Del Val CURA³,
Steve Tsham Mpinda ATAKY⁴

¹Computer Department, UFSCar – Federal University of São Carlos, São Carlos, Brazil

²IFMS – Federal Institute of Mato Grosso do Sul, Corumbá, Brazil

³UNIFACCAMP – University Center of Campo Limpo Paulista, Campo Limpo Paulista, Brazil

⁴ÉTS – École de Technologie Supérieure, Université du Québec, Montréal (Québec), Canada
diego.saqui@ufscar.br

Abstract—Hyperspectral images (HIs) are characterized by a higher spectral resolution than other images and have applications in various fields, to wit, medicine, agriculture, mining, among others. Segmentation can be obtained from the pixel classification and it is a powerful tool for object identification. Notwithstanding, the problems of the curse of dimensionality and the demand for computational resources occur due to the number of bands. Techniques that reduce dimensionality, such as genetic algorithms, are promising, but they cannot assure a balance between conflicting objectives such as improving classification and reducing the number of bands. Multiobjective band selection can be applied to search for tradeoff solutions that have this balance. Therefore, in this manuscript, we propose a novel method called Incorporated Decision-Marker-based multiobjective band selection (IDMMoBS) that tries to find tradeoff solutions using spectral and spatial information. In the experiments, the IDMMoBS reduced the number of bands between 85.4% and 85.8% of the total and it outperformed the majority of other methods compared in this criterion. For the pixel classification, the IDMMoBS presented better results than all compared cases taking into account all evaluated metrics using SVM classifier. Accordingly, the IDMMoBS is suitable for band selection.

Index Terms—remote sensing, hyperspectral imaging, image segmentation, image classification, evolutionary computation.

I. INTRODUCTION

Hyperspectral images (HIs) are concerned with the measurement, analysis, and interpretation of spectra acquired from a given scene (or specific object) at a short, medium or long distance by an airborne or satellite sensor [1]. These images are characterized by their spatial and spectral resolutions (number of spectral bands) and contain information of materials beyond the visible spectrum range [2]. The advantage of using HIs is due to their unique bands that facilitate the discrimination of different materials [2-3]. Each pixel has a signature consisting of a large number of bands representing the wavelength and different reflectance values. This spectral signature varies according to the type of the object, allowing tasks of identification and analysis. HIs are useful in many applications, to wit, medical imaging, agriculture, among others [4].

Land cover classification is one of the different tasks in

Remote Sensing (RS), modeled as part of a supervised process based on previous Ground Truth (GT), which is concerned with the identification of different land cover objects [1]. This identification is possible through the pixel classification and segmentation of HIs. The segmentation is the process of subdivision of the image into its constituent regions. For pixel classification and subsequent segmentation, algorithms with peculiar statistical properties such as Support Vector Machines (SVM), Gaussian Maximum Likelihood Classifier (GMLC) and Random Forest (RF) are required [5-7]. These algorithms can be used directly for segmentation or can be complemented using image processing techniques.

HIs have a high dimensionality of data due to the number of bands [8], and because of this some challenges are common, such as:

- i) Hughes's phenomenon/curse of dimensionality, described by the existence of high dimensionality for a low number of samples. This problem is common in HIs because of the cost and difficulty of elaborating GTs, and it can result in poor classification accuracy [9];
- ii) information redundancy caused by the high correlation between neighboring bands;
- iii) computation complexity for classifiers based on conventional statistics [10];
- iv) the high cost of HIs sensors [11]; and
- v) difficulty in data transmission and storage. Modern data capturing devices, such as Drones, cannot have many computational resources to meet the demand for HIs processing.

Thereof, HIs classification and segmentation are often preceded by dimensionality reduction techniques which are categorized as feature selection or feature extraction [12].

Feature extraction performs a feature space transformation for a new space with a smaller dimension, and some examples of this category are principal component analysis (PCA) [13] and independent component analysis (ICA) [14]. Feature selection is a routine of selecting a subset of the original features, resulting in a smaller number of features where the original representation of the variables is not changed. Therefore, feature selection is preferred when the original meaning of the features is demanded [15].

For HIs, there is an interest in maintaining the original

meaning of the information of bands. Thus, feature selection is often applied and for this purpose, it is called band selection. The advantage of this routine is the possibility of identifying bands that characterize objects more efficiently than feature extraction, in this light, avoiding problems like the curse of dimensionality. With fewer identified bands, it is also possible to construct specific and cheaper sensors, which require less computational resources for classification, transmission, and storage of obtained data.

There is a great diversity of methods for band selection. Some examples are a clustering-based method developed using each channel map as a data point [16], sparse representation methods used to select the band subset [17-18], and structure-aware measures for band informativeness and independence [19]. A traditional clustering-based method for band selection is Ward's Linkage Strategy using Mutual Information (WaLuMi). The WaLuMi groups the bands according to the mutual information and then it selects a representative band from each group. This was one of the first clustering-based algorithms in the literature and it has shown good results even when compared to recent methods [20]. In [21], the Maximum information and minimum redundancy with clonal selection algorithm (MIMR-CSA) defines a criterion that maximizes the amount of information for the selected band combination while removing redundant information. The authors compare the classification capacity, and the proposed method presented better results than others in the literature.

Methods based on evolutionary algorithms have shown satisfactory results for hyperspectral band selection. Some examples are the Firefly algorithms [22-23], Genetic Algorithms (GAs) [24-25] and Particle Swarm [26-27]. A hybrid method of Wrapper and Filter approaches called Information Gain-Grey Wolf Optimization (IG-GWO) has been proposed, where Information Gain and Gray Wolf Optimizer were combined. IG is an important metric that can measure how much information the features could contribute to the classification. This method was compared with five other state-of-the-art methods and it showed better accuracy results [26]. All these optimization algorithms establish a single objective function, and for band selection application, some measure of segmentation is conducted individually or combined with the reduction of the number of bands in a single objective function. Particularly in [25], a supervised/Wrapper method for band selection based on GA with Support-Vector Machines (GA-SVM) was proposed combining two metrics in fitness function. The GA-SVM uses a weighted sum strategy between accuracy and the inverse of the number of bands to try to find solutions with a balance between these metrics and shows good overall accuracy results. However, the difficulty of this method and the others is to find balanced tradeoff solutions, for example, solutions that simultaneously have a satisfactory classification performance with a reduced number of bands.

The need for reducing the number of bands and improve classification performance makes the band selection problem be characterized as a multiobjective problem since both objectives can be conflicting. For this, multiobjective optimization band selection (MOBS) can be applied [4].

MOBS are usually based on evolutionary algorithms and

use the Pareto Frontier that can contain more than one optimal solution. Pareto Frontier (tradeoff) solutions cannot be considered better among each other and therefore cannot be sorted, but these solutions are better than others in the search space.

Different unsupervised MOBS, which can also be categorized as filter methods, can be found in the literature. Non-dominated Sorting Genetic Algorithm 2 (NSGA2) was proposed to find solutions that optimize cumulative texture and spectral information of a band combination using a minimum number of bands [12]. Another method is the Tchebycheff Decomposition Strategy (TDS) for the multiobjective evolutionary algorithm based on decomposition (MOEA/D). This method was studied to optimize the number of bands and entropy contained in the selected band subsets [4].

A state-of-the-art method called incorporated rank-based multiobjective band selection (IRMoBS) using TDS was proposed to optimize three objectives, band number, variance and information entropy [29]. Three band selection methods [17-19], are compared with IRMoBS which presented better classification using SVM, extreme learning machine (ELM) or K-Nearest Neighborhood (KNN) in the pixels of HIs with selected bands.

If the GT is available, supervised/Wrapper methods can be used to fit a classification model based on a priori knowledge; thereby, it is expected to obtain better results in band selection. Wrapper methods usually require a high computational cost but they present better results than filter methods because they incorporate the classifier into the selection process [30]. This type of method, together with multiobjective algorithms, can be an alternative for the band selection problem. Nonetheless, there were little studied in the literature.

Another feature little explored in the literature is the use of spatial information with band selection algorithms. A study of the application of filters for spatial information is on HIs is performed in [31]. In this study the authors show that even simple filtering can greatly improve the performance of classifiers in HIs. Therefore, spatial filtering can be introduced to improve the classification accuracy of HIs.

In this paper, a novel method based on MOEA/D is modeled for MOBS of HIs. The method is called incorporated decision-maker-based multiobjective band selection (IDMMoBS) and aims to deal with the dimensionality of HIs through the search for tradeoff solutions. The IDMMoBS is modeled as a Wrapper strategy and therefore the SVM classifier is used. The IDMMoBS incorporates a decision-maker (DM) to select only one solution from Pareto Frontier and a repair method based on WaLuMi [20] and IG/IG-GWO [26]. Another differential of this study is that the use of spatial information was considered. The DM, the repair technique and the use of spatial information are little explored in band selection, and this is quite innovative in this study.

The main contributions of this manuscript are:

- i) A new Wrapper MOBS to search for tradeoff solutions since these types of methods have not been explored enough;
- ii) Use of DM and a repair method that are rarely

discussed in band selection methods;

- iii) Use of pixel neighborhood information in each band;
- iv) Comparisons with other supervised (GA-SVM [25]), semi-supervised (IG-GWO [26]) and unsupervised (WaLuMi [20], MIMR-CSA [21], IRMoBS [29]) state-of-the-art methods;
- v) Comparisons with HIs with all bands;
- vi) Analysis of the classification of the pixels of HIs with bands selected by the different methods compared, being this classification performed by SVM;
- vii) Use of spatial information.

The remainder of this text is organized as follows. Section II presents a detailed description of the proposed method. The experiments are conducted and the results reported in Section III. At last, in Section IV, the main contribution of this manuscript is summarized and future work direction is pointed.

II. METHODOLOGY

Multiobjective optimization is related to solutions with contradictory objectives to be optimized simultaneously searching for tradeoff solutions. Methods for multiobjective optimization are based on the Pareto Frontier and use the non-dominance relation to compare solutions [4][12][29].

A solution x dominates a solution y , if x is better than y in at least one objective and equal or better in all other objectives. The notation x that dominates y is given by $x \succ y$. Solutions that are not dominated by any other are Pareto-optimal solutions and constitute the Pareto Frontier.

A Pareto-optimal solution to a multiobjective optimization problem could be an optimal solution of a scalar optimization problem where the objective is an aggregation of all the objectives to be optimized. Therefore, the approximation of the Pareto Frontier can be decomposed into a number of scalar objective optimization subproblems [32].

There are several approaches for converting the Pareto Frontier approximation problem into scalar optimization subproblems; and one explored in unsupervised MOBS is the TDS [4, 29]. In this paper, the TDS is used in a supervised and Wrapper MOBS to compose the IDMMoBS.

In IDMMoBS proposed herein, each band of a solution generated from a HI is represented by a binary code in a vector as shown in Fig. 1. The value 1 indicates the presence of the band equivalent to that position and 0 otherwise.

0	1	0	0	1	1	0	...	number of bands
---	---	---	---	---	---	---	-----	-----------------

Figure 1. Representation of the solution.

Two objective functions, f_1 and f_2 , that should be maximized, are designed for IDMMoBS. f_1 is based on the overall accuracy of each solution x and it is shown in eq. (1).

$$f_1(p, \hat{p}) = \frac{1}{n} \sum_{i=1}^n (1 * (p_i = \hat{p}_i)) \quad (1)$$

where n is the number of pixels, i represents each pixel, p_i is the original value of the class in the GT, and \hat{p}_i is the predict value of the class. The choice for overall accuracy is

because it is the most used metric in the studies found to evaluate the pixel classification in images with selected bands and it is directly related to the success rate of classification of pixels.

For the reduction of bands, it is considered in f_2 the difference of 1 and the ratio between the number of bands of solution x and the total number of bands possible, as shown in eq. (2).

$$f_2(x) = 1 - \frac{t(x)}{\text{len}(x)} \quad (2)$$

where $\text{len}(x)$ is the size of the array of the solution x and $t(x)$ is the total of bands present in the solution x as shown in eq. (3).

$$t(x) = \sum_{n=0}^{\text{len}(x)} x_n \quad (3)$$

The TDS used in the IDMMoBS aggregates the m objectives, which in this case are f_1 and f_2 , into a single scalar objective function. This aggregation is performed through the difference in fitness values of f_1 and f_2 between a solution x and an ideal point $z^* = [z_1^*, \dots, z_m^*]^T$. For maximizing values, $z_j^* = \max\{f_j(x) \mid x \in \Omega\}$ for j -th objective and Ω is the decision space.

The single scalar objective function $g^{\text{tds}}(x \mid w, z^*)$ shown in eq. (4).

$$g^{\text{tds}}(x \mid w, z^*) = \max_{1 \leq j \leq m} \{w_j \mid f_j(x) - z_j^*\} \quad (4)$$

$x \in \Omega$

where $w = [w_1, \dots, w_m]^T$ is the weight vector for the m objectives with $w_j \geq 0$ and $\sum_{j=1}^m w_j = 1$. $x^* = [x_1^*, \dots, x_m^*]^T$ is the ideal solution corresponding to the ideal point z^* .

In the subproblem evolution process, the single scalar objective function in eq. (4) of the neighboring solutions and new solutions generated by mutation and crossover operations are calculated to further determine which solutions could be updated. Solutions that have larger g^{tds} values are more likely to be replaced by others that are closer to the ideal solution [29].

A. IDMMoBS

Prior to the execution of the IDMMoBS algorithm, a mean filter was applied to each spectral band to smooth out band noise and also allow the use of pixel spatial information. A mode filter was applied concurrently to the GT with classes of pixels used for training, where the same mask was used in both cases. To apply these filters, a 2x2 mask was used that worked in accordance with the sample capture strategy presented in the experiments section.

The IDMMoBS is shown in Algorithm 1, where in **Step 1 - Initialization**, the population P is initialized from *OrigHI* (**Step 1.1**) and fitness is established. In this study the WaLuMi algorithm (based on Mutual Information) was used to generate groups and then randomly a band was selected from each group to generate solutions for the initial population P . This procedure decreases the correlation

between bands and it is repeated until a number N of initial solutions are reached.

In **Step 1.2** the EP is set to empty and in the **Step 1.3** the weight vectors W are established. In **Step 1.3** are also computed the Euclidean distances between the weight vectors and then it is assign the NS nearest solutions to each solution i based on the weight vectors creating the neighborhood $NG(i)$. In **Step 1.4**, the ideal point z^* is determined from the maximum fitness values f_1 and f_2 present in the initial population.

In **Step 2**, the update is performed by a $NGEN$ number of generations. In each generation the DM selects a solution from the EP for display. For each subproblem from $i=1$ to N , in **Step 2.1 - Reproduction and Repair**, the neighborhood solutions $NG(i)$ are selected to generate new solutions NI through crossover and mutation operators. After the crossover operation and before the mutation operation, a repair method is applied. The heuristic of the repair method is explained in the following sections. Then the fitness values of the new solutions NI are established.

The ideal point z^* is updated after comparing the objectives of the new solutions NI with the z^* (**Step 2.2**). Then in **Step 2.3**, the objective values are computed based on eq. (4) for the neighborhood solutions $NG(i)$. Among the neighborhood solutions, the ones whose objective values are larger than g^{tds} , that is, those that are more distant from the z^* , are replaced by solution $y' \in NI$. Then, in **Step 2.4**, the EP is updated with non-dominated solutions.

Finally, in **Step 3** the DM selects the solution according to the established heuristic from EP and returns a trained classification model.

Algorithm 1 IDMMoBS

Input:

$NGEN$ (number of generations),
 N (population size),
 NS (neighborhood size),
 GT ,
 p (threshold of the amount of bands),
 $OrigHI$ (Pixels/samples with all bands),
 mr (mutation rate)
and cr (crossover rate).

Output:

Solution selected by the DM and the trained classification model.

Step 1 – Initialization:

Step 1.1 – Creates an initial population P generating solutions from $OrigHI$ using strategy with WaLuMi.

* Establish fitness values using (1) and the mean obtained from cross-validation and calculate the fitness for the number of bands using (2) of solutions in P .

Step 1.2 – Creates an external population $EP = \emptyset$.

Step 1.3 – Set of weight vector $W = [w_1, \dots, w_N]^T$.

For each $i=1 \dots N$,

$$NG(i) = \{i_1, \dots, i_{NS}\},$$

where $w_{i_1}, \dots, w_{i_{NS}}$ are the NS closest weight vectors to w_i .

Step 1.4 – Initialize z^* .

Step 2 – Update:

For $ngen = 1, \dots, NGEN$ do (Stopping Criteria),

DM selects and output a solution of EP ,

For $i = 1, \dots, N$, do,

Step 2.1 – Reproduction and repair:

For each i_a, i_b from $NG(i)$:

Applies the crossover operator and generates new solutions y'_a and y'_b in NI .

Repairs the solutions in NI .

Applies the mutation operator.

* Establish fitness values of NI .

Step 2.2 – Update of z^* :

For each $y' \in NI$,

For each $j = 1, \dots, m(\text{objectives})$,

if $f_j(y') > z_j^*$, then

$$\text{set } z_j^* = f_j(y')$$

Step 2.3 – Update of Neighboring Solutions:

For each index $k \in NG(i)$,

if $g^{tds}(y' | w_k, z^*) \leq g^{tds}(x_k | w_k, z^*)$ then

$$\text{set } x_k = y', \quad f_1(x_k) = f_1(y')$$

$$\text{and } f_2(x_k) = f_2(y').$$

Step 2.4 – Update of EP :

Remove all solutions dominated by y' from EP .

Add y' to EP if no solutions in EP dominate y' .

Step 3 – Ending:

DM selects and outputs a solution of EP and the trained classification model.

B. Support Vector Machines (SVM)

SVM is based on the theory of statistical learning [34] with principles for finding an optimal hyperplane as a decision function in a large space [35][36]. In the case of classification problems with two distinct classes that can be linearly separable, given infinite possibilities for obtaining the separator function; the SVM selects the best one. This is the simplest scenario for classification by the SVM. The SVM considers the best separator, the one that minimizes the generalization error. Thus, the selected separator is the one with the largest margin between two classes. For nonlinear surfaces, the feature vector is mapped to a larger Euclidean feature space using a kernel function in the nonlinear SVM [37].

The SVM algorithm is used to classify the pixels of each solution for later the average accuracy in eq. (1) be applied. During the execution of IDMMoBS, whenever some pixel classification is performed, 3-Fold cross-validation is applied. The pixels are alternately selected from the HI using the bands of the solution being classified and inserted into one of the folds (systematic sampling). Thus, it avoids that the pixels of the same class be allocated only in the same fold and also allows that the pixels of the same class of different regions in a HI be present in the different folds.

SVM was chosen because it is one of the most commonly used algorithms for HIs and does not require assumption about data distribution (nonparametric). One of its disadvantages for use in conjunction with Wrapper strategy is its training time, but this cannot be a serious problem for the band selection process because it is performed prior to the use of bands in practical applications. Its main advantages are the high classification rate, and that this algorithm requires few training samples to operate which is important for HIs applications [38].

C. Repair strategy and Centroid-based DM

In [30] it is discussed that the repair method is to make the solutions feasible. For the band selection problem a maximum band threshold may be desired and therefore for this problem a set of feasible solutions may be desired. Therefore, in this study the following heuristic is used:

- i) Based on [20], this strategy applies WaLuMi to group the highly correlated bands using Mutual Information of each solution;
- ii) based on [26], selected the band with the highest information gain (IG) of each group;
- iii) sort the set of bands according to the IG;
- iv) and select the p bands with the highest IG.

This heuristic limits the search for solutions with balanced objectives, but only when they have less than p bands. p acts as a threshold amount of bands. To avoid the high computational cost, the IG of each band can be calculated before the band selection process and the mutual information between the bands has already been calculated to initialize the population in the WaLuMi.

Even with the p threshold, many solutions can emerge at the Pareto Frontier and so a centroid-based decision-maker (DM) is designed. This DM is based in the method proposed in [33] that is based on centroid and a threshold, and it is adapted to the IDMMoBS.

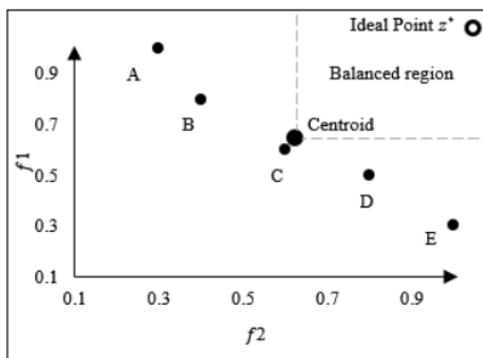


Figure 2. Representation of the DM.

In [33], the centroid is calculated from the f_1 and f_2 of the Pareto-optimal solutions and acts as a threshold to create a region of balanced solutions (balanced region). Fig. 2 shows an example of balanced region (in the dashed region) where the centroid is the largest circle in the lower left corner of the balanced region and the ideal point z^* is the circle without filling. A-E are solutions of the decision space that constitute the Pareto Frontier.

Having the centroid calculated different rules can be established for decision making which solution to choose. In this paper we use the following:

- 1) If there are solutions in the balanced region then the solution in that region closest to the ideal point z^* is selected;
- 2) If there is no solution in the balanced region, then the solution in the Pareto Front with value greater than the centroid of one of the functions that presents the best value for another solution is selected.
- 3) otherwise, the solution of all Pareto-optimal solutions closest to the centroid is selected.

Through this strategy it is possible to select only one solution of the Pareto Frontier.

III. EXPERIMENTS AND RESULTS

A. Experimental setup

Three different HIs/datasets were used to evaluate the IDMMoBS and compare with other methods. The images used were Indian Pines, Salinas, and Pavia University and information about them are shown in Table I.

TABLE I. INFORMATION ABOUT HIS

Name	Resolution	# of pixels	# of classes	# of bands
Indian Pines	145x145	21025	16	200
Salinas	512x217	111104	16	204
Pavia University	610x340	207400	9	103

Indian Pines and Salinas datasets were collected by the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) sensor. Examples of images Indian Pines and Salinas with their respective GTs are shown in Figs. 3 and 4.

Originally the Indian Pines image had 220 bands and Salinas 224, but 20 water absorption bands were removed. The spatial resolution of these images is 20 m/pixel and the wavelength range from 0.4 to 2.5 μm . The classes of these images represent mainly crops characterizing themselves as agricultural areas. Both images have 16 classes represented by different colors in the GTs, and some Indian Pines classes represent Alfalfa, Corn, Grass, Oats, Soybean, Woods, among others. Salinas classes represent Broccoli, Corn, Fallow, Lettuce, Vineyard, among others.

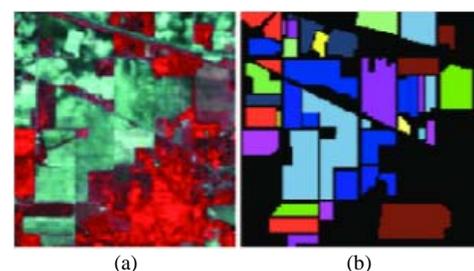


Figure 3. (a) Indian Pines HI. (b) GT.



Figure 4. (a) Salinas HI. (b) GT.

The Reflective Optics System Imaging Spectrometer (ROSIS) obtained Pavia University dataset (shown in Fig. 5) with a spatial resolution of 1.3 m/pixel and their classes represent urban areas, and it has 9 classes, some of them being Water, Trees, Asphalt, among others.

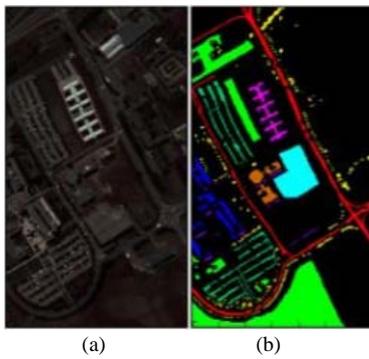


Figure 5. (a) Pavia University HI. (b) GT.

Six different band selection methods/cases were used for comparison:

- i) GA-SVM [25];
- ii) IRMoBS [29];
- iii) WaLuMi [20];
- iv-) MIMR-CSA [21];
- v) IG-GWO [26];
- vi) and HIs with all bands.

To avoid the influence of the mean filter on the pixels used in IDMMoBS, always the lower right pixel was used as a sample (systematic sampling). For each band, the mean value was obtained from the previous three pixels using the 2x2 mask. An example of the sampled pixels is shown in Fig. 6, where the black background pixel was used as a sample for IDMMoBS and the four pixels with the same number were used to calculate their mean value in each band.

1	1	2	2	3	3	4	4
1	1	2	2	3	3	4	4
5	5	6	6	7	7	8	8
5	5	6	6	7	7	8	8
9	9	10	10	11	11	12	12
9	9	10	10	11	11	12	12

Figure 6. Example of pixels used as samples (with black background) in experiments with the IDMMoBS.

For IDMMoBS, 20% of the sample pixels obtained by the previous method, that is, using black background pixel, were used for testing with 5-Fold Cross-Validation. For all other methods compared, 20% of all pixels of each HI were used for testing with 5-Fold Cross-Validation. An important observation is that the special information was considered only for IDMMoBS as part of the purpose of this paper, and this occurred both in the band selection process as well as in testing. For the other methods only the spectral information proposed in the original works was used.

The experiments with the comparative methods were similar to those performed in [26], so the comparison of these methods was possible. Only for the GA-SVM and IRMoBS further executions were necessary to adjust the HIs and amount of samples used. For the GA-SVM, the fitness function was set with 0.8 and 0.2 respectively for accuracy and the inverse of the number of bands. Other parameters of these two algorithms were the same as the original proposals in [25] and [29].

The IDMMoBS, IRMoBS and GA-SVM have been set to run for 500 generations with 100 solutions (subproblems) in the population. The SVM was applied for the calculation of the overall accuracy with 3-Fold Cross-Validation. These values were empirically selected based on the runtime of each generation on the tested hardware. The crossover rate and mutation were 50% and 5% respectively. Other parameters defined in the IDMMoBS were neighborhood size $NS = 20$, p , the threshold of the amount of bands, with 30 (Indian Pines and Salinas) or 20 (Pavia University) and the weights w_1 and w_2 were randomly defined for each i -th subproblem. In the experiments, 80% of each HI pixels different from 20% testing pixels were used in the process of band selection in IDMMoBS.

After the band selection performed by each method, a classification was performed on the 20% testing pixels by the SVM using 5-Fold Cross-Validation to evaluate the final result. The optimal parameter C of the RBF kernel was determined via 5-Fold Cross-Validation.

In the GA-SVM, IRMoBS and IDMMoBS methods the bands were obtained automatically and their quantities are shown in the experiments. In the IG-GWO method the original number of bands obtained in [26], that is, 30, 28 and 18 bands were respectively used for the Indian Pines, Salinas and Pavia U. images. In the other three selection methods, the number of bands selected is 30 for the Indian Pines and Salinas datasets and 20 for the Pavia University dataset.

Three metrics were used to evaluate the success rate of classification of pixels:

- i) Overall accuracy (OA) refers to the number of correctly classified instances divided by the total number of testing samples;
- ii) Average accuracy (AA) is a measure of the mean value of the classification accuracies of all classes;
- iii) The kappa coefficient (KC) is a statistical measurement of consistency between the ground truth map and the final classification map [26].

In addition to these metrics the number of bands used in each method is also discussed.

B. Results

The results obtained by the different methods and for each considered HI are shown in Tables II-IV. The first column shows the name of each method and also in parentheses is shown the total amount of bands used. The three results from the following columns respectively represent the overall accuracy (OA), the average accuracy (AA) and the kappa coefficient (KC). Best results in these tables are highlighted in bold.

Table II refers to results obtained for Indian Pines image. This Table shows that bands selected by IDMMoBS outperform the other six cases for classification using the SVM classifier. The good results provided by the IDMMoBS are repeated for the different metrics evaluated, that is, OA, AA and KC. In particular, it is observed that even the GA-SVM method, which is biased towards searching for solutions that directly improve the pixel classification, has no classification results similar to those provided by the bands selected by IDMMoBS. A plausible explanation for the good results achieved by the IDMMoBS in this image is because of the use of spatial information obtained by the mean filter. The results are also superior to other state-of-the-art literature methods such as IG-GWO and also for the images that use of all spectral bands.

TABLE II. OA, AA AND KC OBTAINED FROM DATA WITH BANDS SELECTED FROM INDIAN PINES BY DIFFERENT METHODS AND CLASSIFIED BY THE SVM USING 5-FOLD CROSS-VALIDATION

Indian Pines			
Method	OA(%)	AA(%)	KC(%)
GA-SVM (131 bands)	86.8 ±1.5	81.0 ±2.1	84.9 ±1.7
IRMoBS (44 bands)	80.0 ±1.7	74.6 ±4.6	77.2 ±1.9
WaLuMi (30 bands)	80.6 ±0.8	68.2 ±2.0	77.7 ±1.0
MIMR-CSA (30 bands)	84.5 ±0.9	79.4 ±1.8	82.3 ±1.0
IG-GWO (30 bands)	85.2 ±0.8	82.6 ±6.6	83.1 ±0.9
All Bands (200 bands)	84.9 ±1.5	75.2 ±4.0	82.8 ±1.7
IDMMoBS (29 bands)	91.4 ±1.1	91.4 ±3.9	90.1 ±1.3

TABLE III. OA, AA AND KC OBTAINED FROM DATA WITH BANDS SELECTED FROM SALINAS BY DIFFERENT METHODS AND CLASSIFIED BY THE SVM USING 5-FOLD CROSS-VALIDATION

Salinas			
Method	OA(%)	AA(%)	KC(%)
GA-SVM (38 bands)	92.4 ±0.3	96.2 ±0.2	91.5 ±0.4
IRMoBS (32 bands)	88.3 ±0.6	93.5 ±0.5	86.9 ±0.6
WaLuMi (30 bands)	93.0 ±0.2	96.4 ±0.2	92.2 ±0.2
MIMR-CSA (30 bands)	93.5 ±0.2	96.8 ±0.1	92.7 ±0.2
IG-GWO (28 bands)	93.9 ±0.2	97.0 ±0.7	93.2 ±0.2
All Bands (204 bands)	92.8 ±0.3	96.5 ±0.2	92.1 ±0.4
IDMMoBS (29 bands)	94.4 ±0.4	97.3 ±0.1	93.7 ±0.4

TABLE IV. OA, AA AND KC OBTAINED FROM DATA WITH BANDS SELECTED FROM PAVIA UNIVERSITY BY DIFFERENT METHODS AND CLASSIFIED BY THE SVM USING 5-FOLD CROSS-VALIDATION

Pavia University			
Method	OA(%)	AA(%)	KC(%)
GA-SVM (70 bands)	93.9 ±0.5	92.9 ±1.1	92.0 ±0.7
IRMoBS (24 bands)	89.5 ±0.3	88.9 ±1.0	86.2 ±0.5
WaLuMi (20 bands)	90.1 ±1.0	86.7 ±1.2	86.7 ±1.4
MIMR-CSA (20 bands)	92.5 ±0.3	89.1 ±0.6	90.1 ±0.3
IG-GWO (18 bands)	94.2 ±0.2	92.3 ±1.0	92.3 ±0.2
All Bands (103 bands)	94.1 ±0.3	92.8 ±0.9	92.2 ±0.4
IDMMoBS (15 bands)	96.2 ±0.3	95.5 ±0.5	94.7 ±0.4

Tables III and IV refer to Salinas and the Pavia University images respectively. For these images most methods performed well for classification, often with metrics greater than 90%. These tables show that the IDMMoBS also

presented better results when compared with the other methods in the classification criterion. The most competitive method to IDMMoBS was the IG-GWO which for Salinas and Pavia University images presented very close results.

In addition to comparing the OA, AA, and KC metrics, the mean value of the last solution found by each method and used to generate the data from Tables II-IV are shown in Table V. This Table shows the data on the number of bands reduction percentage for each evaluated method. These data can be calculated from the number of bands shown in the first column of Tables II, III and IV.

As shown in Table V, IDMMoBS also showed better results in the number of bands of the Indian Pines and Pavia University images, but for the Salinas image the IG-GWO performed better.

TABLE V. PERCENTAGE REDUCTION OF THE NUMBER OF BANDS DEFINED OR OBTAINED IN EACH METHOD.

Method	Indian Pines	Salinas	Pavia University
GA-SVM	34.5%	81.4%	32.0%
IRMoBS	78%	84.3%	76.7%
WaLuMi	85%	85.3%	80.6%
MIMR-CSA	85%	85.3%	80.6%
IG-GWO	85%	86.3%	82.5%
IDMMoBS	85.5%	85.8%	85.4%

By analyzing data classification and number of bands shown in the previous tables, in many methods simultaneously reducing the number of bands can lead to improved classification metrics. This behavior is present mainly when the IDMMoBS and IG-GWO methods are applied, which for many of the cases presented better results than the use of all bands and can suggest some improvement of the Hughes phenomenon.

IV. CONCLUSION

In this study, a novel supervised MOBS method called IDMMoBS was proposed. This method uses evolutionary algorithms and it is based on TDS, SVM, a repair strategy, spatial information and a centroid-based DM. The IDMMoBS searches for tradeoff solutions with the balance between the overall accuracy and the number of bands.

Experiments were conducted on HIs with different regions and diverse areas. HIs with selected bands by the proposed method were compared to pixel classification of original HIs with all bands and five other band selection methods. For most of the test cases conducted, the IDMMoBS, showed higher performance for the number of bands and better overall accuracy when applying the SVM classifier. In just one case, the method called IG-GWO achieved a slightly lower number of bands than the proposed method, but this method had a worse classification rate. Based on the tests performed the IDMMoBS has been able to find significant tradeoff solutions for specific classifiers and appear as a good alternative MOBS of HIs. IDMMoBS may have promoted good results in the classifier, mainly due to the use of spatial information incorporated in the band selection process and subsequent classification.

In future works, the band selection process will be evaluated with other algorithms such as decision trees and deep learning. Different parameters will also be tested, such as the number of generations, crossing rate, mutation, and

fitness functions. Other filter methods will also be considered to improve segmentation.

REFERENCES

- [1] A. Plaza, J. A. Benediktsson, J. W. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, M. Marconcini, J. C. Tilton, G. Trianni, "Recent advances in techniques for hyperspectral image processing", *Remote Sensing of Environment*, vol. 113, pp. S110-S122, 2009. doi:10.1016/j.rse.2007.07.028.
- [2] M. J. Khan, H. S. Khan, A. Yousaf, K. Khurshid, A. Abbas, "Modern Trends in Hyperspectral Image Analysis: A Review," *IEEE Access*, vol. 6, pp. 14118-14129, 2018. doi: 10.1109/ACCESS.2018.2812999.
- [3] M. Attas, E. Cloutis, C. Collins, D. Goltz, C. Majzels, J. R. Mansfield, H. H. Mantsch, "Near-infrared spectroscopic imaging in art conservation: investigation of drawing constituents", *Journal of Cultural Heritage*, vol. 4, n. 2, pp. 127-136, 2003. doi:10.1016/S1296-2074(03)00024-4.
- [4] M. Gong, M. Zhang and Y. Yuan, "Unsupervised Band Selection Based on Evolutionary Multiobjective Optimization for Hyperspectral Images," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 1, pp. 544-557, 2016. doi: 10.1109/TGRS.2015.2461653
- [5] T. M. Lillesand, R. W. Kiefer, J. W. Chipman, "Remote Sensing and Image Interpretation", pp. 550-562, 5th ed., Ed. John Wiley & Sons, 2004. doi: 10.2307/634969.
- [6] S. Amini, S. Homayouni, A. Safari, A. A. Darvishsefat, "Object-based classification of hyperspectral data using Random Forest algorithm", *Geo-spatial Information Science*, vol. 21, pp. 127-138, 2018. doi: 10.1080/10095020.2017.1399674.
- [7] J. Xia, P. Ghamisi, N. Yokoya, A. Iwasaki, "Random Forest Ensembles and Extended Multi-Extinction Profiles for Hyperspectral Image Classification." *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, n. 1, pp. 202-2016, 2018. doi: 10.1109/TGRS.2017.2744662.
- [8] D. A. Landgrebe, "Signal Theory Methods in Multispectral Remote Sensing", John Wiley and Sons, pp. 237-239, 2003. doi: 10.1002/0471723800.
- [9] G. Hughes, "On the mean accuracy of statistical pattern recognizers". *IEEE Trans. Inf. Theory*, vol. 14, n. 1, pp. 55-63, Jan. 1968. doi: 10.1109/TIT.1968.1054102.
- [10] X. Zhang, Q. Sun, J. Li, "Optimal band selection for high dimensional remote sensing data using genetic algorithm", *Proceedings of SPIE - The International Society for Optical Engineering.*, October 2009. doi: 10.1117/12.847907.
- [11] D. Saqui, J. H. Saito, L. A. C. Jorge, E. J. Ferreira, D. C. Lima, J. P. Herrera, "Methodology for band selection of Hyperspectral Images using Genetic Algorithms and Gaussian Maximum Likelihood Classifier", *International Conference on Computational Science and Computational Intelligence*, Las Vegas, EUA, pp. 733-738, 2016. doi: 10.1109/CSCI.2016.0143.
- [12] M. Kumar, "Feature Selection for Classification of Hyperspectral Remotely Sensed data using NSGA-II", *Water Resources Seminar*, Citeseer, 2004.
- [13] M. D. Farrell, R. M. Mersereau, "On the impact of PCA dimension reduction for hyperspectral detection of difficult targets", *IEEE Geoscience and Remote Sensing Letters*, vol. 2, no. 2, pp. 192-195, Apr. 2005. doi: 10.1109/LGRS.2005.846011.
- [14] N. Falco, J. A. Benediktsson, L. Bruzzone, "A study on the effectiveness of different independent component analysis algorithms for hyperspectral image classification", *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2183-2199, Jun. 2014. doi: 10.1109/JSTARS.2014.2329792.
- [15] M. Masaeli, G. Fung, J. G. Dy, "From transformation-based dimensionality reduction to feature selection". *ICML'10 Proceedings of the 27th International Conference on International Conference on Machine Learning*, pp. 751 - 758, 2010.
- [16] K. Sun, X. Geng, L. Ji, "Exemplar component analysis: A fast band selection method for hyperspectral imagery", *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 5, pp. 998-1002, May 2015. doi: 10.1109/LGRS.2014.2372071.
- [17] W. Sun, L. Zhang, B. Du, W. Li, Y. M. Lai, "Band selection using improved sparse subspace clustering for hyperspectral imagery classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2784-2797, Jun. 2015. doi: 10.1109/JSTARS.2015.2417156.
- [18] W. Sun, L. Zhang, L. Zhang, Y. M. Lai, "A dissimilarity-weighted sparse self-representation method for band selection in hyperspectral imagery classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 9, pp. 4374-4388, Sep. 2016. doi: 10.1109/JSTARS.2016.2539981.
- [19] G. Zhu, Y. Huang, J. Lei, Z. Bi, F. Xu., "Unsupervised hyperspectral band selection by dominant set extraction," *IEEE Geoscience and Remote Sensing Letters*, vol. 54, no. 1, pp. 227-239, Jan. 2016. doi: 10.1109/TGRS.2015.2453362.
- [20] A. Martínez-Usó, F. Pla, J. M. Sotoca, P. García-Sevilla, "Clustering-Based Hyperspectral Band Selection Using Information Measures". *IEEE Transaction Geoscience Remote Sensing* 2007, vol. 45, pp. 4158-4171. doi 10.1109/TGRS.2007.904951.
- [21] J. Feng, L. Jiao, F. Liu, T. Sun, X. Zhang, "Unsupervised feature selection based on maximum information and minimum redundancy for hyperspectral images". *Pattern Recognition* 2016, 51, 295-309. doi: /10.1016/j.patcog.2015.08.018.
- [22] R. Y. M. Nakamura, L. M. G. Fonseca, J. A. dos Santos, R. da S. Torres, X.-S. Yang, J. P. Papa, "Nature-inspired framework for hyperspectral band selection," *IEEE Trans. Geoscience Remote Sensing*, vol. 52, no. 4, pp. 2126-2137, Apr. 2014. doi: 10.1109/TGRS.2013.2258351.
- [23] H. Su, B. Yong, Q. Du, "Hyperspectral band selection using improved firefly algorithm," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 1, pp. 68-72, Jan. 2016. doi: 10.1109/LGRS.2015.2497085.
- [24] C. Vaiphasa, A. K. Skidmore, W. F. Boer, T. Vaiphasa. "A hyperspectral band selector for plant species discrimination". *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 62, n. 3, p. 225-235, 2007. doi: 10.1016/j.isprsjprs.2007.05.006
- [25] L. Zhuo, J. Zheng, F. Wang, X. Li, B. Ai, J. Qian, "A Genetic Algorithm Based Wrapper Feature Selection Method for Classification of Hyperspectral Images Using Support Vector Machine". *Proceedings of SPIE 7147, Geoinformatics 2008 and Joint Conference on GIS and Built Environment: Classification of Remote Sensing Images*, vol. 7147, pp. 397-402, 2008. doi: 10.1117/12.813256.
- [26] X. Zhang, W. Wang, Y. Li, L. C. Jiao, "Pso-based automatic relevance determination and feature selection system for hyperspectral image classification", *Electronics Letters*, vol. 48, n. 20, pp. 1263-1265, 2012. doi: 10.1049/el.2012.0539.
- [27] M. Zhang, J. Ma, M. Gong. "Unsupervised Hyperspectral Band Selection by Fuzzy Clustering with Particle Swarm Optimization". *IEEE Geoscience and Remote Sensing Letters*, vol. 14, n°5, pp. 773-777. 2017. doi: 10.1109/LGRS.2017.2681118.
- [28] F. Xie, F. Li, C. Lei, L. Ke, "Representative Band Selection for Hyperspectral Image Classification". *ISPRS International Journal of Geo-Information* 2018, 7, 338. doi:10.3390/ijgi7090338.
- [29] X. Xu, Z. Shi, B. Pan, "A New Unsupervised Hyperspectral Band Selection Method Based on Multiobjective Optimization," in *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 11, pp. 2112-2116, Nov. 2017. doi: 10.1109/LGRS.2017.2753237.
- [30] N. Sánchez-Maróño, A. Alonso-Betanzos, M. Tombilla-Sanromán, "Filter methods for feature selection - A comparative study". *International Conference on Intelligent Data Engineering and Automated Learning. Lecture Notes in Computer Science*, vol. 4881, Springer, Berlin, Heidelberg, 2007. doi:10.1007/978-3-540-77226-29.
- [31] X. Cao, B. Ji, Y. Ji, L. Wang, L. Jiao, "Hyperspectral image classification based on filtering: A comparative study". *Journal of Applied Remote Sensing*, vol. 11. doi: 10.1117/1.JRS.11.035007.
- [32] Z. Qingfu, Li, Hui, "MOEA/D: A Multiobjective Evolutionary Algorithm Based on Decomposition". *IEEE Transactions on Evolutionary Computation*, Vol. 11, N. 6, 2007. doi: 10.1109/TEVC.2007.892759.
- [33] D. Kimovski, R. Matha, S. Ristoy, R. Prodan, "Multiobjective service oriented network provisioning in ultra-scale systems". *European Conference on Parallel Processing, Lecture Notes in Computer Science*, vol 10659. Springer, Cham pp. 529-540, 2017. doi: 10.1007/978-3-319-75178-8_43.
- [34] V. N. Vapnik, "The Nature of Statistical Learning Theory". New York: Springer-Verlag, 1995. doi: 10.1007/978-1-4757-3264-1.
- [35] B. Boser, I. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proc. 5th Annu. Workshop Comput. Learn. Theory*, 1992, pp. 144-152. doi: 10.1145/130385.130401.
- [36] N. Cristianini and J. Shawe-Taylor. Cambridge, U.K.: Cambridge Univ. Press, 2000. doi: 10.1017/CBO9780511801389.
- [37] M. Pal and G. M. Foody, "Feature Selection for Classification of Hyperspectral Data by SVM," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 5, pp. 2297-2307, May 2010. doi: 10.1109/TGRS.2009.2039484.
- [38] P. Ghamisi, J. Plaza, Y. Chen, J. Li, and A. Plaza, "Advanced spectral classifiers for hyperspectral images: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 1, pp. 8-32, Mar. 2017. doi: 10.1109/MGRS.2016.2616418.