

## MINERAÇÃO DE DADOS APLICADA A IMAGENS HIPERESPECTRAIS E MULTIESPECTRAIS NA CLASSIFICAÇÃO DO USO E OCUPAÇÃO DO SOLO

LAÍZA C. DE A. SILVA<sup>1</sup>, EDUARDO M. RAMBO<sup>2</sup>, WILLYAN R. BECKER<sup>3</sup>,  
HUMBERTO J. COMINETI<sup>4</sup>, CARLOS A. DA SILVA JÚNIOR<sup>5</sup>, JERRY A.  
JOHANN<sup>6</sup>

<sup>1</sup> Discente de Mestrado do Programa de Pós-Graduação em Engenharia Agrícola, PGEAGRI, Universidade Estadual do Oeste do Paraná, UNIOESTE, Cascavel, PR, Brasil, Fone: (45)3220-7320, {laiza.cavalcante@hotmail.com}

<sup>2</sup> Discente de Mestrado do Programa de Pós-Graduação em Engenharia Agrícola, PGEAGRI, Universidade Estadual do Oeste do Paraná, UNIOESTE, Cascavel, PR, Brasil, Fone: (45)3220-7320, {eduardomrambo@hotmail.com}

<sup>3</sup> Discente de Doutorado PGEAGRI, UNIOESTE, Cascavel, PR, Brasil, Fone: (45) 3220-7320, {willyanbecker@hotmail.com}

<sup>4</sup> Discente de graduação, UNIOESTE, Cascavel, PR, Brasil, Fone: (45) 3220-7320, {humbertocominetti@gmail.com}

<sup>5</sup> Prof. Dr., Laboratório de Geotecnologia Aplicada em Agricultura e Floresta, GAAF, Universidade do Estado do Mato Grosso, UNEMAT, Cáceres, MT, Brasil, Fone: (66) 3511-2158, {carlosjr@unemat.br}

<sup>6</sup> Prof. Dr., Orientador, PGEAGRI, UNIOESTE, Cascavel, PR, Brasil, Fone: (45) 3220-7320, {jerry.johann@unioeste.br}.

Apresentado no  
XLIX Congresso Brasileiro de Engenharia Agrícola - CONBEA 2020  
23 a 25 de novembro de 2020 - Congresso On-line

**RESUMO:** Este trabalho teve como objetivo: comparar a eficiência de quatro agrupamentos de bandas pelo algoritmo de classificação *K-means*, no procedimento de redução de dimensionalidade para dados hiperespectrais (sensor AisaFENIX); realizar a classificação do uso e cobertura do solo utilizando sensores multiespectrais (Sentinel-2 e Landsat-8/OLI) e avaliar os classificadores Random Forest (RF) e Support Vector Machine (SVM), quanto a suas acurácias para uma região de Alta Floresta/MT. Os mapeamentos pelo sensor hiperespectral AisaFENIX resultaram em métricas de acurácia maiores que os sensores multiespectrais. Seu desempenho superior pode ser explicado em função da maior presença de dados sobre cada alvo, reunidos em novas bandas, possibilitando maior capacidade de distinção entre as classes.

**PALAVRAS-CHAVE:** AisaFENIX, SVM, Random Forest, K-means

### DATA MINING APPLIED TO HYPERSPECTRA AND MULTIESPECTRAL IMAGES CLASSIFICATION IN THE LAND USE LAND COVER

**ABSTRACT:** This work aimed to compare the efficiency of K-means on dimensionality reduction for hyperspectral data (AisaFENIX sensor) for a scene in Alta Floresta-MT, Brazil; perform classification of land use and cover by multispectral sensors (Sentinel-2 and Landsat-8/OLI) and hyperspectral (AisaFENIX) using Random Forest (RF) and Support Vector Machine (SVM) algorithms. The classification for hyperspectral sensor AisaFENIX resulted in greater accuracy metrics than the multispectral sensors. The higher performance can be explained due to the greater presence of data on each target, gathered in new bands, allowing greater capacity to distinguish between targets.

**KEYWORDS:** Aisafenix, SVM, Random Forest, K-means.

**INTRODUÇÃO:** Sensores hiperespectrais fornecem informações mais detalhadas, discriminadas e precisas em comparação com sensores multiespectrais, possibilitando a

aquisição simultânea de dados em centenas de bandas espectrais, compreendendo o espectro do visível ao infravermelho, e conseqüentemente um maior detalhamento dos alvos. Todavia, o maior detalhamento espectral gera enorme volume de dados, elevando os custos computacionais de processamento (ZHANG & QIU, 2012). Neste sentido, o processo de classificação do uso e cobertura de áreas gera entraves em função do fenômeno de Hughes, o qual descreve que um aumento do número de atributos reduz a precisão do classificador quanto a sua predição (HUGHES, 1968). Logo, anterior a classificação é necessário a extração de dados irrelevantes ou redundantes, principalmente ao analisar dados com alta dimensionalidade.

Desta forma, tem-se como objetivo comparar a eficiência de quatro agrupamentos de bandas, pelo algoritmo de classificação *K-means*, no procedimento de redução de dimensionalidade para dados hiperespectrais (sensor AisaFENIX); realizar a classificação do uso e cobertura do solo utilizando sensores multiespectrais (Sentinel-2 e Landsat-8/OLI); avaliar os algoritmos Random Forest (RF) e Support Vector Machine (SVM) para a classificação, quanto a suas acurácias para uma região de Alta Floresta/MT.

**MATERIAL E MÉTODOS:** A metodologia foi conduzida conforme as etapas estabelecidas por Fayyad et al. (1996): Pré-processamento, Transformação, Mineração de dados, Avaliação e Interpretação (Figura 1). Em função das características distintas entre os sensores utilizados, este estudo possui duas abordagens metodológicas distintas (Figura 1a e 1b).

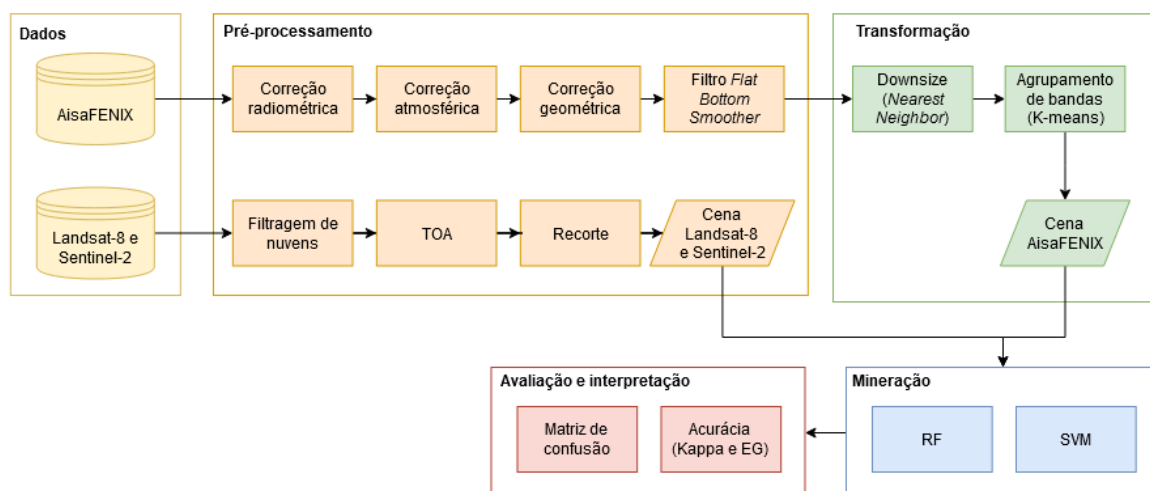


FIGURA 1. Fluxograma descrevendo as etapas de Análise e Descoberta de Conhecimento (KDD) para as bases de dados AisaFENIX e Landsat-8 e Sentinel-2.

A área sob estudo situa-se entre 9°44'S e 56°12'W (Figura 2), no município de Alta Floresta/MT, o qual tem como vegetação característica predominante a floresta ombrófila, pertencente ao bioma amazônico. A área imageada pelo sensor AisaFENIX compreende uma faixa de aproximadamente 0,9x5km, obtida através do sobrevoo a 726m de altitude, resultando numa resolução espacial de 1,1m. A faixa espectral abrange de 380 a 2500 nm, totalizando 363 bandas. Foram usadas cenas dos sensores Landsat-8/OLI (órbita/ponto 227/067), e Sentinel-2 (órbita/ponto 21LWK), nas datas 06/10/2017 (Landsat-8) e 14/10/2017 (Sentinel-2).

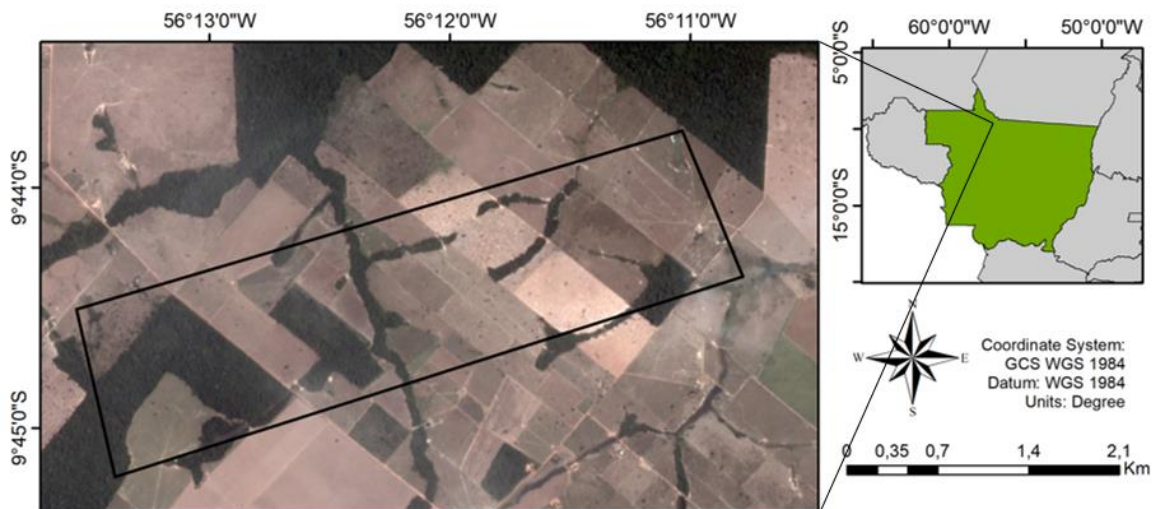


FIGURA 2. Mapa de situação da faixa sob estudo do município de Alta Floresta/MT.

A cena do sensor AisaFENIX foi adquirida por um sobrevoo da área, a 726 metros de altitude, entre 13:20:24 horas e 13:21:36 horas do dia 09/10/2017, com uma resolução espacial de 1,1 metro. O sensor AisaFENIX realiza a coleta de dados na faixa espectral de 380 a 2500 nm, apresentando resolução espectral de 6,8 nm entre 380 e 970 nm e 5,46 nm entre 970 e 2505 nm (totalizando 363 bandas). O campo de visão (FOV) é de 32,3° e o campo de visão instantânea (IFOV) é de 0,084° e podendo chegar a 1,0 m de resolução espacial com altitude de voo de 660 m (SPECIM, 2017).

Cenas dos satélites Landsat-8/OLI (órbita/ponto 227/067), produto Level-C1/TOA e Sentinel-2 (órbita/ponto 21LWK), produto Level-C1/TOA foram adquiridas pela plataforma do Google Earth Engine. Em virtude do não imageamento destes sensores no mesmo dia de sobrevoo com o sensor AisaFENIX, utilizou-se as imagens com menor porcentagem de nuvens mais próximas a data de imageamento, sendo: 06/10/2017 (Landsat-8) e 14/10/2017 (Sentinel-2). Tal condição não é um problema pelo fato de a reflectância dos alvos sob estudo possuir baixa variação em pequenos intervalos (SILVA, 2017).

O pré-processamento da cena hiperespectral abrangeu as etapas de: correção radiométrica pelo *software* CaliGEO Pro 2.2, conforme Campos (2017), correção atmosférica (*software* ATCOR-4) e correção geométrica. Além disso, aplicou-se o filtro *Flat Bottom Smoother* para suavizar as curvas de reflectância de cada banda, pela comparação entre três pontos (antecessor, atual e posterior), caso o ponto atual possua o menor valor de reflectância de seus vizinhos este será substituído pelo menor valor destes (SILVA, 2017). Após isto, para reduzir a dimensionalidade dos dados, aplicou-se o algoritmo *K-means* visando agrupar bandas hiperespectrais semelhantes (SANTOS & PEDRINI, 2016; PENTREATH, 2015) conforme a distância euclidiana entre elas. Todavia, anterior ao processo de agrupamento executou-se o *downsizing* das bandas, pelo interpolador vizinho mais próximo (*Nearest Neighbor*), para reduzir o tempo de processamento. Testou-se o agrupamento das bandas hiperespectrais aglomeradas em 5, 10, 15 e 20 grupos (Figura 3). Os procedimentos citados consistem na transformação dos dados, sendo realizados por scripts em *Python*.

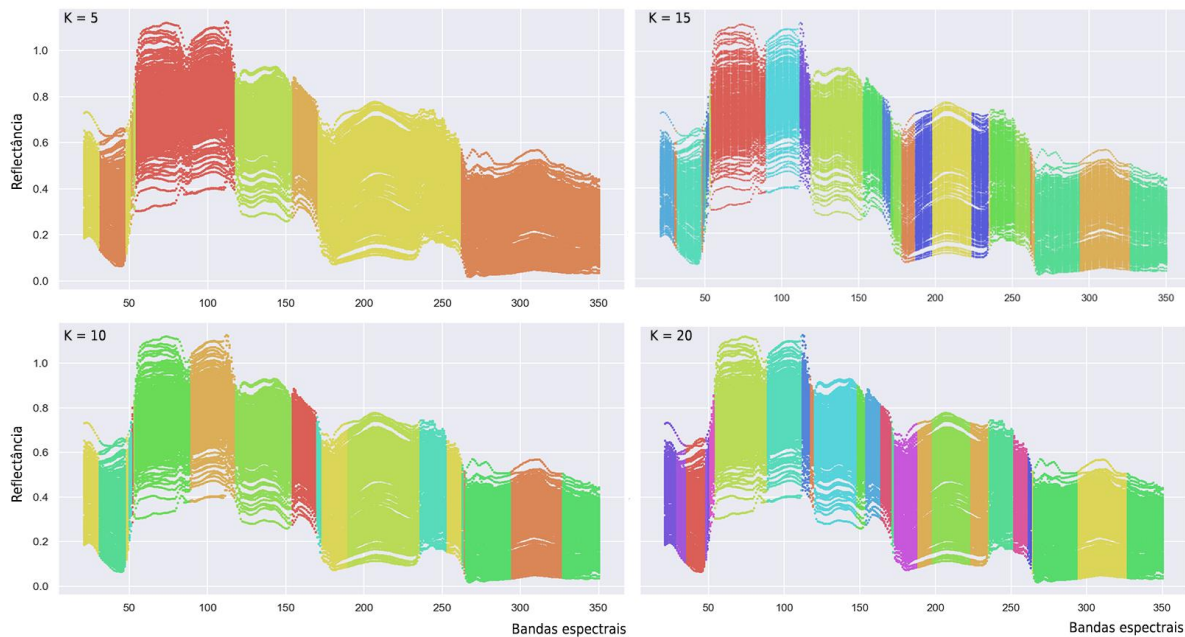


FIGURA 3. Agrupamento das bandas espectrais do sensor AisaFENIX pelo método K-means para os grupos 5, 10, 15 e 20.

As subetapas que compreendem a Mineração são: seleção de amostras de teste e treinamento, treinamento dos algoritmos e validação. Para avaliar os quatro agrupamentos gerados pelo K-means fez-se a classificação por ambos os algoritmos e posteriormente mensurou-se a Exatidão Global (EG) e o índice Kappa. Ambos os índices foram calculados pixel a pixel, a partir de uma vetorização da área sob estudo por imagens Google Earth Pro. A seleção das amostras de treinamento ocorreu pela identificação visual dos alvos nas composições coloridas falsa-cor para os três sensores sob estudo (AisaFENIX, Sentinel-2 e Landsat-8/OLI). A partir das amostras de treinamento, cada conjunto de dados (cenas dos sensores Sentinel-2 e Landsat-8 e as bandas agrupadas do sensor AisaFENIX) foram classificados pelos algoritmos RF e SVM.

### RESULTADOS E DISCUSSÃO:

Para o sensor AisaFENIX, ambos os classificadores (RF e SVM) possuem desempenho semelhante para todos os agrupamentos testados (5, 10, 15, 20), com variações de Kappa inferiores a 0,28 (Tabela 1). Logo, levando em consideração apenas as métricas de acurácia, tem-se que para todos os agrupamentos seu desempenho é semelhante. Além disso, segundo Congalton & Green (1998) todos os mapeamentos são considerados excelentes, com variação decimal entre os resultados. Todavia, vale ressaltar o incremento do tempo de processamento conforme a quantidade de bandas utilizadas. O SVM demanda maior tempo e custo computacional considerável em comparação ao RF, principalmente com maior número de classes. Tal condição também ocorre para o RF, embora a diferença interna do classificador seja menor.

Tabela 1. Índices Kappa e EG para a classificação por RF e SVM.

Sensor	Agrupamentos	RF		Tempo	SVM		Tempo
		Kappa	EG (%)	(s)	Kappa	EG (%)	(s)
AisaFENIX	5	83,21	92,47	231,51	83,58	92,64	276,75
	10	83,57	92,64	218,31	83,79	92,73	424,71
	15	83,51	92,60	243,07	83,77	92,72	634,69
	20	83,34	92,53	258,39	83,79	92,74	856,49
Landsat-8	--	81,74	91,81	1,85	82,18	91,96	0,24
Sentinel-2	--	74,76	88,10	4,17	73,05	87,14	5,77

RF: Random Fores, SVM: Support Vector Machine, EG: Exatidão Global.

A classificação do sensor AisaFENIX pelos classificadores RF (Figura 4) e SVM (Figura 5) diferenciam-se na detecção do alvo açude, localizado na parte leste da área de estudo. Embora exista pequenos açudes/banhados localizados na transição dos alvos de floresta e pastagem, esta porção é o local de maior destaque dentre os pequenos alvos ao longo do local. Esta característica gerou pequenas confusões entre os alvos pastagem e floresta. Para o alvo edificação, a confusão se deu entre a pastagem, em função da presença de bebedouros de animais ao longo da classe, como também pela mistura espectral no envolto das áreas construídas. Além disso, analisando as matrizes de confusão, o classificador SVM apresentou os maiores erros de inclusão e omissão entre os alvos floresta e pastagem em todas as classificações.

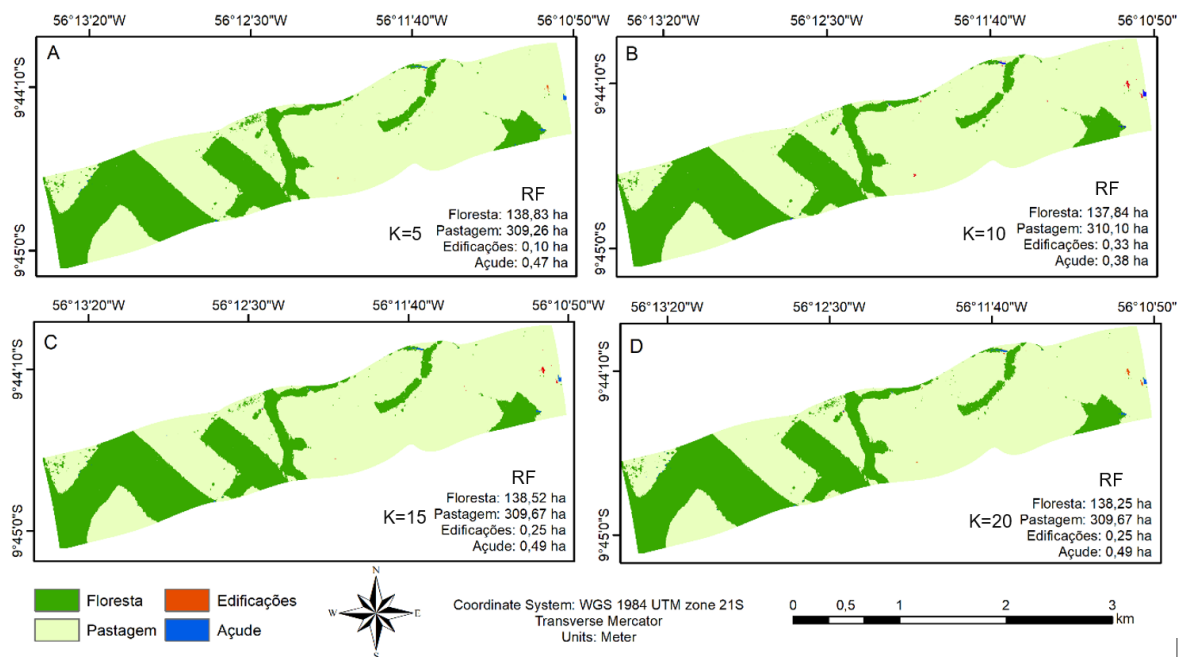


FIGURA 4. Mapa temático com as classificações pelo *Random Forest* nos diferentes agrupamentos K-means de bandas hiperespectrais do sensor AisaFENIX: (A) 5 grupos; (B) 10 grupos; (C) 15 grupos; (D) 20 grupos.

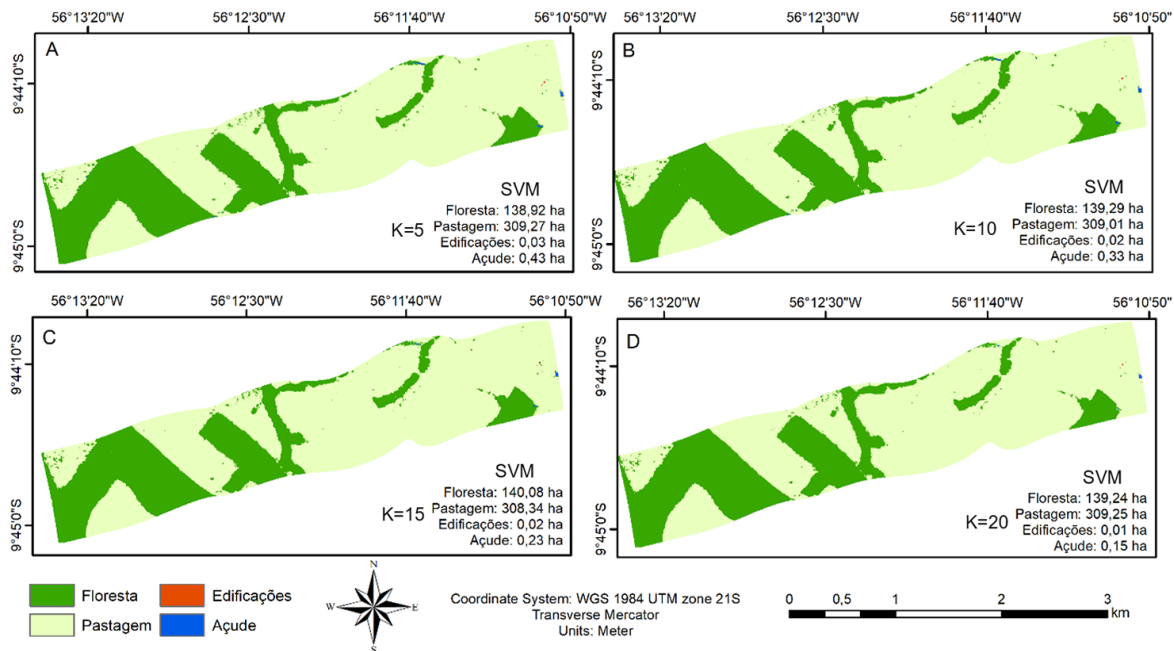


FIGURA 5. Mapa temático com as classificações pelo *Support Vector Machine* nos diferentes agrupamentos K-means de bandas hiperespectrais do sensor AisaFENIX: (A) 5 grupos; (B) 10 grupos; (C) 15 grupos; (D) 20 grupos.

Para as cenas dos sensores multiespectrais não foi executado a etapa de agrupamento em virtude da quantidade inferior de bandas quando comparadas ao sensor AisaFENIX. Os critérios de avaliação dos mapeamentos também foram o índice Kappa e EG. Ambos os sensores possuem resoluções espaciais e espectrais distintas (Sentinel-2 com 10m e 4 bandas e Landsat-8/OLI com 30m e 7 bandas). Apesar da resolução espacial proporcionar melhor definição dos alvos e distinção entre estes, esta não influenciou positivamente na classificação do uso e ocupação do solo. O sensor Sentinel-2 apresentou desempenho inferior do Kappa (74,76) e EG (88,10), com maiores erros de omissão e inclusão do que o Landsat-8 para o classificador RF (Tabela 1).

Embora a classificação do Sentinel-2 pelo RF (Figura 6) se enquadre na categoria muito bom (CONGALTON & GREEN, 1998), denota-se uma elevada confusão entre os alvos edificação e pastagem na porção sudoeste da área de estudo. Para o alvo açude também houve confusão com mata e pastagem o que pode ser atribuído a mistura espectral destes alvos visto que a presença de água não é marcante e se encontrava sempre em pequenas porções. Tais condições se dão pela baixa quantidade de pixels na amostra de treinamento, quais foram reduzidos em função da resolução espacial. A menor quantidade de bandas deste sensor pode ter afetado a distinção entre esses alvos, o que explica o desempenho superior para o sensor Landsat-8/OLI (Figura 6).

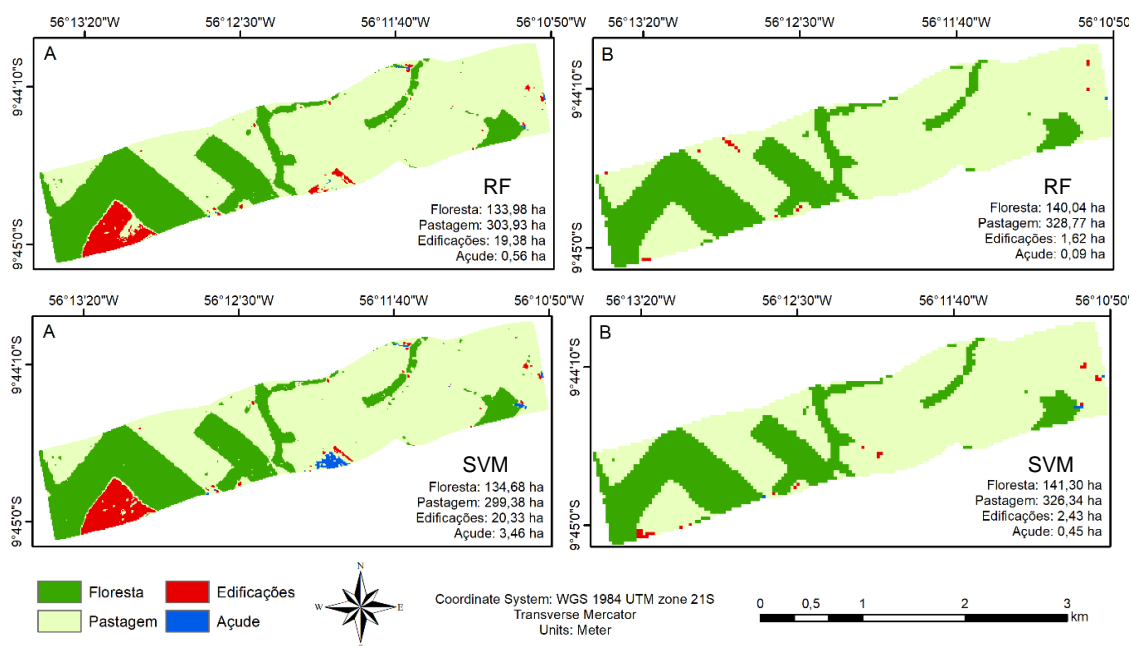


FIGURA 6. Mapa temático com as classificações pelo *Random Forest* e *Support Vector Machine* para as cenas multiespectrais dos sensores (A) Sentinel-2 e (B) Landsat-8.

Em todos os casos avaliados, os mapeamentos resultantes do sensor hiperespectral AisaFENIX resultaram em métricas de acurácia maiores que os sensores multiespectrais (Tabela 1). Seu desempenho superior pode ser explicado em função da maior presença de dados sobre cada alvo, reunidos em novas bandas, possibilitando assim maior capacidade de distinção entre os alvos sob estudo, assim como sua resolução espacial superior.

É necessário estudos mais aprofundados na etapa de redução de dimensionalidade das bandas hiperespectrais, para a criação de faixas de maior distinção entre os alvos. Inúmeros trabalhos investigam a capacidade de distinção de diferentes alvos em ambientes complexos utilizando sensores hiperespectrais (BRADLEY; KELLER; WEINMANN, 2018; CAMPOS, 2017; GUIDICI & CLARK, 2017), que evidenciam as vantagens destes sensores para este fim. Em contrapartida, dados hiperespectrais são de complexo manuseio, necessitando maior ênfase no processo de pré-processamento, sendo custoso computacionalmente e monetariamente, em função da necessidade de planejamento dos voos. Neste sentido, os sensores multiespectrais são vantajosos por exigirem menor custo computacional e monetário na determinação do uso e ocupação do solo em ambientes com poucos alvos e de fáceis distinção. Outra forma de elevar a capacidade de distinção de alvos é adotar classificadores mais robustos, como redes neurais, para aumentar esta capacidade Mahdianpari et al. (2018).

Embora os algoritmos possuam pequena diferença entre si, o SVM destaca-se pela melhor capacidade de distinção dos alvos. Wieland & Pittore (2014) ao avaliar performance de quatro classificadores (*Normal Bayes*, *K-Nearest Neighbors*, *Random Trees* e *Support Vector Machines*) para reconhecimento de áreas urbanas ratifica o desempenho superior do classificador SVM na área sob estudo. Andreola e Haertel (2009) também concluíram que o classificador SVM apresenta potencial na classificação de dados hiperespectrais AVIRIS, utilizando o banco de dados *Indian Pines*. De acordo com (ABE, 2005), o SVM apresenta vantagens em comparação com outros classificadores convencionais, especialmente quando o número de amostras de treinamento é pequeno e a dimensionalidade dos dados é grande.

A baixa quantidade de alvos na cena sob estudo e sua redução para possibilitar a comparação com demais sensores, inviabilizou uma maior seleção de amostras de treinamento, o que demanda um algoritmo que seja capaz de se adequar a essa condição. Desta forma, algoritmos de aprendizagem baseados em kernel tem-se mostrado mais eficazes que os tradicionais métodos de aprendizagem, em função destes transformarem os dados originais para um novo espaço de características onde se consegue uma melhor separação entre as classes.

Melgani & Bruzzone, (2004) aplicaram o algoritmo de aprendizagem SVM a imagens hiperespectrais, o qual se revelou uma efetiva e poderosa ferramenta para a interpretação desses dados, superando métodos tradicionais como: Regularized Radial Basis Function Neural Networks (RegRBFNN) e KNN. Camps-Valls & Bruzzone (2005) analisaram e compararam, em relação à interpretação de dados hiperespectrais, vários algoritmos baseados em kernel como Reg-RBFNN, SVM, Kernel Fisher Discriminant Analysis (KFDA), e Regularized AdaBoost (Reg-AB). Esse estudo revelou a superioridade do algoritmo SVM que, desde então, tem sido indicado por diversos autores pelo seu bom comportamento face à escassez e alta dimensionalidade das amostras de treinamento. Sua eficácia é considerada como referência em vários trabalhos da literatura (TARABALKA, 2010).

**CONCLUSÕES:** Os mapeamentos resultantes do sensor hiperespectral AisaFENIX resultaram em métricas de acurácia maiores que os sensores multiespectrais. Seu desempenho superior pode ser explicado em função da maior presença de dados sobre cada alvo, reunidos em novas bandas, possibilitando maior capacidade de distinção entre os alvos.

**AGRADECIMENTOS:** A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001, pelo apoio financeiro, à Universidade Estadual do Oeste do Paraná - UNIOESTE/Campus Cascavel, à Universidade Estadual do Mato Grosso - UNEMAT/Campus Cáceres e ao Núcleo de Pesquisa em Geotecnologias e Ciência de Dados (GeoScience) pelo apoio técnico e científico.

#### **REFERÊNCIAS:**

- ABE, S.; **Support Vector Machines for Pattern Classifications**. Kobe, Japão: Ed. Springer, 2005.
- ANDREOLA, R.; HAERTEL, V. Support Vector Machines na Classificação de Imagens Hiperespectrais. In: Simpósio Brasileiro de Sensoriamento Remoto, 14, 2009, Natal, Brasil. **Anais...** São José dos Campos: INPE, p. 6757-6764. 2009.
- CONGALTON, R.; GREEN, K. **Assessing the Accuracy of Remotely Sensed Data: Principles and Practices**. Boca Raton: CRC/Taylor & Francis. 183p. 2009.
- CAMPOS, T. L de L. Discriminação de espécies arbóreas nativas da Floresta Estacional Semidecidual e da exótica *Eucalyptus urograndis* W. Hill ex Maiden utilizando dados hiperespectrais. **Dissertação** (Mestrado em Agronomia). Universidade Estadual de Maringá (UEM), Maringá PR, 121f, 2017.
- CAMPS-VALLS, G; L. BRUZZONE, L. Kernel-based methods for hyperspectral image classification. **IEEE Transactions on Geoscience and Remote Sensing (TGARS)**. V. 43, n. 6, 2005.
- FAYYAD U, PIATETSKY-SHAPITO G, SMYTH P. From data mining to knowledge discovery in databases. **American Association for Artificial Intelligence Menlo Park**. V. 17, n. 3, 1996.

- GUIDICI, D.; CLARK, M. L. One-dimensional convolutional neural network land-cover classification of multi-seasonal hyperspectral imagery in the San Francisco Bay Area, California. **Remote Sensing**, v. 9, n. 6, 2017.
- HUGHES, G.F. On the mean accuracy of statistical pattern recognizers. **IEEE Transactions on Information Theory** v. 14, n. 1, 1968.
- MAHDIANPARI, M.; SALEHI, B.; REZAEI, M.; MOHAMMADIMANESH, F.; ZHANG, Y. Very deep convolutional neural networks for complex land cover mapping using multispectral remote sensing imagery. **Remote Sensing**, v. 10, n. 7, 2018.
- MELGANI and L. BRUZZONE. **Classification of hyperspectral remote sensing images with Support vector machines**. *TGARS*, 42(8):1778–1790, 2004.
- PENTREATH, N. **Machine learning with Spark**. Packt Publishing: Birmingham, UK, 2015.
- SANTOS, A. C. S.; PEDRINI, H. A combination of  $k$ -means clustering and entropy filtering for band selection and classification in hyperspectral images, **International Journal of Remote Sensing**, 37:13, 3005-3020, 2016.
- SILVA, L. C. D. A. Análise e categorização dos padrões espectro-temporais de índices de vegetação de alvos agrícolas e permanentes oriundos do sensor Modis entre os anos-safras 2013/2014 e 2016/2017. **Trabalho de conclusão de curso** (Engenharia agrícola) Universidade do oeste do paran (UNIOESTE), Cascavel PR, 128p, 2017.
- TARABALKA. **Classification of Hyperspectral Data Using Spectral-Spatial Approaches**. Tese (Engenharia Eltrica), University of Iceland and Grenoble Institute of Technology, 2010.
- WIELAND, M.; PITTORE, M. Performance evaluation of machine learning algorithms for urban pattern recognition from multi-spectral satellite images., v. 6, n. 4, p. 2912–2939, 2014.
- ZHANG, C.; QIU, F. Mapping individual tree species in an urban forest using airborne lidar and hyperspectral imagery. **Photogrammetric Engineering & Remote Sensing**, v. 78, n. 10, p. 1079 - 1087, 2012.