

Algoritmos de Aprendizado de Máquina para classificação de soja em imagens Landsat-8

Laíza Cavalcante De Albuquerque Silva¹, Jonathan Richetti, Willyan Ronaldo Becker²,
Weverton Rodrigo Verica², Alex Paludo², Jerry Adriani Johann^{1,2}

¹Curso de Engenharia Agrícola, Universidade Estadual do Oeste do Paraná - Cascavel, PR.

²Programa de Pós-graduação em Engenharia Agrícola (PGEAGRI), Universidade Estadual do Oeste do Paraná - Cascavel, PR.

Email autor correspondente: laiza.cavalcante@hotmail.com

Artigo enviado em 10/07/2017, aceito em 25/09/2017.

Resumo: O estado do Paraná ocupa uma posição importante no cenário nacional na produção de soja, sendo responsável por mais de 18% da produção brasileira, produzindo mais do que o quarto maior produtor mundial, a China. Para monitorar a produção agrícola, a informação de área é fundamental neste processo. Diversas técnicas e métodos podem ser empregados, incluindo algoritmos de Aprendizado de Máquina (*Machine Learning*). Logo, este trabalho tem como objetivo comparar quatro técnicas de aprendizado de máquina para mapear a área de soja a partir de imagens do sensor Landsat-8 no estado do Paraná, durante a safra de verão 2013/2014. Os algoritmos empregados no processo foram o *Random Forest* (RF), *Model Averaged Neural Network* (MANet), *Classification and Regression Trees* (CART) e o *Extreme Learning Machine* (ELM). Todos receberam os mesmos dados de treinamento (38 áreas de soja e 28 áreas não-soja) e as bandas de 1 a 7 do sensor Landsat-8 no *tile* 223/77 do dia 18/12/2013. A avaliação da acurácia (Exatidão Global e índice Kappa) foi realizada para cada classificação, sendo que técnica MANet apresentou os melhores resultados (Kappa = 0,9993 e EG = 0,9505) e o ELM o pior (Kappa = 0,9980 e EG = 0,8855).

Palavras-chave: Random Forest, Neural Network, Extreme Machine Learning, CART, culturas de verão.

Comparative analyses techniques among Machine Learning to classify soybean areas in Western region of Paraná from Landsat-8 images

Abstract: The state of Paraná is an important soybean producer in the national scenario, being responsible for more than 18% of the Brazilian production. This share alone is bigger than China's, the fourth world producer. In order to monitor crops, it is essential to have concise information about the cultivated area, which can be calculated with several available techniques, including machine learning. Thus, this paper aims to compare four machine learning algorithms for mapping the soybean area, using Landsat-8 images of the west of Paraná during the 2013/2014 season. The algorithms compared in this study were the Random Forest (RF), Model Averaged Neural Network (MANet), Classification and Regression Trees (CART), and Extreme Learning Machine (ELM). All algorithms were trained with the same data (38 soybean areas and 33 non-soybean areas) and used the bands 1 to 7 from the Landsat-8 tile 223/77, from December 18th, 2013. Each method was rated per its accuracy (Global Accuracy – GA –

and Kappa index), where higher numbers mean the algorithm is considered better. The results showed that MANet presented the higher results (Kappa = 0,9993 and EG = 0,9505) and ELM technique the lowest (Kappa = 0,9980 e EG = 0,8855).

Keywords: Random Forest, Neural Network, Extreme Machine Learning, CART, summer crops.

Introdução

A produção agrícola brasileira representa uma parcela expressiva da economia do país, alavancando o desenvolvimento de regiões e inovações tecnológicas voltadas para o agronegócio, abrangendo indústrias de insumo, distribuição e logística, como também a unidade produtiva. Segundo a FAO (OECD/FAO, 2015) o setor agroindustrial brasileiro deve atingir, em 2024, o patamar de 69,4 milhões de hectares plantados com oleaginosas, grãos brutos, arroz, dentre outras culturas. Atualmente as principais culturas de verão paranaense são a soja e o milho (CONAB, 2016). Em virtude da importância agrícola, se faz necessário o acompanhamento agrícola para obtenção de estimativas confiáveis de safra, que são informações cruciais para os integrantes da cadeia agropecuária, desde o produtor até o consumidor (COUTINHO et al., 2013; JOHANN et al., 2012). Além de avaliar o potencial produtivo de cada região para atender a demanda crescente de alimento e também a gestão dos recursos ambientais (ATZBERGER, 2013; THENKABAIL, et al., 2012).

Em função da aquisição de informações a baixo custo, com maior antecedência e menor custo operacional, o uso do sensoriamento remoto associado a técnicas de mineração de dados para otimizar o mapeamento de culturas agrícolas apresenta-se em diversos trabalhos como uma alternativa viável. Silva Junior et al. (2017), Grzegozewski et al. (2016); Johann et al. (2012) e Souza et al. (2015) elaboraram metodologias para o

monitoramento de áreas agrícolas utilizando o sensor MODIS (*Moderate Resolution Imaging Spectroradiometer*). Embora eficazes, estas metodologias são onerosas, envolvendo vários processos manuais, demandando tempo de aperfeiçoamento para o resultado final. Além disso, estas metodologias utilizam o sensor MODIS, com resolução espacial de 250m, para execução do mapeamento.

Com finalidade de incrementar a representatividade dos alvos em estudo, pode ser uma opção, utilizar sensor Landsat 8, com resolução espacial de 30 m. De forma que, em busca de metodologias mais rápidas e automatizadas, os algoritmos de aprendizagem de máquina (*Machine Learning*) apresentam possibilidades a serem exploradas no setor agrícola.

O aprendizado de máquina é uma das etapas da Descoberta de Conhecimento em Banco de Dados (*Knowledge Discovery From Data - KDD*). A metodologia KDD baseia-se em uma sequência de processos iterativos e tem como base, encontrar relações ou padrões ocultos em bases de dados, a fim de converter dados brutos em informações úteis e conhecimento (FAYYAD et al., 1996; TAN et al., 2009). O KDD voltado a classificação de imagens pode ser operado de forma supervisionada e não supervisionada. Nos métodos supervisionados o usuário dispõe de informações que permite a identificação das classes, enquanto que o software enquadra os pixels das imagens nas classes informadas. Já o método não supervisionado, o próprio algoritmo define as classes por

semelhanças e homogeneidades dos pixels.

Neste sentido, algoritmos como *Random Forest* (RF), *Model Averaged Neural Network* (MANet), *Classification and Regression Trees* (CART) e *Extreme Learning Machine* (ELM) apresentam a capacidade de utilizar as informações de dados de treinamento (*TrainData*) para construir modelos que permitem a classificação dos dados de interesse. Neste contexto, o objetivo deste trabalho foi comparar os algoritmos de aprendizagem de máquina em relação à sua capacidade de mapear áreas de cultivo de soja utilizando-se de imagens Landsat-8 para a região oeste do Paraná.

Material e Métodos

Área em estudo

A área de estudo foi o tile 223/077 (Órbita/Ponto) compreendido

na mesorregião Oeste (Figura 1), localizada no estado do Paraná. O mapeamento da área de soja, para a safra 2013/2014, empregou a cena do dia juliano 352 de 2013 (18/12/2013) proveniente do sensor Landsat-8, abrangendo uma área de 5.288,89 mil ha. A cena selecionada contempla alvos de soja em diferentes estádios vegetativos, pelo fato do Paraná apresentar diferentes datas de semeadura (JOHANN et al., 2016), com intuito de representar os alvos de soja. Outro fator preponderante na seleção da cena do sensor é a aquisição de uma imagem com baixa incidência de nuvens, visto que estas prejudicam a identificação da cultura pelo algoritmo, afetando a sua capacidade de classificação de alvos (RUDORFF et al., 2007; SUGAWARA; RUDORFF; ADAMI, 2008).

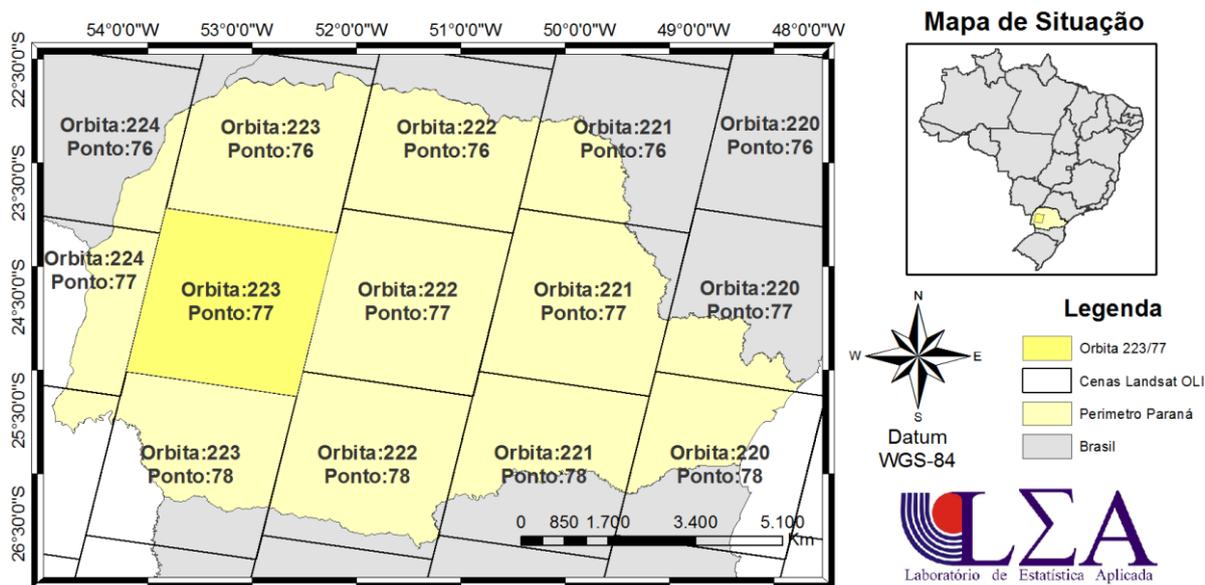


Figura 1. Mapa de localização do tile 223/77 na região Oeste do Estado do Paraná.

Dados de Entrada (Input)

A seleção dos dados de treinamento, *TrainData*, (Figura 2) empregou a composição RGB-564 (falsa cor) do Landsat-8, que proporcionou a identificação visual dos alvos em função de sua interação com a radiação

eletromagnética (RIZZI; RUDORFF, 2005). Logo, selecionou-se de forma aleatória e representativa, 38 áreas de soja e 28 áreas que não apresentam a cultura da soja (abrangendo alvos de milho, floresta, corpos hídricos, cidades e solo exposto). Além disso, não houve

validação de campo com os alvos selecionados.

Assim, todos os algoritmos receberam o mesmo *TrainData* de forma

dicotômica. Isto é, áreas de soja foram marcadas como 1 e áreas que não continham soja foram marcadas como 0 no *TrainData*.

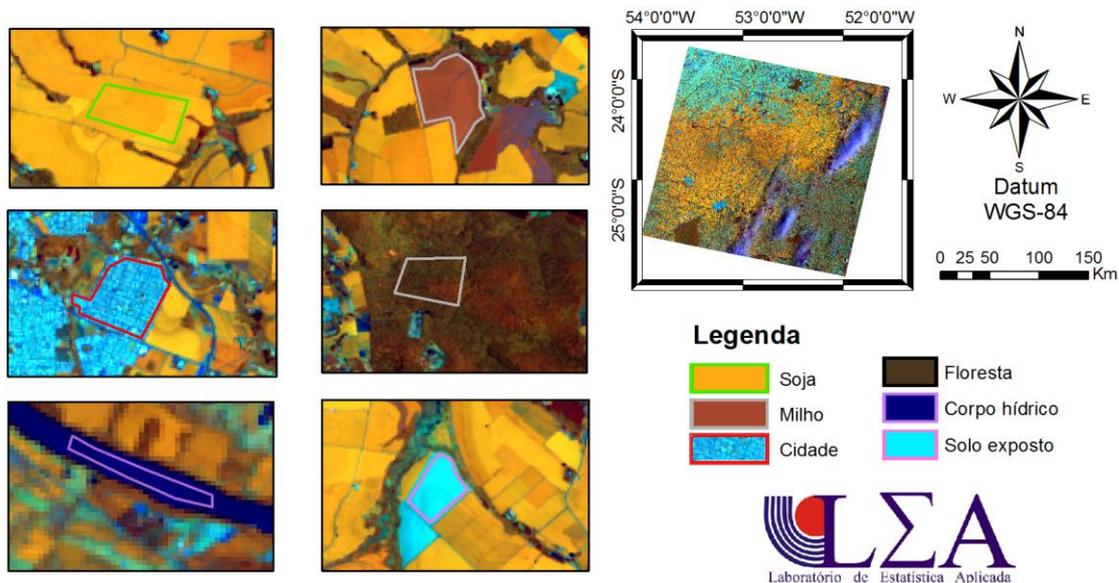


Figura 2. Composição RGB564 do sensor Landsat-8 exemplificando a seleção de dados de treinamento (*TrainData*).

O *input* dos algoritmos foram as bandas 1 a 7 do sensor Landsat-8 e o *TrainData*. Utilizou-se o software R versão 3.3.1 (R Development Core Team, 2016) e os métodos de classificação *Random Forest* (RF), *Model Averaged Neural Network* (MANet), *Classification and Regression Trees* (CART) e *Extreme Learning Machine* (ELM) em sua configuração *default*. Para isto, foram utilizados os pacotes: *raster* (HIJMANS, 2016); *caret* (KUHN et al, 2016); *rgdal* (BIVAND et al., 2016); *elm* (GOSSO, 2015); *rpartScore* (GALIMBERTI et al., 2012) e *nnet* (VENABLES & RIPLEY, 2002).

Random Forest (RF)

O algoritmo *Random Forest*, disponibilizado por Santacruz (2015), é um método de aprendizado não supervisionado que se propõe agrupar variáveis de entrada de dados por meio de várias árvores de decisão, construídas no momento de

treinamento do método (*TrainData*) (OSHIRO, 2013). O algoritmo cria múltiplas árvores de decisão, treinadas por meio da seleção aleatória de dois terços dos atributos contidos no *TrainData*, enquanto o um terço restante é empregado na validação cruzada da árvore gerada (BREIMAN, 2001). A saída do classificador é dada pela classe que foi retornada como resposta pela maioria das árvores pertencentes às classificações (TAN et al., 2009). Após a execução dos processos, o classificador tem como saída o índice Kappa, Exatidão Global (EG) e o mtry (número de preditores aleatoriamente utilizados, parâmetro do modelo de classificação), definido pela seleção aleatória de preditores como candidatos para realizar a separação de cada nó.

Model Averaged Neural Network (MANet)

No algoritmo MANet, o mesmo modelo de rede neural é ajustado

utilizando diferentes sementes de números aleatórios (vetores quaisquer do *TrainData*), apresentando pontuações no modelo que são calculadas pela média, sendo traduzidas posteriormente para cada uma das classes previstas (VENABLES & RIPLEY, 2002). Este modelo possui os seguintes parâmetros: *Nhid* e *decay*, que corresponde ao número de unidades escondidas e a perda de pesos, respectivamente.

Classification and Regression Trees (CART)

CART é um algoritmo não paramétrico, com o objetivo de investigar possíveis relações entre os dados de origem à procura de critérios de divisão. Estes critérios são organizados por meio de tomadas de decisão (nós), sendo subdivididos conforme o grau de similaridade entre as variáveis analisadas. A árvore de classificação é construída utilizando a função de impureza generalizada Gini, em que os custos de classificação errada são dados pelas diferenças absolutas ou quadradas nos escores atribuídos às categorias da resposta (GALIMBERTI; SOFFRITTI; MASO, 2012). Entretanto, estas árvores de classificação podem apresentar elevados níveis de complexidade, necessitando de uma otimização. A redução dos nós e ramificações (folhas) insignificantes da árvore (TIMOFEEV, 2004) é executada pelo parâmetro *cp* (GALIMBERTI; SOFFRITTI; MASO, 2012).

Extreme Learning Machine (ELM)

Extreme Machine Learning (ELM) é uma rede neural, organizada em três camadas: entrada (*input*), intermediária

(denominada como *Simple Hidden Layer Feedforward Networks* – SFLN) e saída (*output*), transmitindo os sinais em uma única direção (entrada-saída). O método de treinamento do SFLN não necessita de ajuste iterativo, nem de parâmetros de ajuste como taxa de aprendizado, etc., quando comparado a outros algoritmos tradicionais de aprendizagem baseados em gradiente (HUANG et al., 2012). O conceito principal do método está na inicialização aleatória dos pesos utilizados na camada SFLN, que tem como objetivo alcançar o menor erro de treinamento e a menor norma dos pesos de saída para alcançar boas generalizações (PENG; LI; TANG, 2013). Estes pesos são atribuídos às conexões entre os neurônios de entrada e os pesos contidos na SFLN, e entre as conexões dos neurônios contidos na SFLN, e os pesos de saída, não necessitando de ajuste inicial (HUANG; SIEW, 2005). Ao longo das camadas o método utiliza uma função de ativação, que tem como tarefa normalizar os valores de entrada num intervalo entre 0 e 1. Neste algoritmo conforme o *Nhid* (*Number of hidden neurons selected*) emprega-se as seguintes funções de ativação: seno (*sin*), linear (*purelinear*), Radbas (Base radial, ou RBF), Tansig (tangente hiperbólica).

Avaliação das classificações

Os procedimentos de classificação foram realizados com as sete bandas do Landsat-8 sem nenhum pré-processamento e utilizou-se o método *Bootstrap* com 25 repetições para amostragem e cálculo dos índices Kappa (Equação 1) e Exatidão Global (EG – Equação 2) (CONGALTON & GREEN, 1999).

$$IK = \frac{\left[n * \sum_{i=1}^r x_{ij} - \sum_{i=1}^r (x_{i.} \cdot x_{.j}) \right]}{\left[m^2 - \sum_{i=1}^r (x_{i.} \cdot x_{.j}) \right]} \quad (1)$$

$$EG = \frac{A}{m} * 100 \quad (2)$$

Em que: n = número de amostras; m = número de pontos amostrais; r = número de linhas da matriz de erro; x_{ij} = observação na linha i e coluna j; $x_{i.}$ = total marginal da linha i; $x_{.j}$ = total marginal da coluna, e A = acerto geral (pontos amostrais com acertos).

As classificações e as métricas do Kappa e Exatidão Global foram executadas a partir de uma adaptação do *script* disponibilizado por Santacruz (2015), elaborado na linguagem R. A partir destes índices, adotou-se como critério de avaliação dos métodos os que apresentaram a maior acurácia e EG. Com o intuito de avaliar os possíveis erros de omissão e inclusão nos mapeamentos elaborados, estes foram avaliados visualmente.

Resultados e Discussão

Visando a redução do tempo necessário para elaborar o mapeamento do uso e ocupação do solo, técnicas de

aprendizado de máquina vem sendo utilizados (CECHIM-JUNIOR; JOHANN; ANTUNES, 2017; SALMONA; PAIVA; MATRICARDI, 2016; SOUZA et al., 2016). Os quatro algoritmos analisados seguem duas heurísticas distintas em procedimentos de classificação em tarefas de mineração de dados; redes neurais e árvores de decisão, em que MANet e ELM definem-se como redes neurais, enquanto RF e CART como árvores de decisão. Embora sigam metodologias distintas, os métodos utilizados classificaram as áreas cultivadas com soja no *tile* 223/77 do sensor Landsat-8 (Tabela 1), apresentando EG e Kappa superiores a 0,95 e 0,88, respectivamente.

Tabela 1. Estatísticas de Exatidão Global e Kappa dos algoritmos CART, ELM, MANet e RF¹.

Algoritmo	Parâmetros		Exatidão Global	Kappa
	Nhid	Funções de Ativação		
CART		Cp		
		0,0	0,9980	0,9949
		0,5	0,9980	0,9949
		1,0	0,7398	0,0000
ELM	1	Sin	0,5662	0,0642
	1	Radbas	0,7656	0,0000
	1	Purelin	0,6670	0,0497
	1	Tansig	0,6647	0,1464
	3	Sin	0,5740	0,0724
	3	Radbas	0,7655	0,0002
	3	Purelin	0,9416	0,8619
	3	Tansig	0,7875	0,1250
	5	Sin	0,5768	0,0872
	5	Radbas	0,7651	0,0010
	5	Purelin	0,9505	0,8855

	5	Tansig	0,7869	0,0956
	Size	Decay		
MANet	1	0,0000	0,7351	0,0000
	1	0,0001	0,7444	0,0372
	1	0,1000	0,9566	0,8395
	3	0,0000	0,7351	0,0000
	3	0,0001	0,8481	0,4217
	3	0,1000	0,9993	0,9980
	5	0,0000	0,7431	0,0349
	5	0,0001	0,8724	0,5181
	5	0,1000	0,9991	0,9978
	Mtry			
RF		2	0,9990	0,9975
		4	0,9985	0,9961
		7	0,9985	0,9961

¹Os valores em negrito foram os adotados para a classificação.

Os critérios de escolha para selecionar o melhor método de classificação foram a Exatidão Global e o índice Kappa. Neste sentido, o algoritmo MANet apresentou maiores índices de acurácia (EG = 0,9993 e Kappa = 0,9980).

O algoritmo RF obteve o segundo melhor desempenho (EG = 0,9990 e Kappa = 0,9975) dentre os quatro métodos de aprendizado de máquina empregados. Este classificador é amplamente utilizado com dados de sensoriamento remoto para a classificação do uso e ocupação do solo, seja utilizando dados multiespectrais ou hiperespectrais, apresentando resultados satisfatórios em comparação aos demais métodos, conforme exposto na revisão bibliográfica sobre o método elaborada por Belgiu e Drăgu (2016). Abdel-Rahman (et al., 2014) comparou a eficiência dos classificadores RF e SVM (*Supported Vector Machine*) para detectar infestação de pragas em áreas de *pinus* através de imagens hiperespectrais, destacando que ambos apresentam resultados semelhantes (EG de 74,50% na classificação elaborada por RF e 73,50% para SVM, empregando todas as bandas espectrais do sensor

AISA Eagle). Na classificação do uso e ocupação do solo, integrando imagens do sensor Landsat-5 a um modelo digital do terreno do local em estudo, este também apresentou um EG de 92% e Kappa de 0,92, conforme exposto por (RODRIGUEZ-GALIANO et al., 2012).

Além disso, RF é considerado um algoritmo adequado para classificação de dados espectrais, apresentado a vantagem de não ser tão sensível ao sobre-ajuste dos dados de entrada (BELGIU; DRĂGU, 2016). Gislason, Benediktsson e Sveinsson (2006) elucidaram que este método é mais rápido no treinamento de dados, quando comparado com o método *boosting*. Embora CART e RF adotem árvores de decisão para executar o processo de classificação, o desempenho superior de RF justifica-se pelo seu método de classificação, já que este adota um conjunto de árvores de decisão para elaborar sua classificação, enquanto CART conta com uma única árvore (GISLASON; BENEDIKTSSON; SVEINSSON, 2006).

ELM é considerado um algoritmo robusto por apresentar elevada gama de funções de ativação e rápido processo de aprendizagem, quando comparado a

outros métodos como BP (*Back Propagation*) e SVM (WANG; CAO; YUAN, 2011). Chang (et al., 2010) empregou o classificador PL-ELM (uma variação do método ELM original) para classificação do uso e ocupação do solo através do sensor SPOT-5, resultando no Kappa de 0,8976 e EG de 92%. Acredita-se que o desempenho baixo, quando comparado aos demais métodos empregados, se justifica pelo processo de treinamento dos dados de entrada (*TrainData*), em razão deste ser responsável pela construção da classificação da imagem associado as funções de ativação empregadas, como também o número de camadas ocultas (CHEN et al., 2014).

O algoritmo ELM classificou uma área superior aos demais, totalizando 1.139 ha de soja, ou seja, um acréscimo de 8,17% de área em relação ao método MANet (Tabela 2 e Figura 3). Os algoritmos MANet, CART e RF resultaram em áreas de 1.053 ha, 1.027 ha e 952 ha, respectivamente. Embora a diferença de área entre as técnicas adotadas sejam relativamente pequenas (Tabela 2), uma avaliação visual dos mapeamentos – pela composição falsa-cor RGB-564 – evidencia erros de inclusão e omissão de alvos. A técnica ELM incluiu erroneamente áreas de milho (na tonalidade marrom) na

estimativa (Figura 4), resultando no acréscimo indevido de área. O algoritmo RF apresentou a menor área em função da não inclusão de alvos de soja no mapeamento (Figura 5).

Já Chan e Paelinckx (2008) avaliaram a classificação em mapeamentos de regiões ecologicamente homogêneas, com *airborne hyperspectral imagery*, e obtiveram EG de 0,6880 para RF e 0,6950 com o algoritmo *Adaboost*.

Mapeamentos de milho e soja com base na fenologia das culturas, adotando séries temporais provenientes do sensor MODIS, obtiveram menor acurácia a nível estadual. Zhong et al. (2016) obteve uma acurácia de 87,2% e índice Kappa de 0,8040. Grzegozewski et al. (2016) obteve EG de 0,8775 e Kappa de 0,7500 para o ano safra 2013/2014 e EG de 0,8675 e Kappa de 0,7400 para o ano safra 2012/2013 para a cultura da soja. Também para o Paraná, Souza et al. (2015) obteve EG de 0,8600 e Kappa de 0,7800 para o ano safra de 2010/2011 utilizando o índice WDRVI (*Wide Dynamic Range Vegetation Index*). Observa-se, portanto, que os algoritmos utilizados apresentaram índices de acurácia elevados e podem ser utilizados para mapeamentos de soja na região oeste do Paraná.

Tabela 2. Área da cultura da soja obtida pelos algoritmos CART, ELM, MANet e RF no mapeamento da cena 352 do sensor Landsat-8.

Cultura	Algoritmo	Área (mil ha)	Diferença MANet e Algoritmos (%)
Soja	CART	1.027	2,47
	ELM	1.139	8,17
	MANet	1.053	-
	RF	952	9,59

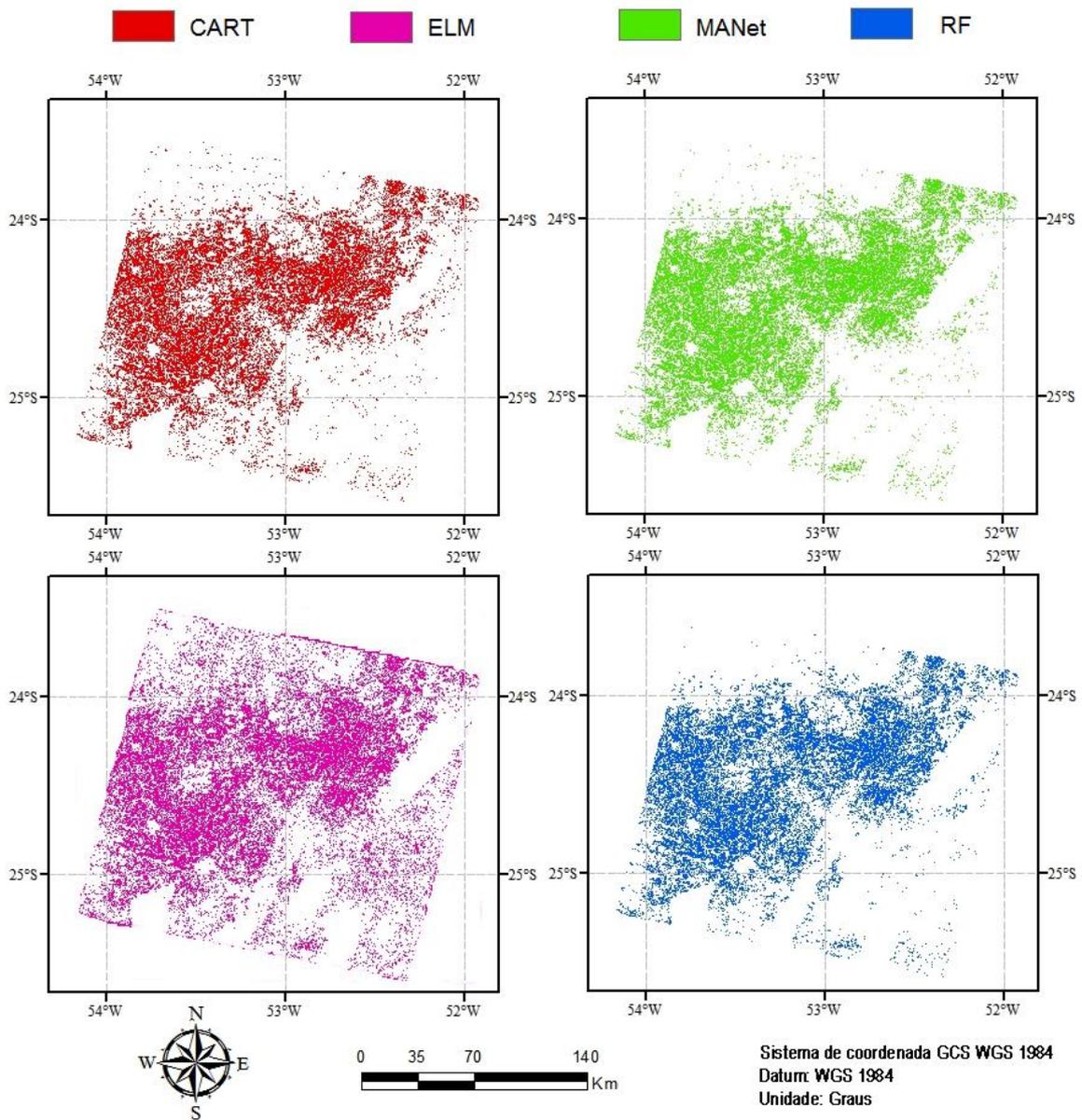


Figura 3. Mapeamento de soja do tile 223/077 na região Oeste do Estado do Paraná por meio dos algoritmos de aprendizado de máquina RF, CART, ELM e MANet.

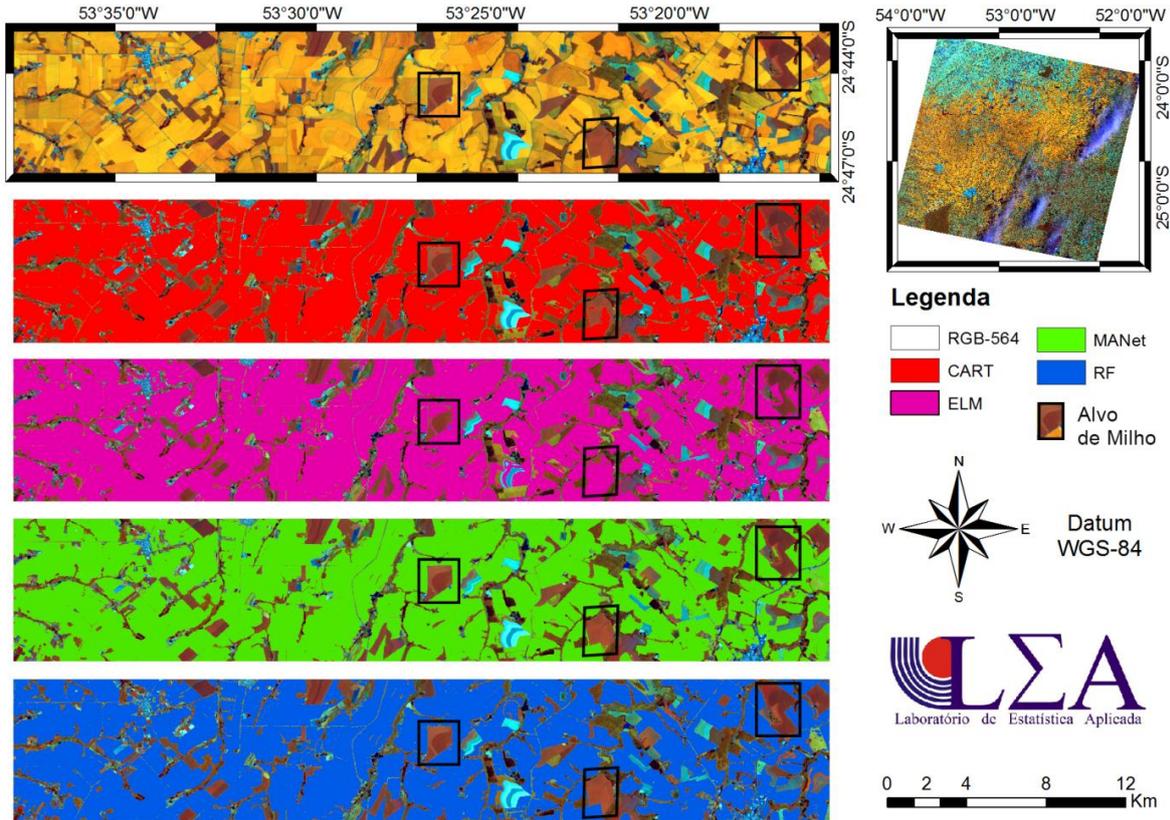


Figura 4. Comparativo dos alvos de milho entre os mapeamentos do tile 223/077 na região Oeste do Estado do Paraná.

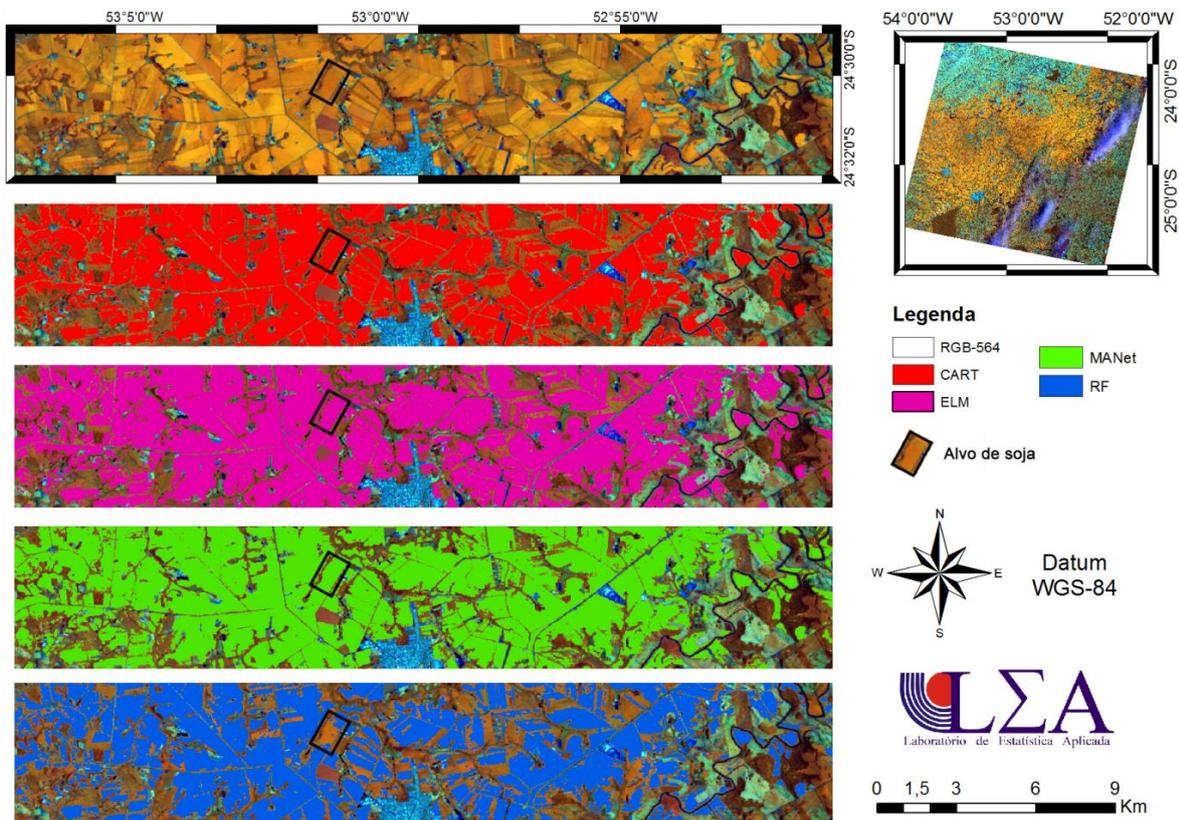


Figura 5. Comparativo dos alvos de soja entre os mapeamentos do tile 223/077 na região Oeste do Estado do Paraná.

Conclusões

O uso de algoritmos de aprendizagem de máquina em imagens de alta resolução do satélite Landsat-8 apresenta elevada capacidade de classificação de áreas cultivadas com soja no oeste do Paraná. Além disso, tem-se que a) todos os algoritmos apresentaram elevados valores de Exatidão Global e Kappa, indicando que todos podem ser utilizados para o mapeamento de soja na safra de verão; b) o algoritmo MANet (*Model Averaged Neural Network*) apresentou maiores valores de EG e Kappa; c) o algoritmo ELM (*Extreme Machine Learning*) apresentou os menores índices de acurácia entre os algoritmos estudados. As técnicas adotadas neste estudo demonstram vantagens na identificação de pequenos talhões de soja, em virtude da média resolução temporal do satélite e da qualidade dos algoritmos empregados. As técnicas de aprendizagem de máquina se mostraram eficazes e ágeis na tarefa de mapeamento de culturas agrícola e sua aplicação em escala estadual contribuirá no acompanhamento agrícola paranaense.

Referências

- ABDEL-RAHMAN, E. M.; MUTANGA, O.; ADAM, E.; ISMAIL, R. Detecting Sirex noctilio grey-attacked and lightning-struck pine trees using airborne hyperspectral data, random forest and support vector machines classifiers. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 88, p. 48–59, 2014.
- ATZBERGER, C. Advances in remote sensing of agriculture: Context description, existing operational monitoring systems and major information needs. **Remote Sensing**, v. 5, n. 2, p. 949–981, 2013.
- BELGIU, M.; DRĂGU, L. Random forest in remote sensing: A review of applications and future directions. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 114, p. 24–31, 2016.
- BIVAND, R.; KEITT, T.; ROWLINGSON, B. rgdal: Bindings for the Geospatial Data Abstraction Library. R package version 1.2-6. Disponível em: '<https://CRAN.R-project.org/package=rgdal>'. Acesso em 24 de mar. de 2017.
- BREIMAN, L. Random forests. **Machine Learning**, v. 45, n. 1, p. 5–32, 2001.
- CECHIM-JUNIOR, C.; JOHANN, J. A.; ANTUNES, J. F. G. Mapping of sugarcane crop area in the Paraná State using Landsat/TM/OLI and IRS/LISS-3 images. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 21, n. 6, 2017.
- CHANG, N.; HAN, M.; CHEN, L.; XU, S. Change detection of land use and land cover in an urban region with SPOT-5 images and partial Lanczos extreme learning machine. **Journal of Applied Remote Sensing**, v. 4, p. 1–15, 2010.
- CHAN, J. C. W.; PAELINCKX, D. Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. **Remote Sensing of Environment**, v. 112, n. 6, p. 2999–3011, 2008.
- CHEN, J.; ZHENG, G.; FANG, C.; ZHANG, N.; CHEN, H.; WU, Z. Time-series processing of large scale remote sensing data with extreme learning machine. **Neurocomputing**, v. 128, p. 199–206, 2014.
- CONAB (Companhia Nacional de Abastecimento). **Acompanhamento da safra brasileira de grãos, v.3 – safra**

2015/16, n. 4. Quarto levantamento, Brasília, p. 1-154, janeiro de 2016.

CONGALTON, R.G.; GREEN, K. **Assessing the accuracy of remotely sensed data: principles and practices**. Boca Raton: CRC Press, 1999. 160p.

COUTINHO, A. C.; ESQUERDO, J. C. D. M.; OLIVEIRA, L. S.; LANZA, D. A. Methodology for systematical mapping of annual crops in Mato Grosso do Sul state (Brazil). **Geografia**, Rio Claro, v. 38, p. 45-54, 2013.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. Knowledge Discovery and Data Mining: Towards a Unifying Framework. In: Proceedings of Second International Conference on Knowledge Discovery and Data Mining, 1996, Portland, Oregon. p.82-88.

GALIMBERTI, G.; SOFFRITTI, G.; MASO, M. Di. Classification Trees for Ordinal Responses in R: The rpartScore Package. **Journal of Statistical Software**, v. 47, n. 10, p. 25, 2012.

GISLASON, P. O.; BENEDIKTSSON, J. A.; SVEINSSON, J. R. Random forests for land cover classification. **Pattern Recognition Letters**, v. 27, n. 4, p. 294-300, 2006.

GOSSO, A. ElmNN: Implementation of ELM (Extreme Learning Machine) algorithm for SLFN (Single Hidden Layer Feedforward Neural Networks). R package version 1.0. 2015

GRZEGOZEWSKI, D. M.; JOHANN, J. A.; URIBE-OPAZO, M. A.; MERCANTE, E.; COUTINHO, A. C. Mapping soya bean and corn crops in the State of Paraná, Brazil, using EVI images from the MODIS sensor. **International Journal of Remote Sensing**, v. 37, n. 6, p. 1257-1275, 2016.

HIJMANS, R. J. Raster: Geographic data analysis and modeling. R package version 2.5-8. Disponível em: 'https://CRAN.R-project.org/package=raster'. Acesso em 12 de fev. de 2017.

HUANG G.-B.; ZHOU H.; DING X.; ZHANG R. Extreme Learning Machine for Regression and Multiclass Classification **IEEE Transactions on Systems, Man, and Cybernetics - part B: Cybernetics**, vol. 42, n. 2, 513-529, 2012

HUANG, G.; SIEW, C. Extreme Learning Machine with Randomly Assigned RBF Kernels. **International Journal of Information Technology**, v. 11, n. 1, p. 16-24, 2005.

JOHANN, J. A.; ROCHA, J. V.; DUFT, D. G.; LAMPARELLI, R. A. C. Estimativa de áreas com culturas de verão no Paraná, por meio de imagens multitemporais EVI/Modis. **Pesquisa Agropecuária Brasileira**, Brasília, v. 47, n. 9, p. 1295-1306, set. 2012.

JOHANN, J. A. et al. Uso de imagens do sensor orbital modis na estimação de datas do ciclo de desenvolvimento da cultura da soja para o estado do Paraná - Brasil. **Engenharia Agrícola**, Jaboticabal, v. 36, n. 1, p. 126-142, fev. 2016.

KUHN, M. Contributions from WING, J. WESTON, S. WILLIAMS, A. KEEFER, C. ENGELHARDT, A. COOPER, T. MAYER, Z. KENKEL, B. R Core Team, BENESTY, M. LESCARBEAU, M. ZIEM, A. SCRUCICA, L. TANG, Y. CANDAN, C. CARET: Classification and Regression Training. R package version 6.0-71. Disponível em: 'https://CRAN.R-project.org/package=caret'. Acesso em 02 de fev. 2017.

OECD/FAO. **Agricultural Outlook**

2015-2024: HIGHLIGHTS. Disponível em: 'http://www.fao.org/3/a-i4738e.pdf' Acesso em: 3 de jun. de 2017.

OSHIRO, T. M. **Uma abordagem para a construção de uma única árvore a partir de uma Random Forest para classificação de bases de expressão gênica.** 101p. Dissertação (Mestrado) – Universidade de São Paulo, Ribeirão Preto, 2013.

PENG, J.; LI, L.; TANG, Y. Y. Combination of activation functions in extreme learning machines for multivariate calibration. **Chemometrics and Intelligent Laboratory Systems**, v. 120, p. 53–58, 2013.

R Development Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2016.

RIZZI, R.; RUDORFF, B. F. T. Estimativa da área de soja no Rio Grande do Sul por meio de imagens Landsat. *Revista Brasileira de Cartografia*, v. 57, n. 3, p. 226–234, 2005.

RODRIGUEZ-GALIANO, V. F.; GHIMIRE, B.; ROGAN, J.; CHICA-OLMO, M.; RIGOL-SANCHEZ, J. P. An assessment of the effectiveness of a random forest classifier for land-cover classification. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 67, n. 1, p. 93–104, 2012.

RUDORFF, C. de M.; RIZZI, R.; RUDORFF, B. F. T.; SUGAWARA, L. M.; VIEIRA, C. A. O. Superfícies de resposta espectro-temporal de imagens do sensor MODIS para classificação de área de soja no Estado do Rio Grande do Sul. **Ciência Rural**, Santa Maria, v. 37, n. 1, p. 118–125, 2007.

SALMONA, Y. B.; PAIVA, A. O.; MATRICARDI, A. T. Estimativas futuras de desmatamento e emissões de CO₂ equivalente no Oeste baiano. **Revista Brasileira de Cartografia**, v. 68, n. 7, p. 1385–1395, 2016.

SANTACRUZ, A. **Image Classification with Random Forests in R (and QGIS).** Disponível em: 'http://amsantac.co/blog/en/2015/11/28/classification-r.html'. Acesso em: 30 de out. 2016.

SILVA JUNIOR, C. A. DA; NANININ, M. R.; TEODORO, P. E.; SILVA, G. F. C. Vegetation Indices for Discrimination of Soybean Areas: A New Approach. **Agronomy Journal**, v. 109, n. 4, p. 1–13, 2017.

SOUZA, C. G.; CARVALHO, L.; AGUIAR, P.; ARANTES, T. B. de aprendizagem de máquina e variáveis de sensoriamento remoto para o mapeamento da cafeicultura. **Boletim de Ciências Geodésicas**, v. 22, n. 4, p. 751–773, 2016.

SOUZA, C. H. de W.; MERCANTE, E.; JOHANN, J. A.; LAMPARELLI, R. A. C.; URIBE-OPAZO, M. A. Mapping and discrimination of soya bean and corn crops using spectro-temporal profiles of vegetation indices. **International Journal of Remote Sensing**, v. 36, n. 7, p. 1809–1824, 2015.

SUGAWARA, L. M.; RUDORFF, B. F. T.; ADAMI, M. Viabilidade de uso de imagens do Landsat em mapeamento de área cultivada com soja no Estado do Paraná. **Pesquisa Agropecuária Brasileira**, Brasília, v.43, n. 12, p. 1777–1783, 2008.

TAN, P.; STEINBACH, M.; KUMAR, V. **Introdução ao datamining: mineração de dados.** Rio de Janeiro:

Editora Ciência Moderna, 2009. 928p.

THENKABAIL, P. S.; KNOX, J. W.; OZDOGAN M.; GUMMA M. K.; CONGALTON, R. G.; WU, Z.; MILESI, C.; FINKRAL, A.; MARSHALL, M.; MARIOTTO, I. M.; YOU, S.; GIRI, C. NAGLER, P. Assessing future risks to agricultural productivity, water resources and food security: how can remote sensing help? **Photogrammetric Engineering and Remote Sensing**, v. 78, n. 8, p. 773-782, 2012.

TIMOFEEV, R. **Classification and regression trees (CART) Theory and Applications**. 2004. 39p. Dissertação (mestrado) – Humboldt University, Berlin, 2004.

VENABLES, W. N. & RIPLEY, B. D. **Modern Applied Statistics with S**. Nova Iorque: Editora: Springer-Verlag, 2002. 498p.

ZHONG, L; HU, L; YU, L; GONG, P, BIGING, G. S. Automated mapping of soybean and corn using phenology. **ISPRS Journal of Photogrammetry and Remote Sensing**, v.119 p. 151-164, 2016.

WANG, Y.; CAO, F.; YUAN, Y. A study on effectiveness of extreme learning machine. **Neurocomputing**, v. 74, p.2483-2490, 2011.