



Análise do Uso e Cobertura da Terra Utilizando Imagens Sentinel-2A e Inteligência Artificial Analysis Land Use and Land Cover Using Sentinel-2A Images and Artificial Intelligence

Matheus Frigo Wolfer¹; Juliana Marchesan²; Elisiane Alba³, Mateus Schuh²;
Dionatas Henrique Honnef¹; Helena Silva Oliveira¹ & Rudiney Soares Pereira¹

¹Universidade Federal de Santa Maria, Departamento de Engenharia Rural, Av. Roraima, n° 1000,
Campus UFSM, Prédio 44, Bairro Camobi, CEP 97105-900, Santa Maria, RS, Brasil

²Universidade Federal de Santa Maria, Programa de Pós-graduação em Engenharia Florestal, Av. Roraima,
n° 1000, Campus UFSM, Prédio 44, Bairro Camobi, CEP 97105-900, Santa Maria, RS, Brasil

³Universidade Federal Rural de Pernambuco, Unidade Acadêmica de Serra Talhada, Av. Gregório Ferraz
Nogueira, S/N - José Tomé de Souza Ramos - CEP: 56909-535, Serra Talhada, PE, Brasil
E-mails: matheuswolfer@gmail.com; marchesan.ju@gmail.com; elisiane.alba@ufrpe.br; mateuschuh@gmail.com;
dionatas.honnef@gmail.com; oliveira.h23@gmail.com; rudiney.s.pereira@gmail.com

Recebido em: 15/06/2020 Aprovado em: 01/10/2020

DOI: http://dx.doi.org/10.11137/2020_4_395_403

Resumo

O objetivo do presente estudo consistiu em avaliar a eficiência dos algoritmos de aprendizado de máquina *Random Forest* (RF), *Support Vector Machine* (SVM) e *Artificial Neural Network* (ANN) na classificação supervisionada de usos e cobertura da terra utilizando imagens MSI/Sentinel-2A na bacia hidrográfica do Rio Pardo – RS. O processo de classificação foi conduzido por meio do *software* R utilizando o pacote *caret*. Os algoritmos foram avaliados utilizando os valores de acurácia de 30 repetições da validação cruzada. Para cada repetição foram atribuídos pesos para ordenamento dos algoritmos de acordo com a eficiência, de modo a verificar se os mesmos se diferiram estatisticamente utilizando o teste de Friedman e Nemenyi. O algoritmo RF apresentou o melhor resultado de acurácia média (0,9973) seguido do SVM (0,9937) e ANN (0,5628). Os algoritmos se diferiram estatisticamente entre si, sendo possível inferir que o RF pode ser utilizado de forma eficiente para classificação do uso e cobertura da terra, por meio de imagens MSI/Sentinel-2A. Neste contexto, o presente estudo demonstrou que os algoritmos de aprendizado de máquina aliados às imagens de média resolução espacial possibilitam adquirir resultados de boa precisão e confiabilidade, os quais poderão auxiliar em estudos ambientais futuros, permitindo a obtenção de resultados de forma rápida e eficiente.

Palavras-chave: *Aprendizado de máquina; Classificação supervisionada; Linguagem R*

Abstract

The aim of the present study was to evaluate the efficiency of machine learning algorithms *Random Forest* (RF), *Support Vector Machine* (SVM) and *Artificial Neural Network* (ANN) supervised classification of land use and land cover using images MSI/Sentinel-2A in the hydrographic basin of Rio Pardo – RS. The classification process was conducted using the *software* R using the *caret* package. The algorithms were evaluated using the accuracy values of 30 repetitions of cross-validation. Weights were assigned to each repetition for ordering the algorithms according to efficiency, in order to verify if they differed statistically using the Friedman and Nemenyi test. The RF algorithm presented the best result of average accuracy (0.9973) followed by SVM (0.9937) and ANN (0.5628). The algorithms differed statistically from each other, being possible to infer that the RF can be used efficiently to classify land use and land cover, through images MSI/Sentinel-2A. In this context, the present study demonstrated that machine learning algorithms combined with medium spatial resolution images acquire results of good precision and reliability, which can assist in future environmental studies, allowing results to be obtained quickly and efficiently.

Keywords: *Machine learning; Supervised classification; Language R*

1 Introdução

As mudanças ocorridas no bioma Mata Atlântica tiveram seu ponto de partida na era colonial do Brasil, primeiramente com o ciclo do Pau Brasil logo em seguida com o ciclo da cana de açúcar, onde este último usava lenha proveniente do bioma para alimentar as caldeiras visando beneficiamento do produto (Young, 2006). Durante décadas, a falta de planejamento do uso da terra tem acarretado em diversos impactos negativos para o meio ambiente, como a redução dos remanescentes florestais.

Neste contexto, o mapeamento do uso e cobertura da terra possibilita que sejam realizadas análises da distribuição dos remanescentes naturais presentes na paisagem e das atividades antrópicas que ocorrem na superfície terrestre. Assim, por meio de estudos que visem a caracterização da cobertura da terra, juntamente com suas respectivas formas de manejo e de uso, torna-se importante analisar métodos de avaliar a intensidade e o tipo de mudanças ocorridas ao longo do tempo (Seabra *et al.*, 2015). Para tal, pode-se fazer uso das técnicas de sensoriamento remoto, as quais proporcionam agilidade no processamento dos dados (Vaeza *et al.*, 2010).

Os avanços tecnológicos no sensoriamento remoto têm permitido desde a década de 70, realizar o mapeamento do uso e cobertura da terra em grande escala por meio de imagens orbitais. As imagens provenientes do sensor MSI (*Multispectral Instrument*), a bordo do satélite Sentinel-2A, por serem disponibilizadas gratuitamente e possuírem resolução espacial média tornam-se uma base de dados promissora para estudos que visem mapear o uso e cobertura da terra, além de possibilitarem o monitoramento da superfície terrestre devido à revisita periódica do satélite. Estudos têm sido desenvolvidos nesse âmbito, como por exemplo o de Freires *et al.* (2019) e Abdi (2019), demonstrando que as imagens MSI/Sentinel-2A apresentam resultados satisfatórios quanto ao mapeamento de usos e cobertura da terra.

De modo geral, as técnicas de classificação do uso e cobertura da terra a partir de imagens de satélite podem ser categorizadas como métodos supervisionados ou não supervisionados (Li *et al.*, 2014). Na abordagem supervisionada, o usuário define as classes e as respectivas amostras que irão compor o treinamento (Andrade *et al.*, 2014). Já no processo não supervisionado o algoritmo utiliza estatísticas e técnicas de agrupamento de dados, sem a necessidade do conhecimento prévio das classes a serem criadas (Ge *et al.*, 2020). Recentemente, com os avanços na ciência da computação e nas tecnologias GIS, os métodos de classificação supervisionados têm sido amplamente usados, uma vez que são mais robustos e precisos do que as abordagens não supervisionadas (Niemeyer *et al.*, 2014; Jamali, 2019). Os métodos

supervisionados incluem algoritmos paramétricos, como máxima verossimilhança, distância mínima e classificadores Bayesianos, e algoritmos não paramétricos, como algoritmos de aprendizado de máquina (Phiri & Morgenroth, 2017), que comparativamente aos métodos paramétricos usuais, têm demonstrado precisão e eficiência superior ao trabalhar com dados de maior complexidade e no mapeamento de grandes áreas (Rodríguez-Galiano *et al.*, 2012; Song *et al.*, 2019), tornando-se cada vez mais populares para classificação de dados de sensoriamento remoto (Belgiu & Drăgu, 2016).

Dentre os algoritmos de aprendizado de máquina o *Random Forest* (RF), *Support Vector Machine* (SVM) e *Artificial Neural Network* (ANN) têm demonstrado bom desempenho para a classificação temática da paisagem (Li *et al.*, 2014). O método RF (Breiman, 2001) é um algoritmo de aprendizagem de conjunto projetado para melhorar a classificação linear tradicional e o método de árvore de regressão. Ele integra um grande conjunto de árvores de decisão com base em uma técnica determinística, selecionando um conjunto aleatório de variáveis e uma amostra aleatória dos dados de treinamento (Shao & Zhang, 2016).

Desenvolvido por Cortes & Vapnik (1995) o método SVM é baseado no aprendizado estatístico e no conceito de planos de decisão que define os limites de decisão. As observações que possuem distância mínima até o hiperplano são denominadas vetores de suporte. De acordo com Merschmann (2007), a predição de uma nova observação é feita observando sua posição em relação ao hiperplano de separação ajustado. Os algoritmos de ANN, por sua vez, são sistemas compostos de unidades de processamento densamente conectadas, denominadas neurônios artificiais, os quais computam funções matemáticas e seu treinamento consiste em fornecer à uma arquitetura pré-estabelecida páreas de padrões, sendo um valor de entrada e outro de saída (Monjezi *et al.*, 2010). Uma ANN é caracterizada por dois aspectos fundamentais, sua arquitetura, relacionada ao número e arranjo dos neurônios conectados, e o aprendizado, que se refere às regras de ponderação e ao tipo de informação que deve ser utilizada no ajuste dos pesos (Faceli *et al.*, 2011).

Neste contexto, abrem-se novos caminhos para que sejam testadas técnicas de classificação do uso e cobertura da terra, por meio de algoritmos de aprendizado de máquina, utilizando imagens MSI/Sentinel-2A. Uma vez que, devido ao recente lançamento do satélite Sentinel-2A, existem poucos estudos sobre a utilização do mesmo para classificação do uso e cobertura da terra, principalmente utilizando técnicas de inteligência artificial.

A área de estudo, a bacia hidrográfica do Rio Pardo, situa-se no Rio Grande do Sul e está sob domínio do bioma Mata Atlântica, e vem enfrentando, ao longo dos anos, problemas com o desmatamento, principalmente nas

encostas de seus rios o que acaba por gerar aumento na erosão superficial e posterior assoreamento dos mesmos (SEMA, 2019). Diante disso, o mapeamento do uso e cobertura da terra na bacia possibilitará quantificar os diferentes usos presentes na área, permitindo ações de planejamento para recuperação de áreas ou futuras conversões da cobertura da terra. Desse modo, o objetivo do presente estudo consistiu em avaliar a eficiência dos algoritmos de aprendizado de máquina RF, SVM e ANN na classificação supervisionada de usos e cobertura da terra utilizando imagens MSI/Sentinel-2A na bacia hidrográfica do Rio Pardo – RS.

2 Caracterização da Área de Estudo

A área de estudo consiste na bacia hidrográfica do Rio Pardo, Rio Grande do Sul, a qual abrange os municípios das regiões da depressão central e do planalto meridional como Candelária, Santa Cruz do Sul, Rio Pardo, Soledade e Venâncio Aires. Engloba os cursos de água do Rio Pardinho, Arroio Francisco Alves, Rio Pequeno, Arroio Andréas e Rio Pardo (SEMA, 2019). A bacia possui coordenada geográfica central de 29°30'41,86" de latitude sul e 52°39'10,60" de longitude oeste (Figura 1), abrangendo uma área total de aproximadamente 3.658 Km².

A bacia se encontra sob domínio do bioma Mata Atlântica, tendo como principal tipo fitogeográfico a Floresta Estacional Decidual (SEMA, 2003). Os plantios florestais apresentam importância econômica na região, principalmente o pinus e eucalipto, uma vez que, são utilizados como combustível para secagem do tabaco (Farias, 2010).

A área de estudo está inserida no Conselho Regional de Desenvolvimento (COREDE) Vale do Rio Pardo, no qual a economia dos municípios é de base agrícola, com predomínio dos cultivos de fumo, milho, soja e arroz (IBGE, 2018). Devido à presença de agricultura irrigada o principal uso dos afluentes da região destina-se ao setor agrícola (SEMA, 2019).

3 Material e Métodos

3.1 Delimitação da Bacia

A delimitação da área da área de estudo, a qual consistiu na Bacia do Rio Pardo, foi realizada com o auxílio do *software* Qgis versão 2.16.3 por meio dos módulos do GRASS 7.0.4 utilizando a ferramenta *watershed*, com a qual foi possível extrair os pontos divisores do fluxo de água da bacia. Para tanto, foi utilizado o arquivo vetorial

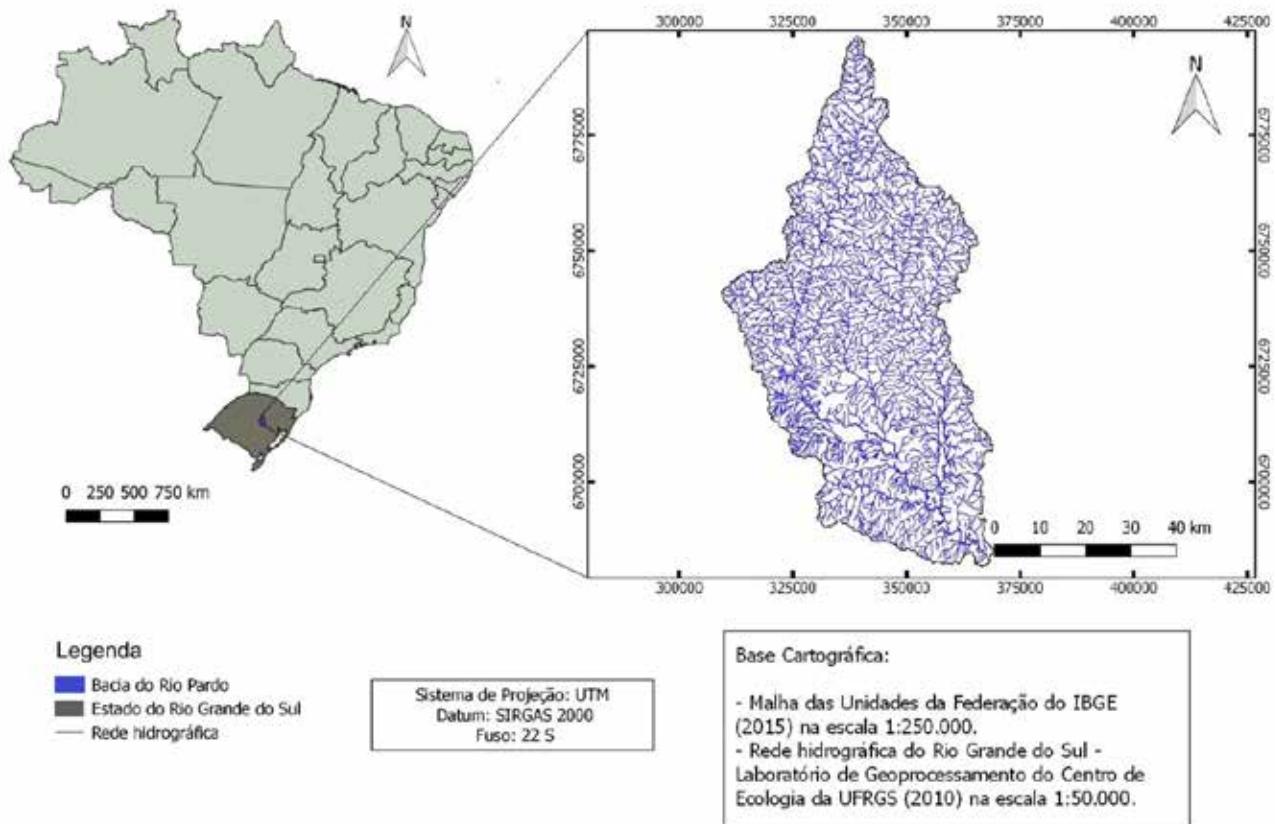


Figura 1 Localização da bacia hidrográfica do Rio Pardo, RS.

da hidrografia do estado do Rio Grande do Sul, zona 22S na escala 1:50.000, disponibilizado pelo Laboratório de Geoprocessamento do Centro de Ecologia da UFRGS (Hasenack & Weber, 2010) e imagens *Shuttle Radar Topography Mission* (SRTM), disponibilizadas pela Embrapa com resolução espacial de 90 metros (Miranda, 2005), em que foi necessário realizar o mosaico destas para que obtivesse a abrangência de toda a bacia.

3.2 Processamento dos Dados Sentinel-2A

As imagens utilizadas foram oriundas do sensor MSI, abordo do satélite Sentinel-2A, com resolução espacial de 10 metros para as bandas B02, B03, B04, B08 e de 20 metros para as bandas B05, B06, B07, B08A, B11 e B12. Para abranger toda área da bacia do Rio Pardo, foram necessárias duas cenas, ambas datadas de 3 de abril de 2019, disponibilizadas pelo *United States Geological Survey* (USGS, 2019). No *software* SPRING versão 5.4.2 foram importadas as dez bandas, nas quais foi aplicada a composição RGB resultando na criação de duas imagens sintéticas denominadas falsa cor e cor verdadeira, para que fosse possível a identificação dos usos e cobertura da terra na área da bacia.

As duas imagens sintéticas geradas foram importadas para o Qgis para que fosse realizada a amostragem das classes de uso e cobertura da terra. Para a realização da amostragem foi criado um arquivo vetorial contendo polígonos correspondentes as seis classes, sendo essas Floresta nativa, Floresta plantada (plantios de pinus e eucalipto), Campo, Agricultura, Água e Solo exposto.

3.3 Classificação Supervisionada por Algoritmos de Aprendizado de Máquina

Para a realização da classificação supervisionada utilizou-se três algoritmos de aprendizado de máquina, sendo eles o SVM, RF e ANN. Para tanto foram utilizados os pacotes *caret* (Kuhn, 2018), *rgdal* (Bivand *et al.*, 2018) e *raster* (Hijmans *et al.*, 2017), disponíveis no *software* R versão 3.5.1 (R Development Core Team, 2018).

Inicialmente foram carregadas, na interface gráfica R Studio, as bandas espectrais do MSI/Sentinel-2A, sendo

estas B2, B3, B4, B5, B6, B7, B8, B8A, B11 e B12. Utilizando o arquivo vetorial contendo as seis classes amostrais, foi gerada uma planilha contendo os valores dos pixels (Números Digitais) referentes a cada uma das bandas espectrais. Os valores obtidos do cruzamento das bandas e do respectivo tipo de uso e cobertura da terra, correspondeu a base de dados do processo de treinamento dos algoritmos.

A escolha do melhor algoritmo de aprendizado de máquina para a classificação do uso e cobertura da terra se deu por meio da validação cruzada, do tipo *k-fold*. A avaliação do ajuste dos algoritmos na validação cruzada, baseou-se em 30 repetições, examinando o valor do índice de acurácia global (exatidão global) de cada um dos algoritmos de aprendizado de máquina. A acurácia global foi o parâmetro utilizado pois informa o quanto um valor está correto e isento de erros, mostrando-se um importante parâmetro para a avaliação de imagens classificadas.

Utilizando os valores de acurácia das 30 repetições para cada um dos algoritmos, foram atribuídos pesos para que fosse ordenado de acordo com a eficiência. Em vista disso, o peso utilizado para o algoritmo de maior ajuste foi o valor 1 e assim sucessivamente. Para comparar os resultados das acurácias e verificar se os algoritmos diferiram significativamente entre si, foi utilizado o teste estatístico de Friedman e Nemenyi (Friedman, 1940; Nemenyi, 1969).

4 Resultados e Discussão

A Tabela 1 demonstra os valores de acurácia global e *kappa* gerados por meio da classificação para os algoritmos de aprendizado de máquina, considerando o conjunto de dados de treinamento. Os algoritmos RF e SVM apresentaram valores de acurácia global e *kappa* próximos quando comparados com o ANN. Observou-se um desempenho inferior do algoritmo ANN ao considerar o conjunto de dados de treinamento quando destinado à identificação dos diferentes usos e cobertura da terra. Visando fornecer uma análise mais completa do processo de modelagem por aprendizado de máquina, a eficiência dos algoritmos foi avaliada por meio do conjunto de dados de validação.

Algoritmos	Acurácia Global	Kappa
RF	0,9956	0,9944
SVM	0,9937	0,9918
ANN	0,6279	0,4966

Tabela 1 Resultado da classificação supervisionada para os algoritmos RF, SVM e ANN a partir do conjunto de dados de treinamento.

Repetições	Acurácia Global			Rank		
	RF	SVM	ANN	RF	SVM	ANN
1	0,9973	0,9937	0,5233	1	2	3
2	0,9973	0,9936	0,6049	1	2	3
3	0,9973	0,9937	0,5064	1	2	3
4	0,9973	0,9936	0,5630	1	2	3
5	0,9972	0,9936	0,5808	1	2	3
6	0,9971	0,9936	0,4995	1	2	3
7	0,9972	0,9937	0,5693	1	2	3
8	0,9973	0,9937	0,6102	1	2	3
9	0,9973	0,9936	0,5216	1	2	3
10	0,9972	0,9937	0,5867	1	2	3
11	0,9971	0,9936	0,5858	1	2	3
12	0,9973	0,9936	0,5494	1	2	3
13	0,9971	0,9937	0,5348	1	2	3
14	0,9972	0,9937	0,5753	1	2	3
15	0,9972	0,9936	0,6644	1	2	3
16	0,9973	0,9936	0,5505	1	2	3
17	0,9971	0,9936	0,5392	1	2	3
18	0,9972	0,9935	0,5365	1	2	3
19	0,9973	0,9937	0,5025	1	2	3
20	0,9971	0,9936	0,5302	1	2	3
21	0,9971	0,9937	0,6032	1	2	3
22	0,9976	0,9936	0,6559	1	2	3
23	0,9974	0,9937	0,5595	1	2	3
24	0,9971	0,9936	0,4880	1	2	3
25	0,9971	0,9936	0,5780	1	2	3
26	0,9973	0,9936	0,5561	1	2	3
27	0,9972	0,9936	0,5831	1	2	3
28	0,9972	0,9936	0,6031	1	2	3
29	0,9971	0,9937	0,5380	1	2	3
30	0,9973	0,9935	0,5845	1	2	3
Média	0,9973	0,9937	0,5628	1	2	3

Tabela 2 Valores de acurácia global e o *rank* obtidos por meio das 30 repetições para os algoritmos RF, SVM, ANN.

4.1 Validação da Classificação

A Tabela 2 mostra os valores que foram obtidos para os índices de acurácia global gerados por meio de 30 repetições e o *rank* dos algoritmos. Os classificadores RF e SVM apresentaram valores de acurácia global médio próximos, os quais foram de 0,9973 e 0,9937, respectivamente, já a ANN apresentou os menores valores quando comparado aos demais, demonstrando que houve uma maior confusão no processo de classificação supervisionada.

O ranqueamento dos algoritmos indicou que o RF obteve os melhores resultados (Tabela 2) em todas as repetições, contudo, para verificar se os algoritmos se diferiram estatisticamente foi utilizado o teste de Friedman e Nemenyi. Por meio do qual foi possível inferir que o RF obteve diferença em relação aos demais algoritmos utilizados no estudo (Figura 2).

Dessa forma, os resultados indicaram que o RF pode ser utilizado de forma eficiente para classificação do uso e cobertura da terra por meio de imagens Sentinel-2A em áreas com características semelhantes à de estudo.

Segundo Farda (2017), os algoritmos de aprendizado de máquina supervisionados baseados em decisões (geralmente chamados de algoritmos lógicos), como o RF e SVM, apresentam maior precisão. Os valores de acurácia global obtidos para a ANN, por sua vez, foram sempre mais baixos que os demais algoritmos, fato este que pode estar relacionado ao número de amostras utilizado. Farda (2017) destacaram que a ANN requer um grande conjunto de amostras de treinamento para melhorar a precisão dos resultados.

Além disso, a ANN pode ter sofrido sobreajuste (*overfitting*) no processo de treinamento, ocasionando assim, baixa capacidade de prever dados nunca vistos (Zhu & Liu, 2015). Nesse sentido, o *overfitting* corresponde ao processo pelo qual o modelo se ajusta ao conjunto de dados de treinamento, apresentando estatísticas elevadas de ajuste, enquanto apresenta dificuldade para descrever o comportamento de um novo conjunto de dados. Já os algoritmos RF e SVM são mais adequados para menores conjuntos de dados (Zhu *et al.*, 2019), como o caso do presente estudo.

Rosa (2018) comparou algoritmos de aprendizado de máquina para a classificação da área urbana da cidade de São Paulo, utilizando como base imagem MSI/Sentinel-2A, e obteve resultados de acurácia global superior para o RF em comparação ao SVM, com valores de 0,872 e 0,843, respectivamente. Estes resultados corroboram com os obtidos na presente pesquisa, o que demonstra um melhor desempenho para o algoritmo RF quando utilizadas imagens Sentinel-2A.

Contudo, o estudo realizado por Gaiad *et al.* (2017) demonstrou que utilizando como base imagem OLI/Landsat 8 para a classificação supervisionada, o SVM resultou em valores de acurácia global de 0,9832, apresentado um melhor desempenho ao comparar com o RF, o qual obteve valor de acurácia global de 0,9526. Assim, levando-se em consideração os valores de acurácia global obtidos pelos autores, nota-se o inverso nos resultados obtidos na presente pesquisa, fato este que pode estar relacionado a base de dados utilizada para a classificação dos usos e cobertura da terra.

4.2 Mapeamento dos Usos e Cobertura da Terra

A Figura 3 apresentam o resultado da classificação supervisionada obtendo as seis classes de uso e cobertura da terra mapeadas neste estudo. Por meio do mapeado foi possível obter a quantificação das classes temáticas para cada algoritmo (Tabela 3).

Conforme discutido anteriormente, a exatidão global demonstrou que o RF e o SVM obtiveram valores próximos, o que corrobora com os resultados encontrados para o mapeamento dos usos e cobertura da terra para a área total de estudo, no qual, por meio da quantificação (Tabela 3), nota-se que em todas as classes temáticas os valores foram semelhantes, enquanto que a ANN obteve valores destoantes. O algoritmo ANN apresentou dificuldades para explicar a variabilidade do comportamento espectral da floresta plantada

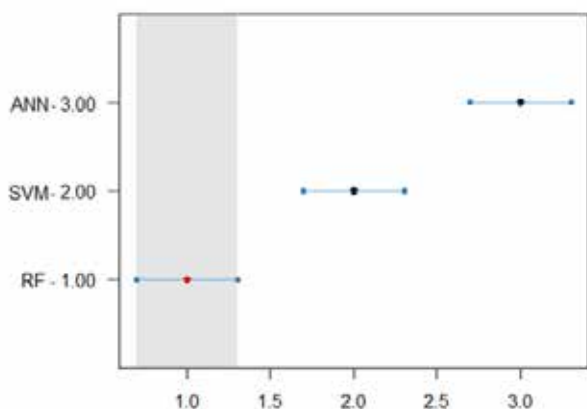


Figura 2 Teste estatístico de Friedman e Nemenyi para os algoritmos RF, SVM e ANN ao nível de significância de 5%.

Classes Temáticas	RF		SVM		ANN	
	ha	%	ha	%	ha	%
Floresta nativa	178.247,42	49,15	175.245,62	48,32	169.375,72	46,07
Floresta plantada	12.553,48	3,46	14.732,94	4,06	0,00	0,00
Solo exposto	31.424,26	8,67	30.769,21	8,48	67.875,98	18,72
Agricultura	33.327,37	9,19	33.612,49	9,27	26.175,72	7,22
Campo	104.544,27	28,83	105.724,47	29,15	99.226,38	27,36
Água	2.557,00	0,71	2.569,07	0,71	0,00	0,00
Total	362.653,80	100	362.653,80	100	362.653,80	100

Tabela 3 Quantificação das classes temáticas de uso e cobertura da terra para os algoritmos RF, SVM e ANN na bacia hidrográfica do Rio Pardo – RS.

Análise do Uso e Cobertura da Terra Utilizando Imagens Sentinel-2A e Inteligência Artificial
Matheus Frigo Wolfer; Juliana Marchesan; Elisiane Alba, Mateus Schuh; Dionatas Henrique Honnef;
Helena Silva Oliveira & Rudiney Soares Pereira

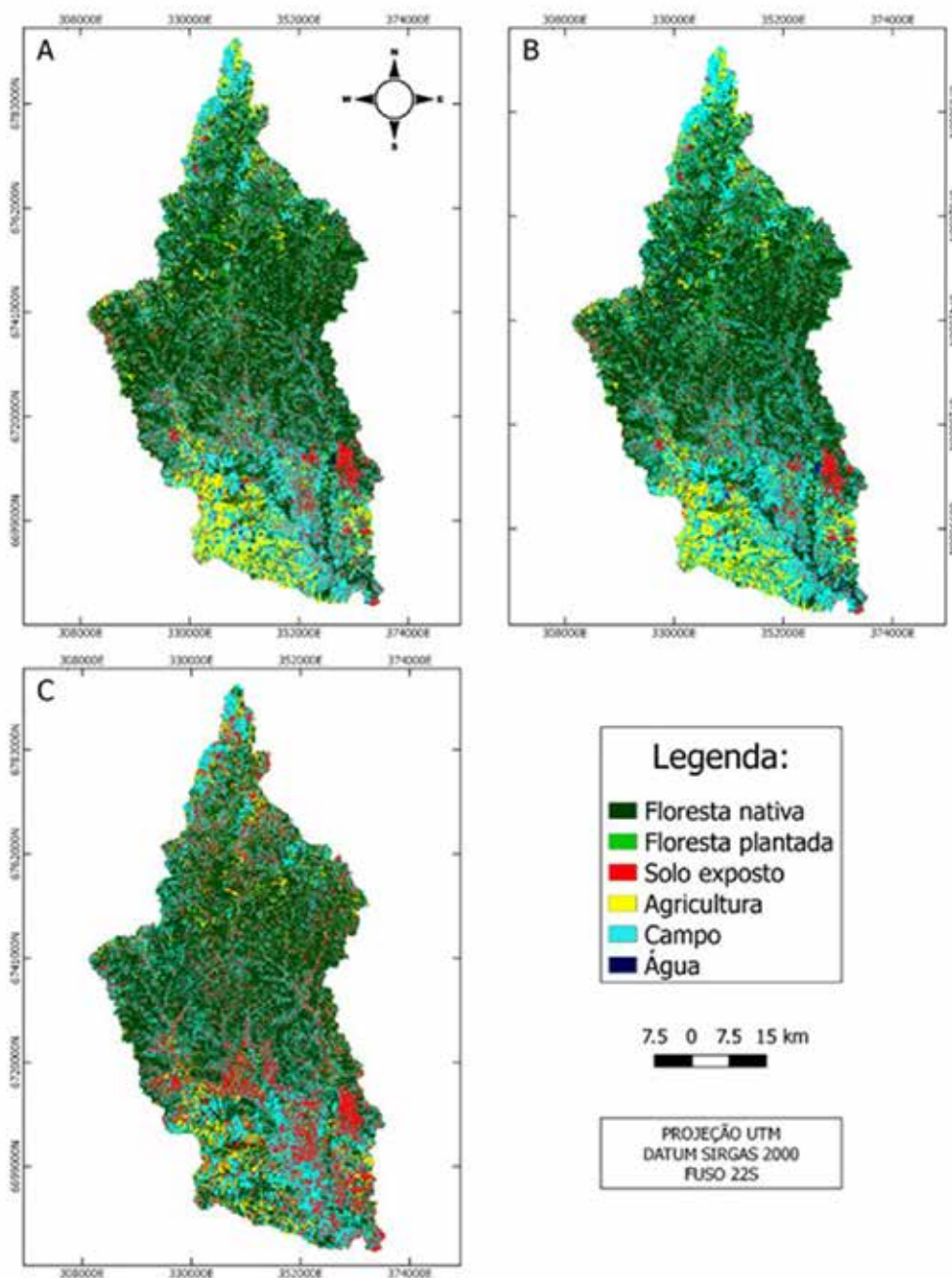


Figura 3 Resultado da classificação supervisionada por meio dos algoritmos RF (A); SVM (B); ANN (C) na bacia hidrográfica do Rio Pardo – RS.

e água, uma vez que foram os usos menos predominantes na área de estudo e por consequência, menor número de amostras foi extraído para o treinamento dos algoritmos. Além disso, não demonstrou eficácia na distinção das classes solo exposto e agricultura. Enquanto que os algoritmos RF e SVM, demonstraram maior ajuste na identificação de padrões espaciais para as classes temáticas estudadas.

Visualmente notou-se que as classificações obtidas pelo RF e SVM foram similares (Figura 3A e 3B), nas quais as classes apresentaram uma grande correlação com a realidade das imagens quando essas foram capturadas, porém houve confusão ao classificar áreas de floresta plantada e nativa devido ao sombreamento causado pelo relevo acidentado em áreas em que esta estava presente.

Assim, algumas áreas de floresta nativa foram classificadas como floresta plantada devido a estas regiões apresentaram semelhanças nos valores digitais dos pixels.

A Figura 3C demonstra a classificação da área de estudo por meio do classificador ANN. Tendo em vista que este foi o algoritmo que apresentou o maior erro na classificação, foi possível observar as confusões geradas pelo classificador, o qual não identificou as classes de floresta plantada e água (Tabela 3). Além disso, notou-se uma diferença na quantificação das classes, principalmente para o solo exposto, o qual totalizou 67.875,98 ha, valor este discrepante do obtido pelo RF (31.424,26 ha) e SVM (30.769,21 ha), indicando que outros usos foram incluídos nesta classe. Ainda, ao analisar a imagem sintética (composição RGB) e a classificação gerada percebeu-se que houve confusão entre os pixels pertencentes à classe campo e água, e ainda, entre as classes de solo exposto e agricultura. Uma possível solução para os problemas encontrados na utilização da ANN seria aumentar a amostragem dessas áreas em que houve maiores confusões.

Além disso, percebeu-se que as classes temáticas que apresentaram cobertura vegetal (floresta nativa, agricultura e campo), com exceção da floresta plantada, tiveram resultados próximos para os três algoritmos. Este fato pode estar relacionado as bandas do Red Edge presentes nas imagens MSI/Sentinel-2A, que permitem melhor discriminação da vegetação. Conforme Abdi (2019) as faixas Red Edge das imagens MSI/Sentinel-2A apresentam vantagens para a classificação do uso e cobertura da terra quando comparado a outros satélites que não capturam essa região do espectro eletromagnético, devido à alta reflectância da vegetação nessa região causada pela concentração de clorofila, que permite a distinção da vegetação entre as classes de usos e cobertura da terra.

Quanto a análise da distribuição espacial das classes temáticas na área de estudo, observou-se predominância da cobertura florestal no entorno dos cursos hídricos, bem como em áreas montanhosas. A agricultura situou-se, em sua maioria, na porção Sul da bacia hidrográfica, associada a condições topográficas mais propícias para essa prática econômica. Do mesmo modo, o solo exposto, que no presente estudo corresponde às áreas cultiváveis em período de pousio ou preparo do solo, englobou uma pequena porção de área, situando-se ao Sul da área de estudo. Outra classe predominante foi o campo, a qual apresentou comportamento de mosaico, distribuído ao longo da bacia hidrográfica.

5 Conclusão

Esse estudo analisou a performance de algoritmos de aprendizado de máquina para a classificação supervisionada de usos e cobertura da terra por meio de imagens MSI/

Sentinel-2A. Sendo possível observar que os mesmos se mostraram eficientes, com destaque para o RF e SVM, pois, apresentaram os melhores parâmetros de acurácia. Estatisticamente o RF foi superior ao SVM, indicando que o mesmo pode ser utilizado com eficiência para classificação do uso e cobertura da terra em áreas similares à de estudo utilizando imagens MSI/Sentinel-2A.

Diante do exposto, o presente estudo demonstrou novas formas de obter a classificação supervisionada de usos e cobertura da terra de forma a combinar técnicas de inteligência artificial com o *software* R. Por meio de algoritmos de aprendizado de máquina foi possível a obtenção de resultados com boa precisão e confiabilidade, podendo ser um processo rápido e eficiente para estudos de cunho ambiental. Assim, a presente pesquisa demonstrou o potencial das técnicas de inteligência artificial aliadas às imagens de média resolução espacial para o mapeamento do uso e cobertura da terra. A qual servirá de base para estudos futuros, bem como para tomada de decisão em ações de planejamento da área.

6 Referências

- Abdi, A.M. 2019. Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data. *GIScience & Remote Sensing*, 57(1): 1-20.
- Andrade, A.C.; Francisco, C.N. & Almeida, C.M. 2014. Desempenho de Classificadores Paramétricos e Não-Paramétricos Na Classificação Da Fisionomia Vegetal. *Revista Brasileira de Cartografia*, 66(2): 349-363.
- Belgiu, M. & Drăgu, L. 2016. Random Forest in Remote Sensing: A Review of Applications and Future Directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114: 24-31.
- Bivand, R.S.; Keitt, T. & Rowlingson, B. 2018. rgdal: Bindings for the Geospatial Data Abstraction Library. R package v. 1.3-4. 2018. Disponível em: <<https://CRAN.R-project.org/package=rgdal>>. Acesso em: 20 nov. 2019.
- Breiman, L. 2001. Random Forests. *Machine Learning*, 45(1): 5-32.
- Cortes, C. & Vapnik, V. 1995. Support-Vector Networks. *Machine Learning*, 20: 273-297.
- Faceli, K.; Lorena, A.C.; Gama, J. & Carvalho, A.C.P.L.F. 2011. *Inteligência Artificial - Uma Abordagem de Aprendizado de Máquina*. Rio de Janeiro, LTC Editora, 394p.
- Farda, N.M. 2017. Multi-temporal Land Use Mapping of Coastal Wetlands Area using Machine Learning in Google Earth Engine. In: THE 5TH GEOINFORMATION SCIENCE SYMPOSIUM, 5, Yogyakarta, 2017. Resumo completo, Yogyakarta, 2017.
- Farias, A.J. 2010. *Atividade Florestal no Contexto da Fumicultura: Oportunidade de Desenvolvimento Regional, Diversificação, Geração de Emprego e Renda*. Programa de Pós-graduação em Engenharia Florestal, Universidade Federal de Santa Maria, Tese de Doutorado, 168p.

- Freires, E.V.; Silva Neto, C.A.; Cunha, D.S.R.; Duarte, C.R.; Veríssimo, C.U.V. & Gomes, D.D.M. 2019. Comparação de Imagens OLI/Landsat-8 e MSI/Sentinel-2 no Mapeamento de Cobertura e Uso da Terra no Maciço de Uruburetama, Ceará. *Anuário do Instituto de Geociências*, 42(4): 427-442.
- Friedman, M.A. 1940. Comparison of alternative tests of significance for the problem of m rankings. *Annals of Mathematical Statistics*, 11: 86-92.
- Gaiad, N.P.; Martins, A.P.M.; Debastiani, A.B.; Dalla Corte, A.P. & Sanquetta, C.R. 2017. Uso e Cobertura da Terra Apoiados em Algoritmos Baseados em Aprendizado de Máquina: O Caso de Mariana – MG. *Enciclopedia Biosfera*, 14(25): 1211-1220.
- Ge, G.; Shi, Z.; Zhu, Y.; Yang, X. & Hao, Y. 2020. Land Use/Cover Classification in an Arid Desert-Oasis Mosaic Landscape of China Using Remote Sensed Imagery: Performance Assessment of Four Machine Learning Algorithms. *Global Ecology and Conservation*, 22: e00971.
- Hasenack, H. & Weber, E. 2010. Base cartográfica vetorial contínua do Rio Grande do Sul. Porto Alegre: UFRGS Centro de Ecologia. Escala 1:50.000. Disponível em: <http://multimidia.ufrgs.br/conteudo/labgeo-ecologia/Arquivos/Downloads/Dados/2010/Base_50k_RS/base_cartografica_vetorial_RS_50k.pdf>. Acesso em: 16 set. 2019.
- Hijmans, R.J. 2017. raster: Geographic Data Analysis and Modeling. R package v. 2.6-7. Disponível em: <<https://CRAN.R-project.org/package=raster>>. Acesso em: 20 nov. 2018.
- IBGE. 2018. Instituto Brasileiro de Geografia e Estatística. Cidades. Produção Agrícola: Lavoura Temporária. Disponível em: <<https://cidades.ibge.gov.br/brasil/rs/panorama>> Acesso em: 04 mar. 2020.
- Jamali, A. 2019. Evaluation and comparison of eight machine learning models in land use/land cover mapping using Landsat 8 OLI: a case study of the northern region of Iran. *SN Applied Science*, 1: 1-10.
- Kuhn, M. 2018. caret: Classification and Regression Training. R package version 6.0-81. Disponível em: <<https://CRAN.R-project.org/package=caret>>. Acesso em: 07 nov. 2019.
- Li, C.; Wang, J.; Wang, L.; Hu, L. & Gong, P. 2014. Comparison of Classification Algorithms and Training Sample Sizes in Urban Land Classification with Landsat Thematic Mapper Imagery. *Remote Sensing*, 6(2): 964-983.
- Merschmann, L.H.C. 2007. *Classificação Probabilística Baseada Em Análise de Padrões*. Programa de Pós-Graduação em Computação, Universidade Federal Fluminense, Tese de Doutorado, 103p.
- Miranda, E.E. 2005. Brasil em Relevô. Embrapa Monitoramento por Satélite. Disponível em: <<http://www.relevobr.cnpm.embrapa.br>>. Acesso em: 15 abr. 2019.
- Monjezi, M.; Bahrami, A. & Varjani, A. Y. 2010. Simultaneous Prediction of Fragmentation and Flyrock in Blasting Operation Using Artificial Neural Networks. *International Journal of Rock Mechanics and Mining Sciences*, 47(3): 476-80.
- Nemenyi, P.B. 1963. *Distribution-free multiple comparisons*. Universidade de Princeton, Tese de doutorado, 254p.
- Niemeyer, J.; Rottensteiner, F. & Soergel, U. 2014. Contextual classification of lidar data and building object detection in urban areas. *ISPRS Journal of Photogrammetry and Remote Sensing*, 87: 152-165.
- Phiri, D. & Morgenroth, J. 2017. Developments in Landsat Land Cover Classification Methods: A Review. *Remote Sensing*, 9(9): 1-25.
- R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing. 2018.
- Rodriguez-Galiano, V.F.; Ghimire, B.; Rogan, J.; Chica-Olmo, M. & Rigol-Sanchez, J.P. 2012. An Assessment of the Effectiveness of a Random Forest Classifier for Land-Cover Classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 67(1): 93-104.
- Rosa, M.R. 2018. Classificação do Padrão de Ocupação Urbana de São Paulo Utilizando Aprendizagem de Máquina e Sentinel 2. *Revista do Departamento de Geografia, Volume Especial: 15-21*.
- Seabra, V.S.; Xavier, R.A.; Damasceno, J. & Dornellas, P.C. 2015. Análise das mudanças de uso e cobertura da terra na bacia do rio Taperoá-PB entre os anos de 1990 e 2009. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 16, João Pessoa, 2015. Resumo completo, São José dos Campos, INPE, p. 108-115.
- SEMA. 2003. Secretaria do Meio Ambiente e Infraestrutura. Inventário florestal contínuo. Porto Alegre: UFSM / SEMA – RS, 2003. Disponível em: <<http://coralx.ufsm.br/ifcfs/frame.htm>>. Acesso em: 04 mar. 2020.
- SEMA. 2019. Secretaria do Meio Ambiente e Infraestrutura. Bacia Hidrográfica do Rio Pardo. Disponível em: <<https://www.sema.rs.gov.br/g050-bacia-hidrografica-do-alto-jacui>>. Acesso em: 12 set. 2019.
- Shao, Z. & Zhang, L. 2016. Estimating Forest Aboveground Biomass by Combining Optical and SAR Data: A Case Study in Genhe, Inner Mongolia, China. *Sensors*, 16(6): 1-16.
- Song, J.; Gao, S.; Zhu, Y. & Ma, C. 2019. A Survey of Remote Sensing Image Classification Based on CNNs. *Big Earth Data*, 3(3): 232-54.
- USGS. 2019. United States Geological Survey. Earth Explorer. Disponível em: <<https://earthexplorer.usgs.gov/>>. Acesso em: 16 set. 2019.
- Vaeza, R.F.; Filho, P.C.O.; Maia, A.G. & Disperati, A.A. 2010. Uso e ocupação do solo em Bacia Hidrográfica urbana a partir de imagens orbitais de alta resolução. *Floresta e Ambiente*, 17(1): 23-29.
- Young, C.F. 2006. Desmatamento e desemprego rural na Mata Atlântica. *Floresta e Ambiente*, 13(2): 75-88.
- Zhu, W.; Sun, Z.; Peng, J.; Huang, Y.; Li, J.; Zhang, J.; Yang, B. & Liao, X. 2019. Estimating maize above-ground biomass using 3D point clouds of multi-source unmanned aerial vehicle data at multi-spatial scales. *Remote Sensing*, 11(22): 1-22.
- Zhu, X. & Liu, D. 2015. Improving forest aboveground biomass estimation using seasonal Landsat NDVI time-series. *ISPRS Journal of Photogrammetry and Remote Sensing*, 102: 222-231.