

Revista Brasileira de Geografia Física



Homepage: https://periodicos.ufpe.br/revistas/rbgfe

Espectrorradiometria na distinção de espécies florestais utilizando Random Forest (RF), Support Vector Machine (SVM) e Artificial Neural Network (ANN)

Igor da Silva Narvaes¹, Mateus Sabadi Schuh², Pábulo Diogo Souza², Matheus Morais Ziembowicz², José Augusto Spiazzi Favarin³, Janderson de Oliveira Silva², Noé dos Santos Hofiço², Laura Camila de Godoy Goergen⁴, Rudiney Soares Pereira⁵.

¹ Pesquisador, Instituto Nacional de Pesquisas Espaciais (INPE)/Coordenação Espacial do Sul (COESU – RS), (55) 999607661. Autor correspondente: igor.narvaes@inpe.br (autor correspondente). ² Programa de Pós-graduação em Engenharia Florestal – Universidade Federal de Santa Maria – RS. mateuschuh@gmail.com; pabulodiogo@gmail.com; mmziembowicz@hotmail.com; janderson_o.silva@hotmail.com; noe.hofico@gmail.com. ³ Programa de Pós-graduação em Engenharia Florestal – Universidade Federal do Paraná – PR. jasflorestal@yahoo.com.br. ⁴ Prefeitura Municipal de Curitiba – PR, Departamento de Investigação e Monitoramento. lauragoergen@yahoo.com.br. ⁵ Professor Titular de Sensoriamento Remoto da Universidade Federal de Santa Maria, RS; Professor e Orientador no Programa de Pós-Graduação em Engenharia Florestal. rudiney.s.pereira@gmail.com.

Artigo recebido em 29/05/2023 e aceito em 27/05/2024

RESUMO

A distinção de espécies florestais na arborização urbana é fundamental para mitigação dos efeitos locais do aquecimento global. Neste sentido, foram utilizados os algoritmos de aprendizado de máquina RF, SVM e de aprendizado profundo ANN. Os elevados valores de acurácia encontrados (F-1 score = 0,989; 0,9434; 0,9346, Acurácia Global = 0,989; 0,9444; 0,9333 e de índice kappa = 0,988; 0,9383; 0,9259) para o algoritmo ANN, SVM e RF, respectivamente. Os erros de classificação para a predição de algumas espécies para os classificadores analisados se dão em geral pela semelhança nos valores de reflectância nas regiões do red edge (700 a 720 nm) relacionados ao conteúdo similar de clorofila e nos comprimentos de onda específicos na região do infravermelho de ondas curtas (1400 e 1420 nm) responsáveis pelas diferenças no conteúdo de água e concentração química na planta e de lignina, respectivamente. Dado a complexidade dos classificadores, em especial o algoritmo de aprendizado de máquina ANN, e também SVM e RF, recomenda-se a alteração de seus hiperparâmetros para se evitar o sobreajuste dos resultados, ou seja, mesmo que o algoritmo esteja adaptado a uma determinada região se torne ineficaz para prever novos resultados.

Palavras-chave: Random forest ; Support vector machine; Artificial neural network (ANN); região do espectro da borda do vermelho; infravermelho de ondas curtas.

Spectroradiometry in Distinguishing Forest Species using Random Forest (RF), Support Vector Machine (SVM) and Artificial Neural Network

ABSTRACT

The distinction of forest species in urban afforestation is essential to mitigate the local effects of global warming. In this sense, the machine learning algorithms used RF, SVM, and deep learning ANN. The high accuracy values found (F-1 score = 0.989; 0.9434; 0.9346, Global Accuracy = 0.989; 0.9444; 0.9333 and kappa index = 0.988; 0.9383; 0.9259) for the ANN, SVM and RF algorithm, respectively. The classification errors for the prediction of some species for the analyzed classifiers are generally due to the similarity in the reflectance values in the red edge regions (700 to 720 nm) related to the similar content of chlorophyll and in the specific wavelengths in the shortwave infrared region (1400 and 1420 nm) responsible for differences in water content and chemical concentration in the plant and lignin, respectively. Given the classifiers' complexity, especially the ANN and also the SVM and RF machine learning algorithms, its hyperparameters alterations are recommended to avoid overfitting the results, i.e., even if the algorithm was adapted to a specified region becomes ineffective to new predictions.

Keywords: Random forest; Support vector machine; Artificial neural network; red edge spectral region; shortwave infrared region.

Introdução

Os índices de vegetação (IV) são modelos matemáticos ou algoritmos baseados no sensoriamento remoto (SR) que visam avaliar e caracterizar a cobertura vegetal, possibilitando assim, observar a sua capacidade de refletir a luz solar de acordo com as características da espécie (He et al., 2022).

As folhas das plantas contêm compostos espectralmente ativos, como pigmentos vegetais, água e proteínas, que sofrem alterações durante qualquer estresse, gerando deteccão de propriedades bioquímicas e biofísicas através da espectroscopia da região do visível e infravermelho próximo - VNIR no monitoramento da tolerância ao estresse hídrico (Das et al., 2017; Das, et al., 2018), índice de área foliar baseada em técnicas quimiométricas por meio do espectrorradiômetro (Das et al., 2020), conteúdo de clorofila (Das et al., 2015) e estresse abiótico (Das et al., 2018).

A espectroscopia VNIR vem sendo frequentemente utilizada com a introdução de índices de vegetação, muitas vezes propostos por estas pesquisas para a avaliação foliar utilizando técnicas de ML (Das et al., 2020), tais como proposto por nossa investigação.

Dentre as diversas finalidades e aplicações do SR, estão entre as mais usuais a detecção de mudanças e o monitoramento da cobertura vegetal (Rêgo et al., 2012; Silva e Galvíncio, 2013; Oliveira et al., 2017). Assim, uma das principais análises está relacionado aos IV, elementos importantes para distinguir as informações espectrais da vegetação dos outros elementos da superfície terrestre utilizando classificadores ou modelos preditivos via algoritmos de ML, além de demonstrar a quantidade e a qualidade da vegetação observada em determinadas épocas do ano (Ribeiro et al., 2016; Lins et al., 2017; Wahla et al., 2023).

Nessa linha, (Dmitriev et al., 2022) relatam que seleção dos IV mais adequados e aplicáveis ao SR para a determinação da composição e *status* das espécies florestais é uma importante tarefa voltada para a avaliação de comunidades vegetais em larga escala.

Paralelo ao avanço do SR, vários algoritmos de *ML* vêm sendo propostos para a classificação de imagens durante as últimas duas décadas (Sheykhmousa et al., 2020), o que tem permitido resolver determinadas lacunas de modelagem, especialmente em cenários complexos. Neste sentido, Lek e Guégan (1999) salientam que o crescimento de análises assistidas por computador, facilmente acessíveis pela comunidade científica tem facilitado as aplicações em modelagens ecológicas.

Alguns algoritmos de ML e aprendizado profundo (DL) vêm ganhando espaço em aplicações florestais e demonstrando eficiência para o manejo de florestas inequiâneas e equiâneas, visando assim, ganhos de precisão e redução de custos de projetos (Souza, 2019).

Associado a isto, a escolha das melhores variáveis para a composição de um bom modelo de classificação utilizando algoritmos SVM, RF e (ANN) vêm demonstrando valores de acurácia superiores aos classificadores por pixel, pois mesclam variáveis oriundas de sensores ópticos e de radar de abertura sintética (Shirmard et al., 2022) aliada a uma estrutura mais sofisticada destes tipos de classificadores.

Por estas razões, estes classificadores baseados em inteligência artificial (AI) vêm sendo utilizados para o mapeamento de mudanças de uso e cobertura da terra (LULCC) utilizando variáveis SAR como entrada para os algoritmos de classificação de *ML* (Diniz et al., 2022) e de *DL* (Ndikumana et al., 2018) ou aplicações utilizando como variáveis dados de sensores ópticos (Shimizu et al., 2023) ou integrando ambos os sistemas sensores (Chen e Bruzzone, 2022; Prudente et al., 2022).

Todavia, dada a complexidade de ambientes analisados a acurácia da classificação e de distúrbios florestais ainda pode produzir resultados de menor acurácia, tais como os encontrados no estudo de (Shimizu et al., 2023) utilizando IVs e segmentação como variáveis do algoritmo RF para cenários pretéritos e futuros.

Em contraposição, com a utilização do algoritmo ANN, mesmo em ambientes complexos foi possível atingir valores de acurácia mais elevados, considerado por Asadi et al. (2022) como classificador eficiente na predição de cenários futuros.

Embora estes classificadores sejam considerados inovadores, mais pesquisas são necessárias já que em uma análise de publicações na área desde 2015 Zhang e Li (2022) salientam que para que haja uma melhora nos resultados se faz necessário mais estudos com fusão de múltiplos dados de diferentes fontes, obtenção de amostras de alta qualidade, entre outros, traduzindo como fundamentais na era de Big Data.

Apesar do vasto campo de evolução destas tecnologias venha sendo comprovado, sobretudo na última década, em específico a coevolução de recursos computacionais e algoritmos de ML

atingiram um pico final com o surgimento de algoritmos de DL na última década (Pichler e Hartig, 2023), desta forma novas abordagens são necessárias para processar e analisar eficientemente esses dados objetivando revelar padrões, tendências e associações relacionadas em classificadores tanto de *ML* como de *DL* (B. Chen et al., 2021; Wang et al., 2022).

A distinção de espécies florestais é fundamental como um recurso útil para o manejo florestal, assim técnicas utilizando AI são fundamentais para tal, capazes de identificar espécies florestais de interesse na floresta com acurácia elevada (Onishi e Ise, 2021). Com o avanço da tecnologia estes algoritmos também vêm sendo utilizados na geração de aplicativos (APPs) considerados eficazes na identificação de espécies (Mäder et al., 2021).

Neste cenário de ambientes complexos e de difícil identificação, classificação, monitoramento e modelagem de cenários futuros, especificamente a identificação de espécies em áreas urbanas permite a consolidação de um trabalho de fronteira entre profissionais e futuros pesquisadores no planejamento do uso da terra ou na aplicação de ferramentas, técnicas espaciais e científicas de informação, se traduzindo como fundamental para que planos de arborização urbana possam ser implementados de forma correta com o auxílio destas ferramentas (Chaturvedi e de Vries, 2021).

Assim, a identificação precisa das espécies de árvores urbanas existentes com a sua localização é a tarefa mais importante para melhorar a qualidade do ar, água e solo; aumento do sequestro de carbono; mitigar os efeitos das ilhas de calor causados pelo aumento de superfícies urbanas expostas e consequente impacto da reflectância solar das superfícies urbanas na temperatura do ar e no fluxo de radiação (Gapski et al., 2023); e proteger o equilíbrio do solo e da água (Cetin e Yastikli, 2022), principalmente em municípios com déficit de árvores, como a área urbana de Santa Maria, RS.

A esse respeito, pequenas variações no espectro eletromagnético analisados via SR em conjunto com algoritmos de AI podem melhorar a identificação de espécies florestais, produzindo resultados superiores em relação à sensores remotos orbitais, em alguns casos de difícil identificação.

No cenário em contexto, o estudo investiga a eficiência de diferentes algoritmos de aprendizado de máquina e de aprendizado profundo na classificação supervisionada de espécies florestais utilizando dados de espectrorradiometria tomados a nível foliar. De modo específico: a) Aplicar os algoritmos RF, SVM e ANN, na construção de modelos preditivos apoiados em um vasto repertório de IV; b) Avaliar o emprego do índice de Gini presente no algoritmo RF como alternativa para seleção de variáveis preditoras para os algoritmos de ML e DL analisados; c) Comparar o desempenho preditivo dos modelos a nível de treinamento e teste utilizando o total de IVs, assim como o subconjunto selecionado pelo RF.

Dado as variações espectrais encontradas nas pesquisas sobre o tema em análise é necessário verificar se tais variações em curvas espectrais de espécies florestais são capazes de serem corretamente classificadas a nível foliar com espectrorradiometria com o auxílio de algoritmos de ML e DL.

Material e métodos

Coleta do material vegetativo

A coleta de amostras foi realizada no Campus da Universidade Federal de Santa Maria (UFSM-RS) nas seguintes coordenadas geográficas centrais: 29°43'11" S e 53°43'27" O, datum WGS-84. Foram coletadas folhas de indivíduos representativos de dez espécies florestais: Cedrela fissilis Vell. (Cedro), Eugenia uniflora L. (Pitangueira), Handroanthus heptaphyllus (Vell.) Mattos (Ipê-roxo), Handroanthus serratifolius (Vahl) S. Grose (Ipêamarelo), Inga marginata Willd. (Ingá), Luehea divaricata Mart. (Açoita cavalo), Melia azedarach L. (Cinamomo), Morus nigra L. (Amora), Platanus acerifolia (Aiton) Willd. (plátano) e Psidium cattleianum Sabine (Araçá).

A amostragem foi realizada no início da manhã, período em que as folhas apresentavam aspecto de plena turgidez. Para cada indivíduo, foram selecionados ramos distribuídos nos quatro quadrantes e na porção periférica da copa. Prontamente, as amostras foram embaladas em sacos plásticos e levadas ao laboratório. Por fim, para tomada da reflectância, foram selecionadas 30 folhas representativas de cada espécie, as quais encontravam-se sadias e na fase adulta.

Medições radiométricas

A resposta espectral das folhas (curva de reflectância) foi registrada por um microcomputador conectado ao espectrorradiômetro FieldSpec®¹ (Zhao, 2006), acoplado a unidade RTS-3ZC (esfera integradora), revestida internamente pelo material Spectralon®

(Labsphere, 2018), que permite a tomada de fatores de reflectância bidirecional a nível de laboratório. As folhas foram posicionadas de forma que apenas a porção adaxial do limbo foliar ficasse exposta ao terminal de leitura da esfera, evitando medições na nervura central medindo a quantidade de energia refletida pela superfície foliar, ou seja, o alvo dividido pela quantidade de energia refletida pela placa de referência (Spectralon®). As informações provenientes do espectrorradiômetro constituem um binário contendo o registro radiométrico de cada comprimento de onda no intervalo de 350nm a 2500nm que serviram para determinação de 237 IV descritos na plataforma index data base (Henrich et 2012) al., https://www.indexdatabase.de/). Esses arquivos foram convertidos no software ASD WiewSpec Pro® (Zhao, 2006), em formato de texto, para então serem tabulados e processados em linguagem R (R Core Team, 2021).

Modelagem

Os IV derivados das medicões radiométricas foram associadas as dez classes rotuladas (espécies florestais), compondo a base de dados para a classificação (Figura 1). Dois algoritmos de ML e um de DL foram testados para construção dos modelos, respectivamente o RF (Breiman, 2001), o SVM (Cortes e Vapnik, 1995) e o ANN (McCulloch e Pitts, 1943). A preparação dos dados, assim como a implementação dos modelos foi realizada em linguagem R, no ambiente da interface gráfica RStudio (RStudio Team, 2021).

A base de dados foi preliminarmente subdividida em treinamento e validação nas respectivas proporções de 70% e 30%, considerando o número de repetições de cada espécie florestal (Neter et al., 1996).



Figura 1. Metodologia de análise para a geração dos classificadores.

Algoritmos

O RF é um método que consiste na construção de muitas árvores de decisão que podem ser usadas para classificar um novo exemplo de forma majoritária. Assim, cada árvore de decisão acaba usando um subconjunto de atributos selecionados de forma aleatória a partir do conjunto de dados originais para atingir um resultado único (Oshiro, 2013).

O SVM aborda conceitos de aprendizado supervisionado, onde ele possui um conjunto de treinamento e teste, sendo muito utilizado para

análise de dados e reconhecimento de padrões, tanto para o caso de classificação, quanto para resolver problemas de regressão (Chamasemani e Singh, 2011).

Neste contexto, é utilizado o conceito de um hiperplano de separação, podendo rotular as observações de uma determinada classe $y_i = 1$ e aqueles da classe $y_i = -1$, levando em consideração a sua magnitude de $f(x^*)$ na definição da proximidade das observações em relação ao limite do hiperplano de separação, o cálculo do hiperplano de margem máxima e posteriormente os vetores de suporte das n classes (p) geradas pode serem vistos em Gareth et al. (2013). Nesta análise foram utilizados os parâmetros fornecidos por default no pacote e1071 do R, em especial os valores de default do polinômio utilizado para o cálculo do kernel para a determinação dos hiperplanos de separação entre as classes os quais foi do tipo polinomial, bem como o valor de tradeoff de 1 (Meyer et al., 2021).

As ANN são técnicas de processamento simples, que apresentam em sua estrutura modelos matemáticos inspirados na capacidade de aprendizado do cérebro humano (Haykin, 2001), porém o comportamento da rede neural é afetado pela regra de aprendizado, arquitetura e função de transferência (Bala e Kumar, 2017).

De forma geral, as unidades escondidas (*hidden units*) análogas aos neurônios no cérebro quando com valores das ativações $A_k = h_k(X)$ são transformações não lineares de combinações lineares de entrada $X_1, X_2,...,X_p$. Portanto, esses A_k não são observados diretamente e as funções hk(.) não são fixas antecipadamente, mas são aprendidas durante o treinamento da rede, onde A_k próximos de 1 estão ativas, enquanto aqueles próximos de 0 estão silenciosos (usando uma função de ativação sigmóide). Por fim, a camada de saída é um modelo linear que usa essas ativações Ak como entradas, resultando em uma função f(x) (Gareth et al., 2013).

Para executar a classificação, nós selecionamos os melhores modelos de cada classificador o qual foi determinado pelo teste de vários modelos criados, variando os valores dos dois parâmetros mais influentes de cada classificador, por exemplo, número de árvores (Ntree) e o número de variáveis amostradas aleatoriamente como candidatas em cada divisão (Mtry) para o RF e a função de penalidade/valor de custo (C) e gamma (Y) para o SVM (Adugna et al., 2022).

Para o classificador ANN foram utilizados os parâmetros estabelecidos em Bala e Kumar

(2017), a saber: *hidden units* e os intervalos de confiança dos pesos de aprendizado utilizando o *parâmetro alfa* para o cálculo do intervalo de confiança na cadeia de processamento.

O método estabelecido para o cálculo da rede neural foi o RPROP (Propagação Resiliente), padrão do pacote neuralnet. Para Riedmiller e Braun (1993) este método é um novo esquema de aprendizado eficiente, que realiza uma adaptação direta da etapa de peso com base nas informações do gradiente local e que tem como diferença crucial as técnicas de adaptação desenvolvida previamente e que este esforço de adaptação não é obscurecido pelo comportamento do gradiente.

Em nossa análise foi utilizada a rede neural ANN (*multi-layer perception*). Esta forma um tipo de rede neural com uma ou mais camadas ocultas além de um determinado número de neurônios internos definidos pelo usuário. A camada oculta possui esse nome porque não é possível prever a saída desejada nas camadas intermediárias e para o treinamento desta rede, o algoritmo básico comumente utilizado é chamado de retropropagação (backpropagation) (Gardner e Dorling, 1998).

Seleção de preditores

Foi utilizada a variável de importância (VI) para o cálculo dos melhores preditores para todos os classificadores empregados pois esta é uma função adicional do RF, métrica explorada em diferentes cenários desde a consolidação do referido classificador, como exemplificado em Belgiu e Drăgut (2016) para a redução do número de dimensões do dado hiperespectral; identificação do dado mais relevante de SR para a classificação de determinada classe, entre outras.

Essa importância é medida pelo índice de Gini, sendo que para tal foi utilizado o gráfico da diminuição média no coeficiente de Gini (Mean Decrease in Gini - MDG), o qual expressa a perda de precisão do modelo por ocasião da exclusão de cada variável, sendo uma medida de como cada variável contribui para a homogeneidade dos nós e folhas na floresta aleatória resultante (Martinez-Taboada e Redondo, 2020), ou seja, como é medida do quanto uma variável reduz a métrica Gini de impureza em uma classe particular (Breiman, 2001).

Nesse sentido, quanto maior o valor do indicador supracitado, maior a importância da variável para o modelo (Yao et al., 2022) de forma que as variáveis a serem consideradas nos três tipos de classificação sejam capazes de

identificar/interpretar as características espectrais das espécies presentes nas classificações geradas.

Para tal foi utilizado um ambiente de programação do R, caracterizado por ser um software livre para computação estatística e gráficos. Os pacotes utilizados para implementação dos modelos RF, SVM e ANN foram respectivamente, *randomForest* (Liaw e Wiener, 2002), *e1071* (Meyer et al., 2021) e *neuralnet* (Günther e Fritsch, 2010).

Avaliação dos modelos

Na avaliação das etapas de treinamento e validação optou-se por utilizar a matriz de confusão e o percentual de acerto, exatidão global (G), índice kappa (Landis e Koch, 1977) e F1-score, esta última recentemente estabelecida e amplamente empregada como uma métrica de acurácia em classificações via ML (Barbosa et al., 2021; Cao et al., 2020). – Eqs. 1 a 4.

$$Acerto = \frac{\sum diag_{matriz}}{\sum matriz} * 100$$
 1

em que: \sum diagon_matriz = somatória dos elementos da diagonal principal da matriz de confusão,

 \sum matriz = somatória dos elementos da matriz de confusão.

$$\boldsymbol{G} = \frac{\sum_{i=1}^{c} x_{ii}}{n}$$

em que: soma da diagonal principal da matriz de erros x_i ; número total de amostras coletadas n.

$$K = \frac{n\sum_{i=1}^{c} x_{ii} - \sum_{i=1}^{c} x_{i+} x_{+i}}{n^2 - \sum_{i=1}^{c} x_{i+} x_{+i}}$$
 3

em que: $x_ii \in o$ valor na linha $i \in coluna i; x_i(i+)$ é a soma da linha $i \in x_i(+i) \in a$ soma da coluna ida matriz de confusão; $n \in o$ número total de amostras e c o número total de classes. Precision =

<u>Σ</u>Verdadeiros positivos (TP) <u>Σ</u>Verdadeiros positivos (TP)+ <u>Σ</u>Falsos positivos(FP)

Recall =

$$F1_i = \frac{2p_i r_i}{p_i + r_i} \tag{4}$$

em que: pi e ri são a precisão e o score recall da classe *i*, respectivamente.

Resultados e discussão

Distinção de espécies utilizando índices de vegetação (n=237)

Ao considerar-se os 237 IV, os algoritmos RF, SVM e ANN apresentaram respectivamente os valores de acerto para treinamento 96,67 %, 100 % e 100 %. Já para etapa de validação os algoritmos apresentaram como acerto 97,78 %, 96,67 % e 82,22 %, respectivamente (Figura 2).

Quando da utilização do RF, em ambientes de alta variação radiométrica em plantios agrícolas distintos na Índia (Nitze et al., 2012) encontraram com a aquisição de 4 imagens em datas similares valores de acurácia global inferior (87%) ao do presente estudo. Embora tendo avaliado um número considerado de plantios agrícolas os autores utilizaram uma variação de 100 a 600 Ntrees na escolha do parâmetro ótimo para a classificação para o RF, portanto o número de árvores de 500 em nosso estudo se mostrou adequado.

Mesmo em ambientes mais complexos e de maior área de análise (Adugna et al., 2022) encontraram valores da magnitude de 86% de acurácia global para o classificador RF quando comparados ao SVM e, que para atingirem estes resultados utilizaram um grande número de imagens e avaliaram a variação do número de árvores de decisão (Ntree) e de número de variáveis (Mtry) na escolha dos parâmetros ótimos para a classificação, considerados por Belgiu e Drăgut (2016) e Thanh Noi e Kappas (2017) como os parâmetros mais efetivos para alcançar a maior acurácia possível na classificação.

Estes autores encontraram como melhor formulação a utilização do *defalult* com Ntree de 100 e Mtry auto, esta última correspondente a raiz quadrada do número de variáveis, valores inferiores aos de *default* do pacote Randomforest no R utilizados em nossa pesquisa (Ntree=500 e Mtry=15).

De acordo com Ray (2019) a árvore de decisão pode encontrar o problema de sobreajuste (over-fitting) para o qual o classificador RF é a solução que se baseia em abordagem de modelagem de conjunto.

Partindo desse pressuposto, a presente pesquisa dada a grande complexidade dos alvos analisados e a consequente variação radiométrica, também foi utilizado o número de árvores de decisão supracitados já que os erros estabilizaram antes dos números de árvores de classificação serem atingidas (Lawrence et al., 2006), com valores similares aos usualmente encontrados na literatura (Nitze et al., 2012; Adugna et al., 2022) e

de número de variáveis sorteadas para compor cada árvore de (Mtry = 15). Outra razão da utilização destes valores, de acordo com Belgiu e Drăgut (2016) é devido ao fato de que 500 é o valor prédefinido no pacote randomForest do R.

Apesar da robustez dos classificadores de ML, alguns IV produziram curvas de distribuição de frequência dos valores dos IV gerados muito semelhantes, refletindo-se na confusão entre algumas classes, especialmente para as espécies de cinamomo e plátano, ambos com acurácia do usuário de 89% utilizando-se o classificador RF e mantendo-se com a mesma acurácia na classificação das mesmas espécies anteriores, acrescido da espécie amora no classificador SVM.

Já para a ANN as espécies plátano e ipê roxo apresentaram os menores valores de acurácia do usuário com apenas 56% dos dados classificados corretamente para ambas as espécies e Ingá com 67%. Neste caso, apenas 67% da espécie foi realmente presente com relação a amostra coletada via equipamento sensor, devido ao comportamento espectral do plátano com função densidade de probabilidade mais distribuída ao longo de sua distribuição de frequência com alta ocorrência de reflectância similar as demais espécies.

Todavia, para as espécies de ipê-roxo e ingá os baixos valores de acurácia do usuário estão relacionados com a similaridade da distribuição de frequência com a maioria das demais espécies, apresentando maior confusão entre estas classes analisadas (Figura 3).

comportamento da distribuição de frequência da reflectância ao longo de todo o espectro de análise da espécie plátano é responsável pelos principais erros de classificação encontrados.

comportamento da função de \mathbf{O} distribuição de frequência da reflectância distinto que compõem os IVs gerados encontrada em nossa análise também está relacionada as diferenças fisiológicas de cada espécie, bem como de sua arquitetura de copa, comprovados pelo estudo de (Goergen et al., 2016) onde os autores analisaram a possibilidade das imagens TM/Landsat 5 em distinguir estas espécies no mesmo povoamento, também comprovado pela utilização dos mesmos IV supracitados acrescidas de algumas bandas espectrais na região do visível, do infravermelho próximo (IVP) e médio (IVM) que também se mostraram eficazes para tal.

Para o classificador SVM, apesar da existência de vários tipos diferentes de kernels, tais como linear, polinomial, função de base radial, e sigmoide, e a função de kernel de base radial (RBF), esta última considerada por diversos autores (Foody e Mathur, 2004; Thanh Noi e Kappas, 2017; Waske et al., 2010; Yang, 2011) como a mais efetiva e comumente utilizada como parâmetro para classificação de imagens em SR. Porém em nossa base de dados, o modelo de kernel polinomial, utilizado como *default* do pacote e1071 do R obteve melhor desempenho.

O número de neurônios utilizado em cada uma das 5 camadas da rede do classificador ANN foi de 139, 91, 60, 40 e 26, seguindo a recomendação de que a primeira camada tenha 2/3 do número total de amostras (i.e 210 amostras), enquanto que para a segunda camada de neurônios deve possuir 2/3 da primeira e assim sucessivamente (Alice, 2015).

Este tipo de classificador de ANN que utiliza a rede neural multi-layer perception (MLP), juntamente com os demais tipos é bastante utilizado, correspondendo a 41% das pesquisas durante os últimos 12 meses no Brasil (www.google.com/trends) e em grande parte dos casos, quando comparados ao SVM e RF tem apresentado resultados superiores de acurácia em relação aos demais classificadores avaliados (Bueno et al., 2022; Raczko e Zagajewski, 2017).

Os principais parâmetros para a ANN são o número de camadas e neurônios nas camadas ocultas (Yao et al., 2022), embora em alguns casos busquem-se o melhor modelo para este classificador na tentativa de obter o melhor ajuste para o seu treinamento e para os melhores efeitos de previsão (Gardner e Dorling, 1998). Em nossa análise optou-se pela função polinomial selecionada por default no pacote neuralnet da linguagem de programação em R.

A complexidade dos alvos a serem avaliados e os produtos utilizados são determinantes na escolha dos hiperparâmetros deste classificador, sendo necessário avaliar as suas modificações para a melhora da eficácia dos modelos (Lek e Guégan, 1999; Ghorbanian et al., 2022; Yao et al., 2022). Desta forma, a sua qualidade da predição (Acurácia global = 0.8222; índice kappa = 0,8025; F-1 score=0,8215) são decorre ntes das escolhas acertadas destes hiperparâmetros. (descritos no item Algoritmos).

Em geral para os três classificadores o



Revista Brasileira de Geografia Física v.17, n.4 (2024) 2582-2605.

Figura 2. Matriz de confusão da etapa de treinamento (entre parênteses) e validação da classificação das espécies florestais com o RF (A), SVM (B) e ANN (C), com todos 237 IV.



Figura 3. Curvas de densidade de frequência, a nível de espécie, para os IV extraídos de (Goergen et al., 2016; 2021), acrescidos do WDVI e SR550_800, calculados utilizando toda base de dados (dados de treinamento e validação) (n = 237; 300 observações).

Aprendizado de máquina na distinção de espécies florestais para os IV selecionados

Random forest

Foram eleitos os 15 preditores ranqueados (índices de vegetação) com base nos valores do índice de Gini para a redução da impureza para a classificação de uma determinada classe (Breiman, 2001) com o intuito de reduzir as incertezas das classificações geradas.

Em especial o MVI e o SB1420 possuem valores superiores aos demais constituindo-se nas variáveis mais importantes para a geração da predição/classificação das espécies florestais escolhidas (Figura 4).

Estes 15 IVs selecionados referem-se aos pontos fora da distribuição "padrão" (outliers)

representada fora dos intervalos do boxplot (intervalo do interquartil até os valores mínimos e máximos), ou seja, são de fato as variáveis cuja relevância dentro do modelo destoa das demais.

O algoritmo RF apresentou os percentuais de acerto de 89,52% e 86,67%, respectivamente. Na avaliação da qualidade da predição foram utilizadas 10 repetições de cada espécie, totalizando 90 observações, correspondentes a 30% do conjunto total de dados para a validação.

A espécie M. nigra foi predita erroneamente como P. acerifolia em 33,3% dos casos, enquanto que M. azedarach também foi classificada erroneamente como P. acerifolia em 22,2% das observações, correspondendo a uma acurácia de 67 e 78%, respectivamente (Figura 5).



Figura 4. Representação do índice GINI em função do número de preditores. Em destaque (cinza) os IV eleitos de maior valor.

Os altos valores de qualidade de interpretação, tais como o nível de concordância classificado como quase perfeito - kappa = 0,93 (Landis e Koch, 1977) e acurácia global de 0,93 são promissores. Todavia de acordo com Bolfe et al. (2004) na classificação de espécies florestais mesmo quando a exatidão global apresente um valor alto este índice mostra-se menos consistente para a aferição de acurácia da classificação por não envolver no valor final todas as células da matriz de erros.

O valor de f1 score de 0,93, comprovou que o modelo de classificação obtido possui

precisão (acurácia do produtor) e recall (acurácia do usuário) altos, neste caso com baixos valores de falsos positivos e falsos negativos gerados.

Este valor se mostrou superior ao encontrado em (Adugna et al., 2022) para a grande classificação de ambientes de complexidade para classificação e de Barbosa et al. (2021) na classificação de áreas de cerrado no Distrito Federal utilizando diferentes combinações de intensidade e de coerência de dados interferométrica das imagens SAR do Sentinel-1 tendo atingido valores de f1-score = 0,79.





Frequência 0,0 2,5 5,0 7,5

AG = 0,9333 Kappa = 0,9259 F1 score = 0,9346

Figura 5. Matriz de confusão da etapa validação da classificação das espécies florestais com o classificador RF (RF) considerando os 15 IV mais importantes.

erros de classificação para o Os classificador RF refletidos nas métricas de acurácia para as classes avaliadas que compõe os índices espectrais do modelo final de classificação, estão ligadas ao comportamento espectral semelhante da reflectância regiões nas do espectro eletromagnético do verde, vermelho e infravermelho de ondas curtas (IVOC) para as espécies M.nigra e M. azedarach.

Ambas as espécies foram erroneamente classificadas como P. acerifolia (Figura 6) devido ao comportamento espectral semelhante destas com P. acerifolia, principalmente nas regiões do verde de comprimento de onda de 550 nm, red edge na faixa de 690 a 730 nm (Jensen, 1996), IVOC na faixa do espectro de 1490 a 1570 nm, regiões presentes da maioria dos IVs selecionados para este modelo de classificação (Tabela 1).

Restrições na distinção de espécies florestais classificadas de forma errônea também

foram evidenciadas por Jackson e Adam (2021) mesmo tendo identificado as regiões do espectroeletromagnético do vermelho, red edge, IVP1 (~770 a 900 nm) e IVP2 (900 a 1050 nm) como fundamentais na identificação de espécies, utilizando o sensor de alta resolução espacial WorldView-2. A ocorrência de sobreposição espectral significativa levaram a erros de classificação entre as espécies analisadas.

A relevância dos IVs na geração da classificação/predição, como nesta análise para o MVI é justificada por sua forte relação com o índice de área foliar (IAF) e o volume da copa de espécies florestais, pois de acordo com Schlerf et al. (2005) o infravermelho médio (IVM) em combinação com o IVP parecia conter mais informações relevantes para a caracterização do dossel florestal do que a combinação das bandas do vermelho e IVP.



Figura 6. Assinatura espectral das espécies classificadas como variáveis preditoras na validação do modelo final (n=90).

A utilização deste índice já havia sido proposta por Fassnacht et al. (1997) na composição de modelos para o cálculo do IAF em florestas de Pinus e floresta de folhosas, tendo demonstrado melhor desempenho em florestas de folhas largas, de maior complexidade estrutural.

A relevância dos IVs para a avaliação do uso e cobertura da terra pode ser verificado em (Talukdar et al., 2020) onde os autores utilizaram a "técnica baseada em índice", introduzida para avaliar e selecionar a melhor técnica de aprendizado de máquina para o mapeamento, neste caso para os índices NDVI (índice de vegetação diferencial normalizado), NDWI (índice de água diferencial normalizado) e NDBI (índice de construção diferencial normalizado) obtendo os melhores resultados para o RF.

Ainda para o mapeamento do LULCC de regiões declivosas Trindade et al. (2023), em um trecho do bioma pampa, verificaram que a utilização do modelo digital de elevação (MDE) do ALOS só se mostrou eficiente quando da inserção conjunta de bandas do Sentinel-2 e IVs como variáveis do algoritmo RF.

A importância da região da região do IVOC pode ser verificada em Persson et al. (2018)

utilizando o RF para a classificação nas quatro estações do ano com dados multitemporais do Sentinel-2 quando da utilização pelo seu ranqueamento desta banda no modelo final, em conjunto com o Red-edge e o IVP, mostrando-se capaz de capturar diferenças fenológicas ocorridas na primavera entre as espécies analisadas, regiões de comprimento de onda similares aos índices de maiores valores de Gini encontrados em nossa análise.

Sob esta perspectiva Hovi et al. (2017) analisando curvas espectrais de espécies florestais em laboratório concluíram que a descoberta mais importante em seu estudo foi que existem grandes diferenças entre as espécies na região do IVOC, também encontrados em nossa pesquisa.

Ao analisar a resposta espectral encontrada em nossa análise (Ban et al., 2022), comprovaram que os valores de maior correlação entre a vegetação, neste caso para o conteúdo de clorofila das folhas de arroz, que a maior relação encontrada com a resposta espectral ocorreu na região do vermelho, seguido do IVP, bem como na faixa de 450-720 nm, em grande parte regiões similares ao encontrado em nossos IVs selecionados.

Índices de vegetação	Formulação	Autor(es)
MVI – Mid- infrared vegetation index	(pred edge (700 nm): ρIVOC (1300 nm)) /(ρIVOC (1570 nm): ρIVOC (1780 nm))	Schlerf et al. (2005)
SB1420 – Single band 1420	ρIVOC(1420 nm)	Curran (1989)
SB1450 – Single band 1450	ρIVOC(1450 nm)	Curran (1989)
SB1120 – Single band 1120	ρIVOC(1120 nm)	Curran (1989)
AFRI1600 – Aerosol free vegetation index 1600	$(\rho IVP - 0,66 * (\rho IVOC (1600 nm))/(\rho IVP + 0,66 * \rho IVOC (1600 nm)))$	Karnieli et al. (2001)
SB1490 – Single band 1490	ρIVOC (1490 nm)	Curran (1989)
SR833/1649 – Simple ratio 833/1649 MSIhyper	(ρIVP (833 nm))/(ρIVOC (1649 nm))	Shibayama et al. (1999)
SB1200 – Single band 1200	ρIVOC (1200 nm)	Curran (1989)
SB1510 – Single band 1510	ρIVOC(1510 nm)	Herrmann et al. (2010)
SR715/705 – Single ratio 715/705	(pred edge(710 nm: 720 nm)) /(pred edge(700 nm: 710 nm))	Le Maire et al. (2004)
SB910	ρIVP (910 nm)	Curran (1989)
SB1400	ρIVOC (1400 nm)	Curran (1989)
MCARI1510	$[(\rho red edge(700 nm) - \rho IVOC(1510 nm)) - 0,2(\rho red edge(700 nm)) - \rho verde(550 nm))]((\rho red edge(700 nm))) /(\rho IVOC (1510 nm)))$	Le Maire et al. (2004)
DSWI-5	$(\rho IVP (800 nm)) - \rho verde (550 nm)) / (\rho IVOC (1660 mnm))$	Apan et al. (2003)
SB1540	$+ \rho verm. (680 nm))$ $\rho IVOC (1540 nm)$	Curran (1989)

Tabela 1. Formulação dos IV selecionados para a geração do	os modelos	finais de j	predição.
---	------------	-------------	-----------

onde: IVP – infravermelho próximo; IVOC – infravermelho de ondas curtas (Limites das várias faixas espectrais baseado em Lorenzzetti (2015).

Support vector machine

Seguindo a mesma tendência do classificador RF, a técnica de ML com algoritmo SVM obteve o mesmo comportamento, a qual a espécie M.nigra foi classificada erroneamente como P. acerifolia em 33% dos casos, todavia um comportamento distinto à primeira foi encontrada com espécie P.cattleianum, classificada a erroneamente como H. serratifolius em 11% dos casos observados, correspondendo a uma acurácia de 67% e 89%, respectivamente (Figura 7).

Os valores ligeiramente superiores de acurácia global (AG = 0.94), índice kappa = 0.94 e F-1 score = 0.94 para o SVM, se comparado ao classificador RF corroboram com o estudo de revisão sistemática de literatura proposto por (Sheykhmousa et al., 2020) onde as análises propostas mostraram que este classificador possui na média das análises acurácia superior, embora autores tenham concluído estes que sua interpretabilidade é inferior se comparado com o RF.

A aplicação do kernel de modelo polinomial mostrou-se eficiente, já que com estes parâmetros o classificador SVM encontrou o hiperplano ótimo em um espaço de classificação ndimensional com a maior margem de separabilidade entre as espécies florestais (classes), conforme sugerido por Raczko e Zagajewski (2017).

Chamasemani e Singh (2011) o qual testaram três tipos diferentes de kernels aplicados ao SVM também verificaram que o método polinomial obteve a maior acurácia global, superior ao presente estudo, tendo os considerado mais robusto, embora os tempos de processamento tenham sido superiores aos demais tipos.

Embora com valores de acurácia ligeiramente inferiores ao nosso estudo, Xi et al. (2022) identificaram a importância da borda do vermelho (red-edge), IVOC e IVs em regiões do espectro eletromagnético semelhantes à nossa análise na classificação de espécies florestais nativas (84,34% para SVM e 81,03% para RF) e plantadas (91,27% para SVM e 88,35% para o RF). Classe alvo



Frequência

AG = 0,9444 Kappa = 0,9383 F1 score = 0,9434

0.0 2,5 5,0 7,5

Figura 7. Matriz de confusão da etapa validação da classificação das espécies florestais com o classificador SVM considerando os 15 IV mais importantes.

Para as regiões do visível os índices MVI, MCARI1510 que possuem em sua formulação a reflectância na região do vermelho (700 nm) e DSWI-5 na região do verde (550 nm) e vermelho (680 nm) demonstram que os erros de predição para as espécies H. serratifolius e P. acerifolia estão

relacionados a maior reflectância foliar na região do verde em relação as demais espécies, o qual está ligado aos impactos causados por alguns estressores, acarretando de acordo com Fraser e Congalton (2021) numa redução na atividade fotossintética, marcado por uma mudança para maiores quantidades de refletância nestas regiões do espectroeletromagnético, responsáveis pela diminuição da acurácia global do classificador SVM.

A importância para a melhora na acurácia de classificação de espécies de árvores com algoritmos de *ML* utilizando IV, que tenham em sua formulação bandas nas regiões da borda do vermelho, IVP e de IVOC foram identificadas como variáveis ótimas por (Jiang et al., 2022), embora os IVs utilizados tenham sido diferentes em relação ao presente estudo.

Em nossa análise os 15 IVs mais significativos tem em sua formulação a banda do red edge (i.e. Single ratio 715/705 – planta saudável reflete mais energia luminosa na faixa de 670-760 nm (Hawryło et al., 2018); IVP (SB910, DSWI-5, SR833/1649 e AFRI1600 – regido pela estrutura e o arranjo celular diferenciado de cada espécie ocasionando uma resposta espectral singular para cada estrutura foliar (Jensen e Epiphanio, 2009; Ponzoni et al., 2012) e o IVOC (MVI, SB1420, SB1450, SB1120, entre outros compreende importantes elementos biofísicos e bioquímicos da vegetação, tais como lignina (Goergen et al., 2021) e celulose (Vicente et al., 2007).

As características de reflectância relacionada as diferenças físico-químicas das espécies florestais são fundamentais para a obtenção de bons resultados quando da utilização de algorítimos de *ML*, desta forma Adelabu et al. (2013) demonstraram que com a inserção da nova banda do red-edge presente no RapidEye, padrão similar a nossa pesquisa, aumentou o potencial de classificação do SVM de espécies florestais em ambientes semi-áridos quando integradas as outras bandas.

Sobre o aspecto de precisão (Caffaratti et al., 2021) utilizando o parâmetro de tipo kernel com modelo linear para o cálculo do hiperplano de separação de classes no SVM, encontraram para o seu modelo de detecção de floresta/não floresta valores superiores a 0,89 de F-1 score, utilizando como dados de entrada IVs e bandas individuais extraídas de imagens do Landsat-8 e MDEs do SRTM. Embora o experimento tenha mostrado resultados promissores em termos de acurácia e de redução da dependência da intervenção humana, estes foram inferiores aos valores de predição de espécies florestais de nossa análise (F1-score = 0,94).

Na mesma linha investigativa de nosso estudo, todavia utilizando dados LiDAR, Cetin e Yastikli (2022) obtiveram os maiores valores de acurácia global (73,75%) para o algoritmo de aprendizado de máquina SVM, os considerando o melhor para a identificação de espécies de árvores urbanas.

Artificial neural networks

A utilização da técnica de DL, também conhecida como redes neurais é um método inspirado no sistema biológico neural, onde modelos computacionais e matemáticos são usados no modelo ANN para imitar os processos do cérebro humano (Narmilan et al., 2022).

A sua eficácia já comprovada em diversos trabalhos científicos nesta temática (Ahmad et al., 2021; Bayat et al., 2021; Park et al., 2022) aliado à tendência crescente de publicações ao longo da última década (Shirmard et al., 2022), motivaram sua utilização nesta análise.

Este classificador produziu os melhores resultados de acurácia, refletindo-se na melhor predição das espécies florestais (Figura 8). A espécie P. acerifolia foi predita erroneamente (falsos positivos) como P. heptaphylius em 11% dos casos observados, obtendo uma acurácia para esta espécie de 89%, resultando nos maiores valores de acurácia em comparação com os classificadores de ML utilizados.

Sheykhmousa et al. (2020) em seu trabalho de investigação a respeito da comparação entre SVM e RF alcançaram acurácia satisfatória, embora em geral o algoritmo de classificação ANN apresente em média os melhores resultados de acurácia em relação as demais (Murat et al., 2017),

Porém este classificador de DL é considerado de menor interpretabilidade, elencando como principais problemas seus layers ocultos e sua natureza de "caixa preta", necessitando de conhecimento avançado (Pichler e Hartig, 2023) e de maior aporte computacional (Pineda-Jaramillo, 2019).

		C. fiss	ilis E. unif	H.hepta	phyllus H.serrat	ifolius 1.mare	inata L.diva	ricata M.azed	arach M.nig	ra P.aceri	folia P.cattl	eianum Acurá
	C. fissilis	9	0	0	0	0	0	0	0	0	0	100%
	E. uniflora	0	9	0	0	0	0	0	0	0	0	100%
Η	heptaphyllus	0	0	9	0	0	0	0	0	0	0	100%
1	H.serratifolius	0	0	0	9	0	0	0	0	0	0	100%
	1.marginata	0	0	0	0	9	0	0	0	0	0	100%
CIASSE	L.divaricata	0	0	0	0	0	9	0	0	0	0	100%
	M.azedarach	0	0	0	0	0	0	9	0	0	0	100%
	M.nigra	0	0	0	0	0	0	0	9	0	0	100%
	P.acerifolia	0	0	1	0	0	0	0	0	8	0	89%
	P.cattleianum	0	0	0	0	0	0	0	0	0	9	100%

Classe alvo

Frequência 0.0 2.5 5.0 7.5

k = 0,9889 Kappa = 0,9877 F1 score = 0,988

Figura 8. Matriz de confusão da etapa validação da classificação das espécies florestais com o classificador ANN considerando os 15 IV mais importantes.

As topologias da ANN e suas especificações consideram um número de camadas e neurônios, algoritmo de aprendizado, tipo de função de ativação e taxa de aprendizado, pois são parâmetros que além de definir a arquitetura do modelo influenciam os resultados da classificação e, portanto, deve ser cuidadosamente determinado para garantir a realização da melhor precisão possível (Ghorbanian et al., 2022).

Nesta mesma análise os melhores resultados foram alcançados com 4 camadas, 36 neurônios para cada camada, algoritmo de aprendizado do tipo de estimação de momento adaptativo (Adam), função de ativação de unidade linear retificada (Relu) e de taxa de aprendizado de 0,001 tendo produzido os mapas mais acurados para ecossistemas de manguezais (F-1 score = 0,97), valores ligeiramente inferiores a presente pesquisa (F-1 score = 0,99).

Em nossa análise, para alcançar estes resultados de predição, foram utilizadas 5 camadas ocultas com respectivamente 139, 91, 60, 40 e 26 neurônios por camada, algoritmo de aprendizado de retrogradação resiliente com retrocesso de peso (rprop+), função de ativação do tipo logística e taxa de aprendizado de -0,5 a 1,2. Considerando que o principal parâmetro para este tipo de classificador de *DL* é o número de camadas ocultas (Lek e Guégan, 1999) para as redes neurais a escolha de 5 mostrou-se em nosso caso adequada.

Já para o modelo desenvolvido por (Yao et al., 2022) na classificação de áreas de plantio agrícola também atingiu elevada acurácia (A.G = 0,93; kappa = 0,90), alcançados após diversos testes de variação dos parâmetros do modelo, obtendo melhor configuração com 3 camadas ocultas; 64, 32 e 16 neurônios por camada e 5000 épocas para o modelo de treinamento.

O número de camadas ocultas e de neurônios por camada inferiores ao da nossa análise e ao estudo de Vieira et al. (2020) também para a identificação de espécies florestais na classificação/identificação de cultivos agrícolas se dá pela sua menor variação espectral (Gava et al., 2022; Pott et al., 2020; Tiruneh et al., 2022).

Dado a sua complexidade florística e estrutural as espécies florestais são de difícil classificação, desta forma as técnicas de *ML* e *DL* vem mostrando resultados expressivos (Xie et al., 2019; Jiang et al., 2022).

Sob esta perspectiva (Raczko e Zagajewski, 2017) utilizaram os mesmos

classificadores de nossa pesquisa na distinção das espécies florestais *Picea alba* L. Karst, *Larix decidua* Mill, *Alnus* Mill, *Fagus sylvatica* L. e *Betula pendula* Roth, com dados em faixas espectrais similares, e também concluíram ao comparar os resultados das classificações a maior acurácia global (A.G = 77%) e índice kappa (0,72) para o classificador ANN, seguidos pelo SVM e RF, embora deva ser levado em consideração o número elevado de dados de entrada que tornaram o classificador ANN mais lento que os demais.

Conclusões

1- Os erros de classificação do algoritmo do RF para a espécie Eugenia uniflora. classificadas erroneamente Luehea como divaricata estão relacionados ao comportamento espectral semelhante de ambas, com valores de fator de reflectância muito próximos em todo espectro de análise, enquanto a confusão na classificação das espécies M. nigra e M.azedarach, ambas erroneamente classificadas como P. acerifolia são devido à valores semelhantes na região do espectro eletromagnético do infravermelho de ondas curtas na região de 1400 nm e 1420 nm na região de absorção nesta faixa, relacionado à flexão e alongamento da ligação H-O em água, bem como sua concentração química, e ao estiramento e deformação da ligação C-H na determinação do conteúdo de lignina, responsáveis pelos valores distintos de reflectância ocasionado pela dispersão das folhas;

2 - Os erros de classificação do algoritmo do SVM em específico para a espécie M. nigra, erroneamente classificada como P. acerifolia é devido a semelhança nos valores de reflectância na região do red-edge como resultado das transições eletrônicas do conteúdo de clorofila. A espécie P. cattleianum erroneamente classificada como H. serratifolius é decorrente a semelhança da reflectância entre ambas na região do visível até 1420 nm do infravermelho de ondas curtas, enquanto que C.fissilis como H. heptaphyllus os erros de classificação estão relacionados à semelhança das curvas espectrais ao longo de todo espectro eletromagnético;

3 - Foi constatado que os erros de classificação do algoritmo ANN ocorreram apenas para a espécie P. acerifolia classificada erroneamente como H. heptaphyllus também devido a semelhança entre os valores de reflectância entre ambas, porém até a faixa de 1400 nm, embora haja variação de reflectância na região de 1150 a 1300 nm, o classificador não foi capaz de discerni-las corretamente;

4 - Embora o algoritmo ANN tenha demonstrado resultados ligeiramente superiores para a discriminação de espécies florestais, os demais algoritmos de machine learning são de mais fácil interpretação e controle dos parâmetros de entrada, em especial o RF, tornando-os de mais fácil uso e interpretação;

5 - Recomenda-se a mudança dos valores dos hiperparâmetros iniciais dos classificadores em questão, já que a heterogeneidade das espécies analisadas aliada a semelhança em determinadas espectro eletromagnéticos pode faixas do da influenciar na acurácia classificação, aumentando a ocorrência de sobreajuste, embora os classificadores tenham apresentado um bom ajuste a este conjunto de dados.

Agradecimentos

Os autores agradecem ao Laboratório de Sensoriamento remoto (LabSere) do Programa de Pós-gradução em Engenharia Florestal da Universidade Federal de Santa Maria (UFSM).

Referências

```
Adelabu, S., Mutanga, O., Adam, E., & Cho, M.
A. (2013). Exploiting machine learning algorithms for tree species classification in a semiarid woodland using RapidEye image. Journal of Applied Remote Sensing, 7(1). https://doi.org/10.1117/1.jrs.7.073480
```

Adugna, T., Xu, W., & Fan, J. (2022).

Comparison of Random Forest and Support Vector Machine Classifiers for Regional Land Cover Mapping Using Coarse Resolution FY-3C Images. *Remote Sensing*, *14*(3). https://doi.org/10.3390/rs14030574

- Ahmad, S., Pandey, A. C., Kumar, A., & Lele, N.
 V. (2021). Potential of hyperspectral AVIRIS-NG data for vegetation characterization, species spectral separability, and mapping. *Applied Geomatics*, 13(3). https://doi.org/10.1007/s12518-021-00355-6
- Alice, M. (2015, September 23). Fitting a neural network in R; neuralnet package.
 Https://Www.r-Bloggers.Com/2015/09/Fitting-a-Neural-Network-in-r-Neuralnet-Package/.

- Apan, A., Held, A., Phinn, S., & Markley, J. (2003). Formulation and assessment of narrow-band vegetation indices from EO-1 Hyperion imagery for discriminating sugarcane disease. *Proceedings of the Spatial Sciences Institute Biennial Conference (SSC 2003): Spatial Knowledge Without Boundaries*. https://research.usq.edu.au/item/9zxwy/form ulation-and-assessment-of-narrow-bandvegetation-indices-from-eo-1-hyperionimagery-for-discriminating-sugarcanedisease
- Asadi, M., Oshnooei-Nooshabadi, A., Saleh, S.-S., Habibnezhad, F.-, Sarafraz-Asbagh, S., Lodewijk, J., & Genderen, V. (2022). Simulation of Urban Sprawl by Comparison Cellular Autom-ata-Markov and ANN. *Preprints*. https://doi.org/10.20944/preprints202208.01 19.v1

Bala, R., & Kumar, D. (2017). Classification using ANN: A review. Int. J. Comput. Intell. Res, 13(7), 1811–1820. https://www.ripublication.com/ijcir17/ijcirv1 3n7_22.pdf

Ban, S., Liu, W., Tian, M., Wang, Q., Yuan, T., Chang, Q., & Li, L. (2022). Rice Leaf Chlorophyll Content Estimation Using UAV-Based Spectral Images in Different Regions. *Agronomy*, *12*(11). https://doi.org/10.3390/agronomy12112832

Barbosa, F. L. R., Guimarães, R. F., Júnior, A. C.
O., & Gomes, A. T. R. (2021). Classificação do uso e cobertura da terra utilizando imagens SAR/Sentinel 1 no Distrito Federal. *Sociedade & Natureza*, *33*, e55954. https://doi.org/10.14393/sn-v33-2021-55954

Bayat, M., Bettinger, P., Heidari, S., Hamidi, S.
K., & Jaafari, A. (2021). A combination of biotic and abiotic factors and diversity determine productivity in natural deciduous forests. *Forests*, *12*(11). https://doi.org/10.3390/f12111450

- Belgiu, M., & Drăgut, L. (2016). Random forest in remote sensing: A review of applications and future directions. In *ISPRS Journal of Photogrammetry and Remote Sensing* (Vol. 114, pp. 24–31). Elsevier B.V. https://doi.org/10.1016/j.isprsjprs.2016.01.01 1
- Bolfe, É. L., Pereira, R. S., Madruga, R. de A., & Fonseca, E. L. da. (2004). Avaliação da classificação digital de povoamentos florestais em imagens de satélite através de índices de acurácia. *Revista Árvore*, 28(1). https://doi.org/10.1590/s0100-67622004000100011
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45, 5–32. https://doi.org/10.1023/A:1010933404324
- Bueno, G. F., Costa, E. A., Finger, C. A. G., Liesenberg, V., & Bispo, P. da C. (2022).
 Machine Learning: Crown Diameter Predictive Modeling for Open-Grown Trees in the Cerrado Biome, Brazil. *Forests*, *13*(8). https://doi.org/10.3390/f13081295
- Caffaratti, G. D., Marchetta, M. G., Euillades, L. D., Euillades, P. A., & Forradellas, R. Q. (2021). Improving forest detection with machine learning in remote sensing data. *Remote Sensing Applications: Society and Environment*, 24, 100654. https://doi.org/https://doi.org/10.1016/j.rsase .2021.100654
- Cao, R., Tu, W., Yang, C., Li, Q., Liu, J., Zhu, J., Zhang, Q., Li, Q., & Qiu, G. (2020). Deep learning-based remote and social sensing data fusion for urban region function recognition. *ISPRS Journal of Photogrammetry and Remote Sensing*, *163*, 82–97. https://doi.org/10.1016/j.isprsjprs.2020.02.01 4
- Cetin, Z., & Yastikli, N. (2022). The Use of Machine Learning Algorithms in Urban Tree Species Classification. *ISPRS International Journal of Geo-Information*, *11*(4). https://doi.org/10.3390/ijgi11040226

Chamasemani, F. F., & Singh, Y. P. (2011).
Multi-class Support Vector Machine (SVM) classifiers - An application in hypothyroid detection and classification. *Proceedings - 2011 6th International Conference on Bio-Inspired Computing: Theories and Applications, BIC-TA 2011*, 351–356. https://doi.org/10.1109/BIC-TA.2011.51
Chaturvedi, V., & de Vries, W. T. (2021). Machine Learning Algorithms for Urban

Land Use Planning: A Review. In *Urban Science* (Vol. 5, Issue 3). MDPI. https://doi.org/10.3390/urbansci5030068

- Chen, B., Xu, B., & Gong, P. (2021). Mapping essential urban land use categories (EULUC) using geospatial big data: Progress, challenges, and opportunities. *Big Earth Data*, 5(3), 410–441. https://doi.org/https://doi.org/10.1080/20964 471.2021.1939243
- Chen, Y., & Bruzzone, L. (2022). Self-Supervised SAR-Optical Data Fusion of Sentinel-1/-2 Images. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1–11. https://doi.org/10.1109/TGRS.2021.3128072
- Cortes, C., & Vapnik, V. (1995). Support-Vector Networks Editor. In *Machine Learning* (Vol. 20). Kluwer Academic Publishers.
- Curran, P. J. (1989). Remote sensing of foliar chemistry. In *Remote Sensing of Environment* (Vol. 30, Issue 3). https://doi.org/10.1016/0034-4257(89)90069-2
- Das, B., Mahajan, G. R., & Singh, R. (2018). Hyperspectral Remote Sensing: Use in Detecting Abiotic Stresses in Agriculture. In S. K. Bal, J. Mukherjee, B. U. Choudhury, & A. K. Dhawan (Eds.), *Advances in Crop Environment Interaction* (pp. 317–335). Springer Singapore. https://doi.org/10.1007/978-981-13-1861-0_12
- Das, B., Manohara, K. K., Mahajan, G. R., & Sahoo, R. N. (2020). Spectroscopy based novel spectral indices, PCA- and PLSRcoupled machine learning models for salinity stress phenotyping of rice. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, 229. https://doi.org/10.1016/j.sep.2010.117082

https://doi.org/10.1016/j.saa.2019.117983 Das, B., Sahoo, R. N., Pargal, S., Krishna, G.,

- Verma, R., Chinnusamy, V., Sehgal, V. K.,
 & Gupta, V. K. (2017). Comparison of different uni- and multi-variate techniques for monitoring leaf water status as an indicator of water-deficit stress in wheat through spectroscopy. *Biosystems Engineering*, *160*, 69–83. https://doi.org/10.1016/j.biosystemseng.2017 .05.007
- Das, B., Sahoo, R. N., Pargal, S., Krishna, G., Verma, R., Chinnusamy, V., Sehgal, V. K., Gupta, V. K., Dash, S. K., & Swain, P. (2018). Quantitative monitoring of sucrose, reducing sugar and total sugar dynamics for

phenotyping of water-deficit stress tolerance in rice through spectroscopy and chemometrics. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, *192*, 41–51. https://doi.org/10.1016/j.saa.2017.10.076

- Das, B., Sahoo, R. N., Pargal, S., Krishna, G., Verma, R., Tiwari, R., Viswanathan, C., Sehgal, V. K., & Gupta, V. K. (2015).
 Spectral based non-invasive estimation of plant chlorophyll content. *J Agric Phys*, *15*, 88–102.
 https://www.researchgate.net/publication/30 4792105
- Diniz, J. M. F. de S., Gama, F. F., & Adami, M. (2022). Evaluation of polarimetry and interferometry of sentinel-1A SAR data for land use and land cover of the Brazilian Amazon Region. *Geocarto International*, 37(5), 1482–1500.
- Dmitriev, P. A., Kozlovsky, B. L., Kupriushkin, D. P., Lysenko, V. S., Rajput, V. D., Ignatova, M. A., Tarik, E. P., Kapralova, O. A., Tokhtar, V. K., Singh, A. K., Minkina, T., Varduni, T. V., Sharma, M., Taloor, A. K., & Thapliyal, A. (2022). Identification of species of the genus Acer L. using vegetation indices calculated from the hyperspectral images of leaves. *Remote Sensing Applications: Society and Environment*, 25, 100679. https://doi.org/10.1016/J.RSASE.2021.1006 79
- Fassnacht, K. S., Gower, S. T., MacKenzie, M. D., Nordheim, E. V., & Lillesand, T. M. (1997). Estimating the leaf area index of North Central Wisconsin forests using the landsat thematic mapper. *Remote Sensing of Environment*, *61*(2). https://doi.org/10.1016/S0034-4257(97)00005-9
- Foody, G. M., & Mathur, A. (2004). A relative evaluation of multiclass image classification by support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, 42(6), 1335–1343.
- Fraser, B. T., & Congalton, R. G. (2021). Monitoring fine-scale forest health using unmanned aerial systems (Uas) multispectral models. *Remote Sensing*, *13*(23). https://doi.org/10.3390/rs13234873
- Gapski, N. H., Marinoski, D. L., Melo, A. P., & Guths, S. (2023). Impact of urban surfaces' solar reflectance on air temperature and radiation flux. *Sustainable Cities and*

Society, 96.

https://doi.org/10.1016/j.scs.2023.104645

Gardner, M. W., & Dorling, S. R. (1998). Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric Environment*, *32*(14–15), 2627–2636. https://doi.org/10.1016/S1352-2310(97)00447-0

Gareth, J., Daniela, W., Trevor, H., & Robert, T. (2013). *An introduction to statistical learning: with applications in R*. Spinger.

Gava, R., Santana, D. C., Cotrim, M. F., Rossi, F. S., Teodoro, L. P. R., da Silva Junior, C. A., & Teodoro, P. E. (2022). Soybean Cultivars Identification Using Remotely Sensed Image and Machine Learning Models. *Sustainability (Switzerland), 14*(12). https://doi.org/10.3390/su14127125

Ghorbanian, A., Ahmadi, S. A., Amani, M., Mohammadzadeh, A., & Jamali, S. (2022).
Application of Artificial Neural Networks for Mangrove Mapping Using Multi-Temporal and Multi-Source Remote Sensing Imagery. *Water (Switzerland)*, 14(2). https://doi.org/10.3390/w14020244

Goergen, L. C. de G., Narvaes, I. da S., & Adami, M. (2021). Estimation of wood volume of eucalyptus dunnii and urograndis of different ages using tm/landsat 5i. *Ciencia Florestal*, *31*(2), 683–704.

https://doi.org/10.5902/1980509834751

Goergen, L. C. de G., Kilca, R. V., Narvaes, I. S., Silva, M. N., Silva, E. A., Pereira, R. S., & Adami, M. (2016). Distinção de espécies de eucalipto de diferentes idades por meio de imagens TM/Landsat 5. *Pesquisa Agropecuaria Brasileira*, 51(1), 53–60. https://doi.org/10.1590/S0100-204X2016000100007

Günther, F., & Fritsch, S. (2010). Neuralnet: training of neural networks. *The R Journal*, 2(1), 30–38. https://svn.rproject.org/Rjournal/trunk/html/_site/archive /2010/RJ-2010-006/RJ-2010-006.pdf

Hawryło, P., Bednarz, B., Wężyk, P., & Szostak, M. (2018). Estimating defoliation of Scots pine stands using machine learning methods and vegetation indices of Sentinel-2. *European Journal of Remote Sensing*, 51(1). https://doi.org/10.1080/22797254.2017.1417 745

Haykin, S. (2001). *Redes neurais: princípios e prática* (L. House, Ed.; 2^a). Bookman Editora.

https://books.google.com.br/books?hl=pt-BR&lr=&id=bhMwDwAAQBAJ&oi=fnd& pg=PP1&dq=+Haykin,+S.+(2001).+Redes+ neurais:+princ%C3%ADpios+e+pr%C3%A 1tica.+Bookman+Editora.&ots=09rlPKLUH v&sig=tJwuxB8hqpFXNRPvo2oXX0QGbfI #v=onepage&q=Haykin%2C%20S.%20(200 1).%20Redes%20neurais%3A%20princ%C3 %ADpios%20e%20pr%C3%A1tica.%20Bo okman%20Editora.&f=false

He, X., Deng, Y., Dong, A., & Lin, L. (2022). The relationship between acoustic indices, vegetation, and topographic characteristics is spatially dependent in a tropical forest in southwestern China. *Ecological Indicators*, *142*. https://doi.org/10.1016/j.ecolind.2022.10922

https://doi.org/10.1016/j.ecolind.2022.10922 9

Henrich, V., Krauss, G., Götze, C., & Sandow, C. (2012). Entwicklung einer Datenbank für Fernerkundungsindizes. *Bochum: AK Fernerkundung*, 15. https://www.indexdatabase.de/info/credits.p hp

Herrmann, I., Karnieli, A., Bonfil, D. J., Cohen, Y., & Alchanatis, V. (2010). SWIR-based spectral indices for assessing nitrogen content in potato fields. *International Journal of Remote Sensing*, 31(19). https://doi.org/10.1080/01431160903283892

Hovi, A., Raitio, P., & Rautiainen, M. (2017). A spectral analysis of 25 boreal tree species. *Silva Fennica*, *51*(4). https://doi.org/10.14214/sf.7753

Jackson, C. M., & Adam, E. (2021). Machine learning classification of endangered tree species in a tropical submontane forest using worldview-2 multispectral satellite imagery and imbalanced dataset. *Remote Sensing*, *13*(24), 4970. https://doi.org/https://doi.org/10.3390/rs1324

4970

Jensen, J. R. (1996). *Introductory digital image* processing: a remote sensing perspective. (Second edition). Prentice-Hall Inc.

Jensen, J. R., & Epiphanio, J. C. N. (2009). Sensoriamento remoto do ambiente: uma perspectiva em recursos terrestres.

Jiang, X., Zhao, S., Chen, Y., & Lu, D. (2022). Exploring Tree Species Classification in Subtropical Regions with a Modified Hierarchy-Based Classifier Using High Spatial Resolution Multisensor Data. Journal of Remote Sensing (United States),

2022.

https://doi.org/10.34133/2022/9847835

- Karnieli, A., Kaufman, Y. J., Remer, L., & Wald, A. (2001). AFRI - Aerosol free vegetation index. *Remote Sensing of Environment*, 77(1). https://doi.org/10.1016/S0034-4257(01)00190-0
- Labsphere. (2018). *Labsphere, Inc. Privacy Policy*. https://www.labsphere.com/aboutlabsphere/
- Landis, J. R., & Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*, *33*(2), 363–374. https://doi.org/10.2307/2529786
- Lawrence, R. L., Wood, S. D., & Sheley, R. L. (2006). Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (randomForest). *Remote Sensing of Environment*, *100*(3), 356–362. https://doi.org/10.1016/j.rse.2005.10.014
- Le Maire, G., François, C., & Dufrêne, E. (2004). Towards universal broad leaf chlorophyll indices using PROSPECT simulated database and hyperspectral reflectance measurements. *Remote Sensing of Environment*, 89(1). https://doi.org/10.1016/j.rse.2003.09.004

Lek, S., & Guégan, J. F. (1999). Artificial neural networks as a tool in ecological modelling, an introduction. In *Ecological Modelling* (Vol. 120). https://doi.org/https://doi.org/10.1016/S0304

-3800(99)00092-7

- Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2*(3), 18–22.
- Lins, F. A. C., Araújo, D. C. dos S., Silva, J. L. B. da, Lopes, P. M. O., Oliveira, J. D. A., & Silva, A. T. C. S. G. da. (2017). Estimativa de parâmetros biofísicos e evapotranspiração real no semiárido pernambucano utilizando sensoriamento remoto. *Irriga*, *1*(1), 64–75. https://doi.org/https://doi.org/10.15809/irriga .2017v1n1p64-75
- Lorenzzetti, J. A. (2015). Princípios físicos de sensoriamento remoto. Editora Blucher.
- Mäder, P., Boho, D., Rzanny, M., Seeland, M., Wittich, H. C., Deggelmann, A., & Wäldchen, J. (2021). The Flora Incognita app – Interactive plant species identification. *Methods in Ecology and Evolution*, 12(7), 1335–1342. https://doi.org/10.1111/2041-210X.13611

- Martinez-Taboada, F., & Redondo, J. I. (2020). The SIESTA (SEAAV Integrated evaluation sedation tool for anaesthesia) project: Initial development of a multifactorial sedation assessment tool for dogs. *PLoS ONE*, *15*(4). https://doi.org/10.1371/journal.pone.023079 9
- McCulloch, W. S., & Pitts, W. H. (1943). A Logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, *5*, 115–133.
- Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., Chang, C.-C., & Li, C.-C. (2021). e1071: Misc Functions of the Department of Statistics, Probability Theory Group. In *CRAN Repository*.
- Murat, M., Chang, S.-W., Abu, A., Yap, H. J., & Yong, K.-T. (2017). Automated classification of tropical shrub species: a hybrid of leaf shape and machine learning approach. *PeerJ*, *5*, e3792. https://doi.org/10.7717/peerj.3792
- Narmilan, A., Gonzalez, F., Salgadoe, A. S. A., Kumarasiri, U. W. L. M., Weerasinghe, H. A. S., & Kulasekara, B. R. (2022).
 Predicting canopy chlorophyll content in sugarcane crops using machine learning algorithms and spectral vegetation indices derived from UAV multispectral imagery. *Remote Sensing*, 14(5), 1140.
 https://doi.org/https://doi.org/10.3390/rs1405 1140
- Ndikumana, E., Minh, D. H. T., Baghdadi, N., Courault, D., & Hossard, L. (2018). Deep recurrent neural network for agricultural classification using multitemporal SAR Sentinel-1 for Camargue, France. *Remote Sensing*, *10*(8). https://doi.org/10.3390/rs10081217
- Neter, J., Kutner, M. H., Nachtsheim, C. J., & Wasserman, W. (1996). *Applied linear statistical models*.
- Nitze, I., Wegener, A., Schulthess, U., Nitze, I., Schulthess, U., & Asche, H. (2012). Comparison of machine learning algorithms random forest, artificial neural network and support vector machine to maximum likelihood for supervised crop type classification. *Proceedings of the 4th GEOBIA*.

https://www.researchgate.net/publication/27 5641579

Oliveira, J. D. A., Medeiros, B. C., Silva, J. L. B., Moura, G. B. A., Lins, F. A. C., Nascimento, C. R., Lopes, P. M. O., Oceanography, in, &

Agricultural Engineerin, in. (2017). Spacetemporal evaluation of biophysical parameters in the High Ipanema watershed by remote sensing. *Journal of Hyperspectral Remote Sensing v*, 7(6), 357–366. www.periodicos.ufpe.br/revistas/jhrs

Onishi, M., & Ise, T. (2021). Explainable identification and mapping of trees using UAV RGB image and deep learning. *Scientific Reports*, *11*(1), 903. https://doi.org/https://doi.org/10.1038/s4159 8-020-79653-9

Oshiro, T. M. (2013). Uma abordagem para a construção de uma única árvore a partir de uma Random Forest para classificação de bases de expressão gênica. Universidade de São Paulo.

Park, G., Lee, Y. G., Yoon, Y. S., Ahn, J. Y., Lee, J. W., & Jang, Y. P. (2022). Machine Learning-Based Species Classification Methods Using DART-TOF-MS Data for Five Coniferous Wood Species. *Forests*, 13(10). https://doi.org/10.3390/f13101688

Persson, M., Lindberg, E., & Reese, H. (2018). Tree species classification with multitemporal Sentinel-2 data. *Remote Sensing*, *10*(11). https://doi.org/10.3390/rs10111794

Pichler, M., & Hartig, F. (2023). Machine learning and deep learning—A review for ecologists. *Methods in Ecology and Evolution*, 14(4), 994–1016. https://doi.org/10.1111/2041-210X.14061

Pineda-Jaramillo, J. D. (2019). A review of Machine Learning (ML) algorithms used for modeling travel mode choice. *DYNA*, 86(211), 32–41. https://doi.org/10.15446/dyna.v86n211.7974 3

Ponzoni, F. J., Shimabukuro, Y. E., & Kuplich, T. M. (2012). Sensoriamento remoto da vegetação (3ª Edição). Oficina de Textos.

Pott, L. P., Amado, T. J. C., Sebem, E., & Schwalbert, R. A. (2020). Spectral curves for identification of weeds in wheat crop. *Revista Engenharia Na Agricultura -REVENG*, 28. https://doi.org/10.13083/reveng.v28i.8154

Prudente, V. H. R., Skakun, S., Oldoni, L. V., A. M. Xaud, H., Xaud, M. R., Adami, M., & Sanches, I. D. (2022). Multisensor approach to land use and land cover mapping in Brazilian Amazon. *ISPRS Journal of Photogrammetry and Remote Sensing*, 189, 95–109.

https://doi.org/10.1016/j.isprsjprs.2022.04.02 5

R Core Team. (2021). R: A language and environment for statistical computing. Foundation for Statistical Computing.

- Raczko, E., & Zagajewski, B. (2017). Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral APEX images. *European Journal of Remote Sensing*, *50*(1), 144–154.
- Ray, S. (2019). A Quick Review of Machine Learning Algorithms. Proceedings of the International Conference on Machine Learning, Big Data, Cloud and Parallel Computing: Trends, Prespectives and Prospects, COMITCon 2019. https://doi.org/10.1109/COMITCon.2019.88 62451
- Rêgo, S. C. A., Lima, P. P. S., Lima, M. N. S., & Monteiro, T. R. R. (2012). Análise comparativa dos índices de vegetação NDVI e SAVI no município de São Domingos do Cariri-PB. *Revista Geonorte*, *3*(5), 1217– 1229.
- Ribeiro, E. P., Nóbrega, R. S., Mota Filho, F. O., & Moreira, E. (2016). Estimativa dos índices de vegetação na detecção de mudanças ambientais na bacia hidrográfica do rio Pajeú. *Geosul*, *31*(62), 59–92.
- Riedmiller, M., & Braun, H. (1993). A direct adaptive method for faster backpropagation learning: The RPROP algorithm. *IEEE International Conference on Neural Networks*, 586–591.
- RStudio Team. (2021). *RStudio: Integrated Development Environment for R* (1.4.1717). RStudio, PBC.
- Schlerf, M., Atzberger, C., & Hill, J. (2005). Remote sensing of forest biophysical variables using HyMap imaging spectrometer data. *Remote Sensing of Environment*, 95(2). https://doi.org/10.1016/j.rse.2004.12.016

Sheykhmousa, M., Mahdianpari, M., Ghanbari,
H., Mohammadimanesh, F., Ghamisi, P., &
Homayouni, S. (2020). Support Vector
Machine Versus Random Forest for Remote
Sensing Image Classification: A MetaAnalysis and Systematic Review. In *IEEE*Journal of Selected Topics in Applied Earth
Observations and Remote Sensing (Vol. 13,
pp. 6308–6325). Institute of Electrical and
Electronics Engineers Inc.

https://doi.org/10.1109/JSTARS.2020.30267 24

- Shibayama, M., Salli, A., Hä me, T., Iso-Iivari, L., Heino, S., Alanen, M., Morinaga, S., Inoue, Y., & Akiyama, T. (1999). Detecting Phenophases of Subarctic Shrub Canopies by Using Automated Reflectance Measurements of plant phenological events called phenophases-for ex. *Remote Sensing* of Environment, 67, 160–180. https://doi.org/10.1016/S0034-4257(98)00082-0
- Shimizu, K., Murakami, W., Furuichi, T., & Estoque, R. C. (2023). Mapping land use/land cover changes and forest disturbances in vietnam using a Landsat temporal segmentation algorithm. *Remote Sensing*, 15(3), 851. https://doi.org/https://doi.org/10.3390/rs1503 0851
- Shirmard, H., Farahbakhsh, E., Müller, R. D., & Chandra, R. (2022). A review of machine learning in processing remote sensing data for mineral exploration. In *Remote Sensing of Environment* (Vol. 268). Elsevier Inc. https://doi.org/10.1016/j.rse.2021.112750
- Silva, L. G., & Galvíncio, J. D. (2013). Análise Comparativa da Variação nos Índices NDVI e SAVI no Sítio PELD–22, em Petrolina– PE, na Primeira Década do Século XXI (Comparative Analysis of Changes in NDVI and SAVI in PELD-22-Petrolina-PE, in the First Decade of the XXI Century). *Revista Brasileira de Geografia Física*, 5(6), 1446– 1456.
- Souza, G. S. A. (2019). Aprendizado de máquina em aplicações de manejo florestal. https://locus.ufv.br//handle/123456789/2784 8
- Talukdar, S., Singha, P., Mahato, S., Shahfahad, Pal, S., Liou, Y.-A., & Rahman, A. (2020). Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations—A Review. *Remote Sensing*, *12*(7), 1135. https://doi.org/10.3390/rs12071135
- Thanh Noi, P., & Kappas, M. (2017). Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. Sensors (Basel, Switzerland), 18(1). https://doi.org/10.3390/s18010018
- Tiruneh, G. A., Meshesha, D. T., Adgo, E., Tsunekawa, A., Haregeweyn, N., Fenta, A. A., & Reichert, J. M. (2022). A leaf

reflectance-based crop yield modeling in Northwest Ethiopia. *PLoS ONE*, *17*(6 June). https://doi.org/10.1371/journal.pone.026979

- Trindade, P. M. P., Peixoto, D. W. B., Silveira, G. V., Kuplich, T. M., & Narvaes, I. da S. (2023). Mapeamento de cobertura e uso da terra no bioma Pampa utilizando diferentes sensores orbitais e classificador Random Forest. *Revista Geografias*, *19*(2), 60–82. https://doi.org/https://doi.org/10.35699/2237 -549X.2023.46915
- Vicente, L. E., Souza Filho, C. R., & Perez Filho, A. (2007). O uso do infravermelho de ondas curtas (SWIR) no mapeamento de fitofisionomias em ambiente tropical por meio de classificação hiperespectral de dados do sensor ASTER. Anais XIII Simpósio Brasileiro de Sensoriamento Remoto, 1915–1922.
- Vieira, H. C., Dos Santos, J. X., Souza, D. V., Rios, P. D., de Muñiz, G. I. B., Morrone, S. R., & Nisgoski, S. (2020). Near-infrared spectroscopy for the distinction of wood and charcoal from fabaceae species: Comparison of ann, knn and svm models. *Forest Systems*, 29(3). https://doi.org/10.5424/fs/2020293-16965
- Wahla, S. S., Kazmi, J. H., & Tariq, A. (2023).
 Mapping and monitoring of spatio-temporal land use and land cover changes and relationship with normalized satellite indices and driving factors. *Geology, Ecology, and Landscapes*, 1–17.
 https://doi.org/https://doi.org/10.1080/24749 508.2023.2187567
- Wang, J., Bretz, M., Dewan, M. A. A., & Delavar, M. A. (2022). Machine learning in modelling land-use and land cover-change (LULCC): Current status, challenges and prospects. *Science of the Total Environment*, 822.

https://doi.org/10.1016/j.scitotenv.2022.1535 59

- Waske, B., Van Der Linden, S., Benediktsson, J. A., Rabe, A., & Hostert, P. (2010). Sensitivity of support vector machines to random feature selection in classification of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 48(7). https://doi.org/10.1109/TGRS.2010.2041784
- Xi, Y., Tian, J., Jian, H., Tian, Q., Xiang, H., & Xu, N. (2022). Mapping tree species in natural and planted forests using Sentinel-2 images. *Remote Sensing Letters*, 13(6).

https://doi.org/10.1080/2150704X.2022.205 1636

- Xie, Z., Chen, Y., Lu, D., Li, G., & Chen, E. (2019). Classification of land cover, forest, and tree species classes with Ziyuan-3 multispectral and stereo data. *Remote Sensing*, *11*(2). https://doi.org/10.3390/rs11020164
- Yang, X. (2011). Parameterizing support vector machines for land cover classification. *Photogrammetric Engineering and Remote Sensing*, 77(1). https://doi.org/10.14358/pers.77.1.27
- Yao, X., Guo, Q., Li, A., & Shi, L. (2022). Optical remote sensing cloud detection based on random forest only using the visible light

and near-infrared image bands. *European Journal of Remote Sensing*, 55(1). https://doi.org/10.1080/22797254.2021.2025 433

Zhang, C., & Li, X. (2022). Land use and land cover mapping in the era of big data. *Land*, *11*(10), 1692. https://doi.org/https://doi.org/10.3390/land11

101692 Zhao, L. (2006). Analytical Spectral Devices, Inc. (ASD). https://www.gep.uchile.cl/Biblioteca/radiom etr%C3%ADa%20de%20campo/TechGuide. pdf