

Aplicação de algoritmos de Random Forest no suporte à identificação das espécies *Handroanthus Serratifolius* (Vahl) S. O. Grose e *Handroanthus Impetiginosus* (Mart. Ex DC.) Mattos (bignoniaceae)

Application of Random Forest algorithm to support in identification of the *Handroanthus Serratifolius* (Vahl) S. O. Grose and *Handroanthus Impetiginosus* (Mart. Ex DC.) Mattos (bignoniaceae) species

DOI:10.34117/bjdv8n5-457

Recebimento dos originais: 21/03/2022

Aceitação para publicação: 29/04/2022

Ygor Luiz Moraes Gonçalves

Bacharelado em Sistemas de Informação

Instituição: Universidade Federal do Pará (UFPA) – Campus Cametá

Endereço: R. Padre Antônio Franco - Matinha CEP: 68400-000

E-mail: ygorluizmoraes@gmail.com

Elton Sarmanho Siqueira

Mestre em Ciência da Computação

Instituição: Universidade Federal do Pará (UFPA) – Campus Cametá

Endereço: R. Padre Antônio Franco - Matinha CEP: 68400-000

E-mail: eltonss@ufpa.br

Cassiana Alves Ferreira

Doutora em Botânica Aplicada

Instituição: Universidad Continental, Huancayo, Perú

Endereço: Av. San Carlos, 1980, Huancayo (12000), Junin, Perú

E-mail: cassianaaf@gmail.com

Maycon da Silva Teixeira

Bacharelado em Engenharia Florestal

Instituição: Universidade do Estado do Pará– Campus Belém

Endereço: Tv. Dr. Eneas Pinheiro, 2626 - Marco, Belém - PA, CEP: 66095-015

E-mail: mayconsilvatx@gmail.com

Pablo da Ventura Correa

Bacharelado em Engenharia Florestal

Instituição: Universidade do Estado do Pará– Campus Belém

Endereço: Tv. Dr. Eneas Pinheiro, 2626 - Marco, Belém - PA, CEP: 66095-015

E-mail: pablov.correa@gmail.com

Cláudia Viana Urbinati

Doutora em Ciência e Tecnologia da Madeira

Instituição: Universidade do Estado do Pará – Campus Belém

Endereço: Tv. Dr. Eneas Pinheiro, 2626 - Marco, Belém - PA, CPE: 66095-015

E-mail: claudiaurbinati@uepa.br

RESUMO

O presente trabalho ressalta a necessidade de estudos de conservação, desenvolvimento sustentável e tecnológico para espécies florestais comercializadas na Amazônia Brasileira e principalmente no estado do Pará, onde a exploração desordenada de espécies madeiras alcança níveis preocupantes. O processo de identificação das madeiras de *Handroanthus serratifolius* e *Handroanthus impetiginosus* conhecidas pelo nome vernacular de ipê (e comercializadas em todo estado, conta com aspectos dificultosos, sendo promissor a utilização de dados anatômicos qualitativos das espécies associados ao desenvolvimento de um modelo inteligente e eficiente utilizando o algoritmo Random Forest. Para tal, duas espécies foram selecionadas e identificadas macroscopicamente e suas características anatômicas foram descritas e utilizadas para o treinamento do modelo. As referidas espécies apresentam semelhanças em sua estrutura anatômica o que torna a identificação pela análise visual complexa, diferindo pela presença de conteúdo obstruindo os vasos, estratificação do parenquima radial, tipo de parênquima e porosidade. Após a etapa de pré-processamento do conjunto de dados (seleção de atributos e balanceamento de classes) com o auxílio do algoritmo de grid search (para otimização de hiperparâmetros), durante a fase de treinamento do modelo, obteve-se uma precisão de 75% de acertos utilizando duas classes dentro de um mesmo conjunto de dados, sendo possível a diferenciação entre as duas espécies.

Palavras-chave: inteligência artificial, aprendizado de máquina, anatomia da madeira, ipê.

ABSTRACT

The present work highlights the need for conservation studies, sustainable development and technology for forest species commercialized in the Brazilian Amazon and mainly in the state of Pará, where the disorderly exploitation of timber species reaches worrying levels. The process of identifying Ipê wood (*Handroanthus serratifolius* and *Handroanthus impetiginosus*) marketed in every state has difficult aspects and the use of qualitative data regarding the anatomy of the species associated with the development of an intelligent and efficient model using the Random Forest algorithm becomes promising. To this end, the species were identified Macroscopically and their anatomical characteristics based on the methodology described by Coradin and Muniz (1991) used to train the model. These species have similarities in their anatomical structure, which makes the identification so complex, differing by the presence of contents obstructing the vessels, regular or irregular stratification, presence of marginal parenchyma, circular or diffuse ring porosity. After the pre-processing step of the data set (attribute selection and balancing class) with the aid of the grid search (for hyperparameter optimization) during the training phase of the model, we obtained an accuracy of 75% using two classes within the same data set, making it possible to differentiate between the two species.

Keywords: artificial intelligence, machine learning, wood anatomy, ipê.

1 INTRODUÇÃO

Estima-se que 2,5 bilhões de metros cúbicos de madeira foram comercializados em todo mundo em 2020, mesmo com importantes fatores socioeconômicos contrários, como a pandemia do Covid-19. As exportações de madeira tropical serrada passaram de

24.200 m³ em junho de 2020 para 40.000 m³ em junho de 2021, com rendimento de US\$14,1 milhões no último período (*International Tropical Timber Organization - ITTO*, 2021). No Brasil a exportação de madeira serrada teve uma queda de 3,3% em volume para o mês de agosto de 2021. Apesar disso, o estado do Acre, localizado na Amazônia Brasileira, foi um dos maiores exportadores de madeira tropical da região, encerrando agosto de 2021 com superavit comercial de US\$2,5 bilhões com a venda de madeira e derivados (ITTO, setembro de 2021).

O estado do Pará, localizado na mesma região, ocupa a terceira posição no ranking dos estados brasileiros que mais exportaram produtos a base de madeira, obtendo uma receita de 211 milhões de dólares (Associação das Indústrias Exportadoras de Madeiras do Estado do Pará - AIMEX, 2020). Somente nesse estado, no primeiro semestre de 2021, foram comercializados um total de 80.272 m³ entre toras e madeira serrada, segundo o Governo do estado do Pará (Boletim de preços mínimos do estado do Pará, Brasil, 2021). Dados com tamanha expressão refletem a necessidade de garantir a correta identificação e a procedência do material comercializado, a fim de conservar as florestas tropicais no Brasil, bem como assegurar que os mecanismos de consumo sustentáveis para comercialização diminuam a pressão sobre espécies que podem se encontrar ameaçadas ou em risco de extinção.

Além da cadeia de custódia, a certificação da madeira após a extração depende também da identificação correta da espécie. Como muitas espécies tropicais são extremamente semelhantes umas às outras, no momento da identificação, podem ocorrer equívocos decorrentes da adoção de nomes vernaculares em detrimento à nomenclatura e identificação botânica, ocasionando entraves na comercialização, agravada pela ineficiente auditoria dos inventários florestais por parte dos órgãos fiscalizadores (*World Wide Fund for Nature - WWF*, 2013).

Nesse cenário, o advento de sistemas computacionais trouxe novas possibilidades de identificação mais rápidas e seguras, a exemplo das técnicas de visão computacional baseadas em padrões de reconhecimento de imagens da estrutura anatômica macro e microscópica, amplamente difundidos na literatura (Hermanson, Wiedenhoeft, 2011; Ibrahim et al., 2016; Da Silva et al., 2017; Andrade et al., 2020). Apesar de já existirem bancos de dados robustos, as metodologias supracitadas não dispensam o auxílio da anatomia forense de identificação e classificação de madeiras baseada na estrutura xilemática de cada táxon, seja ele ao nível de família, gênero e até espécies botânicas. Outras características, bem como a coloração e o cheiro, conhecidas como características

organolépticas da madeira, também são utilizadas de maneira complementar às características anatômicas para segregação e identificação de espécies tropicais. Entretanto, a identificação de espécies tendo como ferramenta a anatomia de madeira além de requerer acesso às coleções biológicas, requer profissionais altamente treinados, demandando tempo e custos adicionais, o que dificulta a atividade.

No estado do Pará, o ipê amarelo (*Handroanthus serratifolius*) é a primeira espécie no *ranking* de exportações de madeiras, porém a partir da análise de dados oficiais, Lentini et al. (2021) apontam para a exploração predatória de ipê (*Handroanthus* spp.) entre os anos de 2007 a 2019, ressaltando que a comercialização de madeiras pertencentes ao gênero duplicou no mesmo período, passando de cerca de 230.000 a 500.000 m³, com aproximadamente 80% da produção concentrada em pólos madeireiros como Prainha, Juruti e Santarém (PA), classificadas pelos autores como “Fronteira Ativa da exploração madeireira no Centro da Amazônia.”

Popularmente conhecido como ipê, pau d'arco, peúva, ipeúna e entre outros. O gênero *Handroanthus* possui aproximadamente 33 espécies, pertence à família botânica Bignoniaceae Juss. da qual compreende cerca de 289 gêneros com ampla distribuição em todo o mundo, em especial na América Latina, em países como Bolívia, Nicarágua, Costa Rica e Brasil (Trópicos, 2021). Estudos visando a segregação através da estrutura xilemática presente em 43 espécies pertencentes aos gêneros *Tabebuia*, *Handroanthus* e *Roseodendron* que antes eram considerados somente *Tabebuia* foram realizados por Santos (2017) que encontrou diferenças anatômicas que corroboram com a segregação sistemática realizada. Entretanto, dentro de um mesmo gênero, em especial o *Handroanthus*, as semelhanças anatômicas xilemáticas são expressivas, causando extrema dificuldade na segregação e identificação entre os táxons comercializados.

Nesse cenário, com base em um grupo de espécies comercializadas sob a denominação única de ipê, o presente trabalho aplica técnicas de *machine learning* com o intuito de segregar espécies com elevada similaridade e pertencentes ao mesmo gênero, utilizando um conjunto de dados qualitativos. Portanto, reconhecendo a importância econômica das espécies pertencentes ao gênero *Handroanthus*, em especial *Handroanthus serratifolius* (Vahl) S.O. Grose e *Handroanthus impetiginosus* (Mart. ex DC.) Mattos, o presente estudo visa a utilização do algoritmo de *Random Forest* no suporte a identificação anatômica das espécies estudadas.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 MACHINE LEARNING

O *Machine Learning* (ML) é uma subárea da Inteligência Artificial que vem apresentando um grande crescimento nas últimas décadas. Trata-se de algoritmos matemáticos, estatísticos e computacionais que são capazes de realizar um processo de inferência por meio de aprendizado baseado em conjunto de dados (ALLENDE-CID, 2019).

Existem diversos métodos de aprendizagem e algoritmos de ML que podem ser utilizados em determinadas situações. No entanto, para cada situação um ou mais algoritmos terão melhor eficiência e trarão melhores resultados.

Os métodos de aprendizagem mais conhecidos e utilizados são o supervisionado e o não supervisionado. De acordo com Çelik (2018), o aprendizado supervisionado é um método em que os dados de entrada são utilizados para alcançar o conjunto de resultados. Existem dois tipos de aprendizagem supervisionada: classificação e regressão. Já no aprendizado não supervisionado, o processo de aprendizagem ocorre utilizando as relações e conexões entre os dados. Existem também dois tipos de aprendizagem não supervisionada: agrupamento e associação.

Para a presente pesquisa, o método utilizado foi algoritmo de classificação denominado *Random Forest*, que segundo Vieira (2015) consiste em um conjunto de árvores de decisão geradas dentro de um mesmo objeto. Cada objeto (conjunto de árvores) passa por um mecanismo de votação (*bagging*), o qual elege a classificação mais votada. A classificação encontra-se nos terminais das mesmas.

Do ponto de vista computacional, o uso de *Random Forest* é atraente devido: ser relativamente rápido de treinar e prever; depende apenas de um ou dois parâmetros de ajuste; apresenta uma estimativa embutida de erro de generalização; pode ser usado diretamente para problemas de alta dimensão; pode ser facilmente implementado em paralelo. Estatisticamente, o *Random Forest* provê recursos adicionais, como: medidas que apresentam a importância de cada variável do conjunto de dados; ponderação diferencial de classe; e visualização e detecção de *outliers* (CUTLER et al., 2011).

2.2 USO DE MACHINE LEARNING PARA IDENTIFICAÇÃO DE ESPÉCIES COM BASE NA ANATOMIA DA MADEIRA

A fim de auxiliar na identificação e segregação de espécies florestais, o uso de *machine learning* como ferramenta de classificação é cada vez mais utilizado e novas

tecnologias e modelos são reportados na literatura, como o reconhecimento de padrões de imagens macro e microscópicas. Esteban et al., (2017) utilizaram redes neurais artificiais para segregar duas espécies de Gimnospermas pertencentes ao gênero *Pinus* L., *Pinus sylvestris* L. e *Pinus nigra* var. *salzmannii* (Dunal) Laguna Lumbreras que são extremamente difíceis de serem identificadas através da análise anatômica forense, sendo identificadas segundo a biometria dos seus elementos xilêmáticos. O modelo obteve uma precisão aproximada de 80%.

No estudo feito por Ravindran et al. (2018) envolvendo a identificação de importantes espécies da família Meliaceae, como *Swietenia macrophylla* King e *Cedrela odorata* L. listadas nos apêndices CITES, foi realizado em duas frentes: uma levando em consideração todas as espécies e outra por agrupamento dos gêneros das respectivas espécies. Utilizando o classificador de imagens por Redes Neurais Convolucionais, o presente trabalho obteve resultados para segregação de gêneros de 97,5% e para espécies de 87,4%.

Espécies do gênero *Swietenia*, principalmente *S. macrophylla* King, *S. mahagoni* (L.) Jacq e *S. humilis* Zucc, também apresentam semelhante dificuldade de distinção quanto os ipês, até mesmo por análise visual da estrutura anatômica macroscópica. Recorre-se, portanto, à aplicação de modelos de *machine learning*, tendo como parâmetro a biometria dos elementos anatômicos, visando diferenciar as três espécies. He et al., (2019) trabalhando com espécies do gênero *Swietenia* obtiveram resultados conflitantes. O mesmo modelo, *Decision Tree* C5.0, apresentou um desempenho de 78,4%, tendo dificuldades para diferenciar a *S. humilis* das demais espécies. No entanto, este mesmo modelo elevou o seu desempenho para 81,4% ao diferenciar apenas as espécies *S. macrophylla* e *S. mahagoni*.

Diferentes ferramentas de identificação como aplicativos para *smartphones* como por exemplo “Madeiras comerciais do Brasil” desenvolvido a partir da Chave de Identificação de Madeiras Comerciais do Brasil (CORADIN et al., 2010) em formato Delta versão 1.0 que traz um guia para identificação com caracteres anatômicos básicos e suas definições e ainda imagens com excelentes resoluções. Também apresenta uma versão para computador composta por uma chave interativa, contando com 157 espécies comercializadas no Brasil, inclusive as espécies do gênero *Handroanthus*.

Outra opção é o aplicativo de “CITESWoodID” que apresenta excelentes imagens e ainda uma chave de identificação com espécies ameaçadas de extinção. Esta está disponível em três idiomas e traz 46 espécies florestais.

3 METODOLOGIA

De modo a construir e avaliar o modelo de classificação baseado em árvore, foi projetado um protocolo experimental desde a coleta de dados até a avaliação do modelo, como mostrado na Figura 1.

Figura 1 - Etapas do desenvolvimento da pesquisa.



3.1 COLETA DE AMOSTRAS DE MADEIRAS

O material analisado foi coletado em uma serraria localizada no Distrito Industrial de Icoaraci, Belém, Pará, Brasil. Obteve-se amostras de 166 indivíduos de ipês provenientes da Microrregião de Santarém, Pará, Brasil. As amostras foram levadas e acondicionadas no Laboratório Multiusuário de Engenharia Florestal (LAMEF), do Centro de Ciências Naturais e Tecnologia da Universidade do Estado do Pará, Belém.

3.2 IDENTIFICAÇÃO DO MATERIAL

As amostras foram encaminhadas à Embrapa (Empresa Brasileira de Pesquisa Agropecuária) Amazônia Oriental para comparação com material de referência oriundo da Xiloteca IAN (Instituto Agrônomo do Norte). O material caracterizado também foi comparado com *e*-xilotecas – *InsideWood* e XiloVAM, bem como literaturas específicas.

3.3 CARACTERIZAÇÃO ANATÔMICA

Para a confecção dos corpos de prova, foi utilizada uma serra fita de bancada da marca *Linus*, modelo SFM 250, do LAMEF. Em que se obteve blocos de madeira da região periférica das toras nas dimensões de 1,5 x 1,5 x 1,5 cm, orientados nos planos anatômicos da madeira: transversal e longitudinal tangencial e radial, respectivamente. Os mesmos foram lixados à mão com auxílio de lixas com gramaturas diferentes,

obedecendo uma ordem crescente (80, 120, 220, 320, 400 e 1200). A descrição anatômica macroscópica foi realizada com base nas orientações de Coradin e Muñiz (1991).

3.4 DESCRIÇÃO DO DATASET

A estrutura do conjunto de dados utilizado é referente às amostras de madeiras de ipê que foram coletadas e posteriormente identificadas. Possuindo um total de 166 registros sendo em sua totalidade dados categóricos, a estrutura do *dataset* pode ser vista no Quadro 1.

Quadro 1 - Atributos do conjunto de dados composto por 17 colunas.

Coluna	Domínio
Vasicentrico	Ausente, Presente, Predominante
Vasicentrico_escasso	Ausente, Presente, Predominante
Aliforme_Extensao_losangular	Ausente, Presente, Predominante
Aliforme_extensao_losangular_aletas_curtas	Ausente, Presente, Predominante
Aliforme_extensão_losangular_confluente	Ausente, Presente, Predominante
Escasso_ausente	Ausente
Marginal_simulando_faixas_marginais	Ausente, Presente
Aneis_semiporosos	Ausente, Presente
Porosidade_difusa	Ausente, Presente
Arranjo_diagonal	Ausente, Presente
Arranjo_padrao_indefinido	Ausente, Presente, Predominante
Conteúdo_TR	Ausente, Presente
Conteúdo coloração TR	Bege, Esverdeada, Amarelada, Ausente, Caramelo, Escura, Negra
Listrado_estratificacao	Regular, Irregular
Linhas_vasculares	Levemente sinuosas, Irregulares, Retilneas, Inclinação, Levemente inclinadas
Espelhado_raios	Contrastado, Pouco contrastado
Especie	<i>Handroanthus barbatus</i> , <i>H. impetiginosus</i> , <i>H. ochraceus</i> , <i>H. sp.</i> , <i>H. serratifolius</i> , <i>H. cf. barbatus</i> , <i>H. cf. impetiginosus</i> , <i>H. cf. serratifolius</i> , e <i>H. cf. ochraceus</i> .

Fonte: Autores

Os dados da coluna espécie possuem algumas peculiaridades em que *Handroanthus* cf. e *Handroanthus* sp. possuem um significado específico para esse estudo. No caso de espécies com “cf. *conferatum*” ou seja, possivelmente é a referida espécie entretanto necessita de confirmação. Já para espécies seguidas de “sp.” significa que até o presente, somente o gênero está confirmado. A distribuição dessas espécies segue da seguinte forma: *Handroanthus* sp.: 107, *Handroanthus* cf. *serratifolius*: 17, *Handroanthus serratifolius*: 14, *Handroanthus* cf. *impetiginosus*: 12, *Handroanthus impetiginosus*: 7, *Handroanthus* cf. *barbatus*: 4, *Handroanthus* cf. *ochraceus*: 2, *Handroanthus ochraceus*: 2, *Handroanthus barbatus*: 1.

3.5 TRATAMENTO DE DADOS

Após a coleta dos dados, e a fim de obter melhores resultados, foi necessário realizar uma análise no *dataset* para identificar e corrigir valores que pudessem não estar nos padrões esperados, dados *outliers* que pudessem influenciar negativamente no treinamento dos dados, tratar valores nulos e dados faltantes (*missing values*). Para que dessa forma, o conjunto de dados estivesse pronto para ser manipulado através dos algoritmos de *Machine Learning*.

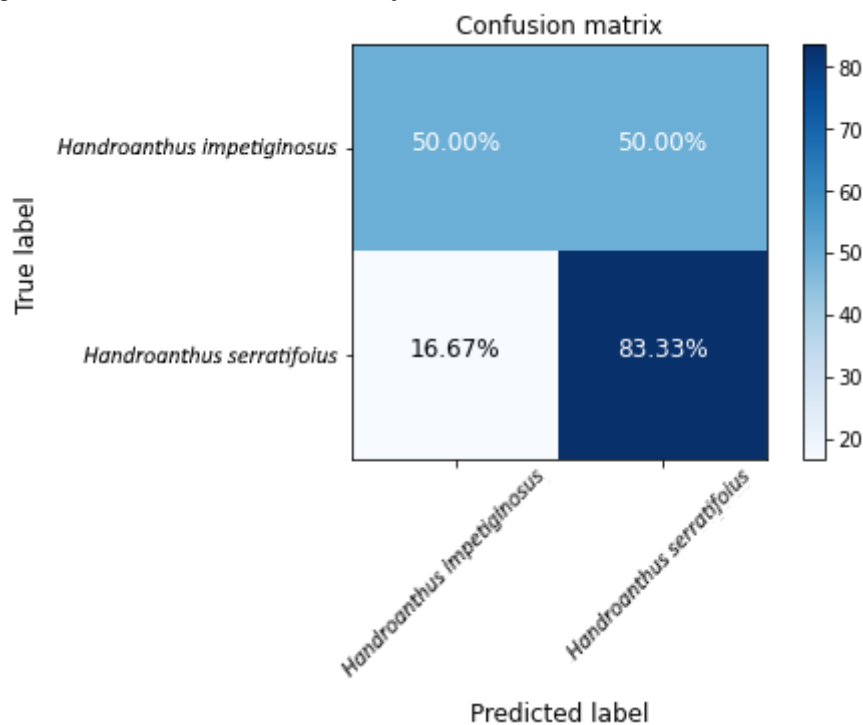
3.6 TREINAMENTO DO MODELO

Nesta etapa, após a coleta, análise e tratamento dos dados, obteve-se as métricas dos resultados de acurácia do modelo de classificação, predição dos dados e da matriz de confusão.

Primeiramente, para realizar o treinamento do modelo, filtrou-se os dados das espécies *Handroanthus serratifolius* e *Handroanthus impetiginosus*, em seguida o modelo de classificação escolhido foi *Random Forest* pois a base de dados utilizada foi composta integralmente por dados categóricos, e por esse método ser um dos mais indicados para problemas que envolvam classificação dessa natureza, ele foi escolhido para esta pesquisa. Com relação aos parâmetros do modelo, foram utilizados os seguintes atributos e valores: {*n_estimators*: 1200, *min_samples_split*: 5, *min_samples_leaf*: 2, *max_features*: 'auto', *max_depth*: 50, *bootstrap* : True}, os quais determinam a melhor configuração do modelo obtidos através do algoritmo de *grid search*, que é uma técnica de ajuste que estima os melhores valores para os hiperparâmetros, realizando uma pesquisa exaustiva sobre os valores específicos de um modelo (LIASHCHYNSKYI et al., 2019).

Para treinar o modelo de classificação, foi necessário dividir o conjunto de dados em duas partes distintas, sendo 75% dos dados para treinamento e 25% para testes. Usando o método *Cross validation* que segundo BERRAR (2018, p.1) “é um método de reamostragem de dados para avaliar a capacidade de generalização de modelos preditivos e para evitar *overfitting*”, com 5 interações adotando o indicador “accuracy”, o modelo obteve a acurácia média de 75%, nesse caso, o modelo de classificação apresentou uma boa precisão para os objetivos da pesquisa. Para esclarecer os detalhes dos resultados do treinamento, utiliza-se uma matriz de confusão (Figura 2), que exibe as frequências de classificação para cada classe do modelo.

Figura 2. Matriz de confusão da validação do modelo de Random Forest construído...



Fonte: Autores

Dessa forma, a matriz de confusão identificou corretamente 50% dos *Handroanthus impetiginosus* e 83.33% dos *Handroanthus serratifolius*, constatando que o modelo consegue predizer de forma satisfatória essas duas espécies.

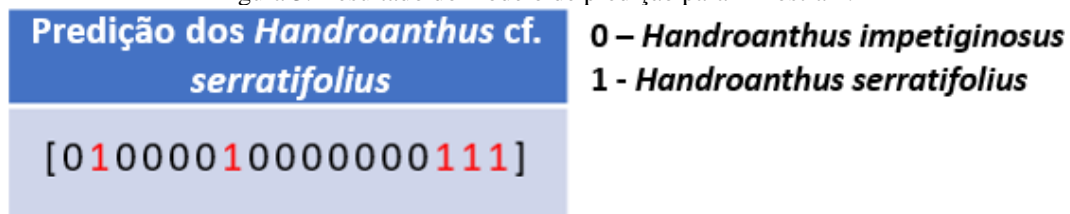
4 RESULTADOS

Após o treinamento do modelo, uma forma de validá-lo foi selecionando algumas amostras não identificadas de ambas as espécies e testá-las. Há duas amostras de validação: Amostra 1: possui 17 registros de *Handroanthus cf. serratifolius*; e Amostra

2: possui 12 registros de *Handroanthus cf. impetiginosus*. Uma vez que o termo “cf.” é designado para as espécies que não possuem uma confirmação laboratorial pelos pesquisadores, desta forma, o modelo treinado pode auxiliar no processo dessas espécies não identificadas no momento.

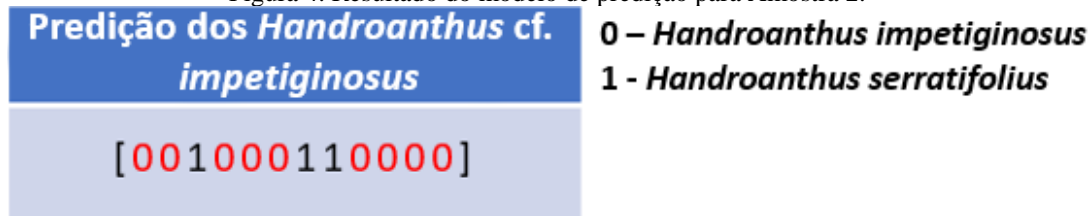
O modelo de classificação desenvolvido reconhece “0” como *Handroanthus impetiginosus* e “1” como *Handroanthus serratifolius*. Desta forma, o processo de validação busca usar somente os dados de entrada das amostras 1 e 2 (que são as mesmas entradas usadas da etapa de treinamento), e verificar quais destas espécies são identificadas a partir do modelo já treinado. Na Figura 3, observa-se o resultado da predição da Amostra 1, em que foi confirmada a identificação de 05 amostras de *Handroanthus serratifolius* (rotulados em vermelho). Enquanto na Figura 4, observa-se o resultado da predição usando Amostra 2, em que ocorreu a confirmação de identificação de 09 registros da espécie *Handroanthus impetiginosus* (rotulados em vermelho). Ressalta-se que os dados dessas amostras de validação não participaram da etapa de treinamento, evitando qualquer viés estatístico. Para esse trabalho, foi possível obter uma matriz de confusão que identificou corretamente 50% dos *Handroanthus impetiginosus* e 83.33% dos *Handroanthus serratifolius*, o que torna o modelo eficiente a predizer as referidas espécies.

Figura 3. Resultado do modelo de predição para Amostra 1.



Fonte: Autores

Figura 4. Resultado do modelo de predição para Amostra 2.



Fonte: Autores

Recentes trabalhos visando o reconhecimento da madeira de espécies tropicais através de redes neurais convolucionais ou ainda modelos de predição obtiveram

resultados semelhantes para diferentes espécies, como Lens et al. (2020) que trabalhou com redes neurais para identificação de imagens microscópicas de espécies neotropicais obteve uma acurácia de 95.6%. No referido trabalho, os autores trabalham com imagens microscópicas que retem uma riqueza de detalhes das características anatômicas das espécies.

5 CONSIDERAÇÕES FINAIS

Neste artigo, foi apresentada uma nova abordagem de identificação de espécies de ipês, utilizando dados categóricos (dados referentes à estrutura anatômica macroscópica) através do algoritmo de *Random Forest*. No entanto, existe a possibilidade de estender essa pesquisa utilizando imagens da estrutura anatômica da madeira, em conjunto com dados categóricos. Ultimamente, tem se demonstrado empiricamente a validação do classificador usando dados qualitativos e/ou quantitativos, aplicando um conjunto de processos estatísticos e algoritmos inteligentes. Os resultados mostraram que é necessário aumentar e diversificar o número de registro de dados das espécies, para que se possa criar um modelo mais robusto no que se diz respeito à identificação das espécies de ipês comercializadas.

Para trabalhos futuros, pode-se considerar a implementação de outros algoritmos de classificação, como SVM (*support Virtual Machine*) ou redes neurais artificiais, junto com algoritmo de *grid search*, para se otimizar a busca de hiperparâmetros ótimos para aumentar nível de precisão do modelo.

REFERÊNCIAS

- ALLENDE-CID, H. Machine Learning: catalisador da ciência. *Computação Brasil*, Belém/PA, v. 39, n. 1, p. 1–46, 2019. Disponível em: <https://www.sbc.org.br>. Acesso em: 15 jul. 2021.
- BERRAR, Daniel. Cross-validation, Japan, p. 1-9, jan 2018. Disponível em: <https://www.researchgate.net/publication>. Acesso em: 9 set. 2021.
- CORADIN, V. T. R., AND MUÑIZ, G. I. B. d. Normas de procedimentos em estudos de anatomia de madeira: Angiospermae e Gymnospermae. (v. 15.), Laboratório de Produtos Florestais - Série Técnica, Brasília, Brasil, 1991.
- CUTLER, A.; CUTLER, D.; STEVENS, J. Random Forests, 2011. Disponível em: <https://www.researchgate.net/publication>.
- Çelik, Özer. (2018). A Research on Machine Learning Methods and Its Applications. 10.31681/jetol.457046. Disponível em: <https://www.researchgate.net/publication>.
- DORMONTT, E. E.; BONERB, M.; BRAUNC, B.; BREULMANND, G.; DEGENE, B.; ESPINOZAF, E.; GARDNERG, S.; GUILLERY, P.; HERMANSONI, J. C.; KOCHJ, G.; LEEK, S. L.; KANASHIROL, M.; RIMBAWANTOM, A.; THOMASN, D.; WIEDENHOEFTO, A. C.; YINP, Y.; ZAHNENQ, J.; LOWE, A. J. Forensic timber identification: it's time to integrate disciplines to combat illegal logging. **Biological Conservation**, v. 191, p. 790–798, 2015.
- ESTEBAN, L. G.; PALACIOS, P.; CONDE, M.; FERNÁNDEZ, F. G.; GARCÍA-IRUELA, A.; GONZÁLEZ-ALONSO, M. Application of artificial neural networks as a predictive method to differentiate the wood of *Pinus sylvestris* L. and *Pinus nigra* Arn subsp. *salzmannii* (Dunal) Franco. **Wood Science and Technology**, v. 51, n. 5, p. 1249-1258, 2017.
- FONSECA, C. N.; LISBOA, P. L. B.; URBINATI, C. V. A xiloteca (Coleção Walter A. Egler) do Museu Paraense Emílio Goeldi. **Boletim do Museu Paraense Emílio Goeldi**, Série Ciências Naturais, v. 1, n. 1, p. 65-140, 2005.
- GENTRY, A. H. *Handroanthus* (Bignoniaceae): a critique. **Taxon**, v. 21, p. 113–114, 1972.
- GENTRY, A. H. *Tabebuia*: The tortuous history of a generic name (Bignon.). **Taxon**, v. 18, p. 635, 1969.
- GROSE, S. O.; OLMSTEAD, R. G. Taxonomic revision in the polyphyletic genus *Tabebuia* s. l. (Bignoniaceae). **Systematic Botany**, v. 32, n. 3, p. 660–670, 2007.
- HE, T.; MARCO, J.; SOARES, R.; YIN, Y., WIEDENHOEFT, A. C. Machine learning models with quantitative wood anatomy data can discriminate between *Swietenia macrophylla* and *Swietenia mahagoni*. **Forests**, v. 11, n. 1, p. 36, 2020.
- HWANG, S.; SUGIYAMA, . Computer vision-based wood identification and its expansion and contribution potentials in wood science: A review. **Plant Methods**, v. 17, n. 1, p. 1–21, 2021.

ŁAWRYNOWICZ, A.; TRESP, V. Introducing machine learning. Perspectives on Ontology Learning, 35–50, 2014. Disponível em: <https://www.researchgate.net/publication>.

LENS, F.; LIANG, C.; GUO, Y.; TANG, X. JAHANBANIFARD, M. ; DA SILVA, F. S. C.; CECCANTINI, G.; VERBEEK, F. J. Computer assisted timber identification based on features extracted from microscopic wood sections. **IAWA Journal** (0), 2020, 1-21.

LEÓN, H. Anatomía del xilema secundario de diez especies de la familia Bignoniaceae de Venezuela. *Acta Botánica Venezuelica*, v. 30 (2), p. 361-384, 2007.

1. LISI, C. S.; SANTANA, C.; MENEZES, N. A.; GONTIJO, A. B.; CORADIN, V. T. R. Madeiras Comerciais do Brasil - Chave interativa de Identificação baseada em caracteres gerais e macroscópicos. O aplicativo disponibiliza um total de 157 espécies de madeiras comerciais do Brasil, Agosto de 2020, versão 1.11.0.

LIASHCHYNSKYI, PETRO & LIASHCHYNSKYI, PAVLO. (2019). Grid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS.

LORENA, A. C.; CARVALHO, A. C. P. L. F. Uma introdução às support vector machines. **Revista de Informática Teórica e Aplicada**, v. 14, n. 2, p. 43–67, 2007.

NYRUD, Q. A; ROOS, A.; RODBOTTEN, M. Product attributes affecting consumer preference for residential deck materials. **Canadian Journal of Forest Research**, v. 38, n. 6, p. 1385–1396, 2008.

PACE, M.R.; LOHMANN, L.G.; OLMSTEAD, R.G.; ANGYALOSSY, V. Wood anatomy of major Bignoniaceae clades. *Plant Systematics and Evolution*, v. 301, p. 967-995, 2015

Patel, H. H.; Prajapati, P. (2018). Study and analysis of decision tree based classification algorithms. **International Journal of Computer Sciences and Engineering**, v. 6, n. 10, p. 74–78. Disponível em: <https://www.researchgate.net/publication>.

RAVINDRAN, P.; COSTA, A.; SOARES, R.; WIEDENHOEFT, A. C.. Classification of CITES-listed and other neotropical Meliaceae wood images using convolutional neural networks. **Plant methods**, v. 14, n. 1, p. 1–10, 2018.

SILVA, N. R.; RIDDER, M.; BAETENS, J. M.; BUCKLE, J. V. den; ROUSSEAU, M.; BRUNO, O. M.; BEECKMAN, H.; ACKER, J. VAN; BAETS, B. de. Automated classification of wood transverse cross-section micro-imagery from 77 commercial Central-African Timber species. **Annals of Forest Science**, 74:30 DOI 10.1007/s13595-017-0619-0, 2017.

TROPICOS (2021). “Missouri botanical garden,” (<http://www.tropicos.org>), accessed August 31, 2021.

VIEIRA, G. Utilização de classificador random forest na detecção de falhas em máquinas rotativas. 2015. 79f. TCC (Graduação) – Curso de Engenharia Eletrônica e de Computação da Escola Politécnica, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2015.

Wang, P. On defining Artificial Intelligence. **Journal of Artificial General Intelligence**, v. 10, n. 2, p. 1–37, 2019. Disponível em: <https://www.sciendo.com/article>.

WEHLE, H. (2017). Machine Learning, Deep Learning, and AI: what's the difference?. Disponível em: <https://www.researchgate.net/publication>.