

Um modelo de classificação de doenças da folha do arroz através de imagens*

Rui Antonio da Rocha Lorensi¹[0000-0002-0215-9547], Mikael Adolfo Maria¹[0000-0002-5071-9799], Anastácio Barcellos Loguercio¹[0000-0002-4154-8422], and Sandro da Silva Camargo¹[0000-0001-8871-3950]

Programa de Pós-Graduação em Computação Aplicada,
Universidade Federal do Pampa, Bagé, RS, Brasil
<http://cursos.unipampa.edu.br/cursos/ppgcap/>
{ruilorensi.aluno,mikaemarkaria.aluno,anastaciologuercio.aluno,
sandrocarmargo}@unipampa.edu.br

Resumo O arroz é um dos alimentos mais consumidos no mundo e o Brasil é o segundo maior produtor mundial. Os cuidados com a produção englobam desde tecnologias de melhoramento dos grãos em laboratório até o monitoramento de pragas e fungos na lavoura. Este trabalho apresenta um estudo acerca da construção de modelos que possam auxiliar na identificação de fungos baseados em imagens coletadas em um estudo anterior, contendo amostras infectadas com Brusone, Mancha Marrom e Mancha Estreita. Testes mostraram que o modelo criado conseguiu classificar estas doenças com acurácia de 100%.

Palavras-chave: árvores de decisão · inteligência artificial · agricultura · processamento de imagens · orizícola

1 Introdução

O Brasil é o maior produtor e consumidor mundial de arroz fora da Ásia. Em média, anualmente, são produzidos cerca de 15 milhões de toneladas de arroz em casca em todo território brasileiro sendo mais de 12 milhões de toneladas destinados para atender ao consumo interno. Além disso, a produção restante é exportada[1]. A região Sul do Brasil, que abrange os estados do Rio Grande do Sul, Santa Catarina e Paraná, concentra em torno de 75% da produção brasileira de arroz. Fatores como a área útil para produção e o desenvolvimento de novas tecnologias de Agricultura de Precisão (AP) possibilitam uma perspectiva de crescimento na produção para atender às demandas internas e externas[1].

O fato da produção ser realizada em larga escala faz com que sejam necessários cuidados com a qualidade da planta em todos os estágios da cultura a fim de minimizar perdas, pois o número de ameaças que podem surgir ao longo da

* Este trabalho foi desenvolvido com apoio financeiro da CAPES/FAPERGS (PDPG), Edital nº 18/2020.

R. A. R. Lorensi et al.

produção podem levar a um número significativo. Ao longo do ciclo de produção do arroz, fatores como insetos, ervas daninhas e doenças são responsáveis por percentuais significativos de prejuízos, em alguns casos irreparáveis [10]. No mundo, mais de 80 doenças podem afetar o cultivo e, no Brasil, esse número ainda não é exato sendo que algumas delas, que ocorrem em escala menor, ainda não foram documentadas [11]. Geralmente, as doenças podem ocorrer em qualquer planta a qualquer momento. Logo, um monitoramento contínuo pode prevenir a disseminação das doenças [7].

Dentre as doenças que causam dano econômico na cultura do arroz, as mais comuns são de origem fúngica, resultando em perdas econômicas significativas na cadeia de produção. A disseminação e o ataque de fungos patogênicos na lavoura são favorecidos em condições adequadas de temperatura e umidade [6]. No Rio Grande do Sul, assim como em diversas partes do Brasil e do mundo, a lista de doenças é extensa, sendo considerada em ordem de importância: 1) Brusone (*Magnaporthe oryzae*) como a mais importante doença do arroz, podendo levar a perda total da produção; e 2) a mancha parda (*Bipolaris oryzae*) que pode causar danos de 12% a 30% na massa dos grãos e de 18% a 22% no número de grãos cheios por panícula. A escaldadura das folhas (*Gerlachia oryzae*), a queima das bainhas (*Rhizoctonia solani*), a mancha-estreita (*Sphaerulina oryzina*) são outras doenças de importância significativa. Segundo [7], cada doença possui elementos que a torna única entre as outras, por exemplo, forma, cor e tamanho são características que representam cada patógeno. Entretanto, em função de algumas similaridades como cor, forma e tamanho, a precisão da identificação de qual fungicida que se deve utilizar pode ser comprometida, acarretando na redução da eficácia do combate aos fungos. Assim, o reconhecimento precoce destas doenças se torna um fator indispensável para aplicar medidas de manejo corretas na tentativa de sanar ou minimizar estes problemas causados por estes patógenos. O reconhecimento de qual doença está atacando a planta pode ser um desafio, porém identificar corretamente que tipo de invasor facilita na aplicação correta de fungicidas para combate destas doenças.

Modelos de classificação baseados em aprendizado de máquina já foram relatados em outros trabalhos publicados na literatura. Sahith e colaboradores [9] utilizaram diferentes algoritmos baseados em árvores de decisão para um conjunto de 120 amostras, criando um modelo que atingiu uma acurácia de 76% com o algoritmo *Random Forest*. Charliepaul [3] apresentou um modelo de classificação de doenças da folha de arroz, baseado em regras, que atingiu 75% de precisão, a partir de uma base de imagens de três doenças diferentes, a fim de extrair características de folhas infectadas. Já Kahar e colaboradores [4] criaram um modelo com um classificador de Máquinas de Vetor de Suporte (SVM) que atingiu uma acurácia de 68% para classificação de três diferentes doenças presentes nas folhas do arroz.

Neste contexto, o objetivo deste trabalho é apresentar um modelo de classificação que auxilie a identificar e classificar corretamente as doenças fúngicas que atacam à folha do arroz prejudicando a produtividade. Este artigo está organizado da seguinte forma: na Seção 2, é exposta a metodologia proposta para

Um modelo de classificação de doenças da folha do arroz...

o desenvolvimento do trabalho. Na Seção 3, são apresentados e discutidos os resultados obtidos no presente estudo. Finalmente, na Seção 4, são apresentadas as conclusões e trabalhos futuros.

2 Material e Métodos

2.1 Base de Dados

Para este estudo, foi utilizada uma base de imagens, disponível em [5], composta de 120 registros fotográficos de folhas de arroz doentes e separadas em três grupos de doenças, com 40 imagens de cada um dos tipos: Brusone, Mancha Parda e Mancha Estreita. O resultado de cada doença na folha do arroz é mostrado na Figura 1. Entretanto, após uma triagem inicial, o total final de amostras foi de 112 conforme apresentado na Tabela 1, pois algumas imagens não tinham a qualidade mínima necessária, incapaz de produzir resultados satisfatórios. Foi necessário também, realizar um pré-processamento destas imagens para padronizar um fundo preto conforme apresentado na Figura 1.

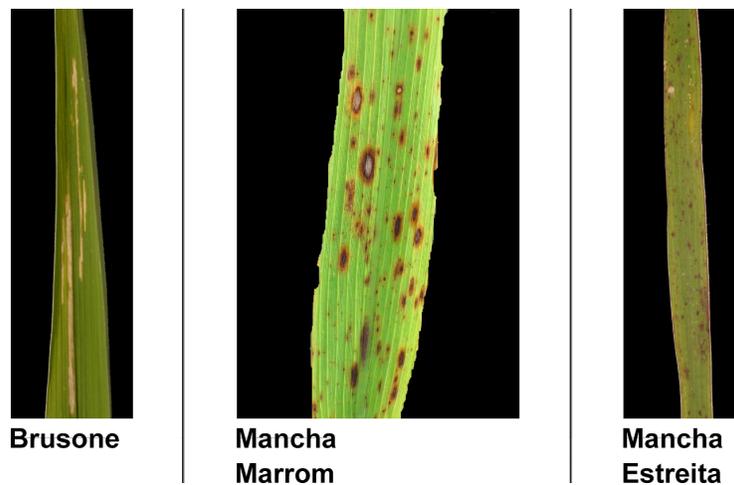


Figura 1. Amostras dos três tipos de doença da folha do arroz.

O *dataset* gerado é composto por 10 atributos, dos quais quatro são coletados das informações apresentadas pela função de histograma do *Software InfraView 64-bit Version*: *averageR*, *averageG*, *averageB*, *averageGray*, conforme apresentado na Figura 2. Outros seis atributos são obtidos através da análise da imagem, sendo eles: *averageRGBGray*, *formato_doenca*, *area_doenca*, *borda_folha_manchada* e *granularidade_manchas*.

Após a triagem inicial e padronização do fundo foi realizada a classificação de acordo com as características apresentadas em cada imagem.

R. A. R. Lorensi et al.

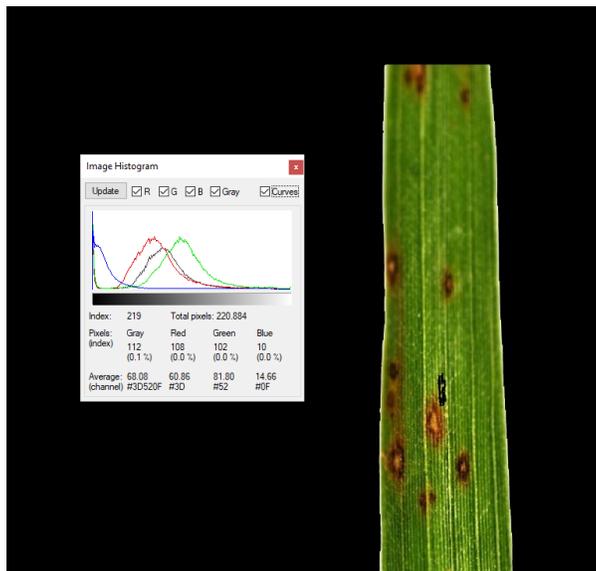


Figura 2. Coleta das médias dos campos $Red(R)$, $Green(G)$, $Blue(B)$, $Gray$.

Tabela 1. Quantidade de amostras disponíveis para cada doença.

Doença	Quantidade de amostras
Brusone	39
Mancha Marrom	39
Mancha Estreita	34

2.2 Classificação com Árvores de Decisão

No presente trabalho, foram aplicadas as técnicas de aprendizado supervisionado, que são uma classe de algoritmos que visam aprender uma função arbitrária que associa os dados de entrada e as classes de saída, que neste caso são os nomes das doenças nas folhas do arroz, previamente identificadas por especialistas. Esta associação geralmente descreve uma função $f_o(x)$ presente de forma implícita em um conjunto de treinamento $D = [x(i), y(i)] \in \mathbb{R} \times \mathbb{R}, i = 1, \dots, l$ consistindo de l pares $(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)$. As entradas x consistem em um vetor n -dimensional onde $x \in \mathbb{R}$, e as saídas y consistem em um vetor 1-dimensional onde $y \in \mathbb{R}$. Durante o processo de treinamento, as amostras são submetidas ao algoritmo de aprendizado. Após o processo de aprendizado supervisionado, é criado um modelo que pode ser utilizado para simular o conhecimento do especialista do domínio.

Neste trabalho, foram aplicadas técnicas de classificação e foi priorizada a construção de modelos do tipo caixa branca, onde se enquadram os algoritmos de árvores de decisão. Apesar destes modelos serem lineares e, tipicamente, terem uma capacidade preditiva menor que modelos não lineares, sua maior vantagem

Um modelo de classificação de doenças da folha do arroz...

é a geração de modelos interpretáveis [2]. Na fase de avaliação dos modelos foi utilizada a técnica *holdout* para permitir a correta mensuração da capacidade preditiva dos modelos. Desta forma, os dados disponíveis foram particionados em dois conjuntos: dados de treino, com 71% das amostras, para criar os modelos; e dados de teste, com os outros 29% das amostras, para avaliar a capacidade preditiva dos modelos.

O algoritmo de geração de árvores utilizado foi o *J48*, que classifica instâncias ordenando-as da raiz da árvore em direção a suas folhas [8]. Os preditores mais relevantes posicionam-se mais perto da raiz da árvore. À medida que a relevância do preditor diminui, ele é posicionado mais longe da raiz e mais próximo às folhas.

2.3 Ferramentas

O processo de padronização do fundo preto das imagens foi realizado através do *Photoshop* na versão 22.1.0, e as médias *RGBGray* das imagens foram analisadas através do *software InfraView 64-bit Version*, utilizando a função de histograma para a verificação dos níveis médios dos canais *Gray*, *Red*, *Green*, *Blue*. Também foram utilizados os ambientes *R-Studio* versão 1.4.1106, e a linguagem R versão 3.6.3 instalado no sistema operacional *Windows 10 Pro - 64 bits* com processador *Intel Core I5-3470* 8,00GB de memória *RAM* e disco sólido interno SSD 128GB. Foram também utilizados os pacotes *Classification and Regression Training* (Caret) versão 6.0-86, *Rpart.plot* versão 3.0.9, *Variable Importance Plots* versão 0.3.2.

3 Resultados e Discussão

O presente estudo teve como finalidade apresentar uma proposta de classificação de patógenos do arroz através da análise de amostras reais e utilização de árvores de decisão e aprendizado de máquina. As amostras foram analisadas de acordo com as características apresentadas nas imagens capturadas e o processamento segmentou o total das 112 amostras, as quais foram divididas em dois grupos, sendo 71% para treino dos modelos e 29% para teste, o que resultou em 80 amostras para treino e 32 amostras para teste.

No conjunto de amostras para treino foi aplicado o algoritmo de classificação *J48* gerando um modelo preditivo para auxiliar na identificação das doenças relacionadas à folha do arroz. Na Figura 3 tem-se o Formato da Doença e a Granularidade das Manchas como atributos principais para a classificação. Após a análise da árvore, nota-se que folhas doentes, que não apresentam formatos Lineares ou do tipo Mix, somam um total de 35% do total de amostras de teste e podem ser classificadas, segundo o algoritmo, com um percentual de 100% como portadoras da doença da Mancha Marrom. No lado esquerdo da árvore, tem-se um total de 35% e 30% de amostras portadoras de Brusone ou Mancha Estreita respectivamente, sendo classificadas de acordo com a presença de Granularidade das doenças. Nesta segunda avaliação dos resultados apresentados pelo algoritmo, ou seja, do ramo esquerdo, tem-se um percentual de 54%

R. A. R. Lorensi et al.

de amostras reconhecidas como portadoras da patologia Brusone (lado esquerdo da árvore) e no outro ramo um percentual de 46% de amostras portadoras da doença Mancha estreita. A Figura 4 apresenta a superfície de decisão gerada para a classificação relacionada à Forma da Doença, que relaciona a presença das formas Oval e Redonda nas folhas com os registros da doença da Mancha Marrom.

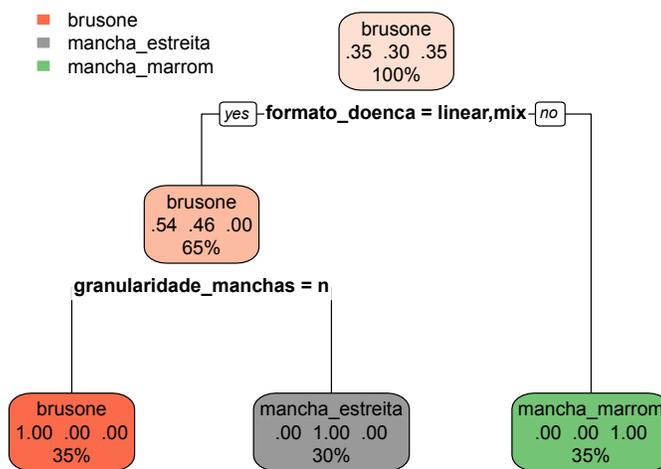


Figura 3. Modelo preditivo de Árvores de Decisão criado pelo algoritmo *J48*.

O segundo ramo de classificação, do lado esquerdo da árvore, divide os resultados de acordo com a presença ou ausência de granularidade, conforme a Figura 5, explicitando a ausência de Granularidade em folhas portadoras da doença de Brusone e apresentando amostras com dois tipos de Granularidade (p e m) para as folhas que estão com a patologia da Mancha Estreita.

A matriz de confusão apresentada na Tabela 2 sumariza as classificações realizadas nos dados de teste. O resultado presente nesta tabela aponta para 32 classificações corretas, sendo 11 para a classe Brusone, 10 para a classe Mancha Estreita e 11 para a classe Mancha Marrom. Não houve erros nas predições realizadas.

A Tabela 3 apresenta as métricas de avaliação do modelo sobre os dados de teste. Os resultados mostram que o modelo atingiu uma acurácia de 100%, calculado a partir de 32 classificações corretas em 32 amostras de teste. Com intervalo de confiança de 95%, estima-se que o modelo pode atingir acurácia de 89,11% a 100%. A taxa sem informação de 34,38% indica a proporção de amostras da maior classe, o que seria o baseline caso o modelo tentasse classificar todas as amostras como a doença mais numerosa. O p-valor deste modelo é

Um modelo de classificação de doenças da folha do arroz...

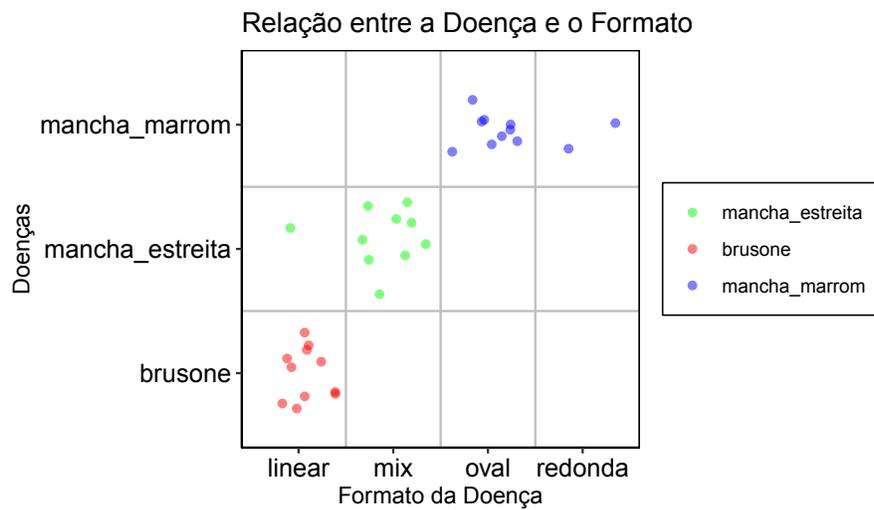


Figura 4. Superfície de Decisão do ramo direito da árvore para folhas portadoras de manchas Ovais ou Redondas.

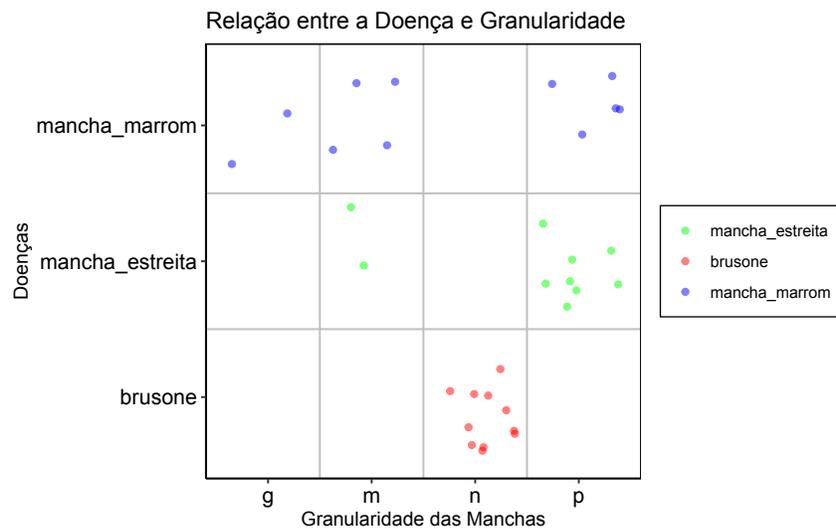


Figura 5. Superfície de Decisão do ramo esquerdo da árvore para os índices de Granularidade das manchas nas folhas.

inferior a 0.05, indicando que os resultados obtidos com as 32 amostras de teste também devem se repetir em populações maiores em mais de 95% das vezes. O coeficiente de Kappa é de 1, indicando um nível de concordância perfeito das decisões do modelo em relação ao diagnóstico real.

R. A. R. Lorensi et al.

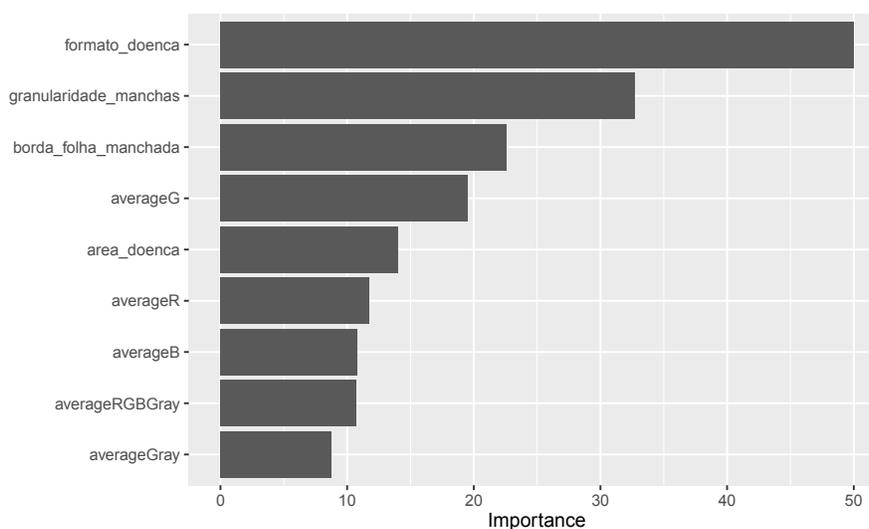
Tabela 2. Matriz de Confusão

Predição	Referência		
	Brusone	Mancha Estreita	Mancha Marrom
Brusone	11	0	0
Mancha Estreita	0	10	0
Mancha Marrom	0	0	11

Tabela 3. Estatísticas do Modelo

Métrica	Valor
Acurácia	1
95% CI	(0.8911, 1)
Taxa sem Informação	0.3438
P-Valor [Acc > NIR]	1.445e-15
Kappa	1

A fim de melhor entender os resultados obtidos no processamento dos dados foi utilizada a biblioteca *Variable Importance Plots* (VIP) para visualizar os atributos em ordem de importância. Analisando a Figura 6 tem-se como as duas principais características o Formato da Doença e a Granularidade das Manchas presentes nas folhas. Estes atributos foram utilizados como variáveis na representação dos gráficos apresentados nas Figuras 4 e 5.

**Figura 6.** Importância dos atributos para identificação das doenças da folha do arroz.

Um modelo de classificação de doenças da folha do arroz...

4 Conclusão

Este trabalho identificou, a partir de dados gerados após coleta estratificada, pré-processamento e análise visual de imagens, que características visuais impactam diretamente na identificação do tipo de doença que acomete a folha do arroz e consequentemente afeta a planta como um todo. A base de dados utilizada é pública e continha 112 amostras das três doenças mais comuns para a cultura do arroz: Brusone, Mancha Estreita e Mancha Marrom. Foi criado um modelo preditivo de árvores de decisão que atingiu uma acurácia de 100% nos dados de teste.

Duas limitações deste trabalho são a quantidade de amostras, que pode ser expandida a partir de novos estudos experimentais, e a extensão para outras doenças comuns na cultura. Outra limitação importante é a dependência da extração manual das características da imagem por um especialista.

Como trabalhos futuros, podem ser avaliadas: 1) a criação de um módulo de pré-processamento de imagens que possa extrair automaticamente as características visuais das amostras, e 2) a utilização de técnicas de *deep learning* que podem fazer a extração de características e o modelo de classificação. Além disso, a coleta de novas imagens com outras doenças comuns a esta importante cultura na economia do sul do Brasil seria uma contribuição importante.

Referências

1. Brazilian Rice: Production profile. <http://brazilianrice.com.br/br/>, acessado em 18 Mai 2021.
2. Camargo, S.S., Azambuja, R.C.C., Feijó, J.O., Corrêa, M.N., Schneider, A., Cardoso, F.F.: Modelagem computacional de indicadores metabólicos para estudo de eficiência reprodutiva em vacas de corte. In: Anais Eletrônicos do X Congresso Brasileiro de Agroinformática. pp. 857–866. Sociedade Brasileira de Agroinformática, Ponta Grossa-PR (2015)
3. C.K.Charliepaul: Classification of rice plant leaf using feature matching. International Journal On Engineering Technology and Sciences **1**, 290–295 (2014)
4. Kahar, M.A.A., Mutalib, S., Abdul-Rahman, S.: Early detection and classification of paddy diseases with neural networks and fuzzy logic. In: Proceedings of the 17th Inter-national Conference on Mathematical and Computational Methods in Science and Engineering. pp. 248–257 (2015)
5. Machine Learning Repository: Rice leaf diseases data set (2021), <https://archive.ics.uci.edu/ml/datasets/Rice+Leaf+Diseases>, acessado em 15 Mai 2021.
6. de Oliveira Neto (Org), A.A.: A cultura do arroz. Tech. rep., Companhia Nacional de Abastecimento, Brasília-DF (2015)
7. Prajapati, H.B., Shah, J.P., Dabhi, V.K.: Detection and classification of rice plant diseases. Journal of Intelligent Decision Technologies **11**(3), 357–373 (2017). <https://doi.org/10.3233/IDT-170301>
8. Quinlan, J.R.: Programs for Machine Learning, vol. 1. Morgan Kaufmann Publishers, New York (1993)

R. A. R. Lorensi et al.

9. Sahith, R., Reddy, P.V.P., Nimmala, S.: Decision tree-based machine learning algorithms to classify rice plant diseases. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)* **9**(1), 5365–5368 (2019). <https://doi.org/10.35940/ijitee.A4753.119119>
10. dos Santos, G.R., Chagas, J.F.R., Tavares, A.T., de Castro Neto, M.D., de Almeida Sarmiento, R., Júnior, A.F.C., do Nascimento, I.R.: Danos causados por doenças fúngicas no arroz cultivado em várzeas no sul do estado do tocamtins. *Fitossanidade* **70**(4), 869–875 (2011). <https://doi.org/10.1590/S0006-87052011000400020>
11. Silva-Lobo, V.L., de Filippi, M.C.C.: Manual de identificação de doenças da cultura do arroz. Tech. rep., Embrapa, Brasília-DF (2017)