Sistema de Classificação do *Capsicum Annuum* Acometido Pela Doença Fúngica *Antracnose* Através do *Random Forest*

Roney Nogueira de Sousa[1], **Rhyan Ximenes de Brito**[1]

[1]Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE)
CE-187, s/n - Estádio, – CEP 62320-000 – Tianguá – CE – Brasil

{nogueiraroney453, rxbrito}@gmail.com

Abstract. *The use of computational techniques of Artificial Intelligence has gained space in the aid of several problems. Based on this, a study was carried out aiming to use the Random Forest classifier, present in the WEKA software package, as an aid tool in the pre-diagnosis of anthracnose fungal disease in a pepper garden. The methodology was composed from a database, collected in the city of Tianguá-CE, in the locality known as Sitio Córrego. 10 training sessions were performed with normalized and balanced data, using k-fold cross validation. The results were analyzed statistically, obtaining a 90.84% correctness measure.*

Resumo. *O uso de técnicas computacionais de Inteligência Artificial tem ganhado espaço no auxílio de diversas problemáticas. Com base nisso foi realizado um estudo visando a utilização do classificador Random Forest, presente no pacote do software WEKA, como ferramenta de auxílio no pré-diagnóstico da doença fúngica antracnose em uma horta de pimentões. A metodologia foi composta a partir de um banco de dados, coletados na cidade de Tianguá-CE, na localidade conhecida como Sitio Córrego. Foram feitos 10 treinamentos com os dados normalizados e balanceados, utilizando a validação cruzada k-fold. Os resultados foram analisados estatisticamente obtendo-se uma medida de acerto de 90,84%.*

1. Introdução

A importância econômica de pimentões do gênero *Capsicum* vem crescendo no Brasil e em diversos países, com o aumento do consumo *in natura* ou através do processamento de molhos, temperos e conservas [de Azevedo et al. 2006]. Deve-se salienar que devido a doenças e pragas que afetam o plantio do fruto, pequenos e grandes agricultores tem observado impactantes danos na produção [de Azevedo et al. 2006].

Nessa perspectiva a antracnose, doença causada por fungos do gênero *Colletotrichum* é uma das doenças de maior importância para as solanáceas como o pimentão *(Capsicum annum)* [de França et al. 2015]. A antracnose pode atingir todos os estágios de desenvolvimento da planta, desde mudas no viveiro até os frutos após a colheita [de Azevedo et al. 2006].

Nesses termos, faz-se necessário a utilização de técnicas de Inteligência Computacional que auxiliem no diagnóstico efetivo e rápido da doença. Esse trabalho tem como principal objetivo realizar um estudo utilizando o algoritmo *Random Forest* disponível

no aplicativo WEKA (*Waikato Environment for Knowledge Analysis*), com o objetivo de utilizá-la para a diagnose efetiva da doença fúngica antracnose em pimentões.

O presente estudo foi realizado a partir do aprofundamento bibliográfico acerca do assunto e na coleta e implementação de um banco de dados composto 368 amostras de plantas, divididas em 9 atributos utilizados como valores de entrada. Os resultados foram analisados através dos resultados e níveis de acurácia do modelo escolhido.

O trabalho em questão foi motivado pela necessidade do aprimoramento de técnicas que garantam o diagnóstico, efetivo e precoce do patógeno visando auxiliar profissionais da agricultura a tomarem medidas que visem o manejo e controle da doença de modo a reduzir perdas na produção.

Este trabalho está organizado da seguinte forma: Seção 2 apresenta alguns trabalhos relacionados, Seção 3 apresenta a fundamentação teórica sobre a doença fúngia antracnose e sobre o classificador *Random Forest*, a Seção 4 apresenta a metododologia, Seção 5 resultados e discussões e a Seção 6 as considerações finais e trabalhos futuros.

2. Trabalhos Relacionados

Esta seção apresenta uma revisão bibliográfica de forma reflexiva sobre diferentes abordagens dentro do escopo da Inteligência Computacional.

[de Sousa et al. 2019] realizou um estudo por meio da implementação da rede neural RBF (*Radial Basis Function*), objetivando usá-la como ferramenta de apoio no diagnóstico da doença HLB (*Huanglongbing*) em um pomar de tangerina poncã. Foram utilizadas 29 amostras de plantas compostas por 8 atributos de uma base de dados. Utilizou-se de 10 treinamentos com os dados balanceados e normalizados com a validação cruzada (*leave-onte-out*). Os resultados foram analisados estatisticamente obtendo-se uma taxa média de acertos de 74,4%.

[Ferreira et al. 2019] fez uso da lógica *Fuzzy* como mecanismo auxiliar na análise do tomateiro para detectar se o mesmo é portador ou poderá desenvolver a murcha de *fusarium*, auxiliando assim profissionais da área da agricultura no diagnóstico da doença antes que possa infectar todo o plantio. O trabalho teve como base pesquisas bibliográficas e dados coletados em uma propriedade rural do interior do estado do Ceará.

[Funck 2019] apresentou um aplicativo *mobile* para a plataforma *Android* que é capaz de identificar a ferrugem asiática da soja quando ela estiver presente na folha da planta, utilizando a câmera de um dispositivo *smartphone*. Assim foram aplicadas técnicas de visão computacional e um modelo de Rede Neural Convolucional para realizar a classificação das imagens. Foram reunidas 2.770 imagens de folhas de soja, utilizando uma parte para os testes e outra para a validação. O aplicativo atingiu uma acurácia de 84,61% na detecção da ferrugem asiática.

[Afonso 2019] desenvolveu um sistema capaz de diagnosticar doenças em folhas de videira através de técnicas de visão computacional. A ideia principal do sistema foi extrair características específicas de uma certa doença, para posteriormente identificar quando encontrá-la novamente. Dessa forma o sistema de visão computacional utilizava imagens de folhas de videira. Com condições de captura específicas obtiveram o diagnóstico sobre o estado da folha, com a doença identificada, ou no caso de diagnóstico negativo, a informação que a folha é saudável.

3. Fundamentação Teórica

Esta seção traz um estudo sobre a doença antracnose abordando os sintomas e formas de contágio assomo como o classificador *Random Forest* com suas particularidades.

3.1. A Doenção Fúngica Antractnose

O pimentão é uma das hortaliças mais apreciadas, consumidas e produzidas no Brasil e seu cultivo tem aumentado com a intensificação do cultivo em casas-devegetação que permitem extensão do tempo de colheita e melhores safras em períodos ambientais adversos [de Azevedo et al. 2006].

Nessa perspectiva o patógeno antracnose pode afetar toda a parte aérea da planta, porém sua ação mais importante se dá nos frutos. Em um ataque severo, até 100% dos frutos podem ser afetados, ocasionando perda total para o produtor [Viana et al. 2007]. A Figura 1 mostra exemplos de frutos infectados com a doença antracnose.

Figura 1. Frutos Infectados Pela Antracnose

O fungo é disseminado pela água da chuva ou vento podendo ser transmitido por sementes, sobrevivendo ainda em restos de cultura [Viana et al. 2007].

A maioria dos relatos de antracnose em pimentões no Brasil aponta o fungo *Colletotrichum gloeosporioide* como o responsável pela doença, embora outras espécies do mesmo gênero possam também causar a doença, tais como, *C. acutatum, C. coccodes, C. capsici e C. dematium* [Viana et al. 2007]. As condições ideais para a ocorrência de infecção são alta umidade relativa do ar e temperatura de, $20\,^{\circ}\mathrm{C}$ a $30\,^{\circ}\mathrm{C}$, porém a doença pode ocorrer em temperaturas de $10\,^{\circ}\mathrm{C}$ a $30\,^{\circ}\mathrm{C}$ [Ghini and Hamada 2008].

Nas hastes afetadas podem ser observadas lesões escuras em forma de estrias, e nas folhas manchas necróticas, secas, irregulares e de coloração parda [Pereira et al. 2013], nos frutos, onde as lesões são mais típicas, essas são deprimidas, circulares, de bordos elevados e de diferentes tamanhos. Nessas lesões, sob condições de elevada umidade, o patógeno produz frutificações na forma de acérvulos escuros com massas de conídios de coloração rósea, salmão ou alaranjada [Viana et al. 2007].

3.2. Classificador *Random Forest*

O algoritmo RF (*Random Forest*) consiste em uma técnica de aprendizado de máquina que agrupa diversas árvores de decisão criadas a partir de uma base de treinamento, de modo que o modelo resultante consolida os resultados de todas as árvores para predições [da Costa Filho et al. 2019].

A criação das árvores é realizado por meio de seleção de atributos aleatória dos subconjuntos, sendo que a RF é formada por uma coleção de árvores de decisão [Silva et al. 2019]. Em geral, as RFs atingem boa acurácia preditiva quando comparadas a outros métodos de aprendizado de máquina supervisionados [Riqueti et al. 2018].

Este tipo de modelo é usualmente utilizado não apenas para classificação, mas também para regressão, estudo de importância e seleção de variáveis, e detecção de *outliers* [de Alvarenga Júnior 2018]. A Figura 2 mostra um exemplo da estrutura do classificador *Random Forest*.

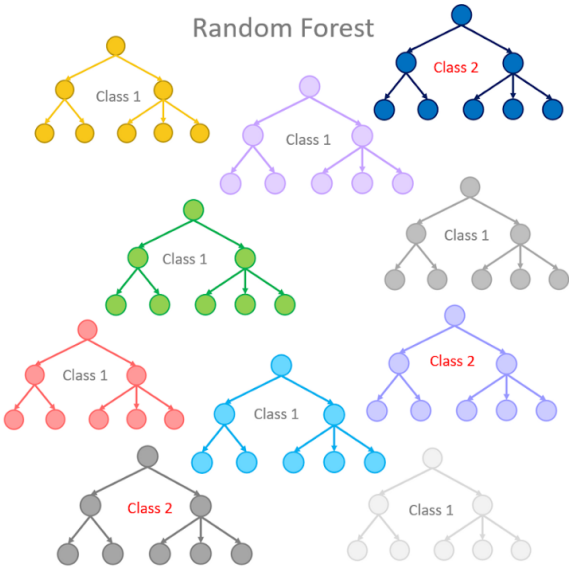

Figura 2. Exemplo da Estrutura da Rede *Random Forest*

4. Metodologia

A metodologia foi estrutura em fases como: (i) coleta e montagem da base de dados; (ii) estudo da ferramenta WEKA e treinamento e teste com a base de dados.

4.1. Base de Dados Utilizada

Os dados foram coletados na zona rural da cidade de Tianguá-CE na localidade conhecida como Sítio Córrego. A base de dados é por 368 amostras de frutos extraídas de 71 plantas, destas amostras foram analisadas 9 características divididas em 3 partes da planta: caule, folhas e frutos. A partir do caule foram observadas as seguintes características, lesões em forma de estrias, necrose e lesões escuras nas hastes. Já nas folhas foram observadas as características, folhas irregulares com coloração parda, manchas necróticas, manchas circulares. Nos frutos, lesão circular de bordos elevados, necrose nos botões florais, lesões nos frutos.

4.2. Treinamento e Teste com a Base de Dados

Para o treinamento e teste utilizou-se o classificador RF, composto por 200 árvores de decisão formadas de forma aleatória. Utilizou-se do algoritmo *Random* pertencente ao pacote de árvores de decisão do *software WEKA*. As amostras foram balanceadas garantindo equidade entre elas.

O *WEKA* é uma suíte de mineração de dados muito popular no meio acadêmico, desenvolvida utilizando a linguagem Java. Criada nas dependências da Universidade de Waikato, Nova Zelândia [Damasceno 2015]. Fornece diversos algoritmos em seu ambiente e, além disso, permite que sejam criados algoritmos de aprendizado. Os principais tipos de algoritmos que estão disponíveis são: regressão, classificação, mineração de regras de associação, seleção de atributos, etc. [Sato et al. 2013].

5. Resultados e Discussão

Os resultados obtidos e analisados tiveram como base os percentuais de acurácia adquiridos nos treinamentos e testes realizados com o classificador RF assim como na discussão dos resultados encontrados por meio das amostras do banco de dados coletado.

Vale salientar que para o treinamento e teste foi utilizado o método de validação cruzada *k-fold*, com o k=10 *folds*, outras configurações foram testadas com k = 5, k = 15, k = 20, com a base de dados normalizada e balanceada.

Como observado na Tabela 1, o melhor caso atingiu uma acurácia de 91,83% de acertos e 8,17% de erros e o pior caso 89,88% de acertos e 10,12% de erros. A taxa média de acertos foi de 90,84% de acertos e 9,16% de erros.

Tabela 1. Resultados dos Testes com o RF

Treinamento	Taxa de Acerto(%)	Taxa de erro(%)	Situação resultado
1	90,63	9,37	
2	90,89	9,11	
3	91,83	8,17	Melhor Caso
4	89,88	10,12	Pior Caso
5	90,43	9,57	
6	90,82	9,18	
7	90,82	9,18	
8	90,63	9,37	
9	91,83	8,17	Melhor Caso
10	90,63	9,37	
total	90,84	9,16	Caso Médio

6. Considerações e Trabalhos Futuros

Este trabalho teve como finalidade mostrar a utilização do classificador *Random Forest* como mecanismo auxiliador no diagnóstico de doenças fúngicas como a antracnose que acomete o pimentão. Para os testes foram utilizados dados reais gerando resultados confiáveis que podem ser utilizados por profissionais da agricultura.

A rede poderá ser usada para auxiliar na detecção precoce da antracnose, pois ao se detectar os primeiros sintomas da doença os cuidados necessários para o combate

serão mais eficientes. Porém o monitoramento das áreas e épocas mais susceptíveis a infestações é um fator importante, assim como o uso de técnicas de Inteligência Computacional que podem ser utilizadas para assegurar estratégias afim de garantir um manejo preventivo com a utilização de fungicidas indicados por um profissional da área.

Como trabalhos futuros é sugerido a realização de um estudo comparativo com outras redes neurais artificiais, tais como a RBF (*Radial Basis Function*) ou a ELM (*Extreme Learning Machine*), verificando assim qual o algoritmo que possui o melhor desempenho para a problemática em questão.

Referências

Afonso, S. (2019). Utilização de algoritmos de visão computacional para diagnóstico de doenças da vinha.

da Costa Filho, S. V. S., Arce, J. E., Montaño, R. A. N. R., and Pelissari, A. L. (2019). Configuração de algoritmos de aprendizado de máquina na modelagem florestal: um estudo de caso na modelagem da relação hipsométrica. *Ciência Florestal*, 29(4):1501–1515.

Damasceno, M. (2015). Introdução a mineração de dados utilizando o weka. *Disponivel: http://connepi. ifal. edu. br/ocs/index. php/connepi/CONNEPI2010/paper/viewFile/258/207, Acesso: setembro*.

de Alvarenga Júnior, W. J. (2018). Métodos de otimização hiperparamétrica: um estudo comparativo utilizando árvores de decisão e florestas aleatórias na classificação binária.

de Azevedo, C. P., Café Filho, A., and REIS, A. (2006). Recomendações de manejo da antracnose do pimentão e das pimentas. *Embrapa Hortaliças-Comunicado Técnico (INFOTECA-E)*.

de França, G. S., da Costa-Carvalho, R. R., Neves, R. P., Araujo, E. R., and Laranjeira, D. (2015). Controle pós-colheita da antracnose do pimentão pela levedura rhodotorula glutinis. *Bioscience Journal*, 31(2).

de Sousa, R. N., de Brito, R. X., and de Sousa Ximenes, J. N. (2019). Aplicação da rede neural rbf como mecanismo auxiliador no diagnóstico da doença hlb em espécies cıtricas. In *Anais da VII Escola Regional de Computação Aplicada à Saúde*, pages 55–60. SBC.

Ferreira, A. F., Moreira, L., de Arruda, G., de Brito, R., and Ferreira, A. V. (2019). Sistema fuzzy como ferramenta auxiliar na detecção da murcha de fusarium no tomateiro. In *Anais da VII Escola Regional de Computação Aplicada à Saúde*, pages 193–198, Porto Alegre, RS, Brasil. SBC.

Funck, F. C. (2019). Detectando a ferrugem asiática na folha da soja utilizando redes neurais convolucionais. B.S. thesis, Universidade Tecnológica Federal do Paraná.

Ghini, R. and Hamada, E. (2008). *Mudanças climáticas: impactos sobre doenças de plantas no Brasil.* Brasília, DF: Embrapa Informação Tecnológica; Jaguariúna: Embrapa Meio . . . .

Pereira, R., PINHEIRO, J., and de CARVALHO, A. (2013). Diagnose e controle alternativo de doenças em tomate, pimentão, curcurbitáceas e cenoura. *Embrapa Hortaliças-Circular Técnica (INFOTECA-E)*.

Riqueti, G., Ribeiro, C., and Zárate, L. (2018). Classificando perfis de longevidade de bases de dados longitudinais usando floresta aleatória.

Sato, L. Y., Shimabukuro, Y. E., Kuplich, T. M., and Gomes, V. C. F. (2013). Análise comparativa de algoritmos de árvore de decisão do sistema weka para classificação do uso e cobertura da terra. *XVI Simpósio Brasileiro de Sensoriamento Remoto*, pages 2353–2360.

Silva, R. J. H., Pazoti, M. A., da Silva, F. A., Pereira, D. R., and de Almeida, L. L. (2019). Autenticação biométrica para sistemas por meio da dinâmica da digitação. In *Colloquium Exactarum. ISSN: 2178-8332*, volume 11, pages 26–33.

Viana, F., FREIRE, F., and Parente, G. (2007). Controle das principais doenças do pimentão cultivado nas regiões serranas do estado do ceará. *Embrapa Agroindústria Tropical-Comunicado Técnico (INFOTECA-E)*.