

Desenvolvimento de um Modelo de Detecção de Doenças em Soja com ML.NET: Análise de Performance

Vinicius Tessele
Faculdade Biopark
Toledo, Brasil

<https://orcid.org/0000-0002-0662-9200>

Marcos Bento de Oliveira Junior
Faculdade Biopark
Toledo, Brasil

<https://orcid.org/0009-0001-0843-8969>

Abstract— This study evaluates the application of ML.NET for detecting diseases in soybeans (*Glycine max*), a crop of significant economic importance, particularly in Brazil where agribusiness accounts for about 30% of GDP. Early identification of diseases is crucial for minimizing financial losses and ensuring global food security. The UN emphasized this importance by designating 2020 as the International Year of Plant Health, estimating that pests and diseases could reduce harvests by up to 40% annually. ML.NET, a machine learning library developed by Microsoft, facilitates the creation of custom models using C# and is compatible with various platforms, including web, mobile, and desktop. In this study, the library will be used to train a model for detecting soybean diseases through the ImageClassification API. The aim is to assess ML.NET's performance in classifying images of soybean leaves affected by diseases, using an external dataset not used in training and validation. The results will be compared with other disease detection methods to validate ML.NET's effectiveness. The choice of ML.NET is justified by its accessibility and ease of integration with different platforms, enabling the development of specialized applications. Implementing an efficient disease detection model for soybeans can reduce losses and contribute to the sustainability of agribusiness, ensuring greater global food security.

Keywords— Machine Learning; AI; Precision Agriculture.

Resumo— Este estudo avalia a aplicação do ML.NET na detecção de doenças na soja (*Glycine max*), um cultivo de alta importância econômica, especialmente no Brasil, onde o agronegócio representa cerca de 30% do PIB. Identificar doenças nas fases iniciais é crucial para minimizar perdas financeiras e garantir a segurança alimentar global. A ONU destacou essa importância ao designar 2020 como o Ano Internacional da Saúde Vegetal, estimando que pragas e doenças podem reduzir as safras em até 40% anualmente. O ML.NET, uma biblioteca de aprendizado de máquina da Microsoft, facilita a criação de modelos personalizados usando C# e é compatível com várias plataformas, incluindo web, móveis e desktop. Neste estudo, a biblioteca será empregada para treinar um modelo de detecção de doenças da soja por meio da API ImageClassification. O objetivo é avaliar o desempenho do ML.NET na classificação de imagens de folhas de soja afetadas por doenças, utilizando um conjunto de dados externo ao usado no treinamento e validação. Os resultados serão comparados com outros métodos de detecção de doenças para validar a eficácia do ML.NET. A escolha do ML.NET se deve à sua acessibilidade e facilidade de integração com diferentes plataformas, permitindo o desenvolvimento de aplicativos especializados. A implementação de um modelo eficiente de

detecção de doenças em soja pode reduzir perdas e contribuir para a sustentabilidade do agronegócio, assegurando maior segurança alimentar global.

Palavras-chave— Aprendizado de máquina; IA; Agricultura de precisão.

I. INTRODUÇÃO

O estudo propõe uma avaliação inicial da aplicação do ML.NET na detecção de doenças da soja (*Glycine max*). O controle de doenças na agricultura é crucial para evitar prejuízos financeiros e garantir a segurança alimentar global. No Brasil, o agronegócio representa aproximadamente 30% do PIB, com perspectivas de crescimento [1]. Identificar e controlar doenças nos estágios iniciais é fundamental para a produtividade e a qualidade das colheitas. A ONU designou 2020 como o Ano Internacional da Saúde Vegetal, destacando que pragas e doenças não controladas podem causar perdas de até 40% nas safras anualmente [2]. O ML.NET, uma biblioteca de código aberto e multiplataforma desenvolvida pela Microsoft, permite a criação de modelos de aprendizado de máquina utilizando C#. Sua integração com aplicativos web, móveis e desktop facilita o desenvolvimento de soluções que incorporam técnicas avançadas de aprendizado de máquina. Neste estudo, utilizaremos a API de classificação de imagem do ML.NET para treinar um modelo destinado à detecção de doenças em folhas de soja. O objetivo é avaliar a eficácia do ML.NET na classificação de imagens e comparar o desempenho do modelo treinado com outros métodos de detecção de doenças. A escolha do ML.NET se deve à sua acessibilidade e facilidade de integração, que permitem o desenvolvimento eficiente de aplicações para o setor agrícola. Considerando a importância econômica da soja para o agronegócio brasileiro e mundial, a implementação de um modelo eficiente de detecção de doenças contribuirá para a redução de perdas e garantirá maior sustentabilidade e segurança alimentar. Além disso, este trabalho está em desenvolvimento e tem como objetivo introduzir o aluno à iniciação científica, proporcionando uma experiência prática no aprendizado de máquina.

II. REFERENCIAL TEÓRICO

As Redes Neurais Artificiais (RNAs) são um ramo da Inteligência Artificial (IA) que desempenha um papel crucial

na tomada de decisões [3]. As RNAs se inspiram na forma como o cérebro humano executa tarefas específicas, buscando replicar esses processos por meio de sistemas computacionais [4]. Essas redes são capazes de coletar, utilizar e armazenar dados, utilizando um modelo de aprendizado baseado em algoritmos. Para isso, empregam funções de ativação que determinam se um neurônio deve ser ativado ou não, avaliando a relevância das informações fornecidas para um modelo específico [4].

Um tipo específico de RNA é a Rede Convolutiva, ou CNN (Convolutional Neural Network), uma arquitetura organizada em camadas, desenvolvida especificamente para o reconhecimento de padrões em imagens [5]. Diferentemente de outras redes neurais, as CNNs possuem três camadas internas principais de processamento: a camada de convolução, que aplica filtros às imagens para extrair características; a camada de Pooling, que reduz a dimensionalidade dos dados para minimizar a complexidade e o tempo de processamento; e a camada totalmente conectada, que conecta todos os neurônios da rede para realizar a classificação final (Figura 1).

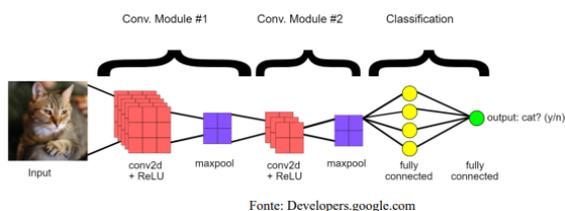


Fig. 1. Estrutura das camadas de uma Rede Convolutiva.

No contexto da IA, a visão computacional é utilizada para resolver problemas ao replicar processos cognitivos humanos, utilizando informações obtidas de imagens, como fotografias, vídeos ou outros sensores [6]. A resolução de problemas por meio da visão computacional segue uma sequência bem definida: captura de imagens, pré-processamento, segmentação, extração e normalização das características relevantes, seguida pela classificação e reconhecimento dos padrões identificados, até que se chegue à solução desejada [4]. Redes neurais, especialmente as redes convolucionais, é a técnica capaz de solucionar problemas de visão computacional, pois são capazes de aprender sobre as características a partir de dados de imagens. Elas permitiram avanços significativos em tarefas complexas, como o reconhecimento de imagens, classificação e a detecção de objetos [6].

III. MATERIAIS E MÉTODOS

Para desenvolver uma Rede Neural Artificial (RNA) capaz de detectar doenças em folhas de soja, utilizamos um conjunto de dados específico disponível no Kaggle, fornecido por Sivaram Mallia. Esse conjunto de dados contém um total de 609 imagens de folhas de soja doentes, classificadas em 10 categorias diferentes: Mossaic Virus (22 imagens), Southern

Blight (62 imagens), Sudden Death Syndrome (110 imagens), Yellow Mosaic (110 imagens), Bacterial Blight (50 imagens), Brown Spot (27 imagens), Crestamento (5 imagens), Ferrugem (65 imagens), Powdery Mildew (137 imagens), e Septoria (21 imagens) [7].

As categorias abrangem uma variedade de doenças que afetam a soja, e oferece uma base de dados de exemplos para treinar a RNA, porém não garantindo que o modelo seja robusto e preciso ao identificar cada uma das doenças presentes nas folhas.

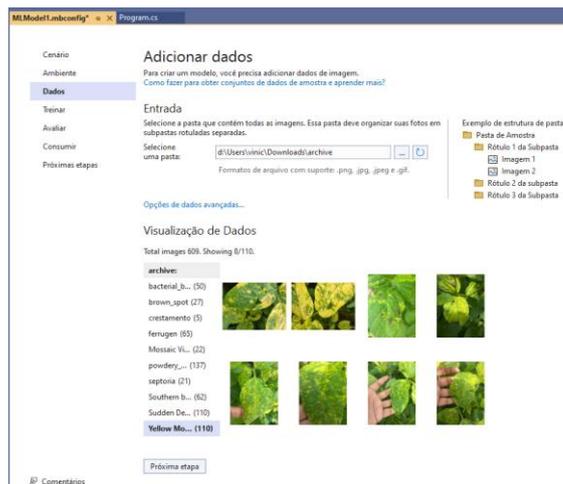


Fig. 2. Interface de entrada de dados.

Na Figura 2, é mostrado o processo de carregamento das imagens na aplicação. Após essa etapa, a opção de divisão 80/20 é selecionada, o que significa que 80% das imagens são alocadas para o treinamento do modelo, enquanto os 20% restantes são utilizados para a validação do desempenho, conforme ilustrado na Figura 3. Essa divisão é importante para garantir que o modelo aprenda de forma eficiente a partir do conjunto de treinamento e seja avaliado adequadamente em dados que não foram vistos anteriormente.

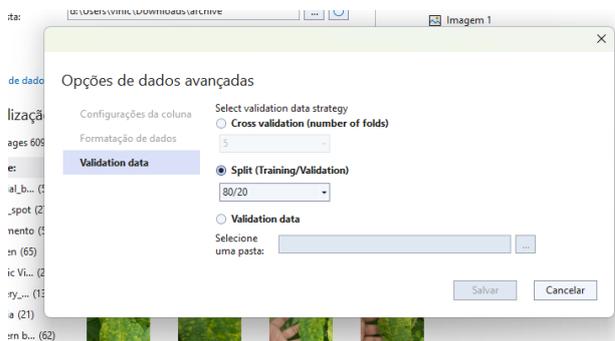


Fig. 3. Configuração 80% treinamento 20% Validação.

Em seguida, foi iniciado o treinamento da RNA. Na Figura 4, os resultados demonstram que o modelo atingiu uma acurácia de 96%, o que mostra um excelente desempenho. Para validar ainda mais o modelo, selecionamos imagens

adicionais, obtidas na internet, que não faziam parte dos conjuntos de treinamento e validação originais. Essas imagens foram usadas para testar o desempenho do modelo em situações reais e verificar se ele mantém sua capacidade de identificar corretamente as doenças, para confirmar se o modelo apresenta bons resultados em diferentes condições e fontes de dados.

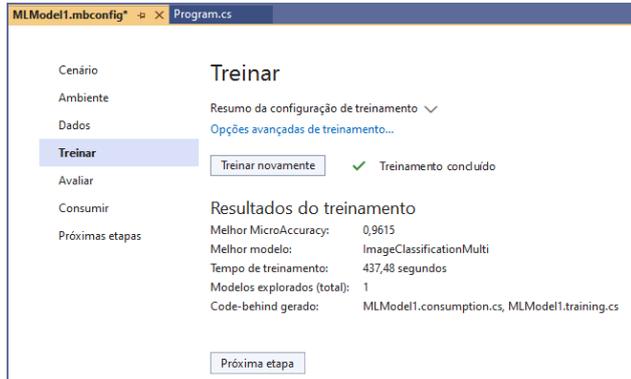


Fig. 4. Resultado do treinamento com a acurácia de 96%.

Na Figura 5, para a primeira avaliação, foi escolhida a doença Síndrome da Morte Súbita (Sudden Death Syndrome - SDS). O modelo conseguiu identificar corretamente a doença, alcançando uma taxa de sucesso de 89%.

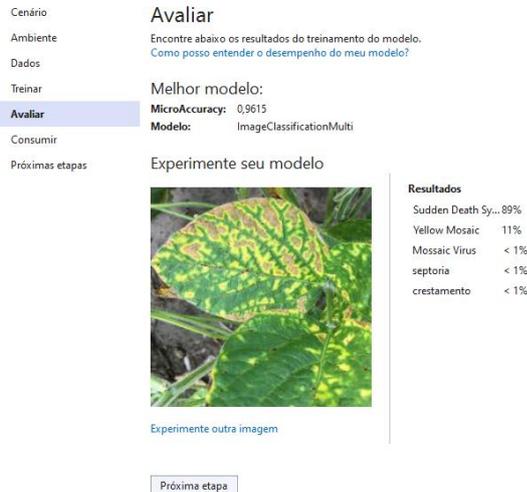


Fig. 5. Sudden Death Syndrome identificada com 89%

Na segunda avaliação, foi selecionada a doença Septoria, causada pelo fungo *Septoria glycines* que afeta as folhas de soja. O modelo também conseguiu identificar a presença da doença, mas com uma taxa de acerto de 75%, conforme ilustrado na Figura 5. Embora o desempenho seja aceitável, este resultado indica que há espaço para melhorias na detecção de Septoria, possivelmente ajustando parâmetros do modelo ou aumentando a quantidade de dados específicos para essa doença.

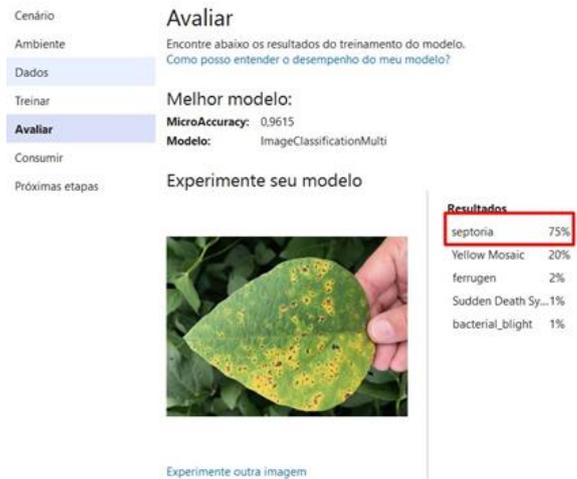


Fig. 6. Septoria identificada com 75%

Essa diferença de desempenho pode ser atribuída similaridade dos sintomas visuais entre as duas doenças. No caso da Síndrome da Morte Súbita (SDS), os sintomas, como manchas bem definidas nas folhas, facilitam a identificação por um modelo de aprendizado de máquina. Ao contrário, os sintomas da Septoriose, caracterizados por manchas marrons mais sutis e menos definidas, exibem maior variabilidade visual, o que pode tornar sua detecção mais difícil para o modelo. Esse fator pode exigir um conjunto de dados mais diversificado ou técnicas de pré-processamento sejam aprimoradas para melhorar a precisão na identificação de doenças cuja as características sejam menos evidentes.

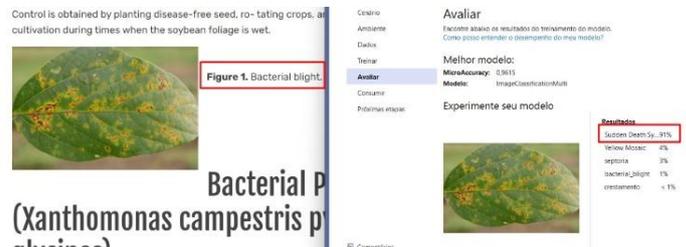


Fig. 7. Bacterial Blight, identificada como Sudden Death Syndrome 91%

Na terceira avaliação, mostrada na Figura 7, foi selecionada a Bacterial Blight, uma doença causada por bactérias que provoca lesões nas folhas cuja a aparência é de "queima". O modelo de RNA treinado apresentou um desempenho insatisfatório, identificando erroneamente a Bacterial Blight como SDS com 91% de precisão. Esse erro pode ser explicado pelo fato de que os sintomas da Bacterial Blight podem ser visualmente similares aos de outras doenças, como SDS e Septoriose.

Na quarta avaliação, mostrada na Figura 8, foi selecionada a Powdery Mildew, que é uma doença fúngica que afeta diversas plantas e sua característica é da presença de um pó branco ou cinza que cobre as filhas, caules e flores das plantas [8].

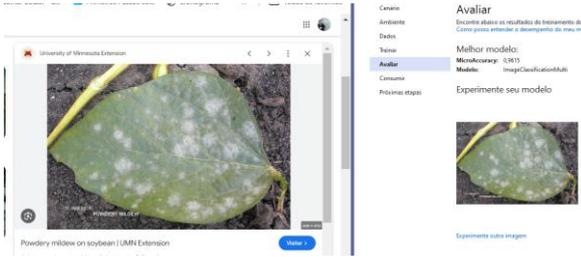


Fig. 8. Powdery mildew, identificada como Sudden Death Syndrome 75%

A semelhança nas características visuais entre essas doenças pode causar confusão no modelo, destacando a necessidade de ajustar o treinamento ou incorporar mais dados específicos para melhorar a diferenciação entre condições com sintomas semelhantes.

IV. CONCLUSÃO

Quando diferentes doenças compartilham características visuais semelhantes, o modelo de aprendizado de máquina pode enfrentar dificuldades na distinção precisa entre elas, resultando em classificações incorretas. Além disso, a variação na acurácia do modelo pode refletir a qualidade e a quantidade dos dados de treinamento disponíveis para cada tipo de doença. Dados insuficientes ou de baixa qualidade podem comprometer a capacidade do modelo de aprender e generalizar de forma eficaz, prejudicando sua capacidade de diferenciar doenças com sintomas semelhantes. Portanto, é essencial garantir que o conjunto de dados seja abrangente e representativo para aprimorar a precisão do modelo e sua capacidade de diagnóstico.

REFERÊNCIAS

- [1] CEPEA – Centro de estudos avançados em economia aplicada. PIB do agronegócio 2021. Disponível em: <https://cepea.esalq.usp.br/br/opiniao-cepea/afinal-quanto-o-agronegocio-representa-no-pib-brasileiro.aspx>. Acesso em: 10 jun. 2024.
- [2] FAO. International Year of Plant Health 2020 | FAO | Food and Agriculture Organization of the United Nations. Disponível em: <https://www.fao.org/plant-health-2020/about/en/>. Acesso em: 10 jun. 2024.
- [3] RODRIGUES, Pedro Asseiro. Classificação de doenças nas plantas utilizando Transfer Learning numa aplicação móvel. 2020. 75 f. Dissertação (Mestrado) - Curso de Sistemas de Informação, Escola Superior de Tecnologia e de Gestão de Bragança, Instituto Politécnico de Bragança, Bragança, 2020. Disponível em: <http://hdl.handle.net/10198/23219>. Acesso em: 10 jun. 2024.

- [4] CARRASCOSA, Alexandre de Freitas. Aplicação de métodos de processamento de imagem e redes neurais na extração de características de doenças em plantas. 2021. 69 f. TCC (Graduação) - Curso de Engenharia Agrícola e Ambiental, Instituto de Ciências Agrárias e Tecnológicas, Universidade Federal de Rondonópolis, Rondonópolis, Mt, 2021. Disponível em: <https://bdm.ufmt.br/handle/1/1798>. Acesso em: 10 jun. 2024.
- [5] ROSA, Matheus Cassali da. Redes neurais convolutivas aplicadas à detecção de ervas daninhas. 2019. 78 f. Dissertação (Mestrado) - Curso de Engenharia Elétrica, Programa de Pós-Graduação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2019. Disponível em: <http://hdl.handle.net/10183/204423>. Acesso em: 10 jun. 2024.
- [6] NEUMANN, Bruno Germano. Identificação de doenças na soja utilizando inteligência artificial por meio de análise de imagens. 2019. 114 f. TCC (Graduação) - Curso de Sistemas de Informação, Centro de Ciências Exatas e Tecnológicas, Universidade do Vale do Taquari, Lajeado, 2019. Disponível em: <http://hdl.handle.net/10737/2543>. Acesso em: 29 jun. 2024.
- [7] MALLIA, S. Soybean Diseased Leaf Dataset. Kaggle. Disponível em: <https://www.kaggle.com/datasets/sivm205/soybean-diseased-leaf-dataset>. Acesso em: 29 jun. 2024.
- [8] HENNING, Ademir Assis; EMBRAPA SOJA. Manual de identificação de doenças de soja. 5. ed. Londrina: Embrapa Soja, 2014. 76 p.: il. color. (Documentos / Embrapa Soja, ISSN 1516-781X; n. 256). Disponível em: <https://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/1158639>. Acesso em: 29 jun. 2024.