Projeto de Pesquisa

Iniciação Científica

Redes neurais convolucionais e *transformers* visuais aplicados à detecção de doenças em imagens de folhas de plantas

Estudante

Heitor Barroso Cavalcante

Bacharelado em Ciência da Computação
Instituto de Matemática e Estatística
Universidade de São Paulo

Orientadora

Nina S. T. Hirata

Departamento de Ciência da Computação
Instituto de Matemática e Estatística
Universidade de São Paulo

Resumo

Técnicas baseadas em *deep learning* constituem atualmente o estado da arte quando se trata de processamento de diversos tipos de dados como imagens, áudio e texto. Na área de Visão Computacional, as redes neurais convolucionais e, mais recentemente, os *transformers* visuais promoveram grandes avanços de métodos que realizam extração de informação de imagens. Este projeto de pesquisa tem como objetivo estudar aspectos teóricos e práticos desses dois modelos. Enquanto o estudo teórico deverá seguir uma abordagem tradicional, o estudo prático será realizado por meio do treinamento e avaliação das redes sobre o problema de detecção de doenças em imagens de folhas de plantas. O exercício prático proporcionará oportunidades para uma avaliação comparativa dos dois tipos de redes, incluindo tanto aspectos quantitativos quanto qualitativos. Espera-se que o conhecimento gerado neste projeto possa ser posteriormente utilizado para abordar um escopo mais amplo do problema de detecção de doenças em plantas.

São Paulo, 6 de março de 2023

1 Introdução

Este projeto de pesquisa foi motivado pelo interesse do estudante em entender como doenças em plantas são detectadas utilizando-se técnicas de processamento de imagens e, em particular, as abordagens baseadas em inteligência artificial.

De acordo com a Organização das Nações Unidas para a Alimentação e a Agricultura (FAO), todo ano, entre 20 e 40% da produção agrícola é perdida em função de pragas (IPPC Secretariat, 2021). Levando tal fato em consideração, é evidente a importância de se diagnosticar doenças em plantas de maneiras que sejam tanto eficientes quanto baratas, visando a minimização dos danos oriundos das pragas na agricultura mundial.

No cenário nacional, recentemente a Embrapa[1] publicou uma matéria intitulada "Inteligência artificial identifica plantas doentes simulando processo cerebral" na qual é discutida uma parceria estabelecida por ela com empresas privadas para desenvolver inovações na detecção automática de doenças em plantas.

A detecção e diagnóstico de doenças em plantas pode ser feita por inspeção visual, avaliação microscópica de características morfológicas, ou ainda técnicas moleculares, sorológicas ou microbiológicas (Mahlein, 2016). Todas essas formas de análise são baseadas em tecnologias ou conhecimentos específicos, sendo portanto altamente dependentes de pessoal qualificado e especialista. Esse processo, além de ser feito em escala reduzida, é custoso e complexo, o que demonstra a necessidade de inovações tecnológicas que auxiliem nesta tarefa e a tornem mais simples e acessível para aqueles que desenvolvem atividades agrícolas.

O artigo de Mahlein (2016) discute o uso de técnicas de imageamento em agricultura e fenotipagem de plantas, destacando a questão de detecção de doenças. O trabalho lista sensores convencionais que geram imagem RGB, câmeras multi ou hiper-espectrais, sensores termais, sensores de fluorescência, e outros sensores capazes de medir biomassa e outras características estruturais das plantas. Dentre essas tecnologias, câmeras convencionais que geram imagem RGB estão amplamente disponíveis, são portáteis e de fácil manuseio. Assim, a inspeção visual poderia ser realizada com o auxílio de processamento computacional dessas imagens. Por outro lado, Bock et al. (2022) aponta que apesar de o uso de algoritmos para detecção de doenças ter sido iniciado na década de 1980, são quase inexistentes produtos derivados de tais estudos. No cenário atual, aplicativos de celular com capacidade de identificar e quantificar doenças podem ser um grande aliado na melhoria da produção agrícola.

[1] https://www.embrapa.br/

Com os recentes avanços na área de Visão Computacional devido ao contínuo melhoramento das técnicas de *deep learning* (Goodfellow et al., 2016), começaram a emergir trabalhos sobre detecção de doenças em imagens de plantas que fazem uso de modelos de redes neurais profundas. Entre esses trabalhos, a maioria considera imagens de folhas de plantas obtidas utilizando-se câmeras convencionais e trata do problema de classificação de doenças (Singh and Misra, 2017; Mohanty et al., 2016; Geetharamani and Arun Pandian, 2019; Thakur et al., 2022). Um conjunto de imagens público bastante usado para o treinamento e avaliação de modelos é o *PlantVillage dataset* (Hughes and Salathé , 2015). Exemplos de imagens desse dataset são mostrados na figura 1.

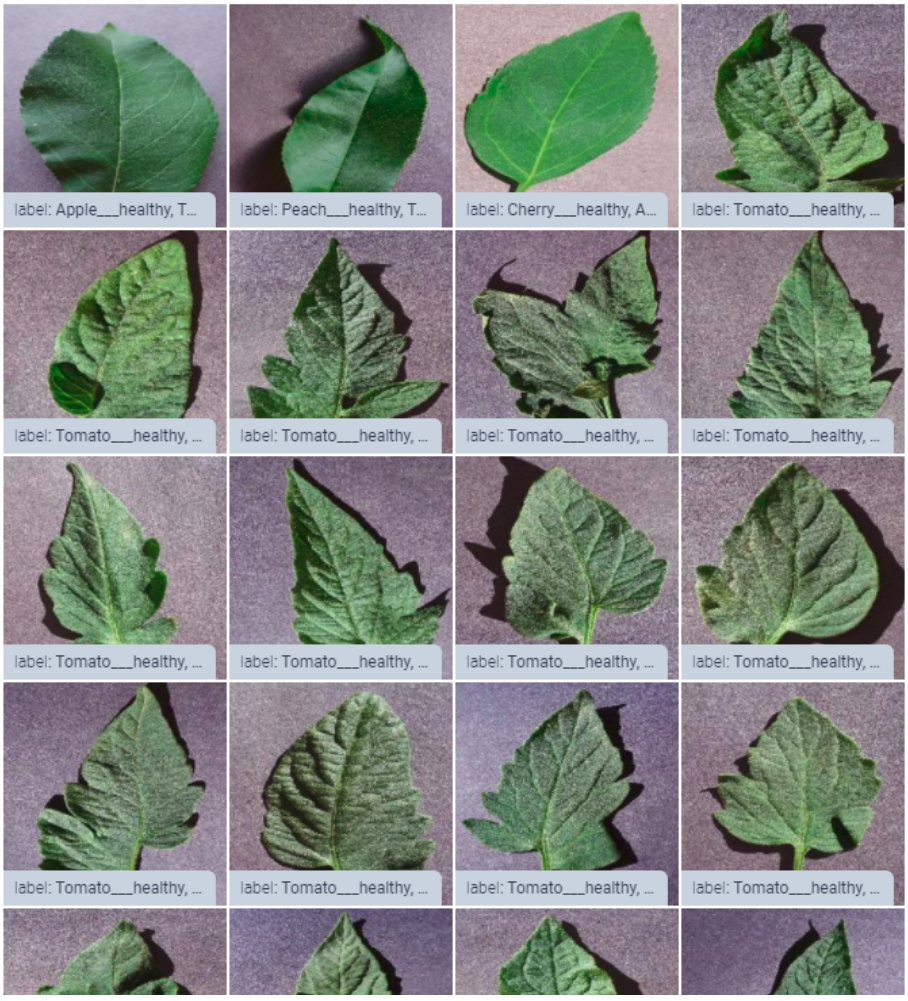

Figura 1: Exemplos de imagens de folhas no dataset *PlantVillage* (extraído de https://paperswithcode.com/dataset/plantvillage).

Os métodos baseados em aprendizado de máquina utilizados inicialmente seguiam uma abordagem tradicional (Singh and Misra, 2017). Tipicamente, abordagens tradicionais incluem uma etapa para a segmentação da região de interesse (folhas, por exemplo), outra para a extração de características (por exemplo cores ou textura), e, por fim, o uso dessas características juntamente com um algoritmo de aprendizado de máquina (tais como SVM ou árvores de decisão) para classificar

a folha quanto a presença ou não de doenças. Ou seja, trata-se de uma abordagem que depende bastante de engenharia manual, tanto para a segmentação das regiões de interesse como para a extração de características.

Em contraste, os trabalhos mais recentes são predominantemente baseados em técnicas de *deep learning* (Mohanty et al., 2016; Geetharamani and Arun Pandian, 2019; Thakur et al., 2022). As técnicas de *deep learning* são baseadas em redes neurais (Nielsen, 2015; Goodfellow et al., 2016) que possuem a capacidade de processar dados em seu formato bruto e extrair automaticamente as representações (características) que são mais eficazes para a inferência final esperada. Essa capacidade, resultante do processo de treinamento da rede neural, faz com que os métodos sejam facilmente adaptados para imagens com diferentes características. Assim, o desenvolvimento de soluções para um novo cenário de aplicação é bastante agilizado.

Muitas das técnicas de *deep learning* empregadas em processamento de imagens são baseadas em Redes Neurais Convolucionais (CNNs, Goodfellow et al. (2016)). Mais recentemente, arquiteturas de redes neurais baseadas no conceito de atenção (Bahdanau et al., 2015), os chamados *transformers* (Vaswani et al., 2017), vem se destacando. Inicialmente desenvolvidos para processamento de textos na área de linguagem natural, os *transformers* logo foram estendidos para aplicações na área de Visão Computacional (Dosovitskiy et al., 2021; Liu et al., 2021), sendo denominados *Visual Transformers* (ViT).

De fato, essas arquiteturas passaram a ser utilizadas recentemente também em problemas de detecção de doenças em plantas. Por exemplo, o trabalho de Geetharamani and Arun Pandian (2019) descreve o uso de uma CNN de 9 camadas que alcançou uma acurácia de 96,46% para classificar doenças presentes no dataset *PlantVillage*. Paralelamente, Thakur et al. (2022) reportam uma acurácia de 98,86% nesse mesmo dataset, sugerindo o sucesso de *Transformers* Visuais quando aliados à blocos de CNN.

2 Objetivos

Conforme exposto na seção anterior, o reconhecimento de doenças em plantas é um problema de grande relevância. Os métodos de classificação baseados em técnicas de *deep learning* e aplicados em imagens de folhas de plantas têm apresentado bons resultados, indicando o potencial dessas técnicas.

O principal objetivo deste projeto de iniciação científica consiste em entender como funcionam as redes neurais do tipo CNN e *Transformers* visuais, assim como

camadas convolucionais combinadas com *transformers*, e como essas redes se comportam e em quais pontos diferem quanto a essa tarefa de classificação de doenças em folhas de plantas.

Os objetivos específicos estão listados a seguir:

- Estudo dos fundamentos teóricos de CNNs e *Transformers* Visuais.
- Implementação, usando bibliotecas de programação da linguagem Python, de uma Rede Convolucional, uma Rede *Transformer* e uma Rede híbrida (*transformer* com camadas iniciais de convolução).
- Treinamento e validação dessas redes para o problema de classificação de doenças em plantas, usando imagens do dataset *PlantVillage* e eventualmente outros.
- Teste dos modelos treinados e análise de performance, comparando aspectos quantitativos e qualitativos.

3 Material e métodos

Seguindo a linha do exposto na seção anterior, é possível agrupar as atividades que serão desenvolvidas ao longo desse projeto em dois grupos, que se complementam e estarão presentes quase que simultaneamente ao longo de todo o processo a ser desenvolvido. O primeiro deles refere-se ao estudo dos fundamentos, ou seja, dos conceitos teóricos que fundamentam as atividades práticas a serem realizadas. Já o segundo, cobre a implementação e experimentação com os modelos em questão — CNN, ViT e CNN+ViT.

3.1 Estudo de fundamentos

O problema a ser tratado neste projeto é um típico problema de Visão Computacional. Desta forma, em termos de fundamentos teóricos é desejável que o estudante possua conhecimentos sobre processamento de imagens (Gonzalez and Woods, 2002) e aprendizado de máquina (Abu-Mostafa et al., 2012). Para estudar e implementar CNNs e Transformers Visuais, é necessário também um conhecimento sólido de redes neurais (Nielsen, 2015).

Neste sentido, é importante ressaltar que, no ano de 2022 o estudante foi bolsista do Programa de Iniciação Científica e Mestrado (PICME), promovido pelo CNPq em conjunto com a CAPES. Durante o programa, sob a orientação da orientadora deste projeto, o estudante foi exposto às noções básicas de imagens digitais

e de processamento de imagens. Além disso, dedicou o segundo semestre ao estudo de fundamentos de aprendizado de máquina, cobrindo Regressão Linear, Regressão Logística e Redes Neurais. O estudo realizado foi documentado na forma de um relatório técnico intitulado "Fundamentos de Redes Neurais", que encontra-se anexo à esta proposta (no sistema SAGe). Esse estudo foi acompanhado de atividades de implementação e teste de código, que podem ser acessados através de links presentes no relatório em questão. Adicionalmente, o estudante cursou uma disciplina sobre Processamento de Sinais Digitais.

Desta forma, o estudante já reúne conhecimentos suficientes necessários para o estudo de CNNs e *Transformers*. Portanto, em termos de fundamentos, o foco do estudo estará nesses dois modelos de redes neurais.

3.1.1 Redes neurais convolucionais

Redes neurais convolucionais (CNNs) são redes neurais que possuem camadas formadas por nós (unidades de processamento) que implementam os filtros de convolução (Gonzalez and Woods, 2002). Esses filtros são parametrizados por kernels, utilizados para realizar processamentos locais ao longo de toda a extensão da imagem. Os kernels, implementados como matrizes de pesos, tem os valores estabelecidos no processo de treinamento da rede. Isto significa que os filtros são "aprendidos" pela rede durante o seu treinamento de forma otimizada para a tarefa-alvo do treinamento.

Alguns possíveis materiais para o estudo de CNNs estão listados a seguir, porém outros deverão ser utilizados à medida que o estudante for ganhando familiarização com o tópico.

- CS231n: Deep Learning for Computer Vision, material disponibilizado pela Universidade de Stanford
- Artigo *Convolutional Networks and Applications in Vision* (LeCun et al., 2010)
- Livro *Deep Learning* (Goodfellow et al., 2016)
- Artigos sobre modelos CNN para classificação de imagem bem conhecidos (Krizhevsky et al., 2012; Simonyan and Zisserman, 2014; Szegedy et al., 2015; He et al., 2016)

3.1.2 *Transformers* e *Transformers* Visuais

O conceito de atenção é um elemento básico nos modelos do tipo *transformers*. Assim, o estudo deverá cobrir inicialmente as redes neurais recorrentes (RNN) e, em

especial, a arquitetura *encoder-decoder* de RNNs pois os mecanismos de atenção em rede neural tornaram-se populares no contexto do emprego desse tipo de rede em problemas de tradução de textos em linguagem natural (Bahdanau et al., 2015). Em seguida, serão estudados os seguintes artigos e outros que se mostrarem adequados:

- *Attention Is All You Need* (Vaswani et al., 2017): este é o artigo que propôs a construção de redes neurais baseadas em auto-atenção para o problema de tradução de textos em linguagem natural, eliminando totalmente a utilização de recorrência, e impulsionou o desenvolvimento de variantes que adotam princípios similares.

- *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale* (Dosovitskiy et al., 2021): este artigo propõe o particionamento da imagem em blocos de 16×16 pixels e o arranjo sequencial dos segmentos, de forma a possibilitar o emprego direto de *transformers* para o processamento.

- *Swin Transformer: Hierarchical Vision Transformer using Shifted Windows* (Liu et al., 2021): este artigo propõe um esquema de particionamento hierárquico da imagem, integrada de forma eficiente em uma arquitetura do tipo *transformer*. Com isso, argumenta-se que essa rede consegue lidar melhor com objetos de diferentes escalas na imagem.

3.2 Parte experimental

A parte experimental deste projeto consiste em implementar, testar e avaliar os três tipos de modelos, a saber CNNs, ViTs, e um híbrido CNN+ViT. As implementações e testes deverão ocorrer em paralelo aos estudos de fundamentos, principalmente na parte inicial em que deverão ser realizadas tentativas de reprodução de resultados conhecidos sobre datasets padrões com as arquiteturas sendo estudadas. Posteriormente as redes serão testadas sobre o problema de classificação de imagens de folhas de plantas. Especificamente, pretendemos utilizar a base de dados pública *PlantVillage* (Hughes and Salathé , 2015).

Os experimentos assim como as avaliações de desempenho a serem realizados com esse dataset serão planejados usando como referência os experimentos descritos nos seguintes artigos (e eventualmente outros que venhamos a encontrar):

- *Identification of plant leaf diseases using a nine-layer deep convolutional neural network* (Geetharamani and Arun Pandian, 2019)

- *A deep learning based approach for automated plant disease classification using vision transformer* (Borhani et al., 2022)

- *Explainable vision transformer enable concolutional neural network for plant disease identification: PlantXViT* (Thakur et al., 2022)

O desempenho das redes devidamente treinadas serão comparados em termos quantitativos (acurácia, *recall* e *precision* de classificação, tempo de inferência, quantidade de memória utilizada) e qualitativos (principais tipos de erros, pontos de atenção, etc).

Os experimentos possibilitarão a prática de treinamento e avaliação dos modelos neurais. Haverá oportunidades para se estudar e trabalhar as estratégias e boas práticas experimentais comumente utilizadas, das quais destacamos algumas a seguir:

- Divisão dos dados em treinamento, validação e teste
- *Transfer learning* + *fine tuning*
- Tratamento de desbalanceamento de classes
- *Data augmentation*
- Otimização de hiperparâmetros e seleção de uma configuração ótima
- Detecção de *overfitting* e formas de combatê-lo
- Métricas de desempenho

3.3 Outras informações

Destacamos que desde meados do ano de 2022 o estudante é participante ativo de um grupo de estudos coordenado pela orientadora. Esse grupo tem como objetivo abordar, de maneira teórica e prática, importantes conceitos de *Machine Learning*, visando a formação sólida dos estudantes nessa área. Discussões e práticas relacionadas às CNNs já estão acontecendo e as relacionadas aos *transformers* também já estão começando a acontecer dentro do grupo. É esperado que com a execução do projeto, o estudante ajude a aprofundar as discussões sobre esses tópicos, tanto do ponto de vista teórico como prático. As reuniões do grupo servirão também para acompanhar o desenvolvimento e progresso deste projeto de pesquisa.

Acreditamos que o estudo de fundamentos e a experimentação prática conforme descritos acima permitirão uma formação sólida do estudante quanto aos tipos de redes estudadas. Ademais, a possibilidade de troca de conhecimentos com os demais membros do grupo facilitará o processo de estudo.

Além disso, este projeto de pesquisa visa também contemplar o exercício de práticas relacionadas à metodologia científica tais como investigação de literatura científica, planejamento de experimentos, comparação de métodos, análise de resultados e escrita científica.

Para a realização dos experimentos computacionais, serão utilizados os equipamentos do Laboratório de Visão Computacional do IME/USP.

Os códigos desenvolvidos serão disponibilizados publicamente via github.

4 Plano de trabalho e cronograma de execução

As atividades a serem desenvolvidas estão agrupadas em cinco itens, conforme descritos a seguir.

- **Estudo de fundamentos:** cobrirá os conceitos que fundamentam os três tipos de redes a serem empregados (CNN, ViT, CNN+ViT) e aqueles relacionados ao treinamento/validação e avaliação de desempenho das redes. Esta será uma atividade a ser realizada durante toda a duração do projeto, com ênfase nos primeiros 6 meses.

- **Trabalhos relacionados:** cobrirá o estudo de trabalhos que tratam de problemas de classificação de imagens, e especificamente de detecção ou classificação de doenças em imagens de plantas. Esta atividade inclui também a investigação de literatura da área, visando complementar o levantamento realizado até este momento.

- **Implementação e treinamento de algoritmos:** esta atividade envolve uma parte inicial de familiarização com códigos e com o treinamento das redes de interesse que deverão ser realizadas de forma concomitante com os estudos teóricos, e uma segunda parte em que as redes serão treinadas e aplicadas ao problema alvo deste projeto. Os trabalhos relacionados estudados servirão como base nesta segunda parte para planejar os experimentos a serem executados.

- **Análise de resultados:** nesta atividade, além de comparações quantitativas, pretendemos também fazer análises qualitativas (por exemplo, analisar os principais tipos de erros, os pontos de atenção de cada tipo de rede, entre outras)

- **Preparação do relatório final:** Como parte das atividades estão previstas o treinamento de escrita científica. Este será por meio da elaboração de re-

latório científico final. Eventualmente, dependendo dos resultados alcançados, poderá também ser elaborado um artigo científico a ser submetido a um fórum adequado. Planejamos também participar de eventos voltados para alunos de graduação tais como o SIICUSP[2] ou *workshops* que acontecem junto a eventos científicos da área.

Um cronograma aproximado para a execução dessas atividades está apresentado a seguir.

	Trimestre			
Atividade	1º	2º	3º	4º
Estudo aprofundado de fundamentos	x	x	x	x
Estudo de trabalhos relacionados	x	x	x	
Implementação e treinamento dos algoritmos	x	x	x	
Análise dos resultados obtidos			x	x
Preparação do relatório final			x	x

5 Forma de análise dos resultados

Os resultados da pesquisa serão avaliados em termos dos seguintes produtos ou atuações resultantes como decorrência direta da execução do projeto de pesquisa:

- Relatório técnico descrevendo os fundamentos de redes neurais convolucionais e *transformers* visuais, a exemplo do texto gerado em 2022 pelo estudante sobre fundamentos de redes neurais.
- Disponibilização pública dos códigos desenvolvidos.
- Submissão ou publicação do trabalho em eventos ou outros veículos pertinentes.
- Apresentação dos resultados em congressos e eventos científicos locais e regionais.
- Divulgação dos conhecimentos adquiridos através de atividades no ambiente universitário, como atividades em grupos de extensão e palestras.

Referências

Abu-Mostafa, Y. S., Lin, H.-T., and Magdon-Ismail, M. (2012). *Learning From Data*. AMLBook.

[2] http://siicusp.prp.usp.br

Bahdanau, D., Cho, K., and Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. In *ICLR*.

Bock, C. H., Chiang, K.-S., and Ponte, E. M. D. (2022). Plant disease severity estimated visually: a century of research, best practices, and opportunities for improving methods and practices to maximize accuracy. *Tropical Plant Pathology*, 47:25–42.

Borhani, Y., Khoramdel, J., and Najafi, E. (2022). A deep learning based approach for automated plant disease classification using vision transformer. *Scientific Reports*, 12(1):11554.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., and Houlsby, N. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. In *ICLR*. OpenReview.net.

Geetharamani, G. and Arun Pandian, J. (2019). Identification of plant leaf diseases using a nine-layer deep convolutional neural network. *Computers & Electrical Engineering*, 76:323–338.

Gonzalez, R. C. and Woods, R. E. (2002). *Digital Image Processing*. Addison-Wesley Publishing Company, second edition.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press. http://www.deeplearningbook.org.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778.

Hughes, D. P. and Salathé , M. (2015). An open access repository of images on plant health to enable the development of mobile disease diagnostics through machine learning and crowdsourcing. *CoRR*, abs/1511.08060.

IPPC Secretariat (2021). *Scientific review of the impact of climate change on plant pests – A global challenge to prevent and mitigate plant pest risks in agriculture, forestry and ecosystems*. FAO on behalf of the IPPC Secretariat, Rome.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Pereira, F., Burges, C., Bottou, L., and Weinberger, K., editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc.

LeCun, Y., Kavukcuoglu, K., and Farabet, C. (2010). Convolutional networks and applications in vision. In *Proceedings of 2010 IEEE International Symposium on Circuits and Systems*, pages 253–256.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., and Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9992–10002, Los Alamitos, CA, USA. IEEE Computer Society.

Mahlein, A.-K. (2016). Plant disease detection by imaging sensors – parallels and specific demands for precision agriculture and plant phenotyping. *Plant Disease*, 100(2):241–251. PMID: 30694129.

Mohanty, S. P., Hughes, D. P., and Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Frontiers in Plant Science*, 7.

Nielsen, M. A. (2015). *Neural Networks and Deep Learning*. Determination Press.

Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556.

Singh, V. and Misra, A. (2017). Detection of plant leaf diseases using image segmentation and soft computing techniques. *Information Processing in Agriculture*, 4(1):41–49.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9.

Thakur, P. S., Khanna, P., Sheorey, T., and Ojha, A. (2022). Explainable vision transformer enabled convolutional neural network for plant disease identification: Plantxvit.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 6000–6010, Red Hook, NY, USA. Curran Associates Inc.