



UNIVERSIDADE FEDERAL DE GOIÁS (UFG)
ESCOLA DE ENGENHARIA ELÉTRICA, MECÂNICA E DE
COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA
ELÉTRICA E DE COMPUTAÇÃO

Vinícios Matheus Oliveira da Silva

HíbridNet: Rede Neural Convolutiva (CNN) Híbrida para Classificação de Doenças em Folhas de Bananeira

Goiânia
2025



UNIVERSIDADE FEDERAL DE GOIÁS
ESCOLA DE ENGENHARIA ELÉTRICA, MECÂNICA E DE COMPUTAÇÃO

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO (TECA) PARA DISPONIBILIZAR VERSÕES ELETRÔNICAS DE TESES

E DISSERTAÇÕES NA BIBLIOTECA DIGITAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a [Lei 9.610/98](#), o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou download, a título de divulgação da produção científica brasileira, a partir desta data.

O conteúdo das Teses e Dissertações disponibilizado na BDTD/UFG é de responsabilidade exclusiva do autor. Ao encaminhar o produto final, o autor(a) e o(a) orientador(a) firmam o compromisso de que o trabalho não contém nenhuma violação de quaisquer direitos autorais ou outro direito de terceiros.

1. Identificação do material bibliográfico

Dissertação Tese Outro*: _____

*No caso de mestrado/doutorado profissional, indique o formato do Trabalho de Conclusão de Curso, permitido no documento de área, correspondente ao programa de pós-graduação, orientado pela legislação vigente da CAPES.

Exemplos: Estudo de caso ou Revisão sistemática ou outros formatos.

2. Nome completo do autor

Vinicios Matheus Oliveira da Silva

3. Título do trabalho

HibridNet: Rede Neural Convolutacional (CNN) Híbrida para Classificação de Doenças em Folhas de Bananeira

4. Informações de acesso ao documento (este campo deve ser preenchido pelo orientador)

Concorda com a liberação total do documento SIM NÃO¹

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. Após esse período, a possível disponibilização ocorrerá apenas mediante:

a) consulta ao(à) autor(a) e ao(à) orientador(a);

b) novo Termo de Ciência e de Autorização (TECA) assinado e inserido no arquivo da tese ou dissertação. O documento não será disponibilizado durante o período de embargo.

Casos de embargo:

- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro;
- Publicação da dissertação/tese em livro.

Obs. Este termo deverá ser assinado no SEI pelo orientador e pelo autor.



Documento assinado eletronicamente por **Vinícios Matheus Oliveira Da Silva, Discente**, em 08/07/2025, às 11:34, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Gelson Da Cruz Junior, Professor do Magistério Superior**, em 08/07/2025, às 16:30, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **5489422** e o código CRC **0795C528**.

Vinicios Matheus Oliveira da Silva

HibridNet: Rede Neural Convolutacional (CNN) Híbrida para Classificação de Doenças em Folhas de Bananeira

Dissertação apresentada ao Programa de Pós-Graduação *Stricto Sensu* em Engenharia Elétrica e de Computação, da Escola de Engenharia Elétrica, Mecânica e de Computação, da Universidade Federal de Goiás como requisito parcial para a obtenção do Título de Mestre em Engenharia Elétrica e de Computação.

Área de Concentração: Engenharia de Computação.

Orientador: Prof. Dr. Gelson da Cruz Junior

Coorientador: Prof. Dr. Tiago do Carmo Nogueira

Goiânia

2025

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UFG.

Silva, Vinícios Matheus Oliveira da
HíbridNet: Rede Neural Convolutiva (CNN) Híbrida para Classificação de Doenças em Folhas de Bananeira [manuscrito] / Vinícios Matheus Oliveira da Silva. - 2025.
LXXIX, 79 f.: il.

Orientador: Prof. Dr. Gélson da Cruz Junior; co-orientador Dr. Tiago do Carmo Nogueira.
Dissertação (Mestrado) - Universidade Federal de Goiás, Escola de Engenharia Elétrica, Mecânica e de Computação (EMC), Programa de Pós-Graduação em Engenharia Elétrica e de Computação, Goiânia, 2025.

Inclui siglas, fotografias, gráfico, tabelas, lista de figuras, lista de tabelas.

1. Folhas de Bananeiras. 2. Doenças foliares. 3. Redes Neurais Convolutivas (CNN). 4. Vision Transformer (ViT). 5. Aprendizado Profundo. I. Junior, Gélson da Cruz, orient. II. Título.

CDU 004



UNIVERSIDADE FEDERAL DE GOIÁS
ESCOLA DE ENGENHARIA ELÉTRICA, MECÂNICA E DE COMPUTAÇÃO

ATA EMC 14/2025

Ata nº 14 da sessão de Defesa de Dissertação de **VINICIOS MATHEUS OLIVEIRA DA SILVA**, que confere o título de Mestre em **Engenharia Elétrica e de Computação**, na área de concentração em **Engenharia de Computação**.

Aos **dezessete dias do mês de junho de dois mil e vinte e cinco**, a partir das **13:30h**, realizou-se a sessão pública de Defesa de Dissertação intitulada “**HíbridNet: Rede Neural Convolutacional (CNN) Híbrida para Classificação de Doenças em Folhas de Bananeira**”. Os trabalhos foram instalados pelo Orientador, Professor Doutor **Gelson da Cruz Junior - (EMC/UFG)** com a participação dos demais membros da Banca Examinadora: Professor Doutor **Tiago do Carmo Nogueira - (IFBaiano)** Coorientador, Professor Doutor **Cássio Dener Noronha Vinhal - (EMC/UFG)** Membro Titular Externo, Professor Doutor **Wesley Pacheco Calixto - (EMC/UFG)** Membro Titular Interno: **cuja participação ocorreram através do link: meet.google.com/ucw-khge-pnc**. Durante a arguição os membros da banca **não fizeram** sugestão de alteração do título do trabalho. A Banca Examinadora reuniu-se em sessão secreta a fim de concluir o julgamento da Dissertação, tendo sido o candidato **aprovado** pelos seus membros. Proclamados os resultados pelo Professor Doutor **Gelson da Cruz Junior**, Presidente da Banca Examinadora, foram encerrados os trabalhos e, para constar, lavrou-se a presente ata que é assinada pelos Membros da Banca Examinadora, aos dezessete dias dos mês de junho de dois mil e vinte e cinco.

TÍTULO SUGERIDO PELA BANCA



Documento assinado eletronicamente por **Vinicios Matheus Oliveira Da Silva, Discente**, em 05/08/2025, às 10:42, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Wesley Pacheco Calixto, Usuário Externo**, em 05/08/2025, às 10:47, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Tiago Do Carmo Nogueira, Usuário Externo**, em 05/08/2025, às 11:52, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Gelson Da Cruz Junior, Professor do Magistério Superior**, em 06/08/2025, às 11:18, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Cassio Dener Noronha Vinhal, Professor do Magistério Superior**, em 06/08/2025, às 11:21, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **5546564** e o código CRC **34AB8243**.

Agradecimentos

Em primeiro lugar, agradeço a Deus por ter me apoiado em cada etapa desta jornada, guiando meus passos, renovando minhas forças nos momentos difíceis e permitindo-me chegar até aqui.

Sou grato à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro concedido, que foi essencial para a concretização deste trabalho. Agradeço ao Prof. Dr. Gelson da Cruz Júnior, meu orientador, pela orientação e pela confiança depositada em mim durante o desenvolvimento desta pesquisa. Sua experiência e comprometimento foram fundamentais para a conclusão deste trabalho.

Agradeço em especial ao Prof. Dr. Tiago do Carmo Nogueira, meu coorientador, que tem sido uma presença constante e inspiradora desde a minha graduação. Foi ele quem me motivou a seguir adiante na vida acadêmica. Sua ajuda foi muito além das correções. Não fosse por seu apoio contínuo, talvez eu não tivesse seguido para o mestrado. Muito obrigado, professor, por fazer parte de toda essa caminhada.

À minha mãe e aos meus avós, que sempre estiveram ao meu lado com amor, paciência e incentivo incondicional, minha eterna gratidão. Vocês são a base de tudo o que conquistei.

Também agradeço aos amigos, colegas de pesquisa e professores que, de forma direta ou indireta, contribuíram para a minha formação e para a realização deste trabalho.

A todos que fizeram parte dessa trajetória: muito obrigado!

*“Consagre ao Senhor tudo o que você faz,
e os seus planos serão bem-sucedidos.
(Bíblia Sagrada, Provérbios 16, 13)*

Resumo

A bananicultura enfrenta desafios significativos devido a doenças foliares, como sigatoka negra, sigatoka amarela, mal-do-panamá e cordona, que comprometem a produtividade e aumentam os custos de produção. Métodos tradicionais de detecção dessas doenças apresentam limitações em precisão e escalabilidade, evidenciando a necessidade de soluções automatizadas. Este estudo propõe a implementação e avaliação de redes neurais convolucionais (CNNs) baseadas nas arquiteturas LeNet e Vision Transformer (ViT). Além disso, é apresentado um modelo híbrido denominado HibridNet, que combina as vantagens estruturais de ambas as arquiteturas. Os resultados experimentais indicam que o HibridNet obteve maior acurácia em comparação com os modelos individuais ViT e LeNet. A abordagem híbrida proposta demonstra elevado potencial para auxiliar no manejo fitossanitário da bananicultura, contribuindo para o aumento da produtividade e a redução dos custos operacionais.

Palavras-chave: Folhas de Bananeira; Doenças Foliares; Redes Neurais Convolucionais (CNN); Vision Transformer (ViT); Aprendizado Profundo; Agricultura de Precisão.

Abstract

Banana cultivation faces significant challenges due to foliar diseases such as black sigatoka, yellow sigatoka, Panama disease, and cordona, which reduce productivity and increase production costs. Traditional disease detection methods are often limited in accuracy and scalability, highlighting the need for automated solutions. This study proposes the implementation and evaluation of convolutional neural networks (CNNs) based on LeNet and Vision Transformer (ViT) architectures. Additionally, a novel hybrid model, named HibridNet, is introduced by combining the strengths of both architectures. Experimental results show that HibridNet achieves higher accuracy compared to individual ViT and LeNet models. The proposed hybrid approach demonstrates significant potential to support disease management in banana cultivation, improving productivity and reducing operational costs.

Keywords: Banana Leaves; Foliar Diseases; Convolutional Neural Networks (CNN); Vision Transformer (ViT); Deep Learning; Precision Agriculture.

Lista de ilustrações

Figura 2.1 – Sigatoka negra em folha de bananeira.	35
Figura 2.2 – Sigatoka amarela em folha de bananeira.	38
Figura 2.3 – Mal-do-Panamá em folha de bananeira.	41
Figura 2.4 – Cordana em folha de bananeira.	44
Figura 3.1 – Amostras do conjunto de dados utilizado na proposta metodológica. . .	48
Figura 3.2 – Arquitetura adaptada baseada na LeNet-5.	51
Figura 3.3 – Arquitetura básica da vision transformer (ViT).	51
Figura 3.4 – Arquitetura proposta denominada HibridNet.	52
Figura 4.1 – Acúrcia do modelo LeNet no conjunto de dados de teste e validação. .	57
Figura 4.2 – Perda do modelo LeNet no conjunto de dados de teste e validação. . .	59
Figura 4.3 – Métricas de avaliação do modelo LeNet no conjunto de dados de treina- mento, teste e validação.	61
Figura 4.4 – Acurácia do modelo ViT no conjunto de dados de teste e validação. . .	65
Figura 4.5 – Perda do modelo ViT no conjunto de dados de teste e validação.	66
Figura 4.6 – Métrica de avaliação do modelo ViT no conjunto de dados de treina- mento, teste e validação.	68
Figura 4.7 – Acurácia do modelo HibridNet no conjunto de dados de teste e validação.	70
Figura 4.8 – Perda do modelo HibridNet no conjunto de dados de teste e validação.	71
Figura 4.9 – Métrica de avaliação do modelo HibridNet no conjunto de dados de treinamento, teste e validação.	72

Lista de tabelas

Tabela 2.1 – Etapas do treinamento de CNNs.	27
Tabela 2.2 – Camadas da arquitetura LeNet.	30
Tabela 2.3 – Etapas da arquitetura vision transformer (ViT).	31
Tabela 3.1 – Distribuição das classes no conjunto de dados após aplicação de aumento de dados (Data Augmentation).	50
Tabela 3.2 – Hiperparâmetros utilizados nos experimentos	53
Tabela 4.1 – Matriz de Confusão do modelo LeNet.	62
Tabela 4.2 – Métricas de desempenho do modelo LeNet.	63
Tabela 4.3 – Matriz de Confusão do modelo ViT.	68
Tabela 4.4 – Métricas de desempenho do modelo ViT.	69
Tabela 4.5 – Matriz de Confusão do modelo HibridNet.	72
Tabela 4.6 – Métricas de desempenho do modelo HibridNet.	73
Tabela 4.7 – Comparação de desempenho entre os modelos LeNet, ViT e HibridNet.	74
Tabela 4.8 – Estudo de ablação dos modelos LeNet, ViT e HybridNet.	74

Sumário

1	Introdução	15
2	Referencial Teórico	18
2.1	Aprendizado Profundo (DL)	18
2.2	Redes Neurais Artificiais (ANNs)	20
2.3	Redes Neurais Convolucionais (CNNs)	23
2.3.1	Autoencoder Convolucional	25
2.3.2	Treinamento das CNNs	26
2.3.3	Classificação das CNNs	29
2.4	Transferência de Aprendizagem (TL)	32
2.5	Doenças Foliares	34
2.5.1	Sigatoka Negra	35
2.5.2	Sigatoka Amarela	37
2.5.3	Mal-do-Panamá	40
2.5.4	Cordana	43
3	Proposta de Modelo Convolucional	47
3.1	Conjunto de Dados	47
3.2	Preparação e Pré-Processamento	48
3.3	Modelos Convolucionais	50
3.3.1	Arquitetura Adpatada da LeNet-5	50
3.3.2	Arquitetura ViT	51
3.3.3	Arquitetura HibridNet	52
3.4	Configurações dos Experimentos	53
3.5	Métricas de Avaliação dos Modelos	53
3.5.1	Acurácia (ACC)	54
3.5.2	F1-Score (F1)	54
3.5.3	Área Sob a Curva (AUC)	55
3.5.4	Erro Médio Quadrático (MSE)	55
3.5.5	Regularização L1 e L2	56
4	Resultados e Discussões	57
4.1	Modelo Baseado na Arquitetura LeNet	57
4.2	Modelo Baseado na Arquitetura ViT	64
4.3	Modelo HibridNet	69
4.4	Estudo de Ablação dos Modelos LeNet, ViT e HibridNet	74
5	Conclusão	76

Referências 77

1 Introdução

A agricultura é uma atividade de grande relevância no Brasil, sendo fundamental tanto para a alimentação da população quanto para o fortalecimento da economia nacional. Nesse contexto, a agricultura constitui uma das principais fontes de renda para muitos países em desenvolvimento. Dentro desse cenário, destaca-se o cultivo da banana, uma fruta rica em minerais como cálcio, manganês, potássio, magnésio e ferro, além de ser uma fonte importante de vitaminas (JUNIOR et al., 2020; MUNAWAR et al., 2025).

As bananas são frutas amplamente consumidas em todo o mundo, sendo reconhecidas por seu alto valor energético. Nessa perspectiva, torna-se essencial o monitoramento das folhas de bananeira, que frequentemente são afetadas por doenças e por alterações climáticas. Esse cuidado é fundamental para mitigar perdas significativas nas lavouras, reduzindo, conseqüentemente, os impactos econômicos associados à produção e exportação (SCHUMANN et al., 2018; PAWAR; DEOSARKAR, 2024; GONZÁLEZ-RODRÍGUEZ et al., 2024).

A bananicultura, portanto, é uma atividade agrícola de grande importância econômica e social em diversas regiões do mundo (MUNAWAR et al., 2025). No entanto, a ocorrência de doenças nas folhas de bananeira - tais como sigatoka negra (*Pseudocercospora fijiensis*), sigatoka amarela (*Pseudocercospora musae*), e mal-do-panamá (*Fusarium oxysporum f. sp. cubense*), entre outras - representa um desafio significativo para os produtores, acarretando perdas na produtividade e o aumento dos custos com o controle fitossanitário (JUNIOR et al., 2020).

Diante desse cenário, destaca-se a aplicação de redes neurais artificiais (do inglês *artificial neural networks* - ANN) como uma solução promissora na identificação de doenças foliares, utilizando o reconhecimento de imagens aliado a técnicas de processamento de linguagem natural (do inglês *natural language processing* - NLP) para a previsão e análise de dados (NATALI, 2021).

Os métodos tradicionais de identificação dessas doenças, geralmente baseados em inspeções visuais feitas por especialistas, estão sujeitos a erros e se tornam inviáveis em larga escala (SINGH; GULERIA; SHARMA, 2024). No entanto, com os avanços da inteligência artificial (IA), as redes neurais convolucionais (do inglês *convolutional neural networks* - CNNs) vêm se destacando no reconhecimento de padrões em imagens, demonstrando alta eficácia na detecção e classificação de doenças em plantas (MILITANTE; GERARDO; DIONISIO, 2019).

Segundo Ergün (2025), muitos modelos baseados exclusivamente em CNNs apresentam limitações no que diz respeito à precisão e à capacidade de generalização, especialmente

quando expostos a cenários mais complexos, que envolvem variações de iluminação, sobreposição de folhas e outras condições ambientais adversas.

Diante desse desafio, surge a necessidade de explorar arquiteturas híbridas, que combinem CNNs com técnicas de aprendizado profundo (do inglês *deep learning* - DL), visando melhorar a acurácia na classificação e na detecção de doenças foliares. Assim, este trabalho busca responder à seguinte questão de pesquisa:

Q01 - Como implementar e analisar um modelo CNN híbrido que proporcione maior precisão e eficiência na classificação das principais doenças em folhas de bananeira, considerando as variáveis e os desafios práticos do campo?

Essa inquietação justifica-se pela necessidade de desenvolver ferramentas tecnológicas mais robustas, capazes de auxiliar no manejo da bananicultura, promovendo maior sustentabilidade, redução das perdas econômicas e aumento da eficiência no controle fitossanitário. Além disso, a adoção de modelos híbridos de CNN representa uma contribuição significativa para a área da visão computacional, abrindo novas possibilidades na detecção e na classificação de doenças em plantas.

Neste contexto, este trabalho propõe o desenvolvimento de um modelo híbrido baseado nas arquiteturas LeNet e Vision Transformer (ViT), cuja escolha fundamenta-se na complementaridade de suas características. A LeNet é reconhecida pela sua eficiência em tarefas de reconhecimento de padrões locais, especialmente em cenários com conjuntos de dados reduzidos, devido à sua arquitetura simples e baixo custo computacional (GANOKRATANAA; KETCHAM; PRAMKEAW, 2023). Por outro lado, a ViT destaca-se pela sua capacidade de capturar relações espaciais globais por meio de mecanismos de autoatenção, sendo particularmente eficaz em situações em que CNNs tradicionais enfrentam limitações, como variações de iluminação, deformações ou sobreposição de elementos (WANG et al., 2025). Dessa forma, a integração dessas arquiteturas visa combinar a capacidade de extração eficiente de características locais da LeNet com o poder de modelagem contextual global da ViT, resultando em um modelo mais robusto e preciso.

Diante disso, o objetivo principal desta pesquisa é implementar e analisar uma CNN híbrida capaz de classificar, de forma eficiente e precisa, as doenças que acometem folhas de bananeira, levando em consideração as variações ambientais e os desafios práticos enfrentados no campo.

Parte-se da hipótese de que a arquitetura proposta, denominada HibridNet, ao combinar elementos da ViT e da LeNet, proporcionará maior precisão e estabilidade na classificação de doenças foliares, mesmo em condições ambientais adversas. Para atingir esse objetivo, os seguintes objetivos específicos são estabelecidos:

- Identificar e analisar as principais doenças que afetam folhas de bananeira, bem como suas características visuais em imagens;
- Desenvolver uma base de dados de imagens de folhas de bananeira, com ênfase nas principais doenças encontradas;
- Avaliar a aplicação prática do modelo proposto, destacando seu potencial como ferramenta de apoio no manejo da bananicultura.

Além desta seção, este trabalho está organizado da seguinte forma: na seção 2 são apresentadas as fundamentações teóricas que embasam a pesquisa; na seção 3 são descritos os procedimentos metodológicos; na seção 4 são apresentados e analisados os resultados obtidos; e, por fim, na seção 5 são discutidas as principais conclusões, as limitações do estudo e as sugestões para trabalhos futuros.

2 Referencial Teórico

Esta seção apresenta os fundamentos teóricos que embasam este estudo, fornecendo suporte conceitual para o desenvolvimento e a análise do modelo proposto. Inicialmente, na subseção 2.1, são abordados os conceitos de aprendizado profundo (do inglês *deep learning* - DL). Na sequência, a subseção 2.2 discute as redes neurais artificiais (do inglês *artificial neural networks* - ANNs) e suas principais características.

A subseção 2.3 aprofunda-se nas redes neurais convolucionais (do inglês *convolutional neural networks* - CNNs), destacando seu funcionamento, incluindo o autoencoder convolucional (2.3.1), os processos de treinamento das CNNs (2.3.2) e suas aplicações em tarefas de classificação (2.3.3).

Na subseção 2.4, aborda-se a técnica de transferência de aprendizagem (do inglês *transfer learning* - TL), que permite reaproveitar modelos previamente treinados em novos problemas.

Por fim, a subseção 2.5 apresenta uma revisão sobre as principais doenças foliares que afetam as plantações, com foco na sigatoka negra (2.5.1), sigatoka amarela (2.5.2), mal-do-panamá (2.5.3) e cordana (2.5.4), que são relevantes para o contexto deste trabalho.

2.1 Aprendizado Profundo (DL)

De acordo com o estudo desenvolvido por Schumann et al. (2018), a DL é uma subárea do aprendizado de máquina (do inglês *machine learning* - ML) que busca modelar e resolver problemas complexos utilizando redes neurais profundas, que são compostas por várias camadas de neurônios artificiais interconectados. Essas redes são inspiradas no funcionamento do cérebro humano, as quais possuem a capacidade de aprender representações hierárquicas e abstratas dos dados (KASINATHAN; UYYALA, 2021).

Na percepção de Natali (2021), o principal diferencial do DL em relação a outros métodos de ML é sua habilidade de realizar a extração automática de características diretamente dos dados brutos, sem a necessidade de pré-processamento manual extensivo.

De acordo com o estudo desenvolvido por Junior et al. (2020), o funcionamento do DL é baseado no conceito de camadas, que processam os dados de forma sequencial. Cada camada extrai informações mais complexas dos dados à medida que são processados. Por exemplo, em uma tarefa de reconhecimento de imagens, as primeiras camadas podem identificar bordas ou formas simples, enquanto camadas mais profundas reconhecem objetos inteiros ou padrões complexos.

Essa hierarquia de aprendizado permite que modelos baseados em DL alcancem resultados superiores em tarefas de visão computacional, NPL, e até mesmo em diagnósticos médicos. Uma característica central é sua dependência de grandes quantidades de dados para alcançar um desempenho ideal (SCHUMANN et al., 2018; JIAO et al., 2024).

Os DLs, devido ao grande número de parâmetros, requerem conjuntos de dados substanciais para evitar problemas com o sobreajuste (do inglês *overfitting*), em que o modelo memoriza os dados de treinamento ao invés de aprender padrões generalizáveis. Além disso, essas redes se beneficiam de grande poder computacional, sendo frequentemente treinadas em unidades de processamento gráfico (do inglês *graphics processing units* - GPUs) ou unidades de processamento tensorial (do Inglês *tensor processing units* - TPUs), que aceleram significativamente o processo de treinamento (MILITANTE; GERARDO; DIONISIO, 2019).

De acordo com Schumann et al. (2018), a base matemática do DL está na otimização de funções de perda, que quantificam o erro do modelo durante o treinamento, sendo que o método mais comumente utilizado para essa otimização é a descida do gradiente, combinado com um processo chamado retropropagação (do inglês **backpropagation**), que ajusta os pesos das conexões da rede para minimizar o erro.

Em tese, o uso de funções de ativação não lineares, como ReLU (do inglês *rectified linear unit* - ReLU), é outro fator importante que permite às redes aprender representações mais complexas, pois essas funções introduzem não linearidades essenciais para resolver problemas realistas (NATALI, 2021). As arquiteturas utilizadas no DL variam de acordo com o tipo de problema a ser resolvido. Nesse sentido, as CNNs são amplamente usadas em visão computacional, pois são capazes de detectar padrões espaciais em imagens, como bordas e texturas, enquanto as redes neurais recorrentes (do inglês *recurrent neural networks* - RNNs) são projetadas para processar dados sequenciais, como séries temporais ou texto, permitindo que o modelo aprenda dependências entre elementos da sequência (MILITANTE; GERARDO; DIONISIO, 2019).

Apesar de o DL oferecer resultados impressionantes, ele também apresenta desafios. O treinamento dos modelos de DL é computacionalmente caro e pode ser demorado. Além disso, a escolha da arquitetura correta, a otimização dos hiperparâmetros e a obtenção de dados suficientemente representativos são aspectos críticos que influenciam o desempenho final do modelo (JUNIOR et al., 2020).

Diante ao que vem sendo apresentado, Natali (2021) descreveu em sua pesquisa que o DL continua sendo uma ferramenta poderosa e inovadora, com aplicações em áreas que vão desde diagnósticos médicos até veículos autônomos, destacando seu papel central na evolução da IA moderna.

2.2 Redes Neurais Artificiais (ANNs)

As ANNs são modelos computacionais inspirados na estrutura e no funcionamento do cérebro humano (KADAM; DHORE, 2024), projetados para processar informações de maneira semelhante à forma como os neurônios biológicos o fazem. As ANNs são amplamente utilizadas em tarefas de ML, como reconhecimento de padrões, previsão de dados e classificação, sendo a base de avanços significativos na IA (REGIANI; DIAS, 2021).

O conceito de redes neurais surgiu na década de 1940, quando Warren McCulloch e Walter Pitts propuseram um modelo matemático para neurônios artificiais, demonstrando que eles poderiam executar funções lógicas básicas. Na década de 1950, Frank Rosenblatt desenvolveu o *perceptron*, uma das primeiras implementações práticas de uma rede neural (MILITANTE; GERARDO; DIONISIO, 2019).

De acordo com Ashraf e Khan (2020), o *perceptron* era capaz de aprender pesos para classificar dados linearmente separáveis, mas suas limitações foram apontadas por Marvin Minsky e Seymour Papert em 1969, quando demonstraram que ele não conseguia resolver problemas não lineares, como o "XOR". Essa limitação resultou no período de estagnação na pesquisa de redes neurais, conhecido como o "inverno da IA".

Nos anos 1980, houve um renascimento do interesse por redes neurais devido à introdução do algoritmo de retropropagação, que permite o treinamento eficiente de redes multicamadas. Esse avanço possibilitou a solução de problemas mais complexos. Nos anos 2000, a disponibilidade de grandes volumes de dados (do inglês *big data*) e o aumento do poder computacional com GPUs, impulsionaram o desenvolvimento de DLs, estabelecendo as redes neurais como uma das tecnologias mais promissoras da atualidade (ASHRAF; KHAN, 2020).

De acordo com o estudo de Bisong et al. (2019), as ANNs são compostas por unidades chamadas neurônios ou nós, organizados em camadas. Essas camadas são classificadas como: camada de entrada que recebe os dados brutos que serão processados; as camadas ocultas que realizam o processamento, extraíndo características e aprendendo padrões a partir dos dados; e a camada de saída, na qual produz o resultado final, como uma classificação ou previsão.

Neste sentido, cada neurônio recebe entradas, as multiplica por pesos, soma esses valores e passa o resultado por uma função de ativação que introduz não linearidades, permitindo que a rede resolva problemas mais complexos. O aprendizado da rede ocorre por meio do ajuste dos pesos com base no erro entre as saídas esperadas e as produzidas. Esse processo é realizado utilizando o algoritmo de retropropagação e otimizadores (ASHRAF; KHAN, 2020).

As ANNs podem aprender padrões e representações complexas diretamente dos dados, tornando-as úteis em tarefas como reconhecimento de imagens, processamento de

linguagem natural e predições. Após o treinamento, as redes podem generalizar para dados que não foram vistos antes, desde que o treinamento seja bem conduzido e os dados sejam representativos (JUNIOR et al., 2020).

As redes podem ser ajustadas para uma ampla variedade de problemas, variando a arquitetura, os hiperparâmetros e as funções de ativação. As camadas ocultas permitem que a rede aprenda representações hierárquicas, desde características simples até padrões complexos. Além disso, elas geralmente requerem grandes conjuntos de dados para treinamento e poder computacional significativo, especialmente em arquiteturas profundas (REGIANI; DIAS, 2021).

As ANNs têm aplicações amplas e diversificadas, incluindo reconhecimento de imagens e detecção de objetos; processamento de linguagem natural, como tradução automática e *chatbots*; diagnósticos médicos, como a detecção de doenças em imagens de raios-X; sistemas de recomendação, como os usados por plataformas de *streaming* e *e-commerce*; e, ainda, veículos autônomos, que dependem de redes para visão computacional e tomada de decisão (BARATLOO et al., 2015; LIU, 2020).

Apesar de seus benefícios, Regiani e Dias (2021), Li, Yan e Xu (2021), relataram que as ANNs apresentam desafios, como a necessidade de grandes volumes de dados para evitar sobreajuste, o alto custo computacional e a dificuldade em interpretar como os modelos tomam decisões (a chamada "caixa-preta"). Ainda assim, seu impacto no avanço da IA é inegável, sendo uma das principais tecnologias que impulsionam inovações na ciência e na indústria.

O funcionamento das ANNs baseia-se em uma estrutura composta por camadas de neurônios artificiais (HARTMANN et al., 2020). Cada camada desempenha um papel específico no processamento da informação. A primeira camada, chamada de camada de entrada, recebe os dados brutos. Em seguida, as camadas ocultas realizam cálculos matemáticos complexos, transformando os dados de forma progressiva. Por fim, a camada de saída fornece a previsão ou classificação final. Esse fluxo de processamento permite que as ANNs extraia características relevantes dos dados de entrada e tome decisões baseadas em padrões identificados (HARTMANN et al., 2020).

Uma das características fundamentais das ANNs é o treinamento supervisionado, no qual um conjunto de dados rotulado é utilizado para ajustar os pesos das conexões entre os neurônios. O algoritmo de aprendizado mais comumente empregado nesse processo é o retropropagação, que calcula o erro entre a saída prevista e a saída real, ajustando os pesos para minimizar esse erro. Além do treinamento supervisionado, existem abordagens como o aprendizado não supervisionado e o aprendizado por reforço, cada um adequado para diferentes tipos de problemas e conjuntos de dados (RAGHU et al., 2019).

A arquitetura das ANNs pode variar significativamente conforme a complexidade

do problema a ser resolvido. As ANNs simples possuem apenas uma ou duas camadas ocultas, enquanto redes mais profundas, conhecidas como redes neurais profundas (do inglês *deep neural networks* - DNNs), contêm várias camadas ocultas e são a base do DL. Essas redes são capazes de aprender representações hierárquicas dos dados, permitindo a resolução de problemas altamente complexos, como reconhecimento facial e tradução automática de idiomas (ZOPH et al., 2018).

Dentre as principais arquiteturas de ANNs, destacam-se as CNNs e as RNNs. As CNNs são amplamente utilizadas no processamento de imagens, pois possuem camadas convolucionais que identificam padrões visuais como bordas, texturas e formas. Já as RNNs são empregadas no processamento de sequências de dados, como texto e séries temporais, pois conseguem manter informações de estados anteriores, permitindo a análise de dependências temporais (ZHU et al., 2019).

Outro elemento essencial das ANNs é a função de ativação, responsável por introduzir não-linearidade no modelo. Sem essa função, as ANNs seriam apenas combinações lineares de entradas, limitando sua capacidade de aprendizado. As funções de ativação mais comuns incluem a sigmoide, ReLU e tangente hiperbólica. A escolha da função adequada depende do tipo de problema e da arquitetura da rede (MAHMOOD et al., 2020).

A capacidade das ANNs de generalização também depende de técnicas de regularização, que evitam o problema do sobreajuste. O sobreajuste ocorre quando a rede aprende padrões específicos excessivos do conjunto de treinamento e tem dificuldade em generalizar para novos dados. Para mitigar esse problema, utilizam-se técnicas como *dropout*, normalização de pesos e aumento de dados, que ajudam a tornar as ANNs mais robustas e adaptáveis a diferentes contextos (RAGHU et al., 2019).

Com o avanço das ANNs e do DL, observa-se um impacto significativo em diversas áreas do conhecimento. Na medicina, por exemplo, ANNs são aplicadas para detectar doenças em exames de imagem, melhorando diagnósticos e auxiliando médicos na tomada de decisão. No setor financeiro, esses modelos são utilizados para prever tendências de mercado e identificar padrões de fraude. Na indústria automobilística, ANNs impulsionam o desenvolvimento de veículos autônomos, permitindo a análise e tomada de decisão em tempo real (HARTMANN et al., 2020).

O futuro das ANNs está diretamente ligado ao aprimoramento de hardware especializado, como GPUs e circuitos integrados de aplicação específica (ASICs), que tornam o treinamento de modelos cada vez mais eficiente. Além disso, a pesquisa avança na direção de ANNs mais interpretáveis, permitindo uma maior transparência nos processos decisórios de modelos de IA, especialmente em aplicações críticas como saúde e justiça (HUANG et al., 2017).

As ANNs representam um marco significativo na evolução da IA e do ML. Seu

funcionamento baseado em camadas de neurônios interconectados permite a resolução de problemas complexos em diversas áreas. Com o desenvolvimento contínuo da tecnologia, espera-se que as ANNs continuem a transformar o mundo moderno, trazendo soluções inovadoras e impactantes para diferentes setores da sociedade.

2.3 Redes Neurais Convolucionais (CNNs)

As CNNs representam um dos avanços mais relevantes e impactantes no campo da IA e do ML. Diferentemente das redes neurais tradicionais (ANNs), que processam cada entrada de forma independente, as CNNs são projetadas especificamente para lidar com dados que possuem estrutura espacial, como imagens, sinais e vídeos, preservando as relações espaciais entre os elementos, como os pixels de uma imagem (HOWARD et al., 2019).

O princípio fundamental das CNNs está na utilização de operações de convolução, que permitem extrair automaticamente características hierárquicas dos dados, desde padrões simples, como bordas e texturas, até composições complexas, como objetos inteiros. A arquitetura típica de uma CNN combina três tipos principais de camadas: camadas convolucionais, responsáveis pela extração de características; camadas de agrupamento (*pooling*), que reduzem a dimensionalidade e aumentam a eficiência computacional; e camadas totalmente conectadas, que transformam os mapas de características em saídas interpretáveis para tarefas como classificação, detecção ou segmentação (TAN; LE, 2019).

O desenvolvimento das CNNs foi profundamente inspirado pelos estudos neurobiológicos de Hubel e Wiesel, que, na década de 1960, identificaram como neurônios no córtex visual de gatos respondiam a estímulos visuais específicos, como linhas e bordas, processando informações de forma hierárquica (ASHRAF; KHAN, 2020). Este conceito biológico motivou pesquisadores a modelar computacionalmente sistemas que imitassem esse tipo de processamento.

Na década de 1980, Yann LeCun introduziu a LeNet, uma das primeiras implementações bem-sucedidas de uma CNN, aplicada ao reconhecimento de dígitos manuscritos, especialmente para leitura automatizada de cheques bancários. No entanto, devido às limitações computacionais da época e à escassez de grandes conjuntos de dados, as CNNs permaneceram por anos restritas a aplicações específicas (BISONG et al., 2019).

Esse cenário começou a mudar drasticamente na década de 2000, impulsionado por dois fatores principais: o aumento exponencial do poder de processamento, especialmente com o uso de GPUs, e a disponibilidade de bases de dados massivas, como o ImageNet. Esses avanços culminaram no desenvolvimento da AlexNet, em 2012, que representou um marco na história das CNNs. AlexNet não só utilizou uma rede muito mais profunda, como também aplicou técnicas inovadoras como a função de ativação ReLU e o *dropout* para

reduzir sobreajuste, conquistando resultados surpreendentes no desafio de classificação de imagens do ImageNet (ABADI et al., 2016).

A partir desse ponto, sucessivas melhorias nas arquiteturas tornaram as CNNs ainda mais eficientes e poderosas. Redes como VGGNet, Inception e ResNet contribuíram significativamente para aumentar a profundidade, a capacidade de generalização e a eficiência das redes. A ResNet, em particular, revolucionou o campo ao introduzir as conexões residuais, que solucionaram o problema de degradação de desempenho em redes muito profundas, possibilitando arquiteturas com centenas ou até milhares de camadas (REGIANI; DIAS, 2021; ALMEIDA, 2017).

Mais recentemente, arquiteturas como MobileNet e EfficientNet foram desenvolvidas com foco na otimização do uso de recursos, tornando possível a utilização de CNNs em dispositivos móveis e em cenários de computação embarcada, sem perda significativa de desempenho (ASHRAF; KHAN, 2020).

As CNNs consolidaram-se como uma das tecnologias mais utilizadas no campo da visão computacional e estão presentes em uma ampla gama de aplicações. Na saúde, são essenciais para análise de exames médicos, como tomografias e ressonâncias magnéticas, auxiliando na detecção precoce de doenças e contribuindo para diagnósticos mais precisos e rápidos (MILITANTE; GERARDO; DIONISIO, 2019). Na indústria automotiva, CNNs são fundamentais para o funcionamento de sistemas de condução autônoma, permitindo o reconhecimento de pedestres, sinais de trânsito e obstáculos em tempo real (HE et al., 2019).

Além disso, CNNs são amplamente aplicadas em sistemas de segurança e vigilância, no monitoramento de padrões agrícolas, no rastreamento ambiental via imagens de satélite e até mesmo em dispositivos móveis, onde modelos compactos como MobileNet possibilitam funcionalidades como reconhecimento facial, processamento de imagens e realidade aumentada (ALMEIDA, 2017).

O impacto das CNNs se estende ainda a plataformas digitais, redes sociais e entretenimento, sendo responsáveis por funcionalidades como classificação de conteúdo, filtros inteligentes e algoritmos de recomendação. No contexto de vídeos, as CNNs viabilizam o reconhecimento de padrões, rastreamento de objetos e identificação de eventos em tempo real (REGIANI; DIAS, 2021).

Apesar de sua ampla adoção e eficácia comprovada, as CNNs ainda apresentam desafios, como a necessidade de grandes volumes de dados rotulados para treinamento e o elevado custo computacional associado às arquiteturas mais profundas. Contudo, esses desafios têm sido mitigados por abordagens como o TL, que permite reutilizar modelos pré-treinados em novas tarefas, e pela integração de técnicas como Redes Gerativas Adversariais (do inglês *generative adversarial networks* - GANs) e aprendizado por reforço, que ampliam

ainda mais o potencial das CNNs (DOSOVITSKIY; AL., 2021).

Nos últimos anos, surgiram modelos alternativos às CNNs, como os Vision Transformers (ViTs), que adotam mecanismos de atenção ao invés de convoluções para processar informações visuais. Embora promissores, os ViTs ainda coexistem com as CNNs, que continuam dominando uma vasta gama de aplicações, especialmente pela sua maturidade, eficiência e ampla validação prática (DOSOVITSKIY; AL., 2021).

Em síntese, como apontam Junior et al. (2020) e Ashraf e Khan (2020), as CNNs constituem uma das bases mais sólidas da IA moderna, destacando-se pela capacidade de processar dados com estrutura espacial e por possibilitar avanços significativos em áreas críticas como saúde, transporte, segurança, meio ambiente e indústria. Sua evolução é fruto de décadas de pesquisa, desde os primeiros *insights* neurobiológicos até as sofisticadas arquiteturas atuais, que seguem moldando o futuro da IA e da transformação digital.

2.3.1 Autoencoder Convolutacional

Os autoencoders convolucionais (*Convolutional Autoencoders* - CAEs) são uma extensão dos autoencodificadores tradicionais, projetados especificamente para lidar com dados estruturados em grade, como imagens. Eles combinam a arquitetura dos autoencoders com as CNNs, permitindo a extração de representações latentes que preservam as relações espaciais dos dados (ALMEIDA, 2017; XIAO et al., 2019).

A arquitetura de um CAE é composta por duas partes principais: o codificador (*encoder*) e o decodificador (*decoder*). O codificador aplica sucessivas camadas convolucionais e de agrupamento para reduzir a dimensionalidade, extraíndo características relevantes dos dados de entrada. O decodificador, por sua vez, utiliza camadas de convolução transposta (*upsampling*) para reconstruir o dado original a partir da representação latente (REGIANI; DIAS, 2021; ALZUBAIDI et al., 2021).

O objetivo principal dos CAEs é minimizar a diferença entre a entrada e a saída reconstruída, geralmente utilizando métricas como o erro quadrático médio (MSE) (ABUHAMAD; WALSH, 2014). Essa capacidade de reconstrução permite aplicações como compressão de imagens, remoção de ruído e detecção de anomalias (XIAO et al., 2019; SCHUMANN et al., 2018; JUNIOR et al., 2020).

Uma das maiores vantagens dos CAEs é sua habilidade de preservar a estrutura espacial dos dados, diferentemente dos autoencoders baseados em camadas densamente conectadas. Isso os torna particularmente eficazes em tarefas de visão computacional, como segmentação de imagens, geração de imagens sintéticas e reconstrução de dados corrompidos (XIE et al., 2017; HARTMANN et al., 2020).

Além disso, os CAEs têm sido integrados com modelos generativos, como as GANs e os *variational autoencoders* (VAEs), ampliando seu uso para geração de dados sintéticos

e aprimoramento de imagens (RAGHU et al., 2019). Na medicina, são amplamente empregados na melhoria da qualidade de imagens diagnósticas, enquanto na segurança digital contribuem para a detecção de fraudes e anomalias (HARTMANN et al., 2020).

O processo de treinamento dos CAEs envolve a otimização dos pesos das camadas convolucionais por meio de algoritmos como Adam e RMSprop, com o uso de técnicas como normalização de lotes e *dropout* convolucional para mitigar sobreajuste (ZOPH et al., 2018; ZHU et al., 2019). A função de perda utilizada varia conforme o objetivo da aplicação, podendo incluir, além do MSE, funções perceptuais para maior qualidade visual na reconstrução (ALMEIDA, 2017).

Contudo, apesar de suas vantagens, os CAEs apresentam alguns desafios. Seu desempenho é altamente dependente de conjuntos de dados grandes e representativos. Além disso, podem sofrer com sobreajuste em bases pequenas e demandam alto custo computacional, especialmente para imagens de alta resolução (MAHMOOD et al., 2020).

De forma geral, os autoencoders convolucionais constituem uma ferramenta essencial no ecossistema do DL, desempenhando papel fundamental em tarefas de compressão, análise e geração de dados visuais. Sua capacidade de capturar e reconstruir informações espaciais faz dos CAEs uma tecnologia central no avanço de aplicações em IA (ALMEIDA, 2017).

2.3.2 Treinamento das CNNs

O treinamento de CNNs constitui um processo central no DL. Esse processo envolve o ajuste dos pesos e parâmetros da rede com o objetivo de minimizar a diferença entre as previsões do modelo e os valores reais dos dados. Ressalta-se que esse treinamento é baseado no aprendizado supervisionado e utiliza métodos matemáticos e computacionais para otimizar o desempenho da rede em tarefas específicas, como classificação de imagens, detecção de objetos e segmentação (REGIANI; DIAS, 2021).

O treinamento de CNNs pode ser dividido em diversas etapas, desde a preparação dos dados até a validação do modelo. Cada uma dessas etapas desempenha um papel fundamental na obtenção de uma ANN precisa, eficiente e com boa capacidade de generalização (BISONG et al., 2019).

A Tabela 2.1 apresenta as principais etapas do treinamento de uma CNN, juntamente com suas respectivas descrições, facilitando a compreensão da abordagem adotada.

Tabela 2.1 – Etapas do treinamento de CNNs.

Etapa	Descrição
Preparação dos dados	<p>A qualidade e a quantidade dos dados de treinamento têm um impacto significativo no desempenho das CNNs. A preparação dos dados envolve:</p> <ul style="list-style-type: none">• Coleta e Rotulagem: Obtenção de conjuntos de dados grandes e bem rotulados.• Pré-processamento: Normalização, redimensionamento e adequação dos dados.• Aumento de Dados: Técnicas como rotação, translação, espelhamento e ajuste de brilho.• Divisão: Separação em conjuntos de treinamento, validação e teste.
Inicialização da rede	<p>Antes do treinamento, a rede precisa ser configurada, incluindo:</p> <ul style="list-style-type: none">• Definição da Arquitetura: Número de camadas, filtros e <i>kernels</i>.• Inicialização dos Pesos: Métodos como He ou Xavier para melhorar a convergência.
<i>Forward Propagation</i>	<p>A entrada passa pelas camadas convolucionais, de <i>pooling</i> e totalmente conectadas para gerar uma saída. Operações realizadas:</p> <ul style="list-style-type: none">• Convolução: Extração de características.• Pooling: Redução de dimensionalidade.• Ativação: Aplicação de funções como ReLU, sigmoid ou tanh.

Continua na próxima página

Tabela 2.1 – continuação da página anterior

Etapa	Descrição
Cálculo da função de perda	<p>Mede a diferença entre a saída prevista e o valor real, utilizando funções como:</p> <ul style="list-style-type: none"> • Cross-Entropy Loss • Mean Squared Error (MSE) • Categorical Hinge Loss
Retropropagação e atualização dos pesos	<p>A escolha depende da tarefa.</p> <p>Processo de ajuste dos pesos da rede:</p> <ul style="list-style-type: none"> • Cálculo dos Gradientes: Usando a regra da cadeia. • Atualização dos Pesos: Aplicação dos gradientes via otimizadores, como: <ul style="list-style-type: none"> – SGD (Stochastic Gradient Descent) – Adam (Adaptive Moment Estimation) – RMSprop
Treinamento em lotes	<p>O treinamento é realizado em lotes, cada um contendo uma fração do conjunto de dados. Isso torna o processo mais eficiente e estável.</p>
Regularização e técnicas contra sobreajuste	<p>Técnicas que ajudam a melhorar a generalização:</p> <ul style="list-style-type: none"> • Dropout • L2 Regularization • Early Stopping

Continua na próxima página

Tabela 2.1 – continuação da página anterior

Etapa	Descrição
Avaliação do modelo	Monitoramento do desempenho durante e após o treinamento. Métricas comuns: <ul style="list-style-type: none">• Acurácia• F1-Score, Precisão e Recall• Loss de Validação
Teste e generalização	Após o treinamento, o modelo é testado em dados não vistos para avaliar sua capacidade de generalizar.

Fonte: Junior et al. (2020).

Apesar de altamente eficazes, as CNNs apresentam desafios, como a necessidade de grandes volumes de dados rotulados, elevados custos computacionais - especialmente em redes profundas -, risco de sobreajuste em conjuntos de dados pequenos ou desbalanceados e sensibilidade à definição de hiperparâmetros, como taxa de aprendizado e tamanho do lote (JUNIOR et al., 2020).

De forma geral, o treinamento de CNNs é um processo complexo, que demanda a integração de técnicas computacionais, estratégias de regularização e métodos de otimização para alcançar resultados robustos. Com os avanços no desenvolvimento de hardware especializado, como GPUs e TPUs, além do surgimento de bibliotecas de DL, como TensorFlow e PyTorch, o treinamento de CNNs tornou-se mais acessível e eficiente, viabilizando aplicações em áreas críticas, como saúde, segurança, transporte e ciência (ALOM et al., 2018).

2.3.3 Classificação das CNNs

A classificação das CNNs refere-se à tarefa de atribuir rótulos ou categorias a uma entrada com base nas características extraídas pela rede. Elas são fundamentais no campo da visão computacional devido à sua capacidade de aprender representações hierárquicas e complexas de dados visuais como imagens (WU; GU, 2015).

A arquitetura das CNNs geralmente é composta por uma sequência de camadas convolucionais, de agrupamento e totalmente conectadas. Cada uma dessas camadas tem como objetivo extrair informações progressivamente mais abstratas da entrada. Com o avanço da área, surgiram diversas arquiteturas inovadoras, como a LeNet - uma das pioneiras - e o ViT, baseado em transformadores, que vem ganhando destaque na visão computacional (BARATLOO et al., 2015).

Segundo Wu e Gu (2015), a LeNet-5 foi uma das primeiras CNNs a obter resultados expressivos em tarefas de classificação de imagens. Desenvolvida por Yann LeCun no final da década de 1980, sua arquitetura foi projetada para reconhecer dígitos manuscritos no conjunto de dados MNIST, um dos *benchmarks* mais utilizados em ML.

A LeNet-5 é considerada uma arquitetura simples, porém eficaz, para classificação de imagens. Sua estrutura é composta por camadas convolucionais, camadas de agrupamento e camadas totalmente conectadas, conforme descrito na Tabela 2.2 (ZHANG et al., 2017).

Tabela 2.2 – Camadas da arquitetura LeNet.

Camada	Descrição
Camadas convolucionais	A primeira camada convolucional aplica 6 filtros de 5×5 , gerando 6 mapas de características de tamanho reduzido. A segunda camada convolucional aplica 16 filtros de 5×5 , produzindo 16 mapas de características ainda mais abstratos.
Camadas de agrupamento	Após cada camada convolucional, uma operação de <i>pooling</i> (tipicamente <i>max pooling</i>) é aplicada para reduzir a resolução espacial e, assim, a complexidade computacional da rede.
Camadas totalmente conectadas	A rede termina com camadas totalmente conectadas (densas), que mapeiam as características extraídas para as classes finais de saída. No caso do MNIST, a rede tem 10 saídas, correspondentes às 10 classes possíveis (dígitos de 0 a 9).

Fonte: Regiani e Dias (2021).

A LeNet foi essencial para consolidar o uso de CNNs, demonstrando que era viável treinar redes profundas para tarefas de visão computacional mesmo com recursos computacionais limitados. Embora simples, sua arquitetura abriu caminho para modelos mais sofisticados desenvolvidos posteriormente (REGIANI; DIAS, 2021).

Por outro lado, o ViT representa uma abordagem mais recente para classificação de imagens, proposta por pesquisadores da Google Research. Diferentemente das CNNs, que utilizam convoluções para extrair características locais, o ViT adapta a arquitetura de transformadores - originalmente desenvolvida para NLP - à visão computacional (BARATLOO et al., 2015).

O ViT processa imagens dividindo-as em pequenos blocos (*patches*), que são tratados de forma análoga às palavras em modelos de NLP (ZHANG et al., 2017). As principais etapas do funcionamento do ViT estão descritas na Tabela 3.1.

O ViT se beneficia consideravelmente do treinamento em larga escala, melhorando seu desempenho à medida que mais dados estão disponíveis. Enquanto as CNNs capturam principalmente relações locais, o ViT é capaz de modelar relações globais entre diferentes

Tabela 2.3 – Etapas da arquitetura vision transformer (ViT).

Etapa	Descrição
Divisão da imagem em <i>patches</i>	A imagem é dividida em pequenos <i>patches</i> quadrados, que são então linearmente projetados em um vetor de características. Cada <i>patch</i> é tratado de maneira semelhante a uma palavra em um modelo Transformer de NLP.
Transformador encoder	Os <i>patches</i> são alimentados em uma pilha de camadas de codificadores do Transformer, que processam os patches de forma sequencial, levando em conta o contexto entre eles. O Transformer usa uma atenção autorregressiva para considerar as interdependências entre os <i>patches</i> , o que lhe permite capturar relações espaciais complexas.
Classificação	Após o processamento no Transformer, a saída é passada por uma camada de classificação para prever a classe da imagem. Uma técnica comum é usar um "token" de classe especial, que coleta as informações dos diferentes <i>patches</i> e produz a predição final.

Fonte: Zhang et al. (2017).

regiões da imagem, o que pode ser vantajoso em tarefas de maior complexidade (ZHANG et al., 2017).

De acordo com Wu e Gu (2015), o ViT apresenta alta versatilidade, podendo ser facilmente adaptado a diferentes tipos de dados e tarefas. Contudo, conforme Schumann et al. (2018), os modelos baseados em transformadores são altamente dependentes de grandes volumes de dados, o que pode representar uma limitação em cenários com dados escassos. Além disso, o ViT tende a demandar maior custo computacional em comparação às CNNs tradicionais, devido às operações intensivas de atenção (REGIANI; DIAS, 2021).

Para Zeiler e Fergus (2014), é relevante estabelecer um comparativo entre a LeNet e o ViT. A LeNet segue uma arquitetura clássica de CNNs, com forte capacidade de capturar padrões locais, enquanto o ViT adota a arquitetura de transformadores, apta a modelar relações globais na imagem.

O ViT, geralmente, supera as CNNs em tarefas mais complexas e com grandes conjuntos de dados. No entanto, em contextos com dados limitados, modelos mais simples como a LeNet podem apresentar desempenho mais satisfatório. Isso ocorre porque o ViT exige grandes volumes de dados para treinamento eficiente, enquanto a LeNet é eficaz mesmo em conjunto de dados menores (BISONG et al., 2019).

Em síntese, a classificação por meio de CNNs abrange diferentes arquiteturas, desde modelos clássicos, como a LeNet, até propostas mais recentes e sofisticadas, como o

ViT. Cada abordagem apresenta vantagens e limitações específicas, sendo a escolha do modelo condicionada à quantidade de dados disponível, aos recursos computacionais e à complexidade da tarefa a ser realizada (MILITANTE; GERARDO; DIONISIO, 2019; NATALI, 2021).

2.4 Transferência de Aprendizagem (TL)

A TL é uma técnica amplamente utilizada na IA, permitindo o reaproveitamento de modelos previamente treinados para novas tarefas. Essa abordagem é especialmente eficaz em CNNs, pois facilita a aplicação de modelos desenvolvidos em grandes bases de dados para contextos específicos. Ao reutilizar pesos e representações aprendidas, a TL reduz o tempo de treinamento e melhora a eficiência dos modelos, mesmo quando o conjunto de dados disponível é limitado (ZOPH et al., 2018).

Segundo Almeida (2017), a TL permite utilizar modelos previamente treinados em grandes bases de dados para resolver novos problemas com menos recursos computacionais e menor volume de dados. Essa técnica é amplamente empregada em DL, especialmente em cenários onde não há dados suficientes para treinar uma rede do zero.

O desenvolvimento da TL foi impulsionado pelo surgimento de redes profundas e pela disponibilidade de grandes bases, como o ImageNet, que contém milhões de imagens categorizadas. Modelos como *AlexNet*, *VGGNet*, *ResNet* e *EfficientNet* foram desenvolvidos com pesos otimizados a partir dessa base e, posteriormente, aplicados em diferentes domínios. Esse reaproveitamento de conhecimento permite a utilização de arquiteturas avançadas sem a necessidade de treinamento do zero, otimizando tempo e recursos computacionais (MAHMOOD et al., 2020).

A TL pode ser aplicada por meio de diferentes estratégias, dependendo da similaridade entre a tarefa original e a nova. Uma abordagem comum é a extração de características, na qual as camadas iniciais de um modelo pré-treinado são mantidas fixas, enquanto as camadas finais são ajustadas para a nova tarefa. Essa técnica é eficaz, pois as primeiras camadas aprendem padrões genéricos, como bordas, texturas e formas, enquanto as últimas camadas são responsáveis por características específicas da tarefa (ALZUBAIDI et al., 2021).

Outra estratégia é o ajuste fino (*fine-tuning*), em que parte ou todas as camadas do modelo são atualizadas no novo conjunto de dados. Essa abordagem é recomendada quando há diferenças consideráveis entre a tarefa original e a nova, e quando há um volume suficiente de dados. Nesse caso, empregam-se taxas de aprendizado reduzidas para preservar o conhecimento previamente adquirido, evitando sobreajuste (DOSOVITSKIY; AL., 2021).

Junior et al. (2020) destacam que as primeiras camadas de uma CNN capturam características genéricas dos dados, como texturas e formas, enquanto as camadas mais profundas aprendem representações específicas da tarefa. Por isso, essas representações podem ser reaproveitadas, desde que haja uma relação entre os domínios.

O processo de TL é composto por três etapas principais: (i) seleção de um modelo pré-treinado, geralmente treinado em grandes bases, como o ImageNet (ex.: VGG, ResNet, Inception, MobileNet, EfficientNet); (ii) adaptação da arquitetura para a nova tarefa, substituindo as camadas finais para refletir o novo espaço de saída; e (iii) treinamento, onde o modelo é ajustado utilizando o novo conjunto de dados, mantendo ou não algumas camadas congeladas, conforme a estratégia adotada (MILITANTE; GERARDO; DIONISIO, 2019; JUNIOR et al., 2020).

As estratégias mais comuns incluem:

- Congelamento total - apenas as camadas finais são treinadas, útil em cenários com poucos dados (NATALI, 2021).
- Ajuste fino (*fine-tuning*) - parte das camadas convolucionais é descongelada e ajustada, permitindo maior adaptação.
- Uso como extrator de características - as saídas de camadas intermediárias alimentam outros classificadores, como SVM ou redes rasas (BISONG et al., 2019).

Entre as principais vantagens da TL, destacam-se: (i) redução do custo computacional; (ii) menor necessidade de dados; (iii) melhor desempenho em cenários com dados limitados; e (iv) ampla aplicabilidade em áreas como diagnóstico médico, agricultura e sensoriamento remoto (ALMEIDA, 2017; SCHUMANN et al., 2018).

Por outro lado, existem limitações, como sensibilidade às diferenças entre os domínios original e de destino. Distribuições de dados significativamente diferentes podem comprometer a capacidade de generalização do modelo. Além disso, se o conjunto de dados da nova tarefa for muito pequeno, pode ocorrer sobreajuste durante o ajuste fino (ASHRAF; KHAN, 2020).

Com os avanços nas arquiteturas e a crescente disponibilidade de modelos pré-treinados, a TL tem se tornado ainda mais eficiente. Modelos como *EfficientNet* e ViT oferecem soluções avançadas, proporcionando maior precisão e melhor aproveitamento dos recursos computacionais (SZEGEDY et al., 2015).

De acordo com Schumann et al. (2018), a TL é uma abordagem poderosa para acelerar o desenvolvimento de modelos baseados em CNNs, reduzindo a necessidade de grandes volumes de dados e recursos computacionais. Sua aplicação tem impactado

significativamente áreas como saúde, agricultura e segurança, consolidando-se como uma ferramenta fundamental no DL.

2.5 Doenças Foliares

As doenças que afetam as folhas da bananeira representam um dos maiores desafios na produção de bananas em todo o mundo, em ressalva que essas doenças podem reduzir significativamente a produtividade e a qualidade da fruta, impactando negativamente o rendimento das lavouras e a economia de produtores (REGIANI; DIAS, 2021).

Em continuidade, muitas dessas doenças são difíceis de controlar, o que exige a implementação de estratégias integradas de manejo para garantir a sanidade das plantas e a sustentabilidade da cultura. As doenças das folhas de bananeira são causadas por uma variedade de agentes patológicos, incluindo fungos, bactérias e vírus (ZEILER; FERGUS, 2014).

O manejo integrado de doenças é uma abordagem que combina diversas estratégias de controle para minimizar os danos causados pelas doenças nas plantações de bananeiras. Entretanto, as práticas recomendadas englobam, o uso de variedades resistentes, em que o cultivo de variedades de bananeira que sejam mais resistentes a doenças específicas é uma das formas mais eficazes de controle (REGIANI; DIAS, 2021).

De acordo com a pesquisa de Szegedy et al. (2015), a rotação de culturas, que condiz em alternar o cultivo de bananas com outras culturas pode ajudar a reduzir a presença de patógenos no solo. O controle químico, com a aplicação de fungicidas e bactericidas é uma prática comum, mas deve ser feita com cautela para evitar resistência dos patógenos.

De acordo com o estudo de Regiani e Dias (2021), o uso de microorganismos antagonistas aos patógenos pode ser uma alternativa viável para o controle de doenças. Em ressalva, ajustar os níveis de irrigação e nutrientes pode ajudar a reduzir a incidência de doenças, já que muitas doenças prosperam em condições de excesso de umidade e desequilíbrio nutricional.

As doenças das folhas de bananeira representam um desafio significativo para a agricultura de banana. As mais comuns, como a sigatoka negra, sigatoka amarela, mal-de-panamá e cordona, podem reduzir drasticamente a produtividade das plantações, afetando tanto a quantidade quanto a qualidade dos frutos (BARATLOO et al., 2015).

A implementação de estratégias de manejo integrado de doenças, incluindo o uso de variedades resistentes, práticas culturais adequadas e controle químico, é essencial para minimizar os danos causados por essas doenças. Além disso, a pesquisa contínua sobre novas variedades e métodos de controle é fundamental para garantir a sustentabilidade e a rentabilidade da produção de bananas (ASHRAF; KHAN, 2020).

2.5.1 Sigatoka Negra

De acordo com Paiva e Fioravanço (2005), a sigatoka negra, causada pelo fungo *Mycosphaerella fijiensis*, é uma das doenças mais devastadoras para a produção de bananas em várias partes do mundo. Ela afeta principalmente as folhas das bananeiras, comprometendo a fotossíntese e reduzindo tanto a qualidade quanto a quantidade da produção.

A doença é particularmente preocupante em regiões tropicais e subtropicais, onde a banana é uma das culturas mais importantes economicamente Paiva e Fioravanço (2005). A Figura 2.1 ilustra a ação da sigatoka negra na folha da bananeira.

Figura 2.1 – Sigatoka negra em folha de bananeira.



Fonte:AGROLINK (2025b).

A doença desenvolve-se nas folhas da bananeira, onde se espalha de maneira significativa. Ela é caracterizada por manchas escuras e necrosadas nas folhas, o que compromete diretamente a capacidade fotossintética da planta, impactando seu crescimento e sua produção (AGROLINK, 2025b).

Os primeiros sinais da sigatoka negra consistem em pequenas manchas de coloração amarelada ou verde-claro nas folhas. Estas manchas evoluem rapidamente para áreas

necróticas de cor marrom ou preta, com margens bem definidas. Com o avanço da doença, as manchas se coalescem, levando as folhas a tornarem-se completamente marrons ou escuras. A destruição das folhas mais velhas compromete diretamente o processo de fotossíntese, essencial para o desenvolvimento da planta e a produção de frutos (ZHANG et al., 2017).

Nos estágios mais avançados, as folhas infectadas tornam-se quebradiças e suscetíveis a danos causados pelo vento ou pela chuva. A desfolha precoce pode reduzir significativamente o rendimento da produção, visto que uma menor área foliar fotossintética resulta em menor quantidade e qualidade dos frutos (PAIVA; FIORAVANÇO, 2005).

De acordo com Paiva e Fioravanzo (2005), a transmissão da sigatoka negra ocorre principalmente por meio dos esporos do fungo, que são disseminados pelo vento e pela água da chuva. O fungo pode sobreviver em esporos presentes no solo ou em folhas infectadas e ser transportado por ferramentas agrícolas, roupas dos trabalhadores ou materiais contaminados. A água da chuva é um dos principais vetores para a dispersão do patógeno, especialmente em ambientes com alta umidade e precipitações frequentes.

O ciclo de vida do fungo está diretamente relacionado às condições climáticas. A doença se espalha mais rapidamente em ambientes com alta umidade e temperaturas elevadas, condições que favorecem a germinação dos esporos e a infecção das folhas, iniciando assim novos ciclos de contaminação (PAIVA; FIORAVANÇO, 2005).

A sigatoka negra é favorecida por ambientes quentes e úmidos. A alta umidade relativa do ar e as chuvas frequentes contribuem para a proliferação do fungo e a dispersão dos esporos. Além disso, a falta de ventilação e o adensamento das folhas nas plantações criam um microclima propício ao desenvolvimento da doença (PAIVA; FIORAVANÇO, 2005).

Segundo Dorgival (2024), a faixa de temperatura ideal para o desenvolvimento da doença varia entre 25°C e 30°C. Em regiões onde essas condições se mantêm durante todo o ano, o fungo pode se espalhar rapidamente, comprometendo grandes áreas de cultivo. O controle da sigatoka negra é desafiador, uma vez que o fungo apresenta alta capacidade de adaptação e resistência a diferentes métodos de controle. No entanto, uma combinação de estratégias pode ser eficaz no manejo da doença, incluindo práticas culturais e o uso criterioso de defensivos químicos (DORGIVAL, 2024).

O uso de fungicidas é uma das medidas mais comuns no controle da sigatoka negra. Tanto os fungicidas sistêmicos quanto os de contato são aplicados nas folhas das bananeiras para prevenir e controlar a infecção. Contudo, o uso excessivo desses produtos pode resultar na seleção de linhagens resistentes do fungo, dificultando o controle futuro.

As aplicações devem ser realizadas em intervalos regulares, especialmente durante o

período chuvoso, quando as condições para a disseminação dos esporos são mais favoráveis. É essencial que a pulverização atinja a face superior das folhas, local onde o fungo se desenvolve. O uso integrado de fungicidas, aliado a outras práticas de manejo, oferece os melhores resultados no controle da doença (PAIVA; FIORAVANÇO, 2005).

As práticas de controle cultural são fundamentais para reduzir a propagação do fungo e melhorar a sanidade das plantas. Entre essas práticas destacam-se: remoção e destruição de folhas infectadas, o que diminui a quantidade de esporos no ambiente; espaçamento adequado entre as plantas, favorecendo a circulação de ar e reduzindo a umidade; e rotação de culturas com espécies não hospedeiras, que contribui para a diminuição da presença do patógeno no solo (PAIVA; FIORAVANÇO, 2005).

O uso de variedades resistentes ou tolerantes à sigatoka negra é uma das estratégias mais eficazes no controle da doença a longo prazo. Nem todas as variedades de banana possuem a mesma suscetibilidade à doença, sendo que algumas apresentam resistência maior ao fungo, reduzindo assim seu impacto (DORGIVAL, 2024).

A pesquisa genética tem desempenhado papel crucial no desenvolvimento de variedades de banana com resistência duradoura à sigatoka negra. O impacto econômico da doença é expressivo, especialmente em países que dependem da banana como principal fonte de renda, tanto para consumo interno quanto para exportação. A queda na produtividade, aliada à redução da qualidade dos frutos, acarreta perdas econômicas significativas. Além disso, os custos adicionais com fungicidas e mão de obra para manejo fitossanitário elevam os custos de produção (DORGIVAL, 2024).

Os produtores de regiões afetadas enfrentam grandes desafios econômicos, particularmente em comunidades cuja subsistência depende fortemente da cultura da banana. A redução da oferta no mercado local pode gerar aumento no preço da fruta, afetando especialmente as populações de baixa renda (PAIVA; FIORAVANÇO, 2005).

De maneira geral, a sigatoka negra representa uma das principais ameaças às plantações de bananeira, exigindo um manejo integrado que combine práticas culturais, controle químico racional, uso de variedades resistentes e monitoramento constante. Conforme destacado por Dorgival (2024), frente ao significativo impacto da doença na produção de bananas, torna-se essencial que os produtores adotem medidas eficazes de manejo para mitigar os prejuízos e assegurar a sustentabilidade da atividade. Pesquisas contínuas e o desenvolvimento de novas tecnologias de controle serão determinantes para enfrentar esse desafio no longo prazo.

2.5.2 Sigatoka Amarela

De acordo com Ferrari e Nogueira (2018), a sigatoka amarela é uma doença causada pelo fungo *Mycosphaerella musicola*. Trata-se de uma fitopatologia relevante que afeta as

folhas das bananeiras, principalmente em regiões tropicais e subtropicais, onde a banana é uma das principais culturas agrícolas. Apesar de ser menos agressiva que a sigatoka negra (*Mycosphaerella fijiensis*), a sigatoka amarela representa uma ameaça significativa, comprometendo a sanidade das plantas e reduzindo a produtividade.

Figura 2.2 – Sigatoka amarela em folha de bananeira.



Fonte: AGROLINK (2025a).

O fungo ataca, principalmente, as folhas, provocando a redução da área foliar saudável e comprometendo a capacidade fotossintética da planta. Embora seja considerada menos devastadora do que a sigatoka negra, a sigatoka amarela pode causar severas perdas na produção, sobretudo quando as condições ambientais são favoráveis ao desenvolvimento do patógeno (AGROLINK, 2025a).

Os sintomas da doença iniciam-se com pequenas manchas de coloração amarelada ou verde-claro nas folhas. Com a progressão da infecção, essas manchas tornam-se mais evidentes, adquirindo coloração amarela intensa e margens bem definidas. Posteriormente, podem coalescer, formando grandes áreas de necrose. Nos estágios avançados, as folhas podem secar completamente, tornando-se quebradiças (FERRARI; NOGUEIRA, 2018).

A doença atinge predominantemente as folhas mais velhas, porém, em casos severos, pode também comprometer folhas mais jovens. A redução da área foliar ativa compromete diretamente o processo fotossintético, o que impacta negativamente o crescimento da

planta e a formação dos frutos. Quando o fungo atinge as folhas novas, os prejuízos à produção tendem a ser ainda mais expressivos.

A disseminação do fungo ocorre, principalmente, por meio dos esporos, que são transportados pelo vento, pela água da chuva ou pelo contato direto entre folhas infectadas e saudáveis. Os esporos possuem alta resistência, podendo permanecer viáveis no ambiente por longos períodos, o que favorece a persistência da doença nas áreas afetadas.

O ciclo de vida do fungo está diretamente condicionado a fatores climáticos. Seu desenvolvimento é acelerado em ambientes de alta umidade e temperaturas moderadas, típicos de regiões tropicais e subtropicais. O ciclo infeccioso inicia-se com a liberação de esporos, que germinam ao entrar em contato com as folhas, penetrando no tecido foliar. A infecção progride internamente, causando a formação de lesões e, posteriormente, necrose. O ciclo completo pode ser concluído em poucas semanas, permitindo rápida disseminação da doença em condições favoráveis (FERRARI; NOGUEIRA, 2018).

As condições ideais para o desenvolvimento da sigatoka amarela incluem temperaturas entre 25°C e 30°C e alta umidade relativa do ar. Segundo Ferrari e Nogueira (2018), a doença tende a se intensificar durante a estação chuvosa, quando há maior facilidade para a dispersão dos esporos. Plantios localizados em regiões com clima úmido e quente estão mais suscetíveis à infecção.

Além disso, o microclima da plantação exerce influência direta no desenvolvimento do patógeno. Fatores como sombreamento excessivo e alta densidade de plantas elevam a umidade nas folhas, criando um ambiente favorável à proliferação do fungo. Plantios mal ventilados e com acúmulo de material vegetal, como folhas secas, apresentam maior vulnerabilidade à doença.

O controle da sigatoka amarela é essencial para minimizar as perdas na produção. A abordagem mais eficiente envolve o manejo integrado, combinando práticas químicas, culturais e, quando possível, biológicas. O controle químico é realizado pela aplicação de fungicidas, tanto de contato quanto sistêmicos, aplicados diretamente nas folhas, especialmente nas áreas mais suscetíveis à infecção. A periodicidade das aplicações varia entre 7 e 15 dias, dependendo da severidade da doença e das condições climáticas (FERRARI; NOGUEIRA, 2018).

Entretanto, o uso intensivo de fungicidas requer cautela, devido ao risco de desenvolvimento de resistência do fungo. A rotação de princípios ativos é uma prática recomendada, visando prolongar a eficácia dos produtos e reduzir a pressão seletiva sobre o patógeno.

O controle cultural, por sua vez, engloba práticas que visam reduzir as condições favoráveis ao desenvolvimento da doença. Entre elas, destacam-se: a remoção e destruição de folhas infectadas, o ajuste do espaçamento entre plantas para melhorar a ventilação, e a adoção de sistemas de irrigação que não aumentem a umidade foliar, como a irri-

gação por gotejamento. Conforme Ferrari e Nogueira (2018), essas medidas contribuem significativamente para a redução da incidência da sigatoka amarela.

Outra estratégia importante é a utilização de variedades de banana com maior resistência ou tolerância à doença. Embora o desenvolvimento de cultivares completamente resistentes ainda seja um desafio, algumas variedades apresentam melhor desempenho frente à infecção, contribuindo para a redução dos danos e dos custos com controle químico.

A pesquisa no campo do melhoramento genético tem se mostrado fundamental para o desenvolvimento de soluções sustentáveis no controle da sigatoka amarela. Nesse contexto, o manejo integrado de doenças, que combina práticas culturais, aplicação racional de fungicidas, uso de variedades resistentes e monitoramento constante, constitui-se como a estratégia mais eficaz para controlar a doença de forma sustentável (FERRARI; NOGUEIRA, 2018).

Do ponto de vista econômico, a sigatoka amarela pode gerar impactos significativos na produção de bananas. A redução da área foliar compromete a fotossíntese, afetando o desenvolvimento dos cachos e a qualidade dos frutos. Além das perdas diretas na produção, há o aumento dos custos operacionais devido à necessidade de aquisição de fungicidas e ao trabalho intensivo para remoção de folhas doentes.

Esses custos podem ser particularmente onerosos para pequenos produtores, que muitas vezes dispõem de menos recursos para investir em tecnologias e práticas de manejo eficiente. Dessa forma, mesmo sendo menos agressiva que a sigatoka negra, a sigatoka amarela continua sendo uma ameaça relevante às lavouras de banana, especialmente em regiões tropicais e subtropicais.

Portanto, o controle eficaz da sigatoka amarela depende da adoção de estratégias integradas, que conciliem práticas culturais adequadas, uso criterioso de fungicidas, emprego de cultivares mais resistentes e constante vigilância fitossanitária. Investir em pesquisa e desenvolvimento de novas tecnologias será essencial para garantir a sustentabilidade e a produtividade das lavouras no enfrentamento dessa doença (FERRARI; NOGUEIRA, 2018).

2.5.3 Mal-do-Panamá

De acordo com Araújo et al. (2018), o mal-do-panamá, também conhecido como *Fusarium wilt*, é uma das doenças mais devastadoras que afeta a produção de bananas em diversas partes do mundo. Causado pelo fungo *Fusarium oxysporum f.sp. cubense*, esse patógeno tem provocado grandes prejuízos econômicos, principalmente em regiões onde a banana é uma das principais culturas agrícolas.

Essa doença é caracterizada pela murcha e pela morte das plantas, afetando principalmente o sistema vascular da bananeira. O comprometimento dos vasos dificulta o

Figura 2.3 – Mal-do-Panamá em folha de bananeira.



Fonte: BRASIL, Secretaria do Desenvolvimento Agrário (2021).

transporte de água e nutrientes, levando ao enfraquecimento e, eventualmente, à morte da planta (BRASIL, Secretaria do Desenvolvimento Agrário, 2021). O fungo é um patógeno vascular que coloniza os vasos de xilema, responsáveis pelo transporte de seiva bruta. Sua capacidade de permanecer no solo por longos períodos, mesmo na ausência de hospedeiros, torna seu controle extremamente difícil (NETO; GUIMARÃES, 2011).

O processo de infecção inicia-se nas raízes, onde o fungo penetra e alcança o sistema vascular. A partir desse ponto, ele bloqueia o fluxo de água e nutrientes, causando murcha e, nos casos mais severos, a morte da planta, impactando diretamente a produtividade da cultura (NETO; GUIMARÃES, 2011).

Os sintomas variam conforme o estágio da doença, mas, geralmente, incluem murchamento, amarelecimento e morte prematura das folhas. O primeiro sinal típico é a murcha progressiva de uma ou mais folhas, começando pelas extremidades, seguida pelo amarelecimento. Com a progressão da infecção, observa-se a queda das folhas e o enfraquecimento geral da planta (NETO; GUIMARÃES, 2011).

Nos estágios avançados, o fungo pode se espalhar para plantas vizinhas, ocasionando o colapso total das bananeiras afetadas. Muitas vezes, as plantas infectadas não chegam a produzir frutos devido ao comprometimento do sistema vascular. Em casos severos, observa-se o colapso estrutural, com o caule tornando-se frouxo e as folhas permanecendo murchas e secas.

Um dos sinais mais característicos é a descoloração dos vasos de xilema, tanto nas

raízes quanto na base do caule, que adquirem tonalidades marrons ou avermelhadas - um indicativo claro da presença do patógeno (NETO; GUIMARÃES, 2011).

O ciclo de vida do *Fusarium oxysporum* é fortemente influenciado por fatores ambientais, especialmente umidade e temperatura. Na forma de esporos, o fungo é capaz de sobreviver no solo por vários anos. Esses esporos podem ser disseminados por meio de ferramentas agrícolas contaminadas, equipamentos, mudas, fragmentos de raízes, animais e até trabalhadores que circulam em áreas infectadas (NETO; GUIMARÃES, 2011).

A disseminação é favorecida por condições de solo úmido e temperaturas elevadas. O patógeno penetra nas raízes, geralmente através de ferimentos ou fissuras, e, uma vez no interior da planta, coloniza os vasos de xilema. Isso provoca o bloqueio do transporte de seiva, levando rapidamente ao murchamento e ao declínio da planta.

Em sistemas de produção comercial, a doença se espalha com maior rapidez quando há utilização de mudas infectadas ou movimentação de solo contaminado. O transporte de frutos e materiais agrícolas também representa uma via de disseminação para novas áreas (NETO; GUIMARÃES, 2011).

As condições climáticas quentes e úmidas são ideais para o desenvolvimento do mal-do-panamá. O fungo apresenta crescimento ótimo em temperaturas entre 25°C e 30°C. Além disso, solos bem drenados, com pH neutro a levemente ácido, favorecem sua proliferação (NETO; GUIMARÃES, 2011).

Nas regiões tropicais e subtropicais, onde essas condições climáticas são comuns, o fungo pode se espalhar rapidamente entre as plantações, especialmente quando as práticas de manejo do solo são inadequadas.

A movimentação de solos contaminados, prática comum em regiões de produção extensiva, contribui significativamente para a expansão da doença. Por esse motivo, o controle do mal-do-panamá deve priorizar a adoção de medidas preventivas, como o uso de mudas certificadas e livres de patógenos.

O manejo da doença é um desafio significativo. Isso se deve à capacidade do fungo de permanecer viável no solo e nas raízes por longos períodos. O controle eficaz demanda uma abordagem integrada, que envolve práticas culturais, medidas químicas, controle biológico e, sobretudo, o uso de cultivares resistentes.

O controle químico, apesar de utilizado, apresenta eficácia limitada, uma vez que o fungo se desenvolve internamente nos vasos da planta e no solo, dificultando a ação dos fungicidas. Além disso, o uso indiscriminado de produtos químicos pode gerar impactos ambientais. Entretanto, a desinfecção de ferramentas e equipamentos agrícolas por meio de agentes químicos é uma prática recomendada para reduzir a disseminação do patógeno.

A adoção de variedades resistentes tem se mostrado a estratégia mais eficaz no

longo prazo. O desenvolvimento de cultivares geneticamente resistentes ou tolerantes representa uma solução promissora. Historicamente, variedades como a Gros Michel foram dizimadas pela doença, levando à substituição por variedades mais resistentes, como a Cavendish.

O impacto econômico do mal-do-panamá nas regiões produtoras de banana é extremamente significativo. A doença reduz drasticamente a produção e compromete a qualidade dos frutos. A morte precoce das plantas e a conseqüente queda na produtividade geram grandes prejuízos, podendo levar ao abandono de áreas de cultivo e à redução da oferta de bananas no mercado (NETO; GUIMARÃES, 2011).

Além disso, os elevados custos com controle e manejo dificultam a sustentabilidade econômica dos produtores. Em países que dependem da exportação de bananas, a presença da doença pode impactar negativamente o comércio internacional, uma vez que muitos mercados impõem restrições sanitárias rigorosas.

2.5.4 Cordana

De acordo com Mendes (2015), a doença conhecida como cordona (Figura 2.4), também chamada de "ferrugem das bananeiras" ou *Pseudocercospora fijiensis*, é um dos principais problemas fitossanitários que afetam a cultura da banana. Trata-se de uma das enfermidades mais prejudiciais à produção de bananas em diversas partes do mundo, especialmente em regiões tropicais.

Segundo Neto e Guimarães (2011), a infecção manifesta-se principalmente nas folhas da planta, podendo ocasionar sérias perdas na produtividade e na qualidade dos frutos. O patógeno dissemina-se, sobretudo, por meio de esporos, os quais podem ser transportados pelo vento, pela água da chuva, por ferramentas agrícolas, pelo contato com plantas infectadas e até pela movimentação de solos contaminados.

O fungo penetra nas folhas da bananeira, colonizando as células e causando danos significativos ao tecido vegetal. Pertencente à família *Mycosphaerellaceae*, esse patógeno afeta predominantemente as folhas, embora, em estágios mais avançados, possa comprometer também a produção de frutos (PLANTIX, 2025).

Após atingirem as folhas, os esporos do fungo estabelecem-se nas estruturas celulares, desenvolvendo-se na forma de manchas e lesões que se expandem, comprometendo progressivamente a saúde da planta. Com o avanço da infecção, a capacidade fotossintética da planta diminui, afetando diretamente seu crescimento e sua produtividade.

Os primeiros sintomas da cordona manifestam-se na face superior das folhas, na forma de manchas pequenas, angulosas, de coloração amarelada a castanho-claro, cercadas por uma borda de tonalidade mais escura. Com o tempo, essas manchas aumentam de tamanho e podem se expandir por toda a lâmina foliar (MENDES, 2015).

Figura 2.4 – Cordana em folha de bananeira.



Fonte: PLANTIX (2025).

À medida que a doença progride, as lesões tornam-se mais evidentes, levando ao amarelecimento e posterior necrose das folhas, que acabam caindo. Nos estágios mais avançados, as plantas acometidas apresentam desfolhamento acentuado, redução no número de folhas funcionais e enfraquecimento estrutural. A carência de folhas saudáveis compromete o processo de fotossíntese, prejudicando tanto o desenvolvimento quanto a capacidade produtiva da bananeira. Além disso, a perda de vigor torna a planta mais suscetível a outras doenças e a estresses ambientais (NETO; GUIMARÃES, 2011).

O ciclo de vida do fungo causador da cordona é altamente adaptado a ambientes tropicais, nos quais a umidade elevada e as temperaturas altas favorecem seu desenvolvimento. Os esporos são disseminados pelo vento, pela chuva, por ferramentas agrícolas contaminadas e por qualquer prática que envolva material infectado.

A infecção tem início quando os esporos entram em contato com a superfície foliar. Uma vez na folha, eles germinam e penetram nas células vegetais, propagando-se pelo tecido foliar, onde causam danos ao sistema vascular, dificultando o transporte de água e nutrientes. Esse ciclo é acelerado em condições de alta umidade e temperaturas amenas, que favorecem tanto a germinação dos esporos quanto a rápida disseminação da doença

(MENDES, 2015).

O ciclo completo do fungo pode durar de uma a duas semanas, dependendo das condições ambientais. Em regiões com alta umidade e chuvas frequentes, a disseminação ocorre de forma rápida e pode gerar infecções generalizadas nas plantações de banana.

A cordona prospera em ambientes quentes e úmidos, características típicas das regiões tropicais e subtropicais onde a banana é cultivada. A umidade relativa elevada, especialmente durante o período chuvoso, cria condições ideais para a propagação e o desenvolvimento do *Pseudocercospora fijiensis*. As temperaturas entre 24°C e 30°C são as mais favoráveis ao desenvolvimento do fungo, embora ele consiga sobreviver, com menor agressividade, em condições térmicas fora desse intervalo.

Fatores como irrigação inadequada também favorecem o agravamento da doença, uma vez que aumentam a umidade ao redor das plantas, criando um ambiente propício para o patógeno. Da mesma forma, a presença de plantas vizinhas infectadas e a movimentação de materiais contaminados, como ferramentas e mudas, contribuem para a disseminação da doença para novas áreas.

Práticas de manejo inadequadas, como o uso de mudas contaminadas, o cultivo de variedades suscetíveis e a ausência de rotação de culturas, agravam ainda mais o problema. A doença tende a se espalhar rapidamente em áreas com alta densidade de plantio e pouca ventilação entre as plantas (NETO; GUIMARÃES, 2011).

A adoção de boas práticas culturais é essencial para o manejo da cordona. Entre elas, destacam-se a rotação de culturas, que ajuda a reduzir a presença do fungo no solo, e a eliminação de folhas infectadas, associada à realização de podas adequadas, o que limita a propagação do patógeno. Além disso, manter espaçamentos adequados entre as plantas contribui para melhorar a circulação de ar e reduzir a umidade, dificultando o desenvolvimento do fungo.

O controle químico, por meio do uso de fungicidas, é uma das estratégias mais empregadas no combate à cordona. Fungicidas à base de cobre e produtos sistêmicos são frequentemente utilizados para reduzir a incidência do fungo nas folhas. No entanto, é fundamental que a aplicação desses produtos ocorra de forma preventiva e que haja rotação entre diferentes princípios ativos, a fim de evitar o desenvolvimento de resistência por parte do patógeno.

O uso de variedades de banana resistentes ao fungo representa uma das estratégias mais eficazes no longo prazo. Atualmente, já existem variedades geneticamente modificadas ou resultantes de melhoramento genético que apresentam menor suscetibilidade ao *Pseudocercospora fijiensis*. O desenvolvimento contínuo de novas variedades resistentes é uma área de grande relevância, sobretudo em regiões onde a doença é endêmica e causa prejuízos significativos à produção.

Os impactos econômicos da cordona são expressivos nas regiões produtoras de banana. A redução na qualidade e na quantidade dos frutos, a morte prematura das plantas e os custos associados ao uso constante de fungicidas elevam significativamente os custos de produção, afetando diretamente a rentabilidade dos produtores. Esse cenário é ainda mais desafiador para pequenos produtores, sobretudo em países em desenvolvimento, onde a banana representa uma cultura de elevada importância social e econômica. As perdas econômicas associadas à cordona podem ser devastadoras em locais onde a produção de banana é uma das principais atividades agrícolas.

3 Proposta de Modelo Convolutacional

Para realizar a detecção de doenças em folhas de bananeira, foram aplicadas técnicas de extração de características por meio da criação e do treinamento de modelos de CNNs. Esta seção está organizada em cinco subseções: a descrição do conjunto de dados utilizado para o treinamento dos modelos CNNs (Subseção 3.1); os procedimentos de preparação e pré-processamento dos dados (Subseção 3.2); os modelos convolucionais para extração de características e classificação das doenças foliares (Subseção 3.3); as configurações dos experimentos realizados (Subseção 3.4); e, por fim, as métricas adotadas para avaliação dos modelos (Subseção 3.5).

3.1 Conjunto de Dados

Na etapa inicial da proposta metodológica, foram analisados separadamente os conjuntos de dados disponibilizados por (HAILU, 2021; MEDHI; DEB, 2022; ARMAN et al., 2023; SILVA et al., 2024). Contudo, observou-se que esses conjuntos, quando utilizados isoladamente, apresentavam um número reduzido de amostras, o que comprometeu a capacidade de generalização do modelo convolutacional proposto.

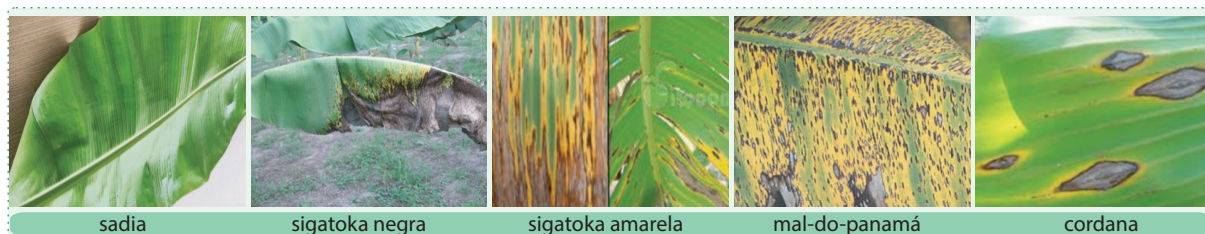
Conforme destacado por Dogan (2024), Chen et al. (2024), conjuntos de dados com quantidade limitada de amostras dificultam que CNNs aprendam padrões robustos e generalizáveis. Essa limitação resulta em baixo desempenho na etapa de validação, especialmente quando o modelo é aplicado a dados reais que não pertencem ao conjunto de dados de treinamento.

Diante desse cenário, visando aumentar a robustez do modelo e mitigar os efeitos de sobreajuste, optou-se por consolidar todos os conjuntos de dados em um único conjunto ampliado, agregando todas as amostras disponíveis nas fontes citadas.

O conjunto de dados resultante contempla cinco classes, correspondentes às seguintes condições: folhas de bananeira sadias; folhas infectadas por sigatoka-negra; folhas infectadas por sigatoka-amarela; folhas afetadas por mal-do-panamá; e folhas infectadas por cordana (Figura 3.1).

A Figura 3.1 apresenta uma amostra representativa do conjunto de dados consolidado, que contém um total de 4569 imagens distribuídas nas seguintes classes: folhas sadias (981 amostras), sigatoka negra (948 amostras), sigatoka amarela (941 amostras), mal-do-panamá (899 amostras) e cordana (800 amostras).

Figura 3.1 – Amostras do conjunto de dados utilizado na proposta metodológica.



3.2 Preparação e Pré-Processamento

O pré-processamento de imagens desempenha um papel crucial no desenvolvimento de modelos de visão computacional, sendo responsável por melhorar a qualidade dos dados e garantir a extração eficiente de características relevantes (SELVARAJ; SIVAPRAKASH, 2025; CHEN et al., 2024). Assim, o conjunto de dados utilizado nesta proposta foi submetido a procedimentos padronizados, com o objetivo de reduzir ruídos, normalizar informações visuais e otimizar a entrada para o treinamento do modelo de CNN.

Inicialmente, as imagens foram organizadas e rotuladas, categorizando-as em duas classes: folhas saudáveis e folhas infectadas por diferentes doenças. Contudo, mesmo após a junção dos conjuntos de dados disponíveis na literatura (HAILU, 2021; MEDHI; DEB, 2022; ARMAN et al., 2023; SILVA et al., 2024), identificou-se um desequilíbrio significativo, especialmente na classe de folhas saudáveis, que apresentava uma quantidade reduzida de amostras. Para mitigar esse problema, foi realizada uma coleta adicional em campo, visando aumentar a representatividade dessa classe.

Durante a coleta, foram utilizados diferentes dispositivos, incluindo câmeras de alta resolução e *smartphones*, de forma a avaliar a robustez do modelo frente à variabilidade dos equipamentos. Este processo considerou diversos fatores, como condições de iluminação (natural e artificial), variações de ângulo, distância de captura e ambientes heterogêneos, de modo a garantir diversidade e generalização do modelo. Todas as imagens coletadas foram devidamente organizadas, armazenadas e anotadas com informações auxiliares, como localização geográfica, data, condições ambientais e confirmação visual da ausência de doenças (folhas saudáveis).

Após esta etapa, aplicou-se uma estratégia de aumento artificial de dados, conforme as diretrizes propostas em Shi et al. (2022). Essa abordagem tem se mostrado fundamental para mitigar o problema de sobreajuste em CNNs, especialmente quando há escassez de dados (XIA et al., 2024; SHI et al., 2022). O desbalanceamento entre as classes, se não tratado, compromete diretamente a capacidade de generalização do modelo (XIA et al., 2024).

Dentre as técnicas de aumento de dados aplicadas, destacam-se as seguintes transformações:

- **Rotação:** permite simular diferentes ângulos de captura da folha. A rotação de um ponto (x, y) em torno da origem por um ângulo θ é definida pela seguinte matriz de transformação:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} \quad (3.1)$$

- **Translação:** desloca a imagem horizontal e/ou verticalmente, simulando mudanças de enquadramento. A operação é representada por:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} x + t_x \\ y + t_y \end{bmatrix} \quad (3.2)$$

onde t_x e t_y correspondem aos deslocamentos nos eixos x e y , respectivamente.

- **Espelhamento (Reflexão Horizontal ou Vertical):** simula inversões da folha na imagem. Para um espelhamento horizontal, a transformação é expressa como:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} \quad (3.3)$$

- **Escala (Zoom):** altera a escala da imagem, simulando aproximações ou distanciamentos da câmera. A matriz de transformação é dada por:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} s_x & 0 \\ 0 & s_y \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} \quad (3.4)$$

onde s_x e s_y são os fatores de escalonamento nos eixos horizontal e vertical, respectivamente.

- **Perturbações Aleatórias:** foram aplicadas pequenas variações no brilho, contraste e ruído, para simular diferentes condições ambientais. A adição de ruído gaussiano, por exemplo, é modelada por:

$$I'(x, y) = I(x, y) + \mathcal{N}(0, \sigma^2) \quad (3.5)$$

onde $I(x, y)$ representa a intensidade do pixel na posição (x, y) e $\mathcal{N}(0, \sigma^2)$ é uma variável aleatória com distribuição normal de média zero e variância σ^2 .

Essas transformações visam não apenas aumentar a quantidade de dados, mas também promover uma maior variabilidade intra-classe, fator essencial para garantir que o modelo aprenda características robustas e generalizáveis. Como resultado, o conjunto de dados final tornou-se mais balanceado e resiliente, conforme detalhado na Tabela 3.1.

Tabela 3.1 – Distribuição das classes no conjunto de dados após aplicação de aumento de dados (Data Augmentation).

Classe	Quantidade de Amostras
folhas saudias	4.905
sigatoka negra	4.740
sigatoka amarela	4.705
mal-do-panamá	4.495
cordana	4.000
Total	22.845

3.3 Modelos Convolutacionais

Os modelos utilizados nesta pesquisa são compostos por uma base convolutacional responsável pela extração de características e por uma camada de classificação, que realiza a inferência final sobre as classes. Para o desenvolvimento e avaliação da proposta, foram adotadas 03 (três) arquiteturas distintas, representativas de diferentes paradigmas no campo da visão computacional: (a) LeNet (subseção 3.3.1); (b) Vision Transformer (ViT) (subseção 3.3.2); e (c) HibridNet (subseção 3.3.3).

Essas arquiteturas CNNs foram selecionadas por sua comprovada eficácia em tarefas de classificação de imagens, reconhecimento de objetos e extração automática de padrões espaciais (LI et al., 2024; HUSSAIN et al., 2025). As CNNs são capazes de modelar relações espaciais locais presentes nas imagens, extraindo características hierárquicas que são fundamentais para a tarefa de diagnóstico visual.

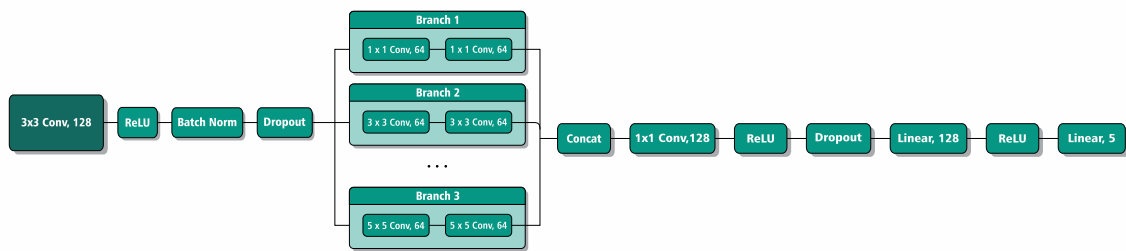
3.3.1 Arquitetura Adpatada da LeNet-5

A primeira arquitetura implementada neste estudo foi uma variação moderna inspirada na clássica CNN LeNet-5, originalmente proposta por (LECUN et al., 1998). Embora mantenha os princípios fundamentais do modelo original como o uso de camadas convolutacionais seguidas por camadas totalmente conectadas, a versão utilizada neste trabalho incorpora uma série de melhorias.

As principais alterações incluem o aumento da profundidade da rede, a introdução de camadas de normalização por lote (*Batch Normalization*) e a aplicação de *Dropout* como técnica de regularização. Além disso, a arquitetura proposta incorpora uma estrutura em ramificações (*branches*), que permite a extração simultânea de características com diferentes tamanhos de *kernel* (1×1 , 3×3 e 5×5), promovendo uma extração de padrões mais rica e diversificada. Essa abordagem remete à ideia de módulos Inception utilizados em redes como a GoogLeNet, mas mantendo uma estrutura geral próxima ao fluxo de dados da LeNet (Figura 3.2).

Dessa forma, embora a arquitetura desenvolvida não represente uma reprodução fiel da LeNet-5 original, ela preserva sua filosofia estrutural e serve como uma versão

Figura 3.2 – Arquitetura adaptada baseada na LeNet-5.



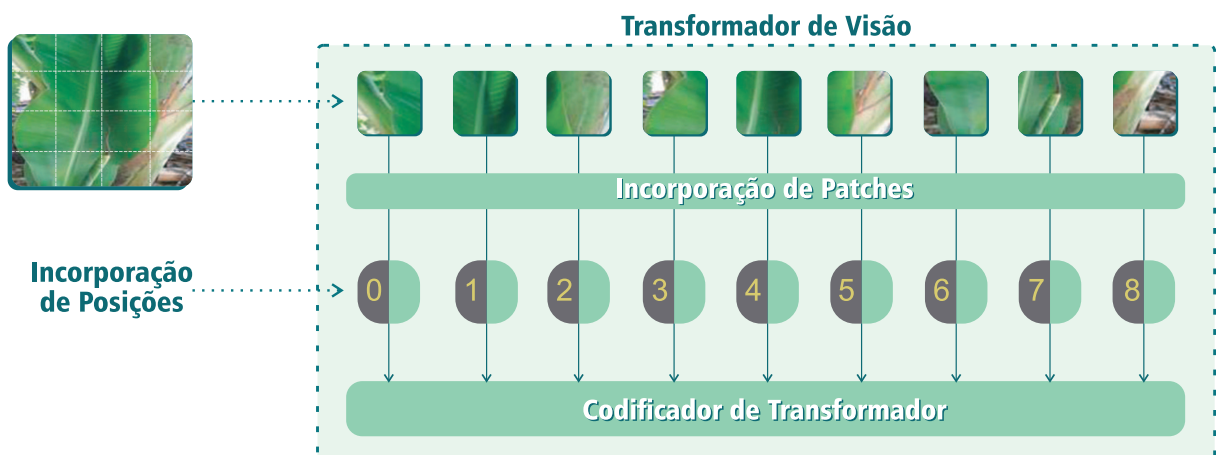
aprimorada, com maior capacidade de generalização e desempenho em conjuntos de dados mais complexos.

3.3.2 Arquitetura ViT

A detecção de doenças em folhas de bananeira é uma tarefa desafiadora, especialmente pela necessidade de identificar padrões visuais sutis e complexos. Embora as CNNs sejam tradicionalmente empregadas em tarefas de classificação de imagens, avanços recentes na área de IA destacam os modelos baseados em transformadores, como o ViT, pela sua eficácia na análise de dados visuais (DOSOVITSKIY et al., 2020).

O ViT adapta o mecanismo de atenção, originalmente desenvolvido para NPL, para o domínio da visão computacional. Diferentemente das CNNs, que dependem de operações de convolução para a extração de características locais, o ViT utiliza mecanismos de autoatenção para capturar relações espaciais globais entre diferentes regiões da imagem (ZHAO et al., 2024; ZHOU et al., 2024). A Figura 3.3 ilustra a arquitetura básica do ViT.

Figura 3.3 – Arquitetura básica da vision transformer (ViT).



Na abordagem ViT, a imagem de entrada é inicialmente dividida em blocos (do inglês *patches*) de tamanho fixo, que são posteriormente linearizados e transformados

em vetores de características (*tokens*). A esses *tokens* são adicionadas informações de codificação posicional, fundamentais para preservar a ordem espacial das partes da imagem. Em seguida, os *tokens* são processados por múltiplas camadas de atenção e de propagação (do inglês *feedforward*), permitindo a captura de padrões visuais em diferentes escalas e contextos. Por fim, uma camada de classificação realiza a predição da classe da imagem (DOSOVITSKIY et al., 2020).

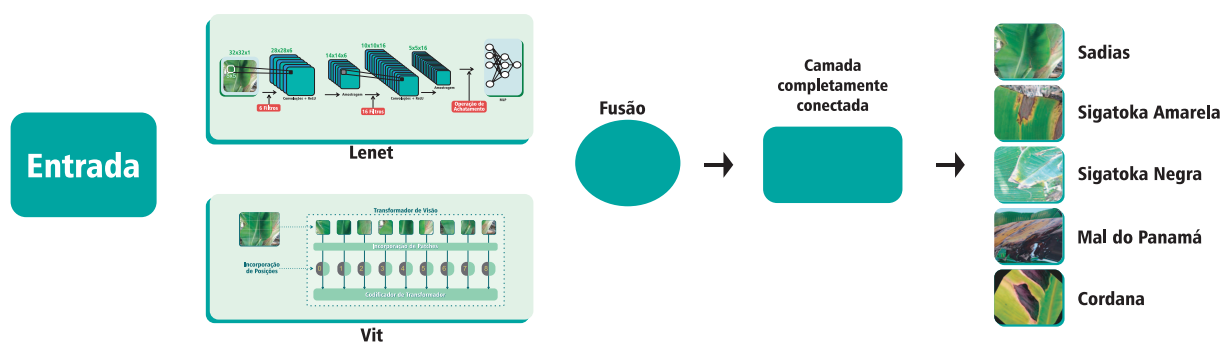
Para este estudo, aplicou-se o ViT na classificação de folhas de bananeira entre saudias e doentes, com o objetivo de identificar de forma precisa os padrões associados às patologias. O *pipeline* de processamento incluiu as etapas de coleta das imagens, pré-processamento - envolvendo redimensionamento, normalização e aumento dos dados por meio de transformações como rotação, espelhamento e ajustes de brilho -, visando aumentar a robustez do modelo e mitigar problemas de sobreajuste.

Embora o ViT demande maior capacidade computacional e volume de dados para treinamento em comparação às CNNs tradicionais (RAMOS; SAPPA, 2024; ZHU et al., 2023), seus resultados destacam seu potencial na análise agrícola de imagens, contribuindo para o monitoramento eficiente e em larga escala das plantações.

3.3.3 Arquitetura HibridNet

Entre os modelos utilizados, a HibridNet destaca-se como uma arquitetura híbrida, que combina diferentes abordagens para aprimorar a precisão na classificação de folhas de bananeira saudáveis e com doenças. Sua principal característica está na capacidade de extrair características tanto locais quanto globais das imagens, utilizando uma integração eficiente entre CNNs e mecanismos de atenção.

Figura 3.4 – Arquitetura proposta denominada HibridNet.



A HibridNet foi projetada para aproveitar o melhor das CNNs - que são altamente eficazes na extração de padrões espaciais locais - juntamente com módulos baseados em atenção, que capturam dependências e relações globais nas imagens. Dessa forma, a arquitetura torna-se mais robusta na identificação de padrões sutis, muitas vezes desafiadores em imagens de folhas acometidas por doenças.

O desenvolvimento desse modelo considerou a necessidade de uma classificação precisa e eficiente, capaz de diferenciar folhas saudáveis daquelas que apresentam sintomas patológicos. Para isso, foi construído um banco de dados específico, contendo imagens devidamente rotuladas, seguido de um processo de pré-processamento que incluiu redimensionamento, normalização dos *pixels* e aumento dos dados através de técnicas como rotação, espelhamento e ajuste de brilho, a fim de melhorar a generalização do modelo e mitigar problemas de sobreajustes.

Portanto, a HibridNet propõe uma evolução nas arquiteturas para classificação de imagens agrícolas, ao combinar o poder das CNNs com os mecanismos de atenção, objetivando análises mais precisas, rápidas e escaláveis no diagnóstico de doenças foliares em bananeiras.

3.4 Configurações dos Experimentos

Para a realização dos experimentos, foram utilizados dois modelos de referência amplamente conhecidos na literatura: o LeNet (LECUN et al., 2002) e o ViT (DOSOVITS-KIY et al., 2020). A partir dos resultados obtidos, foi possível realizar uma comparação com o modelo proposto neste trabalho, denominado HibridNet, que combina características das arquiteturas LeNet e ViT.

Os modelos foram treinados utilizando uma divisão dos dados em 70% para treinamento, 15% para validação e 15% para teste. Esse particionamento visa garantir a avaliação adequada da capacidade de generalização dos modelos.

Na Tabela 3.2 estão apresentados os hiperparâmetros utilizados durante os experimentos.

Tabela 3.2 – Hiperparâmetros utilizados nos experimentos

Hiperparâmetro	Modelo	Valor
Épocas de treinamento	LeNet	60
	ViT	30
	HibridNet	21
Tamanho do lote (<i>batch size</i>)	-	32
Taxa de <i>Dropout</i>	-	0,4
Dimensão da entrada	-	(225, 225, 3)

3.5 Métricas de Avaliação dos Modelos

Para avaliar o desempenho dos modelos, foram utilizadas diversas métricas, cada uma com suas especificidades e contribuições. Essas métricas permitiram uma análise detalhada das arquiteturas aplicadas, assegurando um desempenho confiável na correta classificação entre folhas saudáveis e doentes.

Essas métricas são fundamentais para medir a eficácia e a precisão dos modelos, auxiliando na identificação de erros, viés e na capacidade de generalização das CNNs (LeNet, ViT e HibridNet).

3.5.1 Acurácia (ACC)

A acurácia é uma das métricas mais utilizadas e representa a proporção de previsões corretas em relação ao total de previsões realizadas. Ela considera:

- **TP (*True Positives*)** - folhas doentes corretamente classificadas como doentes;
- **TN (*True Negatives*)** - folhas saudáveis corretamente classificadas como saudáveis;
- **FP (*False Positives*)** - folhas saudáveis incorretamente classificadas como doentes;
- **FN (*False Negatives*)** - folhas doentes incorretamente classificadas como saudáveis.

A fórmula para calcular a acurácia é apresentada na Equação 3.6:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.6)$$

Embora seja de fácil interpretação, a acurácia pode ser pouco representativa em cenários com desequilíbrio entre as classes.

3.5.2 F1-Score (F1)

O F1-Score foi uma métrica essencial na avaliação dos modelos híbridos (HibridNet), especialmente devido ao desbalanceamento das classes, onde uma das categorias possui mais amostras do que a outra.

O F1-Score é a média harmônica entre Precisão e Lembrança, equilibrando os erros de falso positivo e falso negativo. Essa métrica é particularmente relevante em aplicações onde tanto a correta detecção das folhas doentes quanto a não classificação incorreta das folhas saudáveis são cruciais.

A equação que define o F1-Score é apresentada na Equação 3.7:

$$F1 = 2 \cdot \frac{\text{Precisão} \cdot \text{Lembrança}}{\text{Precisão} + \text{Lembrança}} \quad (3.7)$$

A Precisão e a Lembrança são calculadas pelas Equações 3.8 e 3.9, respectivamente:

$$\text{Precisão} = \frac{TP}{TP + FP} \quad (3.8)$$

$$\text{Lembrança} = \frac{TP}{TP + FN} \quad (3.9)$$

A precisão reflete a proporção de verdadeiros positivos entre todas as previsões positivas feitas pelo modelo, enquanto a Lembrança mede a capacidade do modelo em identificar corretamente todos os positivos existentes.

3.5.3 Área Sob a Curva (AUC)

A métrica AUC foi empregada para avaliar a capacidade dos modelos híbridos em distinguir corretamente folhas saudáveis das doentes. Ela mede a performance do classificador ao gerar a curva ROC (do inglês *Receiver Operating Characteristic* - ROC).

Para seu entendimento, utilizam-se dois conceitos principais:

- **Taxa de Verdadeiros Positivos (TPR ou Recall)** - proporção de positivos corretamente identificados:

$$\text{TPR} = \frac{TP}{TP + FN} \quad (3.10)$$

- **Taxa de Falsos Positivos (FPR)** - proporção de negativos incorretamente classificados como positivos:

$$\text{FPR} = \frac{FP}{FP + TN} \quad (3.11)$$

O AUC varia entre 0 e 1, onde valores próximos de 1 indicam excelente capacidade de separação entre classes, enquanto valores próximos de 0,5 indicam desempenho equivalente a um classificador aleatório.

3.5.4 Erro Médio Quadrático (MSE)

O MSE foi utilizado para quantificar a diferença média ao quadrado entre as previsões do modelo e os valores reais. O MSE penaliza mais fortemente os erros maiores, tornando-o sensível a *outliers* e eficaz para ajustar modelos que trabalham com saídas probabilísticas, comuns em CNNs.

A equação que define o MSE é:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3.12)$$

Onde:

- y_i é o valor real;
- \hat{y}_i é o valor previsto;
- n é o número total de amostras.

O MSE é também amplamente utilizado na etapa de treinamento, especialmente como função de perda para modelos de regressão e CNNs, devido à sua propriedade de suavizar gradientes.

3.5.5 Regularização L1 e L2

A regularização é uma técnica fundamental para reduzir o sobreajuste, evitando que os pesos das CNNs assumam valores excessivamente altos. As regularizações L1 e L2 foram aplicadas nas camadas densas da rede, contribuindo para melhorar a capacidade de generalização do modelo, tanto nos dados de treinamento quanto em novos dados.

A regularização L1 adiciona a soma dos valores absolutos dos pesos à função de perda, promovendo a esparsidade dos pesos (alguns se tornam zero), sendo definida pela Equação 3.13:

$$L1 = \lambda \sum_{i=1}^n |w_i| \quad (3.13)$$

A função de perda total com L1 é expressa por:

$$Loss_{total} = Loss_{original} + \lambda \sum_{i=1}^n |w_i| \quad (3.14)$$

A regularização L2 adiciona a soma dos quadrados dos pesos, ajudando a suavizar os pesos sem zerá-los, conforme a Equação 3.15:

$$L2 = \lambda \sum_{i=1}^n w_i^2 \quad (3.15)$$

Sua função de perda total é:

$$Loss_{total} = Loss_{original} + \lambda \sum_{i=1}^n w_i^2 \quad (3.16)$$

A aplicação dessas técnicas são essenciais para garantir que os modelos desenvolvidos (LeNet, ViT e HibridNet) mantenham um equilíbrio entre boa performance no treinamento e uma adequada capacidade de generalização no teste, evitando possivelmente os sobreajustes.

4 Resultados e Discussões

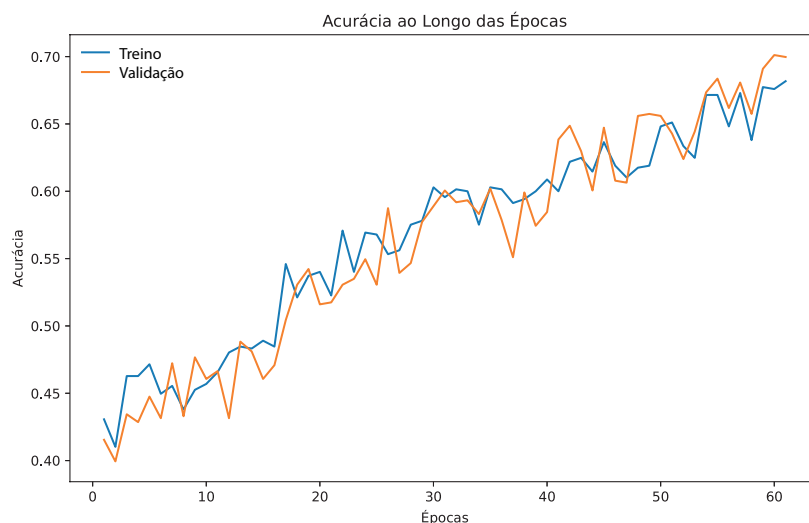
O conjunto de dados de imagens foi dividido em três subconjuntos: treinamento (70%), teste (15%) e validação (15%). Nos gráficos apresentados, foram considerados apenas os subconjuntos de teste e validação. As imagens foram categorizadas em cinco classes: folhas saudias, sigatoka negra, sigatoka amarela, mal-do-panamá e cordana. Na sequência, realizou-se o treinamento dos modelos LeNet e ViT. Após a obtenção dos resultados, foi desenvolvida uma CNN híbrida e avaliou-se o desempenho dos modelos por meio das métricas de acurácia (ACC), F1-Score e erro quadrático médio (MSE).

4.1 Modelo Baseado na Arquitetura LeNet

Avaliou-se a eficiência da adaptação do modelo LeNet para a realização de uma nova tarefa: a classificação de folhas de bananeira, distinguindo entre folhas saudias e aquelas acometidas por sigatoka negra, sigatoka amarela, mal-do-panamá e cordana.

As Figuras 4.1 e 4.2 ilustram o desempenho do modelo, destacando as curvas de perda e de acurácia ao longo dos processos de treinamento e validação. Essas curvas fornecem informações essenciais sobre o aprendizado do modelo, além de auxiliar na identificação de possíveis problemas relacionados a superajuste e subajuste, bem como na análise da estabilidade durante o processo de otimização.

Figura 4.1 – Acurácia do modelo LeNet no conjunto de dados de teste e validação.



A análise das doenças em folhas de bananeira por meio de CNNs envolve as etapas de treinamento e validação, fundamentais para a avaliação do desempenho do modelo. A

acurácia é uma métrica que mede a capacidade do modelo em classificar corretamente as amostras em cada uma dessas etapas, conforme expressa na Equação (4.1).

Com base nos resultados obtidos, a acurácia média no treinamento foi de 0,842, enquanto a acurácia média na validação, calculada pela Equação (4.2), foi de 0,777. A diferença percentual entre essas acurácias, determinada pela Equação (4.3), é de 0,0657. Esse resultado sugere a possibilidade de ocorrência de sobreajuste, uma vez que o modelo apresenta desempenho superior no treinamento em comparação com a validação.

Acurácia Média de Treinamento:

$$UM_{\text{traí}} = \frac{1}{e} \sum_{i=1}^e UM_{\text{traí},i} = 0.842 \quad (4.1)$$

Acurácia Média de Validação:

$$UM_{\text{val}} = \frac{1}{e} \sum_{i=1}^e UM_{\text{val},i} = 0.777 \quad (4.2)$$

Diferença percentual entre Acurácia de Treinamento e Validação:

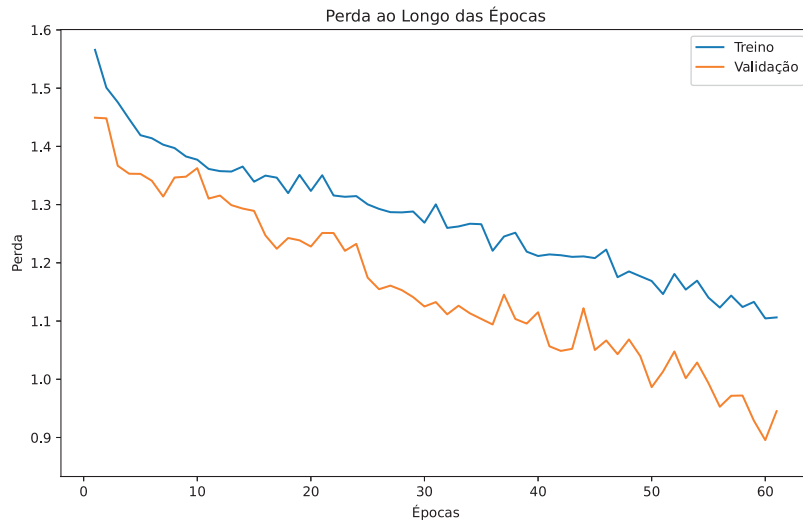
$$\Delta A = UM_{\text{traí}} - UM_{\text{val}} = 0.0657 \quad (4.3)$$

Como esse valor é relativamente alto, isso pode indicar um nível de sobreajuste no modelo.

As Figuras 4.1 e 4.2 apresentam uma análise comparativa entre as etapas de treinamento e validação. A acurácia no treinamento representa a porcentagem de amostras do conjunto de treinamento que o modelo classificou corretamente, sendo que uma acurácia elevada nesse estágio indica que o modelo aprendeu bem os padrões presentes nos dados utilizados durante o aprendizado. Por outro lado, a acurácia na validação corresponde à proporção de amostras do conjunto de validação classificadas corretamente pelo modelo. Uma acurácia elevada na validação sugere que o modelo está conseguindo generalizar adequadamente para dados não vistos anteriormente, o que é fundamental para garantir um bom desempenho em aplicações futuras.

De acordo com Hastie et al. (2009), a validação constitui uma etapa intermediária no processo de desenvolvimento do modelo, sendo aplicada durante o treinamento. O conjunto de validação corresponde a um subconjunto dos dados que não é utilizado diretamente no ajuste do modelo, mas que serve como referência para avaliar sua capacidade de generalização. Esse procedimento permite identificar indícios de sobreajuste, ou seja, quando o modelo se ajusta excessivamente aos dados de treinamento, apresentando dificuldades em generalizar para novos dados.

Figura 4.2 – Perda do modelo LeNet no conjunto de dados de teste e validação.



Na Figura 4.1, observa-se que a acurácia na validação é ligeiramente inferior à acurácia no treinamento, o que indica a ocorrência de sobreajuste. Isso significa que o modelo aprendeu muito bem os padrões, incluindo detalhes e possíveis ruídos presentes no conjunto de treinamento, mas apresenta dificuldade em generalizar para novos dados.

Por sua vez, na Figura 4.2, tanto a acurácia na validação quanto no treinamento são relativamente baixas. Esse comportamento sugere um possível cenário de subajuste, indicando que o modelo não foi capaz de aprender adequadamente os padrões dos dados de treinamento.

Considerando a Equação (4.4), a perda média no treinamento foi de 0,292, enquanto a perda média na validação, calculada pela Equação (4.5), foi de 0,335. A diferença entre elas, determinada pela Equação (4.6), é de -0,043. Apesar dessa diferença ser pequena, seu valor negativo pode ainda sugerir a presença de sobreajuste, uma vez que o modelo apresenta desempenho levemente melhor no conjunto de treinamento em comparação com a validação.

Perda Média de Treinamento:

$$e\bar{u}_{\text{trai}} = \frac{1}{e} \sum_{i=1}^e eu_{\text{trai},i} = 0.292 \quad (4.4)$$

Perda Média de Validação:

$$e\bar{u}_{\text{val}} = \frac{1}{e} \sum_{i=1}^e eu_{\text{val},i} = 0.335 \quad (4.5)$$

Diferença entre Perda de Treinamento e Validação:

$$\Delta L = e\bar{u}_{\text{train}} - e\bar{u}_{\text{val}} = -0.043 \quad (4.6)$$

O aumento na perda de validação sugere a ocorrência de sobreajuste no modelo.

Segundo Xu, Coen-Pirani e Jiang (2023), o subajuste ocorre quando o modelo não é capaz de aprender adequadamente os padrões presentes nos dados, resultando em erros significativos tanto no conjunto de treinamento quanto no de teste. Por outro lado, o sobreajuste representa a situação oposta: o modelo aprende excessivamente os dados de treinamento, incluindo ruídos e características específicas, obtendo resultados muito bons nesse conjunto, porém com baixo desempenho quando exposto a dados não vistos.

A Figura 4.2 apresenta os resultados do modelo baseado na arquitetura LeNet, utilizando as métricas de MSE e F1-Score. Nesse contexto, as métricas de acurácia, perda, erro quadrático médio (MSE) e F1-Score são fundamentais para uma avaliação abrangente do desempenho do modelo.

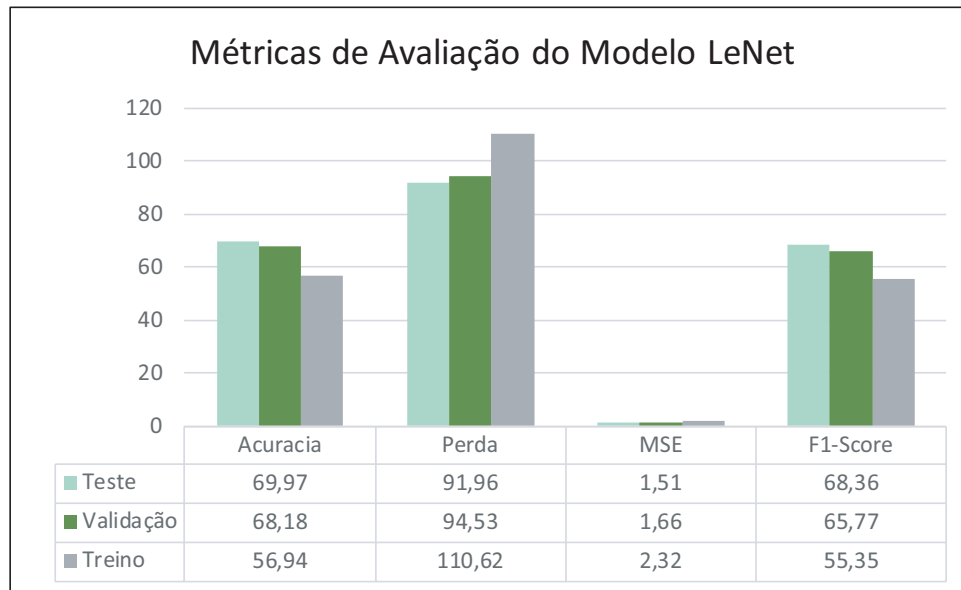
O MSE (Erro Quadrático Médio), embora não seja a métrica padrão para tarefas de classificação, foi empregado como uma medida adicional para quantificar a magnitude dos erros entre as previsões do modelo e os valores reais. Essa métrica é mais frequentemente utilizada em problemas de regressão, mas pode fornecer informações complementares sobre a dispersão dos erros, mesmo em cenários de classificação.

Já o F1-Score é uma métrica que combina precisão e revocação (recall), proporcionando uma avaliação equilibrada do desempenho do modelo, especialmente útil em situações de desequilíbrio entre classes. Isso é particularmente relevante quando há uma quantidade significativamente maior de imagens de folhas saudáveis em relação às folhas com determinadas doenças, ou quando certas doenças ocorrem com muito mais frequência do que outras.

A Figura 4.3 ilustra os resultados das métricas aplicadas:

- Acurácia, que mede a proporção de classificações corretas, mas pode ser uma métrica enganosa na presença de classes desbalanceadas.
- Perda, que permite acompanhar, ao longo do treinamento, se o modelo está efetivamente aprendendo ou não.
- MSE, utilizado como métrica complementar para avaliar a magnitude dos erros.
- F1-Score, que oferece uma avaliação mais robusta do desempenho, sobretudo em cenários de desbalanceamento, por enfatizar a relação entre verdadeiros positivos, falsos positivos e falsos negativos.

Figura 4.3 – Métricas de avaliação do modelo LeNet no conjunto de dados de treinamento, teste e validação.



O MSE representa o quão próximas estão as previsões dos valores de classes corretas. Entretanto, ressalta-se que essa métrica é menos comum em problemas de classificação multiclasse, uma vez que não descreve adequadamente as probabilidades associadas às classes discretas. Ainda assim, pode ser utilizada como métrica complementar para avaliar a magnitude dos erros do modelo.

O MSE, calculado pela Equação (4.7), apresentou o valor de 0,0224, conforme mostrado na Equação (4.8). Esse valor indica que, em média, o erro nas previsões foi relativamente baixo.

Além disso, o F1-Score, obtido pela Equação (4.9), foi de 0,85, o que reflete uma boa capacidade de classificação do modelo, demonstrando equilíbrio entre precisão e revocação, especialmente relevante diante de possíveis desequilíbrios entre as classes.

Erro Médio Quadrático (MSE):

$$MSE = \frac{1}{e} \sum_{i=1}^e (e_i - \hat{e}_i)^2 \quad (4.7)$$

Para esse modelo, obteve-se:

$$MSE = 0.0224 \quad (4.8)$$

Valores menores de MSE indicam melhor ajuste do modelo.

F1-Score Médio:

$$F_1 = 2 \times \frac{\text{precisão} \times \text{lembrança}}{\text{precisão} + \text{lembrança}} \quad (4.9)$$

F1-Score avalia o equilíbrio entre precisão e lembrança. O modelo apresentou uma pontuação média de F1-Score igual a 0,85, indicando uma boa capacidade de classificação.

Esse valor elevado de F1-Score evidencia que o modelo consegue manter um bom equilíbrio entre precisão e revocação, ou seja, é eficiente tanto na detecção correta dos verdadeiros positivos quanto na redução dos erros, como os falsos positivos — quando o modelo indica a presença de doença em uma folha sadia — e os falsos negativos — quando não detecta uma doença presente na folha.

Dessa forma, os resultados obtidos pelas métricas avaliadas fornecem uma compreensão consistente sobre o desempenho do modelo, tanto no que se refere à sua capacidade de aprendizado dos dados de treinamento, quanto à sua capacidade de generalização para novos dados. Isso é essencial para garantir que o modelo seja, de fato, aplicável na prática para a detecção automática de doenças em folhas de bananeira.

A Tabela 4.1 apresenta a matriz de confusão referente ao modelo baseado na arquitetura LeNet. Essa matriz permite avaliar o desempenho do classificador ao comparar as previsões realizadas pelo modelo com as classes reais. No contexto da classificação das folhas de bananeira, a matriz de confusão é uma ferramenta fundamental, pois possibilita identificar a precisão do modelo em cada classe, além de apontar onde ocorrem os principais erros de classificação.

Tabela 4.1 – Matriz de Confusão do modelo LeNet.

Classe	Folhas Sadias	Sigatoka Negra	Sigatoka Amarela	Mal-do-Panamá	Cordana
Folhas Sadias	97	6	0	11	9
Sigatoka Negra	14	103	13	5	3
Sigatoka Amarela	2	5	124	6	10
Mal-do-Panamá	20	32	35	52	8
Cordana	11	5	13	0	102

A matriz de confusão permite avaliar de forma detalhada o desempenho do modelo na classificação das diferentes classes de folhas de bananeira. A seguir, apresentam-se os resultados obtidos para cada classe:

- Folhas Sadias: Das 123 amostras, o modelo classificou corretamente 97, enquanto 26 foram erroneamente atribuídas a outras classes. Observa-se uma confusão relevante com a classe Mal-do-Panamá, na qual 11 amostras foram incorretamente classificadas.
- Sigatoka Negra: De um total de 138 amostras, o modelo identificou corretamente 103. Entretanto, 35 amostras foram classificadas incorretamente, com destaque para a confusão com a classe de folhas sadias, que totalizou 14 casos.

- Sigatoka Amarela: O modelo apresentou bom desempenho nesta classe, com 124 amostras corretamente classificadas e 23 erros, sendo os principais associados às classes Sigatoka Negra e Mal-do-Panamá.
- Mal-do-Panamá: Esta foi a classe com o pior desempenho, registrando apenas 52 classificações corretas. Houve 95 erros de classificação, com predominância de confusões para as classes Sigatoka Negra e Sigatoka Amarela.
- Cordana: Para esta classe, o modelo obteve um desempenho considerado satisfatório, com 102 amostras corretamente classificadas e 29 erros, especialmente relacionados às classes Sigatoka Amarela e folhas sadias.

A interpretação da Tabela 4.1 baseia-se na análise da diagonal principal, que representa as classificações corretas de cada classe. Os valores fora da diagonal indicam os erros de classificação, ou seja, os casos em que o modelo confundiu uma classe com outra.

Assim, a análise dessa matriz permite identificar padrões de confusão recorrentes, como, por exemplo, a tendência do modelo em confundir folhas sadias com Sigatoka Negra ou Mal-do-Panamá, além de destacar as dificuldades específicas na identificação correta da classe Mal-do-Panamá.

Complementarmente, a Tabela 4.2 apresenta um conjunto consolidado de métricas de desempenho, considerando os valores de verdadeiros positivos, falsos positivos e falsos negativos para cada classe. Essa tabela permite uma avaliação mais precisa do desempenho do modelo, incluindo também a comparação percentual entre os valores de acurácia e perda, oferecendo uma visão abrangente da robustez do modelo nas tarefas de classificação das doenças foliares da bananeira.

Tabela 4.2 – Métricas de desempenho do modelo LeNet.

	Folhas Sadias	Sigatoka Negra	Sigatoka Amarela	Mal-do-Panamá	Cordana
TP	103	97	124	52	102
FP	48	47	61	22	95
FN	35	26	23	95	29
Acurácia	55.37%	57.05%	59.61%	30.79%	63.35%
Perda	44.62%	42.94%	40.38%	69.23%	36.64%

As métricas de desempenho são essenciais para avaliar a qualidade do modelo de classificação aplicado às folhas de bananeira. O modelo categorizou as amostras nas classes: folhas sadias, sigatoka negra, sigatoka amarela, mal-do-Panamá e cordana.

Os verdadeiros positivos (TP) representam as amostras que foram corretamente classificadas em suas respectivas classes. Já os falsos positivos (FP) correspondem aos casos em que o modelo classifica incorretamente uma amostra em uma determinada classe, embora ela pertença a outra; por exemplo, quando uma folha sadia é classificada como mal-do-Panamá, isso configura um falso positivo para a classe mal-do-Panamá. Por sua vez,

os falsos negativos (FN) ocorrem quando o modelo não consegue identificar uma amostra como pertencente a uma determinada classe, classificando-a erroneamente fora dela.

A acurácia representa a proporção de classificações corretas, considerando a soma de verdadeiros positivos e verdadeiros negativos em relação ao total de amostras avaliadas. Já a perda (loss) é uma medida do erro do modelo durante o treinamento, indicando o quão distante estão as predições do modelo dos valores reais.

Observou-se que a acurácia variou consideravelmente entre as classes analisadas. A classe Cordana apresentou o melhor desempenho, com uma acurácia de 63,35%, enquanto a classe Mal-do-Panamá obteve a pior performance, com acurácia de apenas 30,79%. Esses resultados sugerem que o modelo encontra dificuldades para distinguir algumas doenças, principalmente aquelas cujas características visuais são mais semelhantes.

As classes Mal-do-Panamá e Sigatoka Negra apresentaram valores relativamente altos tanto de falsos positivos quanto de falsos negativos, indicando uma sobreposição dos padrões visuais detectados pelo modelo e resultando em confusões significativas entre essas doenças e outras classes.

Quanto à perda, a classe Mal-do-Panamá obteve o maior valor, atingindo 69,23%, o que evidencia a dificuldade do modelo em generalizar os padrões dessa doença. Tal desafio pode estar relacionado a um número insuficiente de amostras dessa classe ou às características visuais mais sutis que não foram capturadas adequadamente pelas camadas convolucionais da arquitetura LeNet.

4.2 Modelo Baseado na Arquitetura ViT

No método ViT, o treinamento mostrou-se mais sensível ao tamanho do conjunto de dados e ao poder computacional disponível, pois foi projetado para operar de forma ideal em grandes volumes de dados. Isso o torna especialmente adequado para pré-treinamento em bancos de dados massivos, seguido de ajuste fino (fine-tuning) em conjuntos menores. Durante o pré-treinamento, o modelo aprende representações gerais que podem ser posteriormente refinadas para tarefas específicas.

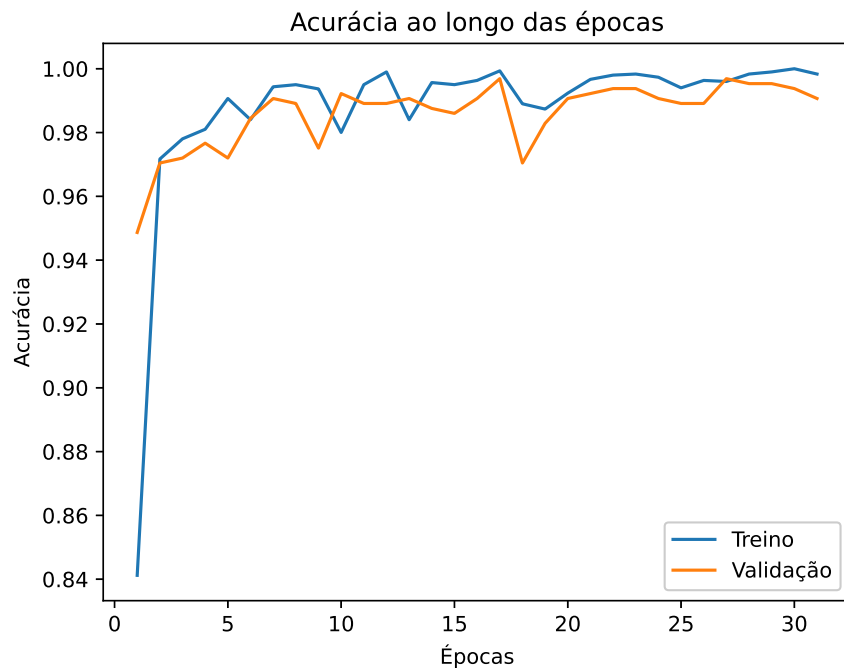
As principais vantagens observadas no ViT incluem:

- Escalabilidade: o modelo escala de forma eficiente com o aumento do volume de dados, apresentando alta precisão em grandes bases.
- Aprendizado de longo alcance: graças ao mecanismo de atenção global, o ViT é capaz de capturar relações de longo alcance entre diferentes regiões (patches) da imagem.
- Flexibilidade: diferentemente das CNNs tradicionais, que utilizam filtros fixos, o ViT não depende exclusivamente das convoluções, sendo mais adaptável a diferentes

tipos de representações visuais.

A Figura 4.4 apresenta o desempenho do modelo ViT, destacando as curvas de perda e acurácia ao longo dos processos de treinamento e validação. Essas curvas fornecem insights essenciais sobre o aprendizado do modelo, auxiliando na identificação de possíveis problemas relacionados ao sobreajuste e subajuste, além de permitir a análise da estabilidade durante a otimização.

Figura 4.4 – Acurácia do modelo ViT no conjunto de dados de teste e validação.



A análise das doenças em folhas de bananeira utilizando o método ViT considerou as etapas de treinamento e validação para a avaliação do desempenho do modelo. A acurácia, que mede a capacidade do modelo em classificar corretamente as amostras em cada fase, apresentou resultados conforme as equações (4.2) a (4.2). A acurácia média de treinamento foi de 0,906, enquanto a acurácia média de validação, dada pela equação (4.2), foi de 0,874. A diferença entre esses valores, calculada pela equação (4.2), é 0,0323, indicando um menor grau de sobreajuste quando comparado ao modelo LeNet.

Acurácia média de treinamento:

$$\bar{UM}_{\text{trem}} = \frac{1}{e} \sum_{eu=1}^e UM_{\text{trem},eu} = 0.906$$

Acurácia média da validação:

$$\bar{UM}_{\text{val}} = \frac{1}{e} \sum_{eu=1}^e UM_{\text{val},eu} = 0.874$$

Diferença percentual:

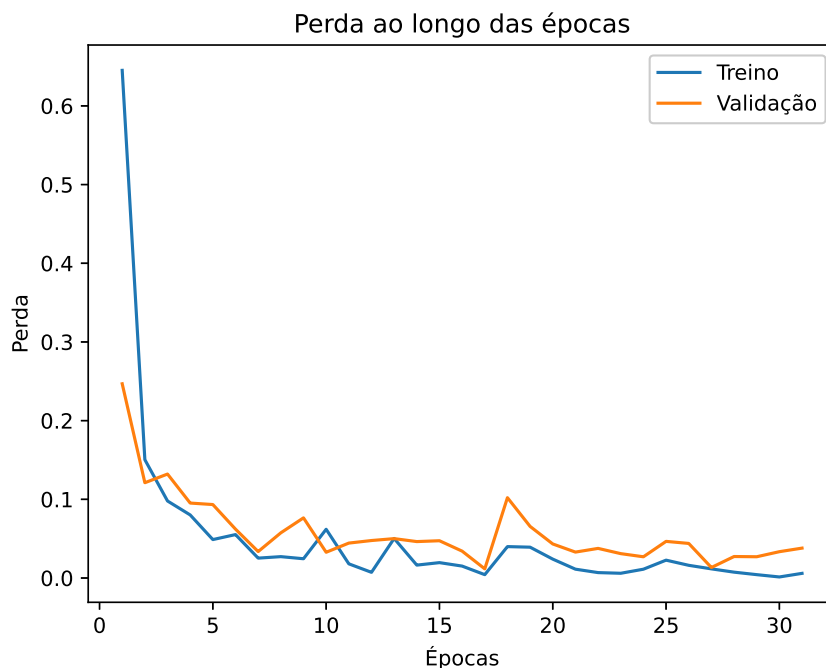
$$\Delta A = \bar{U}M_{\text{trem}} - \bar{U}M_{\text{val}} = 0.0323$$

A Figura 4.4 apresenta a análise comparativa entre o treinamento e a validação. A acurácia no treinamento indica a porcentagem de amostras do conjunto de treino que o modelo classificou corretamente, sendo que uma alta acurácia de treino demonstra que o modelo está aprendendo bem a reconhecer os exemplos durante o processo de aprendizado.

Já a acurácia de validação representa a porcentagem de amostras do conjunto de validação que o modelo classificou corretamente. Uma alta acurácia nessa etapa indica que o modelo está generalizando bem para dados inéditos, o que é essencial para garantir a robustez e aplicabilidade prática do modelo em novos dados.

Observa-se que os resultados obtidos são bastante semelhantes aos apresentados pelo método LeNet, evidenciando que o ViT alcança alta precisão. As especificidades do reconhecimento das características das doenças serão detalhadas posteriormente.

Figura 4.5 – Perda do modelo ViT no conjunto de dados de teste e validação.



Fazendo uma análise comparativa entre as Figuras 4.4 e 4.5, observa-se que, na Figura 4.4, a acurácia de validação é ligeiramente inferior à acurácia de treino, indicando a presença de sobreajuste. Isso significa que o modelo está aprendendo detalhadamente os ruídos e particularidades do conjunto de treinamento, mas pode ter dificuldades para generalizar adequadamente para dados novos.

Por outro lado, na Figura 4.5, tanto a acurácia de validação quanto a de treino são baixas, sugerindo que o modelo está sofrendo de subajuste — ou seja, não conseguiu

aprender os padrões relevantes dos dados de treinamento de forma adequada.

Conforme discutido anteriormente, o MSE (Erro Médio Quadrático), representado no Gráfico 6, indica o quão próximas estão as previsões das classes corretas. Contudo, essa métrica é menos comum para problemas de classificação multiclasse, pois não descreve bem as probabilidades associadas às classes discretas.

Na equação (4.10), a perda média de treinamento é 0,174, enquanto a perda média de validação, apresentada na equação (4.11), é 0,226. A diferença entre elas, calculada pela equação (4.12), é -0,0518, o que indica que o modelo mantém estabilidade durante a validação.

$$\vec{e}_{\text{trem}} = 0.174 \quad (4.10)$$

Perda Média da Validação:

$$\vec{e}_{\text{val}} = 0.226 \quad (4.11)$$

Diferença:

$$\Delta L = -0.0518 \quad (4.12)$$

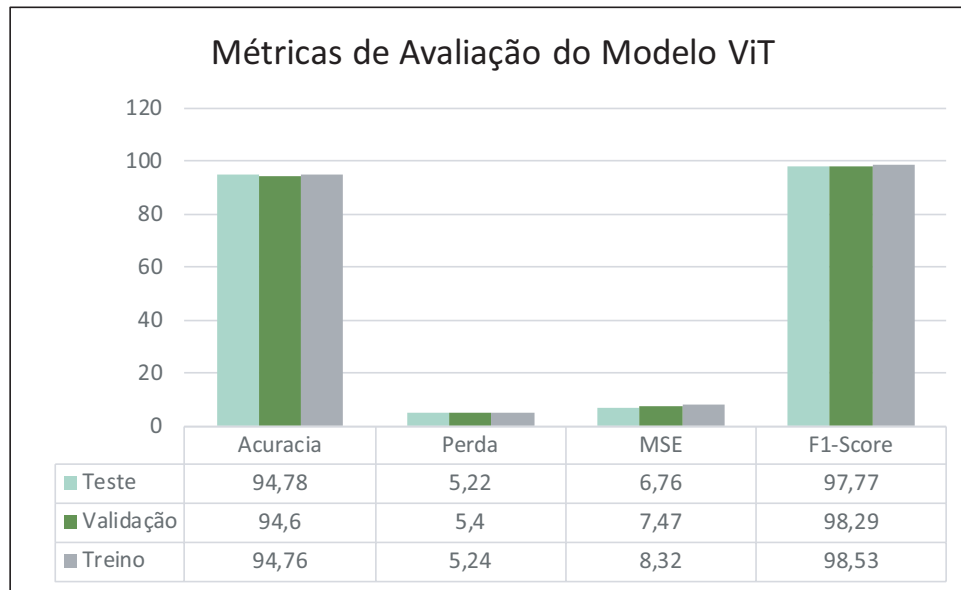
Isso indica que o modelo mantém estabilidade durante a validação. A Figura 4.6 apresenta a estrutura padronizada da divisão do conjunto de dados, onde o conjunto de teste foi utilizado após a conclusão do treinamento para a avaliação final do desempenho do modelo em dados não vistos. Já o conjunto de validação foi empregado para monitorar o desempenho do modelo durante o treinamento, ajudando a prevenir o sobreajuste e permitindo ajustes contínuos. Por fim, o conjunto de treinamento foi utilizado para ajustar os pesos do modelo, expondo o ViT a uma grande variedade de imagens de folhas saudáveis e com doenças.

Percebe-se que pelo método ViT, as informações foram mais precisas, ou seja, em destaque aos dados com precisão correta e o equilíbrio entre a precisão e a revocação.

O cálculo da Área Sob a Curva ROC (AUC) revelou que o modelo obteve um valor de 0,92, indicando excelente desempenho na capacidade de separar corretamente as classes. Em comparação com a arquitetura LeNet, o modelo baseado em ViT apresentou resultados superiores em todas as análises realizadas, destacando-se especialmente em termos de segurança e estabilidade da perda.

A Tabela 4.3, representa a matriz de confusão com base no modelo da arquitetura ViT, avaliando o desempenho de classificação que compara as previsões do modelo com

Figura 4.6 – Métrica de avaliação do modelo ViT no conjunto de dados de treinamento, teste e validação.



as classes reais. No contexto das folhas de bananeira, ela auxilia com a identificação da precisão do modelo ao classificar as doenças.

Tabela 4.3 – Matriz de Confusão do modelo ViT.

Classe	Folhas Sadias	Sigatoka Negra	Sigatoka Amarela	Mal-do-Panamá	Cordana
Folhas Sadias	136	1	3	1	0
Sigatoka Negra	1	155	1	2	1
Sigatoka Amarela	3	0	124	2	0
Mal-do-Panamá	3	1	0	99	3
Cordana	4	2	3	4	127

A matriz de confusão permite avaliar o desempenho do modelo na classificação das diferentes classes de folhas. A seguir, foram os resultados obtidos:

- Folhas Sadias: Das 136 amostras de folhas saudáveis, o modelo classificou: 1 como sigatoka negra, 3 como sigatoka amarela, 3 como Mal-do-Panamá e 4 como cordana. Nos dados preditivos, das 360 amostras de folhas saudáveis, 1 foi classificada como sigatoka negra, 3 como sigatoka amarela, 1 como Mal-do-Panamá e nenhuma como cordana. As principais discrepâncias ocorreram nas doenças Mal-do-Panamá e Cordana.
- Sigatoka Negra: Das 155 amostras de Sigatoka Negra, o modelo classificou 1 como saudável, 0 como sigatoka amarela, 1 como Mal-do-Panamá e 2 como cordana. Nos dados preditivos, as classificações errôneas foram: 1 como sigatoka amarela, 2 como Mal-do-Panamá e 1 como cordana.
- Sigatoka Amarela: O modelo apresentou bom desempenho na identificação da Sigatoka Amarela, com 124 amostras. Entre elas, as classificações errôneas foram: 3

como folhas saudáveis, 0 como sigatoka negra, 2 como Mal-do-Panamá e 0 como cordana. Nos dados preditivos, 3 folhas foram classificadas como saudáveis, 1 como sigatoka negra, nenhuma como Mal-do-Panamá e 3 como cordana.

- Mal-do-Panamá: Essa doença foi a menos prevalente nas folhas de bananeira, com apenas 99 amostras. Nas amostras semelhantes, o modelo classificou 3 folhas como saudáveis, 1 como sigatoka negra, nenhuma como sigatoka amarela e 3 como cordana. Nos dados preditivos, houve 1 folha classificada como saudável, 2 como sigatoka negra, 1 como sigatoka amarela e 4 como cordana.
- Cordana: Das 127 folhas com cordana, o modelo classificou 4 como saudáveis, 2 como sigatoka negra, 3 como sigatoka amarela e 4 como Mal-do-Panamá. Nos dados preditivos, nenhuma folha foi classificada como saudável ou como sigatoka amarela, sendo 1 classificada como sigatoka negra e 3 como Mal-do-Panamá.

Para uma melhor compreensão do que é apresentado na Tabela 4.4, a diagonal principal da matriz contém as previsões corretas, enquanto os valores fora da diagonal representam as classificações incorretas, ou seja, os erros do modelo. A análise apresentada na mesma tabela contribui para a identificação de possíveis confusões entre as classes, como, por exemplo, quando o modelo confunde folhas saudáveis com outras doenças. Além disso, ela fornece as métricas necessárias para a avaliação do desempenho do modelo.

Tabela 4.4 – Métricas de desempenho do modelo ViT.

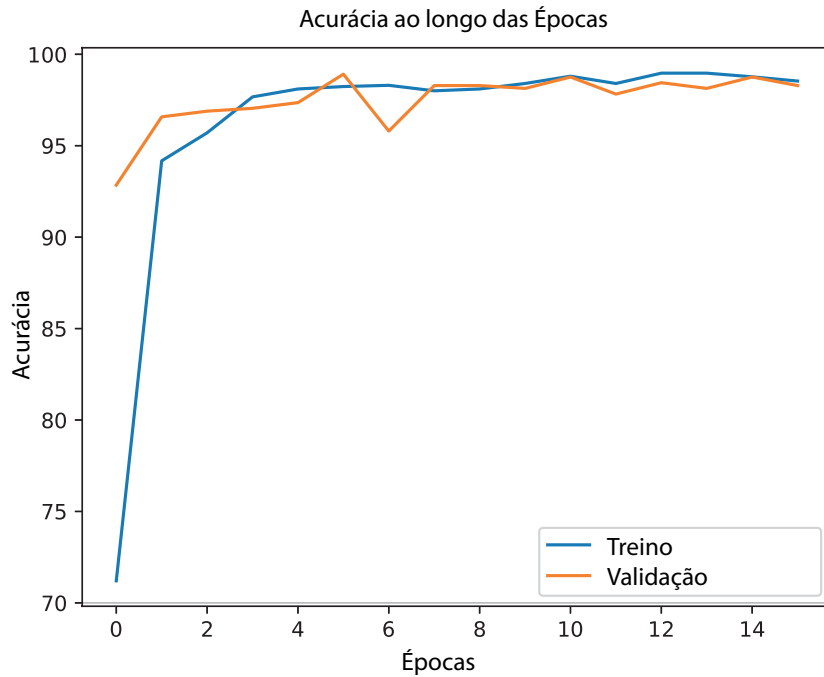
Métrica	Folhas Sadias	Sigatoka Negra	Sigatoka Amarela	Mal-do-Panamá	Cordana
TP	103	97	124	52	102
FP	48	47	61	22	95
FN	35	26	23	95	29
Acurácia	55.37%	57.05%	59.61%	30.79%	63.35%
Perda	44.62%	42.94%	40.38%	69.23%	36.64%

Por fim, com todos os dados analisados, observa-se que, ao comparar com o método LeNet, o modelo ViT apresenta maior precisão, com uma porcentagem de perdas significativamente menor e, conseqüentemente, com previsões corretas mais acentuadas em todas as classes de doenças foliares.

4.3 Modelo HibridNet

O modelo HibridNet foi desenvolvido neste trabalho a partir de uma análise comparativa entre as arquiteturas LeNet e ViT, sendo que esta última apresentou resultados mais precisos e consistentes. A Figura 4.7 ilustra o desempenho do modelo durante as etapas de treinamento e validação, evidenciando resultados similares entre ambos os processos.

Figura 4.7 – Acurácia do modelo HibridNet no conjunto de dados de teste e validação.



Conforme pode ser observado na Figura 4.7, ao longo das épocas utilizadas na análise, a precisão obtida pelo modelo HibridNet demonstrou-se elevada, superando os resultados das CNNs LeNet e ViT. A Figura 4.7 também apresenta as curvas de perda, que indicam os erros em relação aos dados reais para os conjuntos de treinamento e validação.

Ressalta-se que os dados apresentados nas figuras, inicialmente individualizados por modelo, são sintetizados pelas Equações (4.13), (4.14) e (4.15). A acurácia média no treinamento, conforme a Equação (4.13), foi de 0,907; enquanto a acurácia média de validação, dada pela Equação (4.14), foi de 0,903. A diferença entre elas, calculada na Equação (4.15), foi de apenas 0,00385, o que evidencia uma excelente capacidade de generalização do modelo proposto.

Acurácia Média de Treinamento:

$$\overline{e\bar{u}}_{\text{trem}} = 0.907 \quad (4.13)$$

Acurácia Média de Validação:

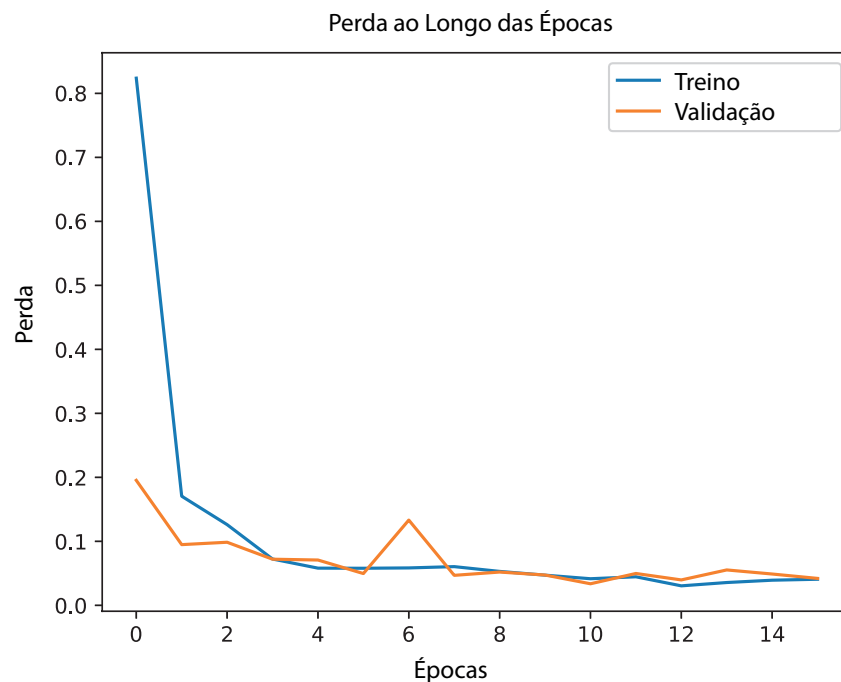
$$\overline{e\bar{u}}_{\text{val}} = 0.903 \quad (4.14)$$

Diferença:

$$\Delta A = 0.00385 \quad (4.15)$$

Isso significa que o modelo HibridNet apresentou uma margem reduzida de perda e uma validação consistente, evidenciando sua capacidade de generalização.

Figura 4.8 – Perda do modelo HibridNet no conjunto de dados de teste e validação.



A Figura 4.8 representa a medida da diferença entre as previsões do modelo e os valores reais, ou seja, os desvios em relação aos dados de referência. Quanto menor essa diferença, melhor o desempenho da CNN utilizada.

De acordo com os cálculos apresentados, a perda média no treinamento, conforme a Equação (4.16), é de 0,134. Já a perda média na validação, apresentada na Equação (4.17), é de 0,162. A diferença entre elas, calculada pela Equação (4.18), é de -0,0275, o que indica um bom ajuste e estabilidade durante a validação do modelo.

Perda Média de Treinamento:

$$\overline{eu}_{\text{trein}} = 0.134 \quad (4.16)$$

Perda Média de Validação:

$$\overline{eu}_{\text{val}} = 0.162 \quad (4.17)$$

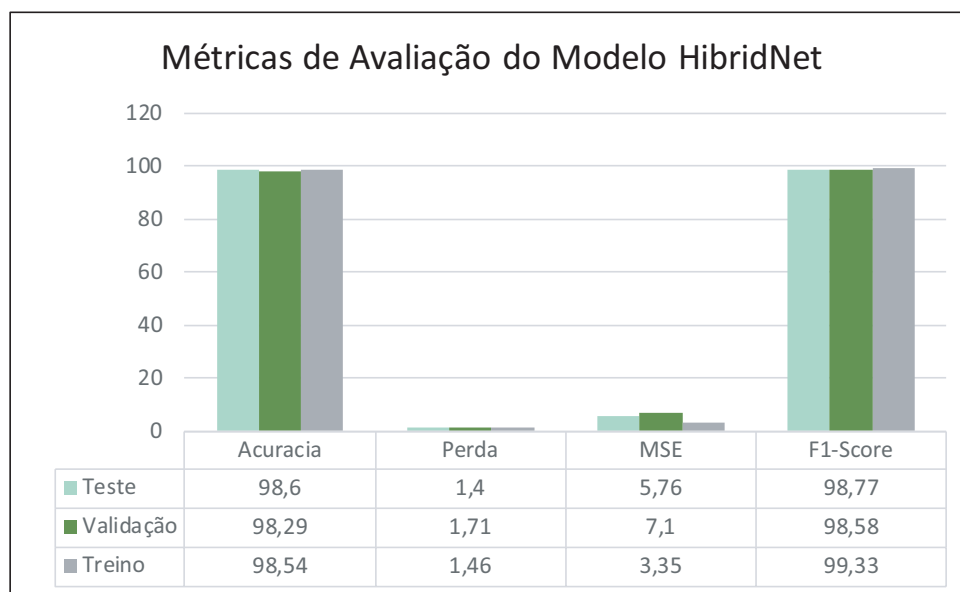
Diferença:

$$\Delta L = -0.0275 \quad (4.18)$$

Tal resultado evidencia a estabilidade do modelo e a sua capacidade de manter um bom ajuste aos dados durante a validação.

Diante dos resultados apresentados, observa-se que o modelo HibridNet fornece previsões mais próximas da realidade em comparação aos demais modelos. A Figura 4.9 ilustra essas informações de forma detalhada.

Figura 4.9 – Métrica de avaliação do modelo HibridNet no conjunto de dados de treinamento, teste e validação.



A matriz de confusão e as métricas de avaliação do modelo indicam um desempenho consistente, com uma precisão média de 0,91, uma lembrança média de 0,89 e um F1-Score médio de 0,90. Além disso, o modelo alcançou uma acurácia global de 90,3%, sendo o melhor desempenho entre os três modelos avaliados.

A Tabela 4.5 apresenta a matriz de confusão gerada a partir do modelo HibridNet, avaliando seu desempenho na tarefa de classificação. Essa matriz compara as previsões do modelo com as classes reais das folhas de bananeira, permitindo verificar a precisão na identificação das diferentes doenças.

Tabela 4.5 – Matriz de Confusão do modelo HibridNet.

Classe	Folhas Sadias	Sigatoka Negra	Sigatoka Amarela	Mal-do-Panamá	Cordana
Folhas Sadias	2280	7	2	27	0
Sigatoka Negra	20	1862	20	6	1
Sigatoka Amarela	2	16	2388	60	1
Mal-do-Panamá	1	1	3	1862	1
Cordana	1	1	3	3	127

A matriz de confusão permite avaliar o desempenho do modelo na classificação das diferentes classes de folhas. A seguir, são apresentados os resultados obtidos:

- **Folhas Sadias:** Das 2280 amostras de folhas saudáveis, o modelo classificou: 20 como sigatoka negra, 2 como sigatoka amarela, 1 como Mal-do-Panamá e 1 como cordana. Nos dados preditivos, das 2280 amostras de folhas saudáveis, 7 foram classificadas como sigatoka negra, 2 como sigatoka amarela, 27 como Mal-do-Panamá e nenhuma como cordana. Os principais erros ocorreram nas classificações para Mal-do-Panamá e Sigatoka Negra.

- **Sigatoka Negra:** Das 1862 amostras de Sigatoka Negra, o modelo classificou 7 como saudáveis, 16 como sigatoka amarela, 1 como Mal-do-Panamá e 1 como cordana. Nos dados preditivos, 7 folhas foram classificadas como saudáveis, 2 como sigatoka amarela, 27 como Mal-do-Panamá e 1 como cordana.
- **Sigatoka Amarela:** O modelo teve um bom desempenho na identificação da Sigatoka Amarela, com 2388 amostras. Destas, as classificações errôneas foram: 2 como folhas saudias, 20 como sigatoka negra, 3 como Mal-do-Panamá e 3 como cordana.
- **Mal-do-Panamá:** O modelo identificou 1862 folhas com Mal-do-Panamá. As classificações erradas incluem: 27 como saudáveis, 6 como sigatoka negra, 60 como sigatoka amarela e 3 como cordana. Nos dados preditivos, houve 1 folha classificada como saudável, 1 como sigatoka negra, 3 como sigatoka amarela e 1 como cordana.
- **Cordana:** Das 127 folhas com cordana, o modelo classificou: 0 como saudáveis, 1 como sigatoka negra, 1 como sigatoka amarela e 1 como Mal-do-Panamá. Nos dados preditivos, 1 folha foi classificada como saudável, 3 como sigatoka amarela, 1 como sigatoka negra e 3 como Mal-do-Panamá.

Para uma melhor compreensão do que é apresentado na Tabela 4.5, a diagonal principal da matriz contém as previsões corretas, enquanto os valores fora da diagonal representam as classificações incorretas, ou seja, os erros do modelo. Por outro lado, a análise representada na Tabela 4.6 contribui para a identificação de possíveis confusões entre as classes. Ela ilustra, por exemplo, como o modelo pode confundir folhas saudáveis com outras doenças, além de fornecer as métricas necessárias para a avaliação do desempenho do modelo.

Tabela 4.6 – Métricas de desempenho do modelo HibridNet.

Métrica	Folhas Sadias	Sigatoka Negra	Sigatoka Amarela	Mal-do-Panamá	Cordana
TP	2280	1674	2388	1862	127
FP	39	16	25	93	3
FN	36	6	79	47	8
Acurácia	96.81%	98.70%	95.82%	93%	92.02%
Perda	3.18%	1.29%	4.17%	6.99%	7.97%

Por fim, a Tabela 4.6 apresenta os dados mais precisos e, em comparação com os modelos LeNet e ViT, mostra as maiores porcentagens de acerto, alcançando quase 100% na análise e classificação das doenças em folhas de bananeira. A Tabela 4.7 compara as métricas dos três modelos (HibridNet, LeNet e ViT), demonstrando a superioridade do modelo HibridNet.

Tabela 4.7 – Comparação de desempenho entre os modelos LeNet, ViT e HibridNet.

	Acurácia	Perda	F1-score	MSE
LeNet	77,70	33,50	65,77	1,66
ViT	94,64	5,40	98,29	7,47
HibridNet	98,29	1,71	98,58	7,10

4.4 Estudo de Ablação dos Modelos LeNet, ViT e HibridNet

Nesta seção, são discutidos os resultados obtidos a partir da comparação entre os modelos LeNet, ViT e HibridNet. O modelo proposto neste trabalho, denominado HibridNet, combina características de ambas as abordagens: convolucional e baseada em atenção. A análise foca na capacidade de generalização dos modelos, bem como em suas limitações e potenciais aplicações práticas.

Conforme demonstrado na Tabela 4.8, a arquitetura HibridNet superou os modelos individuais nas principais métricas de avaliação utilizadas (como acurácia, F1-score, perda e erro quadrático médio), evidenciando a eficácia da fusão entre as abordagens convolucionais e baseadas em atenção. Essa melhora no desempenho indica que a combinação das duas arquiteturas permitiu uma extração de características mais robusta e diversificada, favorecendo a generalização do modelo mesmo diante de variações nos dados.

Tabela 4.8 – Estudo de ablação dos modelos LeNet, ViT e HybridNet.

Modelo	Classe	Perda	Acurácia	TP	FP	FN	F1-Score	MSE
LeNet	Folhas Sadias	44,62%	55,37%	103	48	35	65,77%	1,66%
	Sigatoka Negra	42,94%	57,05%	97	47	26		
	Sigatoka amarela	40,38%	59,61%	124	61	23		
	Mal-do-Panamá	69,23%	30,79%	52	22	95		
	Cordana	36,64%	63,35%	102	95	29		
ViT	Folhas Sadias	44,62%	55,37%	103	48	35	98,29%	7,47%
	Sigatoka Negra	43,94%	57,05%	97	47	26		
	Sigatoka amarela	40,38%	59,61%	124	61	23		
	Mal-do-Panamá	69,23%	30,79%	52	22	95		
	Cordana	36,64%	63,35%	102	95	29		
HibridNet	Folhas Sadias	3,18%	96,81%	2280	39	36	98,58%	7,10%
	Sigatoka Negra	1,29%	98,70%	1674	16	6		
	Sigatoka amarela	4,17%	95,82%	2388	25	79		
	Mal-do-Panamá	6,99%	93,00%	1862	93	47		
	Cordana	7,97%	92,02%	127	3	8		

Apesar dos resultados promissores, algumas limitações devem ser destacadas. O conjunto de dados utilizado apresenta um número reduzido de amostras e um desequilíbrio entre as classes, o que pode comprometer o desempenho dos modelos em contextos mais amplos. Além disso, a presença de ruídos visuais nas imagens pode ter impactado negativamente o desempenho em alguns casos. Outro ponto relevante é o custo computacional da arquitetura HibridNet, que, ao unir dois modelos já existentes, demanda maior poder de processamento e tempo de treinamento, o que pode dificultar sua aplicação em ambientes com recursos computacionais limitados.

Por fim, embora os testes tenham sido realizados com imagens pré-processadas e organizadas em ambiente controlado, não foi conduzida uma validação externa com dados coletados diretamente em campo. Essa limitação compromete a avaliação da aplicabilidade prática do modelo em cenários reais.

5 Conclusão

A presente pesquisa apresentou o desenvolvimento e a avaliação de uma CNN híbrida para a classificação de doenças em folhas de bananeira, com foco em melhorar a precisão e a robustez diante dos desafios práticos encontrados em cenários agrícolas, como variações de iluminação, sobreposição de folhas e diversidade de sintomas.

A partir da criação de um conjunto de dados específica e da integração de técnicas avançadas de DL, foi possível demonstrar que modelos híbridos superam as limitações de abordagens tradicionais baseadas exclusivamente em arquiteturas únicas, alcançando maior desempenho em termos de acurácia, sensibilidade e capacidade de generalização.

Os resultados obtidos confirmam a hipótese, de que a combinação dessas arquiteturas proporciona maior precisão e estabilidade na classificação de doenças foliares, mesmo em condições ambientais adversas, possibilitando a identificação precoce de doenças e contribuindo para a redução de perdas na produção, custos operacionais e impactos ambientais. Além disso, o estudo reforça a relevância do uso da IA na agricultura, destacando sua capacidade de oferecer soluções tecnológicas inovadoras para problemas complexos.

Dessa forma, ao responder à questão de pesquisa (Q01), conclui-se que é possível desenvolver e analisar uma CNN híbrida eficiente ao considerar fatores relevantes, tais como a qualidade da imagem, o desbalanceamento das classes, a capacidade computacional e os ruídos visuais presentes nas folhas. Além disso, técnicas como regularização, amostragem ponderada e parada antecipada foram fundamentais para aperfeiçoar a eficiência do modelo e lidar com desafios práticos. Os resultados demonstram que as CNNs híbridas são ferramentas viáveis e eficientes no manejo fitossanitário da bananicultura, permitindo a identificação precoce de doenças e contribuindo para a redução de perdas na produção, custos operacionais e impactos ambientais.

Portanto, a pesquisa não apenas alcançou seus objetivos, mas também abre novas perspectivas para estudos futuros, como a adaptação do modelo para outras culturas, o aprimoramento da base de dados com imagens de *drones* e a implementação em sistemas móveis para facilitar sua aplicação em campo. Dessa forma, este trabalho contribui significativamente para o avanço da agricultura de precisão e para o fortalecimento da sustentabilidade no setor agrícola, além da realização de testes com dados obtidos diretamente em campo, visando avaliar a robustez do modelo em condições reais.

Referências

- ABADI, M. et al. {TensorFlow}: a system for {Large-Scale} machine learning. In: *12th USENIX symposium on operating systems design and implementation (OSDI 16)*. [S.l.: s.n.], 2016. p. 265–283. Citado na página 24.
- ABUHAMAD, A.; WALSH, E. *Ultrasound in obstetrics and gynecology: a practical approach (-2014)*. [S.l.]: Paula and David Bloomer, 2014. Citado na página 25.
- AGROLINK. *Sigatoka Amarela*. 2025. Acessado em: 02 jan. 2025. Disponível em: <https://www.agrolink.com.br/problemas/mal-da-sigatoka_67.html>. Citado na página 38.
- AGROLINK. *Sigatoka Negra*. 2025. Acessado em: 02 jan. 2025. Disponível em: <https://www.agrolink.com.br/problemas/sigatoka-negra_1719.html>. Citado na página 35.
- ALMEIDA, C. F. P. *Mapeamento da Distribuição Populacional Através da Detecção de Áreas Edificadas em Imagens de Regiões Heterogêneas do Google Earth Usando Deep Learning*. Tese (Tese de Doutorado) — Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), Rio de Janeiro, 2017. Programa de Pós-graduação em Informática da PUC-Rio. Citado 5 vezes nas páginas 24, 25, 26, 32 e 33.
- ALOM, M. Z. et al. Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. *arXiv preprint arXiv:1802.06955*, 2018. Citado na página 29.
- ALZUBAIDI, L. et al. Review of deep learning: concepts, cnn architectures, challenges, applications, future directions. *Journal of big Data*, Springer, v. 8, p. 1–74, 2021. Citado 2 vezes nas páginas 25 e 32.
- ARAÚJO, W. J. da S. et al. Manejo do mal-do-panamá da bananeira utilizando trichoderma sp. e óleos essenciais de lippia sidoides. *Anais da Academia Pernambucana de Ciência Agronômica*, v. 15, n. 2, p. 73–90, 2018. Citado na página 40.
- ARMAN, S. E. et al. Bananalsd: A banana leaf images dataset for classification of banana leaf diseases using machine learning. *Data in Brief*, v. 50, p. 109608, 2023. ISSN 2352-3409. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2352340923006959>>. Citado 2 vezes nas páginas 47 e 48.
- ASHRAF, T.; KHAN, Y. N. Weed density classification in rice crop using computer vision. *Computers and Electronics in Agriculture*, v. 175, p. 105590, 2020. ISSN 0168-1699. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0168169920302003>>. Citado 6 vezes nas páginas 20, 23, 24, 25, 33 e 34.
- BARATLOO, A. et al. Part 1: Simple definition and calculation of accuracy, sensitivity and specificity. *Emerg (Tehran)*, v. 3, n. 2, p. 48–49, Spring 2015. Citado 4 vezes nas páginas 21, 29, 30 e 34.

BISONG, E. et al. *Building machine learning and deep learning models on Google cloud platform*. [S.l.]: Springer, 2019. Citado 5 vezes nas páginas 20, 23, 26, 31 e 33.

BRASIL, Secretaria do Desenvolvimento Agrário. *Mal-do-Panamá: é possível conviver com a doença que afeta as plantações de banana*. 2021. <<https://www.sda.ce.gov.br/2021/09/17/mal-do-panama-e-possivel-conviver-com-a-doenca-que-afeta-as-plantacoes-de-banana/>>. Acessado: 10 jan. 2025. Citado na página 41.

CHEN, Z. et al. A novel imbalanced dataset mitigation method and ecg classification model based on combined 1d_cbam-autoencoder and lightweight cnn model. *Biomedical Signal Processing and Control*, Elsevier, v. 87, p. 105437, 2024. Citado 2 vezes nas páginas 47 e 48.

DOGAN, Y. An innovative approach for parkinson's disease diagnosis using cnn, nca, and svm. *Neural Computing and Applications*, Springer, v. 36, n. 32, p. 20089–20110, 2024. Citado na página 47.

DORGIVAL, J. *Sigatoka: adeal fará inspeção em municípios produtores de banana*. 2024. Defesa agropecuária. Acessado em: 14 jan. 2025. Disponível em: <<https://www.defesaagropecuaria.al.gov.br/noticia/22-rotativas/372-sigatoka-adeal-fara-inspecao-em-municipios-produtores-de-banana>>. Citado 2 vezes nas páginas 36 e 37.

DOSOVITSKIY, A.; AL. et. Uma imagem vale 16x16 palavras: Transformadores para reconhecimento de imagens em escala. In: *International Conference on Learning Representations (ICLR)*. [S.l.]: s.n., 2021. Disponível em: <<https://openreview.net/forum?id=YicbFdNTTy>>. Acesso em: 11 mar. 2025. Citado 2 vezes nas páginas 25 e 32.

DOSOVITSKIY, A. et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. Citado 3 vezes nas páginas 51, 52 e 53.

ERGÜN, E. High precision banana variety identification using vision transformer based feature extraction and support vector machine. *Scientific Reports*, Nature Publishing Group UK London, v. 15, n. 1, p. 10366, 2025. Citado na página 15.

FERRARI, J.; NOGUEIRA, E. M. d. C. Artigo Científico, *Principais doenças fúngicas da bananeira*. 2018. Citado 4 vezes nas páginas 37, 38, 39 e 40.

GANOKRATANAA, T.; KETCHAM, M.; PRAMKEAW, P. Advancements in cataract detection: The systematic development of lenet-convolutional neural network models. *Journal of Imaging*, v. 9, n. 10, 2023. ISSN 2313-433X. Disponível em: <<https://www.mdpi.com/2313-433X/9/10/197>>. Citado na página 16.

GONZÁLEZ-RODRÍGUEZ, V. E. et al. Artificial intelligence: a promising tool for application in phytopathology. *Horticulturae*, MDPI, v. 10, n. 3, p. 197, 2024. Citado na página 15.

HAILU, Y. Banana leaf disease images. *Mendeley Data*, v. 1, 2021. Citado 2 vezes nas páginas 47 e 48.

- HARTMANN, M. et al. Anodal high-definition transcranial direct current stimulation over the posterior parietal cortex modulates approximate mental arithmetic. *Journal of Cognitive Neuroscience*, v. 32, n. 5, p. 862–876, 2020. Citado 4 vezes nas páginas 21, 22, 25 e 26.
- HASTIE, T. et al. *The elements of statistical learning: data mining, inference, and prediction*. [S.l.]: Springer, 2009. v. 2. Citado na página 58.
- HE, T. et al. Bag of tricks for image classification with convolutional neural networks. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2019. Citado na página 24.
- HOWARD, A. et al. Searching for mobilenetv3. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. [S.l.: s.n.], 2019. Citado na página 23.
- HUANG, G. et al. Densely connected convolutional networks. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2017. p. 2261–2269. Citado na página 22.
- HUSSAIN, T. et al. Effresnet-vit: A fusion-based convolutional and vision transformer model for explainable medical image classification. *IEEE Access*, IEEE, 2025. Citado na página 50.
- JIAO, S. et al. Multi-modal contrastive learning for medical image classification with limited training data. In: *2024 International Conference on Machine Learning and Applications (ICMLA)*. [S.l.: s.n.], 2024. p. 1083–1088. Citado na página 19.
- JUNIOR, W. C. et al. Detecção de ferrugem de folhas vegetais por meio de imagens aplicando redes neurais convolucionais. Pontifícia Universidade Católica de Goiás, 2020. Citado 7 vezes nas páginas 15, 18, 19, 21, 25, 29 e 33.
- KADAM, S. R.; DHORE, M. P. Review of deep learning methods. In: *2024 8th International Conference on Computing, Communication, Control and Automation (ICCUBEA)*. [S.l.: s.n.], 2024. p. 1–8. Citado na página 20.
- KASINATHAN, T.; UYYALA, S. R. Machine learning ensemble with image processing for pest identification and classification in field crops. *Neural Computing and Applications*, Springer, v. 33, n. 13, p. 7491–7504, 2021. Citado na página 18.
- LECUN, Y. et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, v. 86, n. 11, p. 2278–2324, 1998. Citado na página 50.
- LECUN, Y. et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, Ieee, v. 86, n. 11, p. 2278–2324, 2002. Citado na página 53.
- LI, Q.; YAN, M.; XU, J. Optimizing convolutional neural network performance by mitigating underfitting and overfitting. In: *2021 IEEE/ACIS 19th International Conference on Computer and Information Science (ICIS)*. [S.l.: s.n.], 2021. p. 126–131. Citado na página 21.
- LI, Z. et al. Local feature acquisition and global context understanding network for very high-resolution land cover classification. *Scientific Reports*, Nature Publishing Group UK London, v. 14, n. 1, p. 12597, 2024. Citado na página 50.

- LIU, J. Survey of the image recognition based on deep learning network for autonomous driving car. In: *2020 5th International Conference on Information Science, Computer Technology and Transportation (ISCTT)*. [S.l.: s.n.], 2020. p. 1–6. Citado na página 21.
- MAHMOOD, F. et al. Deep adversarial training for multi-organ nuclei segmentation in histopathology images. *IEEE Transactions on Medical Imaging*, v. 39, n. 11, p. 3257–3267, 2020. Citado 3 vezes nas páginas 22, 26 e 32.
- MEDHI, E.; DEB, N. Psfd-musa: A dataset of banana plant, stem, fruit, leaf, and disease. *Data in Brief*, Elsevier, v. 43, p. 108427, 2022. Citado 2 vezes nas páginas 47 e 48.
- MENDES, P. Ministra recebe demandas do setor produtivo da banana. *Ministério da Agricultura, Brasília*, v. 22, 2015. Citado 2 vezes nas páginas 43 e 45.
- MILITANTE, S. V.; GERARDO, B. D.; DIONISIO, N. V. Plant leaf detection and disease recognition using deep learning. In: *2019 IEEE Eurasia Conference on IOT, Communication and Engineering (ECICE)*. [S.l.: s.n.], 2019. p. 579–582. Citado 6 vezes nas páginas 15, 19, 20, 24, 32 e 33.
- MUNAWAR, N. H. et al. Counseling on education to prevent anemia in pregnant women with banana and dragon fruit: Penyuluhan tentang edukasi mencegah anemia ibu hamil dengan pisang dan buah naga. *Journal of Community Empowerment for Multidisciplinary (JCEMTY)*, v. 3, n. 1, p. 1–12, 2025. Citado na página 15.
- NATALI, L. *Tipos de aprendizado de máquina e algumas aplicações*. [S.l.]: TerraLAB, 2021. Citado 5 vezes nas páginas 15, 18, 19, 32 e 33.
- NETO, S. D. S.; GUIMARÃES, T. Evolução da cultura da banana no brasil e no mundo. *Planaltina, DF: Embrapa Cerrados*, 2011. Citado 5 vezes nas páginas 41, 42, 43, 44 e 45.
- PAIVA, M.; FIORAVANÇO, J. Sigatoka-negra da bananeira. *Current Agricultural Science and Technology*, v. 11, n. 2, 2005. Citado 3 vezes nas páginas 35, 36 e 37.
- PAWAR, A.; DEOSARKAR, S. Deep learning based mango leaves disease detection for smart agriculture. In: *IEEE. 2024 2nd International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS)*. [S.l.], 2024. p. 499–506. Citado na página 15.
- PLANTIX. *Mancha-de-cordana*. 2025. <<https://plantix.net/pt/library/plant-diseases/100216/leaf-blotch-of-banana>>. Acessado em: 13 jan. 2025. Citado 2 vezes nas páginas 43 e 44.
- RAGHU, M. et al. Transfusão: Compreendendo a aprendizagem de transferência para imagens médicas. In: *Advances in Neural Information Processing Systems (NeurIPS)*. [S.l.: s.n.], 2019. p. 3347–3357. Disponível em: <<https://proceedings.neurips.cc/paper/2019/hash/1d72310edcde4663a0a5a5d9b5c6c2b8-Abstract.html>>. Acesso em: 11 mar. 2025. Citado 3 vezes nas páginas 21, 22 e 26.
- RAMOS, L.; SAPPA, A. D. Multispectral semantic segmentation for land cover classification: An overview. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Ieee, 2024. Citado na página 52.
- REGIANI, E.; DIAS, G. V. *Identificador de doenças em plantas por rede neural convolucional*. São Paulo, 2021. Citado 8 vezes nas páginas 20, 21, 24, 25, 26, 30, 31 e 34.

SCHUMANN, I. S. et al. Controle da podridão negra do relho com o uso de fertilizante de cobre. Instituto Federal Goiano, 2018. Citado 6 vezes nas páginas 15, 18, 19, 25, 31 e 33.

SELVARAJ, P.; SIVAPRAKASH, S. Feature extraction and feature selection in medical images. In: *Intelligent Computing Techniques in Biomedical Imaging*. [S.l.]: Elsevier, 2025. p. 83–97. Citado na página 48.

SHI, Y. et al. Event recognition method based on dual-augmentation for a ϕ -otdr system with a few training samples. *Optics Express*, Optica Publishing Group, v. 30, n. 17, p. 31232–31243, 2022. Citado na página 48.

SILVA, V. M. O. da et al. Detecção de sigatoka amarela em folhas de banana com redes neurais convolucionais (cnn) e aprendizagem por transferência profunda (dtl). *OBSERVATÓRIO DE LA ECONOMÍA LATINOAMERICANA*, v. 22, n. 6, p. e5280–e5280, 2024. Citado 2 vezes nas páginas 47 e 48.

SINGH, G.; GULERIA, K.; SHARMA, S. A deep learning-based fine-tuned resnet50 model for multiclass mango leaf disease classification. In: IEEE. *2024 2nd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)*. [S.l.], 2024. p. 1310–1316. Citado na página 15.

SZEGEDY, C. et al. Going deeper with convolutions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2015. p. 1–9. Citado 2 vezes nas páginas 33 e 34.

TAN, M.; LE, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In: PMLR. *International conference on machine learning*. [S.l.], 2019. p. 6105–6114. Citado na página 23.

WANG, Y. et al. Vision transformers for image classification: A comparative survey. *Technologies*, v. 13, n. 1, 2025. ISSN 2227-7080. Disponível em: <<https://www.mdpi.com/2227-7080/13/1/32>>. Citado na página 16.

WU, H.; GU, X. Towards dropout training for convolutional neural networks. *Neural Networks*, v. 71, p. 1–10, 2015. ISSN 0893-6080. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0893608015001446>>. Citado 3 vezes nas páginas 29, 30 e 31.

XIA, M. et al. Eegnet classification of sleep eeg for individual specialization based on data augmentation. *Cognitive Neurodynamics*, Springer, v. 18, n. 4, p. 1539–1547, 2024. Citado na página 48.

XIAO, N. et al. A novel clustering algorithm based on directional propagation of cluster labels. In: *2019 International Joint Conference on Neural Networks (IJCNN)*. [S.l.: s.n.], 2019. p. 1–8. Citado na página 25.

XIE, S. et al. Aggregated residual transformations for deep neural networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2017. Citado na página 25.

XU, C.; COEN-PIRANI, P.; JIANG, X. Empirical study of overfitting in deep learning for predicting breast cancer metastasis. *Cancers (Basel)*, v. 15, n. 7, p. 1969, Mar 25 2023. Citado na página 60.

- ZEILER, M. D.; FERGUS, R. Visualizing and understanding convolutional networks. In: FLEET, D. et al. (Ed.). *Computer Vision – ECCV 2014*. Cham: Springer International Publishing, 2014. p. 818–833. ISBN 978-3-319-10590-1. Citado 2 vezes nas páginas 31 e 34.
- ZHANG, Q. et al. Image segmentation with pyramid dilated convolution based on resnet and u-net. In: SPRINGER. *Neural Information Processing: 24th International Conference, ICONIP 2017, Guangzhou, China, November 14-18, 2017, Proceedings, Part II 24*. [S.l.], 2017. p. 364–372. Citado 3 vezes nas páginas 30, 31 e 36.
- ZHAO, Z. et al. Hyperspectral image classification using groupwise separable convolutional vision transformer network. *IEEE Transactions on Geoscience and Remote Sensing*, IEEE, 2024. Citado na página 51.
- ZHOU, Y. et al. Dctn: Dual-branch convolutional transformer network with efficient interactive self-attention for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, IEEE, v. 62, p. 1–16, 2024. Citado na página 51.
- ZHU, H. et al. Covc-redrnet: a deep learning model for covid-19 classification. *Machine learning and knowledge extraction*, MDPI, v. 5, n. 3, p. 684–712, 2023. Citado na página 52.
- ZHU, X. et al. Adapting object detectors via selective cross-domain alignment. In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2019. p. 687–696. Citado 2 vezes nas páginas 22 e 26.
- ZOPH, B. et al. Learning transferable architectures for scalable image recognition. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2018. p. 8697–8710. Citado 3 vezes nas páginas 22, 26 e 32.