

UNIVERSIDADE ESTADUAL DE PONTA GROSSA
SETOR DE ENGENHARIAS, CIÊNCIAS AGRÁRIAS E DE TECNOLOGIA
DEPARTAMENTO DE INFORMÁTICA
MESTRADO EM COMPUTAÇÃO APLICADA

EMILI EVERZ GOLOMBIÉSKI

MODELO DE INTEGRAÇÃO DE DADOS CLIMÁTICOS E DE IMAGENS NA
GESTÃO AGRÍCOLA

PONTA GROSSA

2024

EMILI EVERZ GOLOMBIÉSKI

**MODELO DE INTEGRAÇÃO DE DADOS CLIMÁTICOS E DE IMAGENS NA
GESTÃO AGRÍCOLA**

Dissertação apresentada ao Programa de Pós-Graduação em Computação Aplicada da Universidade Estadual de Ponta Grossa, como requisito parcial para a obtenção do título de Mestre em Computação Aplicada.

Área de Concentração: Computação para Tecnologias em Agricultura

Orientadora: Prof^a. Dr^a. Maria Salete Marcon Gomes Vaz

PONTA GROSSA

2024

G627 Golombiéski, Emili Everz
Modelo de integração de dados climáticos e de imagens na gestão agrícola /
Emili Everz Golombiéski. Ponta Grossa, 2024.
91 f.

Dissertação (Mestrado em Computação Aplicada - Área de Concentração:
Computação para Tecnologias em Agricultura), Universidade Estadual de Ponta
Grossa.

Orientadora: Profa. Dra. Maria Salete Marcon Gomes Vaz.

1. Gestão agrícola. 2. Informações climáticas. 3. Imagens - grãos. 4.
Modelo - dados. I. Vaz, Maria Salete Marcon Gomes. II. Universidade Estadual de
Ponta Grossa. Computação para Tecnologias em Agricultura. III.T.

CDD: 004



UNIVERSIDADE ESTADUAL DE PONTA GROSSA
Av. General Carlos Cavalcanti, 4748 - Bairro Uvaranas - CEP 84030-900 - Ponta Grossa - PR - <https://uepg.br>

TERMO

TERMO DE APROVAÇÃO

Emili Everz Golombiéski

"MODELO DE INTEGRAÇÃO DE DADOS CLIMÁTICOS E DE IMAGENS NA GESTÃO AGRÍCOLA"

Dissertação aprovada como requisito parcial para obtenção do grau de Mestre no Programa de Pós-Graduação em Computação Aplicada da Universidade Estadual de Ponta Grossa, pela seguinte banca examinadora:

Prof^ª. Dr^ª. Maria Salete Marcon Gomes Vaz(UEPG/PR - Presidente)

Prof. Dr. Adriano Ferrasa(UEPG/PR)

Prof^ª. Dr^ª. Simone Nasser Matos(UTF/PR)

Ponta Grossa, 19 de dezembro de 2024.



Documento assinado eletronicamente por **Maria Salete Marcon Gomes Vaz, Coordenador(a) do Programa de Pós-Graduação em Computação Aplicada - Mestrado**, em 19/12/2024, às 15:28, conforme Resolução UEPG CA 114/2018 e art. 1º, III, "b", da Lei 11.419/2006.



Documento assinado eletronicamente por **Adriano Ferrasa, Professor(a)**, em 19/12/2024, às 15:28, conforme Resolução UEPG CA 114/2018 e art. 1º, III, "b", da Lei 11.419/2006.



Documento assinado eletronicamente por **Simone Nasser Matos, Usuário Externo**, em 19/12/2024, às 15:28, conforme Resolução UEPG CA 114/2018 e art. 1º, III, "b", da Lei 11.419/2006.



A autenticidade do documento pode ser conferida no site <https://sei.uepg.br/autenticidade> informando o código verificador **2312216** e o código CRC **C9234737**.

“Se eu vi mais longe, foi por estar de pé sobre ombros de gigantes”

– Isaac Newton

Dedico este trabalho ao meu marido,
Emerson Luis Golombiéski pela força,
paciência e amor incondicional que me
mantiveram firme nesta jornada.

AGRADECIMENTOS

A realização desta dissertação só foi possível graças ao apoio e à contribuição de várias pessoas e instituições, às quais expresso meu mais profundo e sincero agradecimento. Em primeiro lugar, sou imensamente grata à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo apoio financeiro e incentivo à pesquisa, que permitiram a concretização deste trabalho.

A meus avós e minha mãe, dedico um agradecimento especial, pois sempre estiveram ao meu lado, oferecendo amor, compreensão e apoio incondicional ao longo de toda a minha trajetória acadêmica. Em particular, quero expressar minha mais profunda gratidão ao meu marido, Emerson Luis Golombiéski, cuja força e constante apoio foram fundamentais para que eu não desistisse nos momentos de maior dificuldade. Seu amor, paciência e encorajamento me deram a energia necessária para superar os obstáculos e alcançar este objetivo.

Não posso deixar de agradecer à minha orientadora, Dr^a. Maria Salete Marcon Gomes Vaz, por todo o apoio, inspiração e clareza ao longo desta jornada. Sua orientação foi muito mais do que apenas acadêmica. Seu papel foi fundamental para o desenvolvimento desta pesquisa e para o meu crescimento enquanto pesquisadora e pessoa, jamais esquecerei tudo que aprendi com você.

Aos meus colegas, minha gratidão pela convivência e pelo compartilhamento de conhecimentos e experiências ao longo desses anos, principalmente a Gabriel Passos de Jesus e Helena Pistune. O apoio mútuo que encontrei em vocês foi um pilar importante durante essa trajetória. Agradeço também ao próprio PPGCA e aos docentes que, com suas valiosas orientações, contribuíram para o meu desenvolvimento acadêmico e para a qualidade deste trabalho. Cada aula, cada conselho e cada palavra de incentivo foram essenciais para que eu chegasse até aqui.

Também expresso minha gratidão ao Laboratório de Sistemas Digitais do Departamento de Informática da Universidade Estadual de Ponta Grossa (UEPG), onde tive a oportunidade de utilizar o estúdio fotográfico para realizar partes importantes desta pesquisa. Em especial, agradeço à Prof^a. Dr^a. Rosane Falate, cujo apoio técnico e orientação foram indispensáveis para o sucesso deste trabalho. A estrutura oferecida pelo laboratório e o suporte técnico foram vitais para que eu pudesse concretizar as análises e resultados apresentados nesta dissertação.

RESUMO

A agricultura enfrenta desafios crescentes devido às mudanças climáticas, que afetam a qualidade dos grãos produzidos. Condições extremas, como secas prolongadas e chuvas intensas, alteram os padrões de crescimento e comprometem a saúde das plantações, resultando em menor rendimento, valor nutricional e aumento de defeitos nos grãos. Para garantir a sustentabilidade e a segurança alimentar, demanda adaptação a essas mudanças e busca de soluções inovadoras para integração de dados climáticos, melhorando a resiliência e a eficiência da gestão agrícola. Deste modo, a pesquisa foi conduzida com o objetivo de desenvolver um modelo de integração de dados climáticos e imagens, visando aprimorar a tomada de decisões na agricultura. A metodologia adotada incluiu uma revisão de literatura, identificando lacunas no conhecimento existente sobre a integração de dados. A partir dessa revisão, foi elaborado um modelo de integração com utilização de técnicas de aprendizado de máquina para analisar dados provenientes de estações meteorológicas e imagens de grãos. O modelo resultante criou um repositório centralizado, facilitando o acesso e a análise em tempo real das informações. Os resultados obtidos mostraram que a integração de dados climáticos e imagens melhora a precisão das previsões de rendimento agrícola, e pode permitir avaliação das condições dos grãos, considerando diversos estados de saúde. A pesquisa incluiu a captura e análise prática de imagens de grãos de soja, documentando defeitos causados por variações climáticas, possibilitando avaliação precisa da saúde das plantas. Além disso, o modelo de integração foi testado em um fluxo de rastreabilidade agrícola, demonstrando a contribuição para acompanhamento eficiente no setor. A utilização de um sistema integrado pode otimizar práticas de manejo, como irrigação e aplicação de defensivos agrícolas, resultando em melhorias na produtividade e na qualidade dos produtos agrícolas. A pesquisa, portanto, apresenta um modelo prático para aplicação na gestão agrícola, promovendo uma abordagem eficiente e adaptável às condições climáticas.

Palavras-chave: Gestão Agrícola, Informações Climáticas, Imagens de Grãos, Modelo de Dados.

ABSTRACT

Agriculture faces growing challenges due to climate change, which impacts the quality of produced grains. Extreme conditions, such as prolonged droughts and heavy rains, alter growth patterns and compromise crop health, resulting in lower yields, reduced nutritional value, and increased grain defects. To ensure sustainability and food security, it is crucial to adapt to these changes and seek innovative solutions for integrating climate data to enhance resilience and efficiency in agricultural management. Thus, this research was conducted with the aim of developing a model for integrating climate data and images to improve decision-making in agriculture. The adopted methodology included a literature review to identify gaps in existing knowledge about data integration. Based on this review, an integration model was developed using machine learning techniques to analyze data from weather stations and grain images. The resulting model created a centralized repository, facilitating real-time access and analysis of information. The results showed that the integration of climate data and images improves the accuracy of agricultural yield predictions and enables the evaluation of grain conditions considering various health states. The research included the practical capture and analysis of soybean grain images, documenting defects caused by climate variations, thus allowing for a precise assessment of plant health. Additionally, the integration model was tested in an agricultural traceability workflow, demonstrating its contribution to efficient monitoring in the sector. The use of an integrated system can optimize management practices such as irrigation and the application of agricultural inputs, resulting in improved productivity and product quality. Therefore, this research presents a practical model for application in agricultural management, promoting an efficient and adaptable approach to changing climatic conditions.

Keywords: Soybean Production Management, Climate Information, Soybean Grain Images, Data Model.

LISTA DE FIGURAS

Figura 1. Cadeia produtiva e pontos de avaliação para rastreabilidade.	15
Figura 2. Fluxo de integração RFID na rastreabilidade agrícola.	16
Figura 3. Imagens de Defeitos de Grãos de Soja.	19
Figura 4. Perfil da ameaça climática à segurança da América do Sul, Central e Caribe, cenário de curto, médio e longo prazo.	22
Figura 5. Transição de El Niño para La Niña.	23
Figura 6. Fluxo de Informações em um data warehouse.	27
Figura 7. Diferença entre carregamento em tempo real e carregamento em lotes.	29
Figura 8. Modelo Estrela.	31
Figura 9. Modelo Floco de Neve.	32
Figura 10. Modelo Constelação.	34
Figura 11. Mapa das estações meteorológicas do INMET, do Agritempo, da ANA e da Epagri, de acordo com as coordenadas disponíveis nos respectivos.	35
Figura 12. Detecção de objetos com YOLO.	39
Figura 13. Reconhecimento de características em imagens com uso de CNN.	40
Figura 14. Arquitetura ResNet50.	41
Figura 15. Processamento de imagem pelo ViT.	43
Figura 16. Ponto de parada antecipada (early stopping).	44
Figura 17. Fluxo da metodologia.	48
Figura 18. Classificação de Grãos em Laboratório.	51
Figura 19. Estúdio Fotográfico.	52
Figura 20. Interface da plataforma OpenWeather.	53
Figura 21. Fórmula para conversão datetime para ISO 8601.	54
Figura 22. Conversão das imagens de CR2 para JPG.	57
Figura 23. Organização dos diretórios de imagens.	58
Figura 24. Classificação de Grãos em Laboratório.	61
Figura 25. Modelo de Integração de Dados Agrícolas.	62
Figura 26. Modelo Dimensional para Integração de Dados Climáticos e de Imagens.	71
Figura 27. Evolução da acurácia e perda ao longo das épocas.	78
Figura 28. Fluxo de Informações Baseado na Integração de Dados.	79

LISTA DE QUADROS

Quadro 1. Influências das variáveis ambientais no desenvolvimento de pragas e doenças nas plantas e suas aplicações na gestão agrícola.....	39
Quadro 2. Processo de Integração e Pré-Processamento de Dados Meteorológicos e Imagens de Grãos.....	67
Quadro 3. Fontes de Dados e Suas Implicações.....	70
Quadro 4. Atributos da Tabela de fatos “Fatos_Imagem”	74
Quadro 5. Atributos da Tabela de Fatos “Fato_Clima”	75
Quadro 6. Dados extraídos da API OpenWeather.....	77
Quadro 7. Amostra de dados climáticos coletados e convertidos.....	78
Quadro 8. Evolução da acurácia e da perda ao longo das épocas.....	79

LISTA DE SIGLAS

AMA – *American Meteorological Association*

API - *Application Programming Interface*

BI – *Business Intelligence*

CRM - *Customer Relationship Management*

CNN - *Convolutional Neural Network*

EF - *Electronic Focusing*

EOS – *Electro Optical System*

ERP - *Enterprise Resource Planning*

ERM – *Environmental Resources Management*

ETL – *Extract, Transform and Load*

EUA – *Estados Unidos da América*

FAO - *Food and Agriculture Organization of the United Nations*

FN – *Forma Normal*

IA – *Inteligência Artificial*

IoT – *Internet of Things*

LED – *Light-Emitting Diode*

MAPA- *Ministério da Agricultura, Pecuária e Abastecimento*

NLP – *Natural Language Processing*

QR - *Quick Response*

RFID - *Radio Frequency Identification*

YOLO - *You Only Look Once*

SUMÁRIO

1 INTRODUÇÃO	12
1.1 OBJETIVOS.....	13
1.1.1 Objetivo Geral	14
1.1.2 Objetivos Específicos.....	14
2 REVISÃO DA LITERATURA	15
2.1. RASTREABILIDADE AGRÍCOLA	15
2.2 CLASSIFICAÇÃO DE GRÃOS.....	17
2.3. GESTÃO DE DADOS AGRÍCOLAS.....	20
2.4 IMPACTOS CLIMÁTICOS NA PRODUÇÃO AGRÍCOLA	22
2.5 CENTRALIZAÇÃO DE DADOS	24
2.6 DATA WAREHOUSE.....	26
2.6.1 Data Warehouse E Sua Aplicação Na Agricultura.....	31
2.6.2 Modelagens Multidimensionais	302
2.7 COLETA DE DADOS EM ESTAÇÕES METEOROLÓGICAS	34
2.8 ALGORITMOS DE RECONHECIMENTO DE PADRÕES DE IMAGENS	40
2.8.1. Algoritmo Yolo.....	40
2.8.2 CNN.....	41
2.8.3 Vision Transformers - Vits.....	43
2.8.4 Funções E Métodos Para Reconhecimento De Padrões.....	44
2.9 TRABALHOS CORRELATOS.....	45
3 MATERIAIS E MÉTODOS.....	50
3.1 REVISÃO DE LITERATURA.....	50
3.2 COLETA DE AMOSTRAS.....	50
3.2.1 Classificação De Grãos – Captura De Imagens	51
3.2.2 Coleta De Dados Meteorológicos.....	52
3.3 MODELAGEM CONCEITUAL.....	54
3.4 VALIDAÇÕES.....	55
4 RESULTADOS E DISCUSSÕES.....	60
4.1 BASE DE IMAGENS DE GRÃOS DE SOJA	60
4.2 MODELO DE INTEGRAÇÃO DE DADOS CLIMÁTICOS E DE IMAGENS.....	62
4.3 MODELO CONCEITUAL	67
4.4 VALIDAÇÃO DA EXTRAÇÃO DE METADADOS.....	75
4.5 INSERÇÃO DO MODELO NO FLUXO DE RASTREABILIDADE AGRÍCOLA.....	80
5 CONCLUSÕES E PESQUISAS FUTURAS.....	81
REFERÊNCIAS	83

1 INTRODUÇÃO

Segundo dados apresentados pela *Food and Agriculture Organization of the United Nations* (FAO, 2020), a produção agrícola é um setor vital para a economia, contribuindo para a garantia da segurança alimentar e para o desenvolvimento socioeconômico global. Dessa forma, demanda possibilitar a produção agrícola do benefício das tecnologias para auxiliar na adaptação às variáveis ambientais, como informações climáticas (Miranda; Silva; Menezes, 2019).

As informações climáticas e agrícolas, quando bem estruturadas, possibilitam decisões, inerentes ao plantio, irrigação e uso de insumos, sejam tomadas de forma mais assertiva, resultando em maior eficiência e produtividade no campo. Essa precisão na tomada de decisões tem impacto direto no rendimento das lavouras e no uso responsável dos recursos, fundamentais para a sustentabilidade da produção agrícola (Campos; Lima; Oliveira, 2021).

Nesse contexto, Santos e Silva (2020) destacam a importância das tecnologias de *data warehouse*, permitindo integrar, organizar e analisar grandes volumes de dados provenientes de diferentes fontes, como sensores climáticos e imagens. Essa capacidade de centralizar informações facilita o acesso a informações estratégicas capazes de aprimorar a eficiência operacional e promover o planejamento agrícola.

A reunião desses dados em um único ambiente oferece um diferencial competitivo ao setor agrícola, pois otimiza o planejamento de safras, permite a gestão de riscos, incluindo a identificação precoce de pragas, condições climáticas adversas, entre outras ameaças (Almeida; Carvalho; Ribeiro, 2021). A integração de dados climáticos e imagens em um *data warehouse* melhora a previsão de safras e a gestão agrícola, garantindo uma abordagem eficiente e rentável (Jin *et al.*, 2019),

Essa integração de dados promove gestão eficaz e torna-se relevante diante dos desafios crescentes enfrentados pela agricultura, devido às mudanças climáticas, afetando diretamente os padrões de cultivo e os rendimentos das colheitas (Rocha *et al.*, 2019). A capacidade de acessar, integrar e analisar dados climáticos em tempo real permite a adaptação e a redução de riscos no setor agrícola (Martins, *et al.*, 2022).

Complementando essa abordagem, Luvisi *et al.* (2021) discutiram sobre a análise de imagens e como é capaz de oferecer informações sobre o estado das plantações, possibilitando a identificação precoce de problemas como pragas e

doenças, e ao combinar essas informações em um ambiente único, os produtores podem tomar decisões estratégicas, otimizando a produtividade e a sustentabilidade da produção agrícola.

Embora os avanços tecnológicos tenham permitido acesso aos dados agrícolas, a ausência de uma infraestrutura tecnológica para tratar com grandes quantidades de dados, como ter sistemas adequados de armazenamento, processamento e análise dos dados, para permitir aos produtores a transformação das informações em conhecimento prático (Karfia; Campos; Nunes, 2021).

Esses desafios comprometem a capacidade de tomar decisões, com otimização do uso de recursos, melhoria na produtividade e redução do impacto ambiental. Assim, mesmo com a crescente disponibilidade de dados, a falta de suporte tecnológico para gerenciá-los impede a muitos agricultores o aproveitamento do potencial das inovações digitais no campo (Eck *et al.*, 2020).

É importante destacar que a fragmentação dos dados, muitas vezes provenientes de diferentes fontes e formatos, pode dificultar a obtenção de uma visão abrangente e precisa da situação do campo (Besen, 2019). Isso reforça a necessidade de um ambiente integrador desses dados, facilitando o acesso às informações de qualidade e em tempo real.

Dado o exposto, a problemática reside na falta de estrutura para integrar e analisar dados climáticos e imagens na agricultura, limitando a capacidade dos produtores de tomar decisões, afetando a gestão dos recursos, a adaptação às variáveis ambientais e a otimização da produtividade diante dos desafios das mudanças climáticas.

O desenvolvimento de uma solução capaz de integrar dados climáticos e imagens para otimizar a gestão agrícola. Ao fornecer uma visão consolidada e acessível dos dados, essa abordagem contribui diretamente para uma agricultura eficiente, resiliente e sustentável, capaz de enfrentar os desafios das mudanças climáticas e das demandas globais por alimentos. Assim, a criação do modelo de integração de dados pode transformar as práticas agrícolas, garantindo o uso estratégico das informações e, conseqüentemente, melhorando a produtividade e a sustentabilidade do setor.

1.1 OBJETIVOS

Tendo em vista o exposto, os objetivos desta pesquisa são os que seguem:

1.1.1 Objetivo Geral

O objetivo geral desta dissertação é criar um modelo de integração de dados para agrupar informações climáticas e imagens na gestão agrícola, com foco na melhoria da tomada de decisões.

1.1.2 Objetivos Específicos

- i. Criar uma base de dados de imagens de grãos de soja em diferentes estados de saúde;
- ii. Desenvolver um modelo de integração de dados, considerando as demandas específicas do setor agrícola; e
- iii. Validar a etapa de extração de metadados na produção da soja.

2 REVISÃO DA LITERATURA

A pesquisa sobre a produção de soja e sua contextualização agrícola, demanda estudos sobre o estado da arte inerente à gestão agrícola. Neste sentido, este capítulo oferece uma análise de trabalhos correlatos aos desafios na gestão de dados agrícolas, como rastreabilidade, classificação de grãos, impactos climáticos na produção agrícola, centralização de dados, *data warehouse* incluindo sua aplicação na agricultura e modelagens multidimensionais, coleta de dados em estações meteorológicas e algoritmos de reconhecimento de padrões.

2.1. RASTREABILIDADE AGRÍCOLA

A rastreabilidade agrícola permite acompanhar o trajeto dos produtos agrícolas, na cadeia de suprimentos, desde o plantio até chegar no consumidor final. Este processo é capaz de processar e garantir a segurança alimentar e promover transparência e confiança entre consumidores e produtores (Golan *et al.*, 2004).

Na agricultura, o sistema de rastreabilidade (Figura 1) passa por várias etapas, iniciando no campo, onde cada lote de cultivo é identificado e registrado com informações, como localização geográfica, métodos de cultivo, uso de insumos agrícolas, datas de plantio e colheita. Esses dados permitem o monitoramento e verificação da conformidade com as normas de qualidade e segurança alimentar, além de possibilitar a rápida identificação e resolução de problemas, como surtos de doenças ou contaminações (Silva, 2018).

Figura 1. Cadeia produtiva e pontos de avaliação para rastreabilidade.

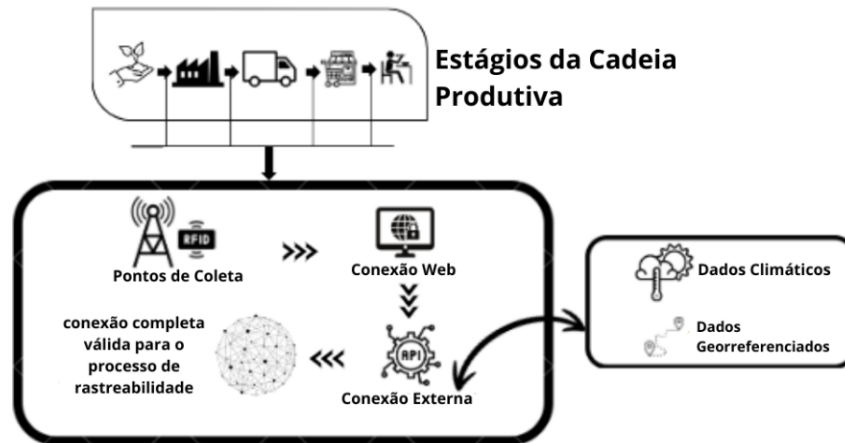


Fonte: Silva, 2018.

Ao longo da cadeia produtiva, transporte e distribuição, a rastreabilidade agrícola utiliza tecnologias como códigos de barras, QR codes, sistemas de RFID - Identificação por Radiofrequência (Figura 2) para registrar e transmitir informações de

forma precisa e em tempo real (Slimani *et al*, 2024). Essas tecnologias permitem rastrear os produtos em cada etapa, desde o ponto de origem até o destino.

Figura 2. Fluxo de integração RFID na rastreabilidade agrícola.



Fonte: Adaptado de Golombiński, Vaz e Guimarães, 2024.

Tratando-se de RFID, a coleta de dados é realizada por meio de pontos de coleta distribuídos ao longo da cadeia de produção, capturando informações em diferentes fases, como plantio, colheita e armazenamento de grãos. Em seguida, os dados coletados são conectados à Web, permitindo a integração com APIs externas. Variáveis como condições climáticas, rotas de transporte e localizações geográficas são analisadas para otimizar o fluxo logístico e minimizar riscos. Todos esses dados são compilados e integrados em um ambiente centralizado, assegurando a validade e a rastreabilidade da cadeia produtiva (Golombiński; Vaz; Guimarães, 2024).

Stankovisk e Ostojić (2023) expõem, aos consumidores, a capacidade da rastreabilidade agrícola oferecer segurança dos produtos adquiridos, com alta qualidade e produção de forma sustentável, onde a rastreabilidade permite acesso às informações sobre a procedência, práticas agrícolas utilizadas, e informações sobre os agricultores responsáveis pela produção.

Além dos benefícios de segurança e transparência, a rastreabilidade agrícola impulsiona a inovação e a melhoria contínua na agricultura. Nguyen *et al.* (2023) expõem a capacidade dos dados coletados ao longo da cadeia de produção poderem ser analisados para identificar padrões, otimizar processos, reduzir desperdícios e maximizar a eficiência dos recursos agrícolas.

2.2 CLASSIFICAÇÃO DE GRÃOS

A classificação de grãos foi instituída pela Lei nº 9.972, de 25 de maio de 2000 e regulamentada pelo Decreto nº 3.664, de 17 de novembro de 2000. O Serviço Nacional de Aprendizagem Rural do Paraná (Senar-PR, 2024) explica que a classificação de grãos permite padronizar e categorizar os produtos agrícolas de acordo com critérios específicos de qualidade, tamanho, teor de umidade, pureza e outras características relevantes.

Essa padronização facilita o comércio internacional, proporciona transparência no mercado, e assegura que os produtos atendam aos padrões exigidos pelos consumidores e reguladores. Além disso, ajuda os produtores na identificação de lotes de alta qualidade para maximizar o valor de mercado e otimizar o uso dos recursos agrícolas (Ribeiro, *et al.*, 2024).

A classificação de grãos é um procedimento manual que ocorre durante o recebimento ou expedição dos produtos agrícolas. Nesse processo, uma amostra do lote é coletada e submetida à análise para avaliar sua condição, englobando aspectos como qualidade, tamanho e teor de umidade dos grãos, além da verificação de impurezas ou defeitos (Brasil, 2006).

A classificação é relevante em um ambiente de comércio internacional, onde a padronização e a transparência são exigidas para atender às demandas dos consumidores e reguladores (Abiove, 2024). Para que os grãos sejam comercializados, interna ou externamente, devem atender a rigorosos padrões de qualidade estabelecidos pela legislação, segundo Instrução Normativa MAPA – Ministério da Agricultura, Pecuária e Abastecimento nº 38/2006.

O descumprimento dos padrões de qualidade pode resultar em penalidades financeiras para produtores e comerciantes, afetando a rentabilidade e a sustentabilidade do negócio. Os principais critérios avaliados durante a classificação incluem percentuais de impurezas, níveis de umidade e a quantidade de grãos com defeitos, sejam graves ou leves, devendo estar dentro dos limites estipulados pela legislação vigente e pela empresa compradora.

Os defeitos encontrados podem ser atribuídos a uma série de fatores, desde condições inadequadas de cultivo e manejo até falhas no armazenamento e transporte. Por exemplo, umidade excessiva durante a colheita pode levar ao

crescimento de fungos, enquanto práticas inadequadas de armazenamento podem resultar em infestação por insetos (Silva; Pereira, 2020).

Os defeitos nos grãos, como mofo, infestação por insetos, danos físicos ou alterações químicas, impactam no valor nutricional e na aceitação comercial. A redução do valor nutricional pode ocorrer devido ao desgaste dos nutrientes, como proteínas, vitaminas e minerais, provocada por fungos (Costa; Silva, 2018).

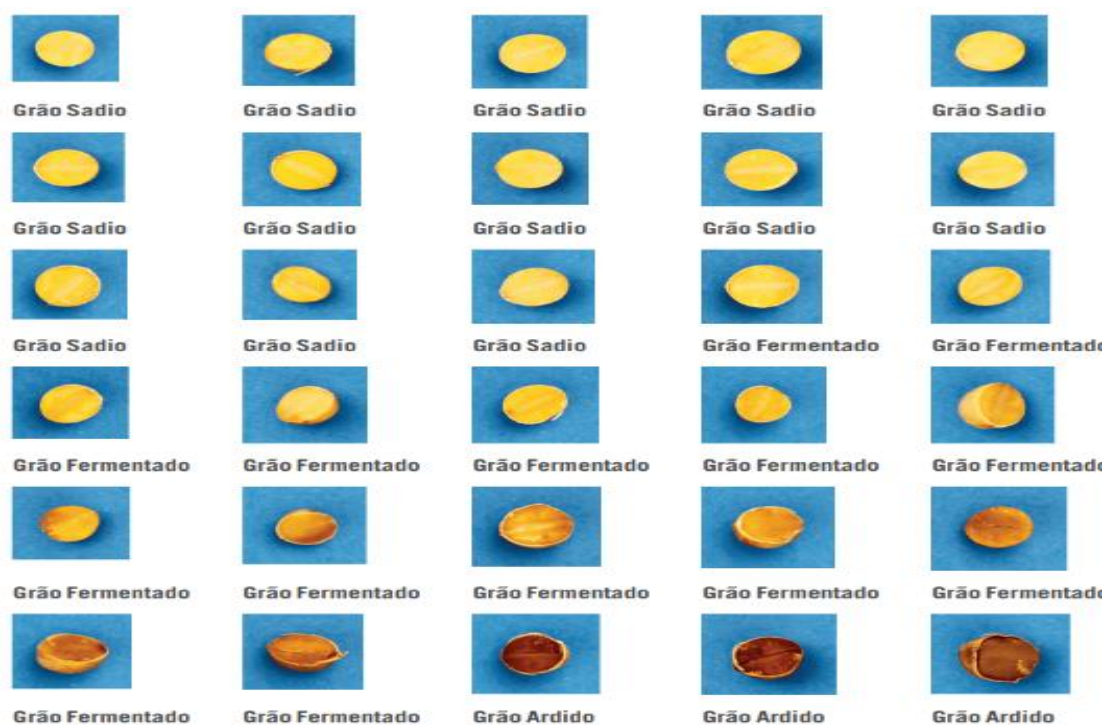
Os grãos defeituosos podem ser contaminados por micotoxinas, produzidas por fungos, cuja ingestão pode acarretar problemas de saúde, levando a uma diminuição da aceitação no mercado, especialmente para consumo humano (Junqueira; Tavares, 2020). A qualidade do produto é afetada, uma vez que grãos com defeitos físicos, como rachaduras ou manchas, são considerados de qualidade inferior, resultando na diminuição do valor comercial e impactando na utilização como produtos alimentícios e/ou ração animal (Nascimento; Oliveira, 2019).

Dependendo da destinação, grãos de consumo humano são avaliados com rigor na aparência e qualidade; assim, grãos com defeitos podem ser redirecionados para ração animal ou para a produção de biodiesel, onde a qualidade é menos crítica, mas ainda assim afetada por contaminações (Borges; Ribeiro, 2021). A presença de defeitos diminui o valor de venda e gera custos adicionais para produtores e comerciantes, podendo gerar inspeções, limpeza ou até descarte de lotes inteiros que não atendam aos padrões de qualidade (Martins e Pereira, 2020).

A implementação de controles rigorosos e práticas adequadas em cada fase do processo agrícola garante a conformidade com padrões de qualidade e minimiza o impacto negativo sobre a comercialização e o valor dos produtos. A eficácia na detecção e correção desses defeitos ajuda a manter a confiança dos consumidores e assegurar a competitividade no mercado global (Vieira; Melo, 2020).

Durante a classificação, diversos defeitos podem ser identificados nos grãos de soja (Figura 3), esses defeitos podem ocorrer em qualquer etapa do processo agrícola e de distribuição, e têm a capacidade de comprometer a aparência e a qualidade dos grãos. Entre os problemas comuns estão a presença de grãos danificados, moídos ou contaminados, os quais afetam a estética do produto e sua qualidade nutricional e segurança alimentar (Abiove, 2024).

Figura 3. Imagens de Defeitos de Grãos de Soja.



Fonte: Abiove, 2024.

Gurieva, Tarasova e Khaba (2023) destacam que durante a colheita, o armazenamento e o transporte, práticas inadequadas de manejo podem resultar em diversos defeitos, afetando a qualidade e o valor comercial dos grãos. Dentre os principais problemas estão as quebras causadas pelo manuseio impróprio; a umidade excessiva durante o armazenamento; e o acúmulo de calor devido ao armazenamento inadequado, podendo promover o crescimento de microorganismos e deteriorar a qualidade dos grãos. Vale destacar a importância de práticas de manejo adequadas para minimizar esses defeitos.

Para reduzir os problemas relacionados à qualidade dos grãos, os produtores de soja podem adotar estratégias preventivas e corretivas ao longo do processo de produção e distribuição, incluindo o uso de práticas agronômicas avançadas para controlar pragas e doenças, bem como o monitoramento das condições de armazenamento e transporte. Com a implementação de tecnologias, como sensores e sistemas de monitoramento em tempo real é possível otimizar essas práticas, garantindo um ambiente adequado para o armazenamento e transporte dos grãos (Souza; Almeida; Pereira, 2023).

2.3. GESTÃO DE DADOS AGRÍCOLAS

A compreensão e enfrentamento de desafios na gestão de dados agrícolas geram eficiência e avanço do setor. O uso de sensores integrados para o monitoramento de pragas e doenças nas lavouras, com o uso de redes de sensores sem fio, resulta na melhoria da detecção precoce de infestações, reduzindo a necessidade de uso intensivo de pesticidas (Rodríguez *et al.*, 2020).

A análise de *big data* aplicada à gestão de fazendas de larga escala, explora como os dados climáticos históricos e previsões futuras podem otimizar o momento do plantio, maximizando a produtividade e reduzindo desperdícios. Ao utilizar dados climáticos históricos e previsões futuras, é possível identificar o melhor momento para o plantio, resultando em um aumento da produtividade das culturas, com redução no uso de insumos, como água e fertilizantes (Smith; Johnson, 2019).

O uso de inteligência artificial permite prever a produtividade agrícola com base em imagens de satélite, com ajustes na aplicação de fertilizantes e irrigação de acordo com a demanda exata de cada área cultivada. O modelo desenvolvido ajusta a aplicação de fertilizantes e a irrigação de acordo com as demandas específicas de cada área de cultivo, resultando em redução no uso de água e insumos, sem comprometer a produtividade. Além disso, os resultados indicaram melhoria na previsão de colheita (Almeida *et al.*, 2021).

Mesmo com tantos avanços, os agricultores e pesquisadores enfrentam obstáculos ao tratar da coleta, organização e análise de dados agrícolas dispersos e heterogêneos (Wang, *et al.*, 2019). Outro desafio consiste em enfrentar a diversidade de fontes de dados; onde agricultores e pesquisadores tratam com informações provenientes de uma variedade de fontes, incluindo sistemas de monitoramento, sensores de campo, registros manuais e registro de imagens. Integrar e consolidar esses dados de maneira eficiente é uma tarefa complexa, requerendo tempo e recursos (Brown; Bell, 2019).

Os dados agrícolas podem variar em formato, estrutura e qualidade, dificultando sua integração e análise. Por exemplo, dados coletados por diferentes tipos de sensores podem ser formatados de maneira diferente, tornando difícil combiná-los para análises integradas, e podendo levar a inconsistências e imprecisões, comprometendo a confiabilidade dos resultados (Yu *et al.*, 2018).

A escala e o volume dos dados agrícolas gerados diariamente é um dos desafios. Com o avanço da tecnologia de sensoriamento remoto e automação agrícola, a quantidade dos dados coletados está aumentando exponencialmente (Zhao *et al.*, 2019). Tratar com grandes volumes de dados de maneira eficiente requer infraestrutura e ferramentas adequadas, o que nem sempre estão disponíveis para os produtores rurais e pesquisadores (O'neal *et al.*, 2018).

A falta de padronização dos dados agrícolas é outro desafio. Os dados coletados por diferentes dispositivos e sistemas podem seguir padrões diferentes de formatação e estrutura, dificultando sua interoperabilidade e compartilhamento, podendo resultar em silos de dados, onde informações importantes ficam inacessíveis (Daniel; Tschorsch, 2021).

Silos de dados referem-se às estruturas ou sistemas isolados, onde dados são armazenados de forma fragmentada, dificultando o acesso e a integração de dados entre diferentes departamentos ou áreas de uma organização, como resultado. Os dados ficam encapsulados em silos, limitando a capacidade da organização de acessar informações relevantes, podendo ser utilizadas para tomada de decisões estratégicas (Lopes; Almeida; Sousa, 2019).

A existência de silos de dados pode levar a redundâncias, inconsistências e falta de colaboração entre as equipes, quando os dados não são compartilhados ou integrados. Há o risco que decisões sejam baseadas em informações incompletas ou desatualizadas, sendo problemático em setores dependentes fortemente de dados, como a agricultura, onde a capacidade de analisar e reagir às informações em tempo real pode impactar na produtividade e na eficiência das operações (Furlan; Piazza; Moreira, 2017).

A falta de capacidade técnica e conhecimento também é um desafio, onde nem todos os produtores e pesquisadores têm acesso a recursos ou treinamento adequados para coletar, organizar e analisar dados de maneira eficaz, podendo levar a subutilização de dados e dificultar a tomada de decisões (Dagar *et al.*, 2021).

Em face desses desafios, é necessário a busca por soluções inovadoras e colaborativas para a gestão de dados, com a implementação de padrões de formatação e interoperabilidade, investimentos em infraestrutura tecnológica e a promoção do desenvolvimento de habilidades e capacidades.

2.4 IMPACTOS CLIMÁTICOS NA PRODUÇÃO AGRÍCOLA

A agricultura está enfrentando uma série de desafios devido à mudança climática (Lobell *et al.*, 2018). A análise da mudança climática na produção agrícola tem demonstrado influência de fatores como temperatura, umidade e padrões de precipitação nas safras futuras (Figura 4). A compreensão desses padrões permite o planejamento agrícola e a adoção de práticas de manejo adequadas.

Figura 4. Perfil da ameaça climática à segurança da América do Sul, Central e Caribe, cenário de curto, médio e longo prazo.



Fonte: Távora, França e Lima (2022) adaptado de TCCS (2020).

Silva *et al.* (2021) discutem as alterações nos padrões climáticos, o aumento da temperatura, os padrões de precipitação irregulares e os eventos climáticos extremos que estão impactando a produção agrícola. Os fenômenos climáticos El Niño e La Niña afetam a dinâmica atmosférica e oceânica, influenciando os padrões climáticos globais e, consequentemente, a agricultura em diversas regiões.

El Niño e La Niña (Figura 5) são fenômenos climáticos, resultantes de variações na temperatura da superfície do mar, no Oceano Pacífico equatorial, influenciando os padrões climáticos globais, El Niño ocorre quando as águas do Pacífico equatorial aquecem anormalmente, afetando os ventos alísios e alterando a circulação atmosférica, podendo ocasionar mudanças nas precipitações, aumentando a probabilidade de secas em algumas regiões e chuvas intensas em outras. Por outro lado, La Niña é caracterizada pelo resfriamento das águas do Pacífico equatorial, reforçando os ventos alísios e resultando em padrões climáticos opostos, como chuvas acima da média em áreas tropicais e secas em outras regiões, especialmente na costa oeste das Américas (Mcculloch, 2021).

Figura 5. Transição de El Niño para La Niña.



Fonte: Agros, 2024.

As mudanças climáticas estão agravando a frequência e a intensidade de eventos climáticos extremos, incluindo El Niño e La Niña (Hussain; Ahmad; Rashid, 2021). O aumento das temperaturas globais, causado pela emissão de gases de efeito estufa, tem a capacidade de influenciar os padrões oceânicos e atmosféricos que governam esses fenômenos. Por exemplo, temperaturas mais elevadas da superfície do mar podem intensificar o El Niño, aumentando a probabilidade de ocorrências severas. Além disso, a alteração das correntes oceânicas e dos ventos alísios pode levar a um comportamento menos previsível de La Niña, tornando a previsão e a gestão dos impactos climáticos desafiadoras (Kumar *et al.*, 2022).

A inversão dos fenômenos climáticos El Niño e La Niña pode ter efeitos na agricultura, pois altera os padrões de precipitação e temperatura. Durante um evento El Niño, os agricultores podem enfrentar secas severas em áreas tradicionalmente férteis, resultando em baixas colheitas e aumento da vulnerabilidade a pragas e doenças. Além disso, a ocorrência de chuvas intensas pode causar inundações, afetando a qualidade do solo e a saúde das culturas. Em contrapartida, La Niña pode causar chuvas excessivas em regiões sofrendo com a saturação do solo, comprometendo a colheita e causando erosão (Tadesse, *et al.*, 2022).

Silva *et al.*, (2019) evidenciaram que o aumento da temperatura altera os ciclos de crescimento das plantas, afetando o tempo de floração e frutificação, além das chuvas irregulares e as secas prolongadas prejudicarem o desenvolvimento das plantas, resultando em menores rendimentos e qualidade inferior das produções.

Além disso, tempestades severas e ondas de calor, podem causar danos às lavouras, resultando em perdas para os agricultores.

Diante desse cenário, a agricultura precisa se adaptar às mudanças climáticas. Uma abordagem para enfrentar esse desafio é a centralização de dados agrícolas e a utilização de tecnologias de análise de dados para prever e diminuir os impactos da mudança climática na produção agrícola (Valin *et al.*, 2020).

Com os dados centralizados, os agricultores podem tomar decisões proativas para gerenciar suas lavouras, em face às mudanças climáticas. Por exemplo, a análise de históricos de padrões climáticos pode ajudar os agricultores na identificação de períodos de risco para eventos climáticos extremos e medidas preventivas, como investir em infraestrutura de irrigação e implementar práticas agrícolas resistentes ao clima (Ferreira *et al.*, 2020).

A utilização de modelos de previsão climática pode ajudar os agricultores na antecipação de mudanças nas condições climáticas e ajustar suas estratégias de cultivo. Se uma previsão indicar um período de seca prolongada, os agricultores podem adotar práticas de conservação de água e selecionar variedades de cultura tolerantes à seca (Harrison *et al.*, 2020).

Santos *et al.*, (2021) ressaltam a colaboração entre agricultores, pesquisadores, instituições acadêmicas e governamentais, para enfrentar os desafios impostos pela mudança climática na agricultura. Essa colaboração inclui a coleta e compartilhamento de dados, o desenvolvimento de modelos de previsão climática precisos e a implementação de estratégias de adaptação climática nas práticas agrícolas, bem como o desenvolvimento de políticas públicas.

A mudança climática representa ameaça na produção agrícola, no entanto, ao centralizar dados e adotar abordagens baseadas em análise de dados, os agricultores podem se posicionar de forma resiliente, a fim de garantir a segurança alimentar e a sustentabilidade agrícola a longo prazo (Zhang *et al.*, 2022).

2.5 CENTRALIZAÇÃO DE DADOS

A centralização de dados na agricultura tem ganho destaque, principalmente devido à crescente quantidade de informações geradas por tecnologias integradas nas práticas agrícolas. A implementação de tecnologias de sensoriamento remoto, Internet das Coisas (IoT), estações meteorológicas e dispositivos inteligentes cria um grande volume de dados que, se mal gerenciado, pode levar a desafios na análise e

utilização desses dados. Diante disso, a centralização de dados é considerada uma solução para otimizar a tomada de decisão e a gestão das operações agrícolas (Martins *et al.*, 2019).

Quando os dados não são centralizados, torna-se difícil para os agricultores acessarem informações precisas e atualizadas, o que prejudica a eficiência das operações agrícolas. A centralização de dados promove a simplificação do acesso e gestão das informações e contribui para a transparência e a rastreabilidade na cadeia produtiva agrícola, útil no contexto da segurança alimentar, onde a precisão dos dados garante práticas agrícolas de acordo com os padrões de qualidade exigidos (Carvalho *et al.*, 2020).

Além disso, a centralização de dados facilita a colaboração entre diferentes agentes do setor agrícola. Ao centralizar as informações é possível acessar os mesmos dados e trabalhar de maneira colaborativa na identificar soluções para melhoria da produtividade agrícola (Pereira *et al.*, 2019). O compartilhamento de dados é chave para o avanço da agricultura de precisão. A centralização dos dados permite que modelos de análise preditiva e ferramentas de *big data* sejam aplicados de maneira eficaz, aumentando a capacidade para identificar padrões e prever tendências climáticas e de produção (Basso; Antle, 2020).

A centralização e o uso de tecnologias de análise de dados, como inteligência artificial e aprendizado de máquina, podem otimizar a tomada de decisão ao permitir uma análise rápida e precisa dos dados coletados em campo, relevante para a gestão de recursos como água, fertilizantes e pesticidas, uma vez que permite aos agricultores agirem em resposta às variáveis que impactam diretamente a produtividade das colheitas (Rose *et al.*, 2018).

A centralização de dados promove a integração das informações provenientes de diferentes fontes, resultando uma visão abrangente das operações agrícolas, fundamental para promover a sustentabilidade no setor agrícola, uma vez que o uso eficiente de dados pode ajudar a reduzir desperdícios e otimizar o uso de insumos agrícolas (Duncan *et al.*, 2021). A capacidade de centralizar e analisar dados em tempo real pode ser um fator relevante para a resiliência das práticas agrícolas (Huang *et al.*, 2020).

A implementação de sistemas centralizados requer investimentos em infraestrutura e capacitação tecnológica, no entanto, os benefícios em termos de eficiência e produtividade podem compensar esses custos a longo prazo,

especialmente à medida que as tecnologias se tornam mais acessíveis (Pagliari *et al.*, 2019). Há a necessidade de garantir que os sistemas de centralização de dados sejam robustos e seguros, a fim de proteger a integridade e a confidencialidade das informações sensíveis do setor agrícola (Jones *et al.*, 2021).

A centralização de dados tem implicações nas políticas públicas. A implementação de políticas para adoção de tecnologias de centralização de dados pode promover uma agricultura sustentável e eficiente, ao fornecer subsídios e incentivos financeiros para a adoção de tecnologias de centralização. Os governos podem ajudar a superar as barreiras econômicas e tecnológicas que dificultam a integração de sistemas de dados agrícolas (Silva *et al.*, 2020).

Dessa forma, a centralização de dados é importante para o avanço da agricultura. Ao proporcionar maior transparência, eficiência e colaboração entre os diferentes atores da cadeia produtiva, contribui para uma agricultura mais sustentável e resiliente, capaz de enfrentar os desafios crescentes impostos pelas mudanças climáticas e pelas demandas globais por segurança alimentar.

2.6. DATA WAREHOUSE

Segundo Paliwal e Saraswat (2022), um *data warehouse* é uma área de armazenamento de dados, que permite o gerenciamento de atividades de *business intelligence*, especificamente consultas complexas e análises de grandes volumes de dados. Há integração de dados de várias fontes heterogêneas em um repositório centralizado e organizado, proporcionando uma visão unificada e consistente das informações. Os *data warehouses* são otimizados para a leitura de dados e são utilizados para armazenar dados históricos, permitindo análises de tendências e a geração de relatórios detalhados para auxiliar a tomada de decisões estratégicas dentro das organizações.

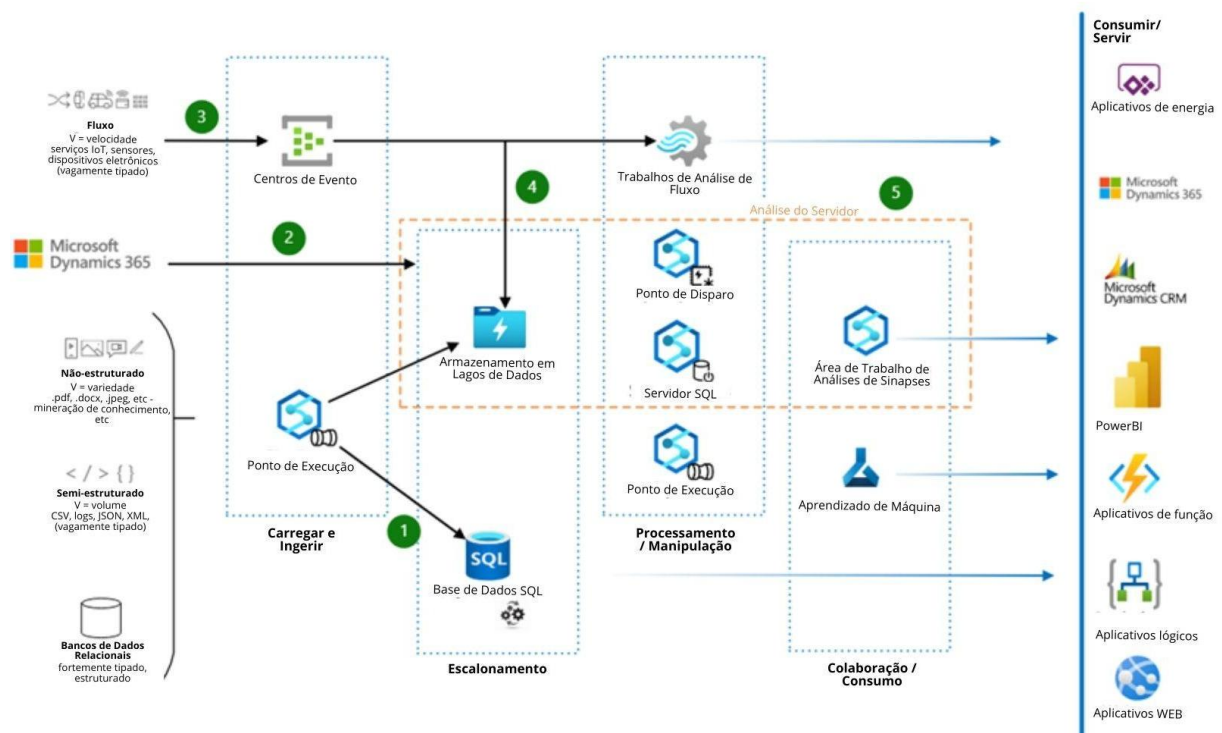
Business Intelligence (BI) é o conjunto de processos, tecnologias e ferramentas que transformam dados brutos em informações para a tomada de decisões estratégicas (Sharma; Agarwall, 2021). No setor agrícola, sua importância está na necessidade de decisões baseadas em dados, aproveitando informações geradas por sensores, drones e sistemas de IoT para otimizar o plantio, irrigação e colheita (Shukla; Kapoor; Jain, 2020).

O BI melhora a eficiência operacional, auxiliando na gestão de recursos como água e fertilizantes, e ajuda prever tendências e riscos, como mudanças climáticas ou

flutuações nos preços. Além disso, o BI promove práticas agrícolas sustentáveis, como o uso consciente da água e a rotação de culturas, contribuindo para a preservação ambiental. Também, é relevante na rastreabilidade e conformidade, garantindo que agricultores cumpram as regulamentações de segurança alimentar e monitorem todas as etapas da cadeia produtiva, facilitando auditorias e garantindo a qualidade dos produtos (Villas-Boas *et al.*, 2021).

O fluxo de informações em um *data warehouse* (Figura 6) é um processo estruturado, envolvendo várias etapas para coleta, transformação e armazenamento de dados provenientes de diferentes fontes (Microsoft Learn, 2024). O processo inicia com extração de dados de bases operacionais, bancos de dados transacionais e fontes heterogêneas, como arquivos de log, aplicativos de ERP (*Enterprise Resource Planning*) e sistemas CRM (*Customer Relationship Management*).

Figura 6. Fluxo de Informações em um data warehouse.



Fonte: Adaptado de Microsoft Learn, 2024.

Na etapa de extração, os dados são coletados de suas fontes originais e preparados para a próxima fase, um passo crítico, pois os dados podem estar em formatos variados e precisar de uma padronização inicial (Mannino, 2020). Ferramentas de coleta de dados como sensores, API's, entre outros, são utilizadas para realizar essa função, garantindo a extração de dados de maneira eficiente e segura (Ta'a *et al.*, 2022).

A etapa de transformação, dados extraídos passam por limpeza, filtrados e formatados de acordo com as demandas do ambiente de *data Warehouse*. Os dados são normalizados ou desnormalizados, agregados, e enriquecidos para garantir consistência e qualidade, onde as transformações incluem remoção de duplicatas, correção de valores incorretos e conversão de unidades de medida, assegurando precisão e consistência dos dados, para a etapa de análise (Mannino, 2020).

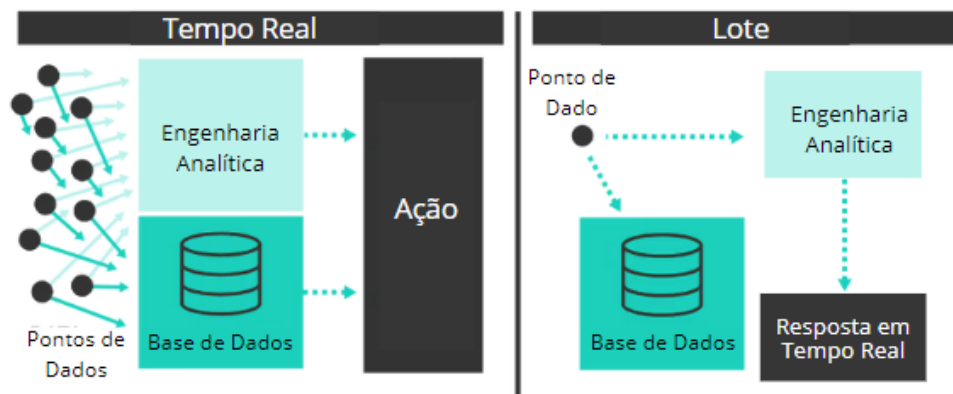
Após a transformação, os dados são carregados no *data warehouse*, envolvendo a inserção de dados processados em estruturas de armazenamento específicas, otimizando o acesso e a análise (Mannino, 2020). Essas estruturas incluem as tabelas fato, contendo informações base para análises. Paralelamente, existem as tabelas dimensão, usadas para armazenar atributos descritivos, como categorias de produtos, datas ou localização, permitindo aos dados serem organizados de forma a facilitar a categorização e a consulta (Marr, 2021).

O processo de carregamento de dados no *data warehouse* pode variar dependendo do volume de dados e da frequência de como esses dados precisam ser atualizados, quando se trata de grandes volumes de dados ou situações em que as atualizações não são imediatas. É comum a utilização do método de carregamento em lotes (*batch loading*).

Nesse cenário, os dados são acumulados ao longo de um período, como horas ou dias, e então processados e inseridos no *data warehouse* em uma única operação. Esse método é eficiente para sistemas que podem lidar com um certo atraso nas atualizações, permitindo carregar grandes quantidades de dados de uma só vez, sem sobrecarregar o sistema constantemente (Suresh; Saravanan, 2021).

Por outro lado, o carregamento em tempo real (*real-time loading*) é utilizado quando é necessário manter os dados atualizados e disponíveis para os usuários. Os dados são inseridos no *data warehouse* conforme são gerados, sem intervalos significativos. Esse método é essencial para cenários onde decisões precisam ser tomadas com base em dados atualizados continuamente, como em sistemas monitoradores de condições climáticas, mercados financeiros, ou operações logísticas, onde qualquer atraso pode prejudicar a eficiência das decisões (Gupta e Sharma, 2021). A principal diferença entre *batch loading* e *real-time loading* reside na forma como os dados são processados (Figura 7)

Figura 7. Diferença entre carregamento em tempo real e carregamento em lotes.



Fonte: Adaptado de Richman (2024).

A escolha entre carregamento em lotes e em tempo real depende das necessidades específicas do sistema e dos usuários. Para situações que exigem dados imediatos e em constante atualização, o *real-time loading* é mais adequado. Já para casos em que a análise pode ser feita de forma periódica, o *batch loading* é mais eficiente, pois economiza recursos de processamento e evita atualizações contínuas, podendo sobrecarregar o sistema.

2.6.1 Data Warehouse E Sua Aplicação Na Agricultura

A aplicação de tecnologias de data warehouse na agricultura integra e gerencia grandes volumes de dados agrícolas, oferecendo armazenamento, processamento e análise de informações oriundas de múltiplas fontes, e oferece conhecimentos aos agricultores e gestores de agronegócio. A implementação de *data warehouse* na agricultura visa melhorar a eficiência operacional, otimizar a gestão de recursos e facilitar a tomada de decisões (Paliwal; Saraswat, 2022).

A introdução de *data warehouses* na agricultura representa um avanço no gerenciamento de dados agrícolas em larga escala. Essas ferramentas são projetadas para centralizar e organizar grandes volumes de dados de maneira eficiente, proporcionando análises avançadas e tomada de decisões no setor agrícola, capazes de tratar dados heterogêneos e complexos provenientes de diversas fontes, incluindo sistemas de monitoramento automatizados, sensores agrícolas, registros manuais e imagens (Machado *et al.*, 2020)

Os data warehouses integram e consolidam dados, oriundos de diferentes formatos e estruturas, em um único local acessível, permitindo análise integrada dos dados agrícolas, proporcionando uma visão ampla e coesa da produção. Os agentes

do setor agrícola podem acessar facilmente essas informações e realizar análises detalhadas sobre vários aspectos da atividade, como produtividade, saúde das plantas, condições ambientais e eficiência operacional (Smith, 2021).

Com a capacidade de armazenar o histórico de dados agrícolas, permite análise longitudinal dos dados e possibilita a avaliação de tendências sazonais, variações climáticas e impactos de intervenções agrícolas ao longo das safras. Essa perspectiva temporal permite compreensão do desempenho agrícola ao longo do tempo e identificação de oportunidades de melhoria (Machado, 2022).

A aplicação dessa tecnologia permite a centralização de informações sobre as condições ambientais que afetam o cultivo, incluindo dados meteorológicos, como temperatura, umidade e precipitação, informações sobre o solo e o relevo da área de plantio (Karanikolas *et al.*, 2021). Ao integrar esses dados com informações sobre o ciclo de cultivo é possível avaliar o impacto das condições ambientais na saúde e no rendimento das plantas, permitindo ajustes oportunos nas práticas de manejo.

Os *data warehouses* permitem aos agricultores o armazenamento e análise de imagens, fotografias aéreas e dados de sensoriamento remoto para avaliar a saúde das plantas ao longo do ciclo de cultivo, podendo ajudar na identificação precoce de doenças, pragas ou deficiências nutricionais, e permitindo intervenções rápidas e eficazes para proteger e promover o crescimento saudável das plantas (Jones *et al.*, 2021).

Outro benefício de *data warehouse* é sua utilização no monitoramento e otimização do uso de insumos agrícolas, como fertilizantes e defensivos. Ao centralizar informações da aplicação de insumos e condições do solo, os agricultores avaliam a eficácia das práticas de manejo e identificar oportunidades de redução de custos e diminuição do impacto ambiental (Silva *et al.*, 2022).

O data Warehouse se aplica a toda cadeia produtiva, dando suporte à rastreabilidade e à certificação de produtos. Ao registrar e rastrear dados do processo de produção, demonstram a conformidade com padrões de qualidade e segurança alimentar, agregando valor aos produtos (Dupont *et al.*, 2021).

2.6.2 Modelagens Multidimensionais

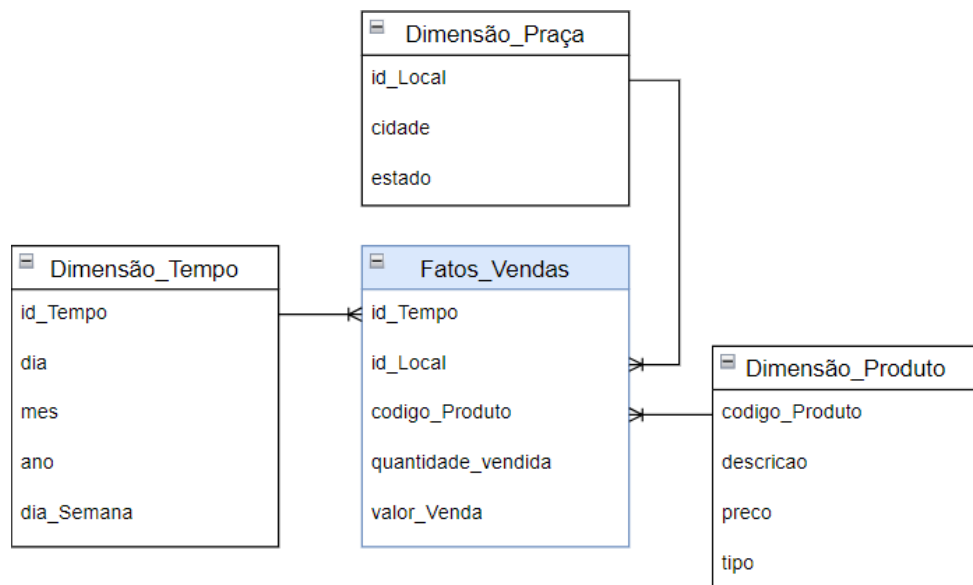
A modelagem multidimensional é uma abordagem utilizada na organização e análise de dados em sistemas de data warehouse, facilitando a visualização e compreensão das informações por meio de uma estrutura que permite múltiplas

perspectivas de análise. Esse tipo de modelagem é eficaz para dar suporte às decisões de negócios, organizando os dados em um formato facilitador de consultas complexas e análises detalhadas (Golfarelli; Gokcel, 2020).

Dentro do *data warehouse*, os dados são organizados em esquemas, geralmente em forma de estrela ou floco de neve, facilitando a navegação e a realização de consultas complexas (Akid *et al.*, 2022). Esses esquemas estruturam os dados de maneira lógica, permitindo análise eficiente e geração de relatórios.

No contexto de *data warehouses*, os esquemas estrela e floco de neve são modelos multidimensionais para organização de dados para otimizar a eficiência das consultas e análises (Akid *et al.*, 2022). O esquema estrela (Figura 8) é comumente utilizado para organizar dados com uma tabela de fatos central e várias tabelas dimensões ao seu redor.

Figura 8. Modelo Estrela.

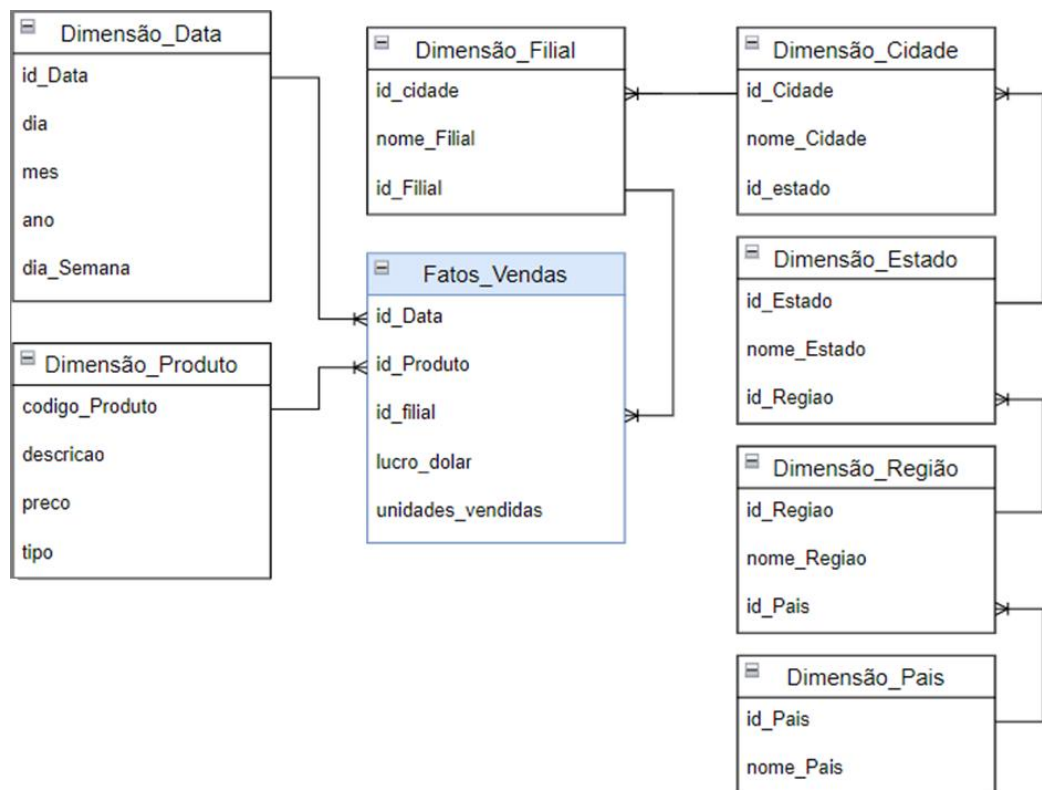


Fonte: Adaptado de Fonseca (2023).

No modelo estrela, a tabela de fatos quantifica os dados analisados, como vendas, transações ou eventos. Cada registro na tabela de fatos contém chaves estrangeiras, referenciando as tabelas dimensões e métricas quantitativas, como quantidade vendida ou receita gerada. As tabelas dimensões armazenam dados descritivos relacionados às métricas da tabela de fatos. O esquema estrela facilita a execução de consultas complexas devido às tabelas dimensões acessíveis pela tabela de fatos, permitindo execução junções de forma eficiente (Fonseca, 2023).

O esquema em floco de neve (Figura 9) é uma extensão do esquema estrela com tabelas dimensões normalizadas e detalhadas. Diferente do esquema estrela, as tabelas dimensões no esquema em floco de neve são detalhadas e normalizadas. As informações nas tabelas dimensões são divididas em múltiplas tabelas, eliminando redundâncias. Uma tabela dimensão de produtos é dividida em tabelas, armazenando categorias de produtos e detalhes de fornecedores (Ciferri, 2022).

Figura 9. Modelo Floco de Neve.



Fonte: Adaptado de Ciferri, 2022.

A normalização é um processo importante no desenvolvimento de bancos de dados, visando organizar os dados de forma a reduzir a redundância e melhorar a integridade, sendo dividida em três formas normais (Date, 2019).

Uma tabela está na primeira forma normal (1FN) se todos os seus atributos contêm apenas valores atômicos, ou seja, cada coluna deve armazenar um único valor, e não um conjunto de valores ou uma lista. Além disso, cada linha deve ser única, alcançado através de uma chave primária. Para alcançar a 1FN, é necessário eliminar colunas contendo múltiplos valores e criar uma chave primária para identificar exclusivamente cada registro.

Uma tabela está na segunda forma normal (2NF) se estiver na 1NF e todos os atributos não-chave forem completamente dependentes da chave primária, não deve haver dependências parciais, onde um atributo depende apenas de parte da chave primária. Para atingir a 2NF, deve-se identificar essas dependências parciais e dividir a tabela em mais tabelas, onde cada tabela contém apenas informações que dependem completamente da chave primária.

Uma tabela está na terceira forma normal (3FN) se estiver na 2NF e não houver dependências transitivas entre os atributos não-chave. Isso significa que um atributo não-chave não deve depender de outro atributo não-chave. Para alcançar a 3NF, é necessário identificar e remover essas dependências transitivas, criando tabelas que relacionem os dados de maneira adequada.

A normalização até a terceira forma normal garante a integridade e a consistência dos dados em um sistema de banco de dados, embora a normalização possa levar a consultas complexas devido à necessidade de mais junções entre tabelas. Também, proporciona economia de espaço de armazenamento e uma redução na redundância de dados, tornando a manutenção e a atualização das informações mais eficientes.

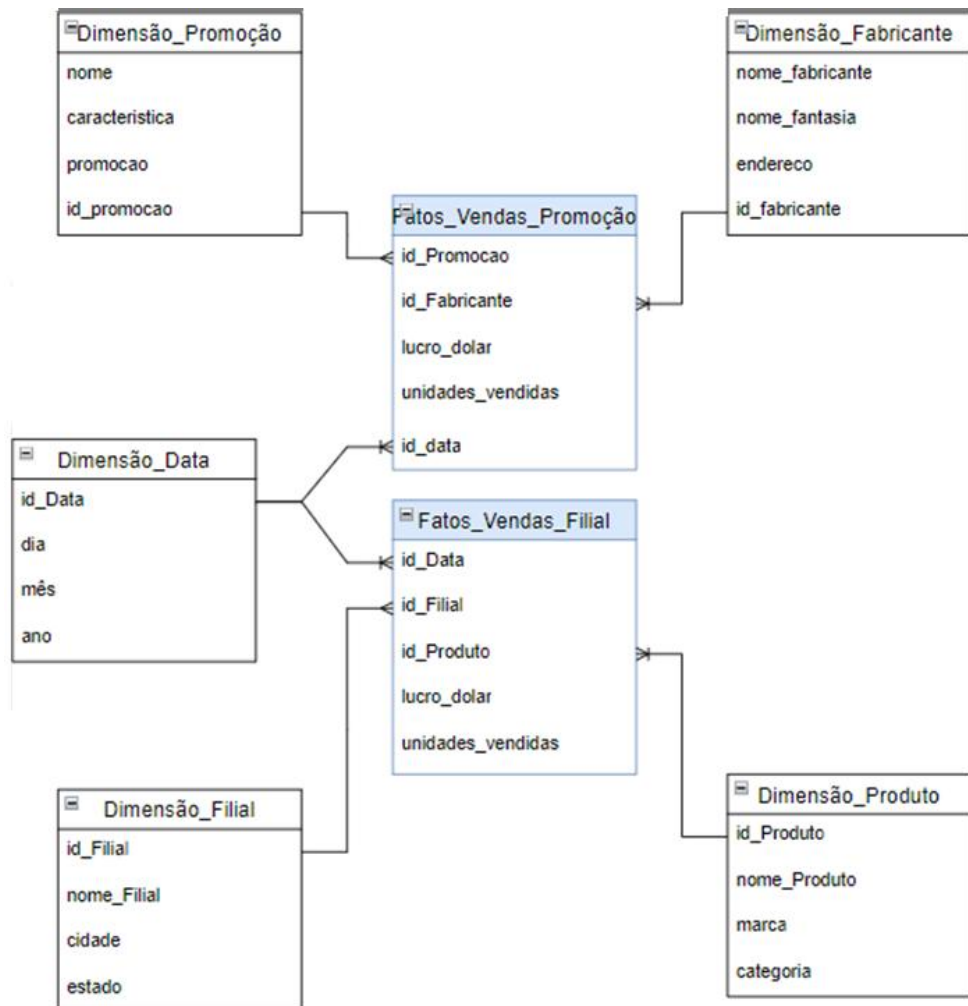
Enquanto a normalização até a terceira forma normal é necessária para assegurar a integridade e a consistência dos dados, para a escolha entre um esquema estrela e um floco de neve é necessário equilibrar a simplicidade e a eficiência das consultas com a normalização e a redução da redundância de dados (Vassiliadis; Sellis, 2020).

O esquema estrela é simples e eficiente para consultas, pois requer menos junções entre tabelas, sendo ideal para *data warehouses* onde a simplicidade e a velocidade das consultas são prioritárias. Por outro lado, o esquema floco de neve oferece maior normalização e economia de espaço, sendo útil em cenários passíveis de violação de integridade dos dados e redundância de dados. Ambos os esquemas, estruturam os dados de modo diferente para atender às demandas e objetivos da análise de dados (Date, 2019).

É possível mesclar os esquemas estrela e floco de neve com criação de modelo híbrido, esquema constelação ou estendido (Figura 10), no esquema híbrido, algumas dimensões podem ser organizadas de forma simples e direta como no esquema estrela, onde cada dimensão está ligada à tabela de fatos. Outras dimensões, no entanto, podem ser normalizadas como no floco de neve, onde as

informações são divididas em várias tabelas para reduzir redundâncias e melhorar a integridade dos dados (Ni *et al.*, 2022).

Figura 10. Modelo Constelação.



Fonte: Adaptado de Ciferri, 2022.

O modelo constelação permite consultas que exigem rapidez e simplicidade possam utilizar a modelo estrela, enquanto as que demandam análise detalhada e precisa possam utilizar o modelo floco de neve. Dessa forma, o esquema híbrido oferece flexibilidade maior, balanceando a simplicidade e a eficiência das consultas com a integridade e a consistência dos dados (Ciferri, 2022).

2.7 COLETA DE DADOS EM ESTAÇÕES METEOROLÓGICAS

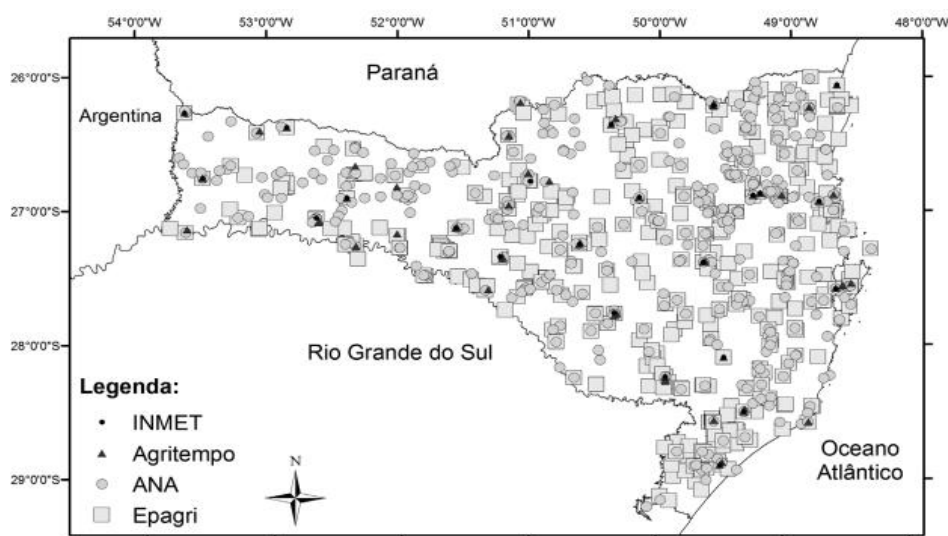
A coleta de dados em estações meteorológicas permite a obtenção de informações precisas sobre as condições atmosféricas, necessárias para análise e tomada de decisão em diversas áreas, incluindo a agricultura. Este processo envolve

a instalação e manutenção de equipamentos, a coleta e transmissão de dados em tempo real e a análise dos dados para gerar informações (WMO, 2021).

As estações meteorológicas são compostas por uma variedade de sensores e instrumentos, medidores de diferentes parâmetros atmosféricos. Os equipamentos típicos incluem termômetros para medir a temperatura do ar, higrômetros para avaliar a umidade relativa, barômetros para monitorar a pressão atmosférica, anemômetros para medir a velocidade do vento e pluviômetros para quantificar a precipitação (Oliveira *et al.*, 2021). A escolha e a calibração dos sensores garantem a precisão dos dados coletados (Wei *et al.*, 2019).

A instalação de estações meteorológicas deve ser realizada em locais capazes de garantir a precisão e a representatividade dos dados (Figura 11). A localização é escolhida com base em critérios como a ausência de obstruções passíveis de afetar medições, a estabilidade do solo e a distância de fontes de poluição. Após a instalação, será realizada manutenção regular para assegurar o funcionamento dos equipamentos de forma correta e os dados coletados precisos. Isso inclui a calibração periódica dos sensores, a limpeza dos componentes e a verificação dos sistemas de transmissão de dados. A manutenção preventiva ajuda a minimizar erros e garantir a continuidade da coleta de dados (Vianna *et al.*, 2017).

Figura 11. Mapa das estações meteorológicas do INMET, do Agridempo, da ANA e da Epagri, de acordo com as coordenadas disponíveis nos respectivos.



Fonte: Vianna *et al.*, 2017.

Uma característica das estações meteorológicas é a capacidade de coletar e transmitir dados em tempo real. Os sensores medem continuamente os parâmetros

atmosféricos e enviam as informações para uma unidade central de processamento, podendo estar localizada na própria estação ou em um centro de controle remoto. A transmissão dos dados pode ser realizada por meio de diversos métodos, incluindo redes celulares, satélites ou conexões via internet. A coleta em tempo real permite monitoração contínua das condições meteorológicas e tomada de decisões rápida, especialmente em situações de eventos climáticos extremos (Fu *et al.*, 2023).

Os dados coletados são armazenados em bancos de dados, onde são organizados e processados para análise posterior. O armazenamento pode ser feito localmente na estação ou em servidores remotos, dependendo da configuração do sistema. O processamento dos dados envolve a aplicação de algoritmos para filtrar e analisar as informações, a fim de gerar relatórios e gráficos representativos das condições atmosféricas (Fu *et al.*, 2023). Técnicas de processamento de dados incluem correção de erros e interpolação de valores para preencher lacunas nas séries temporais (Hyndman e Athanasopoulos, 2021).

Para obter uma visão completa das condições climáticas, os dados das estações meteorológicas são frequentemente integrados com informações de outras fontes, como imagens de satélite, modelos climáticos e dados de estações meteorológicas vizinhas (Zhao; Chen; Whang, 2020).

A integração de diferentes fontes de dados permite análise robusta e precisa das condições atmosféricas e suas implicações para a agricultura, com técnicas de fusão de dados e modelos de assimilação sendo utilizadas para combinar informações de diversas fontes e melhorar a precisão das previsões e análises (Pawłowski; Wróblewska; Sysko-Romańczuk, 2023).

Na agricultura, os dados coletados em estações meteorológicas são utilizados para monitorar condições que afetam diretamente a produção de culturas. Informações sobre temperatura, umidade e precipitação permitem previsão de doenças, planejamento de irrigação e escolha de práticas de manejo. A umidade do solo e a temperatura do ar podem influenciar o desenvolvimento de pragas e doenças nas plantas, enquanto os dados sobre precipitação ajudam a determinar as necessidades de irrigação (Ogunjo *et al.*, 2021).

A compreensão das variáveis ambientais e suas influências no desenvolvimento de pragas e doenças nas plantas é necessária para a gestão agrícola eficiente. O Quadro 1 resume como fatores apresentados por Ogunjo *et al.* (2021) afetam a saúde

das culturas e apresenta aplicações práticas, para adoção por parte dos agricultores, para otimizar suas decisões e promover a sustentabilidade nas operações agrícolas.

Quadro 1. Influências das variáveis ambientais no desenvolvimento de pragas e doenças nas plantas e suas aplicações na gestão agrícola.

Variável Ambiental	Influência no Desenvolvimento de Pragas e Doenças	Uso em Gestão Agrícola
Umidade do Solo	Alta umidade pode favorecer o crescimento de fungos e bactérias.	Monitorar para ajustar a irrigação e evitar excesso de umidade.
Temperatura do Ar	Temperaturas elevadas podem aumentar a reprodução de insetos e patógenos.	Usar dados para selecionar variedades mais resistentes a temperaturas extremas.
Precipitação	Precipitações excessivas podem causar alagamentos, promovendo doenças radiculares.	Ajustar a programação de irrigação e monitorar áreas propensas a alagamentos.
Luminosidade	A falta de luz pode estressar as plantas, tornando-as mais suscetíveis a pragas.	Ajustar a cobertura de plantas e a disposição das culturas.
Ventos	Ventos fortes podem danificar plantas e facilitar a dispersão de esporos.	Implementar barreiras de vento ou selecionar locais de plantio estratégicos.
Umidade Relativa do Ar	Alta umidade pode criar condições favoráveis para o desenvolvimento de doenças foliares.	Monitorar e, se necessário, aplicar fungicidas preventivos.

Fonte: Baseado em Ogunjo *et al.*, 2021.

Apesar dos avanços tecnológicos, a coleta de dados em estações meteorológicas enfrenta desafios e limitações. Entre eles estão problemas relacionados à precisão dos sensores, à variabilidade dos dados, à necessidade de calibração constantes; condições ambientais e a manutenção dos equipamentos, podendo afetar a qualidade dos dados. Além disso, a transmissão de dados em tempo real pode ser afetada por problemas técnicos ou falhas na conectividade, podendo resultar em lacunas nas séries temporais (Costa *et al.*, 2021).

Os avanços tecnológicos têm contribuído para a melhoria da coleta de dados em estações meteorológicas, a incorporação de sensores precisos, a utilização de tecnologias de comunicação avançadas e a implementação de sistemas de análise automatizados, aprimorando a qualidade e a eficiência da coleta de dados. Além disso, o desenvolvimento de plataformas de monitoramento integradas permite a visualização e a análise dos dados em tempo real, facilitando a tomada de decisões e a gestão de informações (Haris *et al.*, 2023).

Para garantir a qualidade e a consistência dos dados coletados, as estações meteorológicas devem seguir protocolos e normas estabelecidos por organizações internacionais, como a Organização Meteorológica Mundial (OMM) e a Associação Americana de Meteorologia (AMA). Esses protocolos definem padrões para a instalação, calibração e operação dos equipamentos, bem como para o processamento e a interpretação dos dados. O cumprimento dessas normas assegura a confiabilidade e a comparabilidade das informações meteorológicas.

2.8. ALGORITMOS DE RECONHECIMENTO DE PADRÕES DE IMAGENS

Para os algoritmos, os padrões em imagens referem-se às características visuais recorrentes, nos quais os algoritmos de processamento de imagem identificam e analisam para executar tarefas como reconhecimento de objetos, segmentação ou detecção de anomalias. Esses padrões podem incluir texturas, bordas, contornos, formas geométricas, cores e características locais, como pontos de interesse. Por exemplo, a detecção de bordas permite que os algoritmos identifiquem limites de objetos, enquanto a análise de cores ajuda a distinguir diferentes regiões em uma imagem.

Em aprendizado profundo, os algoritmos aprendem a extrair automaticamente esses padrões, o que permite a identificação de objetos e cenários com alta precisão. Tais padrões são essenciais para a aplicação de algoritmos de visão computacional em áreas como vigilância, agricultura de precisão e robótica (Gonzalez; Woods, 2018).

2.8.1. Algoritmo YOLO

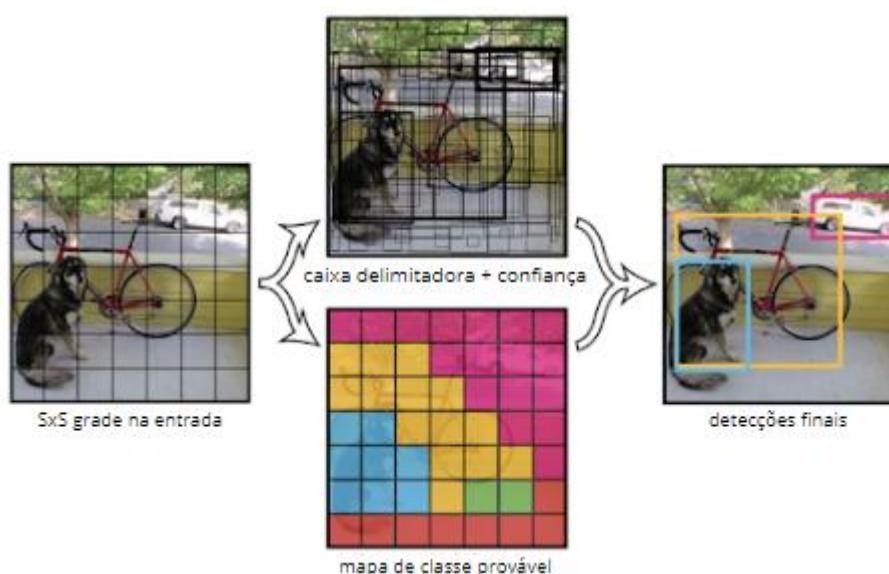
No contexto do reconhecimento de padrões de imagens para aplicações avançadas, como a integração de dados climáticos e imagens agrícolas, deve-se compreender como os defeitos de grãos, apresentados em imagens são capturados por meio de algoritmos de reconhecimento de padrões. Kate *et al.* (2023) apontaram os métodos e algoritmos *You Only Look Once* (YOLO), Redes Neurais Convolucionais (CNNs) e *Vision Transformers* (ViTs), cada um possui características distintas, adequados para diferentes aspectos do reconhecimento de padrões e permitindo aos modelos de integração de dados trabalhar com qualquer fonte.

Nazir e Wani (2023) descrevem *You Only Look Once* (YOLO) como um algoritmo de detecção de objetos em tempo real. YOLO tem uma abordagem rápida e eficiente para identificar e classificar objetos em uma única passagem pela rede neural. Em vez

de usar uma abordagem de duas etapas, faz detecção e classificação simultaneamente, resultando em processamento mais rápido.

O funcionamento do YOLO baseia-se na divisão da imagem em grades de células (Figura 12). Cada célula prevê as localizações de possíveis objetos e suas respectivas classes. O modelo gera uma série de caixas delimitadoras e associa cada caixa a uma classe, com uma probabilidade de confiança. Então refina as previsões por meio de técnicas de supressão de não-máximos, eliminando redundâncias e mantendo as mais relevantes (Nazir; Wani, 2023).

Figura 12. Detecção de objetos com YOLO.



Fonte: Adaptado de Lapix, 2018.

Os dados são capturados utilizando uma arquitetura de detecção de objetos em tempo real, processando a imagem em uma única passagem e permitindo ao YOLO alcançar altas taxas de detecção, em tempo real. O formato dos dados de entrada, geralmente, é uma imagem em RGB, redimensionada para um tamanho fixo antes do processamento (Nazir; Wani, 2023).

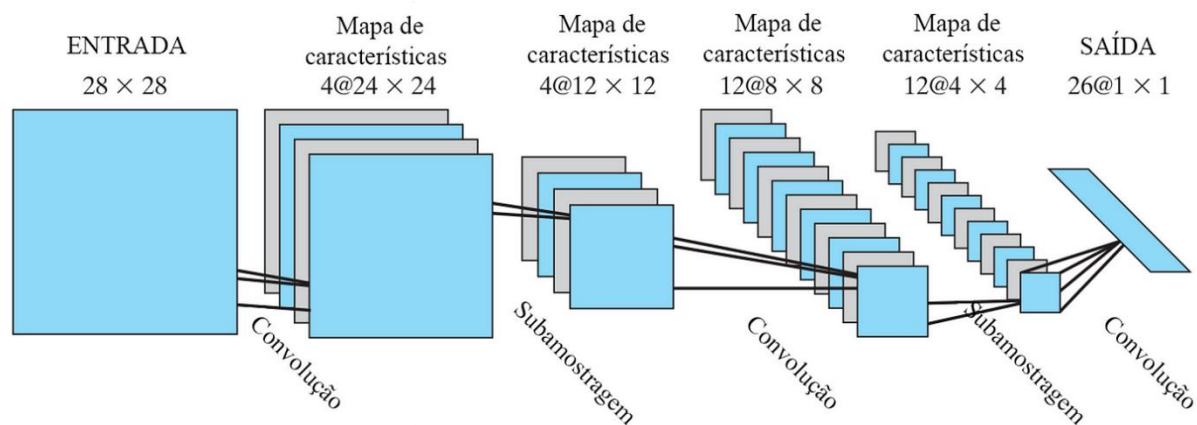
A principal vantagem do YOLO em relação a outros métodos é sua velocidade e eficiência, tornando-o ideal para aplicações que requerem detecção em tempo real, como monitoramento de segurança e análise de vídeo, com capacidade de realizar detecções rápidas e precisas em ambientes dinâmicos. Além disso, sua arquitetura permite o processamento de grandes volumes de dados de imagem sem comprometer a velocidade (Redmonm; Farhadi, 2018).

2.8.2 CNN

Derry, Krzywinski e Altman (2023) destacaram as Redes Neurais Convolucionais (CNNs) como uma das técnicas mais utilizadas para o reconhecimento de padrões de imagens. Inspiradas pela estrutura do córtex visual dos animais, são projetadas para identificar padrões visuais por meio de camadas de convolução, *pooling* e camadas totalmente conectadas. Cada camada convolucional aplica um filtro para detectar características específicas, como bordas e texturas, e as camadas subsequentes combinam essas características para formar representações mais complexas.

As CNNs capturam dados de imagens em formato de matrizes tridimensionais altura x largura x canais de cor (Figura 13). Cada filtro extrai características específicas da imagem, e a operação de *pooling* reduz a dimensionalidade dos dados, mantendo as características relevantes. Esse processo permite que a rede aprenda representações hierárquicas das imagens, desde características simples até padrões complexos (Derry, Krzywinski; Altman, 2023).

Figura 13. Reconhecimento de características em imagens com uso de CNN.



Fonte: Poloni, 2022.

Os dados de entrada para uma CNN pode ser, dentre outros, uma imagem RGB ou em escala de cinza, que é processada em múltiplas camadas convolucionais. As CNNs são eficazes para tarefas de classificação de imagens, detecção de objetos e segmentação semântica. A escolha das CNNs para reconhecimento de padrões é baseada em sua capacidade de aprender e generalizar a partir de grandes volumes de dados de imagem, oferecendo alta precisão e flexibilidade em diversas aplicações (Derry, Krzywinski; Altman, 2023).

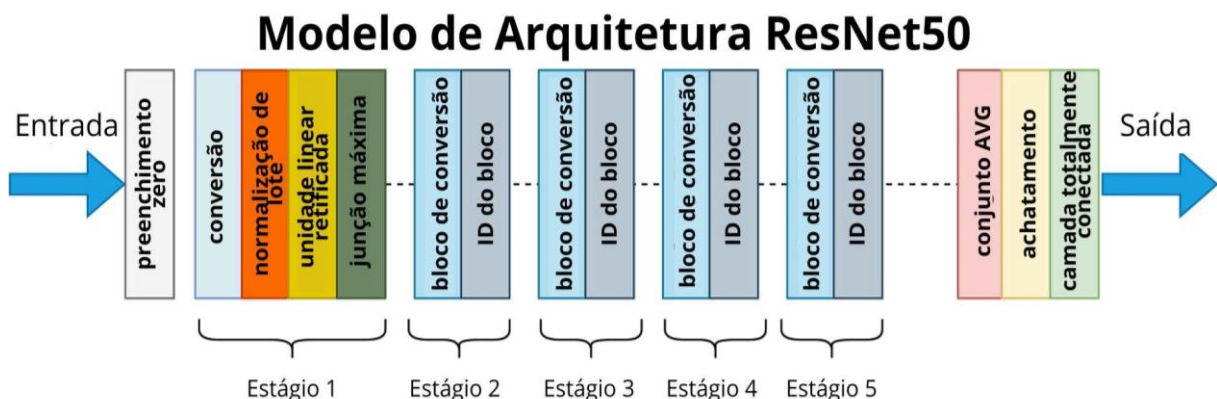
Uma vantagem das CNNs é sua capacidade de extrair e aprender características relevantes diretamente das imagens, sem a necessidade de engenharia de características manual. O processo de selecionar, transformar e criar novas

características a partir de dados brutos para melhorar o desempenho de modelos de aprendizado de máquina, envolve a aplicação de conhecimento especializado sobre o problema para identificar quais informações dos dados podem ser úteis para o modelo (Baldominos; García; Roggen, 2019).

Dentre as CNNs destaca-se a ResNet50 (*Residual Networks*), proposta por He *et al.* (2016), consolidada como uma das redes neurais convolucionais mais eficazes para tarefas de classificação de imagens, especialmente devido à sua capacidade de diminuir o problema do gradiente em redes profundas, através da introdução de blocos residuais (Li, 2020). Seu uso tem sido amplamente explorado em diversas áreas da visão computacional, na última década. A ResNet50 tem sido aprimorada e aplicada em diferentes contextos, incluindo agronegócio, saúde, e outras áreas que demandam análises precisas de imagens (Du *et al.*, 2023).

A ResNet50 destaca-se pela capacidade de treinar redes extremamente profundas sem ocorrer degradação no desempenho (Figura 14), algo que limitava o uso de arquiteturas mais profundas. É importante em situações que envolvem imagens de alta resolução e complexidade visual, como aplicações na agricultura para a identificação de doenças em plantas e defeitos em grãos. A arquitetura residual permite que a rede preserve informações relevantes, mesmo em camadas profundas, resultando em maior precisão na classificação (Zhang *et al.*, 2021).

Figura 14. Arquitetura ResNet50.



Fonte: Adaptado de Murkerjee, 2022.

No contexto do agronegócio, Kumar *et al.* (2020) exploraram o uso da ResNet50 para a classificação de doenças em folhas de milho, demonstrando uma acurácia de 98%, evidenciando a eficiência do modelo em identificar padrões visuais relacionados a condições de saúde das plantas. O estudo comparou a ResNet50 com outros modelos, como VGG16 e InceptionV3, e concluiu que, embora essas redes

apresentem bons resultados, a ResNet50 oferece um melhor equilíbrio entre profundidade e capacidade de generalização.

Chen *et al.* (2022), aplicaram a ResNet50 para a detecção de defeitos em produtos alimentícios, demonstrando que, mesmo em cenários com imagens ruidosas ou com variações de iluminação, a rede conseguiu manter altos níveis de precisão. Esse resultado foi alcançado, em parte, devido à capacidade da ResNet50 de extrair características mais robustas em múltiplas escalas, relevante para a identificação de pequenas variações que possam indicar defeitos ou anomalias.

Wu *et al.* (2023) investigaram a aplicação da ResNet50 no contexto da medicina, na detecção precoce de câncer de pele. O estudo destacou que, ao utilizar imagens de alta resolução de lesões cutâneas, a ResNet50 superou modelos mais simples, com uma acurácia de 92%, contribuindo para diagnósticos mais precisos e rápidos. O uso de blocos residuais permitiu que a rede se aprofundasse em detalhes visuais complexos, o que foi determinante para a classificação.

A ResNet50 tem sido comparada a arquiteturas como o Vision Transformer (ViT). Embora a ViT apresente uma abordagem inovadora, utilizando mecanismos de atenção para tratar com a sequência de pixels das imagens, Zhao *et al.* (2023) apontam que a ResNet50 se sobressai em tarefas de visão computacional em cenários com menor quantidade de dados rotulados, devido à sua capacidade de aproveitar melhor os dados disponíveis. Assim, a ResNet50 continua sendo uma opção robusta e confiável, especialmente em contextos com dados limitados.

A ResNet50 tem apresentado resultados promissores em aplicações industriais. Segundo Liu *et al.* (2020), a rede foi implementada em sistemas de controle de qualidade em fábricas de manufatura para a detecção de defeitos em peças produzidas, com um desempenho altamente confiável, mesmo em condições de variação de luz e perspectiva. Essa robustez em diferentes ambientes de captura de imagem reforça a sua versatilidade e aplicação em contextos que exigem análise visual precisa e eficiente.

A ResNet50 continua sendo amplamente utilizada e relevante em várias áreas de pesquisa e aplicação, desde a agricultura até a medicina e a indústria, mantendo-se como um dos principais modelos de aprendizado profundo para a análise de imagens complexas, mesmo frente a arquiteturas mais recentes.

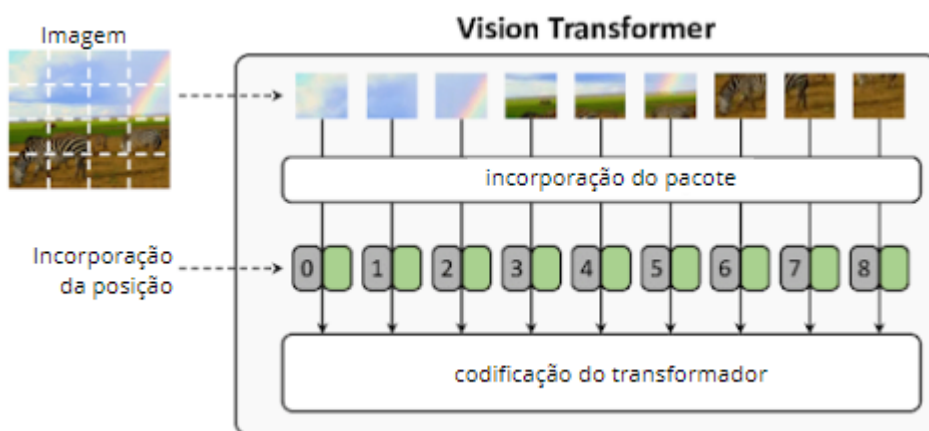
2.8.3 Vision Transformers - ViTs

Azad *et al.* (2023) explicam que os *Vision Transformers* (ViTs) representam uma abordagem recente e inovadora para o reconhecimento de padrões de imagens, utilizando uma arquitetura baseada em *transformers*, originalmente desenvolvida para tarefas de processamento de linguagem natural (NLP). Os ViTs dividem a imagem em pequenos blocos, chamados de *patches*, e tratam cada *patch* como uma unidade de entrada para a rede *transformer*, permitindo que o modelo capture informações de longo alcance e complexas dependências espaciais.

O funcionamento dos *Vision Transformers* baseia-se na aplicação de camadas de atenção, que avaliam a importância relativa de diferentes *patches* na imagem. Cada *patch* é representado por um vetor de características, e a atenção é usada para ajustar as representações desses vetores com base em suas interações. O modelo aprende a identificar padrões e relações entre os *patches*, permitindo uma análise detalhada e precisa das imagens (Azad *et al.*, 2023).

Os dados são capturados através da divisão da imagem em *patches* de tamanho fixo, que são processados e integrados para formar uma representação global da imagem (Figura 15). O formato de entrada para os ViTs é semelhante ao dos *transformers* em NLP, onde os *patches* são transformados em *embeddings*, representações vetoriais, os quais representam objetos do mundo real, como palavras, imagens ou vídeos, em uma forma que os computadores podem processar. Após, são processados pela rede. (Azad *et al.*, 2023).

Figura 15. Processamento de imagem pelo ViT.



Fonte: Adaptado de Miranda, 2024.

A escolha dos *Vision Transformers* é motivada por sua eficácia em capturar relações complexas e sutis em imagens, tarefa desafiadora para arquiteturas

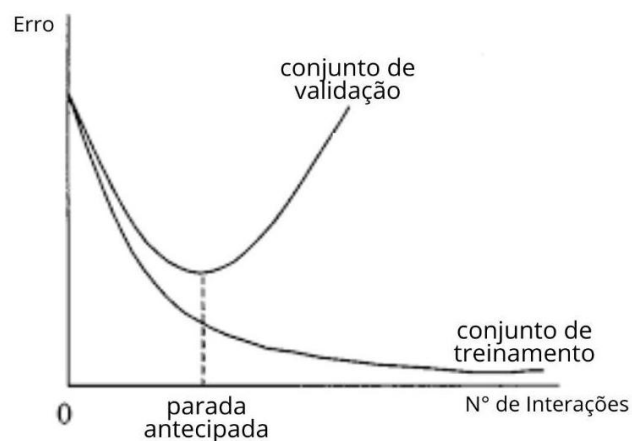
baseadas em convolução. Embora sejam mais recentes e exijam grandes volumes de dados e recursos computacionais para treinamento, os ViTs têm mostrado desempenho superior em tarefas de reconhecimento de padrões de imagens em comparação com métodos tradicionais, sendo úteis para tarefas que envolvem imagens grandes e complexas, como reconhecimento de objetos em cenas variadas e análise de imagens de drones sobre plantações (Anjos; Vieira; Almeida, 2022).

2.8.4 Funções E Métodos Para Reconhecimento De Padrões

A Função *ColorJitter* permite alterar aleatoriamente as cores das imagens, ajudando no reconhecimento de padrões, de forma mais robusta (Redgate, 2021). A Função *Resize* ajusta as imagens para um tamanho uniforme, garantindo consistência nos dados de entrada, facilitando o processamento pelo modelo e evitando problemas relacionados a dimensões variadas (Rosebrook, 2017). Já a Função *CenterCrop* realiza um corte central da imagem, focando nos detalhes relevantes e eliminando áreas menos importantes (Dey, 2019). Essa etapa de pré-processamento melhora a eficácia do modelo durante o treinamento e aumenta sua capacidade de generalização ao lidar melhor com novos dados.

Para garantir o desempenho é necessário empregar a técnica de *early stopping* (parada antecipada), utilizada em treinamento de modelos de aprendizado de máquina, especialmente em redes neurais, para evitar o *overfitting* (sobreajuste). A ideia principal é monitorar o desempenho do modelo em um conjunto de validação durante o treinamento e interromper o processo quando o desempenho começa a piorar. Isso garante que o modelo esteja em seu melhor estado, antes que comece a se ajustar demais aos dados de treinamento (Bakhshandeh; Mousavi, 2023).

Figura 16. Ponto de parada antecipada (early stopping).



Fonte: Adaptado de Gençay, 2023.

O *early stopping* (Figura 16) pode ser implementado de várias maneiras, mas geralmente envolve as seguintes etapas: divisão dos dados, onde os dados são separados em conjuntos de treinamento e validação; treinamento do modelo, que é realizado em múltiplas iterações (épocas); monitoramento do desempenho, que consiste na avaliação do modelo no conjunto de validação após cada época; e um critério de parada, onde se o desempenho não melhorar após um número predefinido de épocas (conhecido como *patience*), o treinamento é interrompido. Essa técnica é amplamente utilizada para melhorar a generalização de modelos de aprendizado de máquina (Prendergast, 2020).

2.9 TRABALHOS CORRELATOS

A análise de trabalhos correlatos revela diversas diferenças entre eles, refletindo abordagens e metodologias distintas. No entanto, todos compartilham a semelhança com a presente pesquisa ao focar na integração de imagens e dados climáticos. Essa interconexão otimiza a tomada de decisões no setor agrícola, permitindo um entendimento das relações entre variáveis ambientais e desempenho das culturas. Portanto, a convergência dessas pesquisas destaca a integração de múltiplas fontes de dados para aprimorar as práticas agrícolas.

Fishman *et al.* (2010) investigaram os danos causados pelo ozônio nas safras de soja na sub-região ao norte do Centro-Oeste dos Estados Unidos, utilizando dados de satélite e medições terrestres. O estudo revelou que altas concentrações de ozônio atmosférico estão associadas a danos nas culturas de soja, afetando tanto a qualidade quanto o rendimento. Ao integrar observações de satélite com dados de monitoramento local, os pesquisadores mapearam a extensão geográfica desses danos de forma precisa, destacando áreas vulneráveis e padrões sazonais de exposição ao ozônio.

A importância deste estudo reside no entendimento dos impactos ambientais na agricultura, fornecendo informações para a gestão de culturas em regiões afetadas por poluentes atmosféricos. Além disso, ao demonstrar a eficácia das tecnologias de sensoriamento remoto na avaliação de danos ambientais, abre espaço para desenvolvimento de futuros monitoramentos e redução de impactos semelhantes em outras regiões agrícolas.

Hampf *et al.* (2020) exploraram as implicações das mudanças climáticas e avanços tecnológicos nos sistemas de cultivo duplo, com foco em milho e soja no Brasil, investigando como variações climáticas e inovações tecnológicas influenciam a viabilidade econômica e ambiental desses sistemas agrícolas. Com projeto de modelos integrados de projeções climáticas com cenários de adoção de tecnologia.

Essa abordagem permitiu identificar desafios, como a necessidade de adaptação às mudanças na disponibilidade de água e ao aumento das temperaturas, ao mesmo tempo em que destacou oportunidades para aumentar a eficiência produtiva e reduzir os impactos ambientais negativos por meio de práticas agrícolas sustentáveis. O estudo permite orientar políticas públicas e estratégias de manejo agrícola para promover a resiliência frente às mudanças climáticas e avanços tecnológicos na agricultura, contribuindo para o desenvolvimento agrícola.

Gao *et al.* (2018) utilizaram imagens de satélite de alta resolução para avaliar a variabilidade na produção de milho e soja no estado norte-americano de Iowa. Foram empregadas análises de múltiplos satélites para mapear com precisão as variações temporais e espaciais na produtividade agrícola. Ao investigar fatores como a qualidade do solo, práticas agrícolas e condições climáticas locais, foram identificados padrões de variabilidade, influenciando a produtividade das culturas.

Este estudo avançou em técnicas de monitoramento agrícola e forneceu informações para melhorar a gestão de culturas e a previsão de safra em escalas regionais, contribuindo para agricultura sustentável. Ao integrar tecnologias de sensoriamento remoto com modelos agronômicos, abriu caminho para o desenvolvimento de práticas agrícolas precisas e adaptáveis às condições locais variáveis, melhorando a segurança alimentar e reduzindo os impactos ambientais.

Jin *et al.* (2017) examinaram diferentes abordagens para melhorar a precisão da estimativa de rendimento agrícola, utilizando dados de satélite de alta resolução. Foram comparados métodos escaláveis para estimar o rendimento, explorando técnicas avançadas de sensoriamento remoto e modelagem.

Neste estudo, ao testar múltiplas abordagens, foram identificadas estratégias promissoras para aumentar a precisão das previsões de rendimento, oferecendo novas perspectivas para a agricultura de precisão. O aprimoramento de técnicas de monitoramento agrícola baseadas em satélite beneficia as aplicações práticas na gestão agrícola e na segurança alimentar. Ao propor métodos robustos e escaláveis

para estimar o rendimento agrícola, contribuíram para o desenvolvimento de políticas agrícolas eficazes e sustentáveis, adaptadas às demandas específicas de diferentes regiões e sistemas agrícolas.

Chen *et al.* (2022) focaram na melhoria da desagregação espacial do rendimento das culturas, integrando aprendizado de máquina e dados de múltiplas fontes. Foram demonstrados como modelos avançados de aprendizado de máquina podem ser aplicados para prever o rendimento do milho na China, com maior precisão, utilizando informações de sensoriamento remoto e variáveis ambientais.

Ao desenvolver metodologias para captura de variações espaciais complexas na produtividade das culturas, ofereceu novas ferramentas para otimizar o manejo de culturas e adaptar a agricultura às condições locais variáveis. A exploração do potencial do aprendizado de máquina na agricultura de precisão, abriu caminho para melhorias na segurança alimentar e na sustentabilidade agrícola. Ao integrar dados de múltiplas fontes com técnicas de aprendizado de máquina, avançaram na precisão das previsões de rendimento e promoveram gestão agrícola eficiente e adaptável às mudanças climáticas.

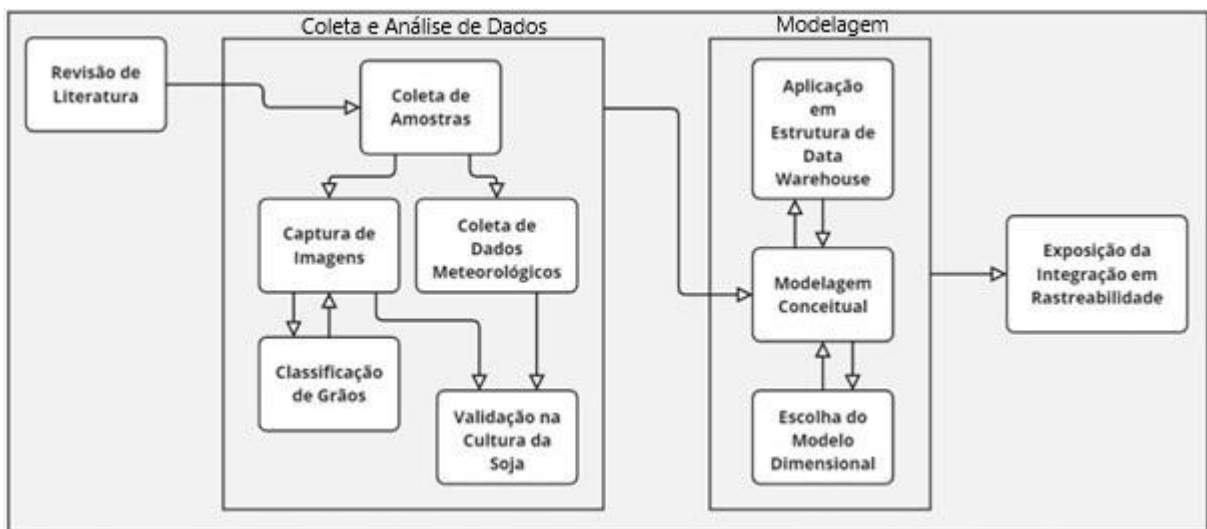
Apesar do avanço na integração de dados climáticos e de imagens na pesquisa agrícola, muitas pesquisas têm se concentrado em resolver questões específicas, como problemas de irrigação para a produção de milho ou a análise isolada de imagens de satélite. Essa abordagem segmentada limita a aplicação dos resultados, pois não aborda o contexto mais amplo e as interações complexas que podem ocorrer entre diferentes fatores climáticos e agronômicos.

Além disso, nenhum dos trabalhos analisados propôs um modelo de integração de dados. As integrações existentes serviam apenas a propósitos específicos, sem explorar como uma abordagem generalizada poderia otimizar a gestão agrícola de forma eficaz. Essa lacuna evidencia a demanda de desenvolver um modelo de integração que una dados climáticos e imagens, proporcionando uma ferramenta para a gestão agrícola, considerando a multiplicidade de variáveis interconectadas e suas influências no desempenho das culturas.

3 MATERIAIS E MÉTODOS

Nesta pesquisa foi especificado um modelo de integração de dados climáticos e de imagens na gestão agrícola, com o objetivo de aprimorar a tomada de decisão. A metodologia adotada incluiu uma série de etapas fundamentais (Figura 17), da revisão de literatura até a proposta de integração no processo de rastreabilidade. O foco principal foi a construção de um ambiente centralizado que permite a integração e a utilização de dados para otimizar a gestão agrícola.

Figura 17. Fluxo da metodologia.



Fonte: A autora.

3.1 REVISÃO DE LITERATURA

Foi realizada uma revisão de literatura para identificar lacunas e fundamentar a pesquisa. A revisão das Seções 2.1 a 2.8 se basearam nos seguintes critérios:

- Os artigos não deveriam ser anteriores a 2013, visto que as mudanças climáticas se intensificaram nos últimos dez anos (IPCC, 2021);
- Os artigos deveriam incluir estudos de diferentes regiões geográficas para entender como a integração de dados climáticos e de imagens pode variar, conforme as condições locais e os tipos de cultivo.

Foram utilizadas as bases de dados Scopus, CAPES, Google Acadêmico, e sites de conferências com submissão de trabalhos, como fontes primárias para a busca de artigos. Foram aplicadas palavras-chave relacionadas ao tema da pesquisa, como "rastreabilidade agrícola", "classificação de grãos", "integração de dados climáticos", "data warehouse em agricultura", entre outras. Após a busca, os artigos foram selecionados com base nos critérios pré-definidos. Os artigos que não se

encaixavam nesses critérios foram excluídos para garantir que a revisão refletisse o estado mais atual e relevante da pesquisa sobre a integração de dados climáticos e imagens para gestão agrícola.

Em relação à revisão realizada acerca dos trabalhos correlatos (Seção 2.9), mantiveram os critérios estabelecidos para as seções anteriores acrescidos dos seguintes critérios de seleção:

- Os estudos deveriam abordar a integração de dados provenientes de diversas fontes, como sensores de campo, estações meteorológicas e imagens, enfatizando os desafios e as soluções para a unificação desses dados;
- Os artigos deveriam abordar a integração de dados a partir de diferentes áreas, como agronomia, meteorologia, ciência da computação e ciências ambientais, para enriquecer a discussão sobre a gestão agrícola;
- Os artigos deveriam apresentar estudos de caso ou exemplos de aplicação prática que demonstrasse a eficácia da integração de dados climáticos e de imagens na gestão agrícola, destacando resultados mensuráveis e impactos diretos na produção; e
- Os artigos, obrigatoriamente, deveriam conter integração de dados climáticos e de imagens.

Foram empregadas diversas palavras-chave e suas combinações para identificar trabalhos correlatos, envolvendo a integração de dados climáticos e imagens para a gestão agrícola. Essas palavras-chave foram selecionadas com o objetivo de abranger os principais conceitos e tecnologias inerentes ao tema, assegurando que a pesquisa representasse as soluções mais relevantes para o setor. Foram utilizadas as seguintes palavras-chave e suas traduções, com as diversas combinações com o uso do conector lógico AND: Integração de dados climáticos; Imagens agrícolas; Imagens de satélite; Clima; Imagem; Mudanças climáticas; Integração de dados climáticos; Variação Climática; Modelos de previsão agrícola; Projeções climáticas; e Gestão de culturas com dados multivariados.

O uso do conector lógico AND justifica-se por restringir os resultados da pesquisa, buscando trabalhos que contenham ambas as palavras-chave. No campo da lógica computacional e da pesquisa em bases de dados, o conector AND é utilizado para filtrar informações, retornando registros que satisfaçam as condições especificadas. Quando utilizado em consultas, o operador AND exige que ambos os

termos de busca estejam presentes em um documento, o que torna a pesquisa mais específica e focada (Souza e Carvalho, 2020).

A busca inicial resultou em 23 trabalhos promissores. Artigos que tratavam de temas muito específicos ou que não se enquadravam nos objetivos gerais da integração de dados foram descartados. Assim, restaram seis estudos que se alinhavam diretamente à pesquisa, fornecendo uma base sólida para o desenvolvimento do modelo de integração de dados climáticos e imagens.

3.2 COLETA DE AMOSTRAS

A coleta de amostras foi realizada tanto para as imagens quanto para os dados climáticos de forma rigorosa e padronizada. Como o modelo havia sido planejado para ser validado utilizando imagens de grãos de soja, todas as imagens coletadas estão relacionadas a este tipo de grão, capturando diferentes estágios de qualidade e condições físicas. Paralelamente, os dados climáticos foram obtidos a partir de API's de estações meteorológicas, abrangendo variáveis como temperatura, precipitação e umidade, essenciais para a análise do impacto das condições ambientais no desempenho agrícola.

Durante a coleta de dados, foi realizada uma busca por bases de imagens de soja em seus diferentes estados de saúde. O objetivo era compilar uma coleção abrangente que refletisse a diversidade de condições que os grãos de soja podem apresentar, desde os saudáveis até aqueles afetados por doenças e pragas. Diversas bases de dados foram consideradas, incluindo o *Soybean Disease Image Collection*, *PlantVillage*, *The Crop Image Gallery* e *Grain Image Database*. Cada uma dessas fontes oferece uma variedade de imagens para análise e treinamento de modelos de aprendizado de máquina.

Entretanto, durante a pesquisa, foi detectada a falta de atualização das imagens disponíveis, especialmente aquelas que representam defeitos nos grãos. As mudanças climáticas e as novas doenças emergentes têm impactado as culturas de maneiras que não estão suficientemente documentadas nas bases existentes. Essa desatualização impede a criação de modelos de reconhecimento eficazes que possam tratar com as variações atuais nas condições das plantações. Assim, a necessidade de uma nova base de dados se tornou evidente, levando à decisão de criar uma coleção de imagens para captura dos efeitos das mudanças climáticas nos grãos de

soja. Essa nova base foi fundamental para fornecer dados atualizados e representativos, contribuindo para avanços na pesquisa e na prática agrícola.

3.2.1 Classificação De Grãos – Captura De Imagens

Como a pesquisa teve foco na compreensão das variações climáticas, utilizou-se captura de imagens em laboratório de grãos de soja, da Safra 2022/2023, de uma fazenda localizada na Cidade de Rio Azul, Município Paranaense, a fim de apresentar imagens com maior (ou menor) índice de defeitos do que as apresentadas em bases de dados não atualizadas nos últimos dez anos.

Os grãos de soja pertencentes a amostra passaram por um processo de classificação de grãos (Figura 18), para garantir a existência de defeitos, com a classificação realizada conforme normas vigentes, avaliando qualidade, tamanho e entre outros parâmetros, e utilizando uma Lupa com Luminária HL-500 LED Solver, com aumento de lente de 8x e 6W de consumo.

Figura 18. Figura 18. Classificação de Grãos em Laboratório.



Fonte: A autora.

Para captura das imagens, foi utilizado o estúdio fotográfico da Universidade Estadual de Ponta Grossa (Figura 19), bem como os equipamentos que seguem: Câmera Canon EOS Rebel T5i; lente Canon EF 100mm f/2.8L Macro IS USM; 3 (três) Refletores Super Lamp Holder SLH 3; mesa com material translúcido branco, 8 (oito) Lâmpadas de 40W Taschibra.

Figura 19. Estúdio Fotográfico.



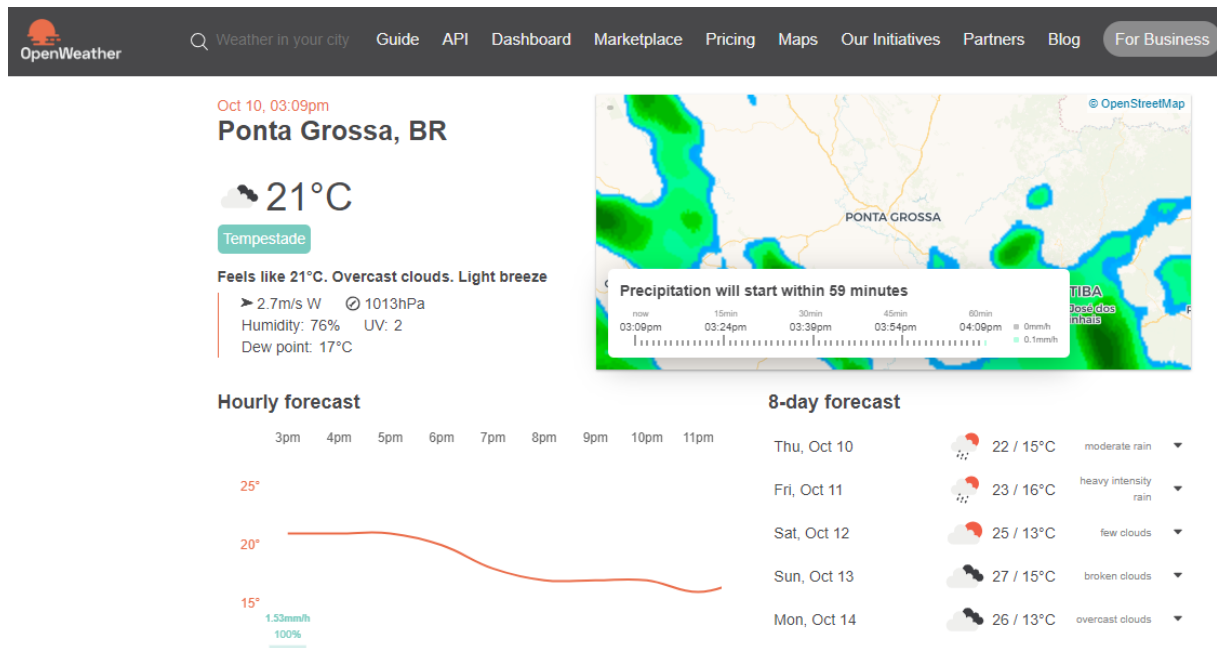
Fonte: A autora.

As imagens coletadas em laboratório passaram por um pré-processamento manual, para garantia da qualidade e da precisão dos dados inseridos na base. Esse procedimento incluiu a remoção de imagens com baixa nitidez, as quais poderiam comprometer a detecção de defeitos nos grãos. Além disso, imagens que continham fatores externos prejudiciais à captura, como poeira sobre a lente ou reflexos indesejados, foram descartadas.

3.2.2 Coleta De Dados Meteorológicos

Simultaneamente, houve a coleta de dados meteorológicos de estações via API *OpenWeatherMap* (Figura 20), e foi realizada conexão com a API, fornecendo dados em tempo real e históricos sobre o clima. A comunicação com a API ocorreu por meio de requisições HTTP, possibilitando a definição de parâmetros, como localização desejada e o tipo de informação necessária. As respostas da API foram retornadas em um Formato JSON (*JavaScript Object Notation*).

Figura 20. Interface da plataforma OpenWeather.



Fonte: OpenWeather, 2024.

Após envio da requisição, a resposta retorna como um objeto JSON, contendo detalhes sobre o clima, tais como temperatura, umidade, pressão atmosférica e descrições do clima. Essas informações fornecem dados quantitativos e qualitativos podendo ser cruzados com imagens.

Foram realizadas conversões dos dados coletados para garantir a uniformidade e a consistência das informações, permitindo uma integração eficaz no modelo de dados. Cada tipo de dado passou por um processo de conversão.

A temperatura foi inicialmente fornecida em Kelvin, formato utilizado em APIs meteorológicas internacionais. Para garantir dados mais intuitivos e úteis para análises agrícolas, a temperatura foi convertida para Celsius (Fórmula 1).

$$\text{Temperatura em Celsius} = \text{Temperatura em Kelvin} - 273,15 \quad (1)$$

Essa conversão foi implementada por meio de uma função que percorreu todos os registros de temperatura retornados pela API, aplicando a fórmula e armazenando os resultados em uma nova coluna no banco de dados.

A umidade foi recebida em formato decimal, representando a porcentagem da umidade relativa do ar. Esta medida é normalmente expressa em uma escala de 0 a 1, onde um valor de 0,75 corresponde a 75% de umidade. Para manter a consistência,

foi mantido o formato decimal sem realizar conversão adicional, armazenando a unidade diretamente no banco de dados.

Em relação à pressão atmosférica, os dados foram obtidos em hectopascals (hPa), unidade padrão na meteorologia. Para integrá-los de forma consistente, foram convertidos para atmosferas (atm), unidade comumente utilizada nos sistemas brasileiros. Para converter hPa para atm, o valor da pressão em hPa é dividido por 1013,25 (Fórmula 2).

$$\text{Pressão em atm} = \frac{\text{Pressão em hPa}}{1013,25} \quad (2)$$

A velocidade do vento foi fornecida em metros por segundo, unidade de medida amplamente utilizada. Como essa unidade é adequada para as análises, não houve necessidade de conversão, e os dados foram armazenados diretamente.

O *timestamp* foi obtido em Formato Unix, representando a quantidade de segundos desde o início da coleta. Para converter esse valor para o Formato ISO 8601, norma internacional para representação de data e hora. Foi utilizada a biblioteca *datetime* em Python (Figura 21). O timestamp Unix foi convertido em um objeto de data e hora e formatado no Padrão ISO 8601, assegurando que as datas e horas fossem apresentadas de maneira clara e consistente.

Figura 21. Fórmula para conversão datetime para ISO 8601.

```
from datetime import datetime

unix_timestamp = 1617205200 # Exemplo de timestamp
iso_timestamp = datetime.utcfromtimestamp(unix_timestamp).isoformat()
```

Fonte: Python Software Foundation, 2024.

Essas conversões garantiram que todas as informações climáticas coletadas da *API OpenWeather* fossem integradas de forma homogênea no banco de dados, possibilitando análises robustas e fundamentadas.

3.3 MODELAGEM CONCEITUAL

Para consolidar esses dados, foi realizada a escolha do modelo dimensional constelação para a criação do modelo conceitual, com foco na criação de uma arquitetura que refletisse as inter-relações entre os dados climáticos e de imagens.

Este modelo é uma das etapas do modelo de integração de dados climáticos e de imagens desenvolvido nesta pesquisa.

A modelagem foi realizada utilizando o software gratuito *draw.io*. Esta ferramenta foi escolhida devido à sua interface intuitiva e às funcionalidades robustas que permitem a criação de diagramas e modelos complexos de forma colaborativa. O *draw.io* possibilita a visualização clara e dinâmica dos esquemas de dados, facilitando a integração de múltiplas dimensões e tabelas de fato.

A modelagem conceitual incluiu a criação de tabelas de fatos relacionadas a imagens e dados climáticos, bem como tabelas de dimensões correspondentes. A capacidade do *software* de oferecer diferentes formas de visualização ajudou a evidenciar as interações entre as diversas dimensões e a estrutura geral do modelo de dados, permitindo um *design* eficiente para o armazenamento e análise das informações. Além disso, a escolha do *draw.io* foi motivada pela sua acessibilidade e pelo suporte a múltiplos formatos de exportação, o que facilita a documentação e o compartilhamento dos modelos desenvolvidos.

As variáveis contidas nas tabelas foram selecionadas com base nas informações extraídas da revisão de literatura, garantindo que o modelo de integração de dados fosse robusto e alinhado com as necessidades práticas.

Para criar o modelo de integração de dados climáticos e de imagens foi feita modelagem baseada na estrutura de um *data warehouse*, garantindo a capacidade de armazenar e processar grandes volumes de dados com eficiência.

3.4 VALIDAÇÕES

O modelo foi validado com imagens da cultura da soja, considerando as especificidades desse cultivo. Para a validação do modelo de captação de dados meteorológicos, foi utilizada exclusivamente a *API OpenWeather*. Essa escolha foi justificada pela abrangência e confiabilidade da *OpenWeather*, que oferece dados meteorológicos precisos e em tempo real de uma vasta rede de estações. A *API* fornece informações detalhadas, como temperatura, umidade, precipitação e velocidade do vento, para o monitoramento climático. Além disso, a *OpenWeather* é amplamente reconhecida pela qualidade e robustez dos seus dados, sendo utilizada em diversas aplicações meteorológicas e ambientais.

A decisão de utilizar uma única *API* também foi baseada na necessidade de simplificar o processo de integração e validação do sistema. Como a *OpenWeather* já

atende aos requisitos de precisão, atualidade e diversidade de dados climáticos, considerada suficientes para garantir a eficácia do sistema. A API oferece suporte a uma grande quantidade de estações meteorológicas, permitindo a coleta de dados representativos de diferentes regiões geográficas e condições climáticas, eliminando a necessidade de múltiplas fontes para comparação. Essa abordagem reduz a complexidade no desenvolvimento e validação do sistema, garantindo que as informações captadas sejam confiáveis e consistentes ao longo do tempo.

Para validar a criação da base de dados utilizada no modelo de análise de grãos, foram coletadas imagens em laboratório, para garantir que os defeitos representados fossem coerentes com os estados de saúde dos grãos. A escolha por imagens laboratoriais permitiu um controle rigoroso das condições de captura, como iluminação, ângulo e resolução, assegurando que os detalhes visuais relevantes para a identificação de defeitos estivessem claramente visíveis e bem definidos.

Durante o processo de validação, foram selecionados grãos de soja expostos a uma variedade de condições ambientais, visando influenciar seu estado de saúde, como umidade excessiva, danos físicos e deterioração natural. Essa abordagem permitiu a obtenção de amostra diversificada de defeitos, como grãos quebrados, mofados, com alterações na coloração ou texturas inadequadas. Ao garantir que a base de dados contivesse representações fiéis dos problemas comuns observados na prática agrícola, foi possível assegurar ao modelo a capacidade de identificar e classificar os defeitos de grãos.

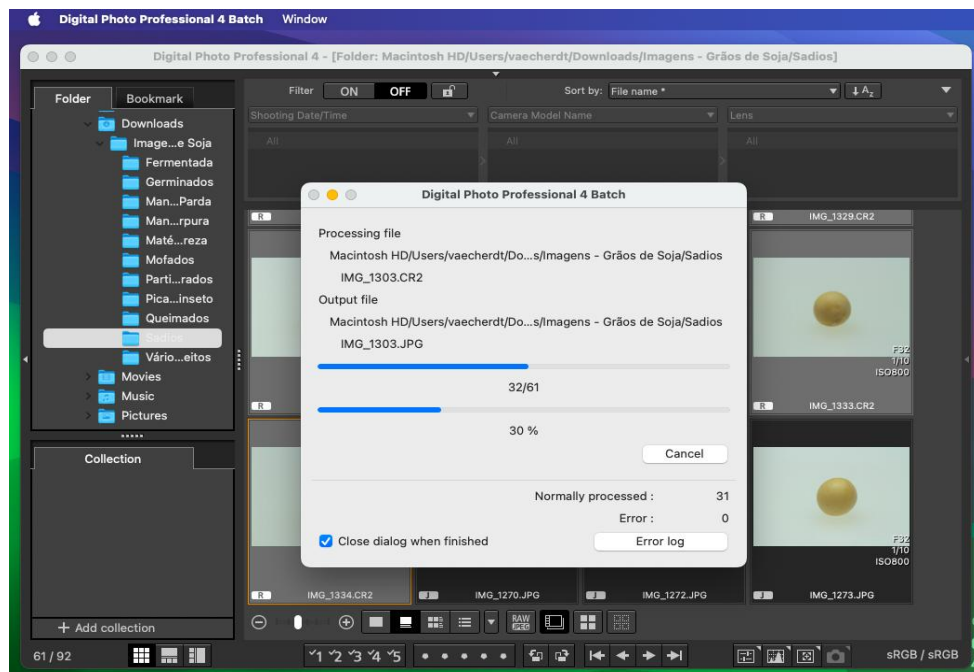
A validação também envolveu a comparação dos defeitos observados nas imagens com padrões estabelecidos em estudos de qualidade de grãos, garantindo que a base de dados estivesse alinhada com as normas do setor agrícola. Isso conferiu maior credibilidade ao modelo, assegurando que as imagens utilizadas fossem representativas de situações reais enfrentadas por agricultores e indústrias alimentícias, e que os resultados da análise fossem aplicáveis em cenários de produção e comercialização de grãos.

A validação da conversão dos formatos de imagem foi uma etapa importante no desenvolvimento do modelo, assegurando que as imagens fossem compatíveis com os algoritmos de reconhecimento de padrões em imagens, utilizando Formato JPG. As imagens capturadas originalmente no formato proprietário CR2 da Canon, oferecendo alta qualidade e detalhes ricos, foram convertidas para o Formato JPG,

possibilitando enquadramento na execução futura dos algoritmos de processamento e classificação de imagens.

A conversão foi realizada utilizando o *Software Photo Professional 4*, também da Canon, especializado no tratamento e exportação de imagens em alta qualidade. O uso dessa ferramenta garantiu as características visuais, como a precisão das cores e a definição dos contornos dos grãos, e preservadas durante o processo de conversão. O *Photo Professional 4* foi escolhido por sua capacidade de realizar conversões em lote, otimizando o tempo e garantindo consistência do formato final de todas as imagens processadas (Figura 22).

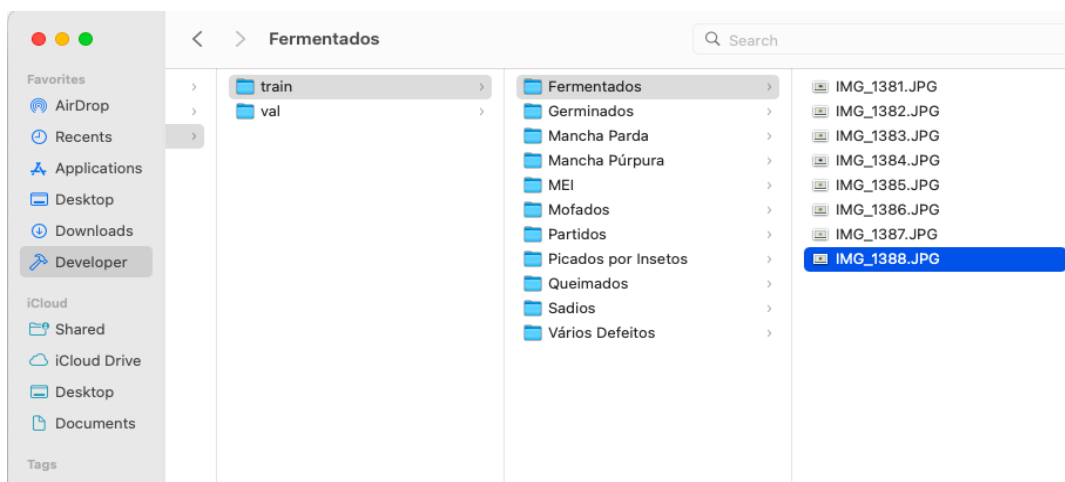
Figura 22. Conversão das imagens de CR2 para JPG.



Fonte: A autora.

As imagens convertidas foram organizadas em um conjunto de dados estruturados, dividido em duas subpastas: *train* e *val*. A Pasta *train* contém 80% das imagens, destinadas ao treinamento do modelo, enquanto a Pasta *val* tem 20% das imagens, voltadas para validação. Essa divisão garante ao modelo aprender a partir de uma maior quantidade dos dados, enquanto reserva uma parte para avaliar seu desempenho em imagens não visualizadas anteriormente. As imagens foram organizadas em diretórios (*dataset*) para cada tipo de defeito ou doença (Figura 23). Esse processo assegura ao modelo a generalização do aprendizado, aprimorando sua precisão na identificação de defeitos em novas imagens de grãos.

Figura 23. Organização dos diretórios de imagens.



Para validação da extração de metadados foi utilizada a *ResNet50*, uma rede neural convolucional profunda amplamente reconhecida por sua capacidade de realizar tarefas de classificação de imagens com alta precisão. As classes, para o reconhecimento e classificação de defeitos e doenças nos grãos, foram definidas com base nos tipos de anomalias observadas, permitindo que o modelo identificasse diferentes padrões visuais associados a cada defeito.

A configuração do treinamento do modelo foi realizada utilizando a função de perda *CrossEntropyLoss*, que mede a diferença entre as previsões do modelo e os rótulos corretos, ajudando a ajustar os pesos da rede neural durante o aprendizado. O otimizador escolhido foi o *Stochastic Gradient Descent* (SGD), com taxa de aprendizado de 0.0001 e um *momentum* de 0.9, garantindo um ajuste gradual dos pesos do modelo e evitando atualizações abruptas. O objetivo dessa configuração foi justar os pesos do modelo durante o treinamento, minimizando a perda e, consequentemente, melhorando o desempenho nas classificações.

O algoritmo de extração foi treinado por 25 (vinte e cinco) épocas, onde cada época representa uma iteração completa sobre o conjunto de dados de treinamento. A escolha por 25 épocas foi baseada na análise do comportamento da acurácia e da perda durante o processo de treinamento. Esse número foi considerado suficiente para que o modelo aprendesse os padrões dos dados sem sofrer de *overfitting*, ou seja, sem se ajustar demais ao conjunto de treinamento, a ponto de perder sua capacidade de generalização.

A cada época, o desempenho foi avaliado com base em sua capacidade de identificar corretamente os defeitos e doenças nos grãos, ajustando seus parâmetros para minimizar a função de perda. Após testes iniciais, foi observado que a partir da

25ª época a acurácia se estabilizou e a perda não apresentou melhorias. Portanto, continuar o treinamento além desse ponto poderia resultar em *overfitting*, prejudicando a capacidade do modelo de se generalizar para novos dados.

O modelo foi exposto como parte integrante do processo de rastreabilidade, permitindo compreender sua participação no processo de verificação da cadeia produtiva, de forma transparente e precisa, fortalecendo a gestão agrícola com dados consistentes e rastreáveis.

4 RESULTADOS E DISCUSSÕES

Neste capítulo, são apresentados os resultados desta pesquisa, abordando as etapas de desenvolvimento do modelo de integração. Primeiramente, é discutida a base de dados criada, de imagens de grãos de soja. Em seguida, é descrito o modelo de integração de dados climáticos e de imagens, foco principal desta pesquisa. Após, é apresentado o modelo conceitual que advém do modelo principal, detalhando as estruturas e interrelações das variáveis envolvidas. Ainda, são apresentadas as validações de extração de metadados necessárias e, por fim, é analisada a inserção do modelo no fluxo de rastreabilidade agrícola, demonstrando como pode contribuir para acompanhamento preciso e eficiente no setor.

4.1 BASE DE IMAGENS DE GRÃOS DE SOJA

As pesquisas em bases de imagens de grãos de soja revelaram uma lacuna na atualização de imagens representativas de defeitos causados por variações climáticas. Esse problema, agravado pelas mudanças climáticas, afeta a capacidade de monitorar com precisão a saúde e a qualidade das plantas, comprometendo a eficiência agrícola e a segurança alimentar.

Para abordar essa deficiência, foi realizada a captura e análise prática de imagens de grãos de soja, representando diversos estados de saúde dos grãos (Figura 24). A análise envolveu a documentação de defeitos como ardido, chocho, fermentado, germinado, mancha púrpura, mofado, partido/quebrado, picado por inseto, queimado, sadio e mancha parda. A captura das imagens atualizadas permitiu avaliação detalhada e precisa das condições dos grãos, em resposta a diferentes condições climáticas extremas, proporcionando uma base de dados robusta e representativa para futuras pesquisas e aplicações práticas.

Figura 24. Classificação de Grãos em Laboratório.



Fonte: A autora.

As imagens foram catalogadas de acordo com os critérios de classificação de defeitos de grãos estabelecidos pelo Ministério da Agricultura, Pecuária e Abastecimento (MAPA), abrangendo todos os defeitos reconhecidos, como grãos danificados, quebrados e mofados. Além desses defeitos, foram incluídas as condições de mancha parda e mancha púrpura, que são consideradas doenças e, não faziam parte dos critérios formais de classificação.

A inclusão dessas doenças na catalogação é relevante, pois mesmo não utilizadas diretamente nos critérios de classificação, podem afetar a qualidade dos grãos, impactando sua aceitação comercial e seus aspectos de armazenamento.

A amostra analisada consistiu em 169 grãos, dos quais foram capturadas 209 imagens aproveitadas após o processo de pré-processamento. Essa quantidade de imagens é atribuída ao fato de muitos grãos apresentarem mais de um tipo de defeito, resultando em múltiplas capturas por grão. Além disso, algumas imagens foram descartadas durante a fase de pré-processamento, devido a critérios para garantir a

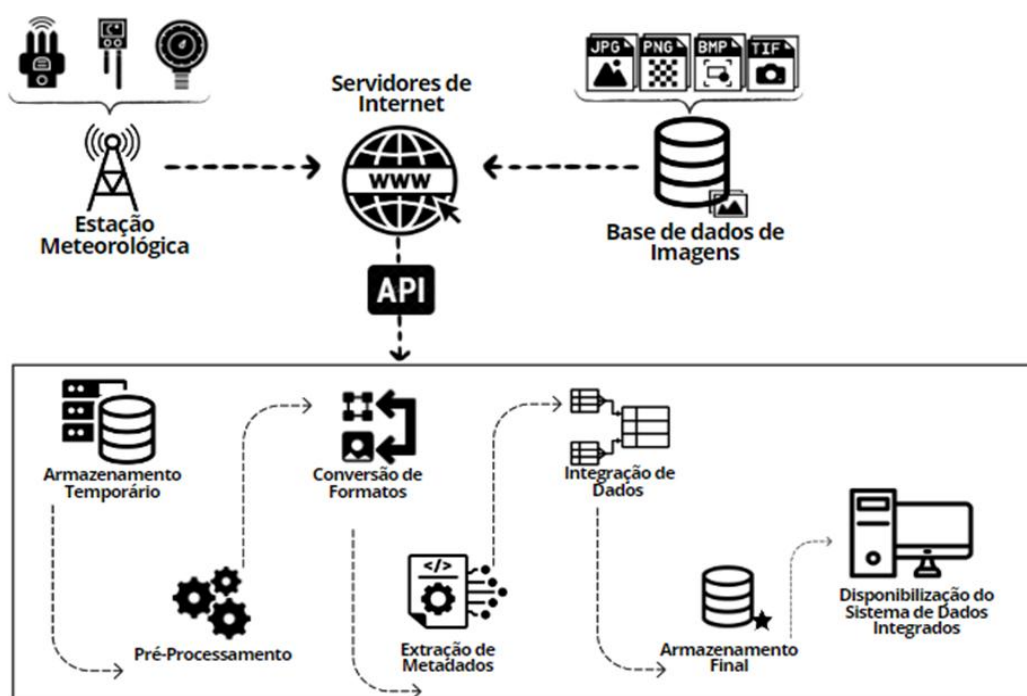
qualidade e a nitidez das imagens. Esse cuidado no pré-processamento assegura que as imagens utilizadas na análise fossem representativas, permitindo uma avaliação precisa dos defeitos presentes nos grãos.

Além da coleta de dados visuais, foi realizada uma discussão teórica sobre a utilização desses dados para o monitoramento da saúde das plantas e a qualidade dos grãos. Essa discussão motivou a integração de tecnologias de imagem com sistemas de monitoramento climático para prever e reduzir os impactos das condições extremas. Ao utilizar imagens e análise de dados, os agricultores podem tomar decisões sobre práticas de manejo, como aplicação de defensivos agrícolas e irrigação, melhorando a produtividade e a qualidade dos produtos agrícolas.

4.2 MODELO DE INTEGRAÇÃO DE DADOS CLIMÁTICOS E DE IMAGENS

O modelo (Figura 25) foi desenvolvido para integrar informações provenientes de estações meteorológicas e bases de dados em um único repositório. Essa integração permite a centralização dos dados climáticos e de produção agrícola, facilitando acesso e gerenciamento das informações. Ao unificar diferentes fontes de dados, o modelo oferece uma estrutura organizada e eficiente, capaz de dar suporte a análises complexas e consultas em tempo real, para o monitoramento contínuo das condições que afetam a produção de grãos.

Figura 25. Modelo de Integração de Dados Agrícolas.



Fonte: A autora.

O modelo de integração (Figura 25) abrange etapas para garantir uma gestão eficiente dos dados coletados e dar suporte adequado às análises desejadas. Cada etapa foi planejada para otimizar coleta, organização e processamento de informações climáticas e de produção. As etapas incluem desde a extração dos dados de estações meteorológicas e de bases agrícolas até a transformação e armazenamento em um formato consistente e acessível. Além disso, o modelo contempla processos de validação, garantindo a precisão e a confiabilidade dos dados integrados, os quais servirão de base para a geração de informações por meio de relatórios e análises automatizadas.

As estações meteorológicas coletam dados em tempo real sobre várias condições atmosféricas, como temperatura, umidade, pressão atmosférica, velocidade do vento, entre outras. Esses dados são capturados por sensores e armazenados em um sistema local. Os dados coletados são formatados para garantir compatibilidade com os padrões exigidos pelos sistemas de transmissão e pelos servidores que irão processá-los. A formatação pode incluir a conversão dos dados em formatos estruturados, como CSV (*Comma-Separated Values*), JSON (*JavaScript Object Notation*) ou XML (*eXtensible Markup Language*).

Os dados são transmitidos das estações meteorológicas para os servidores na Internet, por meio de vários métodos, em conexões de rede, como Ethernet ou Wi-Fi. A transmissão pode ser feita por protocolos de comunicação como HTTP (*HyperText Transfer Protocol*) ou FTP (*File Transfer Protocol*). O protocolo depende da arquitetura do sistema e dos requisitos de segurança.

Em áreas remotas, onde a conexão com a internet via cabo pode não estar disponível, as estações podem usar redes móveis (como 3G, 4G ou 5G) para enviar dados. Em oceanos ou regiões polares, os dados podem ser enviados via satélite. A estação meteorológica transmite dados para um satélite, que os retransmite para os servidores terrestres.

Durante a transmissão, os dados são codificados em formato binário ou textual. A codificação binária é usada para compactar e proteger os dados, enquanto a codificação textual (como JSON ou XML) pode ser utilizada para garantir compreensão e integração dos dados com outros sistemas. Os dados chegando aos servidores de Internet, são armazenados em bancos de dados e processados conforme demanda. Os dados podem ser analisados, visualizados em gráficos ou mapas, e utilizados para previsões meteorológicas. Sistemas de gerenciamento de

dados e ferramentas analíticas são empregados para interpretar os dados e apresentar informações aos usuários finais.

Os dados processados são disponibilizados em sites de internet, aplicativos móveis ou plataformas de visualização. Estes sites podem fornecer interfaces gráficas interativas, relatórios detalhados e recursos para consultas às informações meteorológicas, tais como API's para coleta de dados. Essas etapas garantem a transmissão dos dados meteorológicos coletados, de forma eficiente e segura, para análise e visualização online.

A coleta de dados de API's de estações meteorológicas na Internet segue um processo específico, envolvendo várias etapas e tecnologias. Primeiramente, é necessário identificar a API da estação meteorológica ou do serviço meteorológico fornecedor dos dados. Essas API's são oferecidas por serviços meteorológicos, plataformas de dados climáticos ou diretamente por fabricantes de estações meteorológicas.

Algumas APIs exigem autenticação para garantia de acesso aos usuários autorizados. Isso pode envolver o uso de chaves de API (*API keys*), *tokens* de acesso ou outros métodos de autenticação. O usuário precisa obter essas credenciais, fornecidas pelo provedor da API ao registrar-se no serviço.

Para coletar dados, é necessário configurar e enviar uma solicitação HTTP para a API. A solicitação pode ser feita usando Métodos HTTP, como GET (para obter dados) ou POST (para enviar dados). A URL da solicitação inclui parâmetros com especificação dos dados desejados, como localização, intervalo de tempo, e tipo de informações meteorológicas. A solicitação é enviada ao servidor da API, para processar a solicitação e retornar os dados solicitados. O envio pode ser feito usando ferramentas de linha de comando, como *cURL*, ou bibliotecas de programação, como *requests* em *Python*, *axios* em *JavaScript*, entre outras.

A resposta da API é recebida em Formato JSON, XML, ou outro formato de dados estruturados. O formato mais comum é JSON. A resposta contém os dados meteorológicos solicitados, como temperatura, umidade, precipitação, etc. Após receber os dados, são analisados e processados conforme demanda. Isso pode incluir a extração de informações específicas, a conversão de dados em um formato adequado para armazenamento, ou a integração com outras fontes de dados.

Os dados podem ser armazenados em bancos de dados, planilhas ou sistemas de gestão de dados para uso futuro. Os dados coletados podem ser apresentados em

aplicativos, websites ou plataformas de visualização. Isso pode incluir gráficos, tabelas, ou mapas interativos que permitem aos usuários visualizar e analisar as condições meteorológicas.

Imagens de grãos podem ser capturadas usando diferentes tipos de câmeras, cada uma com características distintas, como formatos, resoluções e tipos de arquivos. Essas variações influenciam a qualidade e a aplicabilidade das imagens na análise agrícola. Câmeras digitais comuns podem gerar imagens em formatos como JPEG, PNG ou TIFF, enquanto câmeras avançadas, como câmeras hiperespectrais, podem produzir dados em formatos específicos, incluindo informações em várias bandas espectrais. A resolução das imagens pode variar, desde imagens de baixa resolução, oferecendo uma visão geral dos grãos, até imagens de alta resolução, capturando detalhes minuciosos para análises precisas.

Após a captura, as imagens são enviadas para servidores de Internet. Esse processo de transmissão pode ocorrer por meio de diversos protocolos, como FTP ou HTTP, dependendo da configuração do sistema e das demandas específicas. As imagens são enviadas em pacotes, e a comunicação pode envolver a compressão dos dados para otimizar a largura de banda e a velocidade de *upload*. Durante o envio, os dados de imagem podem ser acompanhados por informações adicionais, como *timestamps* e identificadores de origem, para garantir a integridade e o rastreamento das imagens.

No servidor, as imagens recebidas são integradas ao sistema, receptor de dados de estações meteorológicas. Para análise eficaz, tanto os dados meteorológicos quanto as imagens precisam passar por um processo de pré-processamento e extração de metadados (Quadro 2). O pré-processamento envolve normalização e ajuste das imagens e dados meteorológicos, preparando-os para análise. Isso pode incluir a correção de distorções de imagem, a remoção de ruídos e o ajuste de brilho e contraste, bem como a conversão dos dados para formatos compatíveis.

Quadro 2. Processo de Integração e Pré-Processamento de Dados Meteorológicos e Imagens de Grãos.

Etapas	Descrição	Tecnologias Úteis
Recepção dos Dados	Recebimento das imagens de grãos e dos dados meteorológicos no servidor.	Protocolos de Comunicação (FTP, HTTP)

Etapa	Descrição	Tecnologias Úteis
Armazenamento Inicial	Armazenamento temporário dos dados recebidos para processamento.	Banco de Dados, Armazenamento em Nuvem
Pré-Processamento de Imagens	Correção de distorções, remoção de ruídos, ajuste de brilho/contraste, etc.	Software de Processamento de Imagens (OpenCV, MATLAB)
Conversão de Formatos	Conversão das imagens e dados meteorológicos para formatos compatíveis.	Ferramentas de Conversão de Dados
Extração de Metadados	Extração de informações relevantes, como resolução de imagem, tipo de câmera.	Scripts Personalizados, Ferramentas de Metadados
Integração de Dados	Combinação de dados meteorológicos com as imagens processadas.	Ferramentas de Integração de Dados
Armazenamento Final	Armazenamento dos dados processados e metadados para análise futura.	Bancos de Dados, Repositórios de Imagens
Disponibilização para Análise	Preparação e disponibilização dos dados para análise e tomada de decisões.	Data Warehouse

Fonte: A autora.

No pré-processamento de imagens, várias técnicas são aplicadas para garantir imagens em estado ideal para análise. Isso inclui a correção de distorções, remoção de ruídos e ajustes de brilho e contraste, utilizando ferramentas como *OpenCV* ou *MATLAB*. Paralelamente, ocorre a conversão dos dados para formatos compatíveis com os sistemas de análise utilizados.

A extração de metadados corresponde à coleta de informações como a resolução das imagens e o tipo de câmera utilizada. Essas informações servem para compreender o contexto dos dados e realizar análises precisas. Após a extração, os dados meteorológicos e as imagens de grãos são integradas, permitindo uma análise conjunta, podendo revelar correlações entre as condições ambientais e as características dos grãos.

Os dados processados e seus metadados são armazenados em um repositório, ficando disponíveis para futuras análises. A última etapa envolve a preparação e disponibilização dos dados em um data warehouse, podendo ser utilizados para processos de tomada de decisão, por meio de plataformas de análise, como *Python*, *R* ou *Tableau*.

O uso de data warehouse permite a centralização e organização eficiente dos dados, facilitando consultas rápidas e integradas para dar suporte às análises complexas e detalhadas. Essa estrutura organizada e detalhada possibilita ao modelo realizar análises eficazes e robustas, integrando diferentes fontes de dados para gerar informações e melhorar a gestão agrícola.

A centralização dos dados em data warehouse garante a consistência e integridade das informações, tornando a análise confiável e permitindo melhor utilização dos dados para a otimização de recursos e tomada de decisões estratégicas no setor agrícola.

A integração de dados meteorológicos e imagens de grãos permite uma análise abrangente e detalhada. Por exemplo, os dados meteorológicos podem fornecer informações sobre as condições climáticas que afetam o crescimento e a qualidade dos grãos, enquanto as imagens permitem a observação direta de características físicas e de saúde dos grãos. A combinação desses dados, após o devido pré-processamento e extração de metadados, oferece informações valiosas a ser usadas para melhorar práticas agrícolas, otimizar a gestão de recursos e tomar decisões estratégicas.

A integração de dados meteorológicos e de imagens de grãos foi realizada por meio do modelo estrela, o qual organiza as informações de forma centralizada, permitindo análise detalhada. No centro do modelo, encontra-se a tabela de fatos, a qual armazena os dados de condições climáticas registradas pelas estações meteorológicas e as características físicas observadas nas imagens dos grãos. As tabelas de dimensão, conectadas à tabela de fatos, contêm os metadados e informações contextuais, como a localização da estação meteorológica, o tipo de câmera utilizada, a data e hora da captura das imagens, entre outros.

4.3 MODELO CONCEITUAL

O modelo conceitual criado é parte do modelo de integração desta pesquisa, podendo ser visualizado na etapa “integração de dados”. O Quadro 3 apresenta uma visão das diversas fontes de dados utilizadas no contexto da agricultura de precisão e suas respectivas implicações para a construção de um modelo constelação. Este quadro permite entender como diferentes tipos de dados são integrados e utilizados em um *data warehouse* para otimizar a tomada de decisões agrícolas.

Quadro 3. Fontes de Dados e Suas Implicações

Fonte de Dados	Tipo de Dados Extraídos	<i>Data Warehouse</i>	Utilização
APIs de Serviços Meteorológicos	Dados climáticos (temperatura, precipitação, umidade etc.)	Tabela de Dados Climáticos	Permite acesso em tempo real a dados climáticos, essencial para a tomada de decisões rápidas.
Algoritmos de Reconhecimento de Padrões	Imagens de grãos (saudáveis e afetados)	Tabela de Imagens Agrícolas	Facilita a identificação precoce de problemas nas plantações, como pragas e doenças.
Aprendizado de Máquina	Dados históricos de produção e climáticos	Modelos preditivos	Melhora a precisão das previsões sobre rendimento e condições climáticas futuras.
ETL	Dados de várias fontes (climáticos, imagens, produção)	Tabelas de Fatos e Dimensões	Garante a qualidade e consistência dos dados, permitindo análises confiáveis.
Análise Estatística	Dados de rendimento, clima e pragas	Relatórios e gráficos de análise	Ajuda a entender as relações entre variáveis e a eficácia das intervenções agrícolas.
Modelos de Dados (Estrela, Floco de Neve)	Estruturas de dados organizadas	Estruturas de dados no Data Warehouse	Facilita a consulta e análise de dados, otimizando a eficiência das operações.

Fonte: A autora.

As API's de serviços meteorológicos fornecem dados climáticos em tempo real, incluindo temperatura, precipitação, umidade, entre outras variáveis. Esses dados são armazenados em uma tabela de dados climáticos no *data warehouse*. A disponibilidade de informações climáticas permite aos agricultores, a tomada de decisões rápidas e precisas, como a programação de irrigação e a aplicação de defensivos agrícolas, melhorando a eficiência das operações agrícolas.

A utilização de algoritmos de reconhecimento de padrões permite a extração de imagens de grãos, tanto saudáveis quanto defeituosos. Esses dados são organizados na tabela de imagens agrícolas do *data warehouse*. A análise dessas imagens facilita a identificação precoce de problemas nas plantações, possibilitando intervenções rápidas e eficazes para reduzir danos e garantir a saúde das culturas.

Os dados históricos de produção e climáticos são utilizados em modelos preditivos desenvolvidos por técnicas de aprendizado de máquina. Esses modelos melhoram a precisão das previsões sobre o rendimento das culturas e as condições climáticas. A capacidade de prever esses fatores com maior acurácia permite aos

agricultores planejem suas atividades com antecedência, reduzindo riscos e maximizando a produtividade.

O processo ETL é responsável por extrair, transformar e carregar dados de várias fontes, incluindo dados climáticos, imagens e informações de produção, em tabelas de fatos e dimensões no *data warehouse*. Esse processo garante a qualidade e a consistência dos dados, tornando as análises confiáveis. Dados integrados permitem a construção de modelos analíticos e a geração de informações para agregar na tomada de decisão agrícola.

A análise estatística de dados de rendimento, clima e incidência de pragas resulta na criação de relatórios e gráficos. Esses relatórios ajudam a entender as relações entre diferentes variáveis e a eficácia das intervenções agrícolas. Compreender essas interações permite o desenvolvimento de estratégias para o manejo das culturas, otimizando os recursos e melhorando os resultados.

Para assegurar a rastreabilidade agrícola, os dados coletados são armazenados de maneira estruturada, tal como um *data warehouse*. A estruturação dos dados em modelos adequados facilita a consulta e análise, permitindo a rastreabilidade completa do ciclo produtivo.

A capacidade de rastrear e monitorar cada etapa da produção agrícola, desde a semeadura até a colheita e distribuição, garante a segurança e qualidade dos produtos agrícolas. Dados precisos permitem identificar rapidamente a origem de problemas, como contaminação ou surtos de pragas, e tomar medidas corretivas. Além disso, a rastreabilidade facilita a conformidade com regulamentos de segurança alimentar e padrões de certificação, exigidos pelos consumidores e autoridades regulatórias.

A organização dos dados em modelos estruturados, como estrela e floco de neve, no *data warehouse*, facilita consulta e análise dos dados. Essas estruturas otimizam a eficiência das operações, permitindo acesso rápido e intuitivo às informações para a tomada de decisões. A utilização de modelos estruturados é necessária para a implementação de um ambiente de *data warehouse*.

A modelagem de dados foi abordada por meio da análise de diferentes modelos, como estrela e floco de neve, para entender suas vantagens e limitações no contexto da integração de dados climáticos e imagens agrícolas. A pesquisa destacou a necessidade de um modelo flexível para dar suporte à complexidade e variedade dos dados, permitindo a integração de informações de diferentes fontes e formatos. A

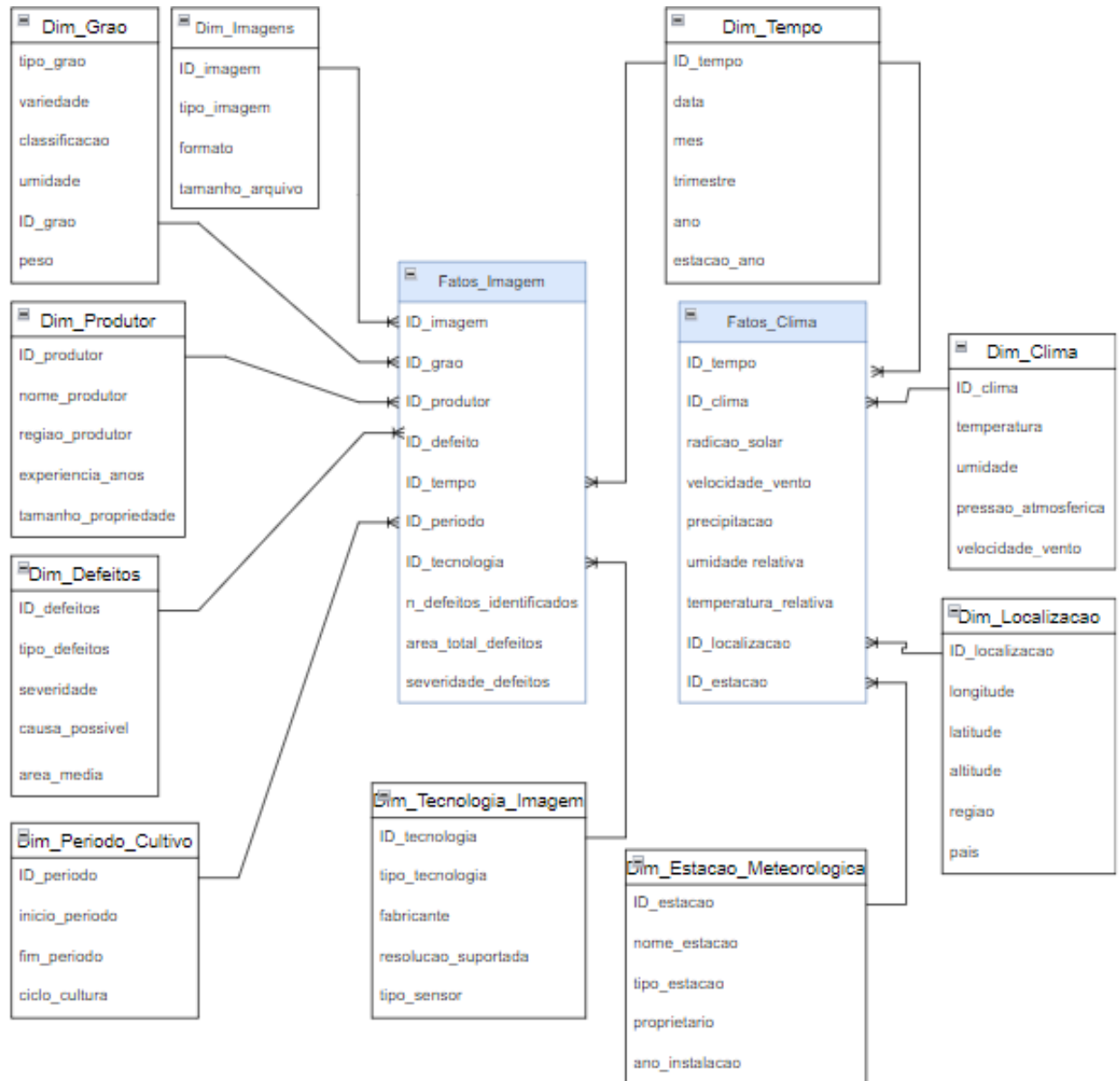
escolha de utilizar o modelo constelação na integração de dados climáticos e imagens é justificada pela complexidade e diversidade dos dados a serem analisados.

O modelo constelação combina características dos esquemas estrela e floco de neve, permitindo algumas dimensões organizadas de forma simples e direta, enquanto outras foram normalizadas. Isso é útil no contexto agrícola, onde diferentes tipos de dados (climáticos, de produção e imagens) podem ter diferentes níveis de complexidade e detalhamento

O modelo constelação facilita a análise longitudinal dos dados, permitindo avaliar tendências sazonais e variações climáticas ao longo do tempo, para entendimento de como as mudanças climáticas impactam a produção agrícola e desenvolvimento de estratégias de redução de danos. Além do mais, o modelo é escalável, permitindo novos dados e dimensões serem adicionados, sem comprometer a estrutura existente.

O modelo constelação (Figura 26) é uma estrutura de banco de dados para dar suporte à análise integrada de dados climáticos e imagens de defeitos em grãos. Este tipo de modelo é útil em grandes volumes de dados heterogêneos, os quais precisam ser analisados em conjunto. A capacidade de integrar informações de fontes variadas (clima e imagens) permite análises complexas, fundamentais para a tomada de decisões baseadas em dados, em cenários que exigem alta precisão e resposta rápida, como a detecção precoce de problemas em culturas agrícolas.

Figura 26. Modelo Dimensional para Integração de Dados Climáticos e de Imagens.



Fonte: A autora.

Para a construção das tabelas de dimensão do modelo conceitual, os atributos foram extraídos das análises realizadas. A Dimensão Imagens foi alimentada com dados oriundos da conversão de formatos de imagem, garantindo que os atributos relevantes fossem integrados. Já a tabela de grãos foi composta com base nas informações de classificação de grãos, definidas pelas instruções vigentes, abrangendo os defeitos identificados e os períodos de cultivo, para avaliação completa da qualidade dos produtos agrícolas.

Para a tabela do produtor, foram utilizadas informações obtidas por meio de pesquisa sobre rastreabilidade agrícola, garantindo que os dados necessários sobre o produtor estivessem presentes em todas as etapas da cadeia produtiva. Essa

inclusão assegura transparência e qualidade da produção. As tabelas relacionadas à tecnologia de imagem, estação meteorológica, tempo e localização foram fundamentadas nas pesquisas sobre dados coletados em estações meteorológicas, permitindo a integração dos aspectos climáticos e contextuais. Essa abordagem assegura uma representação abrangente e precisa do modelo de dados integrado.

A tabela de fatos “Fato_Imagens”, projetada para armazenar informações relacionadas às imagens capturadas de grãos com defeitos, possui atributos (Quadro 4) para maximizar o uso das análises, realizadas a partir desses dados.

Quadro 4. Atributos da Tabela de fatos “Fatos_Imagem”.

Atributo	Descrição
ID_Imagem	Identifica cada registro de imagem dentro da tabela, necessário para manter integridade dos dados e permitir correta indexação.
ID_Grão	Conecta a tabela de fatos a dimensão Dim_Grão, permitindo análises, correlacionando defeitos nas imagens com características dos grãos.
ID_Defeito	Conectada à Dim_Defeito, possibilita análises quantitativas dos defeitos e identificação dos tipos de defeitos.
ID_Tempo	Armazena data em que a imagem foi capturada e se conecta à Dim_Tempo, permitindo análises temporais.
Resolução	São atributos específicos da imagem, para qualificar a imagem e a gravidade do problema visualizado.
Número de Defeitos Identificados	
Area_Total_Defeitos	Oferecem medidas quantitativas, permitindo análises sobre a gravidade dos problemas observados nas imagens, influenciando decisões, tais como a necessidade de intervenção na lavoura.
Severidade_Media_Defeitos	

Fonte: A autora.

A Tabela de Fatos “Fato_Clima” foi criada para registrar as condições climáticas de diferentes locais e momentos, oferecendo base para análise da influência do clima sobre a produção agrícola. Cada atributo (Quadro 5) maximiza a utilidade das análises, podendo ser realizadas a partir desses dados.

Quadro 5. Atributos da Tabela de Fatos “Fato_Clima”.

Atributo	Descrição
ID_Clima	Identifica cada registro de dados climáticos.
ID_Estacao	Conecta a tabela de fatos à Dim_Estação_Meteorológica, associando dados climáticos à estação específica onde os dados foram coletados.
ID_Localizacao	Conecta a tabela de fatos à Dim_Localização , permitindo análise de dados, em termos geográficos.
ID_Tempo	Viabiliza a análise dos dados climáticos ao longo do tempo
Temperatura Media	São atributos climáticos que oferecem uma visão completa das condições ambientais, podendo ser cruzados com dados de defeitos em grãos, para identificar correlações ou causas potenciais de problemas.
Umidade_Relativa	
Precipitacao	
Velocidade_Vento	
Radiacao_Solar	

Fonte: A autora.

Em relação às tabelas de dimensão do modelo, estão organizadas em três módulos principais, cada um desempenhando um papel na integração dos dados climáticos e de imagens de defeitos em grãos. No primeiro módulo, as tabelas de dimensão conectadas à tabela de fatos “Fatos_Imagem” são necessárias para análise dos dados inerentes às imagens de grãos.

A **Dim_Grão** é responsável por armazenar dados sobre os tipos e variedades de grãos, permitindo a identificação de padrões e tendências na qualidade dos grãos, com base em suas características intrínsecas. A **Dim_Defeito** detalha os diferentes tipos de defeitos observados nas imagens, como manchas, rachaduras, ou descolorações, facilitando a análise dos problemas que afetam a qualidade dos grãos. A **Dim_Imagem** descreve as características técnicas das imagens capturadas, como resolução e formato, para avaliar como a qualidade da imagem pode influenciar a detecção e análise dos defeitos.

A **Dim_Tecnologia_Imagem** fornece informações sobre a tecnologia utilizada para capturar as imagens, permitindo a avaliação do impacto da tecnologia no processo de análise. A **Dim_Técnica_Análise** permite a descrição das técnicas

específicas de análise aplicadas às imagens, ajudando a identificar quais métodos são eficazes para detectar determinados tipos de defeitos. A **Dim_Produtor** armazena informações sobre os produtores agrícolas, oferecendo uma perspectiva sobre como diferentes práticas de cultivo podem influenciar a qualidade dos grãos e a presença de defeitos.

No segundo módulo, as tabelas dimensão conectadas à tabela de fatos “Fatos_Clima” permitiram análise dos dados meteorológicos. A **Dim_Estação_Meteorológica** contém informações sobre as estações meteorológicas, onde os dados climáticos são coletados, incluindo localização, tipos de sensores, e outros detalhes técnicos. Isso permite correlacionar as condições climáticas com dados específicos de localização, ajudando a identificar como diferentes microclimas podem afetar os grãos.

A **Dim_Localização** fornece a base geográfica para a análise, permitindo segmentar os dados climáticos por região e identificar padrões regionais, como a influência do clima em diferentes áreas de cultivo. A conexão dessas tabelas com os dados climáticos é necessária para entender como as variações meteorológicas impactam o desenvolvimento e a qualidade dos grãos ao longo do tempo.

No terceiro módulo, que se conecta a ambas as tabelas de fatos, destacam-se as dimensões para integração dos dados de imagem e dados climáticos. A **Dim_Tempo** corresponde à análise temporal, permitindo a segmentação dos dados em diferentes períodos, como dias, meses, estações ou anos. Essa tabela permite a identificação de tendências sazonais ou mudanças ao longo do tempo, podendo influenciar tanto a qualidade dos grãos quanto as condições climáticas.

Essa organização em módulos e a interconexão entre as tabelas dimensão e de fatos permitem análise integrada dos dados, possibilitando a identificação de padrões, os quais poderiam não ser evidentes em um ambiente de dados não integrado. Cada módulo e conexão foram planejados para proporcionar informações detalhadas sobre os diversos fatores que afetam a qualidade dos grãos, permitindo realizar correlações entre clima, práticas de cultivo, tecnologia de imagem, entre outros fatores. Isso resulta em um modelo robusto, dando suporte às decisões estratégicas de gestão agrícola.

O modelo constelação facilita a integração de dados climáticos e de imagens de defeitos dentro de um *data warehouse*, proporcionando visão holística e multidimensional dos fatores que afetam a qualidade dos grãos. A partir dessa

estrutura, é possível realizar análises cruzadas complexas, considerando variáveis climáticas, geográficas, temporais, tecnológicas e de manejo agrícola.

4.4 VALIDAÇÃO DA EXTRAÇÃO DE METADADOS

Nesta seção, apresentamos as validações realizadas para as etapas integrantes do modelo de integração de dados semi-automatizado, com foco específico na utilização da *API OpenWeather*.

Essa API fornece uma variedade de dados climáticos para análise e tomada de decisões no contexto agrícola. A validação dessas informações garante que os dados coletados sejam consistentes, precisos e adequados para a integração com outras fontes de dados, como as imagens dos grãos. O Quadro 6 apresenta as informações extraídas da API, incluindo o tipo de dado, formato original, formato convertido e a utilidade de cada dado no modelo de integração.

Quadro 6. Dados extraídos da API OpenWeather.

Dado	Tipo de Dado	Formato Original	Formato Convertido	Utilidade do Dado
Temperatura	Numérico	Kelvin	Celsius	Utilizado para análise de temperatura média nas regiões de cultivo.
Umidade	Percentual	Decimal	Decimal	Avalia a umidade do ar, para prever condições de armazenamento e saúde das plantas.
Pressão Atmosférica	Numérico	Hectopascal	Pascal	Ajuda a entender as condições climáticas que podem impactar a produção agrícola.
Velocidade do Vento	Numérico	Metros por segundo	Metros por segundo	Utilizado para análise de riscos de vento forte, podendo afetar a colheita.
<i>TimeStamp</i>	Data/Hora	Unix Timestamp	ISO 8601	Registra a data e hora da coleta dos dados, permitindo a correlação temporal com outros dados.

Fonte: A autora.

As informações coletadas da *API OpenWeather* foram analisadas e convertidas para um formato uniforme, garantindo que os dados pudessem ser integrados ao banco de dados. A conversão dos formatos foi realizada para assegurar a consistência e facilitar o processamento das informações. Por exemplo, a temperatura foi convertida de Kelvin para Celsius, permitindo uma interpretação intuitiva dos dados climáticos. Da mesma forma, o *timestamp* foi convertido de um formato Unix para o padrão ISO 8601, o que facilita a leitura e o manuseio das datas em aplicações futuras.

As informações coletadas da API foram referentes à região da Cidade de Rio Azul - PR, uma vez que as imagens analisadas provêm de uma fazenda localizada naquela região, relacionadas à Safra 2022/2023. O plantio foi realizado no dia 1º de outubro de 2022, e a colheita ocorreu em 17 de abril de 2023, segundo proprietário da fazenda. Essa localização e o período permite correlacionar as condições climáticas com o desenvolvimento da cultura de soja e avaliar o impacto dessas condições sobre os grãos capturados nas imagens.

O Quadro 7 apresenta de exemplos de dados recolhidos pela API sobre o dia do plantio, um dia intermediário no desenvolvimento da cultura e o dia da colheita. Esses exemplos permitem ilustrar como os dados meteorológicos são apresentados pela API, oferecendo uma visão clara de parâmetros como temperatura, umidade, velocidade do vento e pressão atmosférica. É importante destacar que a API oferece a possibilidade de coletar dados meteorológicos diariamente e em tempo real, o que permite um monitoramento contínuo e preciso das condições climáticas ao longo de toda a safra. Isso contribui para análise da correlação entre o clima e dos defeitos observados nos grãos durante o pós-processamento das imagens.

Quadro 7. Amostra de dados climáticos coletados e convertidos.

Dado	01/08/2022	08/12/2022	17/04/2023
Temperatura	15°C	28°	21°C
Umidade	0,85	0,77	0,81
Pressão Atmosférica	1 atm	1 atm	1 atm
Velocidade do Vento	1,91 m/s	4,16 m/s	3,33 m/s
TimeStamp	01/08/2022 12:00:04	08/12/2022 12:00:03	17/04/2023 12:00:06

Fonte: A autora.

Em relação à extração de metadados das imagens, os resultados obtidos ao longo das 25 épocas de treinamento foram promissores e indicativos de um aprendizado no processo de classificação das imagens dos grãos de soja. Observou-se um aumento gradual e consistente na acurácia, refletindo a capacidade do modelo de aprender e se adaptar à complexidade dos dados de entrada. Essa melhoria na acurácia é um ponto positivo, sugerindo que o modelo está se tornando cada vez mais eficiente em distinguir entre diferentes tipos de defeitos apresentados nos grãos, como ardido, chocho, fermentado, germinado, mancha púrpura, mofado, partido/quebrado, picado por inseto, queimado, sadio e mancha parda.

Por outro lado, a perda, representando a diferença entre a saída prevista pelo modelo e a saída real, mostrou tendência de redução ao longo das mesmas 25

épocas. Essa diminuição da perda é um indicativo claro de que o modelo está ajustando seus parâmetros, minimizando os erros na classificação e, conseqüentemente, aprimorando seu desempenho. O Quadro 8 ilustra essa dinâmica, possibilitando a visualização da correlação entre o aumento da acurácia e a redução da perda, e evidenciando a eficácia do processo de aprendizado.

Quadro 8. Evolução da acurácia e da perda ao longo das épocas.

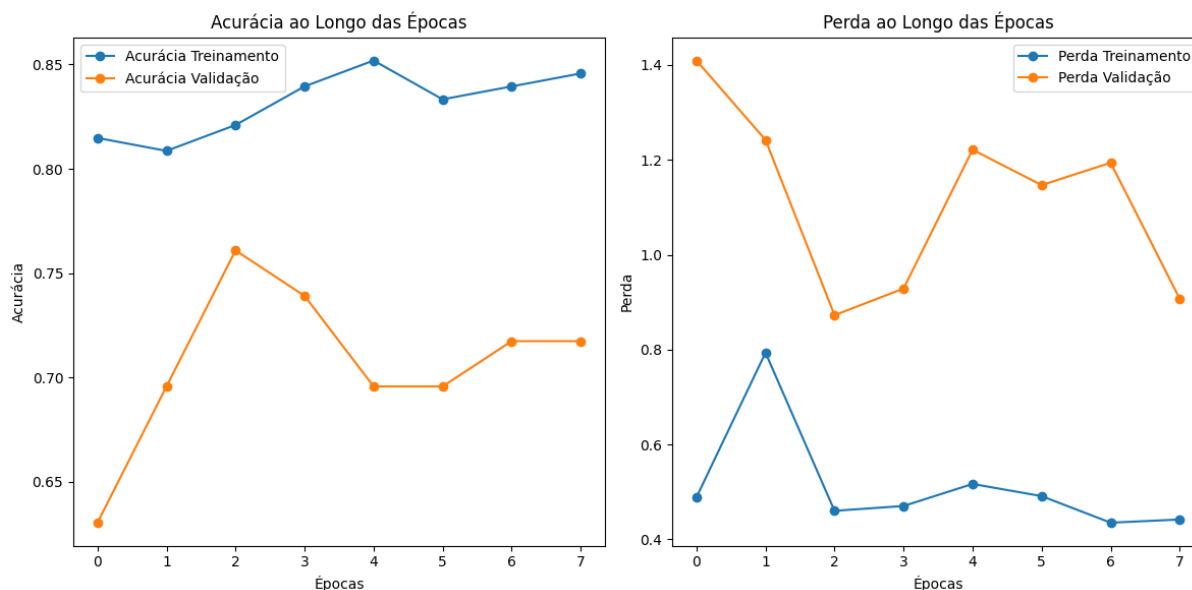
Época	Perda train	Acurácia train	Perda val	Acurácia val
0	0.4888	0.8148	1.4095	0.6304
1	0.7935	0.8086	1.2421	0.6957
2	0.4601	0.8210	0.8726	0.7609
3	0.4703	0.8395	0.9288	0.7391
4	0.5167	0.8519	1.2215	0.6957
5	0.4911	0.8333	1.1468	0.6957
6	0.4351	0.8395	1.1940	0.7174
7	0.4418	0.8457	0.9074	0.7174

Fonte: A autora.

Embora o modelo estivesse configurado para treinar por 25 épocas, o processo foi interrompido devido ao *early stopping*. No total, o modelo foi treinado por 8 épocas, pois o *early stopping* foi acionado após a 7ª época, quando não houve melhoria na perda de validação. Os dados demonstram que o modelo começou a estabilizar sua acurácia e reduzir a perda, evidenciando seu potencial para aprender os padrões, sem comprometer a capacidade de generalização para novos dados.

É possível verificar a evolução da acurácia e perda ao longo das épocas (Figura 27), embora a acurácia de validação tenha flutuado ao longo das épocas. O modelo alcançou uma taxa de acerto de 76,09% no conjunto de validação. Essa taxa de acerto reflete a porcentagem de previsões corretas feitas pelo modelo ao identificar defeitos e doenças nas imagens de grãos não visualizadas durante o treinamento. O pico de acurácia no conjunto de validação, de 76,09%, foi alcançado na terceira época.

Figura 27. Evolução da acurácia e perda ao longo das épocas.



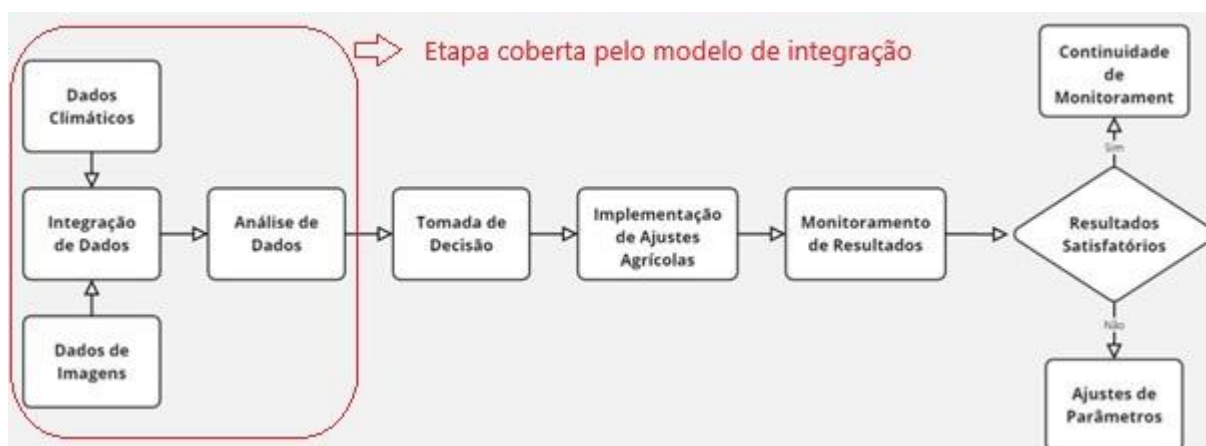
Fonte: A autora.

A variação entre a acurácia do conjunto de treino (85,19%) e de validação (76,09%) pode ser atribuída à heterogeneidade dos dados. Para melhorar a generalização do modelo, seria ideal contar com uma base de dados com mais de 209 imagens. No entanto, durante a avaliação, optou-se por não incorporar imagens desatualizadas, já que poderia introduzir ruídos e comprometer a validade dos resultados. Portanto, a limitação na quantidade e na atualização das imagens pode impactar a capacidade do modelo de reconhecer padrões e classificar corretamente os dados de validação.

4.5 INSERÇÃO DO MODELO NO FLUXO DE RASTREABILIDADE AGRÍCOLA

O modelo apresentado pode ser aplicado no fluxo geral de informações, como ilustrado na Figura 28, permitindo análises e a geração de relatórios para a gestão agrícola. Ele possibilita, por exemplo, a avaliação de tendências na produção de soja em relação às condições climáticas, a elaboração de relatórios detalhados sobre a produção em diferentes períodos, e a identificação de padrões por meio de algoritmos de aprendizado de máquina. Além disso, a criação de *dashboards* integradores de dados climáticos e de produção facilitará a visualização clara e intuitiva, auxiliando os gestores agrícolas na tomada de decisões.

Figura 28. Fluxo de Informações Baseado na Integração de Dados.



Fonte: A autora.

O modelo de integração de dados é a primeira etapa do fluxo de informações para a tomada de decisão no contexto agrícola. Nesse modelo, dados de imagens e informações climáticas são combinados para fornecer uma base para a análise de dados, especialmente quando integrada com princípios de inteligência artificial.

Algoritmos de aprendizado de máquina podem ser utilizados para prever o momento para o plantio, analisando históricos de padrões climáticos e condições atuais. Além disso, técnicas de redes neurais podem ser aplicadas para otimizar o armazenamento de grãos, identificando o ponto ideal de umidade e temperatura para minimizar perdas e maximizar a qualidade do produto ao longo do tempo.

Através do fluxo de integração e análise, a tomada de decisão agrícola se torna mais precisa. Os dados podem indicar a demanda de ajustes nas datas de plantio com base nas previsões de condições climáticas, ou sugerir mudanças nos métodos de irrigação para otimizar o uso da água em períodos de seca.

Decisões de ajustes do espaçamento entre plantas, de aplicação de fertilizantes ou de controle de pragas podem ser orientadas pela análise integrada de dados de campo e climáticos, propiciando uma abordagem proativa e adaptativa.

Após a implementação das decisões, é importante manter monitoramento contínuo dos resultados obtidos. Se os resultados forem satisfatórios, o monitoramento deve continuar para garantir a manutenção dos resultados ao longo do tempo. Se a alteração na época de plantio resultar em um aumento da produtividade, continuar o monitoramento das condições para ajuste rápido em caso de mudanças imprevistas.

No caso de os resultados não ser satisfatórios, é necessário realizar ajustes nos parâmetros para corrigir as discrepâncias. Isso pode incluir a recalibração dos algoritmos de previsão de colheita, o ajuste na quantidade de insumos utilizados, ou a alteração de práticas de manejo, como a profundidade de plantio ou a frequência de irrigação. Além disso, a análise dos dados históricos e a comparação com os resultados atuais podem fornecer informações sobre novos ajustes, permitindo uma adaptação contínua e dinâmica às condições variáveis do campo.

Os ajustes e o monitoramento contínuo são necessários para otimização dos processos agrícolas e para obtenção de resultados consistentes e sustentáveis. A capacidade de adaptar as práticas agrícolas com base em uma análise de dados robusta e fundamentada em inteligência artificial coloca os gestores agrícolas em uma posição de vantagem, permitindo a correção de desvios e a antecipação de problemas, promovendo uma agricultura eficiente e resiliente.

5 CONCLUSÕES E PESQUISAS FUTURAS

Nesta dissertação apresentou-se uma análise sobre a interconexão entre dados climáticos e imagens, destacando a relevância desta integração para a otimização da gestão agrícola, argumentando que, apesar da crescente disponibilidade de dados, a fragmentação e a falta de suporte tecnológico eficaz limitam a capacidade dos agricultores de tomar decisões, impactando diretamente na produtividade e na sustentabilidade das práticas agrícolas. A pesquisa revela que a utilização de técnicas de aprendizado de máquina, aliada à integração de múltiplas fontes de dados, pode transformar a forma como os agricultores gerenciam suas culturas, promovendo uma agricultura adaptável e resiliente.

A captura e análise de imagens atualizadas dos grãos de soja representam um avanço no monitoramento agrícola. A pesquisa preenche uma lacuna na base de dados existente e abre possibilidades para a aplicação de novas tecnologias na agricultura. A inovação e atualização permitem enfrentar os desafios impostos pelas mudanças climáticas e garantir a sustentabilidade e resiliência do setor.

Para enfrentar os desafios identificados, propõe-se o desenvolvimento de uma plataforma integrada que permita a coleta, análise e visualização em tempo real de dados climáticos e imagens agrícolas. Essa solução facilitaria o acesso a informações e capacitaria os agricultores a tomarem decisões com informações relevantes e abrangentes, alinhadas às variáveis ambientais em constante mudança.

Os resultados indicam que a integração de dados climáticos e de imagens pode melhorar a eficiência da produção de soja e contribuir para a sustentabilidade agrícola. A capacidade de prever e responder a eventos climáticos extremos, pode reduzir perdas e aumentar a qualidade dos grãos, beneficiando tanto os agricultores quanto os consumidores. A pesquisa enfatiza a relevância de explorar tecnologias para integração de dados, sendo diferencial competitivo no setor.

A identificação de padrões climáticos e suas correlações com a produção de soja permitem a construção de um arcabouço teórico que pode ser explorado em futuras pesquisas. Os dados coletados podem ser integrados com tecnologias avançadas, como inteligência artificial e Internet das Coisas, para criar sistemas automatizados de monitoramento e análise. Algoritmos de Inteligência Artificial podem ser treinados para reconhecer padrões nos dados visuais e prever a ocorrência de defeitos com base em condições climáticas futuras.

Para que as soluções propostas sejam efetivas, é necessário investir em educação e capacitação dos produtores rurais. Programas de treinamento para utilizar as tecnologias de integração de dados e interpretar as informações geradas podem maximizar os benefícios e garantir que os produtores estejam preparados para enfrentar os desafios futuros. A colaboração entre setor público, privado e instituições de pesquisa contribui para a implementação de modelos de integração de dados.

Embora este estudo tenha se concentrado na produção de soja, há um potencial para aplicar o modelo de integração de dados a outras culturas agrícolas. As variáveis climáticas e as imagens podem ser integradas para otimizar a produção de diferentes tipos de grãos, frutas e vegetais, adaptando as soluções às necessidades específicas de cada cultura.

Estudos que analisem tendências climáticas e suas correlações com a produtividade das culturas podem fornecer informações para o desenvolvimento de estratégias de adaptação e redução de danos. Fomentar a colaboração interdisciplinar, como agronomia, computação, meteorologia e ciências sociais, pode enriquecer pesquisas na área agrícola. Essa abordagem pode levar a soluções inovadoras, abordando aspectos técnicos, dimensões sociais e econômicas.

"Integrar dados é semear o futuro da agricultura." Essa frase resume a essência do trabalho, enfatizando que a verdadeira inovação no setor agrícola reside na capacidade de unir informações dispersas em um todo coeso e funcional. Ao retornar à ideia central apresentada na introdução, fica claro que a integração de dados climáticos e de imagens não é apenas uma necessidade técnica, mas uma questão de sobrevivência e prosperidade para a agricultura.

Em um mundo onde as mudanças climáticas ameaçam a segurança alimentar e a sustentabilidade, esta pesquisa apresenta um futuro em que a tecnologia e a natureza coexistem em harmonia. Ao adotar uma abordagem integrada, é possível otimizar a produção agrícola e cultivar um planeta mais saudável e resiliente para as gerações futuras. A transformação da agricultura não é apenas uma possibilidade; é uma responsabilidade que devemos abraçar com urgência e determinação.

REFERÊNCIAS

- ABIOVE. **Manual de Boas Práticas de Classificação de Soja**. Brasil. 2024.
- AGROS. **Impacto do La Niña na Agricultura Brasileira: Estratégias para Safra 24/25**. Disponível em: <https://www.agrosnutrition.com.br/impacto-do-la-nina-na-agricultura-brasileira-estrategias-para-safra-24-25>. Acesso em: 09 de out. de 2024.
- AKID, H. *et al.* Desempenho de implementações de gráficos NoSQL de esquemas estrela vs. floco de neve. **IEEE Access**. 2022.
- ALMEIDA, F. S.; *et al.* Predicting Crop Yields Using Satellite Images and Machine Learning: A Study in Brazilian Agriculture. **Computers and Electronics in Agriculture**, v. 182, p. 105–115, 2021.
- ALMEIDA, F. L.; CARVALHO, J. P.; RIBEIRO, L. S. A importância da integração de dados na agricultura: inovações para a gestão de riscos e aumento da produtividade. **Revista de Administração e Inovação**, v. 18, n. 2, p. 45-60, 2021.
- ANJOS, E. M.; VIEIRA, L. M.; ALMEIDA, F. M. Vision Transformers: uma revisão sobre o desempenho em reconhecimento de padrões. **Revista de Ciências da Computação**, v. 12, n. 3, p. 123-145, 2022.
- AZAD, R. *et al.* Avanços na análise de imagens médicas com transformadores de visão: uma revisão abrangente. **Análise de imagens médicas**. 2023.
- BASSO, B.; ANTLE, J. Digital agriculture to design sustainable agricultural systems. **Nature Sustainability**, v.3, n.4, p.254-256, 2020.
- BAKHSHANDEH, M., e MOUSAVI, H. Early Stopping: A Simple Yet Powerful Technique in Neural Networks. **IEEE Transactions on Neural Networks and Learning Systems**. 2023.
- BALDOMINOS, P.; GARCÍA, M.; ROGGEN, D. Feature learning for Human Activity Recognition using Convolutional Neural Networks. **CCF Transactions on Pervasive Computing and Interaction**, 2019.
- BESSEN, G. R.; *et al.* A integração de dados em sistemas de informação agrícola: um estudo de caso no Brasil. **Revista Brasileira de Agricultura de Precisão**, v. 12, n. 2, p. 165-177, 2019.
- BORGES, A. L.; RIBEIRO, A. R. **Grãos: Aspectos de Qualidade e Destinação**. 1. ed. Brasília: Embrapa, 2021.
- BRASIL. **Decreto nº 3.664, de 17 de novembro de 2000**. Diário Oficial [da] República Federativa do Brasil, Brasília, DF. 2000.
- BRASIL. **Lei nº 9.972, de 25 de maio de 2000**. Diário Oficial [da] República Federativa do Brasil, Brasília, DF. 2000.
- BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. **Instrução Normativa nº 38, de 4 de agosto de 2006**. Estabelece os procedimentos e critérios para a classificação de grãos de soja. Diário Oficial da União, Brasília, 2006.
- BROWN, P. S.; BELL, M. Challenges in Agriculture: A Diverse Data Universe. **Nature**, v. 572, p. 27-29, 2019.

CAMPOS, F. F.; LIMA, T. A.; OLIVEIRA, A. M. Uso de tecnologias de informação e comunicação na agricultura: impactos das mudanças climáticas na produtividade de culturas. **Revista Brasileira de Agrometeorologia**, v. 29, n. 1, p. 1-12, 2021.

CARVALHO, F. R.; SOUZA, L. M. Transparência e Prestação de Contas na Agricultura: O Papel da Centralização de Dados. **Revista de Gestão Agrícola**, v. 8, n. 1, p. 30-42, 2020.

CHEN, J.; *et al.* Defect Detection in Food Products Using Deep Learning: Application of ResNet50. **Journal of Food Science**, v. 87, n. 4, p. 1123-1131, 2022.

CIFERRI, C. D. A. Modelagem Multidimensional: processamento analítico de dados. **Periódicos USP**. 2022.

COSTA, C. E. *et al.* Desafios na coleta de dados em estações meteorológicas: um estudo sobre a precisão e a variabilidade dos dados. **Revista Brasileira de Meteorologia**, v. 36, n. 3, p. 401-412, 2021.

COSTA, E. A.; ALMEIDA, L. L.; SILVA, M. A.; SOUZA, J. A. Uso de dados climáticos e tecnologias de informação para a agricultura: impactos na produtividade e sustentabilidade. **Revista Brasileira de Ciências Agrárias**, v. 15, n. 1, p. 87-95, 2020.

COSTA, N. P.; SILVA, A. F. **Qualidade dos Grãos: Aspectos Físicos, Químicos e Nutricionais**. 2. ed. Brasília: Embrapa, 2018.

DAGAR, V. *et al.* Variações na eficiência técnica de agricultores com tamanhos de terra distintos em zonas agroclimáticas: evidências da Índia. **Journal of Cleaner Production**. 2021.

DANIEL, E. E TSCHORSCH, F. IPFS e amigos: uma comparação qualitativa de redes de dados ponto a ponto de próxima geração. **IEEE Communications Surveys & Tutorials**. 2021.

DATE, C. J. **Database System Concepts**. 8. ed. New York: McGraw-Hill, 2019.

DERRY, A., KRZYWINSKI, M., & ALTMAN, N. Redes neurais convolucionais. **Nature Methods**. 2023.

DEY, S. **Hands-On Image Processing with Python**. 1. ed. Birmingham: Packt Publishing, 2019.

DU, X. *et al.* A method for detecting the quality of cotton seeds based on an improved ResNet50 model. **PLOS ONE**, v.18. 2023.

DUNCAN, J.; RIZZO, D. M.; SINGH, G. Data-driven agriculture and its implications for the future of farming. **Agricultural Systems**, v.191, p.103-109, 2021.

DUPONT, P. *et al.* Integração de Modelos de Previsão de Safras de Soja com Data Warehouses Climáticos. **Revista Brasileira de Tecnologia Agrícola**, v. 14, n. 3, p. 78-89, 2021.

ECK, G.; *et al.* Big Data: A construção de um modelo para a gestão da produção agrícola. **Revista Brasileira de Agricultura de Precisão**, v. 13, n. 1, p. 1-12, 2020.

FAO - Food and Agriculture Organization of the United Nations. **The State of Food Security and Nutrition in the World 2020**. Rome: FAO, 2020.

FERREIRA, T. R. *et al.* A Importância da Centralização de Dados na Agricultura: Um Estudo de Caso sobre Condições Climáticas e Padrões de Crescimento de Plantas. **Revista Brasileira de Agricultura de Precisão**, v. 11, n. 2, p. 45-52, 2020.

FONSECA, G. H. G. Modelagem de Data Warehouse. **Periódico da Universidade Federal de Ouro Preto**, 2023.

FU, X. *et al.* **Estudo sobre transmissão de comunicação via satélite de informações meteorológicas entre estações principais com base em transferência de estado.** Anais da 8ª Conferência Internacional sobre Tecnologia Eletrônica e Ciência da Informação. 2023.

FURLAN, L. R.; PIAZZA, R.; MOREIRA, T. D. A influência dos silos de dados na eficiência organizacional: evidências do setor agrícola. **Revista de Gestão e Tecnologia**, v. 17, n. 4, p. 245-263, 2017.

GENÇAY, R. Early stopping based on cross-validation. **ResearchGate**, 2023.

GOLAN, E. *et al.* Traceability in the US food supply: economic theory and industry studies. Washington, DC: United States Department of Agriculture. **Economic Research Service**, 2004.

GOLFARELLI, M.; GOKCEL, A. **Data Warehouse Design: Modern Principles and Methodologies.** Cambridge: Cambridge University Press, 2020.

GOLOMBIÉSKI, E. E.; VAZ, M. S. M. G.; GUIMARÃES, A. M. Integration of RFID Technology in Agricultural Traceability to Optimize Control and Quality of Products. **Iberoamerican Journal of Applied Computing**, v.12, n.1, 2024.

GOMES, L. S.; MARTINS, F. C. S.; OLIVEIRA, R. B. A Importância da Soja na Segurança Alimentar Global: Uma Revisão de Literatura. **Revista Brasileira de Agricultura Sustentável**, v. 10, n. 2, p. 45-52, 2020.

GONZALEZ, R. C. e WOODS, R. E. **Digital Image Processing** (4th ed.). Pearson. 2018.

GOOGLE. **Google Acadêmico.** Disponível em: <<https://scholar.google.com>>. Acesso em: 10 out. 2024.

GURIEVA, K., TARASOVA, E., E KHABA, N. Revisão analítica de métodos para determinar danos mecânicos em grãos de trigo sarraceno. Tovaroved prodovolstvennykh tovarov. 2023.

HARIS, M. *et al.* **Design e desenvolvimento de estação de monitoramento de clima e qualidade do ar baseada em IoT.** Conferência Internacional de 2023 sobre Robótica e Automação na Indústria (ICRAI). 2023.

HARRISON, J. M., *et al.* The role of seasonal climate forecasts in agriculture: Adapting to climate variability and change. **Weather and Climate Extremes**, v. 27, p. 100268, 2020.

HE, K.; *et al.* **Deep residual learning for image recognition.** In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), p. 770-778. 2016.

HUANG, S.; FAN, J.; WANG, Q. Resilience in agriculture through data centralization: A review. **Journal of Agricultural Data Science**, v.6, n.3, p.123-132, 2020.

HUSSAIN, A.; AHMAD, B.; RASHID, M. Climate change impacts on the El Niño Southern Oscillation (ENSO) phenomenon: A review. **Environmental Science and Pollution Research**, v. 28, p. 6357–6372, 2021.

HYNDMAN, R. J.; ATHANASOPOULOS, G. **Forecasting: Principles and Practice**. 3. ed. OTexts, 2021.

INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE (IPCC). Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. **Cambridge University Press**, 2021.

JIN, W.; LI, L.; XU, Z. Integration of Climate Data and Satellite Images for Crop Prediction Using Data Warehousing Techniques. **Journal of Agricultural Informatics**, v. 10, n. 2, p. 45-58, 2019.

JONES, R. S.; MARTINS, A. C.; OLIVEIRA, F. R. Benefícios do Monitoramento Detalhado da Saúde das Plantas de Soja por Meio de Data Warehouses. **Journal of Agricultural Science**, v. 38, n. 2, p. 112-125, 2021.

JONES, A.; SMITH, K.; TAYLOR, P. Data security in the agricultural sector: Ensuring safe and reliable centralized systems. **Computers and Electronics in Agriculture**, v.183, p.106-112, 2021.

JUNQUEIRA, N. T. V.; TAVARES, D. C. **Micotoxinas: Efeitos na Saúde Humana e Animal**. 1. ed. São Paulo: Livraria Varela, 2020.

KARANIKOLAS, N. *et al.* Data management and analytics for precision agriculture: A systematic review. **Agricultural Systems**, v. 190, p. 103091, 2021.

KARFIA, A.; CAMPOS, A. A.; NUNES, F. A tecnologia da informação e a agricultura de precisão: um estudo sobre a utilização de dados para tomada de decisão na agricultura. **Revista Brasileira de Agricultura de Precisão**, v. 14, n. 2, p. 99-112, 2021.

KUMAR, A.; *et al.* Maize Leaf Disease Detection Using ResNet50. **Computers and Electronics in Agriculture**, v. 170, p. 105-114, 2020.

KUMAR, K., *et al.* Influence of climate change on the characteristics of El Niño and La Niña: A review. **Earth-Science Reviews**, v. 234, p. 104131, 2022.

LAPIX (Santa Catarina). UFSC. **Deep Learning::: detecção de objetos em imagens. Detecção de Objetos em Imagens**. 2018. Disponível em: <https://lapix.ufsc.br/ensino/visao/visao-computacionaldeep-learning/deteccao-de-objetos-em-imagens/>. Acesso em: 01 ago. 2024.

LI, Z. Practice of Gesture Recognition Based on Resnet50. **Journal of Physics: Conference Series**, 1574. 2020.

LIU, P.; *et al.* Quality Control in Manufacturing using Deep Learning: A Case Study with ResNet50. **Journal of Manufacturing Processes**, v. 45, p. 312-320, 2020.

LOBELL, D. B. *et al.* Prioritizing climate change adaptation needs for food security in 2030. **Science**, v. 319, n. 5863, p. 607-610, 2018.

LOPES, J. A.; ALMEIDA, J. P.; SOUSA, A. R. Silos de dados: causas e soluções para a integração de informações em organizações. **Revista Brasileira de Gestão e Desenvolvimento Regional**, v. 15, n. 2, p. 45-58, 2019.

LUVISI, A. *et al.* Unmanned Aerial Vehicle (UAV) and Digital Soil Map: A Preliminary Study on Nematode Detection. **Sustainability**, v. 13, n. 12, p. 1-12, 2021.

MACHADO, R. S. OLIVEIRA, A. L. SOUZA, P. C. Avanços na Gestão de Dados Agrícolas: O Papel dos Data Warehouses na Tomada de Decisões na Agricultura de Grande Escala. **Revista Brasileira de Agricultura de Precisão**, v. 12, n. 3, p. 241-256, 2020.

MACHADO, R. S. Capacidade de Armazenamento Longitudinal de Dados Agrícolas em Data Warehouses: Avaliação de Tendências e Variações Climáticas. **Revista Brasileira de Agricultura de Precisão**, v. 14, n. 3, p. 45-58, 2022.

MANNINO, Mark V. **Data Warehousing: Architecture and Implementation**. 3rd ed. New York: Springer, 2020.

MARR, B. **Data Strategy: How To Profit From A World Of Big Data, Analytics And The Internet Of Things**. 1st ed. London: Kogan Page, 2021.

MARTINS, *et al.* Integração de dados climáticos e análise em tempo real na agricultura: desafios e oportunidades. **Revista Brasileira de Agricultura e Sustentabilidade**, v. 11, n. 4, p. 123-138, 2022.

MARTINS, A. C.; OLIVEIRA, J. R.; PEREIRA, M. F. Centralização de Dados na Agricultura: Estratégias para Evitar Inconsistências e Melhorar a Integração. **Revista Brasileira de Agricultura Sustentável**, v. 6, n. 2, p. 45-57, 2019.

MARTINS, C. F.; PEREIRA, J. R. **Economia do Agronegócio: Impactos da Qualidade dos Grãos nos Custos e Preços**. 1. ed. São Paulo: Editora Rural, 2020.

MICROSOFT LEARN. **Data warehouse moderno para pequenas e médias empresas**. Microsoft. 2024.

MIRANDA, A. N. **Arquitetura dos Vision Transformers**. 2024. Disponível em: <https://norven.com.br/arquitetura-dos-vision-transformers/>. Acesso em: 10 jun. 2024.

MIRANDA, F. D. A.; SILVA, C. R. F.; MENEZES, T. R. Impactos das tecnologias de informação na produção de soja: uma análise das variáveis climáticas e da sustentabilidade. **Revista Brasileira de Ciência do Solo**, v. 43, 2019.

MUKHERJEE, S. **The Annotated ResNet-50: explaining how resnet-50 works and why it is so popular. Explaining how ResNet-50 works and why it is so popular**. 2022. Disponível em: <https://towardsdatascience.com/the-annotated-resnet-50-a6c536034758>. Acesso em: 21 ago. 2024.

NASCIMENTO, A. A.; OLIVEIRA, R. L. **Qualidade dos Grãos e sua Relação com a Aceitação do Mercado**. 1. ed. Brasília: Embrapa, 2019.

NAZIR, A., E WANI, M. **You Only Look Once - Modelos de detecção de objetos: uma revisão**. 10ª Conferência Internacional sobre Computação para o Desenvolvimento Global Sustentável (INDIACom). 2023.

- NGUYEN, Q. *et al.* Towards a Standardized Approach for the Geographical Traceability of Plant Foods Using Inductively Coupled Plasma Mass Spectrometry (ICP-MS) and Principal Component Analysis (PCA). **Foods**, v.12. 2023.
- NI, C.; *et al.* Diagrama de constelação tridimensional com um design de nível hierárquico para transmissão multinúcleo. **Optics express**. 2022.
- OGUNJO, S. T. *et al.* Análise multifractal das temperaturas do ar e do solo. **Chaos**. 2021.
- OLIVEIRA, M. *et al.* Calibração de algoritmo de temperatura da superfície terrestre por meio de estações meteorológicas. **Revista Pesquisa Agropecuária Brasileira**. 2021.
- O'NEAL, M. R. *et al.* Evaluating UAV-based ultraviolet and visible imagery for estimating soybean biomass and nitrogen status. **Precision Agriculture**, v. 19, n. 5, p. 757-772, 2018.
- PAGLIARI, L.; BARTLETT, P.; SHAW, M. Challenges of implementing centralized agricultural data systems. **Agricultural Informatics**, v.8, n.2, p.145-154, 2019.
- PALIWAL, M. E SARASWAT, P. Abordagens De Data Warehousing E Suas Aplicações: Uma Revisão. **International Journal of Innovative Research in Computer Science & Technology**. 2022.
- PEREIRA, A. L. Rastreabilidade de Alimentos na Agricultura: Impacto da Precisão dos Dados na Segurança e Qualidade dos Produtos Agrícolas. **Revista de Segurança Alimentar**, v. 5, n. 2, p. 75-88, 2019.
- POLONI, K. **Redes Neurais Convolucionais: uma breve história sobre as arquiteturas. Uma breve história sobre as arquiteturas**. 2022. Disponível em: <https://medium.com/itau-data/redes-neurais-convolucionais-2206a089c715>. Acesso em: 06 jun. 2024.
- PRENDERGAST, P., *et al.* Early stopping for deep learning: A review. **Neural Computing and Applications**, v. 32, n. 12, p. 8935-8951, 2020.
- REDGATE. **Oracle SQL: A Beginner's Guide**. McGraw-Hill, 2021.
- RIBEIRO, C. G. *et al.* Qualidade dos grãos de soja armazenados em diferentes condições. **Ciência Rural**, v. 52, n. 1, p. 1-9, 2024.
- RICHMAN, Jeffrey. **Batch Processing vs Stream Processing: Pros, Cons, Examples**. 2024. Disponível em: <https://estuary.dev/batch-processing-vs-stream-processing/>. Acesso em: 09 out. 2024.
- ROCHA, T. K. F. *et al.* **Agricultura de precisão: desafios e perspectivas para a agricultura tropical**. Lavras: Editora UFLA, 2019.
- RODRÍGUEZ, L.; *et al.* Wireless Sensor Networks in Pest Control: A New Approach to Crop Monitoring. **Agricultural Systems**, v. 182, p. 101–110, 2020.
- ROSE, D. C.; CLARK, M.; BRAUNHOLTZ, S. Artificial intelligence in agriculture: Potential impacts and challenges. **Nature Sustainability**, v.1, n.3, p.253-255, 2018.
- ROSEBROCK, A. **Deep Learning for Computer Vision with Python**. 1. ed. New York: PyImageSearch, 2017.

SANTOS, M. R. C. *et al.* Colaboração Multidisciplinar na Adaptação Agrícola à Mudança Climática: Um Estudo de Caso na Agricultura de Soja. **Revista de Estudos Agroambientais**, v. 8, n. 1, p. 78-89, 2021.

SANTOS, R. A., & SILVA, M. C. **Utilização de Tecnologias de Data Warehouse na Agricultura: Uma Revisão Sistemática da Literatura**. Anais do XX Simpósio Brasileiro de Sistemas de Informação (SBSI). São Paulo, Brasil, 2020.

SCOPUS. **Scopus Database**. Disponível em: <<https://www.scopus.com>>. Acesso em: 10 out. 2024.

SENAR-PR. **Classificação de grãos de soja**. Sistema FAEP, 2024. Disponível em: <https://www.sistemafaep.org.br>. Acesso em: 9 fev. 2024.

SLIMANI, H., MHAMDI, J., & JILBAB, A. Assessing the advancement of artificial intelligence and drones' integration in agriculture through a bibliometric study. **International Journal of Electrical and Computer Engineering (IJECE)**. 2024.

SHARMA, R.; AGARWAL, M. Business Intelligence: A Comprehensive Approach. **International Journal of Information Management**, v. 58, p. 101-114, 2021.

SHUKLA, S.; KAPOOR, S.; JAIN, A. Smart Agriculture using IoT, Drones and Data Analytics. **Journal of Agriculture & Food Information**, v. 21, n. 3, p. 215-236, 2020.

SILVA, P. A.; COSTA, M. R.; PEREIRA, J. S. Climate Change Impacts on Agricultural Production: A Review. **Journal of Agricultural Science**, v. 9, n. 13, p. 17-28, 2021.

SILVA, J. A.; PEREIRA, M. R. **Qualidade dos Grãos: Fatores que Influenciam e Práticas de Manejo**. São Paulo: Editora Agronômica, 2020.

SILVA, A. B.; SANTOS, C. D. A.; SOUZA, E. F. Impacto do Aquecimento Global no Ciclo de Desenvolvimento de Plantas Cultivadas. **Revista Brasileira de Ciências Agrárias**, v. 14, n. 3, p. 89-97, 2019.

SILVA, J. A. **Ontologia na Rastreabilidade de Dados Agrícolas**. Dissertação (Mestrado) - Curso de Mestrado em Computação Aplicada, Programa de Pós-Graduação em Computação Aplicada, Universidade Estadual de Ponta Grossa. 2018.

SILVA, C. A.; GARCIA, M. L.; SOUZA, J. P. Monitoramento e Otimização do Uso de Insumos Agrícolas por Meio de Data Warehouses: Estudo de Caso na Cultura da Soja. **Revista Brasileira de Agricultura Sustentável**, v. 7, n. 3, p. 45-58, 2022.

SILVA, R. M.; VASCONCELOS, L. O.; ALMEIDA, R. P. Políticas públicas para a centralização de dados agrícolas: Um caminho para a sustentabilidade. **Revista Brasileira de Políticas Agrícolas**, v.15, n.4, p.89-97, 2020.

SMITH, J. R. Data Warehousing in Agriculture: Integration and Consolidation of Diverse Data Formats for Informed Decision Making. **Journal of Agricultural Informatics**, v. 13, n. 2, p. 78-91, 2021.

SMITH, T.; JOHNSON, R. Big Data in Agriculture: A Case Study of Precision Farming Techniques. **Journal of Precision Agriculture**, v. 14, n. 3, p. 251–266, 2019.

SOUZA, J. M.; ALMEIDA, R. T.; PEREIRA, L. F. Uso de tecnologias na gestão da qualidade de grãos na cadeia produtiva da soja. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 27, n. 5, p. 347-355, 2023.

SOUZA, M. C.; CARVALHO, A. P. **Técnicas de Revisão Bibliográfica e Operadores Lógicos**. São Paulo: Editora Acadêmica, 2020.

STANKOVSKI, S. e OSTOJIC, G. **Traceability In Agri-Food Production And Supply Chains**. 19th International Scientific Conference on Industrial Systems (2023).

SURESH, S., & SARAVANAN, V. A survey of batch loading techniques for big data. **Journal of Ambient Intelligence and Humanized Computing**, v. 12, p. 2321-2332, 2021.

TA'A, A. *et al.* Análise De Impacto Do Processo De Extração, Transformação E Carga Para Manutenção Do Sistema De Data Warehouse. **Journal of Information System and Technology Management**. 2022.

TÁVORA, F. L.; FRANÇA, F. F.; LIMA, J. R. P. A. **Impactos das Mudanças Climáticas na Agropecuária Brasileira, Riscos Políticos, Econômicos e Sociais e os Desafios para a Segurança Alimentar Humana**. Brasília. Núcleo de Estudos e Pesquisas/CONLEG/Senado, dezembro 2022 (Texto para Discussão nº313). 2022.

THE CENTER FOR CLIMATE AND SECURITY (TCCS). **A Security Threat Assessment of Global Climate Change: how likely warming scenarios indicate a catastrophic security future**. Washington, DC: 2020.

VALIN, H., *et al.* The role of data and technology in adapting agriculture to climate change. **Environmental Science & Policy**, v. 113, p. 31-39, 2020.

VASSILIADIS, P.; SELLIS, T. **Data Warehouse Systems: Design and Implementation**. 2. ed. Cham: Springer, 2020.

VIANNA, L. F. N. *et al.* Bancos de Dados Meteorológicos: Análise dos Metadados das Estações Meteorológicas no Estado de Santa Catarina, Brasil. **Revista Brasileira de Meteorologia**. 2017.

VIEIRA, J. R.; MELO, J. C. Práticas de Controle de Qualidade na Produção Agrícola: Impactos na Comercialização e na Confiança do Consumidor. **Revista Brasileira de Agroecologia**, v. 14, n. 1, p. 45-56, 2020.

VILLAS-BOAS, S. B.; *et al.* Business Intelligence aplicado à Agricultura: Eficiência, Sustentabilidade e Rastreabilidade na Cadeia Produtiva. **Journal of Agribusiness and Information Technology**, v. 15, n. 2, p. 152-170, 2021.

WANG, X. *et al.* Big data in agriculture: A review. **Journal of Cleaner Production**, v. 227, p. 876-889, 2019.

WMO. **Manual de Meteorologia: Observações Meteorológicas**. Genebra: Organização Meteorológica Mundial, 2021.

WU, X.; *et al.* Early Skin Cancer Detection Using ResNet50: A Clinical Study. **Medical Imaging and Diagnosis**, v. 48, n. 1, p. 53-59, 2023.

YU, C. *et al.* Challenges of Big Data Analysis in Agriculture: A Review. **Geospatial Information Science**, v. 21, n. 3, p. 202-210, 2018.

ZHANG, A. H. *et al.* Modelagem do impacto de fatores climatológicos e revolução tecnológica na produtividade da soja: evidências de 13 grandes províncias da China. **International Journal of Environmental Research and Public Health**. 2022.

ZHANG, L.; *et al.* Enhancing Image Classification with Residual Networks: A Review of ResNet50 and its Applications. **Pattern Recognition Letters**, v. 148, p. 45-53, 2021.

ZHAO, C. *et al.* Data volume and scale of IoT-based agricultural data: Cases in China. **Computers and Electronics in Agriculture**, v. 163, p. 104858, 2019.

ZHAO, Q.; *et al.* A Comparative Analysis of Vision Transformers and Convolutional Neural Networks. **IEEE Transactions on Neural Networks and Learning Systems**, v. 34, n. 2, p. 1183-1190, 2023.

ZHAO, Y.; CHEN, M.; WANG, J. Integrating Meteorological Data with Remote Sensing for Climate Analysis: A Review. **Remote Sensing**, v. 12, n. 22, p. 3742, 2020.