

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA  
DE MINAS GERAIS (IFMG)  
*CAMPUS BAMBUÍ*  
BACHARELADO EM ENGENHARIA DE COMPUTAÇÃO

Marçal Henrique Moreira

**DESENVOLVIMENTO DE UM SISTEMA PARA A LEITURA DE CÓDIGOS DE  
ETIQUETAS DE CASEÍNA EM QUEIJOS MINAS ARTESANAL UTILIZANDO  
TÉCNICAS DE VISÃO COMPUTACIONAL E INTELIGÊNCIA ARTIFICIAL**

BambuÍ – MG  
2025

MARÇAL HENRIQUE MOREIRA

**DESENVOLVIMENTO DE UM SISTEMA PARA A LEITURA DE CÓDIGOS DE ETIQUETAS DE CASEÍNA EM QUEIJOS MINAS ARTESANAL UTILIZANDO TÉCNICAS DE VISÃO COMPUTACIONAL E INTELIGÊNCIA ARTIFICIAL**

Trabalho de conclusão de curso apresentado ao Curso de Bacharelado em Engenharia de Computação do Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais (IFMG) – *Campus* Bambuí para obtenção do grau de Bacharel em Engenharia de Computação.

Orientador: Me. Calebe Giaculi Júnior

Coorientadora: Dra. Fabiana Aparecida Couto

---

**Catálogo na Fonte Biblioteca IFMG - *Campus Bambuí***

---

M838d Moreira, Marçal Henrique.

Desenvolvimento de um sistema para a leitura de códigos de caseína em queijos Minas Artesanal utilizando técnicas de visão computacional e inteligência artificial [manuscrito] / Marçal Henrique Moreira – 2025.

67 f. : il.

Possui anexos.

Orientador: Calebe Giaculi Júnior.

Coorientadora: Fabiana Aparecida Couto.

Trabalho de Conclusão de Curso (Bacharelado em Engenharia de Computação) – Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais. *Campus Bambuí*, 2025.

1. Visão computacional. 2. Redes neurais artificiais. 3. Queijos Canastra. 4. Etiquetas de caseína. 5. Reconhecimento óptico de caracteres. I. Giaculli Júnior, Calebe. II. Couto, Fabiana Aparecida. III. Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais – *Campus Bambuí*. IV. Título.

CDD 006.32

---

**Catálogo: João Batista Rodrigues - CRB-6/2022**

Marçal Henrique Moreira

**DESENVOLVIMENTO DE UM SISTEMA PARA A LEITURA DE CÓDIGOS DE ETIQUETAS DE CASEÍNA EM QUEIJOS MINAS ARTESANAL UTILIZANDO TÉCNICAS DE VISÃO COMPUTACIONAL E INTELIGÊNCIA ARTIFICIAL**

Trabalho de conclusão de curso apresentado ao Curso de Bacharelado em Engenharia de Computação do Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais (IFMG) – *Campus* Bambuí para obtenção do grau de Bacharel em Engenharia de Computação.

Aprovado em 17 de Junho de 2025 pela banca examinadora:

Me. Calebe Giaculi Júnior – IFMG – *Campus* Bambuí – (Orientador)

Dra. Fabiana Aparecida Couto – Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais – (Coorientadora)

Me. Francisco Heider Willy dos Santos – Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais

Dr. Gustavo Augusto Lacorte – Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais



Documento assinado eletronicamente por **Calebe Giaculi Junior, Professor**, em 17/06/2025, às 18:36, conforme Decreto nº 10.543, de 13 de novembro de 2020.



Documento assinado eletronicamente por **Fabiana Aparecida Couto, Professora**, em 17/06/2025, às 18:37, conforme Decreto nº 10.543, de 13 de novembro de 2020.



Documento assinado eletronicamente por **Francisco Heider Willy dos Santos, Professor**, em 17/06/2025, às 18:37, conforme Decreto nº 10.543, de 13 de novembro de 2020.



Documento assinado eletronicamente por **Gustavo Augusto Lacorte, Professor**, em 17/06/2025, às 18:37, conforme Decreto nº 10.543, de 13 de novembro de 2020.



A autenticidade do documento pode ser conferida no site <https://sei.ifmg.edu.br/consultadocs> informando o código verificador **2351322** e o código CRC **FA0B999A**.

---

A minha família e amigos.

## **AGRADECIMENTOS**

Agradeço primeiramente aos meus pais Raquel e Marçal e minha irmã Amanda que sempre estiveram presentes durante toda a minha trajetória acadêmica. Agradeço aos meus amigos, Paulo Vitor, Júlio César e Bruno Oliveira pelo companheirismo e amizade durante todos estes anos. Agradeço também ao Prof. Me. Calebe Giaculi Júnior e a Prof. Dra. Fabiana Aparecida Couto e a todos que contribuíram de alguma forma para realização deste trabalho.

“É hora de esquecer os erros do passado e começar a planejar os erros do futuro.”  
Agostinho Carrara

## RESUMO

Este trabalho explora o uso de técnicas de visão computacional para detecção e leitura de caracteres presentes em etiquetas de caseína aplicadas em Queijos Minas Artesanal, utilizando imagens como base. As etiquetas de caseína são elementos essenciais para o controle de qualidade e rastreabilidade na cadeia produtiva. O estudo propôs o desenvolvimento de um sistema composto por modelos de aprendizado de máquina para localizar automaticamente as etiquetas e reconhecer os caracteres nelas presentes. Foram utilizados conceitos de aprendizado profundo e redes neurais convolucionais, aplicados a um conjunto de dados de imagens de queijos Canastra. O modelo de detecção baseado na arquitetura YOLOv8 alcançou o desempenho médio de  $mAP@0.5$  igual a 91,9% nas tarefas de detecção. O sistema OCR baseado na PaddleOCR obteve uma acurácia média de 81,8% e CER médio de 18,2% demonstrando viabilidade técnica para automatizar o processo de registro. Os resultados indicam que a solução proposta contribui para eficiência e confiabilidade na certificação da produção artesanal na região da Serra da Canastra.

**Palavras-chave:** visão computacional, redes neurais artificiais, queijo canastra, etiquetas de caseína, reconhecimento óptico de caracteres

## ABSTRACT

*This work explores the use of computer vision techniques for the detection and reading of characters present on casein tags applied to Queijo Minas Artesanal, using images as input. Casein tags are essential elements for quality control and traceability within the production chain. The study proposed the development of a system composed of machine learning models capable of automatically locating the tags and recognizing the characters printed on them. Concepts of deep learning and convolutional neural networks were applied to a dataset of images of Canastra cheese. The detection model, based on the YOLOv8 architecture, achieved an average mAP@0.5 of 91.9% in the detection tasks. The OCR system, based on PaddleOCR, obtained an average accuracy of 81.8% and a mean Character Error Rate (CER) of 18.2%, demonstrating the technical feasibility of automating the registration process. The results indicate that the proposed solution contributes to the efficiency and reliability of the certification process of artisanal production in the Serra da Canastra region.*

**Keywords:** *computer vision, artificial neural networks, Canastra cheese, casein tags, optical character recognition*

## LISTA DE FIGURAS

Figura 1 - Arquitetura de um neurônio . . . . .	25
Figura 2 - Arquitetura de uma MLP . . . . .	25
Figura 3 - Fases CRISP-DM . . . . .	35
Figura 4 - Exemplo de imagem original e correspondente anotação no <i>CVAT</i> . .	41
Figura 5 - Exemplo de aplicação de brilho e contraste. . . . .	42
Figura 6 - Exemplo de aplicação da <i>CLAHE</i> . . . . .	43
Figura 7 - Exemplo de aplicação da <i>HSV</i> . . . . .	43
Figura 8 - Exemplo de aplicação do ruído gaussiano. . . . .	44
Figura 9 - Exemplo de aplicação do <i>motion blur</i> . . . . .	44
Figura 10 -Exemplo de aplicação do <i>median blur</i> . . . . .	45
Figura 11 -Exemplo de aplicação do <i>gaussian blur</i> . . . . .	46
Figura 12 -Exemplo de transformações geométricas aplicadas. . . . .	47
Figura 13 -Fluxograma de treino . . . . .	48
Figura 14 -Fluxograma de inferência . . . . .	50
Figura 15 - <i>Boxplot</i> da métrica <i>mAP@50</i> . . . . .	54
Figura 16 - <i>Boxplot</i> da métrica <i>Precision</i> . . . . .	54
Figura 17 -Resumo das métricas de treino . . . . .	55
Figura 18 -Resumo das métricas de treino . . . . .	56
Figura 19 -Curva de precisão . . . . .	57
Figura 20 -Curva de <i>Recall</i> . . . . .	58
Figura 21 -Exemplo de inferência. . . . .	59
Figura 22 -Processo de saída do <i>OCR</i> . . . . .	59

## LISTA DE TABELAS

Tabela 1 - Comparação dos trabalhos sobre visão computacional . . . . .	32
Tabela 2 - Resumo do conjunto de imagens coletadas . . . . .	40
Tabela 3 - Configuração dos parâmetros utilizados no treinamento do modelo YOLOv8n . . . . .	49
Tabela 4 - Métricas obtidas nas 10 execuções do modelo com <i>data augmentation</i> . . . . .	52
Tabela 5 - Estatísticas descritivas das métricas obtidas nas 10 execuções . .	53
Tabela 6 - Resultados da inferência do OCR para detecção das etiquetas de caseína . . . . .	57
Tabela 7 - Estatísticas descritivas dos resultados da inferência do OCR . . .	58
Tabela 8 - Métricas do modelo escolhido por época . . . . .	67

## LISTA DE SIGLAS

- INPI – Instituto Nacional da Propriedade Industrial
- FAO – Food and Agriculture Organization*
- PDI – Processamento Digital de Imagens
- CLAHE – Contrast Limited Adaptive Histogram Equalization*
- HSV – Hue, Saturation, Value*
- OCR – Optical Character Recognition*
- RNA – Rede Neural Artificial
- MLP – Multilayer Perceptron*
- CRISP-DM – Cross Industry Standard Process for Data Mining*
- CER – Character Error Rate]*
- IoU – Intersection over Union*
- GIoU – Generalized IoU*
- CIoU – Complete IoU*
- DFL – Distribution Focal Loss*
- mAP – Mean Average Precision*
- TPA – True Positive Accuracy*

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>15</b>
1.1	Objetivos Geral	15
1.2	Objetivos Específicos	16
1.3	Justificativa	16
1.4	Organização do Documento	17
<b>2</b>	<b>FUNDAMENTOS TEÓRICOS</b>	<b>18</b>
2.1	Queijo Minas Artesanal Canastra	18
2.2	Rastreabilidade e Etiqueta de Caseína	18
2.3	Visão Computacional	19
2.3.1	<i>Data augmentation em visão computacional</i>	20
2.3.2	<i>Arquitetura YOLO para detecção de objetos</i>	20
2.3.3	<i>Reconhecimento óptico de caracteres</i>	21
2.4	Processamento Digital de Imagens	21
2.4.1	<i>Equalização de histograma</i>	22
2.4.2	<i>Ajuste de brilho e contraste</i>	22
2.4.3	<i>Transformações no espaço de cores HSV</i>	22
2.4.4	<i>Adição de ruído gaussiano</i>	23
2.4.5	<i>Filtros de desfoque</i>	23
2.5	Aprendizado de Máquina	23
2.6	Aprendizado Profundo	24
2.7	Redes Neurais Artificiais	24
2.8	Treinamento de Redes Neurais Artificiais	24

<b>2.9</b>	<b>Métricas de Validação</b>	<b>26</b>
<b>2.9.1</b>	<b><i>Precisão</i></b>	<b>26</b>
<b>2.9.2</b>	<b><i>Recall</i></b>	<b>27</b>
<b>2.9.3</b>	<b><i>Box Loss</i></b>	<b>27</b>
2.9.3.1	Intersection over Union (IoU)	27
2.9.3.2	Generalized IoU (GIoU)	28
2.9.3.3	Complete IoU (CIoU)	28
<b>2.9.4</b>	<b><i>Distribution Focal Loss</i></b>	<b>29</b>
<b>2.9.5</b>	<b><i>mAP@0.5</i></b>	<b>29</b>
<b>2.9.6</b>	<b><i>mAP@0.5:0.95</i></b>	<b>30</b>
<b>2.9.7</b>	<b><i>Character Error Rate (CER)</i></b>	<b>31</b>
<b>2.10</b>	<b>Trabalhos Correlatos</b>	<b>32</b>
<b>3</b>	<b>METODOLOGIA</b>	<b>34</b>
<b>3.1</b>	<b>Classificação da Pesquisa</b>	<b>34</b>
<b>3.2</b>	<b>Metodologia de Desenvolvimento</b>	<b>34</b>
<b>3.3</b>	<b>Materiais e Tecnologias</b>	<b>36</b>
<b>3.4</b>	<b>Plano de execução</b>	<b>37</b>
<b>4</b>	<b>DESENVOLVIMENTO</b>	<b>39</b>
<b>4.1</b>	<b>Entendimento do Negócio</b>	<b>39</b>
<b>4.2</b>	<b>Entendimento dos Dados</b>	<b>39</b>
<b>4.3</b>	<b>Preparação dos Dados</b>	<b>40</b>
<b>4.3.1</b>	<b><i>Anotação das regiões de interesse</i></b>	<b>40</b>
<b>4.3.2</b>	<b><i>Padronização das imagens</i></b>	<b>40</b>
<b>4.3.3</b>	<b><i>Aumento de dados</i></b>	<b>41</b>

4.3.3.1	Brilho e contraste . . . . .	42
4.3.3.2	<i>CLAHE</i> . . . . .	42
4.3.3.3	<i>HSV</i> . . . . .	42
4.3.3.4	Ruído gaussiano . . . . .	43
4.3.3.5	<i>Motion blur</i> . . . . .	44
4.3.3.6	<i>Median blur</i> . . . . .	45
4.3.3.7	<i>Gaussian blur</i> . . . . .	45
4.3.3.8	Transformações geométricas aplicadas durante o treinamento . . . . .	45
<b>4.4</b>	<b>Modelagem . . . . .</b>	<b>47</b>
4.4.0.1	Treinamento . . . . .	47
4.4.0.2	Inferência . . . . .	49
<b>5</b>	<b>RESULTADOS . . . . .</b>	<b>52</b>
<b>5.1</b>	<b>Treinamentos . . . . .</b>	<b>52</b>
<b>5.2</b>	<b>Inferência . . . . .</b>	<b>56</b>
<b>6</b>	<b>TRABALHOS FUTUROS . . . . .</b>	<b>60</b>
<b>7</b>	<b>CONSIDERAÇÕES FINAIS . . . . .</b>	<b>61</b>
	<b>REFERÊNCIAS . . . . .</b>	<b>62</b>
	<b>ANEXOS . . . . .</b>	<b>67</b>

## 1 INTRODUÇÃO

O queijo é um alimento presente em diversas culturas, com variações de técnicas e ingredientes que refletem as características regionais de produção (NARDY; CARVALHO; ROCHA, 2019).

No Brasil, a indústria de laticínios ocupa uma das principais posições no setor alimentício, sendo os queijos artesanais parte importante da economia de regiões rurais (GUIMARÃES *et al.*, 2011).

Entre os queijos artesanais brasileiros destaca-se o Queijo Minas Artesanal Canastra, produzido na região da Serra da Canastra, em Minas Gerais. Essa produção é tradicionalmente associada a práticas herdadas entre gerações, que agregam identidade cultural e relevância econômica à região (JUNIOR *et al.*, 2009).

Para atender às exigências legais e garantir a rastreabilidade, a produção do Queijo Canastra utiliza etiquetas de caseína. Essas etiquetas são aplicadas diretamente no produto, contendo um código alfanumérico único que permite a identificação de origem e a conformidade com normas sanitárias (REVISTA MENU, 2019).

A rastreabilidade está prevista na legislação brasileira, que estabelece diretrizes para os Procedimentos Operacionais Padronizados e Boas Práticas de Fabricação, exigindo sistemas capazes de monitorar os produtos ao longo da cadeia produtiva (AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA, 2002).

No entanto, o processo atual de anotação das informações contidas nas etiquetas de caseína é manual, sujeito a falhas, perda de dados e atrasos. As informações são registradas fisicamente pelos produtores e digitalizadas posteriormente pelas associações, em um processo que pode levar mais de 15 dias até estarem disponíveis para consulta online.

Este trabalho propõe o desenvolvimento de um *pipeline* automatizado para detecção de etiquetas de caseína e leitura de caracteres em imagens, utilizando técnicas de visão computacional e aprendizado de máquina. A proposta busca avaliar a viabilidade técnica da automação desse processo no contexto da rastreabilidade do Queijo Minas Artesanal Canastra, contribuindo para a redução de falhas operacionais e otimização do controle de produção.

### 1.1 Objetivos Geral

Desenvolver um *pipeline* automatizado para a detecção de etiquetas de caseína e a leitura dos caracteres presentes em imagens de Queijos Minas Artesanal Canastra, utilizando técnicas de visão computacional e aprendizado de máquina.

## 1.2 Objetivos Específicos

Os objetivos específicos foram estabelecidos como:

- Construir um conjunto de dados de imagens contendo registros de queijos Canastra com etiquetas de caseína visíveis.
- Realizar a anotação das regiões de interesse nas imagens, marcando as etiquetas para uso no treinamento do modelo.
- Aplicar técnicas de processamento digital de imagens para aumentar e melhorar o conjunto de dados, bem como otimizar a qualidade das imagens utilizadas no *pipeline* de detecção.
- Treinar um modelo de detecção automática de etiquetas de caseína.
- Avaliar a acurácia e desempenho do modelo desenvolvido, por meio de métricas adequadas de avaliação.
- Validar a capacidade de leitura automática dos caracteres detectados, assegurando a efetividade do *pipeline* na extração das informações presentes nas etiquetas.

## 1.3 Justificativa

A produção do Queijo Minas Artesanal Canastra é uma atividade tradicional da região da Serra da Canastra e está sujeita a regulamentações específicas que exigem mecanismos de identificação e rastreabilidade dos produtos comercializados.

As etiquetas de caseína cumprem o papel de identificar cada unidade produzida, sendo exigidas por órgãos de fiscalização sanitária. Atualmente, o processo de anotação dos dados dessas etiquetas é manual, o que pode resultar em inconsistências, perdas de informação e dificuldades na rastreabilidade.

A proposta de um *pipeline* automatizado para detecção de etiquetas e leitura de caracteres atende à necessidade de sistematização desse processo. No campo tecnológico, o uso de técnicas de visão computacional e inteligência artificial possibilitam a automatização de uma etapa crítica da cadeia produtiva. Do ponto de vista científico, a aplicação prática dessas técnicas em um contexto específico contribui para a avaliação de métodos computacionais em problemas reais, permitindo replicação e estudos futuros.

Em termos econômicos, a redução de erros operacionais e o ganho de eficiência podem gerar economia de tempo e de recursos no controle de produção. No aspecto social, a padronização do processo de identificação contribui para o fortalecimento da rastreabilidade, atendimento às exigências legais e valorização do produto perante o consumidor, com impacto direto na comercialização e na sustentabilidade da atividade local.

## 1.4 Organização do Documento

Este documento está organizado em quatro seções, incluindo este. Na seção 2 são expostos os fundamentos necessários para a compreensão deste trabalho, além dos trabalhos correlatos a este. Na seção 3 é descrita a classificação da pesquisa, metodologia de desenvolvimento, materiais e tecnologias, plano e execução e métricas de validação. Na seção 4 encontra-se a descrição detalhada do desenvolvimento do sistema proposto.

Na seção 5 são apresentados e analisados os resultados obtidos com a aplicação do sistema, demonstrando o desempenho nas tarefas de detecção e reconhecimento, bem como a avaliação das métricas utilizadas para a validação.

Na seção 7 são apresentadas as considerações finais deste trabalho, destacando as contribuições alcançadas, as limitações identificadas. Por fim na seção 6 contém sugestões para trabalhos futuros que possam aprimorar ou expandir a abordagem proposta.

## 2 FUNDAMENTOS TEÓRICOS

A seção atual apresenta os fundamentos teóricos necessários para o entendimento do presente trabalho. As seções 2.1 e 2.2 abordam os conceitos relacionados a queijos e rastreabilidade de alimentos. A seção 2.3 trata os conceitos relacionados à visão computacional. Em seguida, a seção 2.4 trata do processamento digital de imagens. A seção 2.5 discute os principais conceitos de aprendizado de máquina. A seção 2.7 discute redes neurais artificiais. A seção 2.8 aborda conceitos para treinamentos de uma rede neural artificial.

Por fim a seção 2.10 apresenta os trabalhos correlatos, destacando pesquisas que possuem relação com o presente estudo.

### 2.1 Queijo Minas Artesanal Canastra

De acordo com o Ministério da Agricultura, Pecuária e Abastecimento - MAPA (1996), entende-se por queijo o produto fresco ou maturado obtido por separação parcial do soro do leite, coagulados por meio de processos físicos, enzimáticos ou bacterianos, podendo conter aditivos, especiarias e corantes, desde que aptos ao consumo humano.

O Queijo Minas Artesanal, em especial o Queijo Canastra, produzido na região da Serra da Canastra, no sudoeste de Minas Gerais, é um dos principais representantes desse modo de produção. Reconhecido nacional e internacionalmente, o Queijo Canastra é produzido a partir de práticas artesanais que se perpetuam há gerações, utilizando leite cru e técnicas tradicionais de fabricação, o que confere ao produtor uma identidade cultural e um valor econômico significativo para a região (JUNIOR *et al.*, 2009).

A importância do Queijo Canastra extrapola o aspecto econômico, compondo o patrimônio cultural e gastronômico brasileiro. O produto possui Indicação de Procedência, certificação concedida em 2012 pelo Instituto Nacional da Propriedade Industrial (INPI), que assegura sua procedência e o método tradicional de produção, além de reforçar sua identidade no mercado (INSTITUTO NACIONAL DA PROPRIEDADE INDUSTRIAL, 2012).

### 2.2 Rastreabilidade e Etiqueta de Caseína

A rastreabilidade na cadeia produtiva de alimentos é um elemento fundamental para assegurar a qualidade, a segurança alimentar e a confiança dos consumidores. De acordo com a Organização das Nações Unidas para Alimentação e Agricultura (FAO do inglês *Food and Agriculture Organization*), a rastreabilidade permite

identificar e monitorar a trajetória de um produto desde a origem até o consumidor final, sendo essencial para ações corretivas em casos de não conformidade e para valorização de produtos de origem controlada (FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS, 2019).

No Brasil, a legislação sanitária estabelece diretrizes claras para a rastreabilidade de alimentos. AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA (2002), define os Procedimentos Operacionais Padronizados e Boas Práticas de Fabricação como instrumentos obrigatórios para garantir a qualidade sanitária dos alimentos, reforçando a necessidade de sistemas que permitam a identificação e o monitoramento dos produtos ao longo de toda cadeia produtiva.

No caso do Queijo Minas Artesanal Canastra, a etiqueta de caseína constitui um mecanismo para identificação formal e rastreabilidade do produto. Essa etiqueta, feita de uma proteína láctea natural, é aplicada diretamente sobre o queijo no momento da produção, contendo um código alfanumérico único que possibilita rastrear o produto até o seu produtor, assegurando sua autenticidade e procedência (REVISTA MENU, 2019).

O uso da etiqueta de caseína no Queijo Canastra é parte do caderno de normas de associações com o objetivo de coibir a falsificação e fortalecer os mecanismos de certificação de origem, especialmente após o reconhecimento da Indicação de Procedência concedida pelo INPI (INSTITUTO NACIONAL DA PROPRIEDADE INDUSTRIAL, 2012).

Além de assegurar a rastreabilidade, a etiqueta também é um instrumento de comunicação com os consumidores, que podem verificar informações sobre a origem e as condições de produção do queijo. Esse mecanismo está alinhado com as tendências internacionais de fortalecimento dos sistemas de rastreabilidade como estratégia de promoção da qualidade e segurança dos alimentos (FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS, 2019).

### **2.3 Visão Computacional**

Visão computacional (*computer vision*) é a área responsável por possibilitar que sistemas computacionais interpretem e processem informações extraídas de imagens ou vídeos. Essa capacidade permite que os sistemas reconheçam, manipulem e analisem objetos presentes em imagens (BALLARD; BROWN, 1982).

Essa área de pesquisa abrange desde registros de dados brutos até extração de padrões de imagem e interpretação de informações visuais (PATEL *et al.*, 2012), combinando técnicas e conceitos de processamento digital de imagens, reconhecimento de padrões, inteligência artificial e gráficos computacionais (COSIDO *et al.*, 2014).

O objetivo da visão computacional é criar modelos e extrair informações de imagens, enquanto o processamento digital de imagens se concentra na manipulação de propriedades visuais, como nitidez, contraste e brilho (BABATUNDE *et al.*, 2015).

Visão computacional é composta por uma combinação de técnicas que permitem o reconhecimento de padrões em imagens, resultando na extração de informações para tomadas de decisões (LONG; SHELHAMER; DARRELL, 2015).

O reconhecimento de padrões, como um ramo da visão computacional, aborda o processo de identificação de objetos por meio de transformações de imagem. Este processo é baseado em técnicas de aprendizado de máquina, aprendizado profundo, redes neurais artificiais, entre outras (BISHOP; NASRABADI, 2006).

Entre as técnicas empregadas em visão computacional, destacam-se a detecção de objetos e o reconhecimento óptico de caracteres (*OCR*, do inglês *Optical Character Recognition*). A detecção de objetos visa localizar e identificar instâncias específicas em imagens (REDMON *et al.*, 2016). Enquanto o *OCR* busca a extração automática de textos presentes em imagens (SMITH, 2007).

Ambas as técnicas são essenciais para o presente trabalho, sendo utilizadas na identificação das etiquetas de caseína aplicadas ao Queijo Canastra e na leitura automatizada dos caracteres nelas contidos.

### **2.3.1 Data augmentation em visão computacional**

Em aplicações de visão computacional, especialmente em tarefas de detecção de objetos, é comum a utilização de técnicas de *Data Augmentation*. Essas técnicas consistem na aplicação de transformações sistemáticas ou aleatórias nas imagens de treinamento, com o objetivo de ampliar a diversidade do conjunto de dados e aumentar a robustez do modelo frente a variações naturais, como iluminação e orientação (SHORTEN; KHOSHGOFTAAR, 2019).

### **2.3.2 Arquitetura YOLO para detecção de objetos**

Entre as diversas arquiteturas de detecção de objetos, destaca-se a abordagem *You Only Look Once (YOLO)* proposta por Redmon *et al.* (2016). A principal característica do *YOLO* é tratar detecção como um problema de regressão direta, dividindo a imagem em uma grade e para cada célula predizendo simultaneamente a presença ou ausência de um objeto, as coordenadas da caixa delimitadora, a confiança da detecção e a classe do objeto detectado.

Essa abordagem possibilita a realização de detecção de objetos em tempo real.

### **2.3.3 Reconhecimento óptico de caracteres**

O Reconhecimento Óptico de Caracteres (*OCR*) é uma técnica que visa a extração automática de textos presentes em imagens ou documentos digitalizados. Essa técnica consiste em identificar e converter símbolos gráficos contidos em uma imagem em dados textuais editáveis (SMITH, 2007).

O processo de *OCR* envolve etapas como o pré-processamento da imagem, a segmentação para a localização de regiões com texto, o reconhecimento de padrões e, eventualmente a correção de erros por meio de modelos linguísticos (SMITH, 2007).

Neste presente trabalho, o *OCR* foi empregado para realizar a leitura automatizada dos códigos alfanuméricos presentes nas etiquetas de caseína aplicadas ao Queijo Canastra.

## **2.4 Processamento Digital de Imagens**

O Processamento Digital de Imagens (PDI) é uma área da ciência da computação e da engenharia que se dedica ao desenvolvimento de métodos para manipulação e transformação de imagens no formato digital. Esses métodos permitem a melhoria da qualidade visual, a extração de informações relevantes e a preparação das imagens para outras tarefas computacionais, como a análise automatizada ou modelagem (GONZALEZ; WOODS, 2008).

Segundo Castro e Madeo (2020) as operações realizadas no PDI podem ser classificadas em três níveis:

- Pré-processamento: envolve técnicas destinadas a melhorar a qualidade da imagem, como remoção de ruído, filtragem e correção de iluminação.
- Processamento de nível médio: inclui a segmentação de regiões de interesse, extração de bordas e detecção de formas, preparando a imagem para análises mais complexas.
- Pós-processamento: refere-se à aplicação de transformações para realce ou visualização podendo também incluir compressão ou codificação de imagens.

Em aplicações baseadas em aprendizado de máquina, como modelos de detecção de objetos, o PDI permite a normalização das imagens, padronização de tamanhos e a aplicação de técnicas de *data augmentation*, que visam ampliar a diversidade dos dados de treinamento e aumentar a robustez do modelo (SHORTEN; KHOSHGOFTAAR, 2019).

O PDI é considerado uma etapa necessária na maioria dos sistemas que envolvem a análise automática de imagens, contribuindo para eficiência, precisão e generalização dos modelos desenvolvidos.

Dentre as técnicas de pré-processamento e aumento de dados aplicadas

neste trabalho, destacam-se aquelas relacionadas ao ajuste de brilho e contraste, transformações no espaço de cores, equalização de histograma, adição de ruído e aplicação de filtros de desfoque. Na sequência, são apresentadas as principais técnicas utilizadas, com foco em seus fundamentos e aplicações em visão computacional.

#### **2.4.1 Equalização de histograma**

A Equalização de Histograma é uma técnica utilizada para melhorar o contraste global ou local, redistribuindo os níveis de intensidade. A versão adaptativa conhecida como *Contrast Limited Adaptive Histogram Equalization (CLAHE)* aplica a equalização em pequenas regiões (ou *tiles*) da imagem, permitindo o realce de detalhes em áreas específicas sem amplificar excessivamente o ruído (ZUIDERVELD, 1994).

O *CLAHE* limita o contraste através de um parâmetro de *clipping* (limite superior que restringe a amplificação do contraste para evitar realce excessivo de ruído), evitando que áreas homogêneas sejam excessivamente realçadas (ZUIDERVELD, 1994).

#### **2.4.2 Ajuste de brilho e contraste**

O ajuste de brilho consiste em alterar a intensidade luminosa global da imagem, enquanto o ajuste de contraste modifica a diferença relativa entre tons claros e escuros (SHORTEN; KHOSHGOFTAAR, 2019).

Essas transformações são utilizadas para simular diferentes condições de iluminação, como exposição excessiva, sombras ou reflexos, comuns em ambientes não controlados.

#### **2.4.3 Transformações no espaço de cores HSV**

O espaço de cores *HSV (Hue, Saturation, Value)* separa as informações de cor em três componentes: *Hue* (matiz) define a tonalidade dominante da cor, *Saturation* (saturação) determina a intensidade da cor, *Value* (valor) representa o brilho (GONZALEZ; WOODS, 2008).

Transformações nesses componentes permitem simular variações de iluminação e mudanças cromáticas que podem ocorrer devido ao tipo de fonte luminosa, configurações de câmera ou ambiente de captura, aumentando a diversidade cromática do conjunto de dados (GONZALEZ; WOODS, 2008).

#### **2.4.4 Adição de ruído gaussiano**

O ruído gaussiano é um tipo de ruído estatístico que segue uma distribuição normal, caracterizado por pequenas variações aleatórias de intensidade nos *pixels* (GOODFELLOW; BENGIO; COURVILLE, 2016).

Esse ruído é frequentemente introduzido por sensores eletrônicos durante a captura de imagens em condições de baixa iluminação ou devido às interferências no ambiente.

#### **2.4.5 Filtros de desfoque**

Filtros de desfoque são técnicas que suavizam a imagem, reduzindo a presença de ruído de alta frequência e simulando imperfeições óticas. Neste trabalho, foram utilizados três tipos principais: *Motion Blur*, *Median Blur* e *Gaussian Blur*.

*Motion Blur* simula o desfoque causado pelo movimento relativo entre a câmera e o objeto durante a captura, resultando em um aspecto alongado ou borrado em determinada direção. *Median Blur* substitui o valor de cada *pixel* pela mediana dos *pixels* vizinhos, preservando as bordas e removendo ruídos isolados. *Gaussian Blur* aplica uma suavização baseada na função gaussiana, resultando em um desfoque suave e natural, que simula imprecisões de foco (JAIN, 1989).

### **2.5 Aprendizado de Máquina**

Aprendizado de máquina (*machine learning*) é um subcampo da inteligência artificial que se concentra no desenvolvimento de algoritmos e modelos que permitem que os computadores aprendam a partir de dados e tomem decisões com base nesse aprendizado (MITCHELL, 1997).

O conceito de aprendizado de máquina é implementado por meio de três paradigmas principais: supervisionado, não supervisionado e por reforço. Esses paradigmas são utilizados na construção de modelos preditivos, classificação de dados, agrupamento e outras aplicações (ALPAYDIN, 2020).

BISHOP e NASRABADI (2006) define o aprendizado de máquina supervisionado como um processo onde um algoritmo é treinado usando um conjunto de dados rotulados. Esses dados consistem em pares de entrada e saída. Onde o objetivo é aprender uma função que mapeia as entradas para as saídas corretas.

Hastie *et al.* (2009) define o aprendizado de máquina não supervisionado como uma abordagem que lida com dados não rotulados. Isso significa que, diferente do aprendizado supervisionado, não há saídas esperadas para o modelo aprender. O objetivo é encontrar padrões e estruturas nos dados, como agrupamentos ou associ-

ações.

Sutton e Barto (2018) define o aprendizado de máquina por reforço onde um agente interage com um ambiente para aprender a maximizar uma recompensa. Através de interações sucessivas com o ambiente, o agente toma ações em diferentes estados e recebe recompensas com base nessas ações.

## 2.6 Aprendizado Profundo

Aprendizado profundo (*deep learning*) é uma subárea do aprendizado de máquina que se concentra no desenvolvimento de redes neurais artificiais com múltiplas camadas (GOODFELLOW; BENGIO; COURVILLE, 2016).

Essas redes são capazes de aprender representações em diferentes níveis de abstração, permitindo a captura de padrões complexos presentes nos dados. Tal característica torna o aprendizado profundo útil em tarefas como reconhecimentos de imagens, processamento de linguagem natural, entre outras (AGGARWAL, 2018).

O aprendizado profundo se concentra no desenvolvimento de redes neurais artificiais, e tal assunto é abordado na seção 2.7.

## 2.7 Redes Neurais Artificiais

Uma rede neural artificial (RNA do inglês *Artificial Neural Network*) é definida como um modelo de computação inspirado na estrutura e função do cérebro humano. Especificamente, uma RNA é composta por unidades chamadas neurônio artificiais. Neurônio artificial ou *Perceptron* é descrito como uma unidade simples e interconectada que faz parte de um sistema representando um mapeamento não-linear entre um vetor de entrada e um vetor de saída. Esses Neurônios são conectados por pesos, e os sinais de saída são determinados pela soma das entradas para o neurônio, modificada por uma função de transferência não-linear (também conhecida como função de ativação). O efeito coletivo permite que a rede aproxime funções não lineares (GARDNER; DORLING, 1998).

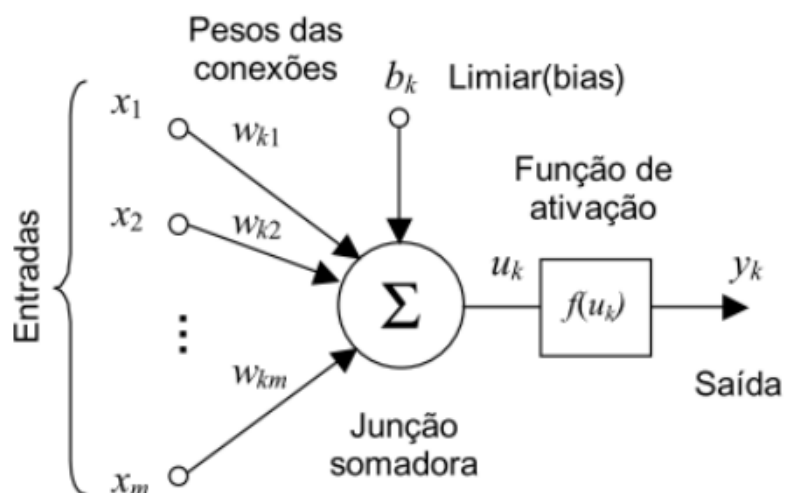
A Figura 1 apresenta a arquitetura de um neurônio.

Um perceptron multicamadas (*MLP* do inglês, *Multilayer Perceptron*) é um tipo de rede neural artificial composto por um sistema de neurônios interconectados. A Figura 2 apresenta a arquitetura de uma rede MLP.

## 2.8 Treinamento de Redes Neurais Artificiais

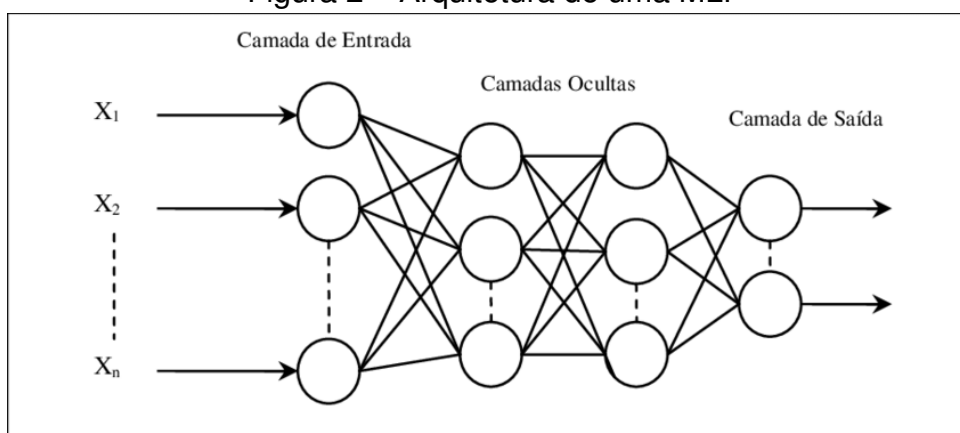
O treinamento de redes neurais artificiais consiste em ajustar os pesos sinápticos por meio de algoritmos de otimização, a fim de minimizar a função de erro

Figura 1 – Arquitetura de um neurônio



Fonte: DEUS, 2019.

Figura 2 – Arquitetura de uma MLP



Fonte: SOBREIRO *et al.*, 2008.

que mede a diferença entre as saídas previstas e os valores reais (HAYKIN, 2009).

Durante esse processo, diversos hiperpâmetros precisam ser definidos. Esses parâmetros influenciam diretamente na eficiência, estabilidade e capacidade de generalização do modelo. Entre os principais, destacam-se: épocas, *batch size*, taxa de aprendizado (*learning rate*), otimização e critério de parada.

Épocas representam o número de vezes que o modelo percorre todo o conjunto de dados de treinamento. Cada época permite que os pesos da rede sejam ajustados com base no erro acumulado (GOODFELLOW; BENGIO; COURVILLE, 2016).

*Batch size* define quantas amostras são processadas antes de uma atualização dos pesos. Um *batch* pequeno tende a gerar atualizações mais ruidosas, porém mais frequentes, enquanto *batches* maiores estabilizam o gradiente, mas exigem maior capacidade computacional (GÉRON, 2019).

*Learning rate* determina o passo dado em cada atualização dos pesos. Ta-

xas muito altas podem fazer o modelo divergir, enquanto taxas muito baixas podem tornar o processo extremamente lento ou fazer o modelo ficar preso em mínimos locais (GOODFELLOW; BENGIO; COURVILLE, 2016).

Otimização refere-se ao algoritmo utilizado para ajustar os pesos com base nos gradientes. Os otimizadores mais comuns incluem o *Stochastic Gradient Descent (SGD)*, que realiza atualizações a partir de mini-lotes, e o *Adam*, que adapta individualmente a taxa de aprendizado de cada parâmetro com base no momento dos gradientes (KINGMA; BA, 2015).

Critério de parada define quando o processo de treinamento deve ser interrompido. Pode-se utilizar um número fixo de épocas ou técnicas como *early stopping*, que interrompe o treinamento quando o desempenho de um conjunto de validação deixa de melhorar, evitando o *overfitting* (GOODFELLOW; BENGIO; COURVILLE, 2016).

## 2.9 Métricas de Validação

Na etapa de modelagem do projeto, o desempenho do *pipeline* foi avaliado utilizando métricas específicas para tarefas de detecção de objetos e *OCR*.

Para a detecção de objetos, foram utilizadas métricas que permitem quantificar o desempenho do modelo na localização e classificação das etiquetas de caixa.

Para a leitura dos caracteres, foram adotadas métricas de *OCR*, que avaliam a precisão da extração automática do texto impresso nas etiquetas.

### 2.9.1 Precisão

POWERS (2020) define precisão (do inglês *Precision*) como a proporção de casos positivos previstos que são positivos. Em outras palavras, é uma medida de quão confiável são as previsões positivas feitas pelo modelo. Também descrevendo precisão como TPA (do inglês *True Positive Accuracy*) que pode ser interpretado como acurácia dos positivos previstos:

$$Precision = \frac{tp}{tp + fp} \quad (2.1)$$

Onde:

- *tp*: Positivos verdadeiros, ou seja, casos em que o modelo previu corretamente a classe positiva.
- *fp*: Falsos positivos, ou seja, os casos em que o modelo previu a classe positiva incorretamente.

POWERS (2020) também destaca que a precisão é uma métrica que se concentra nos exemplos positivos previstos, avaliando a taxa de sucesso dessas previsões.

### 2.9.2 Recall

POWERS (2020) define *Recall* (ou Sensibilidade) como a proporção de casos verdadeiramente positivos que foram corretamente identificados pelo modelo. Em outras palavras, o *Recall* mede a cobertura dos casos positivos reais pela regra de previsão positiva.

$$Recall = \frac{tp}{tp + fn} \quad (2.2)$$

Onde:

- $tp$ : Verdadeiros positivos, ou seja, número de casos reais positivos que foram corretamente previstos como positivos.
- $fn$ : Falsos negativos, ou seja, casos realmente positivos que foram incorretamente previstos como negativos pelo modelo.

### 2.9.3 Box Loss

Em modelos de detecção de objetos, a regressão precisa das coordenadas das *bounding boxes* que são guiadas por funções de perda baseadas em métricas de sobreposição espacial. Tradicionalmente, a *Intersection over Union (IoU)* é a métrica fundamental, mas suas limitações levaram ao desenvolvimento de variações como *GIoU* e *CIoU* (REDMON *et al.*, 2016; ZHAO *et al.*, 2019).

#### 2.9.3.1 Intersection over Union (IoU)

A *IoU* mede a similaridade entre a caixa predita ( $B_p$ ) e a caixa real ( $B_r$ ) através da razão entre a área de interseção e a área de união:

$$IoU = \frac{|B_p \cap B_r|}{|B_p \cup B_r|}, \quad (2.3)$$

onde:

- $|B_p \cap B_r|$ : Área de interseção entre as caixas.
- $|B_p \cup B_r|$ : Área total coberta pelas duas caixas.

A perda associada ( $L_{IoU} = 1 - IoU$ ) varia de 0 (caixas idênticas) a 1 (sem sobreposição). Contudo, a *IoU* falha em fornecer gradientes significativos quando  $B_p$  e  $B_r$  não se intersectam ( $IoU = 0$ ), prejudicando o treinamento (REZATOFIGHI *et al.*,

2019).

### 2.9.3.2 Generalized IoU (GloU)

Para resolver essa limitação, Rezatofighi *et al.* (2019) propôs a *Generalized IoU (GloU)*, que introduz um termo de penalização baseado na área da envoltória ( $C$ ), definida como a menor caixa que contém simultaneamente  $B_p$  e  $B_r$ :

$$GIoU = IoU - \frac{|C \setminus (B_p \cup B_r)|}{|C|}, \quad (2.4)$$

onde  $C \setminus (B_p \cup B_r)$  denota a área da envoltória não coberta pelas caixas. A *GloU* varia no intervalo  $[-1, 1]$ , sendo que:

- $GIoU = 1$ : Indica sobreposição perfeita ( $B_p = B_r$ );
- $GIoU \leq 0$ : Ocorre quando não há interseção, com valores mais negativos indicando maior disparidade espacial.

A função de perda associada é:

$$L_{GIoU} = 1 - GIoU. \quad (2.5)$$

Conforme demonstrado por Rezatofighi *et al.* (2019), a *GloU* garante gradientes não nulos mesmo sem sobreposição, permitindo ajustes contínuos durante o treinamento.

### 2.9.3.3 Complete IoU (CloU)

A *CloU* estende a métrica ao incorporar a distância entre centros ( $\rho$ ) e a consistência de proporções ( $v$ ), sendo mais adequada para objetos com alta variação de *aspect ratio* (ZHENG, Z. *et al.*, 2020):

$$L_{CIoU} = 1 - IoU + \frac{\rho^2(\mathbf{b}_p, \mathbf{b}_r)}{c^2} + \alpha v, \quad (2.6)$$

onde:

- $\rho(\mathbf{b}_p, \mathbf{b}_r)$ : Distância Euclidiana entre os centros das caixas.
- $c$ : Comprimento da diagonal da envoltória  $C$ .
- $v = \frac{4}{\pi^2} \left( \arctan\left(\frac{w_r}{h_r}\right) - \arctan\left(\frac{w_p}{h_p}\right) \right)^2$ : Diferença nas proporções.
- $\alpha = \frac{v}{(1-IoU)+v}$ : Peso adaptativo.

Seleção da Função de Perda: Enquanto a *GloU* resolve o problema dos gradientes nulos, a *CloU* é preferível em cenários que demandam precisão geométrica (ex.: objetos alongados ou pequenos), como demonstrado por Zhaohui Zheng *et al.*

(2020) em detecção de veículos e pedestres.

### 2.9.4 Distribution Focal Loss

A *Distribution Focal Loss (DFL)*, proposta por LI *et al.* (2020), reformula a regressão de coordenadas como uma tarefa de classificação sobre *bins* discretos, onde a perda focaliza nos intervalos mais próximos do valor real. Essa abordagem é particularmente eficaz para detecção de objetos pequenos, pois suaviza gradientes e reduz a sensibilidade a *outliers* (YANG, S. *et al.*, 2021).

Formalmente, a *DFL* minimiza a divergência entre a distribuição prevista  $\hat{p}_i$  e os pesos  $w_i$  calculados com base na proximidade do valor real  $y$  ao centro do *bin*  $c_i$ :

$$w_i = \max\left(0, 1 - \frac{|y - c_i|}{\Delta}\right), \quad (2.7)$$

onde  $\Delta$  é a largura do bin. A perda é então definida como:

$$L_{DFL} = - \sum_{i=1}^N w_i \log(\hat{p}_i). \quad (2.8)$$

Exemplo: Se  $y = 10.3$  e os bins estão centrados em  $[10.0, 10.5, 11.0]$ , os pesos serão  $w_i \approx [0.7, 0.3, 0.0]$ , direcionando o modelo a ajustar as probabilidades  $\hat{p}_i$  nos *bins* relevantes.

Comparada a perdas tradicionais como o *MSE*, a *DFL* oferece maior robustez em cenários com alta variabilidade espacial (LIN *et al.*, 2017).

### 2.9.5 mAP@0.5

O *Mean Average Precision (mAP)* é a métrica padrão para avaliação de modelos de detecção de objetos, sintetizando em um único valor a capacidade do modelo em localizar (regressão de *bounding boxes*) e classificar objetos corretamente (ZHAO *et al.*, 2019; EVERINGHAM *et al.*, 2010).

A métrica é calculada em três etapas:

- a) Curva Precisão-Recall (PR): para cada classe, calcula-se a precisão ( $P$ ) e *Recall* ( $R$ ) em múltiplos limiares de confiança.
- b) *Average Precision (AP)*: a área sob a curva PR, aproximada numericamente via

$$AP \approx \sum_{k=1}^N P(k) \cdot \Delta R(k) \quad (2.9)$$

onde  $P(k)$  e  $\Delta R(k)$  são a precisão e a variação de *Recall* no  $k$ -ésimo limiar.

c) *mAP*: Média das *APs* de todas as classes.

A variante *mAP@0.5* (ou *mAP@50*) considera um limiar fixo de  $IoU \geq 0.5$  para determinar predições corretas:

$$TP = \begin{cases} 1, & \text{se } IoU(B_p, B_r) \geq 0.5 \\ 0, & \text{caso contrário} \end{cases} \quad (2.10)$$

Interpretação da *mAP@0.5* é menos rigorosa que *mAP@0.5:0.95* (média de *IoUs* de 0.5 a 0.95 em degraus de 0.05), sendo adequada para aplicações onde a localização aproximada é suficiente (LIN *et al.*, 2014).

Exemplo: se um modelo tem  $AP = 0.8$  para "carro" e  $AP = 0.6$  para "pedestre", a *mAP@0.5* é  $(0.8 + 0.6)/2 = 0.7$ .

### 2.9.6 *mAP@0.5:0.95*

A *mean Average Precision at IoU thresholds from 0.5 to 0.95* (*mAP@0.5:0.95*) é a métrica padrão em *benchmarks* modernos como *COCO* (LIN *et al.*, 2014), avaliando simultaneamente a precisão de classificação e a qualidade de objetos em múltiplos limiares de *IoU*.

Diferente da *mAP@0.5*, que usa limiar fixo, a (*mAP@0.5:0.95*) calcula a média das *AP* em 10 limiares de *IoU* igualmente espaçados (0.50, 0.55, ..., 0.95):

$$mAP@0.5:0.95 = \frac{1}{10} \sum_{t \in \{0.5, 0.55, \dots, 0.95\}} AP_t \quad (2.11)$$

onde:

-  $AP_t$ : *Average Precision* calculada no limiar de *IoU*  $t$ .

Cada  $AP_t$  é derivada da curva *PR* sob condições específicas:

a) Contagem de Verdadeiro Positivos:

$$TP_t = \begin{cases} 1, & \text{se } IoU(B_p, B_r) \geq t \text{ e classe correta} \\ 0, & \text{caso contrário} \end{cases} \quad (2.12)$$

b) Interpolação da Curva *PR*: a precisão  $P_t(r)$  é interpolada para 101 pontos de *Recall*  $r \in \{0, 0.01, \dots, 1.0\}$ :

$$P_{\text{interp}}(r) = \max_{\tilde{r} \geq r} P(\tilde{r}) \quad (2.13)$$

c) Integração Numérica:

$$AP_t = \frac{1}{101} \sum_{r \in \{0, 0.01, \dots, 1.0\}} P_{\text{interp}}(r) \quad (2.14)$$

### 2.9.7 Character Error Rate (CER)

A *Character Error Rate (CER)* é a métrica padrão para avaliação de sistemas de Reconhecimento Óptico de Caracteres (*OCR*), quantificando a discrepância entre o texto previsto e o texto de referência (SMITH, 2007). É definida como:

$$CER = \frac{S + D + I}{N} \times 100\% \quad (2.15)$$

onde:

- *S*: número de substituições de caracteres (caracteres incorretos)
- *D*: número de deleções (caracteres omitidos)
- *I*: número de inserções (caracteres adicionais)
- *N*: número total de caracteres no texto de referência

O cálculo da *CER* utiliza o algoritmo de *Levenshtein Distance* (distância de edição mínima):

- a) Alinhamento ótimo entre sequências prevista e referência
  - b) Contagem de operações elementares:
    - Substituição: “a” → “o”
    - Deleção: “casa” → “csa”
    - Inserção: “livro” → “livros”
- Faixa de valores:
    - $CER = 0\%$ : Reconhecimento perfeito
    - $CER > 10\%$ : Geralmente inaceitável para documentos estruturados
  - Vantagens:
    - Sensível a pequenos erros em textos curtos
    - Independente de normalização (avalia caracteres brutos)
  - Limitações:
    - Penaliza igualmente erros críticos e não-críticos
    - Não considera semântica ou contexto linguístico

Exemplo Prático:

- Referência: “casa”
- Previsto: “caso”
- Cálculo:  $S = 1, D = 0, I = 0, N = 4 \Rightarrow CER = 25\%$

## 2.10 Trabalhos Correlatos

Esta seção tem como objetivo apresentar os trabalhos correlatos ao proposto. A Tabela 1 apresenta uma comparação de trabalhos encontrados sobre a classificação utilizando técnicas de aprendizado profundo e visão computacional.

Tabela 1 – Comparação dos trabalhos sobre visão computacional

Autor	Modelo	Acurácia	Forma de Validação	Dados
Chen & Chiang (2025)	YOLOv4	mAP 93,29%	Treino, validação e teste	Adquiridos
Yang et al. (2025)	YOLOv8-RSS	mAP@0.5 92,8%	Treino, validação e teste	Adquiridos
Jubayer et al. (2021)	YOLOv5	mAP 99,60%	Treino, validação e teste	Adquiridos
Zhen et al. (2023)	DBNet + CBAM	Reconhecimento 98,1%	Comparação com dados	Adquiridos
Pettersson et al. (2024)	Multimodal	98,1%	Comparação entre modelos	Dataset FineGrainOCR

Fonte: Elaborado pelo Autor, 2025.

A literatura abriga uma maior quantidade de trabalhos que modelam classificadores utilizando conceitos de visão computacional. Em geral os trabalhos apresentam diferentes técnicas para gerar a predição.

O trabalho de CHEN e CHIANG (2025) propõe um sistema de identificação e contagem automática de alimentos, utilizando a técnica de detecção de objetos. O sistema foi treinado com 1836 imagens de 16 tipos de produtos de redes de *fast-food*. Além da detecção, a solução integra um aplicativo móvel que permite ao usuário identificar alimentos em tempo real, acessar informações nutricionais e registrar a ingestão alimentar. Dessa forma, o sistema oferece uma ferramenta para gerenciamento de dieta.

Jiayu Yang *et al.* (2025) desenvolvem um modelo de detecção de objetos para ambientes de armazéns frigoríficos. O estudo utilizou um conjunto de 4.463 imagens coletadas nesse contexto, anotadas em cinco categorias. O modelo foi projetado para detectar pessoas e empilhadeiras em condições de baixa temperatura, visando automatizar o monitoramento de trabalhadores e equipamentos no ambiente.

O estudo de Jubayer *et al.* (2021) propõe uma detecção de mofo em superfícies de alimentos. O conjunto de dados foi composto por 2050 imagens, obtidas

tanto em ambiente laboratorial quanto a partir de fontes online. O modelo foi treinado utilizando pesos pré-treinados. A proposta visa automatizar o processo de identificação de mofo, contribuindo para a detecção precoce e controle de deterioração em alimentos. No qual são utilizadas métricas para avaliar o modelo de detecção e leitura desenvolvidos.

O trabalho de Jishi Zheng *et al.* (2023) desenvolve um método para reconhecimento automático de datas de validade em embalagens de alimentos. O sistema realiza detecção das regiões contendo texto, seguida do reconhecimento de caracteres com uma rede convolucional. O modelo foi treinado e avaliado com um conjunto de mais de 2 mil imagens coletadas manualmente. A solução foi implementada em dispositivos embarcados, demonstrando viabilidade para aplicações industriais.

Pettersson, Riveiro e Löfström (2024) cria uma abordagem multimodal para reconhecimento de produtos de supermercado, combinando imagens de produto e textos extraídos via *OCR*. Foi criado um conjunto de dados denominado *FineGrainOCR*, com 256 classes e diversas amostras de produtos com pequenas diferenças visuais e textuais. O modelo combina representações extraídas de imagens com informações textuais por meio de técnicas de fusão multimodal. A solução propõe um reconhecimento de produtos em ambientes de varejo.

Apesar da variedade de estudos encontrados, não foram identificados trabalhos específicos focados na detecção de etiquetas de caseína e na leitura de caracteres em produtos artesanais, como o Queijo Canastra, contexto que caracteriza a originalidade e relevância da presente pesquisa.

Na seção 3 são descritos os procedimentos metodológicos adotados para o desenvolvimento deste trabalho.

### 3 METODOLOGIA

A seção atual tem como foco descrever a metodologia utilizada no desenvolvimento do presente trabalho. A seção 3.1 trata da classificação da pesquisa. Em seguida, a seção 3.2 explica a metodologia de desenvolvimento. Já, a seção 3.3 aborda os materiais e tecnologias utilizados. A seção 3.4 aborda o plano de execução. Por fim a seção 2.9 explica o sistema de métricas utilizada para validação dos modelos.

#### 3.1 Classificação da Pesquisa

A abordagem deste trabalho é quantitativa, uma vez que, de acordo Gil (2008), uma pesquisa quantitativa é baseada em resultados objetivos e recorre à matemática para descrever causas de um fenômeno e ter suas raízes no pensamento lógico e atributos mensuráveis da experiência humana.

No presente estudo, a classificação quantitativa é evidenciada pela aplicação de técnicas de aprendizado de máquina e visão computacional para detecção e leitura de etiquetas de caseína em imagens de Queijo Minas Artesanal Canastra. Para isso, são utilizadas métricas quantitativas que avaliam o desempenho do modelo desenvolvido.

O foco principal está na quantificação de características visuais e na aplicação de algoritmos matemáticos capazes de identificar padrões que possibilitam a automatização do processo de detecção e leitura.

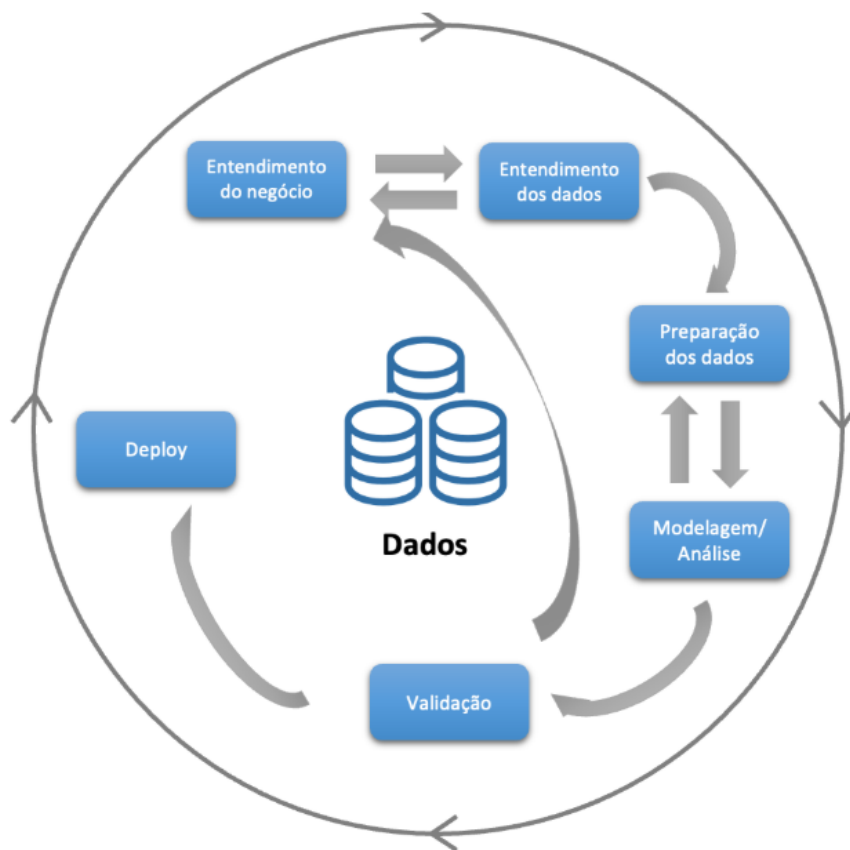
#### 3.2 Metodologia de Desenvolvimento

Como metodologia de desenvolvimento do projeto, foi adotado a metodologia Processo Padrão Intersetorial para Mineração de Dados (do inglês *Cross Industry Standard Process for Data Mining*). Segundo Schröer, Kruse e Gómez (2021) essa metodologia é um modelo de processos retirado da indústria para projetos de mineração de dados e ciência de dados. O modelo consiste em seis fases iterativas: entendimento do negócio, entendimento dos dados, preparação dos dados, modelagem, avaliação e implantação. A Figura 3 ilustra essas fases.

De acordo com Wirth e Hipp (2000) as fases tem as seguintes definições:

- Entendimento do negócio (*Business Understanding*): fase inicial que concentra-se em entender os objetivos e os requisitos do projeto a partir de uma perspectiva de negócios e converte esse conhecimento em uma definição de problema de ciência de dados.
- Entendimento dos dados (*Data Understanding*): coleta inicial dos dados para

Figura 3 – Fases CRISP-DM



Fonte: SOUZA, 2023.

análise de problemas, *insights*, detecção de subconjuntos e formulações de hipóteses.

- Preparação dos dados (*Data preparation*): essa fase abrange todas as atividades como processamento de imagens, limpeza de *dataset* e *data augmentation*. Essa etapa é necessária para construir o conjunto de dados do modelo de detecção e leitura.
- Modelagem (*Modeling*): fase dedicada à modelagem e calibração de parâmetros de treinamento, teste e validação.
- Avaliação (*Evaluation*): etapa que cuida da interpretação, análise e comparação de resultados da etapa anterior. Nesta fase é tomada a decisão sobre o uso dos resultados.
- Implantação (*Deploy*): fase final do ciclo, na qual é feito um relatório final ou realizada a prototipagem do modelo.

Ao utilizar o *CRISP-DM*, o presente trabalho se beneficia de uma estrutura organizada para o entendimento, preparação dos dados, modelagem, avaliação e implantação. Considera-se que, para o desenvolvimento de um modelo, o re-treinamento com diferentes tipos de preparação de dados e a modificação de parâmetros são etapas cruciais para a obtenção de bons resultados.

### 3.3 Materiais e Tecnologias

Todas as implementações apresentadas neste trabalho foram feitas utilizando a linguagem de programação *Python*<sup>1</sup>. O paradigma de programação adotado será o de orientação a objetos e a metodologia *CRISP-DM* descrita na seção 3.2.

Para a detecção das etiquetas de caseína foi utilizado o modelo *YOLOv8*<sup>2</sup> (*You Only Look Once*) implementado a partir do *framework* oficial disponibilizado pela *Ultralytics*.

A leitura automática dos caracteres presentes nas etiquetas foi realizada utilizando o *PaddleOCR*<sup>3</sup>, um sistema de *OCR* baseado em *deep learning*. O *PaddleOCR* integra um pipeline completo de detecção e reconhecimento de texto em imagens, suportando múltiplos idiomas e oferecendo modelos pré-treinados para diferentes cenários.

As etapas de pré-processamento das imagens e o *data augmentation* foram realizadas utilizando a biblioteca *Albumentations*<sup>4</sup>.

Para as anotações das caixas delimitadoras (*bounding boxes*) foi utilizada a ferramenta *CVAT*<sup>5</sup>.

Para operações adicionais foi utilizada a biblioteca *OpenCV*<sup>6</sup>.

O treinamento e a inferência dos modelos foram realizados em ambiente *Docker*<sup>7</sup>, utilizando contêineres configurados com suporte à aceleração por *GPU*. O ambiente de execução foi montado com o *runtime* *NVIDIA*<sup>8</sup>.

O hardware utilizado corresponde a um notebook Acer Nitro 5, equipado com o processador Intel Core i5-12450H e *GPU* *NVIDIA* GeForce RTX 3050, com 4GB de memória dedicada. O ambiente gráfico da *NVIDIA* foi configurado com o *driver* versão 570.133.07 e a *API* *CUDA* versão 12.8.

A utilização do *Docker* proporcionou um ambiente isolado, padronizado e reprodutível, facilitando a gestão de dependências e portabilidade da solução. A configuração do serviço no *Docker* incluiu a ativação do *runtime* *NVIDIA* e a exposição completa dos dispositivos de placa de vídeos para o contêiner de treinamento.

---

<sup>1</sup> <https://www.python.org/>

<sup>2</sup> <https://docs.ultralytics.com/models/yolov8/>

<sup>3</sup> <https://github.com/PaddlePaddle/PaddleOCR>

<sup>4</sup> <https://albumentations.ai/>

<sup>5</sup> <https://www.cvat.ai/>

<sup>6</sup> <https://opencv.org/>

<sup>7</sup> <https://www.docker.com/>

<sup>8</sup> <https://developer.nvidia.com/container-runtime>

### 3.4 Plano de execução

Para aplicar a metodologia *CRISP-DM* no desenvolvimento do *pipeline* automatizado de detecção de etiquetas de caseína e leitura de caracteres em imagens de Queijo Canastra, as seguintes etapas foram realizadas:

a) Entendimento do negócio:

- Definição do objetivo: automatizar a detecção das etiquetas de caseína e leitura de caracteres nelas contidos, com foco na rastreabilidade e certificação do Queijo Canastra.
- Escopo do projeto: o sistema deve localizar a etiqueta de caseína em imagens de queijos e realizar a extração automatizada do código alfanumérico impresso.

b) Entendimento dos dados:

- Coleta de imagens: as imagens foram coletadas com foco na variação de posicionamento, iluminação e orientação das etiquetas de caseína aplicadas ao queijo.
- Foi realizada uma análise das características principais das imagens, considerando aspectos como ruído visual, reflexos, parcialidades e variações na impressão das etiquetas.

c) Preparação dos dados:

- Anotações das etiquetas: as regiões de interesse *bounding boxes* correspondente às etiquetas de caseína foram anotadas manualmente utilizando a ferramenta *CVAT*, seguindo o formato de anotação compatível com o *YOLOv8*.
- Pré-Processamento: as imagens foram redimensionadas e normalizadas para garantir padronização no momento do treinamento.
- *Data augmentation*: técnicas de aumento de dados, como rotação, espelhamento, ajustes de brilho e contraste, foram aplicadas para ampliar a diversidade do conjunto de treinamento, melhorando a robustez do modelo.
- Criação do *dataset* final: o conjunto de dados foi estruturado, separado em treinamento, validação e teste.

d) Modelagem:

- Seleção de arquitetura: foi selecionado a arquitetura *YOLOv8* para realizar a detecção das etiquetas.
- Treinamento do modelo: o modelo de detecção foi treinado utilizando o conjunto de dados preparados, com ajustes de hiperparâmetros como taxa de aprendizado, *batch size* e número de épocas.
- Validação da detecção: durante o treinamento, a validação contínua

permitiu monitorar o desempenho do modelo, ajustando os hiperparâmetros conforme o necessário.

- Configuração do *OCR*: após a detecção, a leitura dos caracteres das etiquetas foi realizada com o *PaddleOCR*, utilizando modelos pré-treinados, ajustados conforme a natureza do texto impresso nas etiquetas.

e) Avaliação:

- Métricas de desempenho: o desempenho da detecção foi avaliado utilizando métricas padrão como *mAP* (*mean Average Precision*), enquanto a leitura de caracteres foi avaliada com a métrica *Character Error Rate* (*CER*). As métricas são detalhadas na seção 2.9.
- Análise dos resultados: a partir dos resultados obtidos, foi realizada uma análise para verificar o nível de acurácia da detecção e da leitura, identificando possíveis pontos de melhoria e ajustes necessários.
- Validação final: foi realizado um teste final com um conjunto de imagens não vistas para validar a generalização do *pipeline*.

f) Implantação

- A etapa de implantação não faz parte do escopo deste trabalho, limitando-se à prototipação do pipeline de detecção e leitura automatizadas.

## 4 DESENVOLVIMENTO

Esta seção apresenta a aplicação prática da Metodologia *CRISP-DM* ao desenvolvimento do *pipeline* automatizado para detecção de etiquetas de caseína e leitura de caracteres em imagens de Queijo Minas Artesanal Canastra. São detalhadas as etapas realizadas, desde a coleta e preparação dos dados, até a modelagem, avaliação e prototipação.

### 4.1 Entendimento do Negócio

Atualmente, o processo de registro dos códigos das etiquetas de caseína é realizado de forma manual: os produtores fazem anotações dos códigos utilizando na maioria dos casos papel e caneta, para posteriormente repassá-los aos responsáveis pelo controle e certificação, como a APROCAN (Associação dos Produtores de Queijo Canastra).

Esse fluxo apresenta diversas limitações, tais como:

- Susceptibilidade a erros humanos: falhas de transcrição ou perda de registros.
- Risco de inconsistências: divergências entre os dados registrados manualmente e os dados efetivos presentes na etiqueta.
- Dificuldade de integração: impossibilita o uso de sistemas automatizados de rastreabilidade.

Diante desse cenário, este projeto visa desenvolver uma automatização para a detecção das etiquetas e a leitura automática dos códigos, eliminando a necessidade do uso de mídia física no processo.

Ao proporcionar a digitalização direta das informações a partir de imagens, o sistema busca não apenas aumentar a eficiência e confiabilidade, mas também criar as bases para a modernização da cadeia produtiva e da rastreabilidade do Queijo Canastra.

### 4.2 Entendimento dos Dados

A coleta de dados foi realizada na Queijaria da Santa, localizada no município de Bambuí (MG), sob a autorização do proprietário Filipe Samir. O conjunto de dados inclui imagens de queijos com etiquetas recém fabricados, imagens do ambiente de produção (*background*) e imagens específicas para teste do *pipeline*. As imagens foram utilizadas para treinar, validar e testar o modelo de detecção e leitura automatizada das etiquetas.

O resumo do conjunto de dados encontra-se na Tabela 2:

As imagens foram capturadas com múltiplos dispositivos móveis, utilizados

Tabela 2 – Resumo do conjunto de imagens coletadas

Categoria	Quantidade
Queijos com etiquetas	33
Apenas etiquetas	39
<i>Background</i>	13
Queijos com etiquetas (teste)	11

Fonte: Elaborado pelo Autor, 2025.

de forma espontânea pelos próprios produtores. Por esse motivo, há uma variação natural nas resoluções e qualidade das imagens, com valores variando entre 1280x960 até 4080x1884 *pixels*.

Essa heterogeneidade reflete o ambiente de produção real e impôs a necessidade de realizar etapas específicas de padronização e pré-processamento, descritas na seção 4.3.

### 4.3 Preparação dos Dados

A preparação dos dados foi uma etapa que visou garantir a qualidade e padronização das imagens utilizadas no treinamento e validação do modelo. As principais atividades realizadas foram:

#### 4.3.1 Anotação das regiões de interesse

As etiquetas de caseína presentes nas imagens foram anotadas manualmente utilizando a ferramenta *CVAT*. O formato adotado para a exportação das anotações foi o padrão *YOLO*, que descreve as *bounding boxes* através de coordenadas normalizadas: classe, centro da caixa ( $x, y$ ), largura e altura.

A Figura 4 ilustra um exemplo de anotação realizada no *CVAT*, destacando a delimitação da região correspondente à etiqueta de caseína.

#### 4.3.2 Padronização das imagens

Devido à diversidade de dispositivos utilizados na coleta, as imagens apresentavam diferentes resoluções e formatos. Para garantir a compatibilidade com o modelo de detecção selecionado, todas as imagens foram redimensionadas para o padrão de entrada *YOLOv8* 640x640 *pixels*.

Figura 4 – Exemplo de imagem original e correspondente anotação no CVAT



(a) Imagem original

(b) Imagem anotada no CVAT

Fonte: Elaborado pelo Autor, 2025.

### 4.3.3 Aumento de dados

Para aumentar a variabilidade do conjunto de treinamento e reduzir o risco de *overfitting* (modelo aprende excessivamente os padrões específicos do conjunto de treinamento, perdendo a capacidade de generalizar para novos dados), foram aplicadas técnicas de *data augmentation* utilizando a biblioteca *Albumentations* e *OpenCV*. Esse processamento *offline* foi realizado previamente sobre as imagens originais, gerando novas amostras a partir de transformações que simulam variação reais nas condições de captura, como alteração de brilho, contraste, cor, adição de ruído e aplicação de filtros de desfoque.

Além disso, durante o processo de treino, o próprio modelo *YOLOv8* realizou um processamento *online*, aplicando automaticamente transformações adicionais sobre os dados em tempo real. Essas transformações incluem principalmente operações geométricas, como rotações aleatórias, cortes, espelhamentos e escalas, aumentando os dados sem a necessidade de armazenar novas imagens no disco.

As técnicas clássicas de processamento de imagens utilizadas foram previamente descritas na seção 2.4, enquanto as transformações geométricas específicas aplicadas pelo modelo durante o treinamento são detalhadas nesta seção.

#### 4.3.3.1 Brilho e contraste

Estas transformações foram aplicadas com o objetivo de simular variações naturais de iluminação, exposição e reflexos que podem ocorrer durante a captura das imagens no ambiente de produção, como visto na seção 2.4.2.

A Figura 5 compara a versão original com o ajuste de brilho e contraste aplicado.

Figura 5 – Exemplo de aplicação de brilho e contraste.



(a) Imagem original



(b) Imagem após a aplicação de brilho e contraste

Fonte: Elaborado pelo Autor, 2025.

#### 4.3.3.2 CLAHE

A técnica de equalização adaptativa de histograma foi utilizada para melhorar o contraste local das imagens, realçando detalhes que poderiam estar ocultos devido à iluminação desigual.

A Figura 6 compara o método com a imagem processada.

#### 4.3.3.3 HSV

As transformações no espaço de cores *HSV* foram aplicadas para simular mudanças cromáticas causadas por diferentes condições de iluminação.

A Figura 7 apresenta a comparação entre a imagem original e a transformação no espaço *HSV*.

Figura 6 – Exemplo de aplicação da *CLAHE*.

(a) Imagem original

(b) Imagem após a aplicação da *CLAHE*

Fonte: Elaborado pelo Autor, 2025.

Figura 7 – Exemplo de aplicação da *HSV*.

(a) Imagem original

(b) Imagem após a aplicação da *HSV*

Fonte: Elaborado pelo Autor, 2025.

#### 4.3.3.4 Ruído gaussiano

A adição do ruído gaussiano foi utilizada para simular imperfeições típicas de sensores eletrônicos ou interferências no ambiente, como ocorre em condições de baixa iluminação ou compressão de imagem como descrito na seção 2.4.4.

A Figura 8 compara a imagem original com a imagem aplicada a técnica.

Figura 8 – Exemplo de aplicação do ruído gaussiano.



(a) Imagem original



(b) Imagem após a aplicação do ruído gaussiano.

Fonte: Elaborado pelo Autor, 2025.

#### 4.3.3.5 Motion blur

O filtro de desfoque por movimento foi aplicado para simular o borramento causado pelo deslocamento da câmera ou do objeto no momento da captura.

A Figura 9 faz a comparação da imagem original com o filtro.

Figura 9 – Exemplo de aplicação do *motion blur*.



(a) Imagem original



(b) Imagem após a aplicação do *motion blur*.

Fonte: Elaborado pelo Autor, 2025.

#### 4.3.3.6 Median blur

O filtro de mediana foi utilizado para reduzir ruídos de alta frequência, preservando as bordas da imagem. Essa transformação cria amostras com suavização que simula certos tipos de imperfeições óticas.

A Figura 10 compara a imagem original e a imagem com o *median blur*.

Figura 10 – Exemplo de aplicação do *median blur*.



(a) Imagem original



(b) Imagem após a aplicação do *median blur*.

Fonte: Elaborado pelo Autor, 2025.

#### 4.3.3.7 Gaussian blur

O desfoque gaussiano foi aplicado para suavizar a imagem de maneira uniforme, simulando perdas de nitidez que podem ocorrer por falhas de foco.

A Figura 11 apresenta a comparação entre a imagem original e a imagem com *gaussian blur*.

#### 4.3.3.8 Transformações geométricas aplicadas durante o treinamento

Durante o treinamento foram aplicadas transformações geométricas e estratégias de composição com o objetivo de aumentar a variabilidade do conjunto de dados e favorecer a generalização do modelo. As transformações geométricas incluem:

- Rotação: variação aleatória de até 5 graus na orientação das imagens, simulando mudanças sutis de ângulo na captura.

Figura 11 – Exemplo de aplicação do *gaussian blur*.



(a) Imagem original



(b) Imagem após a aplicação do *gaussian blur*.

Fonte: Elaborado pelo Autor, 2025.

- Translação: deslocamento aleatório de até 10% da imagem nos eixos horizontal e vertical, simulando alterações na posição dos objetos dentro do quadro.
- Escala: alteração no tamanho dos objetos, com fator de escala de até 50%, permitindo que o modelo aprenda a reconhecer objetos em diferentes proporções.
- Cisalhamento (*Shear*): distorção angular, simulando deformações típicas causadas por ângulos de captura não ortogonais.
- Transformação de perspectiva: alteração na perspectiva das imagens, simulando mudanças na profundidade e no plano de visualização.

Além dessas transformações, foram aplicadas técnicas de composição de imagens:

- *Mosaic*: técnica que combina quatro imagens diferentes em uma única amostra, aumentando significativamente a diversidade das cenas e o contexto dos objetos.
- *MixUp*: fusão de duas imagens com combinações de seus respectivos rótulos e *pixels*, promovendo uma regularização adicional.

Essas transformações e estratégias são implementadas automaticamente no *YOLOv8*, conforme descrito na documentação oficial (ULTRALYTICS, 2023).

Na Figura 12 estão exemplificadas algumas das técnicas descritas anteriormente.

Figura 12 – Exemplo de transformações geométricas aplicadas.



(a) Imagem original



(b) Mosaic



(c) Shear



(d) Rotação

Fonte: Elaborado pelo Autor, 2025.

## 4.4 Modelagem

O desenvolvimento do sistema foi estruturado em dois *pipelines* principais: treino e inferência. O *pipeline* de treino sendo desenvolvido para o treinamento do modelo de detecção e o de inferência para a utilização do modelo de detecção de etiqueta e a inferência do modelo de leitura de caracteres.

### 4.4.0.1 Treinamento

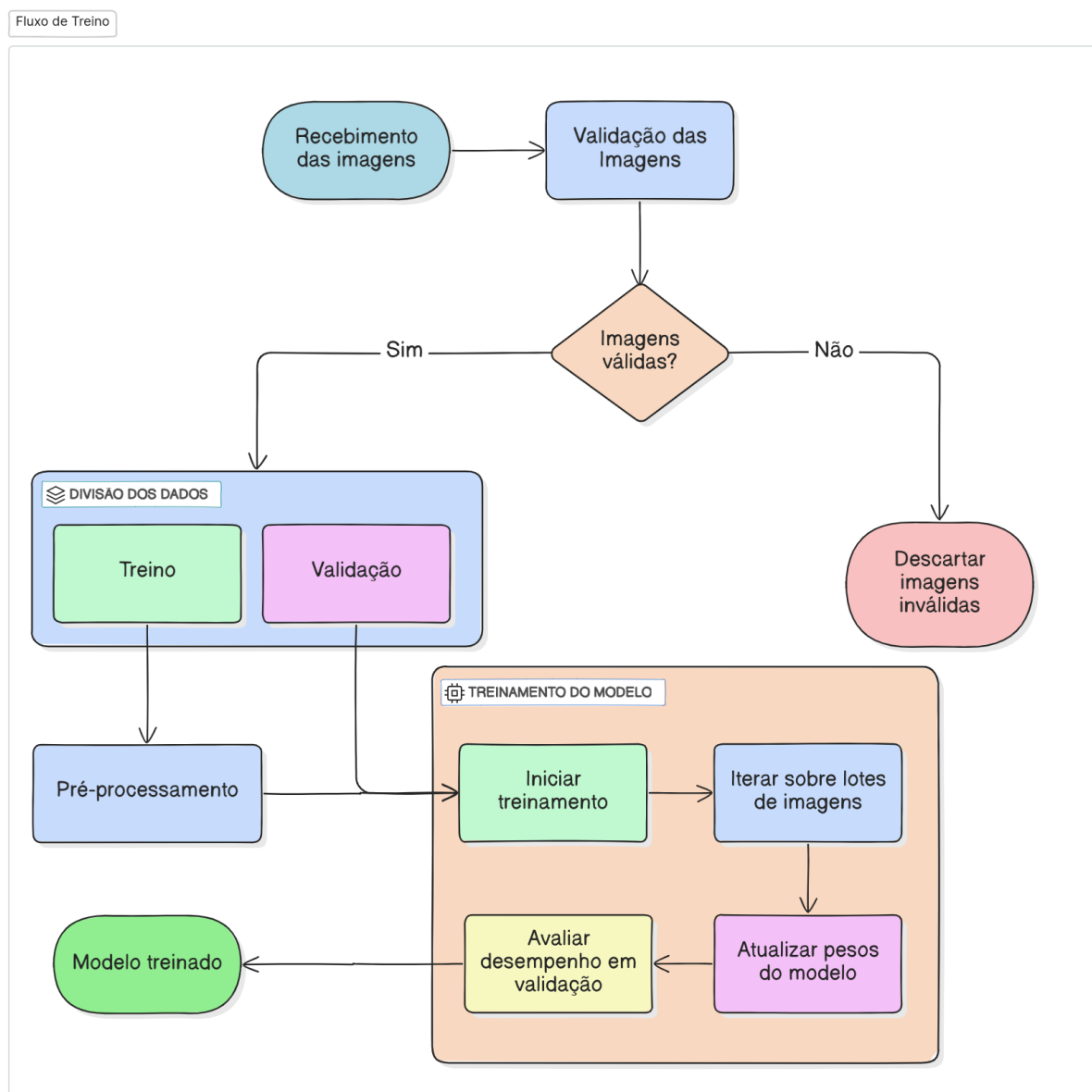
O *pipeline* de treinamento consistiu na preparação e ajuste do modelo de detecção de etiquetas utilizando a arquitetura *YOLOv8*, com o objetivo de identificar

automaticamente as etiquetas de caseína nas imagens.

O modelo foi treinado com aplicação das técnicas descritas na seção 4.3.3, incluindo tanto o processamento *offline* (transformações aplicadas previamente nas imagens) quanto o processamento *online* (transformações geométricas automáticas realizadas pelo *YOLOv8* durante o treinamento).

A Figura 13 evidencia os passos para realizar um treinamento

Figura 13 – Fluxograma de treino



Fonte: Elaborado pelo Autor, 2025.

A arquitetura escolhida foi a *YOLOv8n*, versão *nano* da família de modelos *YOLOv8*. Embora modelos maiores possam oferecer maior acurácia, a escolha se mostrou suficiente para os objetivos deste trabalho, garantindo uma boa relação entre desempenho e custo computacional.

A Tabela 3 sintetiza os principais argumentos de treino, estabelecidos.

Tabela 3 – Configuração dos parâmetros utilizados no treinamento do modelo YOLOv8n

<i>Parâmetro</i>	<i>Descrição e valor configurado</i>
Arquitetura	YOLOv8n (nano)
Resolução de entrada	640x640 pixels
Número de épocas	50
Tamanho do lote (batch size)	4
Taxa de aprendizado inicial (lr0)	0.01
Otimização	Adam
Critério de parada	Early stopping: paciência de 10 épocas
Processamento offline	Utilizando Albumentations e OpenCV
Processamento online	Utilizando transformações geométricas automáticas (YOLOv8)
Transformações geométricas	Rotação até 5°, translação até 10%, escala até 50%, cisalhamento até 0.1, perspectiva 0.0005
Composição de imagens	Mosaic (1.0), MixUp (0.2)
Semente (seed)	Aleatória a cada execução
Número de execuções	10 execuções independentes

Fonte: Elaborado pelo Autor, 2025.

Para garantir a robustez da avaliação, o modelo foi submetido a 10 execuções independentes, com diferentes valores de *seed*, permitindo a aplicação de uma análise estatística das métricas de desempenho, conforme apresentado na seção 5.

#### 4.4.0.2 Inferência

Após o processo de treinamento, foi desenvolvido um *pipeline* de inferência com o objetivo de aplicar o modelo treinado à detecção automática das etiquetas de caseína em novas imagens.

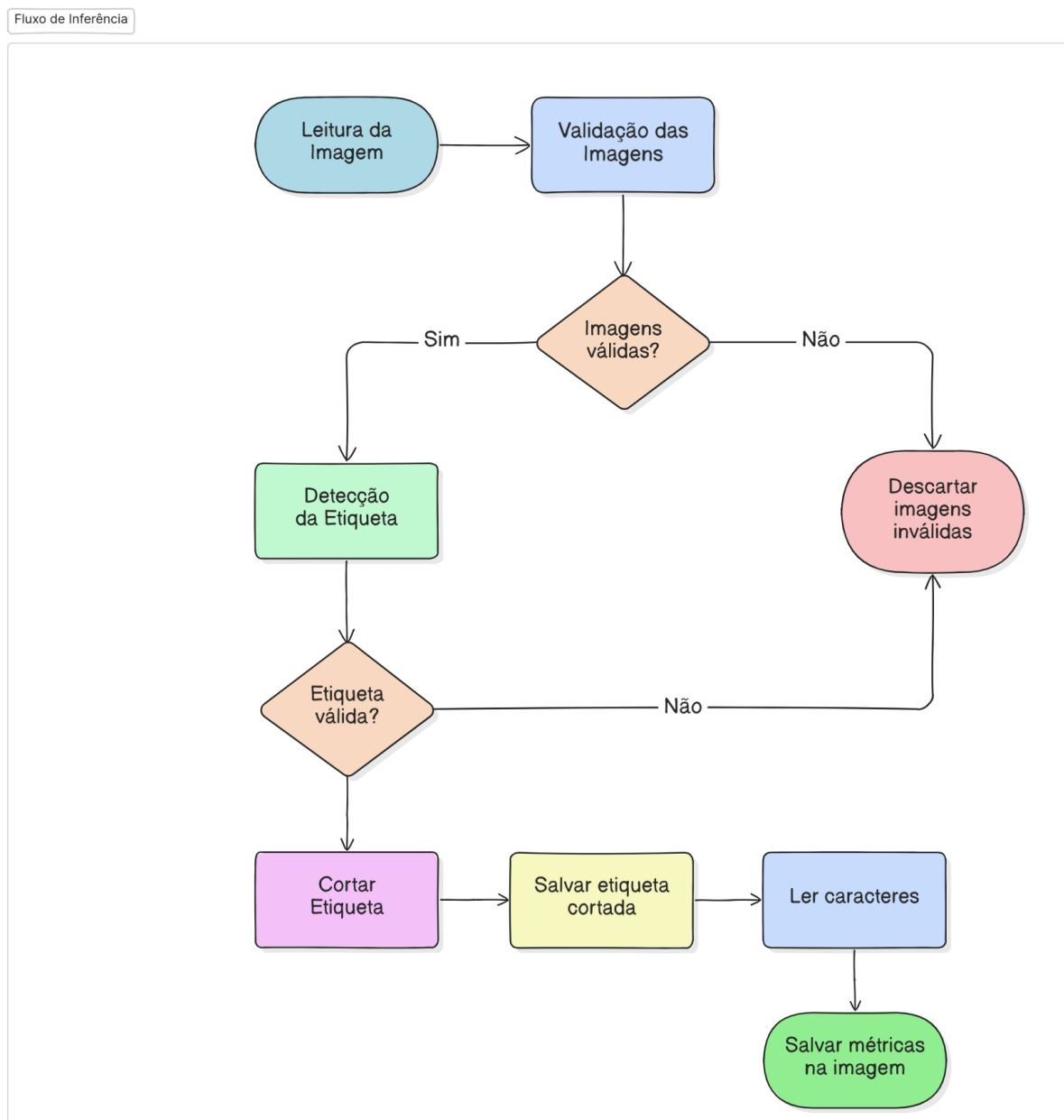
O pipeline de inferência consiste nas seguintes etapas:

- Carregamento do modelo treinado: o modelo *YOLOv8n*, com os pesos obtidos a partir do treinamento, é carregado para execução da inferência.
- Detecção da etiqueta: o modelo realiza a inferência, retornando as caixas delimitadoras associadas às etiquetas de caseína identificadas na imagem, juntamente com as confianças associadas a cada predição.
- Pós-processamento: a partir das coordenadas das *bounding boxes* detectadas, as regiões correspondentes às etiquetas são recortadas e preparadas para a etapa seguinte de leitura.

- Leitura automática do código: as imagens das etiquetas recortadas são processadas por meio do sistema de *OCR* implementado com o *PaddleOCR*, que realiza a extração automática dos caracteres alfanuméricos presentes nas etiquetas.

A Figura 14 mostra o fluxograma de ações que o *pipeline* percorre para o processo.

Figura 14 – Fluxograma de inferência



Fonte: Elaborado pelo Autor, 2025.

Optou-se pela utilização do *PaddleOCR* em razão de sua eficiência (DU *et al.*, 2020).

O treinamento de um modelo de *OCR* requer a utilização de grandes conjuntos de dados anotados, contendo imagens de texto em variadas condições de iluminação, fontes tipográficas, tamanhos e orientações. Conforme destacado por JADERBERG *et al.* (2016) e BAEK *et al.* (2019), a diversidade e o volume de exemplos são fatores determinantes para que o modelo consiga generalizar adequadamente e apresentar um desempenho robusto.

Para modelos de *OCR* amplamente utilizados na literatura foram treinados em conjuntos utilizando *datasets* como o *SynthText* Gupta, Vedaldi e Zisserman (2016) e o *MJSynth* Jaderberg *et al.* (2014), o que evidencia a complexidade e o custo computacional associados à criação de um modelo próprio.

## 5 RESULTADOS

Nesta seção, são apresentados e analisados os resultados obtidos a partir da execução do *pipeline* proposto para a detecção automática das etiquetas de caseína e a leitura automatizada dos caracteres.

Além disso, são discutidas as implicações práticas dos resultados, bem como as limitações observadas durante a execução dos experimentos.

### 5.1 Treinamentos

Nesta seção, são apresentados os resultados quantitativos obtidos durante o processo de treinamento do modelo para a detecção automática das etiquetas de caseína.

O modelo foi treinado com a aplicação de técnicas de *data augmentation*, conforme descrito na seção 4.3. A seguir, são apresentados os resultados das 10 execuções independentes, bem como a análise estatística que caracteriza o comportamento do modelo frente às variações inerentes ao processo de treinamento.

As métricas de avaliação utilizadas, como *mAP@50*, *mAP@50-95*, *Precision*, *Recall*, *DFL Loss* e *Box Loss* são empregadas na literatura para medir o desempenho de modelos de detecção de objetos (ZHAO *et al.*, 2019) e (PADILLA; NETTO; SILVA, 2022). Essas métricas permitem observar a variabilidade e a consistência dos resultados entre as diferentes execuções, conforme apresentado na Tabela 4.

Tabela 4 – Métricas obtidas nas 10 execuções do modelo com *data augmentation*

Treino	<i>mAP@50</i>	<i>mAP@50-95</i>	<i>Precision</i>	<i>Recall</i>	<i>DFL Loss</i>	<i>Box Loss</i>
1	0.99500	0.71826	1.00000	0.99435	0.82892	0.41137
2	0.99071	0.71377	0.99934	0.96667	0.85639	0.44255
3	0.99500	0.74784	0.99101	1.00000	0.78077	0.25622
4	0.99500	0.70094	0.99769	1.00000	0.92164	0.63615
5	0.98353	0.71274	0.96227	0.93333	0.89672	0.55821
6	0.99500	0.73065	1.00000	0.99201	0.96421	0.73656
7	0.99403	0.72486	0.96540	1.00000	0.87341	0.49145
8	0.99500	0.71162	0.98955	1.00000	0.91021	0.62413
9	0.99212	0.71727	0.93746	0.99931	0.85951	0.45791
10	0.99500	0.74243	0.99627	1.00000	0.87362	0.53080

Fonte: Elaborado pelo Autor, 2025.

As estatísticas descritivas das métricas obtidas ao longo das 10 execuções são apresentadas na Tabela 5. A tabela contém medidas de tendência central (média

e mediana), dispersão (desvio-padrão) e extremos (valores mínimo e máximo), permitindo caracterizar o comportamento e a variabilidade dos resultados obtidos pelo modelo.

A análise estatística das métricas, utilizando, por exemplo, o teste de Shapiro e Wilk (1965), é fundamental para verificar a aderência à normalidade e garantir a validade das comparações entre as execuções. Esse procedimento é usual em estudos que envolvem análise de desempenho de modelos de aprendizado de máquina (FIELD, 2013).

Tabela 5 – Estatísticas descritivas das métricas obtidas nas 10 execuções

Métrica	Média	Desvio-padrão	Mínimo	25%	50%	Máximo
<i>mAP@50</i>	0.9930	0.0037	0.9835	0.9926	0.9950	0.9950
<i>mAP@50-95</i>	0.7220	0.0146	0.7009	0.7130	0.7178	0.7478
<i>Precision</i>	0.9839	0.0215	0.9375	0.9714	0.9936	1.0000
<i>Recall</i>	0.9886	0.0220	0.9333	0.9926	0.9997	1.0000
<i>DFL Loss</i>	0.8765	0.0510	0.7808	0.8572	0.8735	0.9642
<i>Box Loss</i>	0.5145	0.1352	0.2562	0.4464	0.5111	0.7366

Fonte: Elaborado pelo Autor, 2025.

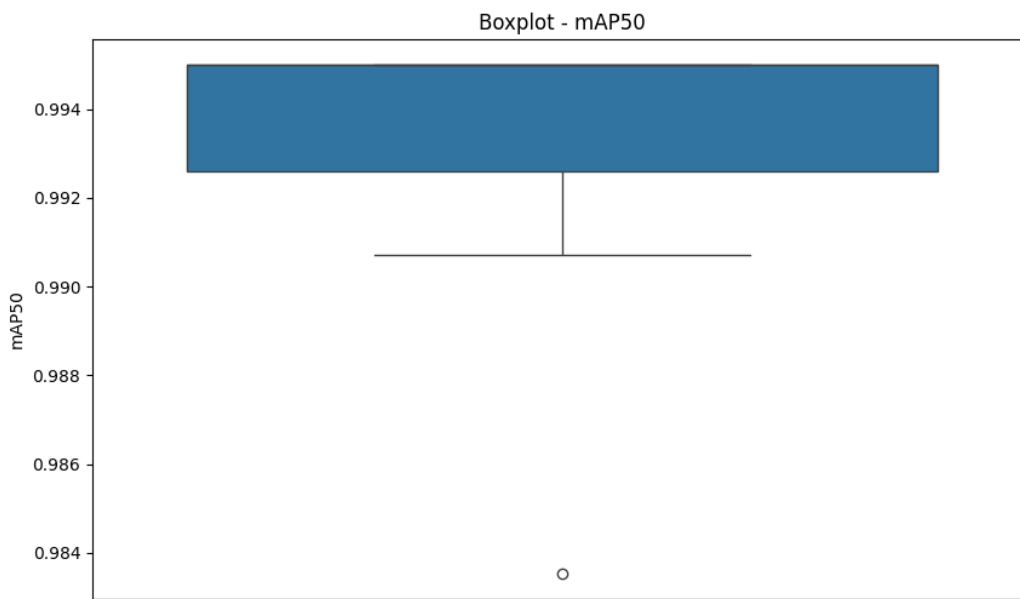
A análise das estatísticas descritivas revela que o modelo apresentou consistência ao longo das 10 execuções. As métricas principais, como *mAP@50* e *Recall*, alcançaram valores médios superiores a 99%, com baixa variabilidade, indicando que o modelo generaliza bem para novas amostras.

Por outro lado, o teste de *Shapiro-Wilk* indicou que algumas métricas, como *mAP@50* e *Precision*, não seguem distribuição normal, evidenciando a presença de pequenas flutuações naturais no processo de treinamento. Apesar disso, essas variações não comprometeram o desempenho geral do modelo, que manteve índices aceitáveis de acurácia e baixa perdas.

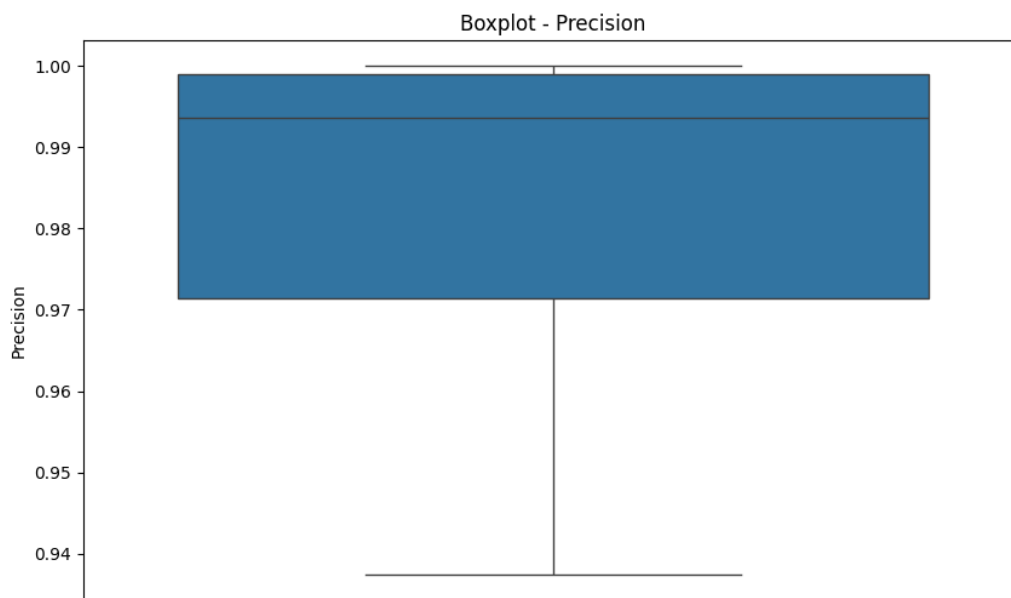
Para visualizar melhor a distribuição e a variabilidade da *Precision* e *mAP@50* entre as execuções, foram elaborados gráficos do tipo *boxplot*, apresentados na Figura 15 e 16. Esses gráficos permitem identificar a dispersão dos dados, possíveis *outliers* e a concentração dos resultados, complementando a análise estatística apresentada anteriormente.

Entre as execuções de treinamento realizadas, o modelo que apresentou o melhor desempenho foi o Treino 3 evidenciado na Tabela 4. Esse treinamento alcançou um *mAP@50* e o maior valor de *mAP@50-95* dentre todas as execuções.

Em relação às perdas associadas à regressão das caixas delimitadoras, o Treino 3 apresentou o menor valor de *Box Loss*, bem como uma das menores *DFL Loss*, evidenciando que o modelo conseguiu realizar uma regressão precisa das *bounding boxes*.

Figura 15 – *Boxplot* da métrica  $mAP@50$ .

Fonte: Elaborado pelo Autor, 2025.

Figura 16 – *Boxplot* da métrica *Precision*.

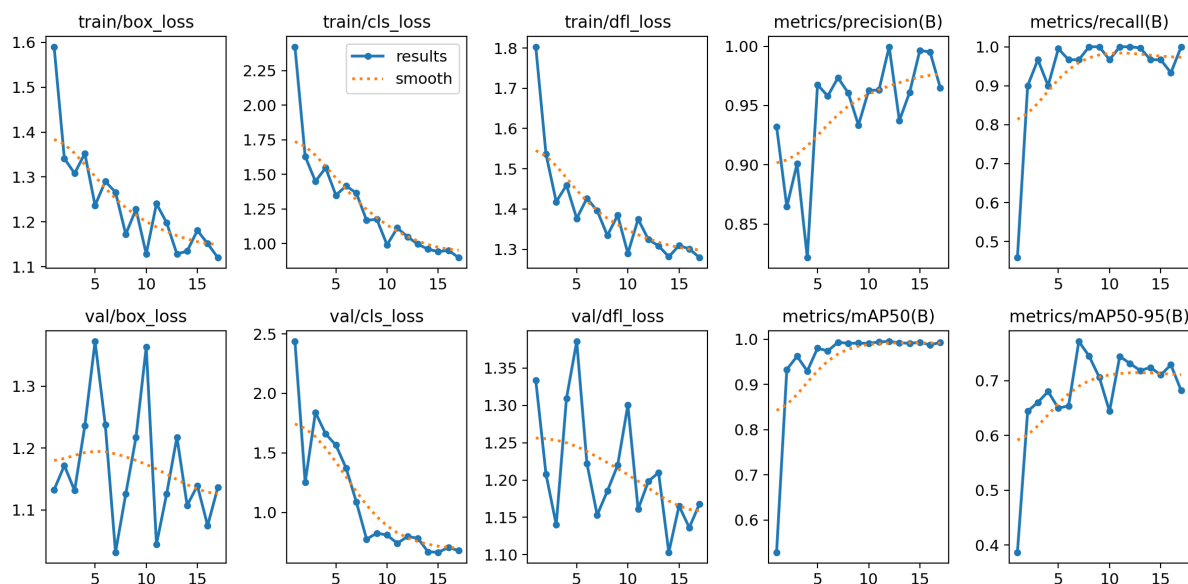
Fonte: Elaborado pelo Autor, 2025.

Diante desses resultados, o modelo resultante do Treino 3 foi selecionado para aplicação prática na fase de inferência.

Adicionalmente, foram gerados gráficos de avaliação específicos do Treino 3, com o objetivo de aprofundar a análise de seu desempenho.

A Figura 17 apresenta um resumo das métricas de treino e validação vistas anteriormente.

Figura 17 – Resumo das métricas de treino



Fonte: Elaborado pelo Autor, 2025.

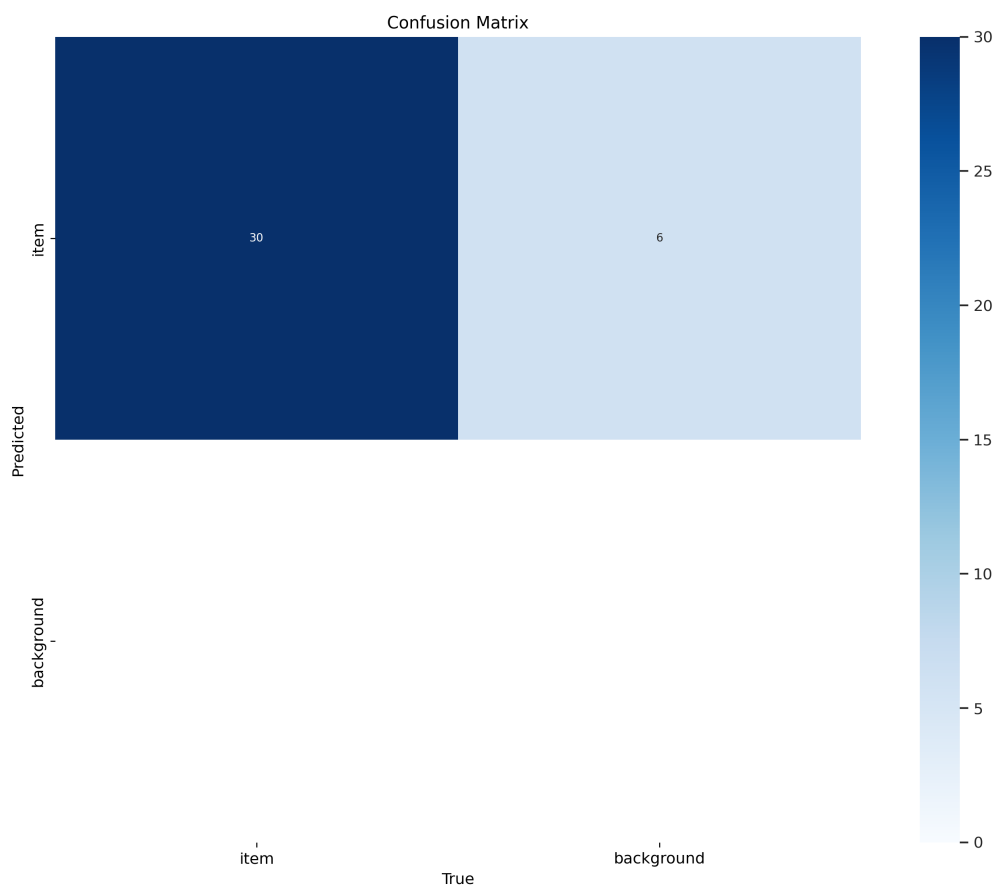
A Figura 18 demonstra por meio de uma matriz de confusão a capacidade do modelo em realizar detecções corretas das etiquetas, com nenhuma ocorrência de falsos positivos e falsos negativos durante o treinamento.

A figura 19 exibe a curva de precisão, indicando a precisão com diferentes valores de limiar de confiança.

A figura 20 apresenta a queda da curva de *Recall* à medida que o limiar de confiança aumenta. Apesar dessa tendência, a curva demonstra que, mesmo sob limiares mais restritivos, o modelo generaliza bem na detecção de etiquetas.

Apesar dos resultados positivos, destaca-se que o conjunto de dados ainda é relativamente pequeno quando comparado a *benchmarks* de detecção, o que pode restringir a capacidade de generalização para casos extremos ou condições de iluminação muito adversas. Entretanto, para o presente trabalho, o volume de dados coletado mostrou-se suficiente para validar a viabilidade da solução, atendendo aos objetos definidos de automatizar a detecção e leitura das etiquetas de caseína em ambientes reais de produção.

Figura 18 – Resumo das métricas de treino



Fonte: Elaborado pelo Autor, 2025.

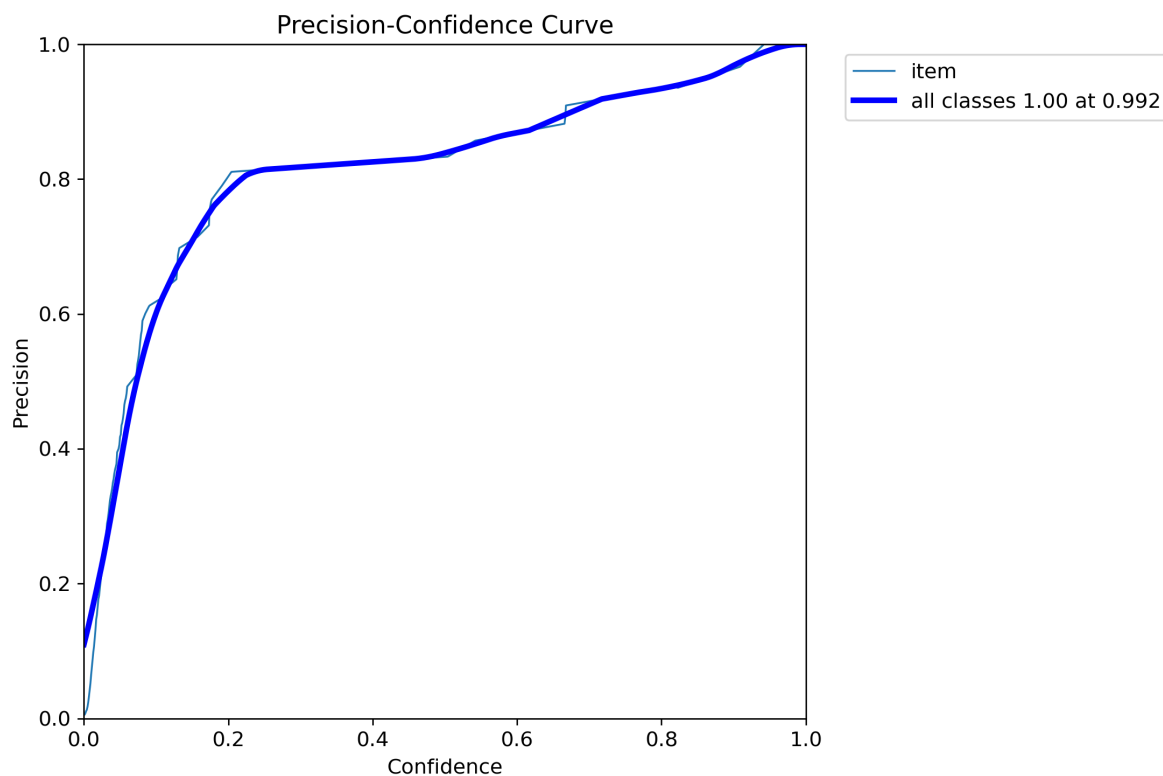
## 5.2 Inferência

A etapa de inferência foi realizada utilizando o modelo resultante do Treino 3, previamente selecionado conforme os critérios de desempenhos descritos na seção 5.1. O objetivo foi aplicar o modelo a um conjunto de imagens não utilizadas durante o treinamento, automatizando a detecção das etiquetas de caseína e, posteriormente, a leitura dos códigos alfanuméricos nelas contidos.

O *pipeline* de inferência foi estruturado conforme descrito na seção 4.4, compreendendo a detecção das regiões de interesse por meio do *YOLOv8*, seguida do recorte das áreas delimitadas e da aplicação do sistema de *OCR PaddleOCR* para extração automática dos caracteres. As Tabelas 6 e 7 resume os resultados obtidos.

A Tabela 7 apresenta as estatísticas descritivas referente aos resultados obtidos na etapa de inferência do sistema de *OCR*, considerando as métricas de confiança da detecção, taxa de erro de caracteres (*CER*) e acurácia. As medidas de tendência central, dispersão e extremos fornecem uma visão geral sobre o desempenho do sistema na leitura automática dos códigos presentes nas etiquetas de caseína,

Figura 19 – Curva de precisão



Fonte: Elaborado pelo Autor, 2025.

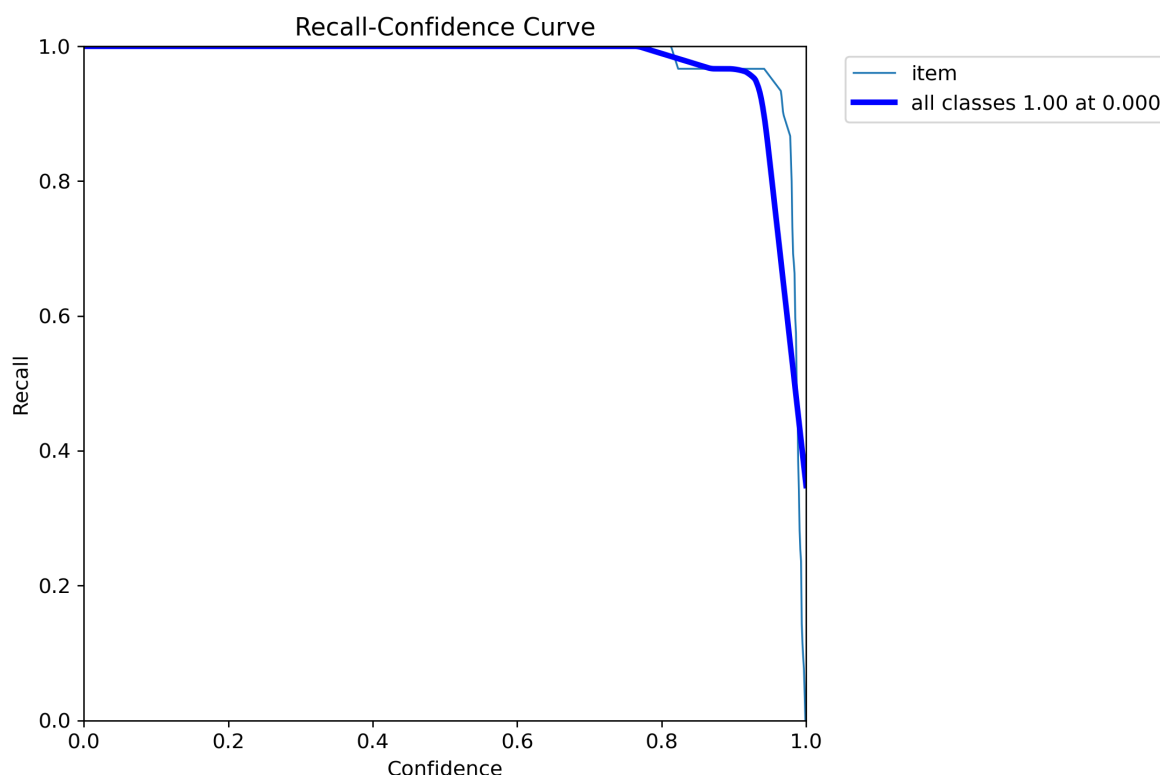
Tabela 6 – Resultados da inferência do OCR para detecção das etiquetas de caseína

Imagem	Confiança Detecção	Ground Truth	Predição	CER	ACC
03523620.jpg	0.9928	03523620	03523620	0.00	1.00
03523582.jpg	0.9817	03523582	03523582	0.00	1.00
03523577.jpg	0.9676	03523577	03523577	0.00	1.00
03523592.jpg	0.9927	03523592	(vazio)	1.00	0.00
03523621.jpg	0.9936	03523621	03523621	0.00	1.00
03523624.jpg	0.9903	03523624	03523624	0.00	1.00
03523622.jpg	0.9961	03523622	3523622	0.13	0.88
03523578.jpg	0.9763	03523578	03523578	0.00	1.00
03523593.jpg	0.9829	03523593	03523593	0.00	1.00
03523579.jpg	0.9920	03523579	03523570	0.13	0.88
03523628.jpg	0.9954	03523628	23628035	0.75	0.25

Fonte: Elaborado pelo Autor, 2025.

destacando tanto os acertos quanto as limitações observadas.

O *pipeline* de inferência alcançou uma acurácia média de 81,8%, indicando que a maioria das etiquetas foi corretamente lida. O *CER* médio foi de 18,2%, com a

Figura 20 – Curva de *Recall*

Fonte: Elaborado pelo Autor, 2025.

Tabela 7 – Estatísticas descritivas dos resultados da inferência do OCR

<i>Métrica</i>	<i>Média</i>	<i>Desvio Padrão</i>	<i>Mínimo</i>	<i>Mediana</i>	<i>Máximo</i>
<i>Confiança Detecção</i>	0.9875	0.0080	0.9676	0.9927	0.9961
<i>CER</i>	0.182	0.314	0.00	0.00	1.00
<i>Acurácia</i>	0.818	0.314	0.00	1.00	1.00

Fonte: Elaborado pelo Autor, 2025.

maioria dos casos apresentando erro zero, mas algumas falhas mais graves elevaram a média. A confiança da detecção foi consistentemente alta em todas as amostras (>96%), demonstrando que o modelo está performando bem.

Observa-se que os erros no *OCR* ocorreram mesmo quando a confiança na detecção foi elevada, reforçando a independência dos dois processos.

Casos como o da imagem 03523592.jpg, onde o *OCR* não retornou nenhum resultado, indicam limitações na generalização do modelo pré-treinado utilizado.

Para exemplificar o funcionamento da etapa de inferência, a Figura 21 apresenta a imagem recortada da etiqueta após a detecção automática, enquanto a Figura 22 ilustra o resultado da leitura automática realizada pelo *OCR*.

Esses resultados demonstram que o *pipeline* é eficaz na detecção das etiquetas de caseína e relativamente eficiente na leitura automática dos caracteres. Con-

Figura 21 – Exemplo de inferência.



(a) Imagem original



(b) Imagem após a detecção

Fonte: Elaborado pelo Autor, 2025.

Figura 22 – Processo de saída do OCR.



03523578 (0.97)

GT: 03523578

CER: 0.00%

Fonte: Elaborado pelo Autor, 2025.

tudo, a presença de erros em algumas amostras sugere a necessidade de ajustes finos ou mesmo a possibilidade de treinar um modelo *OCR* especializado no contexto das etiquetas de caseína. Ainda assim, a abordagem adotada atendeu adequadamente aos objetivos do trabalho, permitindo automatizar o processo de identificação e leitura das etiquetas, validando a proposta como uma solução viável para aplicação em ambientes produtivos reais.

## 6 TRABALHOS FUTUROS

Com base nos resultados obtidos e nas limitações identificadas ao longo do desenvolvimento deste trabalho, algumas direções podem ser exploradas em pesquisas futuras, visando aprimoramento do sistema proposto.

Em primeiro plano, destaca-se a necessidade de ampliação e diversificação do conjunto de dados, incorporando imagens capturadas sob diferentes condições de iluminação, ângulos e ambientes de produção. Essa ampliação permitirá aumentar a capacidade de generalização do modelo, tornando-o mais robusto frente a variações comuns no contexto industrial.

Outra possibilidade relevante consiste no treinamento de um modelo de *OCR* especializado para leitura das etiquetas de caseína, considerando as especificidades do formato e da tipografia utilizados nesse contexto. Tal abordagem pode contribuir para a redução das taxas de erro observadas, melhorando a acurácia do sistema de leitura automática.

Além disso, sugere-se a investigação de técnicas avançadas de pós-processamento, como a aplicação de heurísticas baseadas em regularidades dos códigos das etiquetas ou algoritmos de correção automática de erros, visando mitigar falhas de leitura que ainda possam ocorrer.

## 7 CONSIDERAÇÕES FINAIS

O presente trabalho teve como objetivo desenvolver e avaliar um *pipeline* automatizado para detecção de etiquetas de caseína e leitura dos respectivos códigos alfanuméricos em queijos do tipo Minas Artesanal Canastra. A proposta integrou técnicas de visão computacional e *deep learning*, utilizando a arquitetura *YOLOv8* para detecção e o sistema *PaddleOCR* para leitura dos caracteres.

A aplicação de técnicas de *data augmentation* demonstrou-se eficaz, conforme indicado pelos resultados obtidos. O modelo apresentou uma média de *mAP@50* de 99,30%, com baixo desvio-padrão de 0,37% ao longo de 10 execuções, evidenciando que o aumento da variabilidade no conjunto de treinamento contribuiu para melhorar a capacidade de generalização do modelo e reduzir o risco de *overfitting*.

A avaliação estatística evidenciou a confiabilidade do modelo, sustentada pela baixa dispersão das métricas principais, como o desvio-padrão da *Precision* (2,15%) e do *Recall* (2,20%). Embora o teste de *Shapiro-Wilk* tenha indicado que algumas métricas, como *mAP@50* e *Precision*, não seguem uma distribuição normal, tais flutuações são naturais no processo de treinamento e não comprometeram o desempenho geral, que se manteve elevado e consistente.

A etapa de inferência confirmou a capacidade do sistema em identificar automaticamente as etiquetas de caseína com alta confiança na detecção (média de 98,75%). Embora a leitura automática dos caracteres tenha apresentado algumas limitações, com uma acurácia média de 81,8% e *CER* médio de 18,2%, o sistema demonstrou-se capaz de automatizar o processo de identificação e leitura das etiquetas de forma satisfatória, validando sua aplicabilidade prática.

Em síntese, os resultados obtidos como o alto desempenho na detecção, a confiabilidade estatística, e a acurácia satisfatória na inferência, demonstram o potencial da abordagem proposta como uma ferramenta de apoio à rastreabilidade e ao controle de qualidade na cadeia produtiva do Queijo Minas Artesanal Canastra. A solução contribui para a modernização e eficiência do processo de certificação, alinhando-se com as necessidades de automação em ambientes produtivos reais.

Como principal limitação, destaca-se o tamanho reduzido do conjunto de dados, o que pode restringir a generalização para cenários mais complexos ou condições adversas. No entanto, para os objetivos propostos, o volume de dados coletados foi suficiente para validar a viabilidade da solução.

## REFERÊNCIAS

- AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA. **Resolução RDC nº 275, de 21 de outubro de 2002**. 2002. [https://www.gov.br/servidor/pt-br/siass/centrais\\_conteudo/manuais/resolucao-rdc-anvisa-n-275-de-21-de-outubro-de-2002.pdf/view](https://www.gov.br/servidor/pt-br/siass/centrais_conteudo/manuais/resolucao-rdc-anvisa-n-275-de-21-de-outubro-de-2002.pdf/view). Dispõe sobre o Regulamento Técnico de Procedimentos Operacionais Padronizados aplicados aos Estabelecimentos Produtores/Industrializadores de Alimentos e a Lista de Verificação das Boas Práticas de Fabricação. Acesso em: 22 maio 2025.
- AGGARWAL, C. C. **Neural networks and deep learning**. Cham: Springer, 2018. v. 10. ISBN: 978-3-319-94463-0.
- ALPAYDIN, E. **Introduction to machine learning**. Cambridge: MIT Press, 2020.
- BABATUNDE, O. H. *et al.* A survey of computer-based vision systems for automatic identification of plant species. **Journal of Agricultural Informatics**, v. 6, n. 1, 2015.
- BAEK, Y. *et al.* What is wrong with scene text recognition model comparisons? Dataset and model analysis. In: PROCEEDINGS of the IEEE International Conference on Computer Vision (ICCV). Seoul: IEEE, 2019. p. 4715–4723.
- BALLARD, D. H.; BROWN, C. M. **Computer vision**. Englewood Cliffs: Prentice Hall, 1982.
- BISHOP, C. M.; NASRABADI, N. M. **Pattern recognition and machine learning**. New York: Springer, 2006.
- CASTRO, I. L.; MADEO, F. **Processamento de Imagens: fundamentos e aplicações**. Blucher, 2020.
- CHEN, Y.-C.; CHIANG, H.-C. Deep learning-based automatic food identification with numeric label. **Multimedia Tools and Applications**, Springer, New York, p. 1–23, 2025.
- COSIDO, O. *et al.* Hybridization of convergent photogrammetry, computer vision, and artificial intelligence for digital documentation of cultural heritage: a case study — the Magdalena Palace. In: PROCEEDINGS of the 2014 International Conference on Cyberworlds. Santander: IEEE, 2014. p. 369–376.
- DEUS, G. A. d. **Utilização de aprendizado de máquina para previsão de resultados de jogos de futebol**. 2019. Universidade Tecnológica Federal do Paraná.
- DU, B. *et al.* PaddleOCR: An ultra lightweight optical character recognition system. **arXiv preprint arXiv:2009.09941**, 2020. Disponível em: <https://arxiv.org/abs/2009.09941>. Acesso em: 22 maio 2025.
- EVERINGHAM, M. *et al.* The Pascal Visual Object Classes (VOC) Challenge. **IJCV**, v. 88, p. 303–338, 2010.
- FIELD, A. **Discovering statistics using IBM SPSS statistics**. 4. ed. London: Sage, 2013.

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS. **Food Traceability Guidance**. 2019. Disponível em: <https://www.fao.org/documents/card/en/c/ca8098en>. Acesso em: 22 maio 2025.

GARDNER, M. W.; DORLING, S. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. **Atmospheric environment**, Elsevier, v. 32, n. 14-15, p. 2627–2636, 1998.

GÉRON, A. **Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow**. Sebastopol: O’Reilly Media, 2019.

GIL, A. C. **Métodos e técnicas de pesquisa social**. 6. ed. Editora Atlas SA, 2008.

GONZALEZ, R. C.; WOODS, R. E. **Digital Image Processing**. 3rd: Prentice Hall, 2008.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep learning**. Cambridge: MIT Press, 2016.

GUIMARÃES, J. *et al.* Características físico-químicas do queijo Minas artesanal da Canastra. **Revista do Instituto de Laticínios Cândido Tostes**, v. 66, n. 380, p. 16–22, 2011.

GUPTA, A.; VEDALDI, A.; ZISSERMAN, A. Synthetic data for text localisation in natural images. In: PROCEEDINGS of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. p. 2315–2324.

HASTIE, T. *et al.* **The elements of statistical learning: data mining, inference, and prediction**. Springer, 2009. v. 2.

HAYKIN, S. **Neural Networks and Learning Machines**. Pearson Education, 2009.

INSTITUTO NACIONAL DA PROPRIEDADE INDUSTRIAL. **Queijo Canastra recebe Indicação de Procedência**. 2012.

<https://ocs.ige.unicamp.br/ojs/anais3sigci/article/download/3290/3141/13265>. Registro de Indicação Geográfica assegura a origem e o método tradicional de produção do Queijo Canastra. Acesso em: 22 maio 2025.

JADERBERG, M. *et al.* Reading text in the wild with convolutional neural networks. **International Journal of Computer Vision**, Springer, New York, v. 116, n. 1, p. 1–20, 2016.

JADERBERG, M. *et al.* Synthetic Data and Artificial Neural Networks for Natural Scene Text Recognition. **arXiv preprint arXiv:1406.2227**, 2014.

JAIN, A. K. **Fundamentals of digital image processing**. Englewood Cliffs: Prentice Hall, 1989. ISBN: 978-0133361650.

JUBAYER, F. *et al.* Detection of mold on the food surface using YOLOv5. **Current Research in Food Science**, Elsevier, v. 4, p. 724–728, 2021.

JUNIOR, L. C. G. C. *et al.* Variações na composição de queijo Minas artesanal da Serra da Canastra nas quatro estações do ano. **Revista do Instituto de Laticínios Cândido Tostes**, v. 64, n. 371, p. 13–20, 2009.

KINGMA, D. P.; BA, J. **Adam: a method for stochastic optimization**. 2015. Trabalho apresentado na International Conference on Learning Representations (ICLR), 2015. Disponível em: <https://arxiv.org/abs/1412.6980>. Acesso em: 22 maio 2025.

LI, Z. *et al.* Generalized focal loss: learning qualified and distributed bounding boxes for dense object detection. In: **ADVANCES in Neural Information Processing Systems (NeurIPS)**. Vancouver: Curran Associates, Inc., 2020. p. 21002–21012.

LIN, T.-Y. *et al.* **Microsoft COCO: common objects in context**. 2014. Preprint (arXiv:1405.0312). Disponível em: <https://arxiv.org/abs/1405.0312>. Acesso em: 22 maio 2025.

LIN, T.-Y. *et al.* Focal Loss for Dense Object Detection. In: **PROCEEDINGS of the IEEE International Conference on Computer Vision (ICCV)**. 2017. p. 2980–2988.

LONG, J.; SHELHAMER, E.; DARRELL, T. Fully convolutional networks for semantic segmentation. In: **PROCEEDINGS of the IEEE conference on computer vision and pattern recognition**. 2015. p. 3431–3440.

MINISTÉRIO DA AGRICULTURA, PECUÁRIA E ABASTECIMENTO - MAPA. **Portaria MAPA nº 146, de 07 de março de 1996**. 1996. Acesso em: 19 de março de 2023. Disponível em: <https://www.defesa.agricultura.sp.gov.br/legislacoes/portaria-mapa-146-de-07-03-1996,669.html>.

MITCHELL, T. M. **Machine learning**. New York: McGraw-Hill, 1997.

NARDY, V. P. D. R.; CARVALHO, G. R.; ROCHA, D. T. da. Mercado de leite fluido e queijos no Brasil: uma análise de 2005 a 2016. In: **IN: WORKSHOP DE INICIAÇÃO CIENTÍFICA DA EMBRAPA GADO DE LEITE, 23.**, 2019 . . .

PADILLA, R.; NETTO, S. L.; SILVA, E. A. B. d. A survey on performance metrics for object-detection algorithms. **Neurocomputing**, Elsevier, Amsterdam, v. 505, p. 345–362, 2022.

PATEL, K. K. *et al.* Machine vision system: a tool for quality inspection of food and agricultural products. **Journal of food science and technology**, Springer, v. 49, p. 123–141, 2012.

PETTERSSON, T.; RIVEIRO, M.; LÖFSTRÖM, T. Multimodal fine-grained grocery product recognition using image and OCR text. **Machine Vision and Applications**, Springer, v. 35, n. 4, p. 79, 2024.

POWERS, D. M. W. **Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation**. 2020. Preprint (arXiv:2010.16061). Disponível em: <https://arxiv.org/abs/2010.16061>. Acesso em: 22 maio 2025.

- REDMON, J. *et al.* You only look once: Unified, real-time object detection. In: PROCEEDINGS of the IEEE conference on computer vision and pattern recognition. 2016. p. 779–788.
- REVISTA MENU. **Queijo da Serra da Canastra ganha selo de autenticidade**. 2019. Disponível em: <https://revistamenu.com.br/queijo-da-serra-da-canastra-ganha-selo-de-autenticidade>. Acesso em: 22 maio 2025.
- REZATOFIGHI, H. *et al.* Generalized Intersection over Union: A Metric and A Loss for Bounding Box Regression. In: PROCEEDINGS of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019. p. 658–666.
- SCHRÖER, C.; KRUSE, F.; GÓMEZ, J. M. A systematic literature review on applying CRISP-DM process model. **Procedia Computer Science**, Elsevier, v. 181, p. 526–534, 2021.
- SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). **Biometrika**, Oxford University Press, v. 52, n. 3-4, p. 591–611, 1965.
- SHORTEN, C.; KHOSHGOFTAAR, T. M. A survey on Image Data Augmentation for Deep Learning. **Journal of Big Data**, v. 6, n. 1, p. 60, 2019.
- SMITH, R. An overview of the Tesseract OCR engine. In: IEEE. NINTH international conference on document analysis and recognition (ICDAR 2007). 2007. v. 2, p. 629–633.
- SOBREIRO, V. A. *et al.* **Figura 2: Perceptron Multicamadas. Fonte: Adaptado de LNCC (2008)**. 2008. Disponível em: [https://www.researchgate.net/figure/Figura-2-Perceptron-Multicamadas-Fonte-Adaptado-de-LNCC-2008\\_fig1\\_228432919](https://www.researchgate.net/figure/Figura-2-Perceptron-Multicamadas-Fonte-Adaptado-de-LNCC-2008_fig1_228432919). Acesso em: 27 maio 2025. Referente ao trabalho: Uma estimativa do valor da commodity de açúcar utilizando redes neurais artificiais.
- SOUZA, V. **Entenda o CRISP-DM, suas etapas e como de fato gerar valor com essa metodologia**. Figura: Diagrama das etapas do CRISP-DM. Acesso em: 27 maio 2025. 2023. Disponível em: <https://www.preditiva.ai/blog/entenda-o-crisp-dm-suas-etapas-e-como-de-fato-gerar-valor-com-essa-metodologia>.
- SUTTON, R. S.; BARTO, A. G. **Reinforcement learning: An introduction**. MIT press, 2018.
- ULTRALYTICS. **Ultralytics YOLOv8 Documentation**. 2023. Disponível em: <https://docs.ultralytics.com/>. Acesso em: 27 setembro 2024.
- WIRTH, R.; HIPPEL, J. CRISP-DM: Towards a standard process model for data mining. In: MANCHESTER. PROCEEDINGS of the 4th international conference on the practical applications of knowledge discovery and data mining. 2000. v. 1, p. 29–39.
- YANG, J. *et al.* Lightweight object detection model for food freezer warehouses. **Scientific Reports**, Nature Publishing Group UK London, v. 15, n. 1, p. 2350, 2025.

YANG, S. *et al.* Probabilistic Object Detection: Definition and Evaluation. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, IEEE, v. 44, n. 8, p. 3981–3995, 2021.

ZHAO, Z.-Q. *et al.* Object detection with deep learning: A review. **IEEE transactions on neural networks and learning systems**, IEEE, v. 30, n. 11, p. 3212–3232, 2019.

ZHENG, J. *et al.* Recognition of expiry data on food packages based on improved DBNet. **Connection Science**, Taylor & Francis, v. 35, n. 1, p. 1–16, 2023.

ZHENG, Z. *et al.* Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression. **arXiv preprint arXiv:1911.08287**, 2020.

ZUIDERVELD, K. Contrast limited adaptive histogram equalization. In: HECKBERT, P. S. (Ed.). **Graphics Gems IV**. San Diego: Academic Press Professional, Inc., 1994. p. 474–485.

## ANEXOS

Tabela 8 – Métricas do modelo escolhido por época

Época	Tempo	Box	DFL	Prec.	Recall	mAP50	mAP50-95
1	4.53	1.59	1.80	0.93	0.46	0.53	0.39
2	8.37	1.34	1.54	0.86	0.90	0.93	0.64
3	12.39	1.31	1.42	0.90	0.97	0.96	0.66
4	16.40	1.35	1.46	0.82	0.90	0.93	0.68
5	20.03	1.24	1.38	0.97	0.99	0.98	0.65
6	23.73	1.29	1.43	0.96	0.97	0.97	0.65
7	27.37	1.27	1.40	0.97	0.97	0.99	0.77
8	30.92	1.17	1.33	0.96	1.00	0.99	0.74
9	34.66	1.23	1.38	0.93	1.00	0.99	0.71
10	38.47	1.13	1.29	0.96	0.97	0.99	0.64
11	42.55	1.24	1.37	0.96	1.00	0.99	0.74
12	46.32	1.20	1.32	1.00	1.00	1.00	0.73
13	50.11	1.13	1.31	0.94	1.00	0.99	0.72
14	54.25	1.14	1.28	0.96	0.97	0.99	0.72
15	57.98	1.18	1.31	1.00	0.97	0.99	0.71
16	61.57	1.15	1.30	1.00	0.93	0.99	0.73
17	65.34	1.12	1.28	0.96	1.00	0.99	0.68

Fonte: Elaborado pelo Autor, 2025.