



**MINISTÉRIO DA EDUCAÇÃO
SECRETARIA DE EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO
TOCANTINS - CAMPUS LAGOA DA CONFUSÃO
CURSO SUPERIOR EM ENGENHARIA AGRÔNOMICA**

EUDES ALVES PORTILHO JUNIOR

**DETECÇÃO AUTOMATIZADA DE GRÃOS EM PANÍCULAS DE ARROZ COM
YOLOv11: INFLUÊNCIA DO FUNDO DE CAPTURA NO DESEMPENHO DO
MODELO**

**Lagoa da Confusão - TO
2026**



**MINISTÉRIO DA EDUCAÇÃO
SECRETARIA DE EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO
TOCANTINS - CAMPUS LAGOA DA CONFUSÃO
CURSO SUPERIOR EM ENGENHARIA AGRONÔMICA**

EUDES ALVES PORTILHO JUNIOR

**DETECÇÃO AUTOMATIZADA DE GRÃOS EM PANÍCULAS DE ARROZ COM
YOLOv11: INFLUÊNCIA DO FUNDO DE CAPTURA NO DESEMPENHO DO
MODELO**

Trabalho de Conclusão de Curso apresentado como requisito parcial para obtenção do título de Engenheiro Agrônomo, do Curso Superior de Engenharia Agronômica, do Instituto Federal do Tocantins, Campus Lagoa da Confusão.

Orientadora: Ma. Karen Cristina Leite Silva

Coorientadora: Dra. Lorena Gabriela Coelho de Queiroz

**Lagoa da Confusão - TO
2026**

Dados Internacionais de Catalogação na Publicação (CIP)
Bibliotecas do Instituto Federal do Tocantins

J95d Junior, Eudes Alves Portilho
DETECÇÃO AUTOMATIZADA DE GRÃOS EM PANÍCULAS DE
ARROZ COM YOLOv11 : INFLUÊNCIA DO FUNDO DE CAPTURA
NO DESEMPENHO DO MODELO / Eudes Alves Portilho Junior. –
Lagoa da Confusão, TO, 2026.

39 f.

Trabalho de Conclusão de Curso (Bacharelado em Engenharia
Agrônômica) – Instituto Federal de Educação, Ciência e Tecnologia
do Tocantins, Campus Lagoa da Confusão, Lagoa da Confusão, TO,
2026.

Orientadora: Ma. Karen Cristina Leite Silva Cristina

Coorientadora: Dra. Lorena Gabriela Coelho de Queiroz Gabriela

1. Arroz. 2. YOLOv11. 3. Fenotipagem. I. Cristina, Karen Cristina
Leite Silva. II. Gabriela, Lorena Gabriela Coelho de Queiroz. III.
Título.

CDD 630

A reprodução total ou parcial, de qualquer forma ou por qualquer meio, deste documento é autorizada para fins
de estudo e pesquisa, desde que citada a fonte.

Elaborado pelo sistema de geração automática de ficha catalográfica do IFTO com os dados fornecidos
pelo(a) autor(a).



Ministério da Educação
Secretaria de Educação Profissional e Tecnológica
Instituto Federal de Educação, Ciência e Tecnologia do Tocantins
Reitoria
Campus Lagoa da Confusão
Coordenação do Curso Superior em Engenharia Agrônômica

FOLHA DE APROVAÇÃO

TÍTULO: Contagem automatizada de grãos em panículas de arroz com o modelo YOLOv11: a influência do fundo de captura das imagens no desempenho do treinamento

AUTOR : Eudes Alves Portilho Junior
ORIENTADORA: Karen Cristina Leite Silva

Trabalho de Conclusão de Curso apresentado ao Instituto Federal de Educação, Ciência e Tecnologia do Tocantins, *Campus* Lagoa da Confusão, como parte das exigências para a conclusão do Curso de Bacharelado em Engenharia Agrônômica.

Aprovado em 21 de agosto de 2026.

BANCA EXAMINADORA

Prof^a. Ma. Karen Cristina Leite Silva
Orientadora
IFTO/*Campus* Lagoa da Confusão

Dr^a Lorena Gabriela Coelho de Queiroz
Co-orientadora
Syngenta

Dr. Liomar Borges de Oliveira
Membro
IFTO/*Campus* Lagoa da Confusão

Me. José Robson Mariano Alves
Membro
IFTO/*Campus* Lagoa da Confusão



Documento assinado eletronicamente por **Jose Robson Mariano Alves, Servidor**, em 21/08/2026, às 09:52, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Lorena Gabriela Coelho de Queiroz, Usuário Externo**, em 21/08/2026, às 09:54, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Karen Cristina Leite Silva, Servidora**, em 21/08/2026, às 09:55, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Liomar Borges de Oliveira, Servidor**, em 21/08/2026, às 09:56, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site http://sei.ifto.edu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **3296889** e o código CRC **C40E0E21**.

Rua 02, Quadra 05A, Lote 01, Setor Lagoa da Ilha — CEP 77.493-000 — Lagoa da Confusão/TO
portal.ifto.edu.br — reitoria@ifto.edu.br

Aos meus pais, que nunca mediram esforços para que eu pudesse estudar, e à minha família, que aguentou minha ausência em tantos fins de semana enquanto este trabalho era feito. Sem vocês, eu não teria chegado até aqui.

AGRADECIMENTOS

Agradeço primeiramente a Deus, por ter me dado saúde e paciência para chegar até o fim desta caminhada.

Agradeço à minha orientadora e coorientadora, pela paciência em explicar as coisas quantas vezes fossem necessárias, pelas correções sinceras e por não ter desistido do meu trabalho nos momentos em que ele parecia não sair do lugar. Cada devolutiva que recebi me fez entender melhor o que eu estava fazendo.

Agradeço aos professores do Instituto Federal do Tocantins, Campus Lagoa da Confusão, por todo o conhecimento que dividiram comigo ao longo do curso. Muita coisa que aparece nas próximas páginas eu só consegui entender por causa das aulas de vocês.

Agradeço aos colegas de turma que me ajudaram na parte mais cansativa deste trabalho, que foi marcar grão por grão nas imagens. Foram muitas horas de trabalho repetitivo, e eu não teria conseguido fazer tudo sozinho.

Agradeço à minha família e aos meus amigos, pelo apoio e pela compreensão nos momentos em que eu precisei escolher entre estar com eles e terminar mais uma etapa da pesquisa.

A todos que, de alguma forma, contribuíram para a realização deste trabalho, o meu sincero agradecimento.

RESUMO

O número de grãos por panícula é uma das características associadas à produtividade do arroz (*Oryza sativa* L.) e, por isso, constitui uma informação relevante para a pesquisa e o melhoramento de plantas. Entretanto, sua determinação ainda é frequentemente realizada por contagem manual, um processo demorado, trabalhoso e sujeito a erros. Uma alternativa é a contagem automatizada por meio de visão computacional e redes neurais convolucionais, especialmente com modelos da família YOLO. O desempenho desses modelos pode ser influenciado pelas características das imagens utilizadas no treinamento, incluindo o fundo sobre o qual as amostras são fotografadas. Diante disso, o objetivo deste trabalho foi validar um modelo de automação baseado no YOLOv11 para a detecção de grãos em panículas de arroz e avaliar a influência da composição e da diversidade dos fundos utilizados no treinamento sobre o desempenho do modelo. As panículas foram coletadas aleatoriamente em campo e fotografadas em laboratório sobre quatro tipos de fundo: EVA azul, EVA preto, solo e granito, sendo cada panícula fotografada em apenas uma dessas condições. As imagens foram inseridas na plataforma Roboflow, onde os grãos foram delimitados manualmente por polígonos e classificados em duas categorias: grão cheio e grão chocho. Após o pré-processamento e o aumento de dados, o conjunto totalizou 532 imagens, divididas em treinamento, validação e teste na proporção de 70%, 20% e 10%, respectivamente. Foram estabelecidos quatro cenários de treinamento, compostos por diferentes combinações de fundos, e o desempenho foi avaliado por meio das métricas de precisão, revocação, F1-score e mAP@50. Os resultados demonstraram diferenças de desempenho entre os cenários em função da composição dos fundos utilizados no treinamento. O maior mAP@50 foi obtido no cenário que reuniu todos os fundos (34,4%), seguido pelo cenário com inclusão do granito (33,3%). O cenário composto pelos fundos de EVA apresentou 27,3%, enquanto a inclusão do solo resultou no menor valor, de 26,6%. O F1-score máximo foi próximo de 0,40 nos cenários avaliados, indicando desempenho ainda limitado do modelo. Conclui-se que a composição dos fundos utilizados no conjunto de treinamento influenciou o desempenho do YOLOv11 na detecção de grãos em panículas de arroz, evidenciando que a escolha do fundo de captura deve ser considerada no planejamento metodológico de sistemas automatizados de detecção.

Palavras-chave: *Oryza sativa*. Fenotipagem. Visão computacional. YOLOv11. Aprendizado profundo.

ABSTRACT

The number of grains per panicle is one of the traits associated with rice (*Oryza sativa* L.) yield and, therefore, represents relevant information for research and plant breeding. However, this trait is still frequently assessed through manual counting, a time-consuming, labor-intensive, and error-prone process. Automated counting using computer vision and convolutional neural networks, particularly models from the YOLO family, represents a potential alternative. The performance of these models may be influenced by the characteristics of the images used for training, including the background on which the samples are photographed. Therefore, this study aimed to validate an automation model based on YOLOv11 for grain detection in rice panicles and to evaluate the influence of the composition and diversity of training backgrounds on model performance. Panicles were randomly collected in the field and photographed under laboratory conditions using four background types: blue EVA foam, black EVA foam, soil, and granite, with each panicle photographed under only one of these conditions. The images were uploaded to the Roboflow platform, where grains were manually annotated using polygons and classified into two categories: filled and unfilled grains. After preprocessing and data augmentation, the dataset comprised 532 images, which were divided into training, validation, and test sets at proportions of 70%, 20%, and 10%, respectively. Four training scenarios were established using different combinations of backgrounds, and model performance was evaluated using precision, recall, F1-score, and mAP@50. The results showed differences in model performance according to the composition of the backgrounds used for training. The highest mAP@50 was obtained in the scenario combining all backgrounds (34.4%), followed by the scenario including granite (33.3%). The scenario composed of EVA foam backgrounds achieved 27.3%, whereas the inclusion of soil resulted in the lowest value, at 26.6%. The maximum F1-score was approximately 0.40 across the evaluated scenarios, indicating that model performance was still limited. It was concluded that the composition of the backgrounds used in the training dataset influenced YOLOv11 performance in detecting grains in rice panicles, demonstrating that the choice of image background should be considered when planning automated detection systems.

Keywords: *Oryza sativa*. Phenotyping. Computer vision. YOLOv11. Deep learning.

LISTA DE FIGURAS

Figura 1 - Exemplos de panículas de arroz nos diferentes fundos de captura utilizados para composição do conjunto de dados.....	20
Figura 2 - Gráficos do cenário 1 (EVA preto e EVA azul), com mAP@50 de 27,3%	Erro! Indicador não definido.
Figura 3 - Gráficos do cenário 2 (EVA preto, EVA azul e solo), com mAP@50 de 26,6%	27
Figura 4 - Gráficos do cenário 3 (EVA preto, EVA azul e granito), com mAP@50 de 33,3%	28
Figura 5 - Gráficos do cenário 4 (EVA preto, EVA azul, solo e granito), com mAP@50 de 34,4%	29

LISTA DE TABELAS

Tabela 1 - Composição do conjunto de imagens e condições de captura	21
Tabela 2 - Composição dos cenários utilizados para o treinamento do modelo YOLOv11.....	24
Tabela 3 - Precisão média (mAP@50) obtida para os diferentes cenários Indicador não definido.	Erro!
Tabela 4 - Resumo das boas práticas de anotação	38

SUMÁRIO

1 INTRODUÇÃO	11
2 PROBLEMA DE PESQUISA	12
3 JUSTIFICATIVA	12
4 OBJETIVOS	13
4.1 Objetivo geral	13
4.2 Objetivos específicos	13
5 REFERENCIAL TEÓRICO	13
5.1 A importância do número de grãos por panícula	13
5.2 A contagem manual e as primeiras tentativas de automação	14
5.3 Redes neurais convolucionais e o modelo YOLO	15
5.4 A influência do fundo da imagem no treinamento	16
5.5 Métricas de avaliação e trabalhos recentes	17
6 MATERIAIS E MÉTODOS.....	19
6.1 Tipo de pesquisa.....	Erro! Indicador não definido.
6.2 Coleta das amostras e captura das imagens	19
6.3 Anotação e preparo do conjunto de dados	21
6.4 Treinamento do modelo e definição dos cenários.....	22
6.5 Avaliação do desempenho	23
7 RESULTADOS E DISCUSSÃO.....	24
7.1 Precisão média (mAP@50) obtida em cada cenário.....	24
7.2 Comportamento detalhado das métricas	26
7.3 Discussão dos resultados frente à literatura	31
8 CONSIDERAÇÕES FINAIS	33
REFERÊNCIAS	35
ANEXO A - PASSO A PASSO PARA ANOTAÇÃO NO ROBOFLOW	37
Objetivo.....	37
1 Acessar a plataforma	37
2 Abrir o projeto correto	37
3 Marcar os grãos	37
4 Usar a anotação assistida	37

5 Revisar e ajustar	38
6 Lidar com variações de posição	38
7 Salvar e avançar	38

1 INTRODUÇÃO

O arroz (*Oryza sativa* L.) é uma das culturas alimentares mais importantes do mundo e, por isso, ganhos de produtividade nessa cultura têm reflexo direto na oferta de alimentos (DENG et al., 2021). A produtividade do arroz é definida por um conjunto de componentes de rendimento, entre eles o número de panículas por área, o número de grãos por panícula e o peso dos grãos. Em linhagens e cultivares de arroz de terras altas, o número de grãos por panícula está entre os caracteres mais associados à produtividade (MENDES et al., 2014) e constitui uma informação relevante para as decisões tomadas nos programas de melhoramento (ZOU et al., 2023).

Entretanto, a maneira tradicional de obter esse dado ainda é trabalhosa. O método clássico consiste em debulhar a panícula e realizar a contagem manual dos grãos, tornando o processo demorado, cansativo e sujeito a erros (DENG et al., 2021), especialmente porque a atenção do avaliador pode diminuir ao longo de períodos prolongados de trabalho repetitivo (ZOU et al., 2023).

Diante dessa limitação, pesquisas passaram a buscar métodos automatizados de contagem. Com o avanço do aprendizado profundo, especialmente das redes neurais convolucionais, tornou-se possível desenvolver modelos capazes de aprender características dos objetos a partir de exemplos (LECUN; BENGIO; HINTON, 2015). Entre esses modelos, os da família YOLO se destacaram por combinar velocidade e precisão na detecção de objetos em uma única etapa (REDMON et al., 2016). No arroz, essa família já foi aplicada tanto na contagem de grãos (ZOU et al., 2023) quanto na detecção de panículas em campo, incluindo versões recentes, como o YOLOv11 (PENG et al., 2025; TAN et al., 2025).

Apesar desses avanços, permanece uma lacuna. Os melhores resultados relatados foram obtidos em ambientes controlados, com as amostras posicionadas sobre superfícies escolhidas para favorecer a distinção entre o objeto e o restante da imagem (ZOU et al., 2023; ZHANG et al., 2024). Em condições de campo, nas quais o cenário é mais heterogêneo, o desempenho tende a ser inferior (SUN et al., 2022; WANG et al., 2022). Assim, apesar do bom desempenho dos modelos de detecção de grãos em ambientes controlados, ainda é necessário compreender como a composição e a complexidade do fundo das imagens podem afetar o desempenho do modelo.

Essa lacuna pode estar relacionada à similaridade visual entre o objeto e o fundo, que pode dificultar a discriminação dos grãos pelo modelo e resultar em maior número de detecções incorretas (DENG et al., 2021; LIU et al., 2025). Ainda assim, o fundo raramente é tratado como uma variável do estudo, sendo geralmente apresentado apenas como uma condição metodológica.

Nesse contexto, buscou-se avaliar se a inclusão de imagens obtidas em fundos com diferentes níveis de complexidade visual altera o desempenho do modelo e se o aumento da diversidade de fundos no conjunto de treinamento favorece sua capacidade de generalização.

2 PROBLEMA DE PESQUISA

A automação da contagem de grãos por meio de aprendizado profundo tem se mostrado viável, com estudos relatando elevados valores de precisão na detecção de grãos de arroz (DENG et al., 2021; ZOU et al., 2023). Esses resultados, entretanto, foram obtidos predominantemente em condições controladas, com as amostras posicionadas sobre fundos simples e uniformes, utilizados para favorecer o contraste entre o objeto e o fundo (ZOU et al., 2023). Em condições mais complexas, como em imagens obtidas em campo, a heterogeneidade do fundo pode aumentar a dificuldade de detecção (SUN et al., 2022; WANG et al., 2022).

Nesse contexto, ainda é necessário compreender como as características do fundo utilizado na captura das imagens podem interferir no desempenho dos modelos de detecção. A similaridade visual entre o objeto e o fundo pode dificultar a identificação dos grãos, enquanto a inclusão de fundos distintos no conjunto de treinamento pode produzir respostas diferentes no desempenho do modelo.

Diante disso, estabelece-se a seguinte questão de pesquisa: de que maneira a composição e a diversidade dos fundos utilizados no conjunto de treinamento influenciam o desempenho de um modelo YOLOv11 na detecção de grãos em panículas de arroz?

3 JUSTIFICATIVA

O trabalho se justifica, em primeiro lugar, pela importância da característica estudada e pelo ganho prático que a automação pode trazer. O número de grãos por panícula está entre os caracteres mais associados à produtividade do arroz (MENDES

et al., 2014), e substituir a contagem manual por um processo automático permite avaliar mais amostras em menos tempo e com um padrão mais estável de medição, justamente porque a medição manual é o fator que limita o número de plantas avaliadas em um programa de pesquisa (ZHANG et al., 2024).

Em segundo lugar, justifica-se pelo recorte escolhido. A maior parte dos trabalhos da área concentra esforços em melhorar a arquitetura da rede neural (ZOU et al., 2023; LIU et al., 2025; TAN et al., 2025), enquanto o fundo de captura costuma aparecer apenas como informação de método. Estudar esse ponto de forma direta ajuda a entender uma variável que está ao alcance de qualquer pesquisador, já que trocar o fundo é bem mais simples e barato do que reprogramar uma rede neural (ZHANG et al., 2024).

4 OBJETIVOS

4.1 Objetivo geral

Validar um modelo de detecção de grãos em panículas de arroz baseado no YOLOv11, considerando diferentes fundos de captura de imagens.

4.2 Objetivos específicos

- a) Avaliar o efeito da utilização de diferentes fundos de captura de imagens no treinamento sobre o desempenho e a capacidade de generalização do modelo;
- b) Avaliar o desempenho do modelo por meio das métricas de precisão, revocação, F1 e precisão média (mAP@50).

5 REFERENCIAL TEÓRICO

5.1 A importância do número de grãos por panícula

O arroz constitui um dos principais alimentos da dieta de bilhões de pessoas, desempenhando papel relevante na segurança alimentar mundial. Entre os componentes associados à produtividade da cultura, destaca-se o número de grãos por panícula, característica de interesse em estudos de rendimento e em programas de melhoramento genético (DENG et al., 2021). Dessa forma, a quantificação precisa dos grãos é importante para a avaliação de características relacionadas à

produtividade e pode auxiliar nos processos de seleção (ZOU et al., 2023). Relação semelhante é observada em outras culturas, como o milho, em que o número de grãos constitui um importante componente da produtividade (AMELONG et al., 2015).

O número de grãos por panícula, entretanto, não determina isoladamente o rendimento, que também depende de outros componentes, como a massa dos grãos. Embora representem características distintas, sua avaliação pode envolver a necessidade de quantificação precisa do número de grãos. Nesse contexto, métodos automatizados de detecção e contagem podem contribuir para aumentar a eficiência e a padronização das avaliações, especialmente quando há grande número de amostras a serem analisadas (MENDES et al., 2014; ZOU et al., 2023).

Os avanços obtidos pelo melhoramento genético ao longo das últimas décadas demonstram a importância da seleção e do aprimoramento dos componentes associados ao rendimento das culturas (DUVICK, 2005). No Brasil, por exemplo, a produtividade do milho apresentou aumento expressivo ao longo de sete décadas, alcançando valores aproximadamente quatro vezes superiores aos observados no início desse período (GALVÃO et al., 2014). A continuidade desses ganhos depende da avaliação de um grande número de genótipos nos programas de melhoramento (PRASANNA et al., 2022), o que demanda métodos de fenotipagem capazes de gerar dados de forma precisa, eficiente e padronizada. A qualidade dessas informações é fundamental para aumentar a confiabilidade das decisões de seleção e, consequentemente, contribuir para o progresso genético (CROSSA et al., 2021).

5.2 A contagem manual e as primeiras tentativas de automação.

Apesar da importância do número de grãos por panícula, a forma tradicional de obter esse dado ainda é trabalhosa. O método clássico consiste em debulhar a panícula e contar os grãos um por um, na mão. Já na debulha aparecem problemas, porque alguns grãos se perdem, outros são danificados e outros continuam presos à panícula, e a soma desses pequenos erros vira um erro acumulado (DENG et al., 2021). A contagem também depende da atenção de uma pessoa, que diminui ao longo de horas repetitivas (ZOU et al., 2023), o que compromete as decisões tomadas a partir dela, como acontece na contagem de vagens de soja (LIU et al., 2025), e ainda limita o número de amostras que se consegue avaliar (ZHANG et al., 2024).

Diante desses problemas surgiram as primeiras tentativas de automação. Elas não usavam inteligência artificial moderna, e sim técnicas mais simples de análise de imagens, conhecidas como visão computacional clássica, em que o computador

separa o grão do fundo por regras fixas de cor, tamanho, forma e textura (ZOU et al., 2023). Vários métodos desse tipo foram testados em arroz, mas quase sempre exigiam os grãos já debulhados, o que trazia de volta os erros da debulha (DENG et al., 2021).

Uma das principais limitações dessas técnicas está na dependência de regras previamente definidas para a análise das imagens. No caso dos grãos de arroz, o tamanho reduzido, a semelhança visual e a proximidade entre os grãos dificultam sua individualização (ZOU et al., 2023). Além disso, variações nas condições de iluminação, como presença de sombras e reflexos, podem alterar as características utilizadas na segmentação e comprometer a identificação dos objetos (DENG et al., 2021). Essas limitações contribuíram para a adoção de métodos baseados em aprendizado profundo, capazes de extrair automaticamente características relevantes das imagens (ZHANG et al., 2024).

5.3 Redes neurais convolucionais e o modelo YOLO

O avanço do aprendizado profundo, especialmente das redes neurais convolucionais (Convolutional Neural Networks – CNNs), ampliou as possibilidades de análise automatizada de imagens. Diferentemente dos métodos tradicionais baseados em características e regras previamente estabelecidas, as CNNs são capazes de aprender representações diretamente a partir dos dados utilizados durante o treinamento (LECUN; BENGIO; HINTON, 2015). Entre as primeiras aplicações de destaque dessas redes está o reconhecimento de caracteres manuscritos (LECUN et al., 1998). Posteriormente, seu desempenho em tarefas de classificação de imagens em larga escala contribuiu para a consolidação do aprendizado profundo na área de visão computacional (KRIZHEVSKY; SUTSKEVER; HINTON, 2012; GOODFELLOW; BENGIO; COURVILLE, 2016).

O funcionamento das CNNs baseia-se na extração hierárquica de características das imagens. Nas camadas iniciais, a rede identifica padrões mais simples, como bordas, contrastes e variações de intensidade. À medida que a informação avança pelas camadas, essas características são combinadas em representações mais complexas, como texturas, formas e estruturas relacionadas aos objetos de interesse (GOODFELLOW; BENGIO; COURVILLE, 2016).

O funcionamento de uma CNN pode ser entendido como um processo em camadas. Nas primeiras ela aprende elementos simples, como bordas e variações de

cor; nas seguintes esses elementos viram texturas e formas parciais; nas mais profundas os padrões se juntam e representam o objeto inteiro (LECUN; BENGIO; HINTON, 2015). Isso dá robustez ao modelo, ou seja, a capacidade de continuar funcionando diante de variações de iluminação, ruídos de fundo e grãos sobrepostos (ZOU et al., 2023), e por isso as CNNs viraram a base de várias ferramentas de fenotipagem, que é o trabalho de medir as características da planta (ZHANG et al., 2024).

A contagem de grãos em imagens pode ser abordada como uma tarefa de detecção de objetos, que consiste em localizar e identificar os objetos de interesse presentes em uma imagem (ZOU et al., 2023). Os métodos de detecção podem ser classificados, de modo geral, em detectores de dois estágios e de estágio único. Os detectores de dois estágios inicialmente geram regiões candidatas e, posteriormente, realizam a classificação e o refinamento dessas regiões, abordagem que pode apresentar maior custo computacional (LIU et al., 2025). Em contrapartida, os detectores de estágio único realizam a localização e a classificação dos objetos em uma única etapa. Entre eles, destaca-se a família YOLO (*You Only Look Once*), inicialmente proposta por Redmon et al. (2016), com o objetivo de realizar a detecção de objetos de forma rápida e eficiente. A versão YOLOv11, usada aqui, faz parte dessa evolução e é disponibilizada de forma aberta, já treinada em grandes conjuntos de imagens, o que permite adaptá-la a uma tarefa específica sem construir uma rede do zero (JOCHER; QIU, 2024). O treinamento desses modelos pode ser feito em plataformas de visão computacional que reúnem, em um mesmo ambiente, a anotação das imagens, o aumento de dados e o próprio treinamento, o que reduz a barreira técnica para quem está começando na área (ROBOFLOW, 2026).

5.4 A influência do fundo da imagem no treinamento

O desempenho de um modelo de detecção não depende só da rede escolhida, e sim, em boa medida, das imagens usadas para treiná-lo. Como as redes convolucionais aprendem a partir dos exemplos que recebem, tudo o que aparece nessas imagens influencia o que elas aprendem, inclusive aquilo que não é o objeto de interesse (LECUN; BENGIO; HINTON, 2015). O fundo é justamente tudo o que aparece na imagem e não é o objeto procurado, e desde as técnicas mais antigas a detecção é definida como a separação entre objeto e fundo (ZOU et al., 2023). Um erro acontece quando o modelo aponta um pedaço do fundo como se fosse um grão,

e outro quando ele deixa de encontrar um grão e o trata como fundo (DENG et al., 2021).

A similaridade visual entre objeto e fundo pode dificultar a discriminação dos grãos pelo modelo. Em um estudo de contagem de grãos de arroz, observou-se aumento de erros de detecção nas situações em que o grão e o fundo apresentavam aparência semelhante, assim como em imagens escuras ou desfocadas (DENG et al., 2021). Para reduzir esse risco, muitos trabalhos escolhem de propósito um fundo simples e de alto contraste, como a placa de fundo branco usada na captura de imagens de grãos de arroz (ZOU et al., 2023) ou a placa de base montada para separar visualmente a planta do cenário em soja (ZHANG et al., 2024).

Esse cuidado facilita o treinamento, mas cria um problema novo, porque o modelo fica muito bom naquela situação específica e não em outras, limitação reconhecida pelo próprio trabalho que usou a placa de contraste (ZHANG et al., 2024). Em campo, com fundo heterogêneo, os trabalhos com arroz relatam dificuldade para detectar as panículas, o que motivou modelos específicos para esses ambientes (SUN et al., 2022; WANG et al., 2022; PENG et al., 2025). Existe, então, um equilíbrio a ser buscado, e é para que lado essa balança pende que este trabalho olha.

5.5 Métricas de avaliação e trabalhos recentes

A avaliação de modelos de detecção de objetos é realizada por meio de métricas que permitem quantificar sua capacidade de identificar corretamente os objetos de interesse. Entre as métricas mais utilizadas estão a precisão, a revocação, o F1-score e a precisão média (AP), calculadas a partir das relações entre verdadeiros positivos, falsos positivos e falsos negativos (ZOU et al., 2023).

O verdadeiro positivo (VP) corresponde a um objeto corretamente detectado pelo modelo, enquanto o falso positivo (FP) ocorre quando uma região é identificada incorretamente como objeto de interesse. O falso negativo (FN), por sua vez, representa um objeto presente na imagem que não foi detectado pelo modelo (ZOU et al., 2023).

A precisão expressa a proporção de detecções corretas em relação ao total de detecções realizadas pelo modelo, enquanto a revocação representa a proporção de objetos corretamente identificados em relação ao total de objetos existentes. Essas métricas apresentam uma relação dependente do limiar de confiança adotado, de modo que alterações nesse limiar podem modificar o equilíbrio entre detecções

corretas, falsas detecções e objetos não identificados. O F1-score corresponde à média harmônica entre precisão e revocação, sendo utilizado para representar o equilíbrio entre ambas. A precisão média (AP), por sua vez, sintetiza a relação entre precisão e revocação ao longo da curva precisão-revocação (ZOU et al., 2023).

Nos últimos anos vários trabalhos mostraram que o aprendizado profundo pode contar estruturas das plantas de forma automática, e o arroz está entre as culturas mais estudadas. Na contagem de grãos, um estudo alcançou precisão geral de 99,4 por cento trabalhando diretamente sobre as panículas (DENG et al., 2021), e outro, baseado em YOLOv5, atingiu 99,20 por cento (ZOU et al., 2023). Ambos foram conduzidos em condições controladas de captura.

Na detecção de panículas em condições de campo, os estudos têm explorado estratégias voltadas a ambientes e condições de captura mais complexos. Wang et al. (2022) desenvolveram um método para detecção e contagem de panículas em imagens de grandes áreas, com atenção à identificação de panículas de menor tamanho. Sun et al. (2022), por sua vez, propuseram uma versão aprimorada do YOLOv4 para a detecção de panículas curvas em imagens aéreas obtidas em quatro locais distintos. Mais recentemente, abordagens baseadas no YOLOv11 também foram empregadas nessa tarefa, tanto para a detecção de panículas em condições de campo (PENG et al., 2025) quanto para sua identificação em diferentes estágios, utilizando caixas delimitadoras orientadas (TAN et al., 2025). Abordagens baseadas em visão computacional também têm sido aplicadas a outras culturas. Na soja, Liu et al. (2025) desenvolveram uma ferramenta para contagem de vagens em condições de campo, alcançando 94,1% de desempenho mesmo em situações de sobreposição e interferência do fundo. Zhang et al. (2024), por sua vez, utilizaram o YOLOv8 para a detecção e contagem de grãos no interior de vagens. No milho, Wang e Chen (2025) empregaram uma versão aprimorada do YOLOv8 para a detecção de espigas em condições de campo.

As condições de aquisição das imagens constituem um aspecto relevante na interpretação e comparação do desempenho desses modelos. Estudos realizados em condições controladas de captura relataram valores superiores a 99% em determinadas métricas de desempenho (DENG et al., 2021; ZOU et al., 2023), enquanto trabalhos conduzidos em condições de campo apresentaram valores inferiores (SUN et al., 2022; LIU et al., 2025). Essa diferença evidencia que o desempenho dos modelos deve ser interpretado considerando as condições

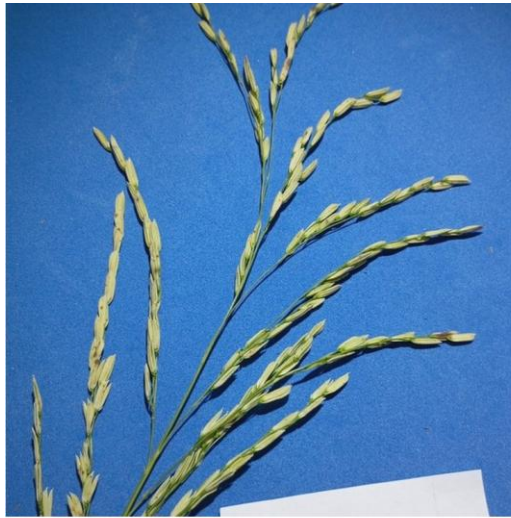
experimentais, as características das imagens e as métricas utilizadas em cada estudo, evitando comparações baseadas exclusivamente nos valores numéricos apresentados (ZHANG et al., 2024).

6 MATERIAIS E MÉTODOS

6.2 Coleta das amostras e captura das imagens

As panículas de arroz foram coletadas em uma fazenda comercial de produção de arroz, localizada no município de Lagoa da Confusão, Tocantins. Foram coletadas aproximadamente 300 panículas em diferentes pontos da área de produção. Após a coleta, as amostras foram encaminhadas ao laboratório para obtenção das imagens, sendo mantidas inteiras, sem debulha. Para a obtenção das imagens, foram utilizados quatro fundos de captura: EVA azul, EVA preto, solo e granito. Cada panícula selecionada foi fotografada em apenas uma condição de fundo. Portanto, os conjuntos correspondentes aos quatro fundos foram constituídos por panículas distintas, não havendo repetição da mesma panícula entre as diferentes condições de fundo. Para cada panícula foram obtidas três fotografias em diferentes ângulos. Posteriormente, foi realizada uma seleção das imagens, priorizando aquelas em que os grãos apresentavam melhor separação visual, com menor sobreposição entre os próprios grãos e menor interferência das ramificações da panícula sobre os grãos. Após essa seleção, foram utilizadas 223 imagens para composição do conjunto de dados. A Figura 1 apresenta exemplos das imagens obtidas nos quatro fundos de captura. Durante a obtenção das imagens, as panículas foram fotografadas juntamente com etiquetas contendo a identificação da amostra e informações previamente registradas, como a contagem manual de grãos e o peso. Essas informações foram mantidas apenas como registro das amostras e não foram utilizadas na avaliação do desempenho do modelo nem para comparação com as contagens obtidas automaticamente.

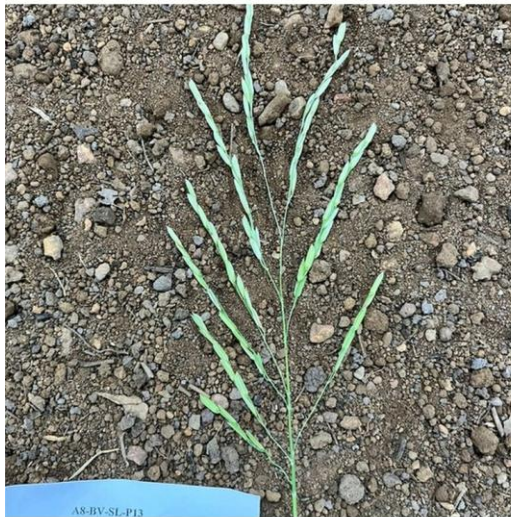
Figura 1 - Exemplos de panículas de arroz nos diferentes fundos de captura utilizados para composição do conjunto de dados.



(a) EVA azul



(b) EVA preto



(c) Solo



(d) Granito

Fonte: Elaborada pelo autor (2026).

Tabela 1 - Composição do conjunto de imagens e condições de captura

Informação	Valor
Número total de imagens capturadas	223
Imagens sobre EVA azul	51
Imagens sobre EVA preto	60
Imagens sobre solo	52
Imagens sobre granito	60
Total de anotações (grãos marcados)	5.153
Anotações da classe grão cheio	2.996
Anotações da classe grão chocho	2.157
Total de imagens após o aumento de dados	532
Imagens no conjunto de treino	372
Imagens no conjunto de validação	107
Imagens no conjunto de teste	53
Proporção da divisão treino/validação/teste	70% / 20% / 10%
Equipamento utilizado na captura	câmera de telefone celular
Resolução original das imagens	não registrada no momento da captura
Distância entre câmera e amostra	ajustada para enquadrar a panícula inteira
Condição de iluminação	ambiente bem iluminado, para facilitar a marcação dos grãos
Resolução usada no treinamento	640 × 640 pixels

Fonte: Elaborada pelo autor (2026).

6.3 Anotação e preparo do conjunto de dados

As imagens selecionadas foram importadas para a plataforma Roboflow, onde foi realizada a anotação dos grãos. As anotações foram efetuadas individualmente por meio da ferramenta de segmentação por polígonos, delimitando o contorno de cada grão presente nas imagens.

Os grãos foram classificados em duas categorias: **grãos cheios** e **grãos chochos**. A classificação foi realizada com base nas características visuais dos grãos, considerando-se como cheios aqueles com maior volume aparente e formato uniforme, enquanto grãos com aparência deformada ou baixo preenchimento foram classificados como chochos. Ao final do processo, foram realizadas **5.153 anotações**, das quais 2.996 corresponderam à classe grão cheio e 2.157 à classe grão chocho.

Para auxiliar o processo de anotação, foi utilizada a ferramenta **Label Assist**, disponível na plataforma Roboflow. As anotações sugeridas automaticamente pela ferramenta foram posteriormente revisadas e, quando necessário, ajustadas manualmente, garantindo a delimitação individual dos grãos e a atribuição da

respectiva classe. O procedimento de anotação seguiu os critérios descritos no Anexo A.

Após a anotação, as imagens foram submetidas às etapas de pré-processamento e aumento de dados (*data augmentation*) na plataforma Roboflow. No pré-processamento, as imagens foram redimensionadas para 640 × 640 pixels. O aumento de dados incluiu transformações de brilho, contraste, rotação e espelhamento, resultando na ampliação do conjunto de 223 imagens originais para 532 imagens.

O conjunto de dados foi posteriormente organizado em subconjuntos destinados ao treinamento, validação e teste do modelo, na proporção de 70%, 20% e 10%, respectivamente.

6.4 Treinamento do modelo e definição dos cenários

O treinamento foi realizado na plataforma Roboflow utilizando o modelo YOLOv11, previamente treinado e posteriormente ajustado ao conjunto de imagens anotadas neste estudo. Os modelos resultantes de cada cenário foram acessados por meio da plataforma, e as métricas e curvas de desempenho foram obtidas utilizando scripts desenvolvidos em linguagem Python.

Para avaliar o efeito da composição dos fundos no desempenho do modelo, foram estabelecidos quatro cenários de treinamento. O cenário 1, considerado como referência, foi composto pelas imagens obtidas sobre os fundos de EVA preto e EVA azul, totalizando 111 imagens originais. O cenário 2 incluiu, além dos fundos de EVA, as imagens obtidas sobre o solo, totalizando 163 imagens. No cenário 3, foram utilizadas as imagens dos fundos de EVA e granito, totalizando 171 imagens. Por fim, o cenário 4 reuniu as imagens obtidas nos quatro fundos avaliados (EVA preto, EVA azul, solo e granito), totalizando 223 imagens originais.

Os cenários foram estruturados de forma cumulativa a partir dos conjuntos de imagens correspondentes aos diferentes fundos, conforme apresentado na Tabela 2. Para cada cenário, foram aplicados os procedimentos de pré-processamento e aumento de dados descritos na Seção 6.3.

Ressalta-se que os cenários diferiram tanto quanto à composição dos fundos quanto ao número de imagens disponíveis para o treinamento. Além disso, as imagens correspondentes a cada fundo foram provenientes de panículas distintas.

Dessa forma, as diferenças observadas entre os cenários foram interpretadas considerando essas fontes de variação.

Tabela 2 - Composição dos cenários utilizados para o treinamento do modelo YOLOv11.

Cenário	Fundos utilizados	Nº de imagens originais
1	EVA preto + EVA azul	111
2	EVA preto + EVA azul + solo	163
3	EVA preto + EVA azul + granito	171
4	EVA preto + EVA azul + solo + granito	223

Fonte: Elaborada pelo autor (2026).

6.5 Avaliação do desempenho

O desempenho dos modelos obtidos em cada cenário foi avaliado por meio das métricas precisão (P), revocação (R), F1-score e mean Average Precision (mAP@50). O valor de mAP@50 foi obtido diretamente na plataforma Roboflow ao final do treinamento, considerando um limiar de *Intersection over Union* (IoU) de 0,50.

Adicionalmente, foram analisadas as curvas de precisão × confiança, revocação × confiança, F1-score × confiança e precisão × revocação, obtidas a partir dos modelos treinados por meio de scripts desenvolvidos em linguagem Python. Essas curvas foram utilizadas para avaliar o comportamento das métricas em função da variação do limiar de confiança.

Para o cálculo das métricas, foram considerados os verdadeiros positivos (VP), correspondentes aos grãos corretamente detectados; os falsos positivos (FP), referentes às detecções incorretas; e os falsos negativos (FN), correspondentes aos grãos presentes nas imagens que não foram detectados pelo modelo. A precisão, a revocação e o F1-score foram calculados conforme as Equações 1, 2 e 3, respectivamente:

$$P = \frac{VP}{VP + FP} \quad (1)$$

$$R = \frac{VP}{VP + FN} \quad (2)$$

$$F_1 = \frac{2 \times (P \times R)}{P + R} \quad (3)$$

$$mAP = \frac{1}{C} \sum_{i=1}^C AP_i \quad (4)$$

Em que P corresponde à precisão, R à revocação, VP aos verdadeiros positivos, FP aos falsos positivos, FN aos falsos negativos e C ao número de classes avaliadas. A *Average Precision* (AP) corresponde à área sob a curva precisão-revocação, enquanto a *mean Average Precision* (mAP) representa a média dos valores de AP obtidos para as classes avaliadas. Neste trabalho, foi considerada a mAP@50, utilizando um limiar de *Intersection over Union* (IoU) de 0,50 para estabelecer a correspondência entre as predições do modelo e as anotações de referência.

A avaliação final do desempenho dos modelos foi realizada utilizando o conjunto de teste, composto por imagens não utilizadas durante as etapas de treinamento e validação. As métricas obtidas nesse conjunto foram utilizadas para comparar o desempenho dos modelos nos quatro cenários avaliados.

7 RESULTADOS E DISCUSSÃO

7.1 Desempenho dos modelos nos diferentes cenários

Os valores de mAP@50 obtidos para os quatro cenários avaliados são apresentados na Tabela 3. O desempenho variou de **26,6% a 34,4%**, com o maior valor observado no cenário 4, composto por imagens dos quatro fundos de captura.

Tabela 3 - Precisão média (mAP@50) obtida para os diferentes cenários.

Cenário	Fundos utilizados no treinamento	mAP@50 (%)
1	EVA preto e EVA azul	27,3%
2	EVA preto, EVA azul e solo	26,6%
3	EVA preto, EVA azul e granito	33,3%
4	EVA preto, EVA azul, solo e granito	34,4%

Fonte: Elaborada pelo autor (2026).

O cenário 1, composto pelos fundos de EVA preto e EVA azul, apresentou mAP@50 de 27,3%. A inclusão das imagens obtidas sobre solo, no cenário 2, resultou em mAP@50 de 26,6%, correspondendo ao menor valor entre os cenários avaliados. Por outro lado, a inclusão do granito no cenário 3 elevou a mAP@50 para 33,3%. O maior valor foi observado no cenário 4, que reuniu os quatro fundos de captura e apresentou mAP@50 de 34,4%, correspondendo a um acréscimo de 7,1 pontos percentuais em relação ao cenário 1. A comparação entre os cenários contendo os fundos naturais apresentou comportamento distinto. A inclusão das imagens obtidas sobre o solo, no cenário 2, resultou em mAP@50 de 26,6%, enquanto o cenário 3, que incluiu o granito, apresentou mAP@50 de 33,3%. Uma possível explicação para essa diferença está relacionada às características visuais dos fundos e ao contraste entre os grãos e a superfície de captura.

Visualmente, o solo apresenta tonalidades próximas às observadas nos grãos de arroz, além de maior heterogeneidade decorrente da presença de partículas de diferentes tamanhos, pequenas pedras e variações de textura. Esses elementos podem reduzir a distinção visual entre os grãos e o fundo e aumentar a complexidade da imagem. Em contraste, o granito utilizado apresentou superfície visualmente mais uniforme e tonalidade distinta dos grãos, o que pode ter favorecido sua diferenciação pelo modelo. A similaridade entre objeto e fundo pode interferir no desempenho de modelos de detecção (DENG et al., 2021). Entretanto, como não foram realizadas análises quantitativas de cor, textura ou contraste, essas características foram avaliadas apenas visualmente e, portanto, não permitem estabelecer uma relação causal com as diferenças de desempenho observadas. Embora a inclusão isolada das imagens obtidas sobre o solo não tenha resultado em aumento da mAP@50 em relação ao cenário de referência, o cenário 4, composto pelos quatro fundos, apresentou o maior valor para essa métrica. Esse comportamento sugere que o desempenho observado pode estar relacionado não apenas às características de um

fundo específico, mas também à maior diversidade visual presente no conjunto de treinamento. A exposição do modelo a diferentes condições de fundo pode favorecer o aprendizado de características associadas ao objeto de interesse, reduzindo sua dependência de elementos específicos do cenário de captura (ZOU et al., 2023). Entretanto, os resultados obtidos não permitem afirmar que a maior diversidade de fundos resultou em maior capacidade de generalização, sendo necessária uma avaliação independente e controlada para verificar essa hipótese.

Ressalta-se, ainda, que cada condição de fundo foi constituída por panículas distintas, de modo que as diferenças observadas entre os cenários não podem ser atribuídas exclusivamente às características dos fundos, uma vez que também existe variabilidade entre as amostras que compõem cada conjunto (LIU et al., 2025). Além disso, a inclusão progressiva de novos fundos foi acompanhada pelo aumento do número de imagens disponíveis para o treinamento, constituindo outro fator que deve ser considerado na interpretação dos resultados.

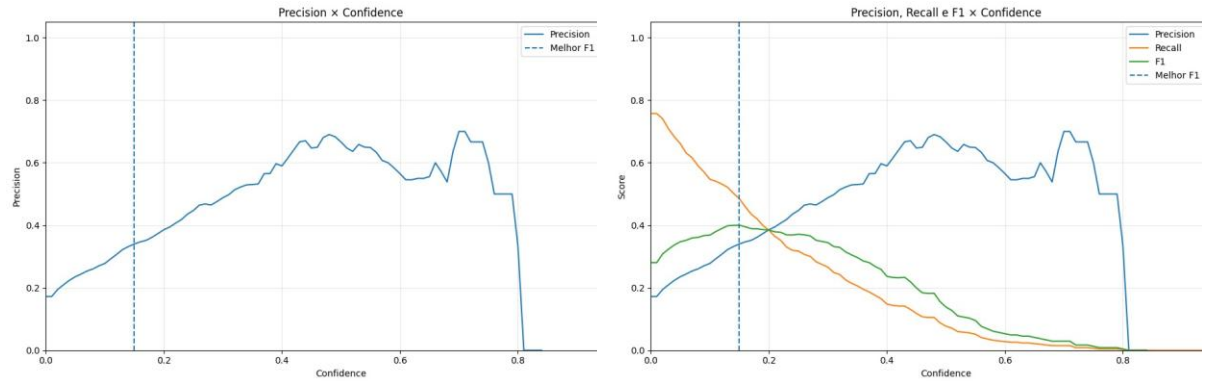
A utilização de fundos visualmente mais heterogêneos também pode aumentar a variabilidade das características presentes nas imagens e, conseqüentemente, a complexidade do conjunto de dados utilizado para o treinamento. Considerando o número de imagens disponível neste estudo, **é possível que** essa maior variabilidade não tenha sido suficientemente representada no conjunto de treinamento. Esse aspecto pode estar relacionado ao comportamento observado nas métricas dos diferentes cenários, incluindo o desempenho do cenário composto apenas pelos fundos artificiais, que apresentavam maior contraste visual entre os grãos e o fundo (Tabela 3). Dessa forma, estudos futuros devem avaliar conjuntos maiores e balanceados de imagens, permitindo investigar de maneira mais controlada a interação entre a diversidade dos fundos, o tamanho do conjunto de treinamento e o desempenho do modelo.

7.2 Comportamento detalhado das métricas

Para cada cenário de treinamento, foram gerados três gráficos que permitem avaliar o comportamento do modelo em diferentes níveis de confiança. O primeiro apresenta a relação entre precisão e confiança. O segundo reúne as curvas de precisão, revocação e F1 em função da confiança. O terceiro apresenta a curva precisão–revocação, cuja área está relacionada à precisão média (AP) (ZOU et al.,

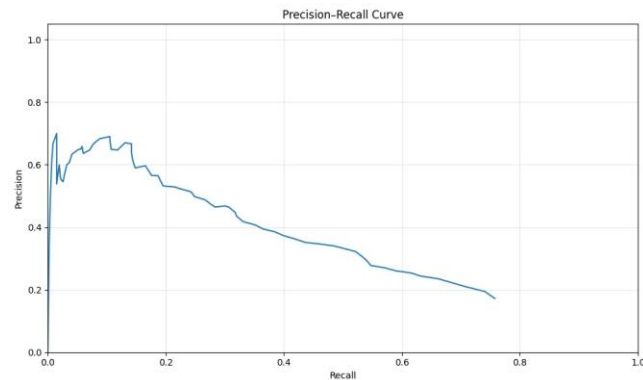
2023). As figuras a seguir apresentam esses gráficos de forma conjunta, organizados em um bloco para cada um dos quatro cenários avaliados.

Figura 2 - Gráficos do cenário 1 (EVA preto e EVA azul), com $mAP@50$ de 27,3%



(a) Precisão × Confiança

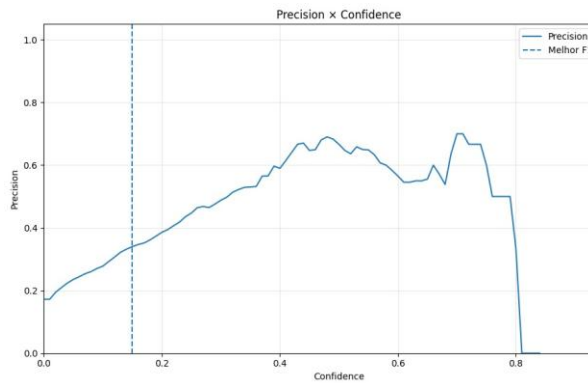
(b) Precisão, Revocação e F1 × Confiança



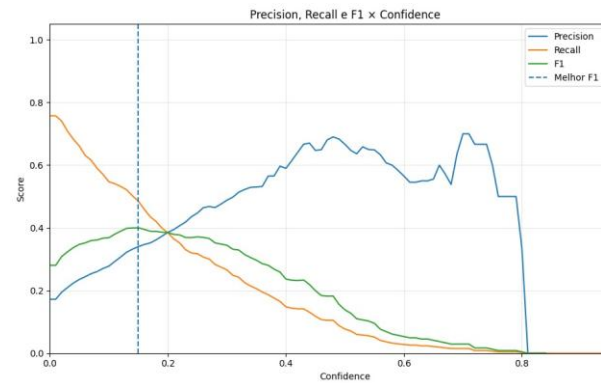
(c) Curva de Precisão e Revocação

Fonte: Elaborada pelo autor (2026).

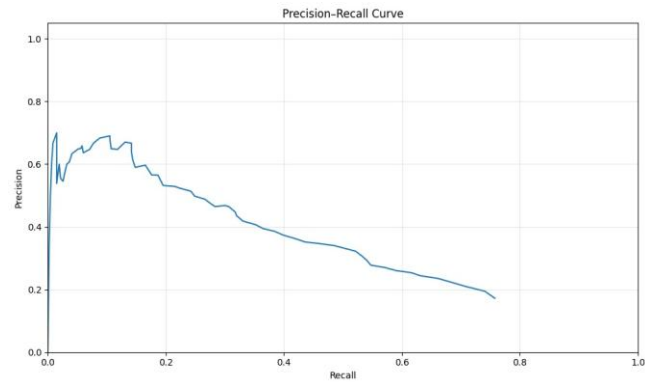
Figura 3 - Gráficos do cenário 2 (EVA preto, EVA azul e solo), com mAP@50 de 26,6%



(a) Precisão × Confiança



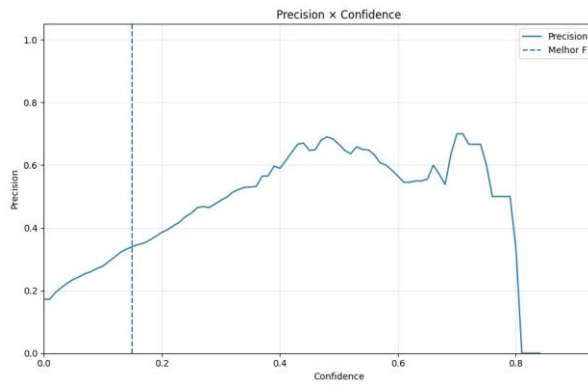
(b) Precisão, Revocação e F1 × Confiança



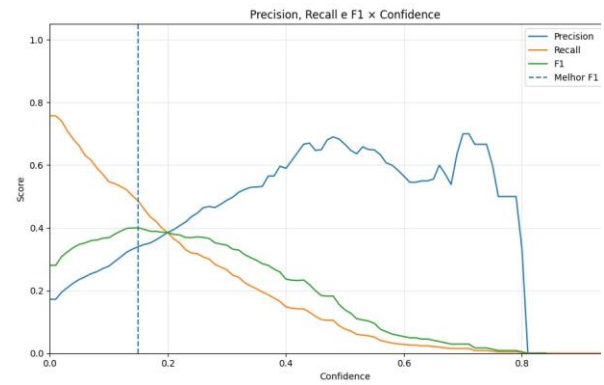
(c) Curva de Precisão e Revocação

Fonte: Elaborada pelo autor (2026).

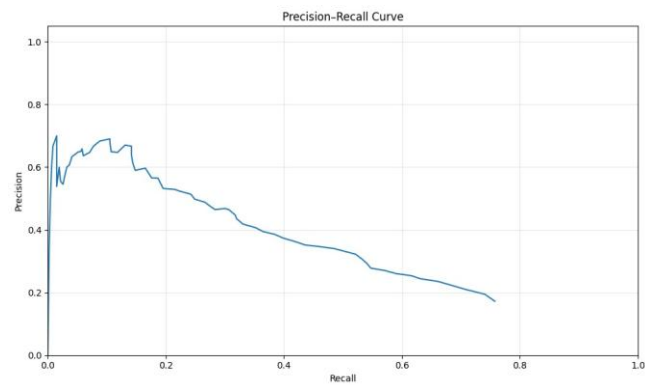
Figura 4 - Gráficos do cenário 3 (EVA preto, EVA azul e granito), com mAP@50 de 33,3%



(a) Precisão × Confiança



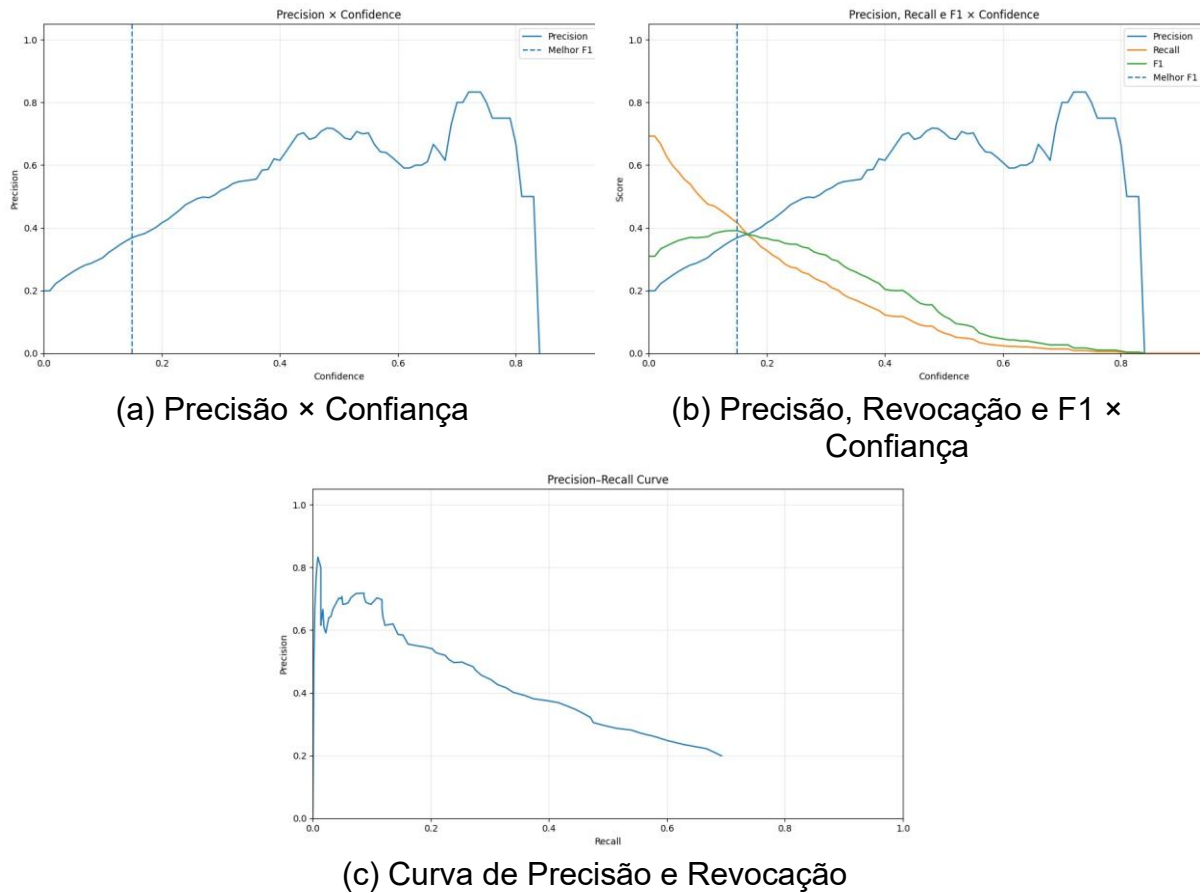
(b) Precisão, Revocação e F1 × Confiança



(c) Curva de Precisão e Revocação

Fonte: Elaborada pelo autor (2026).

Figura 5 - Gráficos do cenário 4 (EVA preto, EVA azul, solo e granito), com mAP@50 de 34,4%



Fonte: Elaborada pelo autor (2026).

No cenário 1, composto pelos fundos de EVA preto e EVA azul, observa-se aumento da precisão à medida que o limiar de confiança se eleva, acompanhado pela redução progressiva da revocação. Esse comportamento demonstra o compromisso entre as duas métricas: ao aumentar o limiar, o modelo se torna mais seletivo, descartando detecções de menor confiança. Como consequência, aumenta a proporção de detecções corretas entre aquelas mantidas, enquanto parte dos objetos verdadeiros deixa de ser detectada. O F1 apresenta seu maior valor em uma faixa de baixa confiança, indicando que o melhor equilíbrio entre precisão e revocação ocorre antes dos limiares que proporcionam os maiores valores de precisão.

No cenário 2, que acrescenta o fundo de solo aos fundos de EVA preto e EVA azul, as curvas apresentam comportamento semelhante ao observado no cenário 1. A precisão aumenta com a elevação do limiar de confiança, enquanto a revocação diminui progressivamente. O F1 também alcança seu máximo em uma faixa de baixa confiança, indicando que a inclusão do solo não alterou de forma expressiva o padrão de resposta das métricas ao aumento do limiar de confiança.

No cenário 3, composto pelos fundos de EVA preto, EVA azul e granito, mantém-se a relação inversa entre precisão e revocação. O aumento do limiar de confiança torna as detecções mais seletivas, resultando em maior precisão e menor revocação. Assim como nos cenários anteriores, o F1 apresenta seu melhor desempenho em limiares relativamente baixos, mostrando que a obtenção de valores elevados de precisão em limiares mais restritivos ocorre acompanhada de perda na capacidade de recuperar os objetos presentes nas imagens.

No cenário 4, que reúne os quatro fundos avaliados, EVA preto, EVA azul, solo e granito, observa-se o mesmo comportamento geral, com aumento da precisão e redução da revocação à medida que o limiar de confiança se eleva. Nesse cenário, a precisão alcança valores mais elevados em limiares de confiança altos; entretanto, esses valores estão associados à redução acentuada da revocação. O maior valor de F1 ocorre em uma faixa de confiança mais baixa, reforçando que o ponto de maior precisão não corresponde necessariamente ao melhor equilíbrio entre a identificação correta dos objetos e a capacidade de recuperar os objetos presentes.

De forma conjunta, os quatro cenários apresentam padrão semelhante de resposta às alterações no limiar de confiança. Em todos eles, o aumento desse limiar favorece a precisão e reduz a revocação, evidenciando o compromisso existente entre essas métricas. Dessa forma, valores elevados de precisão observados em limiares mais restritivos devem ser interpretados em conjunto com a redução da revocação, uma vez que podem representar um modelo mais seletivo, mas que deixa de identificar uma parcela maior dos objetos. Nesse sentido, o F1 auxilia na identificação de uma faixa de confiança que proporciona maior equilíbrio entre precisão e revocação.

7.3 Discussão dos resultados frente à literatura

A análise conjunta dos resultados evidencia o compromisso entre precisão e revocação observado nos diferentes cenários de treinamento. O aumento da precisão em determinados limiares de confiança ocorre acompanhado pela redução da revocação, comportamento característico de modelos de detecção de objetos (ZOU et al., 2023). Dessa forma, a definição do limiar de confiança deve considerar a finalidade da aplicação, uma vez que limiares mais restritivos podem reduzir falsas detecções, mas também aumentar a ocorrência de objetos não detectados. Os valores

obtidos neste trabalho foram inferiores aos relatados em estudos anteriores envolvendo a detecção e a contagem de grãos de arroz. Deng et al. (2021) e Zou et al. (2023), por exemplo, reportaram elevados desempenhos em suas respectivas abordagens. Entretanto, a comparação direta entre esses valores deve ser realizada com cautela, considerando tanto as métricas utilizadas quanto as diferenças metodológicas entre os estudos. As condições de aquisição das imagens, o tamanho e a diversidade do conjunto de dados, a arquitetura empregada e as características dos objetos avaliados podem influenciar diretamente o desempenho dos modelos.

Entre esses fatores, destaca-se a condição de captura das imagens. Estudos realizados em condições controladas e com fundos que proporcionam maior contraste entre os objetos e o ambiente tendem a reduzir a complexidade da tarefa de detecção (ZOU et al., 2023; ZHANG et al., 2024). No presente trabalho, por outro lado, foram avaliados cenários contendo fundos artificiais e naturais, aumentando a variabilidade visual apresentada ao modelo. A similaridade entre objeto e fundo pode dificultar a identificação dos grãos, especialmente em situações de menor contraste, constituindo uma fonte potencial de erros de detecção (DENG et al., 2021; LIU et al., 2025).

Outro aspecto a ser considerado é o conjunto de dados utilizado para o treinamento. Embora o aumento de dados tenha ampliado o número de imagens disponíveis, a quantidade e, principalmente, a diversidade das imagens originais permanecem fatores importantes para a aprendizagem e a capacidade de generalização do modelo. Além disso, neste trabalho foi empregado o YOLOv11 sem modificações estruturais específicas para a detecção dos grãos, enquanto estudos anteriores utilizaram estratégias e adaptações direcionadas à detecção de objetos pequenos e próximos entre si (ZOU et al., 2023; LIU et al., 2025). Essas diferenças metodológicas também devem ser consideradas na interpretação das diferenças de desempenho entre os trabalhos.

As características dos próprios grãos representam outro desafio para a detecção. Por ocuparem pequena proporção da imagem e ocorrerem em grande número e próximos uns dos outros, os grãos podem dificultar tanto sua localização quanto sua individualização pelo modelo. Além disso, variações de forma, textura, coloração e contraste em relação ao fundo aumentam a complexidade da tarefa de detecção (ZOU et al., 2023). Esses fatores ajudam a compreender as limitações observadas no presente estudo, sem atribuir o desempenho a uma única causa.

Apesar dessas limitações, os resultados demonstram que a composição dos fundos utilizados no treinamento influencia o desempenho do modelo. A comparação entre os quatro cenários mostrou que a inclusão de novos fundos não promoveu melhoria uniforme, uma vez que o efeito variou de acordo com o fundo incorporado ao conjunto de treinamento. Esse comportamento reforça a importância da relação entre as características visuais do objeto e do fundo, visto que situações de menor contraste ou maior similaridade podem aumentar a dificuldade de detecção (DENG et al., 2021; LIU et al., 2025).

Assim, a escolha dos fundos de captura deve ser considerada uma etapa importante na construção do conjunto de dados e definida de acordo com a finalidade de aplicação do modelo. Além disso, comparações entre estudos devem considerar as condições de aquisição das imagens e as características dos respectivos conjuntos de dados, uma vez que desempenhos obtidos em condições controladas não necessariamente representam o comportamento do modelo diante de maior variabilidade de fundos (ZHANG et al., 2024).

8 CONSIDERAÇÕES FINAIS

O modelo baseado no YOLOv11 apresentou capacidade de detectar grãos em panículas de arroz, embora o desempenho obtido indique a necessidade de aprimoramento para sua aplicação como ferramenta automatizada de detecção.

Os diferentes fundos utilizados no treinamento influenciaram o desempenho do modelo. A inclusão do solo não proporcionou melhoria em relação ao cenário com fundos artificiais, enquanto a inclusão do granito e a combinação dos quatro fundos resultaram em melhor desempenho, sendo o maior mAP@50 observado no cenário com todos os fundos. Entretanto, os resultados não permitem afirmar que a maior diversidade de fundos aumentou a capacidade de generalização do modelo, uma vez que essa avaliação exigiria uma validação independente.

A análise das métricas de precisão, revocação e F1 também demonstrou que o desempenho varia conforme o limiar de confiança adotado, evidenciando o equilíbrio entre a precisão das detecções e a capacidade de identificar os grãos presentes nas imagens.

Dessa forma, conclui-se que o fundo de captura é um fator relevante para o desempenho do modelo, mas que novos estudos, com maior número e diversidade

de imagens e validação em condições independentes, são necessários para aprimorar o modelo e verificar sua capacidade de generalização.

REFERÊNCIAS

- AMELONG, A. et al. Predicting maize kernel number using QTL information. *Field Crops Research*, Amsterdam, v. 172, p. 119-131, 2015.
- CROSSA, J. et al. The modern plant breeding triangle: optimizing the use of genomics, phenomics, and enviromics data. *Frontiers in Plant Science*, Lausanne, v. 12, art. 651480, 2021.
- DENG, R. et al. Automated counting grains on the rice panicle based on deep learning method. *Sensors*, Basel, v. 21, n. 1, art. 281, 2021.
- DUVICK, D. N. The contribution of breeding to yield advances in maize (*Zea mays* L.). *Advances in Agronomy*, San Diego, v. 86, p. 83-145, 2005.
- GALVÃO, J. C. C. et al. Sete décadas de evolução do sistema produtivo da cultura do milho. *Revista Ceres*, Viçosa, v. 61, supl., p. 819-828, 2014.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep learning*. Cambridge: MIT Press, 2016.
- JOCHER, G.; QIU, J. Ultralytics YOLO11. Versão 11.0.0. [S. l.]: Ultralytics, 2024. Disponível em: <https://github.com/ultralytics/ultralytics>. Acesso em: 11 ago. 2026.
- KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. ImageNet classification with deep convolutional neural networks. In: *ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS (NIPS)*, 2012, Lake Tahoe. *Proceedings [...]*. [S. l.]: Curran Associates, 2012. v. 25, p. 1097-1105.
- LECUN, Y. et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, Piscataway, v. 86, n. 11, p. 2278-2324, 1998.
- LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *Nature*, London, v. 521, n. 7553, p. 436-444, 2015.
- LIU, F. et al. SmartPod: an automated framework for high-precision soybean pod counting in field phenotyping. *Agronomy*, Basel, v. 15, n. 4, art. 791, 2025.
- MENDES, C. A. et al. Análise de associação quanto à produtividade e seus caracteres componentes em linhagens e cultivares de arroz de terras altas. *Pesquisa Agropecuária Brasileira*, Brasília, v. 49, n. 10, p. 771-782, 2014. DOI: 10.1590/S0100-204X2014001000004.
- PENG, X. et al. A field rice panicle detection model based on improved YOLOv11x. *Frontiers in Plant Science*, Lausanne, v. 16, art. 1656505, 2025. DOI: 10.3389/fpls.2025.1656505.
- PRASANNA, B. M. et al. Genetic trends in CIMMYT's tropical maize breeding pipelines. *Scientific Reports*, London, v. 12, art. 20110, 2022.
- REDMON, J. et al. You only look once: unified, real-time object detection. In: *IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR)*, 2016, Las Vegas. *Proceedings [...]*. Piscataway: IEEE, 2016. p. 779-788.
- ROBOFLOW. Roboflow: computer vision platform. [S. l.]: Roboflow, 2026. Disponível em: <https://roboflow.com>. Acesso em: 11 ago. 2026.

SUN, B. et al. Universal detection of curved rice panicles in complex environments using aerial images and improved YOLOv4 model. *Frontiers in Plant Science*, Lausanne, v. 13, art. 1021398, 2022. DOI: 10.3389/fpls.2022.1021398.

TAN, S. et al. OE-YOLO: an EfficientNet-based YOLO network for rice panicle detection. *Plants*, Basel, v. 14, n. 9, art. 1370, 2025. DOI: 10.3390/plants14091370.

WANG, H.; CHEN, J. Maize ear object detection method based on improved YOLOv8 for intelligent breeding unmanned vehicle. In: *CCF CHINESE CONFERENCE ON COMPUTER VISION (CCCV)*, 2024. *Communications in Computer and Information Science*, Singapore, v. 2449, p. 354-366, 2025.

WANG, X. et al. Field rice panicle detection and counting based on deep learning. *Frontiers in Plant Science*, Lausanne, v. 13, art. 966495, 2022. DOI: 10.3389/fpls.2022.966495.

ZHANG, Q.-Y. et al. High-precision automated soybean phenotypic feature extraction based on deep learning and computer vision. *Plants*, Basel, v. 13, n. 18, art. 2613, 2024.

ZOU, Y. et al. Rice grain detection and counting method based on TCLE-YOLO model. *Sensors*, Basel, v. 23, n. 22, art. 9129, 2023.

ANEXO A - PASSO A PASSO PARA ANOTAÇÃO NO ROBOFLOW

Documento elaborado por Lorena Queiroz, com base na experiência prática com o uso do Roboflow em projetos de anotação de imagens de grãos de arroz. Reproduzido com autorização.

Objetivo

Realizar anotações em imagens de panículas de arroz, contornando os grãos de forma precisa e padronizada, para treinar ou validar um modelo de contagem automática.

1 Acessar a plataforma

Acessar o endereço app.roboflow.com e fazer login na conta do projeto.

2 Abrir o projeto correto

Abrir o projeto destinado às anotações manuais e, dentro dele, entrar na aba Annotate. Em seguida, selecionar uma imagem para começar.

3 Marcar os grãos

Selecionar a ferramenta de marcação e contornar cada grão. O recurso de zoom, disponível no canto inferior esquerdo da tela, ajuda a visualizar os grãos de perto e a marcar com mais precisão.

A marcação deve pegar o grão por completo, de modo que as bordas também fiquem dentro da área marcada. Se a marcação acabar encostando em um grão vizinho, não há problema, desde que não se inclua outro grão inteiro dentro dela.

4 Usar a anotação assistida

Para acelerar o trabalho, é possível acionar o Label Assist, que aplica um modelo já treinado e sugere as marcações. Basta escolher o projeto e o modelo desejados e confirmar.

Dois parâmetros precisam ser definidos. A confiança determina o quanto o modelo precisa estar seguro para sugerir uma marcação, de forma que valores mais altos deixam apenas as sugestões mais confiáveis. A sobreposição evita que o mesmo grão seja marcado mais de uma vez.

5 Revisar e ajustar

Depois que as sugestões aparecem, é necessário revisar uma a uma. Com o zoom é possível conferir os detalhes, e clicando sobre uma marcação dá para movê-la, redimensioná-la ou excluí-la.

6 Lidar com variações de posição

As panículas podem aparecer na vertical, na horizontal ou na diagonal, o que faz com que as marcações fiquem em ângulos diferentes e mais próximas umas das outras. Em caso de dúvida sobre como marcar uma imagem específica, recomenda-se comparar com imagens semelhantes já anotadas, para manter o mesmo padrão.

7 Salvar e avançar

Após a revisão, salvar a imagem e avançar para a próxima usando as setas laterais.

Tabela 4 - Resumo das boas práticas de anotação

Tarefa	Orientação
Marcação do grão	Contornar o grão por completo, mesmo que encoste em outro
Grãos encostados	Aceitável, desde que não inclua um segundo grão inteiro
Ferramenta de apoio	Label Assist, com confiança e sobreposição definidas
Revisão	Conferir e ajustar cada marcação sugerida pelo modelo
Dúvida de orientação	Comparar com imagens semelhantes já anotadas
Encerramento	Salvar a imagem e avançar para a próxima

Fonte: Adaptado de Queiroz (2026).