



DETECÇÃO E CLASSIFICAÇÃO DE GRÃOS DE MILHO COM YOLOV8N: UMA ANÁLISE DE DESEMPENHO E REPRODUTIBILIDADE

Gabriel Scherer Souza¹, Alberto Yoshihiro Nakano²

¹Universidade Tecnológica Federal do Paraná (UTFPR), Toledo, Brasil (gscherersouza@gmail.com)

²Universidade Tecnológica Federal do Paraná (UTFPR), Toledo, Brasil

Resumo: Este trabalho avalia o desempenho e a reprodutibilidade da variante nano do YOLOv8 (YOLOv8n) na detecção e classificação binária de grãos de milho (*Zea mays*) em sadios e defeituosos. O modelo foi treinado por 200 épocas em cinco execuções independentes, alcançando mAP@0.5 de $0,906 \pm 0,014$ e *accuracy* de $0,901 \pm 0,011$, com coeficiente de variação inferior a 3%, evidenciando estabilidade e viabilidade para aplicação em dispositivos de borda (*edge device*) na inspeção automatizada de grãos.

Palavras-chave: YOLOv8n; Detecção de objetos; Classificação de grãos; Milho; Reprodutibilidade.

INTRODUÇÃO

A crescente demanda mundial por alimentos posiciona o Brasil como um ator estratégico no cenário agrícola global. O agronegócio nacional constitui um dos principais sustentáculos da economia do país, com impactos expressivos no Produto Interno Bruto (PIB), na geração de empregos e no desempenho da balança comercial. Entre as culturas de maior relevância nesse contexto, o milho se sobressai pela sua amplitude produtiva e pela diversidade de aplicações na cadeia agroindustrial (ARTUZO et al., 2019).

Estimativas para a safra 2025/26 apontam que a produção brasileira total de grãos deve alcançar 353,4 milhões de toneladas, volume 0,3% superior ao registrado na safra anterior (CONAB, 2026). Diante dessa relevância econômica, garantir a qualidade do grão ao longo da cadeia produtiva torna-se uma exigência tanto técnica quanto comercial. A classificação do milho é regulamentada pelo Ministério da Agricultura, Pecuária e Abastecimento (MAPA) e constitui etapa obrigatória nos processos de recebimento e expedição nas unidades armazenadoras (MAPA, 2011).

Tradicionalmente, a classificação dos grãos de milho é executada de forma manual, por meio da inspeção visual realizada por operadores treinados que avaliam amostras representativas do lote conforme os parâmetros estabelecidos pela legislação. Embora consolidado, esse procedimento apresenta limitações relevantes, como a subjetividade inerente ao avaliador, a variabilidade entre diferentes operadores, a fadiga visual decorrente de jornadas prolongadas e a baixa produtividade frente aos elevados volumes movimentados

nas unidades armazenadoras (MAPA, 2011). Tais limitações motivam a busca por métodos automatizados capazes de conferir maior objetividade, repetibilidade e escalabilidade ao processo.

A Figura 1 apresenta, de forma simplificada, o fluxo de classificação manual estabelecido pela Instrução Normativa MAPA nº 60/2011. Após a recepção, amostragem e preparo da amostra, os grãos íntegros são submetidos à inspeção visual (etapa 3), na qual o operador avalia cor, textura, integridade e formato de cada grão. Essa etapa concentra a maior carga de subjetividade do processo e representa o foco de automação proposto neste estudo. As etapas subsequentes envolvem a separação dos defeitos conforme a normativa, a pesagem das frações e a consolidação do laudo técnico.

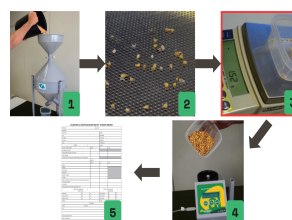


Figura 1: Etapas do processo de inspeção dos grãos simplificado.

A automação desse processo por meio de técnicas de Visão Computacional (VC) tem ganhado crescente espaço como alternativa à inspeção visual tradicional, propiciando maior velocidade, padronização e rastreabilidade nas avaliações (ZHAO et al., 2019).



Nesse contexto, arquiteturas de Aprendizado Profundo (*Deep Learning* — DL) voltadas à detecção de objetos têm se destacado pela capacidade de identificar e localizar instâncias de interesse em imagens com elevada precisão (ALZUBAIDI et al., 2021).

Entre as arquiteturas mais consolidadas para essa finalidade, destaca-se a família *You Only Look Once* (YOLO), que realiza a detecção em uma única passagem pela rede neural, conferindo alta eficiência computacional sem comprometer a acurácia (REDMON et al., 2016). Embora versões mais recentes tenham sido propostas, o YOLOv8, desenvolvido pela Ultralytics (JOCHER; CHAURASIA; QIU, 2023), permanece como uma das alternativas mais amplamente utilizadas na literatura, em razão de sua estabilidade, documentação consolidada e ampla validação experimental. Essa versão disponibiliza variantes de diferentes portes, incluindo a *nano* (YOLOv8n), projetada para ambientes com recursos computacionais restritos (SOHAN; RAM; REDDY, 2024). A aplicação dessas arquiteturas em culturas agrícolas tem demonstrado resultados promissores em tarefas de detecção e controle de qualidade (SÁ et al., 2024).

Nesse contexto, o presente trabalho propõe a aplicação da variante *nano* do YOLOv8 para a detecção e classificação binária de grãos de milho (*Zea mays*), distinguindo grãos sadios de grãos defeituosos — agrupando, nesta última classe, os diferentes tipos de defeito previstos pelo MAPA. O modelo foi treinado sobre o conjunto de imagens disponibilizado por Ferreira (2023a), composto por amostras com diferentes proporções de defeitos simulados. A abordagem privilegia a viabilidade de implantação em ambientes com infraestrutura computacional limitada, característica relevante para cooperativas e unidades armazenadoras de pequeno e médio porte.

Diante do exposto, este trabalho teve como objetivos:

- (i) treinar o modelo YOLOv8n sobre o conjunto de dados disponibilizado por (FERREIRA, 2023a);
- (ii) avaliar seu desempenho por meio das métricas de Precisão (*Precision*), Revocação (*Recall*), F1-Score e *Mean Average Precision* (mAP); e
- (iii) discutir a viabilidade da variante *nano* como solução computacionalmente eficiente para a classificação de grãos em ambientes com recursos limitados de *hardware*.

MATERIAL E MÉTODOS

Base de Dados

O conjunto de dados utilizado foi desenvolvido por Ferreira (2023a), sendo composto por imagens RGB de amostras de grãos de milho da classe amarela (*Zea*

mays), capturadas por meio de um módulo ESP32-CAM. O material está disponível publicamente sob a licença *Creative Commons Attribution 4.0 International* (FERREIRA, 2023b).

As amostras contêm entre 50 e 100 grãos por imagem, com proporções de grãos defeituosos de 0%, 10%, 15%, 20%, 25% e 30%. Os defeitos incluem grãos ardidos, germinados e quebrados, representando condições típicas do processo de classificação comercial.

Anotação das Imagens

O processo de anotação foi conduzido na plataforma *Computer Vision Annotation Tool* (CVAT, 2026), ferramenta de código aberto amplamente adotada em tarefas de rotulagem para Visão Computacional. As imagens do conjunto de dados foram importadas para um projeto dedicado, no qual se definiram duas classes de interesse: **grão sadio** (classe 0) e **grão defeituoso** (classe 1), caracterizando uma abordagem de classificação alinhada aos critérios estabelecidos pela Instrução Normativa MAPA nº 60/2011.

A rotulagem foi realizada de forma manual, grão a grão, por meio do recurso de *bounding boxes* retangulares disponível na ferramenta. Cada caixa delimitadora foi posicionada de modo a circunscrever integralmente o contorno do grão, minimizando a inclusão de áreas de fundo e evitando sobreposições excessivas entre instâncias adjacentes. Grãos parcialmente visíveis nas bordas da imagem ou significativamente ocluídos por outros grãos foram desconsiderados, a fim de preservar a qualidade do conjunto de treinamento.

A atribuição das classes seguiu os critérios visuais definidos pela referida normativa: foram rotulados como *defeituosos* os grãos que apresentavam sinais de ardência, germinação ou quebra, enquanto os demais, íntegros e sem alterações aparentes, foram classificados como *sadios*. A Figura 2 ilustra o resultado do processo de anotação, evidenciando as *bounding boxes* delimitadoras e o sistema de coordenadas normalizadas adotado pelo formato YOLO, no qual a origem (0, 0) corresponde ao canto superior esquerdo da imagem e o ponto (1, 1) ao canto inferior direito.

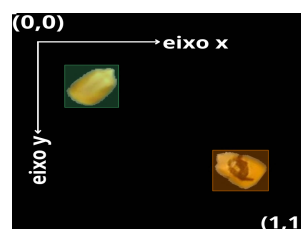
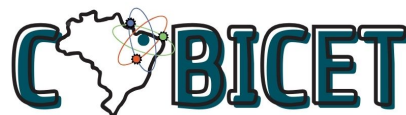


Figura 2: Exemplo de anotação realizada no CVAT, com *bounding boxes*.



Após a conclusão da anotação, o conjunto foi exportado no formato *YOLO 1.1*, que gera, para cada imagem, um arquivo de texto contendo a classe e as coordenadas normalizadas (x_center , y_center , $width$, $height$) de cada *bounding box*, padrão exigido pelo *framework* Ultralytics utilizado na etapa subsequente de treinamento.

Arquitetura YOLOv8

O *You Only Look Once* (YOLO) é uma família de modelos de detecção de objetos em estágio único (*one-stage detectors*), caracterizada por realizar, em uma única passagem da rede, a localização e a classificação dos objetos presentes na imagem. Essa abordagem confere ao YOLO um equilíbrio favorável entre acurácia e velocidade de inferência, tornando-o amplamente adotado em aplicações de tempo real (REDMON et al., 2016).

A versão YOLOv8, desenvolvida pela Ultralytics em 2023, introduz aprimoramentos relevantes em relação às versões anteriores, com destaque para a adoção de uma cabeça de detecção *anchor-free*, a qual prediz diretamente as coordenadas das *bounding boxes* sem depender de caixas-âncora pré-definidas, simplificando o processo de treinamento e melhorando a generalização do modelo (SOHAN; RAM; REDDY, 2024). Sua arquitetura é organizada em três componentes principais:

- **Backbone:** responsável pela extração hierárquica de características, baseado em uma variante do CSPDarknet com módulos C2f (*Cross Stage Partial with two convolutions*), que aprimoram o fluxo de gradiente e a eficiência computacional;
- **Neck:** realiza a fusão de características em múltiplas escalas por meio de uma estrutura do tipo PAN-FPN, que combina a *Feature Pyramid Network* (FPN) com a *Path Aggregation Network* (PAN), permitindo a detecção robusta de objetos de diferentes tamanhos;
- **Head:** composta por ramos desacoplados (*decoupled head*) para classificação e regressão das caixas delimitadoras, utilizando a função de perda *Distribution Focal Loss* (DFL) combinada com *Complete IoU* (CIoU) para refinamento das predições.

A família YOLOv8 disponibiliza cinco variantes — *nano* (n), *small* (s), *medium* (m), *large* (l) e *extra-large* (x) —, que se diferenciam pela profundidade e largura da rede, e, consequentemente, pelo número de parâmetros e custo computacional. Neste trabalho, optou-se pela variante **YOLOv8n** (*nano*), com-

posta por aproximadamente 3,2 milhões de parâmetros, por apresentar a menor demanda computacional da família e, portanto, viabilizar futura aplicação em dispositivos de borda (*edge devices*) com recursos limitados de *hardware*, condição alinhada ao módulo ESP32-CAM utilizado na aquisição das imagens por Ferreira (2023a).

Configuração do Treinamento

O treinamento foi conduzido na plataforma Google Colaboratory (GOOGLE, 2026), com suporte a GPU, utilizando o *framework* Ultralytics com a variante nano do YOLOv8 (YOLOv8n), por meio da técnica de *transfer learning* (TAN et al., 2018), a partir de pesos pré-treinados. A escolha dessa variante se justifica pela sua leveza computacional, tornando-a adequada para ambientes com recursos limitados de *hardware*.

O treinamento foi conduzido por 200 épocas com proporção de divisão dos dados de treino e validação de 80/20. Para garantir a robustez estatística dos resultados e avaliar a reprodutibilidade do modelo, o experimento foi replicado em cinco execuções independentes ($n = 5$), utilizando diferentes sementes de inicialização aleatória dos pesos e mantendo-se inalterados os demais hiperparâmetros e a partição dos dados. As métricas finais são reportadas com média $\pm$ (μ) desvio padrão (σ) sobre as cinco execuções.

Métricas de Avaliação

O desempenho do modelo foi avaliado pelas métricas Precisão (*Precision*), Revocação (*Recall*), F1-Score e *Mean Average Precision* (mAP@0.5 e mAP@0.5:0.95), amplamente consolidadas na literatura para tarefas de detecção de objetos (PADILLA; NETTO; SILVA, 2020; EVERINGHAM et al., 2010; LIN et al., 2014).

Precision (Precisão) mede a proporção de detecções corretas dentre todas as detecções realizadas pelo modelo, sendo definida por

$$Precision = \frac{TP}{TP + FP}, \quad (1)$$

sendo que *TP* (*True Positives*) representa os verdadeiros positivos e *FP* (*False Positives*) os falsos positivos. Essa métrica é particularmente relevante em aplicações nas quais o custo associado a falsos alarmes é elevado.

Recall (Revocação), quantifica a capacidade do modelo de identificar todas as instâncias relevantes presentes nas imagens, conforme

$$Recall = \frac{TP}{TP + FN}, \quad (2)$$

sendo *FN* (*False Negatives*) o número de falsos negativos. Em contextos nos quais a omissão de detecções



compromete a aplicação, a maximização da revocação torna-se prioritária.

O F1-Score corresponde à média harmônica entre *Precision* e *Recall*, oferecendo uma medida equilibrada do desempenho do modelo quando ambas as métricas são igualmente importantes e é dada por

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (3)$$

A *Accuracy* (Acurácia) representa a proporção de predições corretas (positivas e negativas) sobre o total de instâncias avaliadas, dada por

$$Acurácia = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

sendo que *TN* (*True Negatives*) representa os verdadeiros negativos.

Por fim, o *Mean Average Precision* (mAP) constitui a métrica mais consolidada em *benchmarks* de detecção de objetos, tais como o PASCAL VOC (EVERINGHAM et al., 2010) e o MS COCO (LIN et al., 2014). O mAP é calculado a partir da área sob a curva Precisão–Revocação (*AP*) para cada classe, sendo posteriormente mediado entre as *N* classes do conjunto de dados dada por

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (5)$$

O **mAP@0.5** considera uma detecção como verdadeiro positivo quando a Interseção sobre União (*Intersection over Union* – IoU) entre a *bounding box* predita e a anotação de referência é igual ou superior a 0,5. Já o **mAP@0.5:0.95** corresponde à média do mAP calculado em dez limiares de IoU, variando de 0,5 a 0,95 em passos de 0,05, oferecendo uma avaliação mais rigorosa quanto à qualidade da localização espacial das detecções (LIN et al., 2014).

RESULTADOS E DISCUSSÃO

Nesta seção são apresentados e discutidos os resultados obtidos a partir do treinamento do modelo YOLOv8 em sua variante *nano* (YOLOv8n) ao longo de 200 épocas. Visando aumentar a robustez estatística, o experimento foi replicado em cinco execuções independentes. Inicialmente, analisa-se a evolução do treinamento; em seguida, discute-se o desempenho final consolidado e a estabilidade do modelo frente à literatura especializada.

Evolução do Treinamento

A Figura 3 apresenta a evolução das métricas mAP@0.5 e mAP@0.5:0.95 ao longo das 200 épocas

de treinamento, considerando a média e o desvio padrão obtidos a partir das cinco execuções independentes.

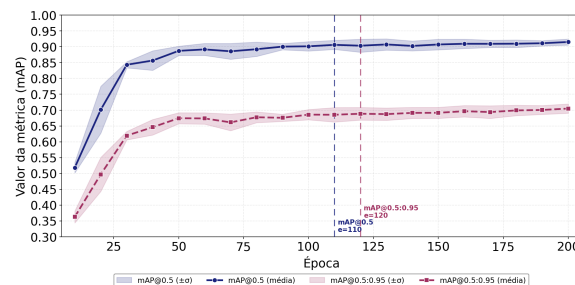


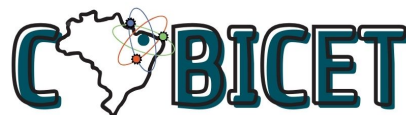
Figura 3: Evolução do mAP e detecção automática de convergência.

A análise da Figura 3 permite identificar três regimes distintos no processo de aprendizado do modelo. No regime inicial (épocas 10–50), observa-se um crescimento acentuado de ambas as métricas, com o mAP@0.5 evoluindo de aproximadamente 0,52 para 0,88 e o mAP@0.5:0.95 de 0,36 para 0,67. Esse comportamento é característico da fase de aprendizado de parâmetros de baixo nível, na qual o modelo ajusta rapidamente seus pesos para capturar representações visuais elementares, como bordas, texturas e gradientes locais (LECUN; BENGIO; HINTON, 2015). No regime intermediário (épocas 50–100), verifica-se uma desaceleração progressiva da taxa de aprendizado, associada ao refinamento das representações semânticas de alto nível e ao ajuste fino dos parâmetros responsáveis pela localização precisa dos objetos. Finalmente, no regime de convergência (a partir da época 110 para o mAP@0.5 e da época 120 para o mAP@0.5:0.95), as curvas exibem um platô bem definido, com ganhos marginais inferiores a 0,5% até o término do treinamento na época 200.

A convergência mais tardia da métrica mAP@0.5:0.95 em relação ao mAP@0.5 é consistente com a literatura, uma vez que aquela exige maior precisão de localização ao avaliar o desempenho sobre múltiplos limiares de IoU (0,50 a 0,95), demandando, portanto, um período adicional de ajuste fino dos parâmetros do detector (LIN et al., 2014).

Adicionalmente, observa-se uma redução progressiva do desvio padrão ao longo do treinamento, indicando que as cinco execuções independentes convergem para um mesmo regime de desempenho. Esse comportamento evidencia a estabilidade numérica e a reprodutibilidade do experimento, atributos desejáveis em estudos comparativos de modelos de aprendizado profundo (HENDERSON et al., 2018).

Um panorama completo da dinâmica do treinamento é apresentado na Figura 4, que reúne as quatro principais métricas de avaliação (*Precision*, *Recall*,



F1-Score e Accuracy) ao longo das épocas e os pontos de convergência identificados.

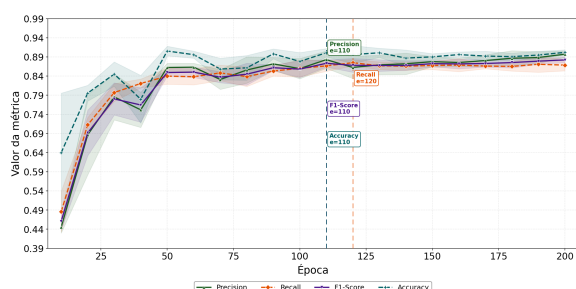


Figura 4: Evolução conjunta de *Precision*, *Recall*, *F1-Score* e *Accuracy* ao longo das épocas.

A análise da Figura 4 revela um comportamento particularmente interessante na fase final do treinamento: enquanto a *Precision* apresenta uma leve tendência ascendente nas últimas épocas, o *Recall* permanece estabilizado em torno de 0,87. Esse padrão sugere que, após a convergência inicial, o modelo passa a refinar prioritariamente sua capacidade de evitar falsos positivos, em detrimento de ganhos adicionais na detecção de instâncias verdadeiras. O *F1-Score* e a *Accuracy*, por sua vez, mantêm-se estáveis em patamares próximos a 0,88 e 0,90, respectivamente.

Observa-se ainda uma convergência harmônica entre *Precision*, *Recall* e *F1-Score* a partir da época 50, evidenciando o equilíbrio alcançado pelo modelo entre sensibilidade e especificidade. A *Accuracy*, calculada a partir da matriz de confusão, apresenta crescimento mais oscilatório nas épocas iniciais (10–50), com estabilização posterior em torno de 0,90 — comportamento coerente com o platô observado nas demais métricas. As épocas de convergência, destacadas pelas linhas verticais tracejadas, situam-se em $e = 110$ para *Precision*, *F1-Score* e *Accuracy*, e em $e = 120$ para o *Recall*, indicando que esta última necessita de um treinamento ligeiramente mais prolongado para atingir sua estabilidade.

Análise da Reprodutibilidade

Um dos principais aportes deste trabalho consiste na avaliação da estabilidade do YOLOv8n frente a cinco execuções independentes. A análise de cada execução individual evidencia que, embora as cinco execuções apresentem comportamentos ligeiramente distintos — sobretudo nas épocas iniciais, onde a variabilidade da inicialização exerce maior influência — todas convergem para uma faixa estreita de desempenho ao final do treinamento, com o $mAP@0.5$ situando-se entre 0,90 e 0,92. O desvio padrão observado nas épocas finais permanece consistentemente abaixo de 0,01, indicando elevada concentração dos resultados em torno

da média.

Esse comportamento reforça a robustez do YOLOv8n quanto à inicialização aleatória dos pesos e demonstra que o modelo oferece desempenho previsível em treinamentos sucessivos — característica desejável para aplicações práticas, nas quais a reprodutibilidade dos resultados é um requisito fundamental. A baixa dispersão entre execuções também sustenta a validade estatística das análises subsequentes, permitindo que as médias reportadas sejam interpretadas como estimativas confiáveis do desempenho esperado do modelo.

Desempenho Final do Modelo

Os valores finais das métricas, consolidados na melhor época de cada indicador e calculados a partir das cinco execuções independentes, são apresentados na Tabela 1.

Tabela 1: Métricas finais do YOLOv8n na melhor época individual ($\mu \pm \sigma$ sobre 5 execuções independentes).

Métrica	Época	Valor ($\mu \pm \sigma$)
$mAP@0.5$	110	$0,906 \pm 0,014$
$mAP@0.5:0.95$	120	$0,689 \pm 0,020$
<i>Precision</i>	110	$0,883 \pm 0,023$
<i>Recall</i>	120	$0,875 \pm 0,008$
<i>F1-Score</i>	110	$0,873 \pm 0,010$
<i>Accuracy</i>	110	$0,901 \pm 0,011$

A análise dos resultados revela três achados relevantes. Primeiro, os baixos valores de desvio padrão observados em $mAP@0.5$ ($\sigma = 0,014$), *F1-Score* ($\sigma = 0,010$), *Recall* ($\sigma = 0,008$) e *Accuracy* ($\sigma = 0,011$) confirmam a estabilidade do modelo entre execuções independentes, indicando que o YOLOv8n apresenta comportamento reprodutível mesmo sob diferentes inicializações aleatórias. Segundo, as métricas $mAP@0.5:0.95$ ($\sigma = 0,020$) e *Precision* ($\sigma = 0,023$) apresentam dispersão ligeiramente superior, refletindo a maior sensibilidade desses indicadores a variações na qualidade da localização espacial das *bounding boxes* e na taxa de falsos positivos, respectivamente. Terceiro, o coeficiente de variação ($CV = \sigma/\mu$) permanece inferior a 3% em todas as métricas avaliadas, evidenciando a robustez geral do modelo e validando a consistência dos resultados obtidos.

Discussão

Os resultados demonstram que o YOLOv8n, mesmo sendo a variante mais leve da família YOLOv8, é ca-



paz de atingir desempenho competitivo na tarefa proposta, com $mAP@0.5$ de $0,906 \pm 0,014$ e *accuracy* de $0,901 \pm 0,011$. Esse achado é particularmente relevante considerando-se o compromisso entre acurácia e custo computacional, dimensão amplamente discutida na literatura recente (A. . . ,).

A análise da evolução das métricas indica que o modelo atinge seu platô de desempenho a partir da época 110, aproximadamente, com ganhos marginais nas épocas subsequentes. Esse comportamento sugere que treinamentos mais longos podem não se justificar em termos de custo computacional, sendo recomendável a adoção de mecanismos de *early stopping* em aplicações práticas, com paciência da ordem de 15 a 20 épocas. Tal observação está em consonância com estudos que reportam saturação semelhante em arquiteturas leves de detecção (WANG; BOCHKOVSKIY; LIAO, 2023; LI et al., 2022).

A diferença esperada entre o $mAP@0.5$ ($\approx 0,91$) e o $mAP@0.5:0.95$ ($\approx 0,69$) é coerente com o comportamento típico de redes neurais de menor capacidade, nas quais a precisão da regressão das *bounding boxes* tende a ser inferior à de modelos mais robustos, como as variantes *small*, *medium*, *large* e *extra-large* do YOLOv8 (JOCHER; CHAURASIA; QIU, 2023).

Um dos principais aportes deste trabalho reside na demonstração empírica da reprodutibilidade do YOLOv8n. As cinco execuções independentes apresentaram desvio padrão inferior a 0,025 para todas as métricas consideradas, com coeficiente de variação inferior a 3%, indicando que o modelo apresenta comportamento previsível frente à variação da inicialização aleatória. Esse resultado é especialmente importante para validar a confiabilidade de relatos científicos baseados em execuções únicas — prática ainda recorrente na literatura, conforme alertam diversos autores no campo de aprendizado profundo (ENDERSON et al., 2018; PINEAU et al., 2021).

Em síntese, os resultados aqui apresentados indicam que o YOLOv8n constitui uma alternativa viável, eficiente e reprodutível para tarefas de detecção de objetos em cenários nos quais a eficiência computacional é um requisito relevante. A investigação da dinâmica de aprendizado, aliada à análise estatística sobre múltiplas execuções, oferece subsídios práticos para a definição de critérios de parada antecipada e para o dimensionamento adequado do treinamento em trabalhos futuros.

CONCLUSÃO

Este trabalho avaliou o desempenho da variante *nano* do YOLOv8 (YOLOv8n) na detecção e classificação binária de grãos de milho (*Zea mays*) da classe amarela, distinguindo grãos sadios de grãos defeituosos em conformidade com os critérios estabelecidos pela Instrução Normativa MAPA nº 60/2011. O modelo foi

treinado ao longo de 200 épocas sobre o conjunto de imagens disponibilizado por Ferreira (2023a), com validação cruzada, garantindo robustez estatística aos resultados reportados.

Os resultados obtidos demonstram que o YOLOv8n, mesmo sendo a variante de menor capacidade computacional da família YOLOv8, alcança desempenho competitivo na tarefa proposta, com $mAP@0.5$ de $0,906 \pm 0,014$, $mAP@0.5:0.95$ de $0,689 \pm 0,020$, F1-Score de $0,873 \pm 0,010$ e *accuracy* de $0,901 \pm 0,011$. A análise da dinâmica de treinamento revelou três fases distintas de aprendizado — crescimento acelerado, refinamento semântico e convergência —, com platô de desempenho estabelecido a partir da época 110, aproximadamente, e ganhos marginais nas épocas subsequentes. Esse achado sugere que estratégias de *early stopping*, com paciência da ordem de 15 a 20 épocas, podem ser empregadas sem prejuízo significativo ao desempenho final, reduzindo o custo computacional do treinamento.

A análise da reprodutibilidade — um dos principais aportes deste estudo — evidenciou desvio padrão inferior a 0,025 e coeficiente de variação inferior a 3% para todas as métricas avaliadas, atestando a estabilidade do modelo frente à variação da inicialização aleatória dos pesos. Esse comportamento previsível constitui um requisito essencial para a confiabilidade de aplicações práticas em ambientes produtivos, especialmente em unidades armazenadoras de pequeno e médio porte que demandam soluções automatizadas, padronizadas e rastreáveis. Adicionalmente, a metodologia adotada contribui para mitigar uma limitação recorrente na literatura de aprendizado profundo, na qual resultados são frequentemente reportados a partir de execuções únicas, sem a devida quantificação da incerteza experimental.

Conclui-se, portanto, que o YOLOv8n configura-se como uma alternativa tecnicamente viável, computacionalmente eficiente e estatisticamente reprodutível para a automação da etapa de inspeção visual no processo de classificação de grãos de milho. Como desdobramentos futuros, vislumbra-se: (i) a avaliação comparativa entre as demais variantes da família YOLOv8 (*small*, *medium*, *large* e *extra-large*), de modo a mapear o compromisso entre acurácia e custo computacional; (ii) a expansão do conjunto de dados com imagens adquiridas em condições reais de operação, contemplando variações de iluminação, fundo e granulometria; (iii) a investigação de estratégias de aumento de dados (*data augmentation*) específicas para a tarefa de classificação de grãos; e (iv) a implantação efetiva do modelo em dispositivos de borda, com destaque para o módulo ESP32-CAM, viabilizando aplicações de inspeção em tempo real diretamente nas unidades armazenadoras e contribuindo para a modernização tecnológica da cadeia produtiva



do milho no Brasil.

REFERÊNCIAS

A comprehensive review of yolo architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS.

ALZUBAIDI, L. et al. Review of deep learning: concepts, cnn architectures, challenges, applications, future directions. *Journal of big Data*, Springer, v. 8, p. 1–74, 2021.

ARTUZO, F. D. et al. O potencial produtivo brasileiro: uma análise histórica da produção de milho. *Revista em Agronegócio e Meio Ambiente*, v. 12, n. 2, p. 515–540, 2019.

CONAB. *Safra de Grãos está estimada em 353,4 milhões de toneladas no ciclo 2025/26*. 2026. Acesso em: 22 de maio de 2026. Disponível em: <<https://www.gov.br/conab/pt-br/assuntos/noticias/safra-de-graos-esta-estimada-em-353-4-milhoes-de-toneladas-no-ciclo-2025-26>>.

CVAT. *CVAT: Computer Vision Annotation Tool*. 2026. Disponível em: <<https://www.cvat.ai/>>. Acessado em: 30 de maio 2026.

EVERINGHAM, M. et al. The pascal visual object classes (VOC) challenge. *International journal of computer vision*, Springer, v. 88, p. 303–338, 2010.

FERREIRA, D. *Classificação de Grãos de Milho da Espécie Zea Mays da Classe Amarela Utilizando Processamento Digital de Imagens*. Trabalho de Conclusão de Curso (Bacharelado em Engenharia Eletrônica) — Universidade Tecnológica Federal do Paraná (UTFPR), Toledo, PR, Brasil, jun. 2023.

FERREIRA, D. *Repositório de imagens para classificação de grãos de milho*. 2023. <<https://github.com/ferreiradiego/maize-classification/tree/main/img/dataset>>. Acesso em: 23 de maio de 2025.

GOOGLE. *Google Colaboratory (Colab)*. 2026. Disponível em: <<https://colab.research.google.com/>>. Acessado em: 30 de maio 2026.

HENDERSON, P. et al. Deep reinforcement learning that matters. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. New Orleans, Louisiana, USA: AAAI Press, 2018. v. 32, n. 1, p. 3207–3214. Disponível em: <<https://ojs.aaai.org/index.php/AAAI/article/view/11694>>.

JOCHER, G.; CHAURASIA, A.; QIU, J. *Ultralytics YOLOv8*. 2023. Disponível em: <<https://docs.ultralytics.com/models/yolov8/>>. Acessado em: 30 de maio 2026.

JOCHER, G.; CHAURASIA, A.; QIU, J. *Ultralytics YOLOv8*. 2023. <<https://github.com/ultralytics/ultralytics>>. Versão 8.0.0. Acesso em: 21 maio 2026.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *nature*, Nature Publishing Group UK London, v. 521, n. 7553, p. 436–444, 2015.

LI, C. et al. YOLOv6: A single-stage object detection framework for industrial applications. *arXiv preprint arXiv:2209.02976*, 2022.

LIN, T.-Y. et al. Microsoft COCO: Common objects in context. In: *European Conference on Computer Vision (ECCV)*. [S.l.]: Springer, 2014. p. 740–755.

MAPA. *Instrução Normativa nº 60, de 22 de dezembro de 2011*. 2011. Acesso em: 22 de abril de 2025. Disponível em: <<https://sislegis.agricultura.gov.br/sislegis/action/detalhaAto.do?method=visualizarAtoPortalMapa&chave=1739574738>>.

PADILLA, R.; NETTO, S. L.; SILVA, E. A. da. A survey on performance metrics for object-detection algorithms. *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, IEEE, p. 237–242, 2020.

PINEAU, J. et al. Improving reproducibility in machine learning research (a report from the neurips 2019 reproducibility program). *Journal of Machine Learning Research*, v. 22, n. 164, p. 1–20, 2021. Disponível em: <<https://www.jmlr.org/papers/v22/20-303.html>>.

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 (CVPR)*. Las Vegas, NV, USA: IEEE, 2016. p. 779–788.

SÁ, P. C. A. et al. YOLOv8 para controle de produção pós-colheita e beneficiamento de frutos. *Revista de Engenharia e Pesquisa Aplicada*, v. 9, n. 1, p. 115–122, 2024.

SOHAN, M.; RAM, T. S.; REDDY, C. V. R. A review on YOLOv8 and its advancements. In: *SPRINGER. International Conference on Data Intelligence and Cognitive Informatics*. Cham, Switzerland: Springer, 2024. p. 529–545.

TAN, C. et al. A survey on deep transfer learning. In: *International Conference on Artificial Neural Networks (ICANN)*. [S.l.: s.n.], 2018. p. 270–279.



WANG, C.-Y.; BOCHKOVSKIY, A.; LIAO, H.-Y. M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. [S.l.: s.n.],

2023. p. 7464–7475.

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.