

**CLASSIFICAÇÃO  
INTELIGENTE DE GRÃOS DE  
MILHO POR VISÃO  
COMPUTACIONAL PARA O  
AGRONEGÓCIO**

**INTELLIGENT CLASSIFICATION OF CORN KERNELS USING COMPUTER  
VISION FOR AGRIBUSINESS**

Ciências Exatas e da Terra, Ciências Agrárias • 19/06/2026

REGISTRO DOI: [10.70773/revistatopicos/781716301](https://doi.org/10.70773/revistatopicos/781716301)

---

João Gabriel Silva dos Santos<sup>1</sup>  
Luis Eduardo José de Carvalho<sup>2</sup>  
João Arthur Pereira de Souza<sup>3</sup>  
Sthefane Dias Silva<sup>4</sup>  
Leandro Sanches Launé<sup>5</sup>  
Benjamim Eliel Nicácio Ribeiro<sup>6</sup>  
Luiz Felipe Cerqueira Guimarães<sup>7</sup>  
Luis Impieri Oliveira<sup>8</sup>  
Johão Vyctor Nunes Campelo<sup>9</sup>  
Isabela Maria Gonçalves<sup>10</sup>  
Giovanni Lucca França da Silva<sup>11</sup>

---

## RESUMO

O milho ocupa posição de destaque na economia e na segurança alimentar do Brasil, sendo o país o terceiro maior produtor e exportador mundial do cereal, com produção distribuída por todas as regiões do território nacional. Contudo, o controle de qualidade industrial ainda depende da classificação manual, um processo lento e suscetível a variações subjetivas entre avaliadores. Para superar esse desafio, este trabalho propõe o uso de Visão Computacional para a triagem automatizada de grãos em três categorias: pure (puro), broken (quebrado) e silkcut (danos por fungos). Avaliou-se um dataset de 11.818 imagens comparando três arquiteturas de redes neurais convolucionais. No conjunto de teste, a EfficientNetV2S obteve o melhor desempenho, alcançando 88% de acurácia, seguida pela MobileNetV2 e pela ResNet50, ambas com 87%. Os resultados obtidos indicam que as três arquiteturas apresentaram desempenho semelhante e demonstram a viabilidade do uso de Visão Computacional e Deep Learning para a automação da triagem de grãos no agronegócio nacional.

**Palavras-chave:** Controle de Qualidade; Visão Computacional; Redes Neurais Convolucionais; Classificação de Grãos; Agronegócio.

## ABSTRACT

Corn holds a prominent position in Brazil's economy and food security, as the country ranks among the world's three largest producers and exporters of the crop, with production spread across all national regions. However, industrial quality control still relies on manual classification, a slow process prone to subjective variation among evaluators. To address this challenge, this study proposes the use of Computer Vision for the automated sorting of corn grains into three categories: pure, broken, and silkcut. A dataset of 11,818 images was evaluated, comparing three convolutional neural network

architectures. In the test set, EfficientNetV2S achieved the best performance, reaching 88% accuracy, followed by MobileNetV2 and ResNet50, both with 87%. The obtained results indicate that the three architectures presented similar performance and demonstrate the feasibility of using Computer Vision and Deep Learning to automate grain sorting in national agribusiness.

**Keywords:** Quality Control; Computer Vision; Convolutional Neural Networks; Grain Classification; Agribusiness.

## 1. INTRODUÇÃO

O milho (*Zea mays* L.) destaca-se como uma das commodities de maior relevância no mercado mundial devido à diversidade de suas aplicações, que abrangem desde o comércio de grãos in natura até o fornecimento de insumos processados para a indústria. No cenário nacional, o cereal consolidou-se recentemente como o quarto produto mais comercializado internacionalmente pelo Brasil, evidenciando sua força na pauta de exportações (ALVES; TEIXEIRA; PINTO, 2022). Além do impacto econômico, a cultura é fundamental para a segurança alimentar, desempenhando um papel central na nutrição humana e animal, bem como na produção de bioetanol (ALVES; TEIXEIRA; PINTO, 2022).

No contexto produtivo, indicadores recentes da Companhia Nacional de Abastecimento (Conab) evidenciam a relevância estratégica da cultura no país. A safra brasileira de milho 2025/2026 apresenta uma área cultivada estimada em aproximadamente 22,4 milhões de hectares, com produtividade média de 6.171 kg/ha e produção total de cerca de 138,27 milhões de toneladas. Apesar da expansão da área plantada em relação à safra anterior, observa-se uma redução na produtividade e na produção total, indicando a influência de

fatores climáticos e operacionais sobre o desempenho da cultura (CONAB, 2026).

O Brasil figura como um dos protagonistas na produção e exportação mundial deste cereal, posicionando-se entre os três maiores produtores e exportadores do grão no mundo, com cultivo distribuído por todas as regiões do território nacional. Esse protagonismo resulta da expansão contínua de novas fronteiras agrícolas, do avanço tecnológico no manejo das lavouras e do crescimento acelerado do agronegócio brasileiro, que transformou o milho em um dos pilares da economia rural do país (DUARTE; MATTOSO; GARCIA, 2021).

Apesar da relevância econômica, a classificação de grãos na etapa de pós-colheita enfrenta desafios críticos de eficiência. Métodos tradicionais de identificação morfológica são limitados por fatores humanos e ambientais, o que reduz a precisão da triagem, enquanto abordagens bioquímicas ou moleculares apresentam custos elevados ou caráter destrutivo (XU et al., 2022). Devido à dificuldade de adaptação dessas técnicas para aplicações em tempo real na indústria, torna-se necessário o desenvolvimento de sistemas não destrutivos, rápidos e eficientes para a classificação automatizada do milho (XU et al., 2022).

Diante desses gargalos, a Visão Computacional destaca-se como uma alternativa não destrutiva capaz de extrair automaticamente características de cor, textura e forma dos grãos para a sua classificação (XU et al., 2022). Integradas ao Aprendizado Profundo (Deep Learning), as Redes Neurais Convolucionais (CNNs) oferecem alta capacidade de autoaprendizagem e generalização, sendo amplamente aplicadas para superar a complexidade da

identificação de padrões e da inspeção de qualidade em sementes de milho (XU et al., 2022).

Diversas arquiteturas de redes neurais têm sido propostas na literatura para tarefas de classificação de imagens agrícolas. Contudo, a aplicação comparativa dessas arquiteturas sob uma mesma base de dados voltada especificamente para grãos de milho ainda carece de investigação detalhada no contexto do agronegócio nacional (MENDONÇA; GUEDES, 2024).

Este trabalho propõe o desenvolvimento e a comparação de modelos de inteligência artificial para a classificação automatizada de grãos de milho em três categorias: pure (puro), broken (quebrado) e silkcut (danos por fungos/insetos). Utilizando um dataset de 11.818 imagens, busca-se identificar o desempenho das arquiteturas avaliadas e compreender as razões estruturais que tornam um modelo mais eficaz que os demais para cada tipo de avaria. Os resultados visam oferecer subsídios técnicos para a automação no agronegócio brasileiro, elevando a precisão e a eficiência na triagem de grãos e contribuindo para a modernização tecnológica do setor.

## **2. REVISÃO DA LITERATURA**

A classificação automatizada de grãos e sementes de milho por meio de técnicas de Inteligência Artificial e Visão Computacional tem recebido atenção crescente na literatura científica, sobretudo devido às limitações dos métodos tradicionais de inspeção. Esses métodos, baseados em análises morfológicas, bioquímicas ou moleculares, tendem a ser lentos, subjetivos, destrutivos e pouco adequados para aplicações industriais em tempo real (XU et al.,

2022). Nesse contexto, as Redes Neurais Convolucionais (CNNs) destacam-se por sua capacidade de extrair automaticamente características visuais relevantes, como cor, textura e formato dos grãos.

Xu et al. (2022) propuseram um sistema de classificação de sementes de milho utilizando redes neurais convolucionais e transfer learning. O estudo avaliou um conjunto de 8.080 amostras e introduziu a arquitetura P-ResNet, comparando seu desempenho com modelos consolidados. Os resultados demonstraram acurácia de 99,70%, evidenciando a alta capacidade do modelo na discriminação das diferentes variedades de sementes.

Alkanan e Gulzar (2024) desenvolveram um modelo baseado no MobileNetV2 para classificação de doenças em sementes de milho, utilizando um dataset com 17.801 imagens. A abordagem combinou técnicas de data augmentation e fine-tuning, alcançando acurácia de 96,27% e superando arquiteturas tradicionais. O estudo destaca a eficiência de modelos leves para aplicações práticas e embarcadas.

Iman, Arabnia e Rasheed (2023) realizaram uma revisão sobre Deep Transfer Learning, destacando sua relevância na adaptação de modelos pré-treinados para novos domínios. Os autores apontam o fine-tuning como principal estratégia e evidenciam que a similaridade entre domínios influencia diretamente o desempenho dos modelos.

Mendonça e Guedes (2024) aplicaram modelos de deep learning na classificação de grãos de milho, avaliando arquiteturas como MobileNetV2, InceptionV3 e EfficientNetV2. Os modelos

apresentaram acurácia entre 93,1% e 95,2%, com destaque para o uso de arquiteturas mais leves em cenários de aplicação prática.

Cavalli e Guadagnin (2021) utilizaram técnicas clássicas de visão computacional para monitoramento de falhas na germinação de milho por meio de imagens aéreas. Embora não apresentem métricas quantitativas padronizadas, os autores demonstram a viabilidade dessas abordagens para suporte à decisão agrícola.

Por fim, Santos et al. (2026) propuseram a comparação de modelos de deep learning para classificação automatizada de grãos de milho em três categorias, utilizando um dataset de 11.818 imagens. Os modelos foram treinados com fine-tuning e técnicas de aumento de dados, alcançando acurácia entre 83,4% e 85,7%. Apesar de inferiores a alguns trabalhos da literatura, os resultados obtidos demonstram a viabilidade da abordagem para aplicações práticas no agronegócio.

Apesar dos avanços significativos apresentados na literatura, observa-se que grande parte dos estudos alcança elevados níveis de acurácia utilizando datasets controlados, com alta qualidade visual e condições bem padronizadas. Em contrapartida, ainda há limitações quanto à generalização desses modelos para cenários reais do agronegócio, nos quais os dados apresentam maior variabilidade, ruído visual e complexidade.

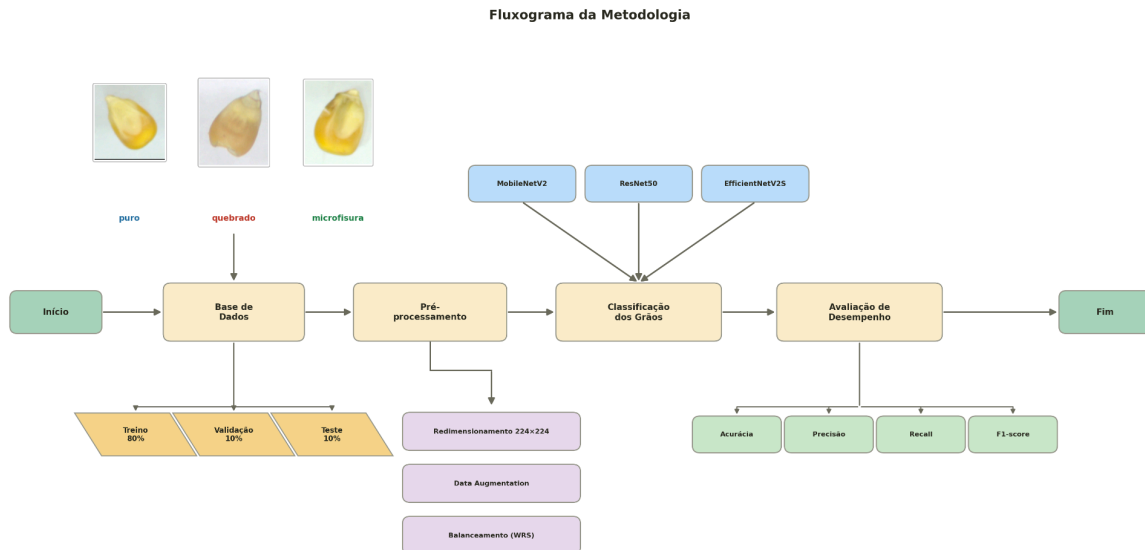
Nesse sentido, destaca-se a necessidade de abordagens que conciliem desempenho, robustez e adaptação a conjuntos de dados mais desafiadores. O presente trabalho busca contribuir nesse contexto ao avaliar arquiteturas de deep learning sob uma base de dados com características mais próximas da realidade prática, além de explorar estratégias de pré-processamento e balanceamento

visando melhorar a capacidade de generalização dos modelos. Dessa forma, pretende-se avançar não apenas em termos de desempenho, mas também na aplicabilidade da solução em ambientes reais de triagem automatizada de grãos.

### 3. METODOLOGIA

Esse artigo utilizou uma metodologia de trabalho aplicada, estruturada em um fluxo sequencial que vai desde a entrada das imagens até a obtenção dos resultados quantitativos por meio de métricas de desempenho. Esse processo é apresentado na Figura 1, em formato de fluxograma, mostrando cada etapa da análise realizada.

**Figura 1:** Fluxograma da Metodologia



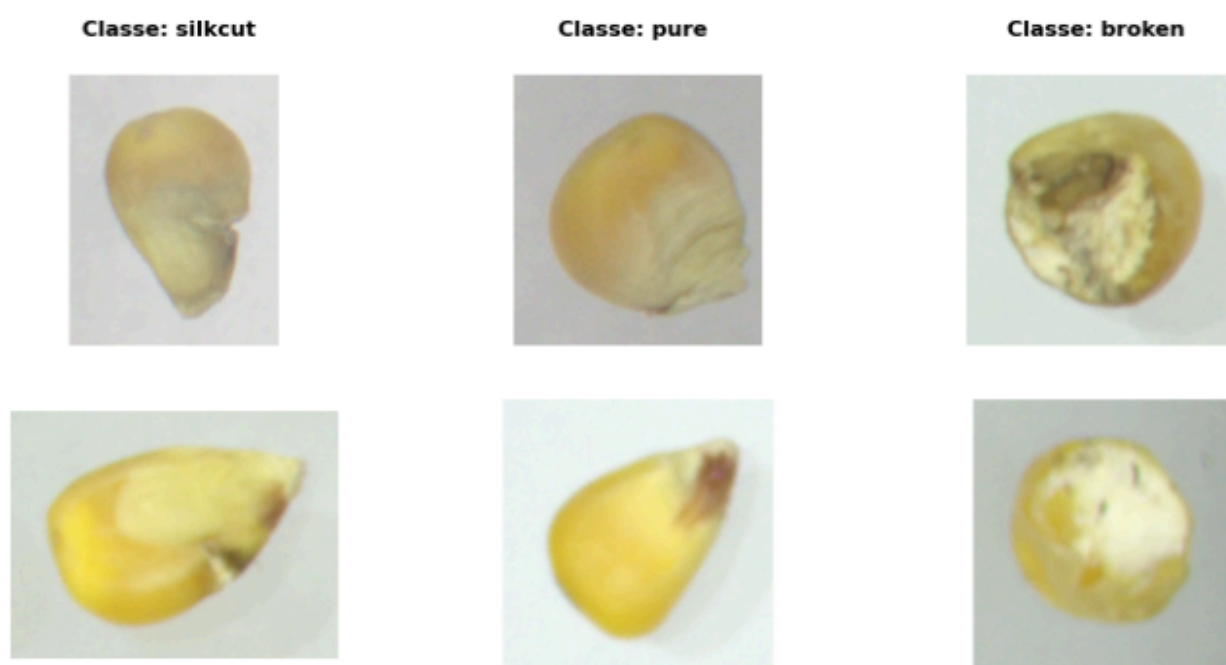
Fonte - Autores (2026)

#### 3.1. Base de Dados

O conjunto de dados utilizado neste trabalho é uma base pública disponível na plataforma Kaggle, composta por 11.800 imagens de grãos de milho individuais, organizadas em três classes: *broken*

(quebrado), *pure* (sadio) e *silkcut* (microfissuras no grão). Essa base foi também utilizada como referência no trabalho de Casadei (2023), que propôs o conjunto CornGrain para classificação automática de grãos, evidenciando sua relevância em pesquisas de visão computacional aplicada ao milho. Cada imagem representa um único grão fotografado sobre fundo padronizado, o que favorece a segmentação e a extração de características visuais relevantes para a classificação (CASADEI, 2023).

**Figura 2:** Representação visual das classes de milho



Fonte: Autores (2026).

### 3.2. Pré-processamento de Imagens

O pré-processamento das imagens constituiu etapa preparatória para o treinamento dos modelos, tendo como objetivo padronizar as entradas e ampliar a diversidade do conjunto de treinamento. Todas as imagens foram redimensionadas para a resolução de 192×192 pixels, compatível com as arquiteturas utilizadas, e submetidas a ajustes de contraste para uniformizar as condições de iluminação entre as amostras. Segundo Gonzalez e Woods (2010), operações de

pré-processamento visam melhorar a qualidade da imagem para as etapas subsequentes de extração de características e classificação, reduzindo variações que não carregam informação relevante para o modelo.

Para ampliar o conjunto de treinamento e reduzir o risco de overfitting, foram adotadas duas estratégias complementares aplicadas exclusivamente ao subconjunto de treino. A primeira foi o balanceamento por `WeightedRandomSampler`, em que pesos inversamente proporcionais à frequência de cada classe foram atribuídos às amostras, de forma que classes minoritárias fossem amostradas com maior probabilidade durante o treinamento, sem descarte de dados (MENDONÇA; GUEDES, 2024).

A segunda foi a aplicação de um pipeline de transformações de data augmentation, composto por: recorte aleatório redimensionado (`RandomResizedCrop`) com zoom leve entre 1,0 e 1,2; espelhamentos horizontal e vertical com probabilidade de 0,5 cada; rotações de até 90°; transformações afins aleatórias com translação de até 10% e escala entre 1,0 e 1,2, aplicadas com probabilidade de 0,75; distorção de perspectiva (`RandomPerspective`) com escala de distorção de 0,2, aplicada com probabilidade de 0,75; e variações de brilho e contraste por meio de `ColorJitter`, com intensidade de 0,1 para ambos os parâmetros, aplicadas também com probabilidade de 0,75.

Por fim, as imagens foram normalizadas com os valores de média e desvio padrão do ImageNet, garantindo compatibilidade com os pesos pré-treinados das arquiteturas utilizadas. Shorten, Khoshgoftaar e Furht (2023) apontam que esse conjunto de transformações aumenta artificialmente a diversidade dos dados, forçando o modelo a aprender representações mais generalizáveis e

invariantes a variações de captura. Para validação e teste, nenhuma transformação de aumento foi aplicada, sendo utilizadas apenas o redimensionamento e a normalização padrão.

### **3.3. Modelagem Preditiva**

Para a tarefa de classificação dos grãos de milho nas classes broken, pure e silkcut, foram treinadas e comparadas três arquiteturas de Redes Neurais Convolucionais (CNNs): MobileNetV2, ResNet50 e EfficientNetV2S. Todos os modelos foram inicializados com pesos pré-treinados no ImageNet e ajustados por meio de fine-tuning, técnica que transfere o conhecimento aprendido em grandes conjuntos de dados para um domínio específico, permitindo alcançar resultados competitivos mesmo com volume limitado de amostras (XU et al., 2022). A escolha dessas arquiteturas justifica-se pela diversidade de estratégias de extração de características que cada uma representa, viabilizando uma análise comparativa abrangente sobre o problema proposto (SOUSA; QUEIROZ; JUNIOR, 2024).

A MobileNetV2 é uma arquitetura projetada para ambientes com recursos computacionais limitados, cuja inovação central é a substituição das convoluções padrão por convoluções separáveis em profundidade (depthwise separable convolutions), o que reduz significativamente o número de parâmetros sem comprometer a acurácia (RYBCZAK et al., 2023).

A ResNet50 introduziu as conexões de atalho (skip connections), que somam a entrada de um bloco ao seu resultado convolucional, solucionando o problema de desvanecimento do gradiente e permitindo o treinamento de redes profundas sem degradação de

desempenho (AGGARWAL et al., 2022). Em estudos com sementes de milho, variantes da arquitetura ResNet alcançaram acurácias de até 99,70%, superando outras CNNs sob as mesmas condições experimentais (XU et al., 2022).

A EfficientNetV2S baseia-se no conceito de escalonamento composto (compound scaling), que ajusta profundidade, largura e resolução da rede simultaneamente em proporções otimizadas, obtendo maior acurácia com menor custo computacional em comparação a arquiteturas de porte equivalente (RYBCZAK et al., 2023).

### **3.3.1. Estratégia de Fine-tuning**

O fine-tuning é uma abordagem de transfer learning em que um modelo pré-treinado em um grande conjunto de dados é reutilizado e ajustado para uma tarefa específica, atualizando seus pesos por meio de retropropagação sobre os novos dados (GOODFELLOW; BENGIO; COURVILLE, 2016). Todos os modelos foram inicializados com pesos pré-treinados no ImageNet, e Xu et al. (2022) demonstraram que essa estratégia permite alcançar resultados comparáveis aos modelos treinados de ponta a ponta, mesmo em conjuntos de dados com volume limitado de amostras.

Para as três arquiteturas avaliadas, o treinamento seguiu duas fases distintas. Na primeira fase, correspondente às cinco primeiras épocas, o backbone permaneceu congelado e apenas a camada de classificação foi treinada com taxa de aprendizado de 0,001, garantindo estabilidade inicial e preservando os pesos pré-treinados (ALKANAN; GULZAR, 2024). A camada final foi substituída por uma sequência composta por Dropout com taxa de 0,3, seguido de uma

camada linear adaptada ao número de classes do problema, adicionando regularização e adequando a saída da rede à tarefa proposta. A partir da época 5, todas as camadas foram descongeladas e o fine-tuning completo foi aplicado com taxa de aprendizado reduzida de  $1 \times 10^{-4}$ .

Em ambas as fases, foi utilizado o otimizador Adam em conjunto com o scheduler ReduceLROnPlateau, configurado para monitorar a acurácia de validação com fator de redução de 0,3. Essa estratégia permite que a taxa de aprendizado seja reduzida automaticamente quando o modelo deixa de evoluir, prevenindo o overfitting ao longo das 40 épocas de treinamento (GOODFELLOW; BENGIO; COURVILLE, 2016).

### **3.4. Avaliação de Desempenho**

O desempenho dos modelos foi avaliado sobre o conjunto de teste, composto por 15% das imagens do dataset, mantido inteiramente separado durante as etapas de treinamento e validação. As métricas utilizadas foram acurácia, precisão, *recall* e F1-score, todas derivadas da matriz de confusão, que organiza as predições do modelo em Verdadeiros Positivos (VP), Verdadeiros Negativos (VN), Falsos Positivos (FP) e Falsos Negativos (FN) (GONÇALVES, 2023).

A acurácia representa a proporção geral de predições corretas sobre o total de amostras, sendo calculada conforme a Equação (3.1):

$$\frac{(VP + VN)}{(VP + VN + FP + FN)}$$

A precisão mede a proporção de predições positivas que foram de fato corretas, obtida por meio da Equação (3.2):

$$\frac{VP}{(VP + FP)}$$

Enquanto o *recall* quantifica a capacidade do modelo de identificar todas as instâncias reais de uma classe, Calculado segundo a Equação (3.3):

$$\frac{VP}{(VP + FN)}$$

O f1-score corresponde à média entre precisão e *recall*, calculado conforme a Equação (3.4), sendo especialmente relevante em problemas multiclasse com possível desbalanceamento entre categorias, pois penaliza modelos que apresentam desempenho assimétrico entre duas métricas (XU et al., 2022).

$$F1\ score = 2 \times \frac{(Precisão \times Recall)}{(Precisão + Recall)}$$

No contexto de classificação de grãos de milho, o recall assume particular importância, pois um modelo com alto recall assegura que a grande maioria dos grãos defeituosos seja corretamente identificada, reduzindo o risco de distribuição de produtos impróprios ao longo da cadeia produtiva (FERREIRA, 2023). Todas as métricas foram calculadas de forma ponderada, a fim de mostrar a contribuição de cada classe no resultado final, permitindo uma comparação justa entre os modelos avaliados.

#### **4. RESULTADOS E DISCUSSÃO**

Neste capítulo, será discutido sobre a configuração da máquina usada no experimento, ferramentas usadas durante o processo, divisão da base de dados, resultados quantitativos de cada arquitetura, tais como gráficos de avaliação e dados de métricas de

classificação em que são mostrados os gráficos do conjunto de validação, onde é possível analisar como cada arquitetura se comportou diante desse conjunto pela acurácia e pela perda, e uma tabela mostrando os resultados das métricas obtidas a partir do conjunto de testes, em que cada arquitetura foi posta à prova com imagens inéditas do dataset.

Após isso, é apresentada a discussão, em que são levantadas as vantagens e desvantagens do método proposto em comparação com os trabalhos relacionados, gerando uma discussão do que ainda pode ser implementado e do que já foi implementado, mas que ainda precisaria de ajustes, assim como também teve implementações que superaram valores de outros trabalhos.

#### **4.1. Máquina e Ferramentas**

O ambiente computacional para a execução dos experimentos e o treinamento dos modelos de *Deep Learning* foi provido integralmente pela plataforma Kaggle. O desenvolvimento e a execução do código ocorreram utilizando o ambiente de notebooks do Kaggle, que forneceu como recurso de hardware acelerador duas unidades de GPU NVIDIA T4, essenciais para o processamento do dataset e para a etapa de fine-tuning das arquiteturas MobileNetV2, ResNet50 e EfficientNetV2S.

#### **4.2. Base de Dados**

Para a divisão do conjunto de dados, adotou-se a proporção de 80% para treinamento, 10% para validação e 10% para teste, totalizando aproximadamente 8.260 imagens de treino, 1.770 de validação e 1.770 de teste. A divisão foi realizada de forma estratificada por

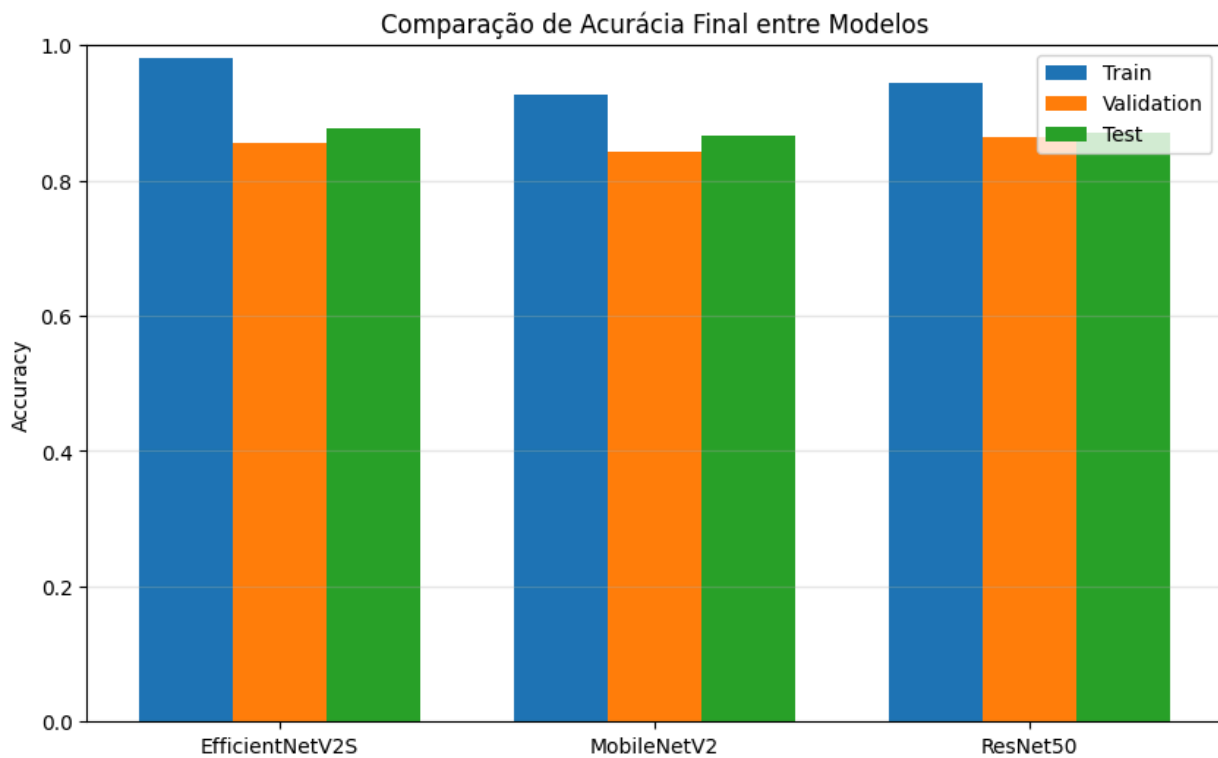
classe, garantindo que a distribuição proporcional entre broken, pure e silkcut fosse mantida em cada subconjunto.

Essa abordagem é recomendada para evitar viés na avaliação do modelo, pois assegura que todas as classes estejam representadas de forma equilibrada tanto no treinamento quanto na validação e no teste (GOODFELLOW; BENGIO; COURVILLE, 2016). A separação do conjunto de validação em relação ao de teste permite o ajuste de hiperparâmetros durante o treinamento sem comprometer a imparcialidade da avaliação final (MENDONÇA; GUEDES, 2024).

### **4.3. Resultados**

Com base na metodologia aplicada, foram treinados modelos para cada uma das arquiteturas avaliadas, comparando-se seus desempenhos finais nos conjuntos de treino, validação e teste para acurácia, e nos conjuntos de treino e validação para perda. A análise conjunta desses gráficos permite observar a capacidade de generalização dos modelos e identificar possíveis indícios de sobreajuste, especialmente quando há grande diferença entre o desempenho no treino e na validação.

**Figura 3:** Gráfico de Acurácia dos modelos

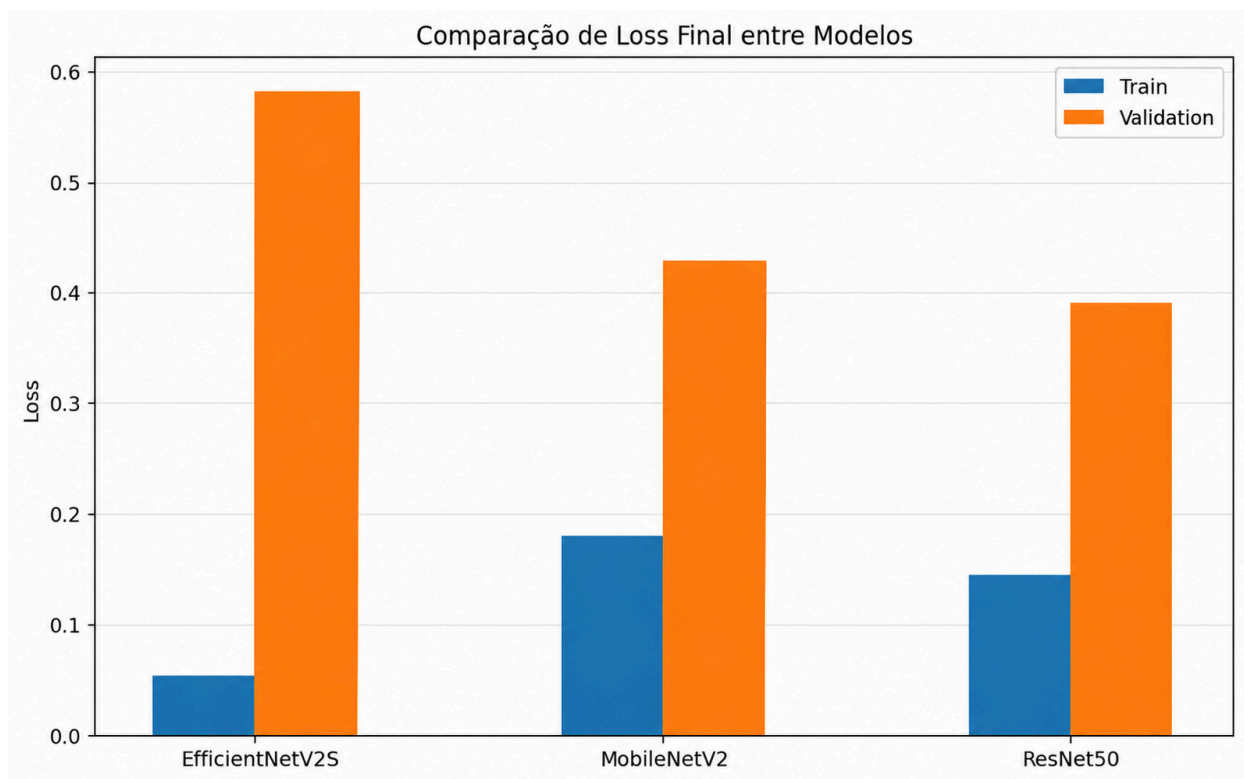


Fonte: Autores (2026).

Na Figura 3, observa-se que todas as arquiteturas apresentaram alta acurácia no conjunto de treino, com destaque para a EfficientNetV2S, que alcançou aproximadamente 98%. Entretanto, no conjunto de validação, os três modelos apresentaram valores mais próximos, variando em torno de 84% a 86%, o que indica desempenho semelhante entre as arquiteturas durante a etapa de validação.

No conjunto de teste, os modelos também mantiveram resultados próximos, com valores entre aproximadamente 86% e 88%. A ResNet50 obteve o melhor desempenho no teste, próximo de 88%, seguida pela EfficientNetV2S e pela MobileNetV2. Apesar da alta acurácia de treino da EfficientNetV2S, a diferença em relação à validação e ao teste sugere maior tendência ao sobreajuste quando comparada às demais arquiteturas.

**Figura 4:** Gráfico de perda dos modelos



Fonte: Autores (2026).

Na Figura 3, que apresenta a perda final dos modelos, a EfficientNetV2S mostrou a menor perda de treino, em torno de 0,05, porém também exibiu a maior perda de validação, próxima de 0,58. Esse comportamento reforça o indício de sobreajuste, pois o modelo parece ter se ajustado muito bem aos dados de treino, mas teve maior dificuldade em manter o mesmo desempenho nos dados de validação.

A MobileNetV2 apresentou perda de treino em torno de 0,18 e perda de validação próxima de 0,43, demonstrando comportamento mais equilibrado. Já a ResNet50 obteve a menor perda de validação, aproximadamente 0,39, além de uma perda de treino também baixa, em torno de 0,15, indicando melhor equilíbrio entre aprendizado e generalização.

No conjunto de testes, todos os modelos alcançaram uma boa acurácia, mas ainda assim, longe do objetivo de 90% ou mais que foi

buscado. Aqui na Tabela 1, são apresentados os resultados das métricas das arquiteturas a partir do conjunto de testes.

**Tabela 1:** Comparação dos resultados das métricas do conjunto de teste

ARQUITETURAS	ACURÁCIA	PRECISÃO	REVOCAÇÃO	F1-SCORE
EfficientNetV2S	88%	86%	87%	87%
MobileNetV2	87%	85%	87%	86%
ResNet50	87%	86%	88%	86%

Fonte: Autores (2026).

Pode-se observar, na Tabela 1, que todas as arquiteturas apresentaram desempenho acima de 85% em todas as métricas avaliadas, indicando resultados consistentes na classificação dos tipos de grãos de milho. Em relação à acurácia, a EfficientNetV2S obteve o melhor resultado, com 88%, enquanto a MobileNetV2 e a ResNet50 alcançaram 87%. Isso mostra uma diferença pequena entre os modelos, com vantagem de apenas 1% para a EfficientNetV2S.

Na métrica de precisão, a EfficientNetV2S e a ResNet50 apresentaram os melhores resultados, ambas com 86%, enquanto a MobileNetV2 ficou ligeiramente abaixo, com 85%. Em relação à revocação, a ResNet50 obteve o maior valor, com 88%, indicando melhor capacidade de identificar corretamente as amostras pertencentes às classes avaliadas. Já a EfficientNetV2S e a MobileNetV2 apresentaram 87% nessa métrica.

Por fim, no F1-score, a EfficientNetV2S apresentou o melhor desempenho, com 87%, enquanto a MobileNetV2 e a ResNet50 obtiveram 86%. De modo geral, os resultados mostram que as três arquiteturas tiveram desempenhos próximos, com leve vantagem para a EfficientNetV2S, que apresentou a maior acurácia e o maior F1-score. Ainda assim, a proximidade entre os valores indica que novos experimentos, ajustes de hiperparâmetros e técnicas de aumento de dados podem contribuir para melhorar ainda mais a capacidade de generalização dos modelos.

#### **4.4. Discussão**

De maneira geral, observa-se que as arquiteturas avaliadas apresentaram desempenhos semelhantes, com diferenças pouco expressivas entre as métricas de acurácia, precisão, revocação e F1-score. Esse comportamento indica uma convergência entre os modelos, sugerindo que as limitações do problema estão mais relacionadas às características do conjunto de dados do que exclusivamente à capacidade das arquiteturas. Ainda assim, a EfficientNetV2S destacou-se levemente nas métricas finais de teste, apresentando a maior acurácia, com 88%, e o maior F1-score, com 87%. Por outro lado, a ResNet50 apresentou o melhor comportamento em termos de perda de validação e a maior revocação, com 88%, indicando maior capacidade de identificar corretamente as amostras pertencentes às classes avaliadas.

Apesar dos resultados satisfatórios, os modelos ainda não atingiram o desempenho observado em estudos da literatura. Trabalhos como Xu et al. (2022), com acurácia de 99,7%, e Alkanan e Gulzar (2024), com cerca de 96%, apresentam resultados superiores. Essa diferença pode ser explicada por aspectos metodológicos distintos.

Enquanto esses estudos priorizam a otimização das arquiteturas, modificações estruturais nas redes e aprofundamento das camadas convolucionais, o presente trabalho concentrou-se principalmente no uso de transfer learning, fine-tuning, balanceamento de classes e técnicas de aumento de dados. Embora essas estratégias tenham contribuído para bons resultados, elas podem ser insuficientes para capturar padrões altamente discriminativos em problemas com baixa separabilidade visual entre classes.

A qualidade do conjunto de dados mostrou-se um fator determinante para o desempenho final. A presença de imagens com possíveis variações de nitidez, iluminação e similaridade visual entre as classes dificulta a extração de características relevantes pelas redes convolucionais, principalmente em relação a padrões finos de textura, borda e danos superficiais nos grãos. Isso reduz a capacidade discriminativa dos modelos e impacta diretamente sua generalização. A proximidade entre os resultados das três arquiteturas reforça que o desafio não está apenas na escolha do modelo, mas também na qualidade, diversidade e representatividade das imagens utilizadas.

A análise das métricas revela equilíbrio entre precisão e revocação, refletido na proximidade entre acurácia e F1-score. A EfficientNetV2S obteve 88% de acurácia, 86% de precisão, 87% de revocação e 87% de F1-score, demonstrando o melhor resultado geral no conjunto de teste. A MobileNetV2 apresentou 87% de acurácia, 85% de precisão, 87% de revocação e 86% de F1-score, mantendo desempenho competitivo. Já a ResNet50 alcançou 87% de acurácia, 86% de precisão, 88% de revocação e 86% de F1-score, destacando-se pela maior revocação entre os modelos. Esse comportamento sugere que as técnicas de balanceamento aplicadas, como o

WeightedRandomSampler, contribuíram para reduzir possíveis efeitos do desbalanceamento inicial entre classes.

Em relação ao processo de treinamento, a estratégia de fine-tuning, com congelamento inicial do backbone seguido da liberação das camadas com menor taxa de aprendizado, contribuiu para a adaptação dos modelos ao domínio do problema. No entanto, a análise conjunta dos gráficos de acurácia e perda indica a presença de possíveis sinais de sobreajuste, principalmente na EfficientNetV2S. Essa arquitetura apresentou acurácia de treino próxima de 98% e perda de treino muito baixa, em torno de 0,05, mas também registrou a maior perda de validação, próxima de 0,58. Esse comportamento sugere que, embora tenha alcançado o melhor desempenho no teste, o modelo pode ter aprendido padrões muito específicos do conjunto de treino.

A ResNet50, por sua vez, apresentou comportamento mais equilibrado na análise da função de perda, com perda de validação próxima de 0,39, a menor entre as arquiteturas avaliadas. Esse resultado indica melhor estabilidade durante a validação e sugere maior capacidade de generalização em comparação às demais arquiteturas. A MobileNetV2 também apresentou desempenho estável, com perda de validação em torno de 0,43 e métricas finais próximas às da ResNet50, embora tenha ficado ligeiramente abaixo nas métricas de precisão e F1-score.

A análise dos erros evidencia uma dificuldade natural na distinção entre classes visualmente semelhantes, especialmente entre grãos puros e grãos quebrados ou com danos sutis. Essa limitação pode ser atribuída à similaridade estrutural entre os grãos e à presença de características visuais muito próximas entre algumas categorias.

Dessa forma, o desafio do problema não está apenas na modelagem, mas também na capacidade do conjunto de dados de representar, com nitidez e diversidade, as variações reais existentes entre as classes.

Mesmo diante dessas limitações, os resultados obtidos demonstram o potencial da aplicação de técnicas de Deep Learning na classificação automatizada de grãos de milho. As acurácias entre 87% e 88% indicam que os modelos avaliados já possuem desempenho promissor para aplicações práticas, especialmente em etapas iniciais de triagem e controle de qualidade. Nesse contexto, a automação pode contribuir para reduzir custos operacionais, diminuir a subjetividade da classificação manual e aumentar a eficiência do processo produtivo no agronegócio.

Por fim, observa-se que ainda há espaço significativo para evolução da abordagem proposta. Trabalhos futuros podem explorar estratégias mais aprofundadas de otimização arquitetural, incluindo ajustes nas camadas convolucionais, uso de modelos modificados, técnicas avançadas de regularização e otimização de hiperparâmetros. Além disso, a ampliação e melhoria da qualidade do conjunto de dados, bem como a utilização de imagens mais representativas do ambiente real de produção, tendem a impactar diretamente o desempenho dos modelos. A integração dessas melhorias, aliada à validação em cenários reais de aplicação, constitui um passo essencial para a consolidação da solução no contexto do agronegócio.

## **5. CONCLUSÃO**

Este trabalho teve como objetivo desenvolver e comparar modelos de Inteligência Artificial baseados em Visão Computacional para a classificação automatizada de grãos de milho nas categorias *pure*, *broken* e *silkcut*. Para isso, foram avaliadas três arquiteturas de Redes Neurais Convolucionais: MobileNetV2, ResNet50 e EfficientNetV2S, utilizando um dataset público composto por 11.818 imagens. A metodologia adotou *transfer learning* com pesos pré-treinados no ImageNet, *fine-tuning*, técnicas de *data augmentation*, balanceamento por *WeightedRandomSampler* e congelamento inicial do *backbone* durante as primeiras etapas do treinamento.

Os resultados obtidos demonstram que todas as arquiteturas foram capazes de aprender características relevantes dos grãos, alcançando desempenho satisfatório no conjunto de teste. A EfficientNetV2S apresentou o melhor resultado geral, com 88% de acurácia, 86% de precisão, 87% de revocação e 87% de F1-score. A MobileNetV2 obteve 87% de acurácia, 85% de precisão, 87% de revocação e 86% de F1-score, enquanto a ResNet50 alcançou 87% de acurácia, 86% de precisão, 88% de revocação e 86% de F1-score. Esses resultados indicam que as três arquiteturas tiveram desempenho bastante próximo, com pequena vantagem da EfficientNetV2S nas métricas finais de acurácia e F1-score.

A análise dos gráficos de acurácia e perda também evidenciou diferenças importantes entre os modelos. A EfficientNetV2S apresentou alta acurácia no treino e baixa perda de treinamento, mas também maior perda de validação, indicando possível tendência ao sobreajuste. Por outro lado, a ResNet50 apresentou menor perda de validação, sugerindo comportamento mais estável durante a generalização. A MobileNetV2 manteve desempenho

intermediário e consistente, com resultados próximos aos das demais arquiteturas.

Em comparação com estudos da literatura, como Xu et al. (2022) e Alkanan e Gulzar (2024), os resultados ainda são inferiores, o que pode estar relacionado à qualidade visual do dataset, à similaridade entre as classes e à ausência de modificações arquiteturais mais profundas. Ainda assim, os resultados demonstram a viabilidade do uso de Deep Learning na triagem automatizada de grãos de milho, especialmente em etapas iniciais de controle de qualidade.

Como trabalhos futuros, recomenda-se a utilização de bases com maior qualidade visual, maior diversidade de amostras e validação em cenários reais de produção. Também devem ser exploradas otimizações arquiteturais, ajustes de hiperparâmetros e técnicas avançadas de regularização, visando aumentar a robustez e a aplicabilidade prática da solução no agronegócio.

## REFERÊNCIAS BIBLIOGRÁFICAS

AGGARWAL, Priya et al. **COVID-19 image classification using deep learning: advances, challenges and opportunities**. Computers in Biology and Medicine, [s. l.], v. 144, 2022.

ALVES, Francys Júnio de Castro; TEIXEIRA, Fábio André; PINTO, Victor Henrique Lana. **Competitividade das exportações de milho de Patos de Minas**, MG. Revista De Política Agrícola, 31(3), 52. Disponível em: <https://rpa.sede.embrapa.br/RPA/article/view/1745>. Acesso em: 10 mar. 2026.

ALKANAN, M.; GULZAR, Y. **Enhanced corn seed disease classification: leveraging MobileNetV2 with feature**

**augmentation and transfer learning. *Frontiers in Applied Mathematics and Statistics*.** v. 9, 12f. 2024.

CAVALLI, H.; GUADAGNIN, E. Aplicação da visão computacional no monitoramento da performance de híbridos de milho. *Conjecturas*, v. 21, n. 7, p. 494–503, 2021.

CONAB – COMPANHIA NACIONAL DE ABASTECIMENTO. Acompanhamento da safra brasileira de grãos: safra 2025/26, 6º levantamento. Brasília: Conab, 2026. Disponível em: <https://www.gov.br/conab>. Acesso em: 10 mar. 2026.

CASADEI, Adriana. **CornGrain: base de dados e classificação automática de grãos de milho.** 58 f. Monografia (MBA em Inteligência Artificial e Big Data) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2023.

FERREIRA, Diego. **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, Toledo, 2023.

GONZALEZ, R. C.; WOODS, R. E. **Processamento digital de imagem.** 3. ed. São Paulo: Pearson, 2010.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep Learning.** Cambridge: MIT Press, 2016.

HOBBS, Sam et al. Corn Kernel Counting Dataset. Kaggle, 2021. Disponível em: <https://www.kaggle.com/datasets>. Acesso em: 10 mar. 2026.

IMAN, M.; ARABNIA, H. R.; RASHEED, K. A review of deep transfer learning and recent advancements. *Technologies*, v. 11, n. 2, art. 40, 2023.

MENDONÇA, Arthur; GUEDES, Elloá. **Classificação e Detecção Inteligentes de Grãos para Agricultura Digital na Cultura do Milho**. Manaus: Laboratório de Sistemas Inteligentes, Escola Superior de Tecnologia, Universidade do Estado do Amazonas, 2024.

POZZO, Beatriz Salandin Dal et al. **ANÁLISE DO RISCO DAS CULTURAS DE SOJA E MILHO NA REGIÃO DO MATOPIBA**. *Revista Econômica do Nordeste*. 55, 4 (out. 2024), 63–85. Disponível em: <https://doi.org/10.61673/ren.2024.1576>. Acesso em: 10 mar. 2026.

PRADO, Juliana. **Uso de técnicas tradicionais e computacionais para caracterização da qualidade de grãos de café**. 2021. Tese (Doutorado em Engenharia Agrícola) – Faculdade de Engenharia Agrícola, Universidade Estadual de Campinas, Campinas, 2021.

ROSIN, Juliana. **Visão Computacional Aplicada na Classificação de Grãos de Milho**. 64 f. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) – Universidade Federal da Fronteira Sul, Chapecó, 2019.

RYBCZAK, Monika; KOZAKIEWICZ, Krystian. **Deep machine learning of the MobileNet, Efficient and Inception models**. Faculty of Marine Electrical Engineering. 2023.

SANTOS, João Gabriel Silva dos, et al. Classificação inteligente de grãos de milho por visão computacional para o agronegócio do Maranhão. 2026. Artigo científico.

SHORTEN, Connor; KHOSHGOFTAAR, Taghi M.; FURHT, Borko. **Data augmentation in classification and segmentation: a survey and new strategies.** *Journal of Imaging*, Basel, v. 9, n. 2, art. 46, fev. 2023.

SOUSA, Jeanne; QUEIROZ, José; JUNIOR, José. **Capacitação avançada em robótica industrial: ferramentas acadêmicas para soluções industriais.** Editora Paisson, Belo Horizonte, 2024.

XU, Peng et al. **Research on Maize Seed Classification and Recognition Based on Machine Vision and Deep Learning.** *Agriculture*, Basel, v. 12, 16f, fev. 2022.

XU, Peng et al. **Research on Maize Seed Classification and Recognition Based on Machine Vision and Deep Learning.** *Agriculture* 2022, 12, 232. Disponível em: <https://doi.org/10.3390/agriculture12020232>. Acesso em: 12 mar. 2026.

---

<sup>1</sup> Graduando em Engenharia de Software, UNDB. E-mail: [acesse o artigo original para visualizar o e-mail](#).

<sup>2</sup> Graduado em Engenharia de Software, UNDB. E-mail: [acesse o artigo original para visualizar o e-mail](#).

<sup>3</sup> Graduando em Engenharia de Software, UNDB. E-mail: [acesse o artigo original para visualizar o e-mail](#).

<sup>4</sup> Graduanda em Análise e Desenvolvimento de Sistemas, UNDB. E-mail: [acesse o artigo original para visualizar o e-mail](#).

<sup>5</sup> Graduando em Engenharia de Software, UNDB. E-mail: [acesse o artigo original para visualizar o e-mail](#).

- <sup>6</sup> Graduando em Engenharia de Software, UNDB. E-mail: [acesse o artigo original para visualizar o e-mail](#).
- <sup>7</sup> Graduando em Engenharia de Software, UNDB. E-mail: [acesse o artigo original para visualizar o e-mail](#).
- <sup>8</sup> Graduando em Engenharia de Software, UNDB. E-mail: [acesse o artigo original para visualizar o e-mail](#).
- <sup>9</sup> Graduando em Engenharia de Software, UNDB. E-mail: [acesse o artigo original para visualizar o e-mail](#).
- <sup>10</sup> Graduanda em Engenharia de Software, UNDB. E-mail: [acesse o artigo original para visualizar o e-mail](#).
- <sup>11</sup> Professor orientador. Dr. em Engenharia Elétrica pela UFMA. E-mail: [acesse o artigo original para visualizar o e-mail](#).