

Automated Correction of Multiple Choice Tests Using Computer Vision

Pedro Henrique Campos Moreira ¹, Bianca Panacho Ferreira ², Marcus Vinicius Diniz dos Reis ³

¹Instituto de Ciências Exatas e Técnicas
Universidade Federal de Viçosa, Campus Rio Paranaíba (UFV-CRP)

pedro.henrique.moreira@ufv.br¹, bianca.p.ferreira@ufv.br², marcus.v.reis@ufv.br³

Abstract. *Automated correction of multiple-choice exams is a challenge in large-scale assessments. Traditional OMR systems require costly hardware, while manual correction is slow and error-prone. This work proposes a low-cost solution based on computer vision, avoiding complex AI methods. Preprocessing includes adaptive thresholding, edge detection, and histogram equalization. The system was tested on 240 answer sheets labeled as blank, correct, or multiple-marked. Results indicate high accuracy and efficiency under varied scanning conditions. Our method also reduces hardware dependency. Thus, it offers an accessible and scalable alternative for educational institutions.*

Resumo. *A correção automatizada de exames de múltipla escolha é um desafio em avaliações de larga escala. Sistemas OMR exigem hardware caro, enquanto a correção manual é lenta e propensa a erros. Este trabalho propõe uma solução de baixo custo com visão computacional, sem uso de IA complexa. O pré-processamento inclui limiarização adaptativa, detecção de bordas e equalização de histograma. O sistema foi testado em 240 folhas classificadas como em branco, corretas ou com múltiplas marcações. Os resultados mostram alta acurácia e eficiência. O método também reduz a dependência de hardware. Assim, é uma alternativa acessível e escalável para instituições de ensino.*

1. Introdução

A avaliação do desempenho acadêmico em larga escala, especialmente por meio de questões de múltipla escolha, é um desafio para os sistemas educacionais em todo o mundo. Exames como o Exame Nacional do Ensino Médio (ENEM), que envolve milhões de estudantes anualmente, exigem processos de correção eficientes, precisos e escaláveis para garantir qualidade e equidade na avaliação [de Assis Zampirolli et al. 2016][Tieppo et al. 2023]. Nesse contexto, o uso de técnicas de correção automatizada, como o Reconhecimento Óptico de Marcas (OMR), tem se mostrado uma solução adequada.

O OMR é amplamente utilizado para a leitura e correção automática de respostas em folhas de resposta, permitindo o processamento rápido de grandes volumes de dados [de Elias et al. 2021][Hafeez et al. 2024]. No entanto, a crescente demanda por escalabilidade e acessibilidade nos sistemas educacionais, especialmente em países com desafios regionais e recursos limitados, como o Brasil, exige o desenvolvimento de soluções mais robustas e adaptáveis. A integração do OMR com técnicas avançadas de visão computacional oferece uma alternativa promissora, capaz de garantir maior precisão e eficiência

na correção de provas, ao mesmo tempo em que minimiza o tempo de processamento e os riscos associados ao erro humano.

O principal objetivo deste trabalho foi desenvolver um sistema de correção automática de testes de múltipla escolha utilizando técnicas de visão computacional com OMR. O objetivo foi melhorar a precisão e a eficiência desses processos, avaliar o desempenho da solução em comparação com métodos tradicionais e explorar sua modularidade para permitir adaptações a diferentes cenários educacionais. Esta proposta inclui a implementação de métodos simples e eficazes, viáveis mesmo para dispositivos eletrônicos com poucos recursos, garantindo acessibilidade em diversas regiões.

2. Trabalhos Relacionados

A correção automatizada de folhas de resposta utilizando técnicas de visão computacional tem sido amplamente explorada na literatura, com diversos estudos focando na otimização da precisão e acessibilidade desses sistemas. No entanto, nosso trabalho se destaca por oferecer uma solução eficiente, em tempo real, capaz de corrigir múltiplas provas por segundo, além de armazenar informações detalhadas sobre os erros e acertos dos candidatos.

Pereira (2019) [Pereira 2019] propôs uma abordagem que utiliza o Reconhecimento Óptico de Marcas (OMR) para correção de folhas de resposta, destacando sua precisão na leitura e segmentação em ambientes educacionais, de forma semelhante à nossa solução. Contudo, nosso sistema vai além ao oferecer correção em tempo real, processando várias provas simultaneamente e exibindo os resultados de forma imediata, sem depender de redes neurais complexas ou altos recursos computacionais.

China et al. (2016) [China et al. 2016] desenvolveram um aplicativo Android para correção de testes de múltipla escolha, utilizando dispositivos móveis como uma solução acessível para instituições com infraestrutura limitada. Embora essa abordagem também se beneficie de tecnologias acessíveis como o OpenCV, nosso sistema se destaca por realizar o processamento e armazenamento dos resultados em tempo real, gerando estatísticas detalhadas sobre erros e acertos à medida que as provas são corrigidas.

Além disso, enquanto muitos sistemas atuais dependem de redes neurais e algoritmos de aprendizado profundo, como o proposto por Silva Filho [Silva Filho 2022], que utiliza CNNs para monitoramento de exames, nosso trabalho adota uma abordagem puramente baseada em visão computacional. Isso torna o sistema extremamente leve e capaz de rodar em praticamente qualquer dispositivo, desde computadores simples até dispositivos móveis, sem a necessidade de hardware especializado ou ambientes controlados.

Em nosso projeto, nossa contribuição vai além da acessibilidade e eficiência do sistema, abrangendo também sua capacidade de realizar correções precisas e armazenamento contínuo dos resultados em tempo real, sem sobrecarga computacional. Isso o torna uma solução prática e escalável para instituições educacionais que buscam otimizar o processo de correção de provas com recursos mínimos.

3. Método Proposto

O método proposto para correção automatizada de testes de múltipla escolha segue cinco etapas principais, conforme ilustrado na Fig. 1.

A **primeira etapa** envolve o cadastro das folhas de resposta corretas e das que serão corrigidas (1), em que as folhas são inseridas no sistema, garantindo a precisão das correções subsequentes, como no projeto proposto por [Zafar 2023].

Na **segunda etapa**, ocorre a captura de vídeo e o redimensionamento das imagens (2), onde as folhas de resposta são processadas digitalmente. Utilizando técnicas avançadas de processamento de imagem, como limiarização e correção de distorções, o sistema consegue ajustar a imagem independentemente da qualidade de entrada, técnica semelhante à utilizada em [Silva et al. 2015] (3). O processamento contínuo garante que o sistema opere de forma eficiente mesmo em dispositivos de baixo custo.

A **terceira etapa** envolve a identificação da folha de resposta (4), na qual as marcações dos estudantes são reconhecidas e extraídas da imagem por meio de algoritmos robustos de visão computacional. Na quarta etapa, o sistema compara as respostas dos alunos (6) com o gabarito previamente registrado.

Todos os dados coletados são armazenados em um banco de dados centralizado (5), onde os erros e acertos são contabilizados e analisados. Por fim, na quinta etapa, o sistema exibe os resultados em uma interface web (7), permitindo que os usuários visualizem em tempo real as estatísticas de acertos, erros e desempenho dos estudantes.

Desenvolvemos o sistema utilizando Python com a biblioteca OpenCV para manipulação de imagens e o framework Flask para construção da interface web interativa. O armazenamento dos dados foi gerenciado com o OpenPyXL, garantindo flexibilidade e precisão no registro das correções. A solução foi projetada para ser leve e capaz de rodar em dispositivos convencionais e móveis, sem necessidade de infraestrutura complexa.

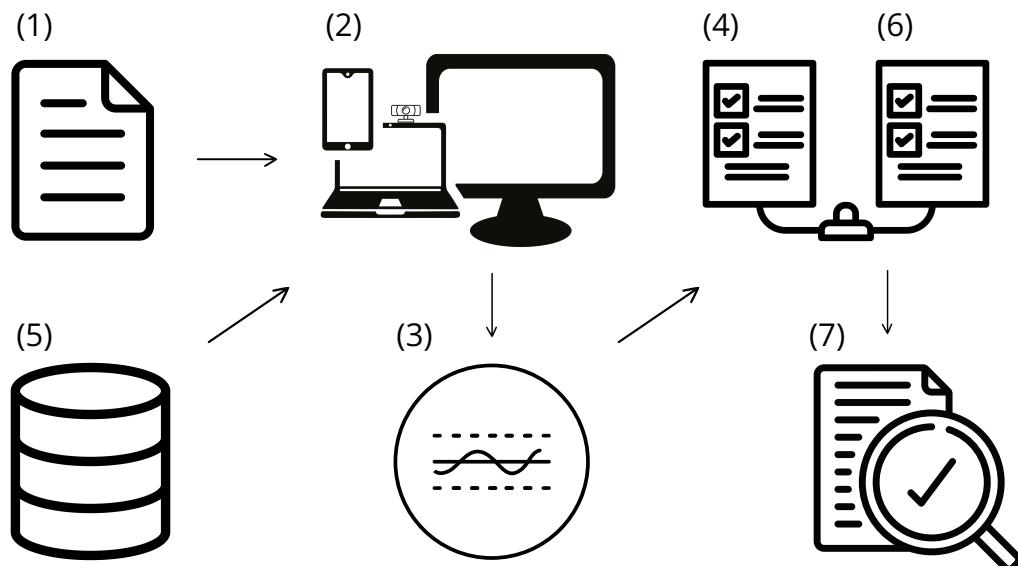


Figura 1. Método Proposto.

3.1. Dataset

As imagens utilizadas foram criadas especificamente para este estudo, totalizando 240 imagens de folhas de resposta provenientes de exames e simulados. Cada imagem continha de quatro a cinco questões marcadas, categorizando as folhas de resposta em três classes: imagens em branco (80 imagens), imagens com informações corretas (80 imagens) e imagens com múltiplas marcações (80 imagens). Para ilustrar o conjunto de dados, uma imagem de exemplo é apresentada na Fig. 2.

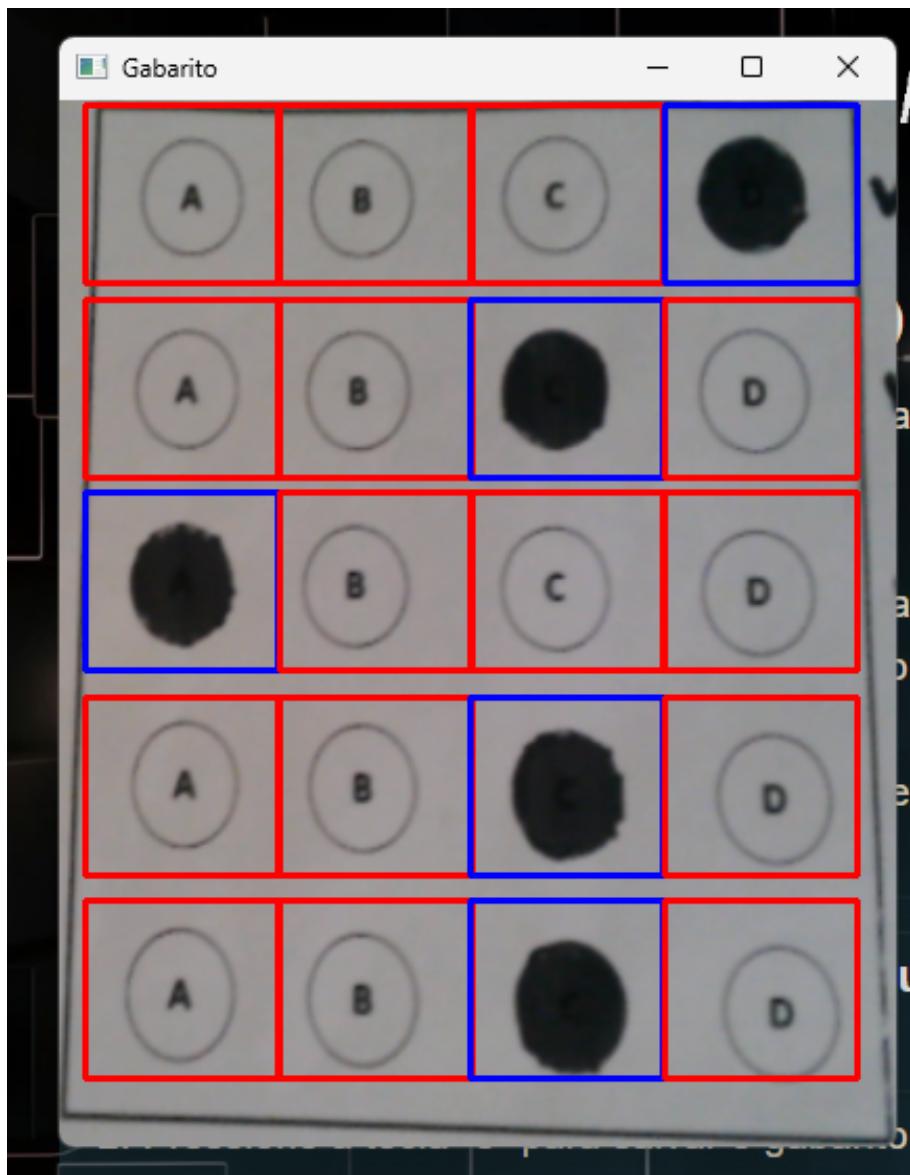


Figura 2. Exemplo das Imagens de Testes.

3.2. Técnicas de Visão Computacional

O método proposto utiliza uma abordagem simples e robusta de visão computacional para processar e analisar folhas de resposta, evitando assim a complexidade das redes neurais, conforme mencionado em [Dias 2018]. Aplicamos filtros de Sobel para detecção de bordas, auxiliando na segmentação das áreas de interesse, enquanto a equalização de

histograma e a limiarização melhoraram a qualidade da imagem e compensaram variações de iluminação e contraste [Gonzalez and Woods 2008]. Essas técnicas simples permitem ao sistema extrair com precisão as informações relevantes das imagens de entrada, como as marcações feitas pelos estudantes.

As imagens processadas (Fig. 3 e 4) ilustram visualmente o funcionamento do sistema. A Fig. 3 mostra a identificação e segmentação das marcações dos alunos, onde as áreas marcadas são destacadas para análise. A versão binarizada à direita facilita a detecção precisa das respostas. A Fig. 4 apresenta o alinhamento e a divisão das folhas de resposta em uma matriz de quadrantes, permitindo uma comparação eficiente entre as respostas dos alunos e o gabarito. Isso possibilitou uma análise clara e detalhada das respostas, garantindo uma correção precisa e em tempo real.

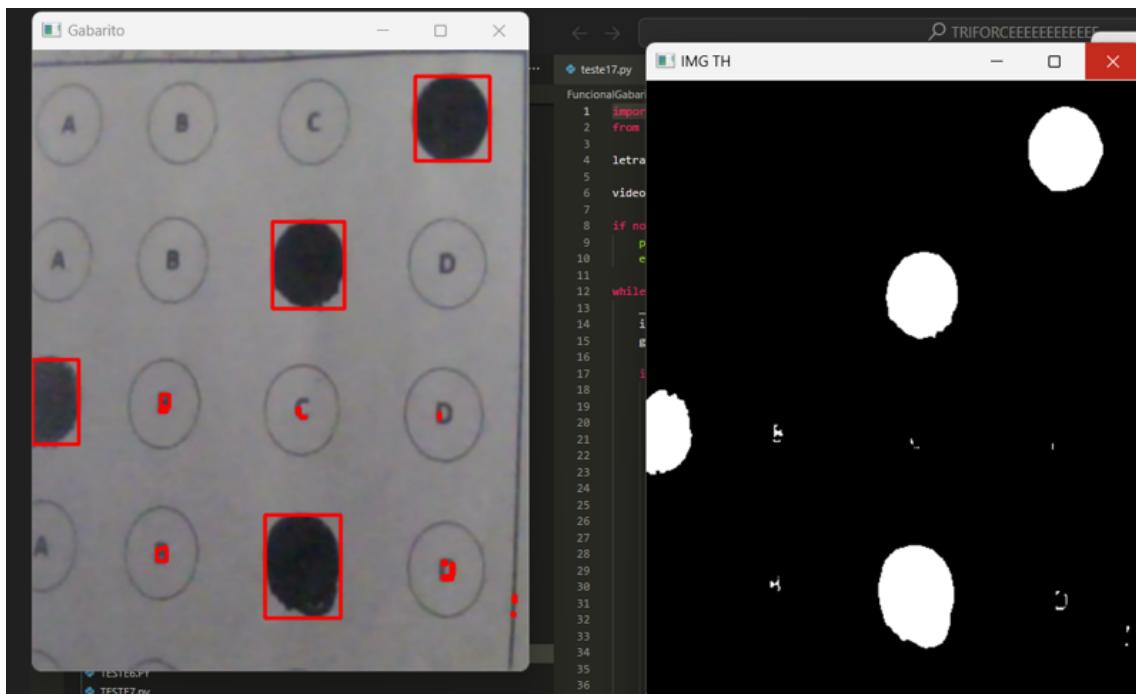


Figura 3. Identificação e Segmentação.

3.3. Métricas de Avaliação

Para a correção em tempo real, consideramos as seguintes métricas: acurácia, precisão, revocação (*recall*) e *F1-Score* [Fawcett 2006]. Essas métricas fornecem uma compreensão abrangente da efetividade na correção das folhas de resposta. As fórmulas consideram os índices derivados da matriz de confusão: verdadeiros positivos (TP), verdadeiros negativos (TN), falsos positivos (FP) e falsos negativos (FN), permitindo a avaliação do desempenho do sistema em diferentes aspectos cruciais.

- Acurácia: mede a correção geral do classificador, calculada pela proporção de verdadeiros positivos (VP) e verdadeiros negativos (TN) em relação ao total de elementos na matriz de confusão (Equação 1).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

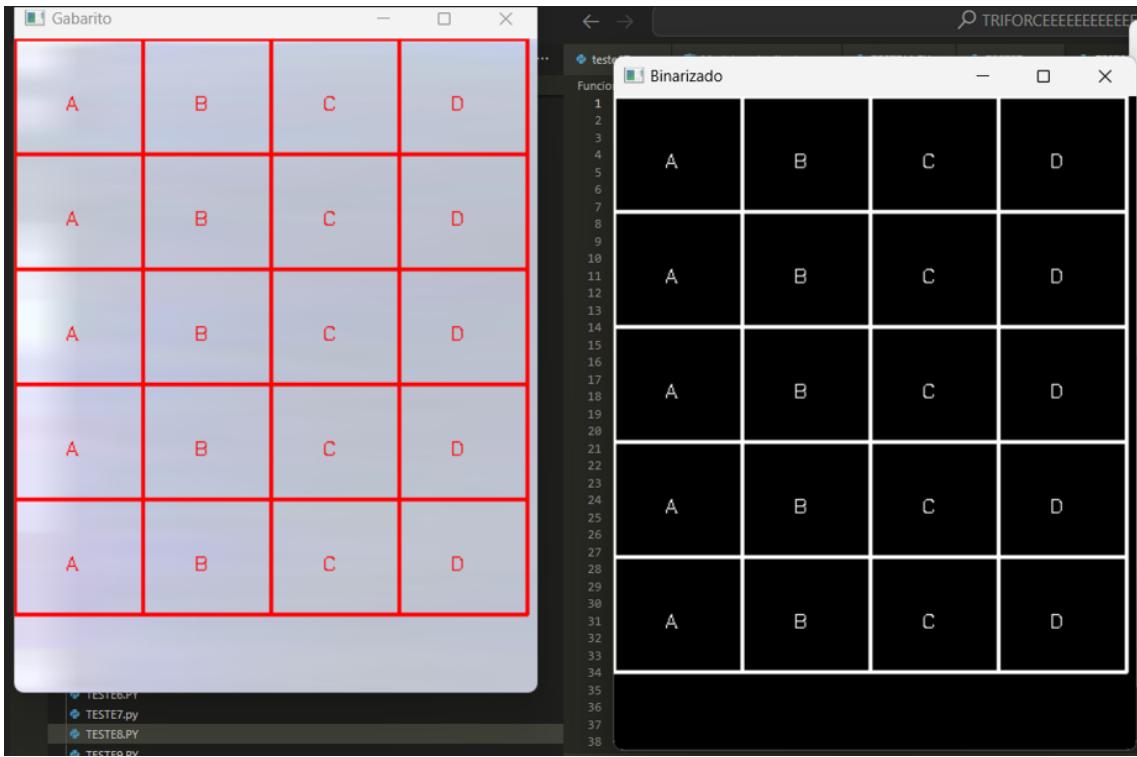


Figura 4. Matriz de quadrantes.

- Precisão: avalia a taxa de casos positivos corretos, sendo particularmente relevante quando a identificação precisa dos casos positivos é crucial (Equação 2).

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

- Recall: mede a capacidade do modelo de identificar corretamente os casos positivos, com foco na minimização de falsos negativos (Equação 3).

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

- F1-Score: é a média harmônica entre precisão e recall, fornece uma visão equilibrada do desempenho, especialmente útil em situações onde ambas as métricas são igualmente importantes (Equação 4).

$$F1 - Score = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (4)$$

4. Resultados

Nosso design experimental foi desenvolvido para testar a arquitetura proposta com foco em sua efetividade em tempo real, sem o uso de redes neurais. Para isso, desenvolvemos e implementamos experimentos abrangendo três áreas principais: avaliação de desempenho, eficiência de processamento e capacidade de generalização.

Primeiramente, realizamos uma Avaliação de Desempenho para tarefas específicas, como a detecção de padrões visuais e a correção automatizada. Utilizamos

transformações de imagem para escala de cinza e técnicas de aumento de dados (data augmentation) para simular diferentes condições de entrada, como variações de iluminação e ângulos de captura. Em seguida, medimos a acurácia da identificação em tempo real e a capacidade do sistema de lidar com essas variações, comparando os resultados com benchmarks da visão computacional.

Em termos de Eficiência de Processamento, avaliamos o tempo de execução das operações em diferentes configurações de hardware, considerando o uso de recursos computacionais e a facilidade de adaptação a novas tarefas. O experimento incluiu a análise de quão rapidamente o sistema pode ser configurado para diferentes contextos, especialmente ao aplicar transformações para escala de cinza e aumento de dados, e como isso afeta os custos computacionais.

Por fim, testamos a Capacidade de Generalização do sistema ao expô-lo a conjuntos de dados com características não previstas durante o desenvolvimento. Variamos elementos como iluminação, ângulos e contextos visuais para avaliar a robustez do sistema em novos cenários. Esse teste foi importante para confirmar a eficácia da arquitetura em condições diversas, sem depender de modelos treinados, combinando essas informações para melhorar a detecção e classificação de elementos importantes nas imagens.

Como mostrado na Tabela 1, nossos resultados experimentais demonstraram alta eficácia e confiabilidade. O sistema alcançou uma taxa de acurácia de 98,8%, superando ou igualando os benchmarks relatados na literatura. Utilizando o OpenCV para detecção e segmentação das respostas, o sistema foi capaz de lidar com variações de iluminação e desalinhamento das folhas de resposta, garantindo alto desempenho na classificação. As marcações foram validadas quando atingiam 30% de preenchimento, proporcionando flexibilidade na interpretação das respostas.

Analisamos o custo de implementação e a complexidade técnica utilizando métodos tradicionais, como a Análise por Pontos de Função (Function Points Analysis — FPA) e a Estimativa de Complexidade Técnica, amplamente utilizados para avaliar o escopo funcional e os desafios técnicos de projetos de software [Sommerville 2015]. Nesse contexto, tanto a FPA quanto a complexidade técnica são medidas em uma escala de 1 a 3: 1 representa tarefas simples, com funcionalidade mínima ou baixa exigência técnica; 2 indica tarefas moderadamente complexas; e 3 corresponde a tarefas altamente complexas, que envolvem funcionalidades avançadas ou desafios técnicos significativos. Nossa solução se destaca por sua simplicidade e desempenho em classificação, conforme ilustrado na Fig. 5.

Além disso, como mostrado na Fig.6, analisamos a distribuição das questões interpretadas corretamente e incorretamente, confirmando a alta taxa de acurácia em todos os cenários.

O sistema demonstrou ser robusto ao lidar com respostas com múltiplas marcações, alcançando 97,9% de precisão e 100% de revocação (*recall*). Esse desempenho foi crucial para garantir a correta classificação de respostas ambíguas, sem comprometer a integridade dos resultados.

A simplicidade do sistema, sua interface web responsiva e sua capacidade de operar em dispositivos convencionais permitem uma adoção fácil, especialmente em cenários que exigem mobilidade. A decisão de validar marcações com 30% de preenchimento foi

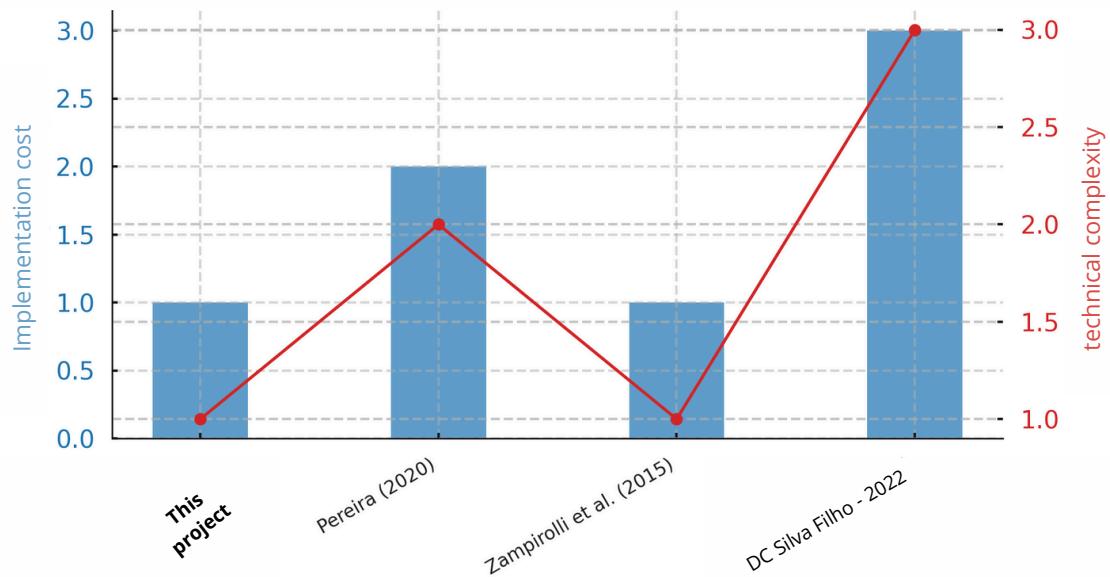


Figura 5. Custo de implementação e complexidade técnica por projeto.

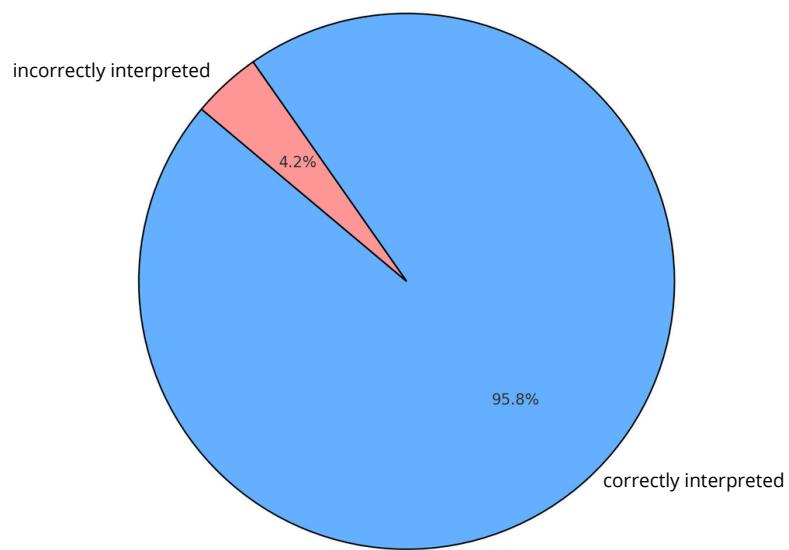


Figura 6. Distribuição de questões interpretadas correta e incorretamente.

Tabela 1. Experimental result.

Classs	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Blank Images	100	100	100	100
Correct Images	100	100	100	100
Images with multiple markings	95.8	93.75	100	96.8
General Average	98.6	97.9	100	98.9

essencial para garantir a precisão em diversas condições de uso, mesmo em situações em que a qualidade da digitalização não era ideal.

Em comparação com métodos tradicionais baseados em IA, nosso sistema oferece menor custo e complexidade, mantendo resultados competitivos. Ele pode ser utilizado em diferentes contextos educacionais sem a necessidade de hardware especializado, o que amplia sua aplicabilidade.

A aplicação de técnicas de visão computacional combinadas com a otimização de parâmetros demonstrou ser uma abordagem eficiente para a correção de folhas de resposta. O sistema é preciso, escalável e acessível, oferecendo uma solução prática para instituições educacionais que buscam automatizar seus processos avaliativos.

5. Conclusão

Os resultados deste estudo confirmam que a aplicação de técnicas de visão computacional oferece uma solução precisa para a correção automatizada de testes de múltipla escolha. Com uma taxa de acurácia demonstrada de 100%, nossa abordagem se mostra superior aos métodos tradicionais em diversos aspectos, especialmente por sua simplicidade, flexibilidade e facilidade de implementação.

Além disso, a interface intuitiva do sistema e sua capacidade de rodar em dispositivos convencionais o tornam acessível, versátil e adequado a uma ampla gama de contextos educacionais. Essa escalabilidade permite atender tanto instituições de pequeno quanto de grande porte, apresentando-se como uma alternativa prática e econômica para a correção automatizada de provas.

Trabalhos futuros devem explorar diferentes técnicas de visão computacional e investigar a integração de funcionalidades como detecção de padrões de erro, análise de desempenho estudantil e mecanismos de feedback adaptativo, com o objetivo de ampliar ainda mais a utilidade do sistema.

Referências

China, R. T., de Assis Zampirolli, F., de Oliveira Neves, R. P., and Quilici-Gonzalez, J. A. (2016). An application for automatic multiple-choice test grading on android. *Revista Brasileira de Iniciação Científica*, 3(2):4–25.

de Assis Zampirolli, F., Batista, V., and Quilici-Gonzalez, J. A. (2016). An automatic generator and corrector of multiple choice tests with random answer keys. In *2016 IEEE Frontiers in Education Conference (FIE)*, pages 1–8.

de Elias, E. M., Tasinazzo, P. M., and Hirata, R. (2021). Optical Mark Recognition: Advances, Difficulties, and Limitations. *SN Computer Science*, 2(5):367.

Dias, N. M. C. (2018). Desenvolvimento de uma nova abordagem baseada em inteligência artificial (rpa) para detecção e correção automática de problemas em processos de suporte de ti.

Fawcett, T. (2006). An introduction to roc analysis. *Pattern recognition letters*, 27(8):861–874.

Gonzalez, R. C. and Woods, R. E. (2008). *Digital Image Processing*. Pearson Prentice Hall, Upper Saddle River, NJ, USA, 3 edition.

Hafeez, Q., Aslam, W., Aziz, R., and Aldehim, G. (2024). An Enhanced Fault Tolerance Algorithm for Optical Mark Recognition Using Smartphone Cameras. *IEEE Access*, pages 1–1.

Pereira, L. V. (2019). Uma abordagem automática para leitura óptica de cartões: um estudo de caso.

Silva, F. A. et al. (2015). Aplicação do pré-processamento de imagens para otimização do reconhecimento de padrões na detecção de deficiência nutricional em espécies vegetais. *Anais XVII Simpósio Brasileiro de Sensoriamento Remoto-SBSR. João Pessoa*.

Silva Filho, D. d. C. (2022). Reconhecimento de caracteres utilizando redes neurais convolucionais para auxiliar nas correções do sistema multiprova. B.S. thesis, Universidade Federal do Rio Grande do Norte.

Sommerville, I. (2015). *Software Engineering*. Pearson, 10th edition.

Tieppo, S. M., Cappelin, A., Zanatta, L. F., Nogueira, C. M. I., and Rezende, V. (2023). Um panorama de situações do tipo misto em provas do Exame Nacional do Ensino Médio. *Ciência & Educação (Bauru)*, 29:e23009.

Zafar, M. F. (2023). Webcam-based optical mark reader for grading assessments. Master's thesis, University of Windsor (Canada).