

Estudo comparativo de técnicas de pré-processamento de imagens para otimização de OCR em rótulos alimentares

Andréia M. Ferreira¹, Ana Trindade Winck²

¹Informática Biomédica – Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSPA)
Porto Alegre – RS – Brazil

²Departamento de Ciências Exatas e Sociais Aplicadas – Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSPA)
Porto Alegre – RS – Brazil

andreia.ferreira@ufcspa.edu.br, anatw@ufcspa.edu.br

Abstract. *The increase in food allergies makes the correct identification of allergens on food labels essential, but this task is made difficult by the wide variation in terminologies and the lack of standardization of labels. This study compared image preprocessing techniques to optimize the results of a text extraction API on food labels, using two optical character recognition (OCR) libraries. Among 87 labels tested, TesseractOCR achieved 37% similarity without preprocessing, 46% with histogram equalization and binarization, and 48% with the addition of Gaussian filtering and mathematical morphology. Despite these improvements, the visual diversity of labels limits the similarity and effectiveness of the technology in real-world scenarios.*

Resumo. *O aumento das alergias alimentares torna essencial a correta identificação de alérgenos nos rótulos alimentares, mas essa tarefa é dificultada pela extensa variação de terminologias e pela falta de padronização dos rótulos. Este estudo comparou técnicas de pré-processamento de imagens para otimizar os resultados de uma API de extração de texto em rótulos alimentícios, utilizando duas bibliotecas de reconhecimento óptico de caracteres (OCR). Em 87 rótulos testados, o TesseractOCR obteve similaridade de 37% sem pré-processamento, 46% com equalização e binarização, e 48% com adição de filtro Gaussiano e morfologia matemática. Apesar dos ganhos, a diversidade visual dos rótulos limita a precisão e a eficácia da tecnologia em cenários reais.*

1. Introdução

Alergia alimentar, segundo a OMS, pode ser definida como uma reação de hipersensibilidade envolvendo o sistema imunológico [Kotchetkoff and Yonamine 2021], sendo frequentemente mediada pela Imunoglobulina E (IgE), e podendo desencadear sintomas como urticária, vômitos e, em casos extremos, levar à morte. No Brasil, estima-se que 6-8% das crianças e cerca de 3% dos adultos apresentem algum tipo de hipersensibilidade alimentar, sendo os alérgenos mais comuns leite, ovos, trigo e soja [Solé et al. 2018], tornando-a um problema de saúde pública.

Evitar alimentos desencadeantes da alergia é a principal medida preventiva, sendo essencial a clareza dos rótulos alimentares. No Brasil, a campanha "Põe no Rótulo", em

2014, resultou na RDC nº 26/2015 [Brasil 2015] da ANVISA (posteriormente RDC nº 727/2022 [Brasil 2022]), que tornou obrigatória a inclusão de advertências legíveis sobre os principais alérgenos em alimentos processados. Entretanto, em [Pereira et al. 2024] foram identificadas variações de nomenclatura para quatro principais alérgenos: 43 para leite, 23 para ovo, 41 para trigo e 35 para soja.

Além da variação de nomenclaturas, variações de texturas, cores, e materiais de embalagem (por exemplo, polipropileno e polietileno) também podem dificultar a leitura dos rótulos, bem como dificuldades individuais. Dados da PNAD Contínua de 2022 [Brasil 2023] indicam que 8,9% da população brasileira apresenta algum tipo de deficiência, incluindo deficiências visuais (3,1%) e dificuldades de aprendizagem (2,6%). Assim, o tamanho da embalagem, a distribuição das informações ou o contraste da fonte podem agravar essas dificuldades.

Assim, garantir que consumidores alérgicos leiam corretamente a lista de ingredientes continua sendo um desafio. Aplicativos móveis, como "Desrotulando" e "Loomos", oferecem auxílio. Este trabalho utiliza um aplicativo [Menezes et al. 2024] projetado para identificar alérgenos a partir de imagens de rótulos alimentares. Esse aplicativo integra APIs de Processamento de Linguagem Natural (PLN) [Gadelha et al. 2024] e Reconhecimento Óptico de Caracteres (OCR) [Menin et al. 2024], onde o OCR extrai o texto das imagens, fazendo a ponte entre as imagens capturadas e o algoritmo de PLN. Embora a API de OCR esteja funcional, o refinamento do pré-processamento de imagens é importante para alcançar maior precisão em rótulos e embalagens diversas.

Este trabalho analisa comparativamente ferramentas de OCR e técnicas de pré-processamento digital de imagens para melhorar a qualidade das imagens, otimizando os resultados da API de OCR e aumentando a precisão na identificação de alérgenos.

2. Metodologia

A metodologia deste estudo contempla a criação de uma base de imagens de rótulos alimentares e o pré-processamento das imagens.

2.1. Banco de Imagens

Não foram encontrados datasets públicos de listas de ingredientes em português, portanto, foram fotografados rótulos de diferentes alimentos, priorizando embalagens diversas, como latas de alumínio, chocolates em blister e recipientes cilíndricos. Essas imagens compõem um dataset que contempla as imagens e a transcrição dos rótulos.

2.2. Pré-processamento de imagens

Nos testes da primeira versão da API de OCR [Menin et al. 2024] do aplicativo, [Menezes et al. 2024] observou-se que embalagens com formatos ou contrastes desfavoráveis apresentaram baixo desempenho. Para otimizar os resultados do OCR, foram aplicadas técnicas de pré-processamento, como conversão para escala de cinzas, equalização de histograma com CLAHE, filtros morfológicos, e binarização adaptativa, substituindo o método global de Otsu. Também foi avaliado o desempenho do EasyOCR, baseado em deep learning e focado em imagens naturais.

Para quantificar as comparações, foram utilizadas métricas de qualidade de imagem e de reconhecimento de texto, como PSNR (Peak Signal to Noise Ratio), SSIM

(Structural Similarity Index Method), CER (Character Error Rate), WER (Word Error Rate), e Distância de Levenshtein, sendo essa última o valor padrão atribuído como percentual de similaridade.

3. Resultados

A obtenção das imagens resultou em um dataset de lista de ingredientes de rótulos alimentícios em português, composto por 87 imagens de rótulos e suas respectivas transcrições, disponível publicamente no Kaggle¹.

Os testes de pré-processamento focaram na redução de ruído e padronização de cores para melhorar o OCR. O pipeline original [Menin et al. 2024] não aplicou técnicas de pré-processamento de imagem, pois processos iguais para todas embalagens prejudicou os resultados. Para mitigar isso, o primeiro algoritmo de pré-processamento testado neste trabalho foi aplicado sobre todas as 87 imagens, convertendo-as para escala de cinza e analisando a cor de fundo original predominante para determinar o nível médio de contraste do rótulo: se o fundo for mais escuro, o algoritmo continua; senão, aplica-se a equalização de histograma. Ao fim, é feita a binarização.

Nos testes iniciais, a equalização adaptativa, especificamente a Equalização de Histograma Adaptativa com Contraste Limitado (CLAHE), mostrou melhores resultados que a global. Após a equalização, realiza-se inversão e binarização, obtendo imagens com fundo branco e texto preto, ideais para OCR. A limiarização adaptativa superou o método de Otsu na binarização.

Mesmo com filtro gaussiano, a similaridade médio ficou em (44%). Adicionar filtros morfológicos para ruídos binários elevou este número para 46%, e a combinação entre filtro gaussiano pré binarização e filtros morfológicos pós alcançou 48%. Apesar de não ultrapassarem 50%, os valores superam a primeira versão, que sem pré-processamento alcançou apenas 37% de similaridade no mesmo conjunto de imagens.

Também foi aplicado o mesmo conjunto de técnicas ao EasyOCR, obtendo 55% de similaridade. Porém, a ferramenta leva em média 1 minuto para processar cada imagem, tornando-a inviável para o contexto de um aplicativo móvel real.

Na Tabela 1 são apresentadas as comparações dos valores de similaridade de Levenshtein entre os diferentes fluxos.

Tabela 1. Comparação de similaridade de Levenshtein entre diferentes versões de OCR.

Versão	Similaridade de Levenshtein
TesseractOCR (v1)	37%
TesseractOCR sem filtros	46%
TesseractOCR com filtros	48%
EasyOCR	55%

4. Conclusão

Embora existam normas legais para a informação de ingredientes, não há padronização em sua apresentação: cores, fontes, tamanhos, ordem de alertas e disposição variam, além da

¹<https://www.kaggle.com/datasets/andreiamf/lista-ingrdientes-ptbr/>

diversidade de materiais e formatos de embalagens. Isso torna cada imagem praticamente única e dificulta a criação de um pipeline unificado de pré-processamento para OCR, já que ajustes que corrigem um caso podem prejudicar outro.

Outro desafio é a parametrização: como cada imagem tem necessidades específicas, não há valores universais. Exemplo disso foi o filtro de nitidez para imagens borras, removido do pipeline final por piorar os resultados em muitos casos. Da mesma forma, limiares e valores de binarização foram definidos por médias para evitar perdas maiores.

A ausência de foco contextual no OCR também limita a detecção em imagens complexas, como embalagens completas, e modelos pré-treinados para identificar o bloco “lista de ingredientes” provavelmente não superarão, a curto prazo, o desempenho de IAs generativas. Contudo, seu uso seria inviável no projeto devido ao alto custo por token, incompatível com o caráter gratuito do aplicativo desenvolvido.

Referências

- Brasil (2015). Resolução da diretoria colegiada nº 26, de 02 de julho de 2015. Diário Oficial da União.
- Brasil (2022). Resolução da diretoria colegiada nº 727, de 15 de julho de 2022. Diário Oficial da União.
- Brasil (2023). Pesquisa nacional por amostra de domicílios contínua: Pessoas com deficiência - 2022. Instituto Brasileiro de Geografia e Estatística.
- Gadelha, G. A., Pereira, R. A., Guedes, F. M., and Winck, A. T. (2024). Processamento de linguagem natural na identificação de alérgenos em rótulos alimentares: Uma aplicação no contexto brasileiro. *Journal of Health Informatics*, 16(Especial).
- Kotchetkoff, E. C. d. A. and Yonamine, G. H. (2021). Reações adversas a alimentos. In Pinotti, R. and Yonamine, G. H., editors, *Alergia Alimentar: Alimentação, nutrição e terapia nutricional*, pages 2–15. Manole, Barueri, SP, 1 edition.
- Menezes, G. P., Pereira, R. A., Guedes, F. M., and Winck, A. T. (2024). Desenvolvimento de um aplicativo móvel para leitura e identificação de alérgenos em rótulos alimentares. In *Anais do XX Congresso Brasileiro de Informática em Saúde*, pages 91–95, São Paulo. Sociedade Brasileira de Informática em Saúde.
- Menin, G., Pereira, R. A., Guedes, F. M., and Winck, A. T. (2024). API rest para reconhecimento óptico de caracteres em rótulos alimentares. *Journal of Health Informatics*, 16(Especial).
- Pereira, R. A., Guedes, F. M., Gadelha, G. A., Menin, G., Menezes, G. P., Amantéa, S. L., and Winck, A. T. (2024). Alérgenos alimentares em rótulos de produtos alimentícios: uma unificação da nomenclatura. *Arquivos de Asma, Alergia e Imunologia*, 8(2):125–128.
- Solé, D., Silva, L. R., Cocco, R. R., Ferreira, C. T., Sarni, R. O., Oliveira, L. C., Pastorino, A. C., Weffort, V., Moraes, M. B., Barreto, B. P., et al. (2018). Consenso brasileiro sobre alergia alimentar: 2018 - parte 2 - diagnóstico, tratamento e prevenção. documento conjunto elaborado pela sociedade brasileira de pediatria e associação brasileira de alergia e imunologia. *Arquivos de Asma, Alergia e Imunologia*, 2(1):39–82.