

Software para Identificação de Ingredientes Alergênicos em Rótulos de Cosméticos

Mariana Emília Rossatto de Moraes¹, Rogerio Xavier de Azambuja¹

¹Instituto Federal do Rio Grande do Sul - Campus Farroupilha - Av. São Vicente, 785
CEP 95.174-274 - Bairro Cinquentenário - Farroupilha - RS - Brasil

mariana.rossatto22@gmail.com, rogerio.xavier@farroupilha.ifrs.edu.br

Abstract. *This study proposes developing a mobile application designed for consumers with allergies or sensitivities to cosmetic products. The goal is to facilitate the identification of ingredients listed on labels by utilizing Optical Character Recognition (OCR) technology to automate reading and consultation based on the International Nomenclature of Cosmetic Ingredients (INCI). The tool aims to expedite and enhance the process of selecting cosmetics, minimizing the risks associated with manual label reading. The study encompasses analyzing requirements, system modeling, implementation, and validation testing, utilizing accuracy and sensitivity metrics to assess the application's performance.*

Keywords: *mobile development, Optical Character Recognition (OCR), cosmetics, allergies.*

Resumo. *Este trabalho apresenta o desenvolvimento de um aplicativo mobile voltado para consumidores com alergias ou sensibilidades a produtos cosméticos. O objetivo é facilitar a identificação dos ingredientes presentes nos rótulos utilizando tecnologia de Reconhecimento Óptico de Caracteres (OCR) para automatizar a leitura e consulta baseada na Nomenclatura Internacional de Ingredientes Cosméticos (INCI). A ferramenta visa tornar o processo de escolha de cosméticos mais rápido e seguro, minimizando os riscos da leitura manual de rótulos. O estudo inclui a análise de requisitos, a modelagem do sistema, a implementação e os testes de validação, utilizando métricas de acurácia e sensibilidade para avaliar o desempenho do aplicativo.*

Palavras-chave: *desenvolvimento mobile, Reconhecimento Óptico de Caracteres (OCR), cosméticos, alergias.*

1. Introdução

O setor de cosméticos tem vivenciado uma expansão notável nas últimas décadas, impulsionado pela crescente demanda global por produtos inovadores e de alta performance, além da motivação por uma aparência mais agradável. Esse crescimento é evidenciado pelo consumo massivo de uma vasta gama de itens, desde maquiagens e perfumes até produtos essenciais de higiene pessoal (Kumari et al., 2024). No contexto brasileiro, essa tendência é ainda mais acentuada, visto que o país consolidou-se como o terceiro maior mercado consumidor de produtos de higiene pessoal, perfumaria e

cosméticos em 2024, atrás apenas da China e dos Estados Unidos (ABIHPEC, 2025). Contudo, essa maior frequência de uso tem sido acompanhada por um preocupante aumento na ocorrência de reações adversas associadas às composições desses produtos, como alergias, dermatites e irritações cutâneas (Kumari et al., 2024).

A Nomenclatura Internacional de Ingredientes Cosméticos (INCI) foi criada para padronizar e facilitar a identificação dos componentes em produtos cosméticos globalmente. Seu objetivo principal é proteger a saúde pública, permitindo que consumidores e profissionais de saúde identifiquem rapidamente substâncias que possam causar danos ou alergias (ANVISA, 2020). Contudo, essa padronização, embora essencial do ponto de vista sanitário e técnico, não elimina a necessidade de verificação manual dos rótulos. Para o consumidor médio, a INCI ainda representa uma linguagem complexa e distante do vocabulário comum, como ressalta a Dra. Fátima Emerson, Coordenadora de Assuntos Comunitários da Associação Brasileira de Alergia e Imunologia (ASBAI): "Ainda temos outras dificuldades, como, por exemplo, os rótulos desses produtos, que trazem informações confusas, em uma linguagem desconhecida do público leigo" (ASBAI, 2017). Essa barreira de compreensão compromete a capacidade do consumidor de identificar os ingredientes aos quais possui sensibilidade, tornando a escolha de produtos seguros um desafio diário.

Diante desse cenário, o Reconhecimento Óptico de Caracteres (OCR) surge como uma tecnologia promissora para automatizar a extração e interpretação de ingredientes em rótulos de cosméticos. Baseado em princípios de visão computacional, o OCR converte textos presentes em imagens para formatos editáveis e processáveis por sistemas computacionais (Gonzalez e Woods, 2017). No contexto dos cosméticos, essa tecnologia pode eliminar a necessidade de leitura manual ao transformar os textos dos rótulos em dados digitais, facilitando a identificação de ingredientes e a verificação de substâncias alergênicas. Embora tradicionalmente limitado por desafios como deformações de imagem e variações de fonte e tamanho dos caracteres (Impedovo et al., 1991), avanços recentes em *machine learning* têm ampliado sua precisão mesmo em condições não ideais, tais como rótulos com reflexos, fontes irregulares ou *layouts* complexos (Hegghammer, 2022). Apesar dessas limitações, o OCR se revela uma ferramenta de grande potencial para ampliar a segurança no consumo de cosméticos, especialmente entre indivíduos com alergias ou sensibilidades específicas.

Este trabalho tem como objetivo geral verificar a viabilidade da utilização da tecnologia de Reconhecimento Óptico de Caracteres (OCR) na identificação de ingredientes alergênicos em rótulos de produtos cosméticos, por meio do desenvolvimento de um aplicativo *mobile*. De forma específica, este estudo visa: (i) mapear os principais ingredientes alergênicos com base na INCI e criar uma base de dados; (ii) integrar uma ferramenta de OCR para extrair o texto de rótulos cosméticos; (iii) implementar um sistema para identificar ingredientes alergênicos a partir do texto

extraído; (iv) projetar a interface do usuário do aplicativo; e (v) avaliar o desempenho do reconhecimento de caracteres e da detecção de alérgenos por meio de testes, utilizando métricas de acurácia e sensibilidade.

A justificativa para a construção deste aplicativo reside na crescente dificuldade que muitos consumidores enfrentam ao ler rótulos de produtos cosméticos, especialmente aqueles com alergias ou sensibilidades. Como destacam Kumar et al. (2023), as soluções existentes, como a inspeção manual de rótulos, são ineficazes e inconvenientes, tornando-se um processo demorado e propenso a erros. Contudo, em um cenário em que o uso de cosméticos faz parte da rotina e as reações adversas têm se tornado mais frequentes, torna-se essencial identificar com precisão os ingredientes desses produtos (Kumari et al., 2024; Canhedo e Reis, 2016). Essa demanda também se reflete em uma vivência pessoal da autora, que convive com a doença celíaca, uma condição autoimune que exige a exclusão total do glúten, inclusive em cosméticos e produtos de higiene, os quais podem causar reações dermatológicas. Ainda assim, essa realidade não é isolada, visto que indivíduos com sensibilidades a conservantes, fragrâncias ou outros componentes enfrentam desafios semelhantes. Portanto, embora motivado por uma necessidade pessoal, o trabalho proposto visa oferecer uma solução acessível e eficaz para um público mais amplo que busca segurança e praticidade na escolha de cosméticos.

O restante deste artigo está organizado da seguinte forma: a Seção 2 apresenta o referencial teórico, oferecendo os conceitos e fundamentos necessários para a compreensão do tema abordado. A Seção 3 descreve os procedimentos metodológicos adotados e as tecnologias empregadas no desenvolvimento deste projeto. Na Seção 4, é detalhado o processo de desenvolvimento do sistema proposto. A Seção 5 apresenta os resultados obtidos, enquanto a Seção 6 traz as considerações finais.

2. Referencial Teórico

Nesta seção, são apresentados os principais conceitos teóricos que fundamentam o trabalho desenvolvido.

2.1. Reconhecimento Óptico de Caracteres (OCR)

Segundo Gonzalez e Woods (2017), a visão computacional é uma área da inteligência artificial que visa simular a visão humana, capacitando os computadores a aprender, fazer inferências e tomar decisões com base em entradas visuais. Uma das aplicações da visão computacional é o OCR, sigla para *Optical Character Recognition* ou Reconhecimento Óptico de Caracteres. Essa tecnologia permite capturar textos presentes

em documentos ou imagens e convertê-los em formatos editáveis para serem utilizados em processamentos computacionais.

De acordo com Mithe, Indalkar e Divekar (2013), o funcionamento geral do OCR pode ser dividido nas seguintes etapas:

- **Digitalização:** a primeira etapa do OCR é a digitalização do material original, que consiste em capturar sua imagem por meio de um *scanner*, câmera ou outro dispositivo. O objetivo é transformar o conteúdo físico em uma imagem digital, que possa ser manipulada *pixel a pixel* pelos algoritmos de OCR.
- **Segmentação:** identifica e separa as áreas que contêm texto de regiões com outros elementos visuais. Essa segmentação é essencial para as próximas etapas, permitindo que os caracteres sejam passados para os mecanismos de reconhecimento de padrões.
- **Pré-processamento:** visa melhorar a qualidade da imagem a fim de facilitar o reconhecimento dos caracteres. Para isso, a imagem é submetida a diversas técnicas, como remoção de ruídos, realce do contraste e correção de distorções, por exemplo. O objetivo é transformar a imagem de modo a eliminar interferências visuais, tornando-a mais simples e limpa.
- **Extração de características:** nesta etapa, o sistema analisa cada caractere e identifica suas características específicas. Essas características podem incluir a forma, tamanho, espaços abertos, curvatura e linhas do caractere. O objetivo é capturar informações essenciais que ajudem a diferenciar um caractere de outro e preparar o sistema para a próxima etapa de reconhecimento.
- **Reconhecimento:** nesta fase final, o sistema de OCR utiliza algoritmos para associar as características extraídas de cada caractere a um símbolo textual correspondente. O resultado é a conversão desses elementos em caracteres digitais que compõem o texto reconhecido.

2.2. Principais Dificuldades e Limitações do OCR

Por um longo período, ferramentas OCR amplamente utilizadas, como o Tesseract, produziam resultados precisos apenas em cenários idealizados, isto é, textos padronizados e sem interferências visuais. No entanto, quando aplicadas a documentos do mundo real com características complexas e degradadas, essas ferramentas apresentavam limitações no reconhecimento do texto, dificultando a extração precisa das informações (Hegghammer, 2022).

Contudo, a precisão do OCR está profundamente ligada à qualidade da imagem capturada. É interessante notar que os problemas identificados por Impedovo et al.

(1991) permanecem relevantes nos dias de hoje, demonstrando a complexidade e a persistência dos desafios relacionados ao OCR. Os problemas apontados incluem:

- Discriminação de formas: a diversidade de fontes e estilos de caracteres, bem como a semelhança visual entre certos caracteres, como “U” e “V” ou “O” e “0”, são especialmente desafiadores para sistemas OCR.
- Deformação da imagem: fatores como ruído, translação e rotação podem distorcer a imagem e dificultar o reconhecimento preciso.
- Variações de tamanho: diferenças no tamanho e na densidade dos caracteres (conhecida como *pitch* ou número de caracteres por polegada) podem comprometer a precisão do OCR.

No contexto mais recente de aplicação do reconhecimento óptico em rótulos e embalagens de produtos, esses desafios permanecem evidentes. Pesquisas atuais apontam como fatores visuais impactam diretamente a eficácia dessa tecnologia em diferentes tipos de produtos. Rimdusit et al. (2024), em sua pesquisa sobre rótulos de medicamentos, identificaram que condições da imagem, como ruídos, desfoques e reflexos, assim como características estruturais dos rótulos e embalagens, incluindo o alinhamento irregular das informações e a curvatura das embalagens, comprometem significativamente o desempenho do OCR. Čakić et al. (2020), por sua vez, observaram que ferramentas de OCR comerciais apresentam dificuldades ao lidar com características visuais de rótulos de garrafas de vinho, como sombras, arredondamentos e linhas horizontais. Por fim, Severo (2024) ao testar ferramentas OCR em rótulos de cosméticos, destacou que a resolução, o tamanho e o espaçamento dos caracteres são cruciais para a extração precisa de textos.

Na última década, os avanços em *machine learning* trouxeram melhorias significativas no desempenho dessas ferramentas (Hegghammer, 2022). Mesmo com os progressos tecnológicos, o reconhecimento óptico de caracteres em rótulos continua sendo uma tarefa complexa e desafiadora. Em de Azambuja et al. (2023), que trabalhou com o processamento de rótulos de vinhos, é informado que o desafio exige vários reprocessamentos até se encontrar um modelo de reconhecimento aceitável e que possa proporcionar bons resultados esperados. Ainda, como observado por Kumar et al. (2023, pág. 5, tradução nossa):

Condições de iluminação ruins, distorções ou irregularidades no posicionamento de rótulos podem afetar a qualidade da captura de imagens, o que, por sua vez, prejudica o desempenho do modelo. Compreender o significado contextual das informações nos rótulos é um desafio. Os produtos podem ter diversos designs, fontes e layouts. Adaptar o sistema para lidar com essa variabilidade e garantir que ele possa processar com precisão informações de rótulos com diferentes estruturas é uma tarefa complexa.

Em síntese, mesmo com os avanços recentes, a aplicação da tecnologia de OCR em ambientes visuais não estruturados, como rótulos de produtos, continua sendo um desafio técnico relevante. Entender essas limitações é fundamental para desenvolver soluções mais eficazes e adequadas às situações do mundo real.

2.3. O Uso de Cosméticos e o Aumento de Reações Adversas

Conforme a Resolução da Diretoria Colegiada (RDC) nº 907/2024 da Anvisa, os produtos de higiene pessoal, cosméticos e perfumes são definidos como:

preparações constituídas por substâncias naturais ou sintéticas, de uso externo nas diversas partes do corpo humano, pele, sistema capilar, unhas, lábios, órgãos genitais externos, dentes e membranas mucosas da cavidade oral, com o objetivo exclusivo ou principal de limpá-los, perfumá-los, alterar sua aparência e ou corrigir odores corporais e ou protegê-los ou mantê-los em bom estado (BRASIL, 2024, Art. 3º, XVI).

Nos últimos vinte anos, o setor de cosméticos tem experimentado uma expansão significativa, com um crescimento anual de aproximadamente 4,5%. A motivação por uma aparência mais agradável, aliada a uma percepção pública por vezes superficial sobre a segurança dos cosméticos, resultou em um aumento massivo em sua utilização. Essa tendência é evidenciada pelo consumo generalizado de uma variedade de produtos, desde maquiagem e perfumes até itens de higiene pessoal como xampus e cremes dentais (Kumari et al., 2024).

No contexto brasileiro, segundo a Associação Brasileira da Indústria de Higiene Pessoal, Perfumaria e Cosméticos (ABIHPEC, 2025), em 2024 o Brasil foi o terceiro maior mercado consumidor de produtos de higiene pessoal, perfumaria e cosméticos do mundo, ficando atrás apenas da China e dos Estados Unidos. Contudo, a maior frequência no uso desses produtos também tem levado a um aumento na ocorrência de reações adversas associadas às suas composições (Kumari et al., 2024).

As reações adversas a cosméticos podem se manifestar de diversas formas, com sintomas que variam desde irritações leves até condições dermatológicas mais complexas. Dados do sistema Notivisa da Anvisa, indicam que, entre 2006 e 2018, os eventos adversos relacionados a cosméticos mais notificados foram irritação (46,0%), alergia (30,5%) e ardor (30,5%) (Teixeira et al., 2019). De acordo com a Associação Brasileira de Alergia e Imunologia (ASBAI, 2017), os sintomas mais comuns das alergias a cosméticos incluem vermelhidão, inchaço, coceira e formação de bolhas. Além disso, Canhedo e Reis (2016) destacam que os cosméticos podem desencadear dermatite de contato (irritativa ou alérgica), desidrose, dermatite atópica, dermatite seborreica, dermatite numular e dermatite de estase, que são doenças inflamatórias da camada superior da pele, apresentando sintomas que variam de rubor discreto a processos inflamatórios graves. Ademais, González-Muñoz et al. (2014) ressaltam que a

dermatite de contato provocada por cosméticos representa de dois a quatro por cento das consultas dermatológicas, sendo aproximadamente 60% dos casos de origem alérgica, com os principais agentes sensibilizantes sendo fragrâncias, conservantes e tinturas capilares.

Diante dos riscos associados às reações alérgicas a cosméticos, a leitura atenta dos rótulos torna-se uma prática fundamental. Canhedo e Reis (2016) destacam que, após a identificação das substâncias alergênicas, é imprescindível que o paciente seja orientado quanto aos nomes químicos desses compostos, bem como sobre a necessidade de verificar sua presença nos rótulos de novos produtos, a fim de evitar a exposição. No entanto, essa orientação esbarra em desafios relevantes, como a dificuldade de compreensão das informações presentes nos rótulos, que frequentemente utilizam termos técnicos ou pouco acessíveis ao público leigo (ASBAI, 2017).

Além disso, há críticas à forma como muitos cosméticos são comercializados, uma vez que campanhas publicitárias de cosméticos induzem os consumidores a acreditarem que os produtos são “naturais, seguros e saudáveis”, mesmo quando essa imagem não corresponde à realidade de sua composição (Samanta, 2022). Pereira e Santos (2024) reforçam essa crítica ao destacarem que o uso indiscriminado de termos como “hipoalergênico”, “pele sensível”, “para crianças”, “natural” e “livre de fragrância” como estratégias de marketing evidencia uma lacuna regulatória, uma vez que tais alegações não garantem menor potencial alergênico. A confiabilidade desses rótulos é, portanto, questionável, pois permite que produtos com potenciais alérgenos sejam apresentados de maneira enganosa ao consumidor.

2.4. Nomenclatura Internacional de Ingredientes Cosméticos (INCI)

A Nomenclatura Internacional de Ingredientes Cosméticos, conhecida pela sigla INCI, é um sistema internacional de codificação que padroniza a identificação dos ingredientes utilizados em produtos cosméticos. Diferente do simples uso do idioma inglês, trata-se de uma nomenclatura internacional, criada com o objetivo de promover a padronização da rotulagem dos ingredientes, facilitando sua identificação por consumidores, profissionais de saúde e órgãos de vigilância sanitária (ANVISA, 2020).

Este sistema surgiu de um esforço conjunto de diversos países, iniciando nos Estados Unidos e expandindo-se para regiões como União Europeia, Mercosul, Comunidade Andina, Japão, Austrália, entre outros. A definição dos nomes das substâncias segue regras específicas, sob a supervisão de um comitê internacional composto por representantes de órgãos reguladores como a Administração de Alimentos e Medicamentos (FDA) dos Estados Unidos, a Comissão Europeia e os ministérios da saúde do Canadá e do Japão (ANVISA, 2020).

Anteriormente à adoção do INCI, os ingredientes cosméticos podiam ser identificados de diferentes formas, seja pela nomenclatura química, pelo nome comercial ou até por nomes variados atribuídos pelos fornecedores. Essa diversidade de denominações dificultava tanto o controle sanitário quanto a compreensão por parte dos consumidores, especialmente daqueles que possuem alergias, sensibilidades ou restrições a determinados componentes (ANVISA, 2020).

O uso da nomenclatura INCI contribui significativamente para a proteção da saúde pública. Ela garante à comunidade médica uma comunicação científica consistente, facilitando a identificação de agentes causadores de reações adversas. Para os cientistas, evita confusões decorrentes de múltiplos nomes para a mesma substância, assegurando precisão na troca de informações. A indústria, por sua vez, consegue monitorar com mais eficiência a segurança e a conformidade regulatória dos ingredientes em escala global. Além disso, oferece transparência aos consumidores, que podem identificar os ingredientes, independentemente do país de origem do produto (Personal Care Products Council, 2025).

2.5. Regulamentação da INCI no Brasil

No Brasil, a INCI foi adotada como parte da harmonização de regulamentações sanitárias no âmbito do Mercosul, sendo a Agência Nacional de Vigilância Sanitária a entidade responsável por sua regulamentação e fiscalização (ANVISA, 2020). Em relação à composição e rotulagem de produtos cosméticos, a Resolução da Diretoria Colegiada (RDC) nº 907/2024 da Anvisa estabelece as seguintes definições:

IX - ingredientes ou composição: descrição qualitativa dos componentes da fórmula através de sua designação genérica, utilizando a codificação de substâncias estabelecida pela Nomenclatura Internacional de Ingredientes Cosméticos (INCI, sigla do inglês);

XX - rótulo ou rotulagem: identificação impressa ou litografada, bem como dizeres pintados ou gravados, decalque sob pressão ou outras técnicas, aplicadas diretamente sobre recipientes, embalagens, invólucros, envoltórios ou qualquer outro protetor de embalagens (BRASIL, 2024, Art. 3º, Seção II).

Recentemente, a Anvisa publicou a RDC nº 898/2024, a qual trata da obrigatoriedade de descrever a composição desses produtos em língua portuguesa. Conforme o Art. 2º da referida RDC, a inclusão da composição em português é obrigatória e deve ser feita sem prejuízo da continuidade da utilização da INCI. A norma estabelece algumas opções para a apresentação da tradução:

- Incluída no rótulo original em etiqueta complementar, desde que esta garanta a integridade das cores do material, a fim de evitar sua remoção (BRASIL, 2024, Art. 2º, § 2º).
- Apresentada em formato digital, acessível preferencialmente via leitura de um código no rótulo por dispositivo móvel, diretamente à visualização da composição do produto específico, devendo sempre ser precedida da frase "Composição (português):" ou "Ingredientes (português):" (BRASIL, 2024, Art. 2º, § 3º).
- Escrita entre parênteses ou após barra, ao lado da respectiva descrição em INCI ou em uma segunda lista, na mesma ordem dos ingredientes em INCI, precedida de "Composição (português):" ou "Ingredientes (português):" (BRASIL, 2024, Art. 2º, § 4º).

Dessa maneira, a tradução da composição dos ingredientes deve utilizar, obrigatoriamente, a lista oficial de ingredientes em português disponibilizada e periodicamente atualizada pela Anvisa, por meio do Painel de Traduções INCI (ANVISA, 2023).

A exigência da tradução entrou em vigor a partir de 1º de novembro de 2023, tornando-se obrigatória para todos os produtos de higiene pessoal, cosméticos e perfumes comercializados no Brasil (ANVISA, 2023). Para viabilizar a adaptação do setor, a norma estabelece regras transitórias, em que, para as empresas, há um prazo máximo de 36 meses para esgotar os estoques de rotulagem com traduções em desacordo, contando a partir da inclusão ou atualização do ingrediente na lista. Contudo, produtos fabricados até 31 de outubro de 2023 podem ser comercializados até a validade. Sendo que, o não cumprimento destas regras constitui infração sanitária (BRASIL, 2024, Art. 3º, § 2º e § 3º; Art. 4º; Art. 6º).

3. Procedimentos Metodológicos

Os procedimentos metodológicos descrevem a sequência de passos adotados para alcançar os objetivos propostos neste trabalho. A engenharia de *software*, de acordo com Sommerville (2018), utiliza uma abordagem sistemática conhecida como processo de *software*, que consiste em uma sequência organizada de atividades para a produção eficaz de um sistema, cujo objetivo central consiste em especificar, projetar, implementar e testar o *software*. Seguindo esse ciclo, o desenvolvimento do aplicativo foi estruturado nas seguintes etapas:

- Análise e classificação dos requisitos: a primeira etapa envolveu uma análise detalhada dos requisitos do sistema, buscando identificar as necessidades do usuário e as funcionalidades necessárias. Os requisitos foram classificados e

organizados para garantir que todos os aspectos essenciais do aplicativo fossem contemplados.

- **Modelagem do sistema:** com base nos requisitos levantados, foi realizada a modelagem do sistema por meio de diagramas que foram fundamentais para visualizar a interação do usuário com o sistema e definir a estrutura de dados e a organização geral do aplicativo.
- **Implementação:** a implementação do sistema seguiu a arquitetura definida na etapa anterior. Para o *backend*, utilizou-se a linguagem de programação Python (Python, 2024) e o *framework* Django (Django, 2024) para a criação da Interface de Programação de Aplicações (API), responsável pela comunicação com a interface e pela interação com o banco de dados no PostgreSQL (PostgreSQL, 2024). No *frontend*, a interface do usuário foi desenvolvida com o *framework* Flutter (Flutter, 2024). Ademais, uma etapa crucial da implementação envolveu a escolha da ferramenta de reconhecimento óptico de caracteres, sendo avaliadas as opções Pytesseract (Pytesseract, 2024) e Cloud Vision AI (Google Cloud, 2024).
- **Testes e validação:** a fase de testes e validação foi realizada de forma contínua durante o desenvolvimento do sistema, com o objetivo de verificar os requisitos e garantir o bom funcionamento do aplicativo. Para medir o desempenho da aplicação, foram empregadas métricas de acurácia e sensibilidade, avaliando a precisão no reconhecimento de caracteres e na identificação de ingredientes alergênicos em rótulos com diferentes *layouts* e características visuais.

3.1. Tecnologias

Para o desenvolvimento deste projeto foram selecionadas tecnologias que atendem às demandas específicas da aplicação. A seguir, descrevem-se as ferramentas escolhidas e suas respectivas funcionalidades.

3.1.1. Interface de Programação de Aplicações (API)

API é a sigla para *Application Programming Interface* (Interface de Programação de Aplicações). Trata-se de um conjunto de operações que disponibiliza por meio de rotas o acesso às funcionalidades de um sistema, permitindo que outros programas as utilizem diretamente, sem necessidade de interação através de uma interface de usuário (Sommerville, 2018).

3.1.2. Python

Python é uma linguagem de programação interpretada, de alto nível e de propósito geral, amplamente reconhecida por sua sintaxe clara e concisa, o que promove a legibilidade do código e agilidade no desenvolvimento. Sua vasta biblioteca padrão e o grande ecossistema de pacotes de terceiros a tornam uma escolha versátil para diversas aplicações (Python, 2024).

3.1.3. PyTesseract

PyTesseract é uma biblioteca Python que serve como interface para o Tesseract OCR, uma das ferramentas de reconhecimento óptico de caracteres de código aberto mais utilizadas para extração de texto a partir de imagens. Ambos são gratuitos, sendo distribuídos sob a licença Apache 2.0, o que permite livre uso, modificação, distribuição e aplicação em projetos comerciais, sem custos (PyTesseract, 2024).

3.1.4. Cloud Vision AI

Trata-se de uma API comercial que permite integrar recursos de detecção de visão em aplicações. Utilizando modelos de aprendizado de máquina de visão computacional pré-treinados do Google, o Cloud Vision AI pode realizar diversas tarefas, como reconhecimento óptico de caracteres (OCR), rotulagem de imagens, detecção facial, dentre outros. Diferentemente de soluções de código aberto, essa ferramenta opera em um modelo pago conforme o uso (Google Cloud, 2024).

3.1.5. Django

Django é um *framework* web de alto nível para Python, conhecido por sua simplicidade e rapidez no desenvolvimento. Neste projeto, foram utilizadas duas ferramentas do Django: o Django Rest Framework (DRF), que facilita a criação de APIs RESTful, e o Django Object-Relational Mapper (ORM), que permite uma interação eficiente com o banco de dados (Django, 2024).

3.1.6. PostgreSQL

PostgreSQL é um sistema de gerenciamento de banco de dados relacional de código aberto que utiliza a *Structured Query Language* (SQL). Além de oferecer diversos recursos para auxiliar desenvolvedores em seus aplicativos, o PostgreSQL é reconhecido pela sua confiabilidade, integridade de dados, conjunto robusto de recursos e extensibilidade (PostgreSQL, 2024).

3.1.7. Flutter

Flutter é um *framework* de código aberto, mantido pela Google, voltado ao desenvolvimento de interfaces de usuário. Ele permite criar aplicativos compilados nativamente para dispositivos móveis, web e *desktop* a partir de uma única base de código usando a linguagem de programação Dart (Flutter, 2024).

4. Desenvolvimento

Esta seção descreve o processo de desenvolvimento do aplicativo proposto, detalhando as etapas práticas realizadas para sua construção. A aplicação foi desenvolvida para plataformas *mobile*, aproveitando funcionalidades nativas dos dispositivos móveis, como a câmera e o armazenamento local, com o objetivo de oferecer maior praticidade ao usuário. O desenvolvimento adotou uma abordagem incremental, cujas etapas são apresentadas nas subseções a seguir.

4.1. Análise e Classificação dos Requisitos

A primeira etapa do desenvolvimento foi a análise e classificação de requisitos. De acordo com Sommerville (2018), os requisitos de um sistema descrevem os serviços que ele deve fornecer e as limitações da sua operação, representando as demandas dos usuários em um contexto específico. O autor classifica esses requisitos em funcionais, não funcionais e de sistema. Os requisitos funcionais descrevem o que o sistema deve fazer e como deve se comportar em situações específicas. Já os requisitos não funcionais impõem restrições ou especificações sobre características do sistema, como desempenho, confiabilidade e uso de memória. Os requisitos de sistema, por sua vez, detalham especificações técnicas relacionadas à implementação. Além disso, as regras de negócio definem como o sistema deve operar dentro de certos limites. As tabelas 1, 2, 3 e 4 apresentam os requisitos do proposto sistema de forma detalhada.

Tabela 1. Relação de requisitos funcionais

Requisitos funcionais	
RF01	O sistema deve permitir três formas de entrada do rótulo: captura de imagem pela câmera, seleção de imagem da galeria ou inserção direta de texto.
RF02	O sistema deve permitir ao usuário selecionar as categorias de alérgenos que deseja verificar na análise, como glúten, corante, fragrância, entre outras, funcionando como filtros de busca.
RF03	O sistema deve utilizar OCR para extrair o texto contido na imagem enviada.

RF04	O sistema deve informar, para cada categoria selecionada, se foram ou não encontrados ingredientes alergênicos, listando-os quando aplicável.
RF05	O sistema deve permitir o armazenamento das imagens ou textos analisados, bem como seus respectivos resultados no histórico do usuário.
RF06	O usuário pode visualizar o histórico de análises anteriores.

Fonte: Elaboração própria, 2024

Tabela 2. Relação de requisitos não funcionais

Requisitos não funcionais	
RNF01	O sistema deve processar a análise dos rótulos em tempo razoável, indicando o progresso ao usuário.
RNF02	O sistema deve ser acessível em dispositivos móveis.
RNF03	O sistema deve ter uma interface intuitiva e fácil de usar.

Fonte: Elaboração própria, 2024

Tabela 3. Relação de requisitos do sistema

Requisitos do sistema	
RS01	O sistema deve integrar tecnologia de OCR para realizar o reconhecimento óptico de caracteres em rótulos de produtos.
RS02	O sistema deve utilizar um banco de dados estruturado por categorias de alérgenos, contendo os ingredientes com nomenclatura INCI e suas traduções para o português.

Fonte: Elaboração própria, 2024

Tabela 4. Relação de regras de negócio

Regras de negócio	
RN01	Para cada categoria de alérgeno selecionada, o sistema deve indicar se o produto é "livre" ou "não livre", com base na presença de ingredientes dessa categoria no rótulo analisado.
RN02	O sistema deve exibir apenas componentes alergênicos que correspondam às categorias selecionadas pelo usuário na análise do rótulo.
RN03	O sistema não deve limitar a quantidade de filtros que o usuário pode aplicar simultaneamente na análise de um rótulo.

RN04	O sistema deve exibir uma notificação de erro ao usuário se não for possível identificar texto na imagem enviada.
RN05	O histórico de análises deve ser armazenado localmente no dispositivo móvel e estar acessível por meio da interface do aplicativo, permitindo consultas futuras.
RN06	O sistema deve seguir as diretrizes da INCI para garantir a correta identificação dos ingredientes alergênicos presentes nos rótulos.

Fonte: Elaboração própria, 2024

4.2. Modelagem

Após a identificação dos requisitos, a etapa seguinte do desenvolvimento foi a modelagem do sistema. Nesta seção são apresentados os artefatos de modelagem produzidos.

4.2.1. Casos de Uso

Com o objetivo de facilitar a visualização das funcionalidades essenciais do sistema, foi desenvolvido um diagrama de casos de uso, conforme ilustrado na Figura 1. De acordo com Sommerville (2018), essa ferramenta de modelagem representa de forma simplificada as interações entre o sistema e seus usuários, contribuindo para a compreensão das funcionalidades e requisitos.

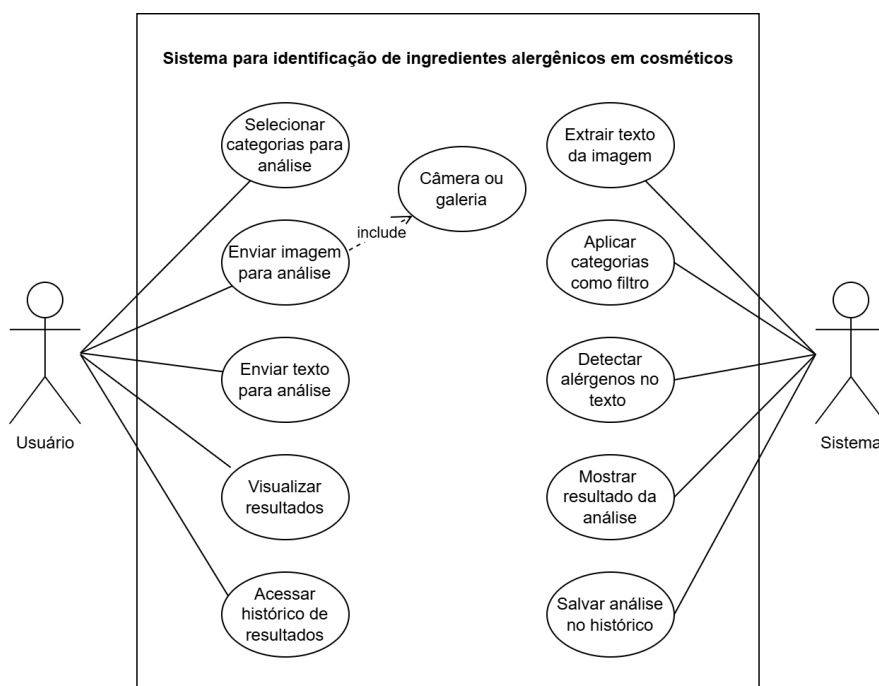


Figura 1. Diagrama de casos de uso

Fonte: Elaboração própria, 2024

A ação central do usuário é enviar imagens ou textos de rótulos de cosméticos para análise. Antes do envio, o usuário seleciona as categorias de alérgenos que deseja verificar, como corantes, fragrâncias e outros. O sistema, caso receba uma imagem como entrada, utiliza ferramentas de OCR para extrair o texto contido nela. Em seguida, realiza a detecção de alérgenos no texto, considerando as categorias escolhidas pelo usuário. Por fim, exibe os resultados da análise e armazena os dados no histórico de consultas, permitindo consultas futuras.

4.2.2. Modelagem Relacional

Diante da necessidade de armazenar e organizar os dados de forma estruturada e eficiente, foi necessário definir uma abordagem adequada para a construção do banco de dados do sistema. Para isso, adotou-se a modelagem relacional, que representa os dados por meio de tabelas, atributos e relacionamentos. Segundo Heuser (1999), essa modelagem é realizada em nível lógico e tem como objetivo preparar os dados para sua implementação em sistemas gerenciadores de banco de dados relacionais.

Para representar essa estrutura foi elaborado um diagrama relacional, apresentado na Figura 2, que descreve as principais tabelas do banco de dados. São elas:

- **Categoria:** representa os diferentes tipos de alérgenos presentes em cosméticos, como corantes, fragrâncias, glúten, entre outros. Essa tabela serve para organizar os ingredientes alergênicos em grupos, sendo que cada categoria pode ter múltiplos ingredientes associados.
- **Ingrediente:** armazena informações sobre os ingredientes alergênicos, incluindo o nome segundo a nomenclatura INCI, a tradução para o português e a categoria à qual pertence. Cada ingrediente pode pertencer a uma ou mais categorias.

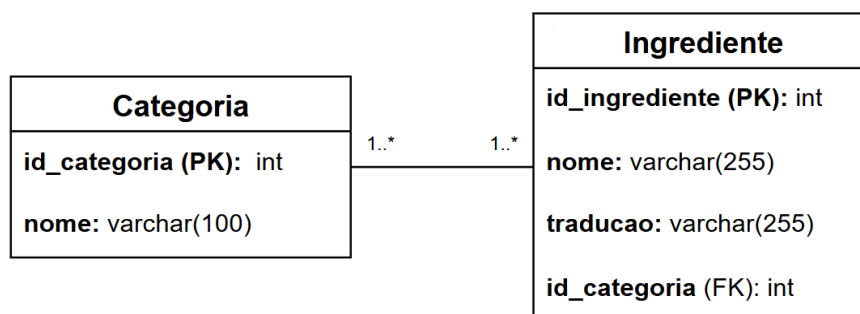


Figura 2. Diagrama relacional do banco de dados

Fonte: Elaboração própria, 2024

4.2.3. Arquitetura do Software

Segundo Sommerville (2018), o projeto de arquitetura tem como objetivo modelar a estrutura geral do *software*. Ele ressalta que este é o estágio inicial do processo de design de sistemas, conectando os requisitos à implementação ao identificar os principais componentes e suas interações. O resultado final é um modelo que descreve a organização do *software* por meio da comunicação entre os componentes.

A arquitetura proposta para o sistema é representada pela Figura 3, que ilustra os principais componentes e a forma como se comunicam. A interface do usuário (IU), desenvolvida em Flutter, é responsável pela interação do usuário com as funcionalidades da aplicação. Essa camada permite o envio de imagens ou textos para análise, a seleção de categorias de alérgenos, a visualização dos resultados e o acesso ao histórico de análises. As informações fornecidas são então enviadas à API, desenvolvida com o *framework* Django, que é responsável por processar e analisar os dados. Quando o dado de entrada é uma imagem, a API aplica o OCR para extrair o texto presente. Em seguida, o sistema verifica se o texto extraído contém algum ingrediente cadastrado no banco de dados que pertence às categorias previamente selecionadas. O banco de dados, criado com PostgreSQL, contém exclusivamente ingredientes alergênicos, organizados por categorias.

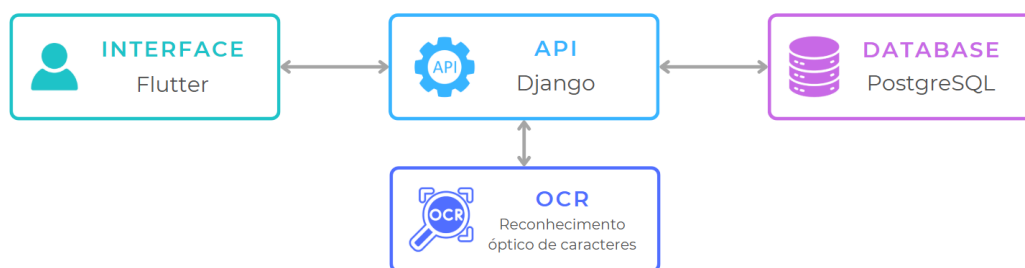


Figura 3. Diagrama de arquitetura de software

Fonte: Elaboração própria, 2024

4.3. Implementação

A implementação do sistema seguiu a modelagem previamente definida e foi organizada em quatro etapas principais. Inicialmente, realizou-se a comparação entre as ferramentas de OCR Pytesseract e Cloud Vision AI, a fim de escolher a mais adequada para o aplicativo proposto. Na segunda etapa, foi implementado o banco de dados contendo os ingredientes alergênicos, devidamente organizados por categoria. Em seguida, desenvolveu-se a API, responsável por intermediar a comunicação entre o banco de dados e a interface. Por fim, foi criada a interface do aplicativo, que integra os recursos descritos. As etapas citadas são detalhadas nos tópicos a seguir.

4.3.1. Avaliação das Ferramentas OCR

A primeira etapa da implementação consistiu em selecionar a ferramenta OCR mais adequada para a extração de textos em rótulos de cosméticos. Para isso, foram analisadas duas opções amplamente utilizadas: Tesseract, por meio da biblioteca Pytesseract, e Cloud Vision AI. A comparação foi feita utilizando um conjunto de 55 imagens reais de rótulos cosméticos, capturadas pela autora através de um smartphone. Essas imagens apresentam variadas características visuais, como cor de fundo, cor da fonte, tipografia, espaçamento e angulação do texto. Algumas dessas imagens podem ser visualizadas no Apêndice A. Ademais, vale ressaltar que as ferramentas foram utilizadas em suas configurações padrão, sem a aplicação de técnicas de pré-processamento ou ajustes manuais.

De acordo com Neudecker et al. (2021), a avaliação da qualidade de sistemas de OCR na comunidade científica baseia-se na manipulação de cadeias de caracteres utilizando algoritmos de distância de edição. Essa abordagem, cujas bases remontam aos estudos de Rice (1996), consiste em calcular o número mínimo de operações para transformar o texto extraído por OCR no texto de referência (*ground truth*).

A acurácia, segundo Rice (1996), é calculada pela seguinte fórmula:

$$\text{Acurácia} = \frac{n - E}{n}$$

Em que:

- n = Número total de caracteres no texto real (*ground truth*).
- E = Número de erros cometidos pelo OCR.

O valor de E é obtido através da distância de Levenshtein entre o texto extraído pelo OCR e o texto real. Esse algoritmo de distância de edição calcula o número mínimo de operações necessárias, sejam inserções, exclusões ou substituições, para transformar uma sequência de caracteres em outra (Rice, 1996).

Nesta análise, a distância de Levenshtein foi calculada utilizando a biblioteca Python de mesmo nome. O cálculo foi realizado entre o texto extraído e normalizado pelas ferramentas de OCR e o texto real obtido a partir de arquivos .txt previamente criados para cada imagem. Esses arquivos contêm o texto correto do rótulo e os alérgenos presentes. A acurácia foi obtida com o seguinte código-fonte em Python, representado pela Figura 4, em que a variável “*distancia*” corresponde ao número de erros cometidos pelo OCR (E) e “*tamanho_real*” ao número total de caracteres do texto real (n).

```

import Levenshtein

def calcular_acuracia_levenshtein(texto_real, texto_ocr_normalizado):
    distancia = Levenshtein.distance(texto_real, texto_ocr_normalizado)
    tamanho_real = len(texto_real)
    acuracia = ((tamanho_real - distancia) / tamanho_real) * 100
    return acuracia

```

Figura 4. Código-fonte em Python para cálculo da acurácia

Fonte: Elaboração própria, 2025

4.3.2. Resultados da Comparação entre as Ferramentas OCR

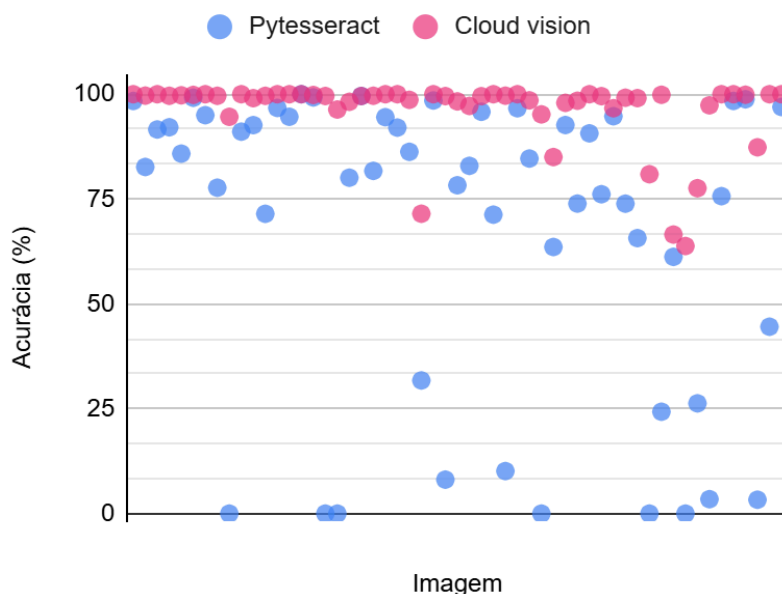
A análise comparativa do desempenho das ferramentas Cloud Vision AI e Pytesseract é apresentada na Tabela 5, que detalha as estatísticas descritivas da acurácia obtida no conjunto de 55 imagens. Os resultados revelam uma diferença significativa no desempenho médio, com o Cloud Vision AI alcançando uma acurácia média de 96,22%, enquanto o Pytesseract obteve 67,72%. A mediana da acurácia também foi superior para o Cloud Vision AI (99,61%) em comparação ao Pytesseract (82,65%), indicando uma tendência geral de melhor desempenho. Ademais, a análise do desvio padrão reforça essa vantagem, com o Cloud Vision AI apresentando maior consistência (8,41), ao passo que o Pytesseract apresentou uma dispersão elevada (35,62).

Tabela 5. Resultados da comparação entre as ferramentas OCR

OCR	Média Aritmética	Mediana	Desvio padrão
Cloud Vision AI	96,22	99,61	8,41
Pytesseract	67,72	82,65	35,62

Fonte: Elaboração própria, 2025

Em complemento, a Figura 5 apresenta um gráfico de dispersão com as acurácias obtidas por ambas as ferramentas no conjunto de imagens analisadas, onde é possível observar a superioridade do Cloud Vision AI ao longo de todo o conjunto, bem como a instabilidade do desempenho do Pytesseract.



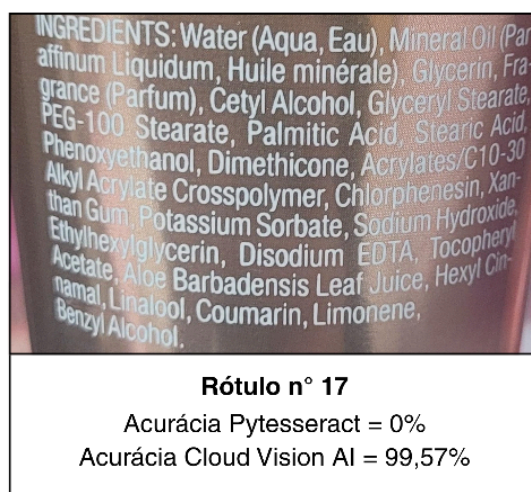


Figura 7. Imagem com maior diferença de acurácia (rótulo n° 17)

Fonte: Elaboração própria, 2025

De modo geral, o Pytesseract apresentou bons resultados em rótulos com fundo claro, fonte escura e texto reto. No entanto, em rótulos com aspectos diferentes a este, sua acurácia decaiu consideravelmente, atingindo 0% de acurácia em seis das 55 imagens. Diante desses resultados, o Cloud Vision AI foi selecionado como a ferramenta OCR mais adequada para o aplicativo proposto, por garantir maior precisão e confiabilidade na extração de textos contidos em rótulos de cosméticos.

4.3.3. Desenvolvimento do Banco de Dados de Ingredientes Alérgenos

Após a seleção da ferramenta OCR, a próxima fase da implementação consistiu na criação de um banco de dados para armazenar ingredientes alérgenos presentes em cosméticos. Conforme a modelagem apresentada na Seção 4.1.3, a estrutura do banco de dados é composta por duas tabelas principais.

A primeira tabela, referente às categorias, lista os tipos de alérgenos que o aplicativo será capaz de identificar. Para este projeto, foram selecionadas seis categorias comuns de alérgenos encontrados em cosméticos: álcool etílico, corante, fragrância, glúten, parabeno e sulfato. A segunda tabela, por sua vez, contém os ingredientes pertencentes a essas categorias, armazenando tanto a sua nomenclatura INCI quanto sua tradução para o português. Com essa estruturação, o sistema pode realizar análises de forma personalizada, considerando apenas os alérgenos de interesse do usuário.

A implementação do banco de dados foi realizada com PostgreSQL e os dados foram obtidos através do Painel de Traduções INCI disponibilizado pela Anvisa. Por meio dessa fonte oficial, foram cadastrados mais de 900 ingredientes alergênicos que se enquadram nas seis categorias definidas.

4.3.4. Desenvolvimento da API

Após a implementação do banco de dados, foi criada a API do sistema utilizando a linguagem Python e o *framework* Django. Além de tratar as requisições feitas pela interface, a API é capaz de acessar o banco de dados por meio de classes que representam categorias e ingredientes, conhecidas como *models*. No Django, os *models* são definidos como classes Python que mapeiam diretamente para tabelas do banco de dados, permitindo que os dados sejam manipulados de forma integrada à lógica da aplicação (Django, 2024).

Para atender os requisitos do projeto, foram implementados dois *endpoints* principais, que são pontos de acesso da aplicação que recebem e respondem às requisições. No contexto deste aplicativo, eles recebem os dados enviados pela interface e, com base no código-fonte Python implementado, acessam o banco de dados, realizam a análise e retornam os resultados.

O primeiro *endpoint* recebe como entrada uma imagem e uma lista com as categorias selecionadas pelo usuário. O processamento da informação ocorre através de três etapas sequenciais. Na primeira etapa, o texto presente na imagem é extraído através do Cloud Vision AI. Em seguida, o texto extraído pelo OCR é normalizado, convertendo todo o conteúdo em letras minúsculas, removendo quebras de linha, espaços em excesso e outros caracteres que possam interferir na análise. Na última etapa, o sistema percorre os ingredientes cadastrados no banco de dados das categorias selecionadas, buscando no texto normalizado correspondências exatas tanto do nome INCI quanto em sua tradução para português. Se um ingrediente for encontrado, ele é adicionado à lista de alérgenos identificados. Ao final, o *endpoint* retorna um objeto JSON (*JavaScript Object Notation*) contendo, para cada categoria, a informação sobre a presença de alérgenos, a lista de ingredientes alergênicos encontrados (quando houver), bem como o texto extraído pelo OCR e o texto normalizado.

O segundo *endpoint* oferece uma funcionalidade similar ao anterior, porém em vez de receber uma imagem, ele recebe diretamente um texto. O processamento segue as mesmas etapas de normalização do texto e verificação no banco de dados, retornando um objeto JSON com informações sobre a presença ou ausência de alérgenos nas categorias selecionadas e, se presentes, quais são os ingredientes identificados, além do texto enviado pelo usuário e o texto normalizado.

4.3.5. Desenvolvimento da Interface do Aplicativo

A última etapa da implementação consistiu no desenvolvimento da interface do aplicativo móvel, criada com o *framework* Flutter. Seu objetivo é oferecer ao usuário uma experiência simples e intuitiva para verificar a presença de alérgenos em cosméticos. O aplicativo é composto por três telas, descritas a seguir.

Tela principal: nesta tela, o usuário pode selecionar as categorias de alérgenos que deseja verificar no produto, sendo que a verificação pode ser feita a partir de uma imagem ou de um texto. No caso da imagem, o aplicativo oferece duas opções: capturar uma nova foto com a câmera ou selecionar uma imagem da galeria do dispositivo. Após a escolha ou captura da imagem, o aplicativo exibe uma ferramenta de recorte (*cropper*), permitindo ao usuário selecionar apenas a área relevante do rótulo, no caso a seção onde são listados os ingredientes. Caso o usuário opte por inserir texto, é possível colar ou digitar manualmente a lista de ingredientes. Essa opção é útil, por exemplo, quando o usuário copia a composição do produto diretamente de uma loja online. A Figura 8 abaixo apresenta capturas da tela principal do aplicativo e a ferramenta de recorte de imagem.

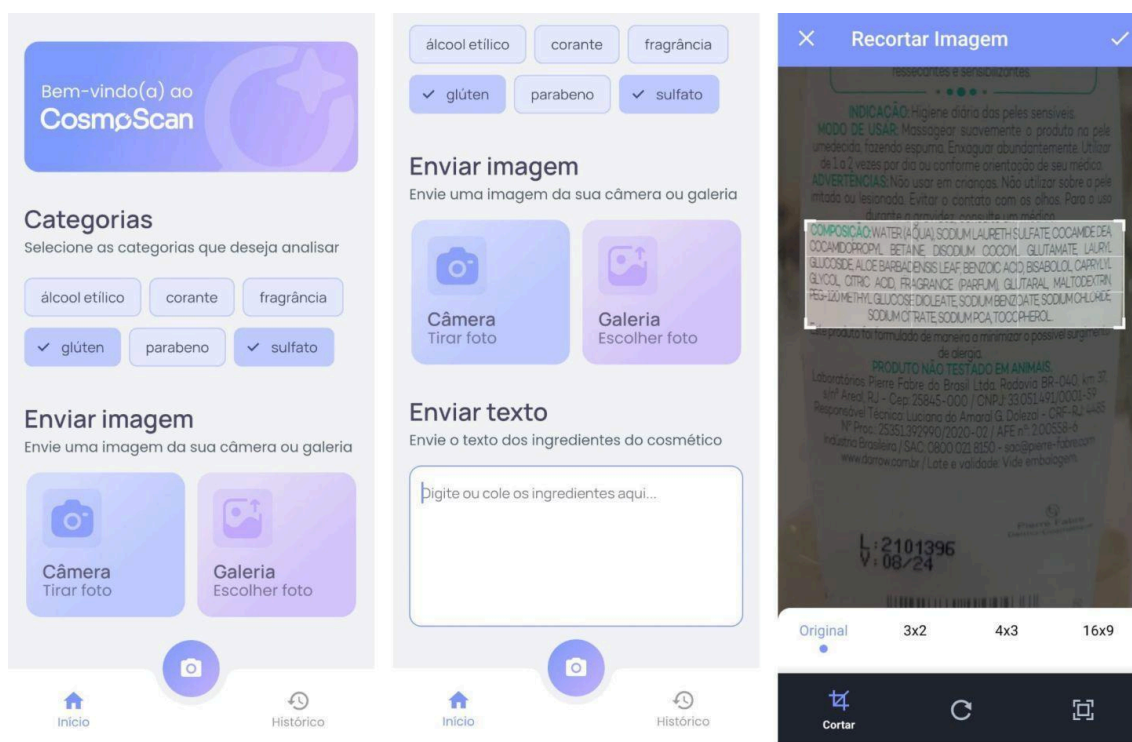


Figura 8. Tela principal do aplicativo e ferramenta de recorte de imagem

Fonte: Elaboração própria, 2025

Tela de resultado: após o envio da imagem ou texto, a API realiza a verificação dos alérgenos e a interface exibe os resultados obtidos. Caso a análise tenha sido feita a partir de uma imagem, a mesma é exibida no topo da tela. Logo abaixo, para cada categoria selecionada, o aplicativo informa o status: “livre”, quando nenhum alérgeno foi identificado, ou “não livre”, quando um ou mais ingredientes associados à categoria foram encontrados. Em casos com alérgenos, são exibidos os nomes dos ingredientes identificados, tanto sua nomenclatura INCI quanto sua tradução para o português,

facilitando o entendimento do usuário. Além disso, o texto enviado ou extraído da imagem também é apresentado nesta tela. Caso ingredientes alérgenos tenham sido identificados, eles são visualmente destacados no texto utilizando um sistema de cores, onde cada cor corresponde a uma categoria de alérgeno específica. Abaixo do texto, há uma legenda que explica a correspondência entre as cores e as categorias. O objetivo desta funcionalidade é fornecer ao usuário o contexto do texto analisado e evidenciar visualmente os alérgenos encontrados, permitindo que ele identifique eventuais falhas. Por fim, o usuário pode optar por salvar a análise e dar um nome a ela ou apenas cancelar, caso não queira armazenar o resultado. A Figura 9 abaixo apresenta capturas dessa tela.

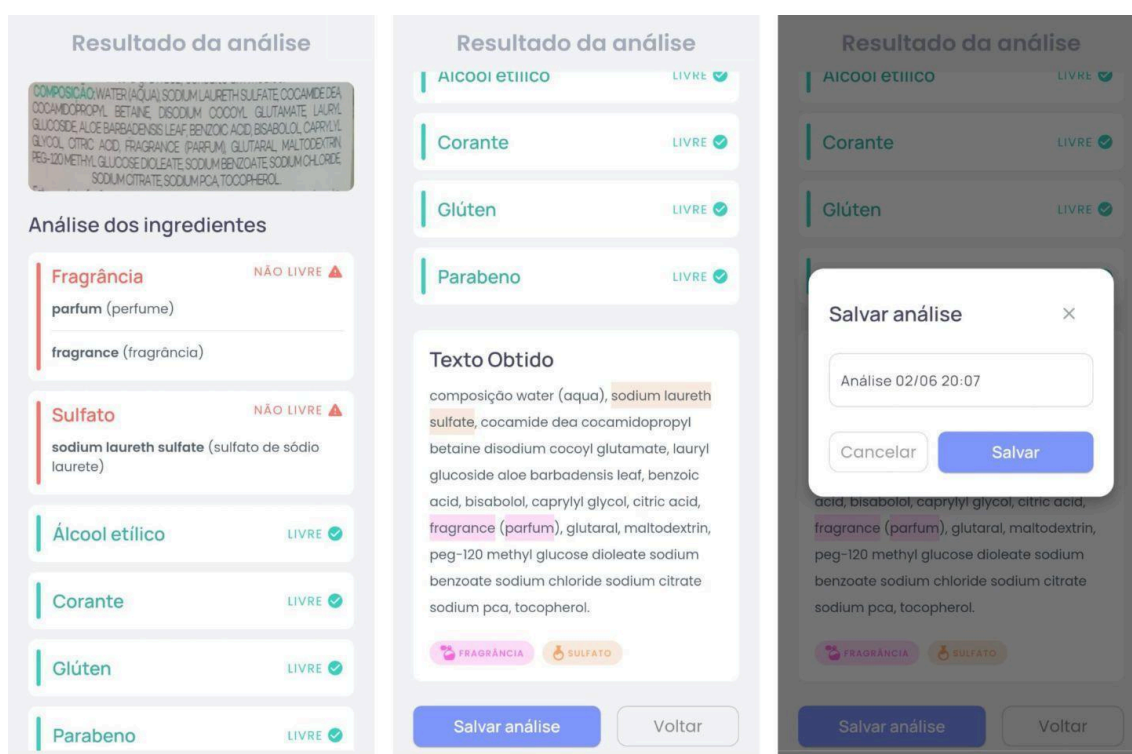


Figura 9. Tela de resultado do aplicativo

Fonte: Elaboração própria, 2025

Tela de histórico: esta tela reúne todas as análises previamente salvas pelo usuário, que são armazenadas localmente no dispositivo. Cada registro exibe uma miniatura da imagem analisada, o nome atribuído à análise, a data e horário de realização e as tags com as categorias de alérgenos identificadas. Ademais, é possível editar o nome da análise, excluí-la, realizar buscas por nome ou ordenar os registros por data, em ordem crescente ou decrescente. Ao clicar em uma análise, o usuário é direcionado a uma tela com os detalhes completos, semelhante à tela de resultado, mas

sem as opções de salvar ou cancelar, já que se trata de uma análise arquivada. Algumas capturas dessa tela são representadas na Figura 10 abaixo.

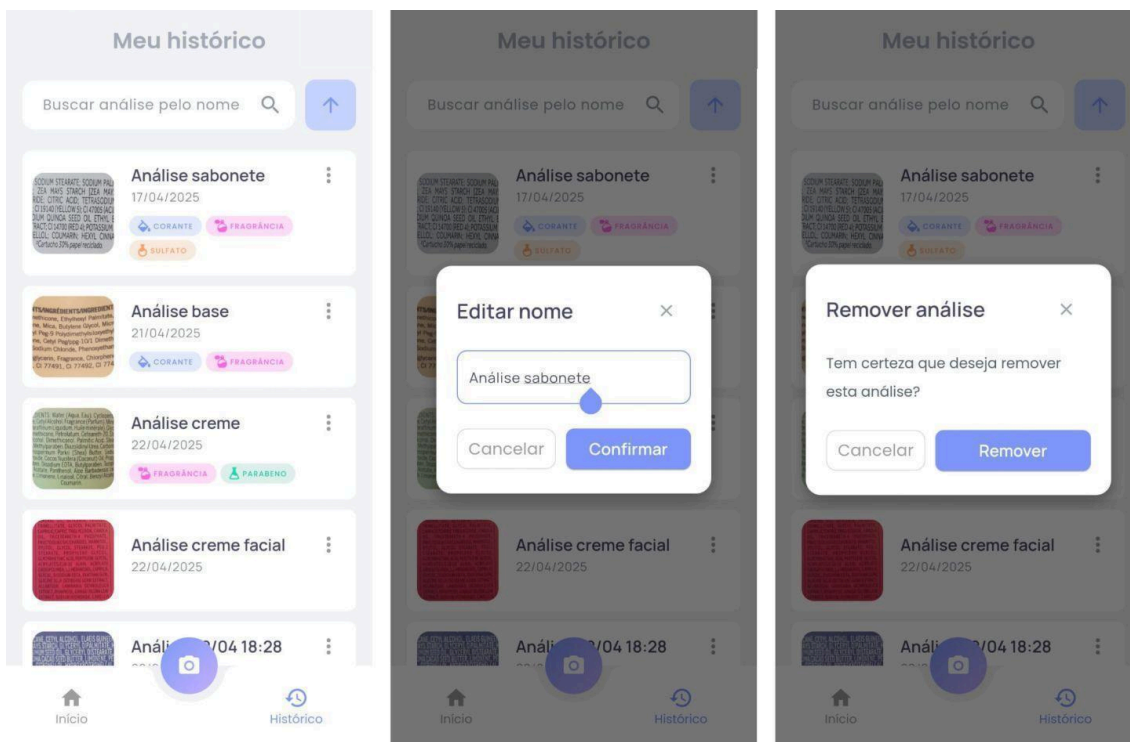


Figura 10. Tela de histórico do aplicativo

Fonte: Elaboração própria, 2025

Além das telas do aplicativo, outro elemento importante da interface é a barra de navegação. Ela permite o acesso direto à tela principal e ao histórico, além de contar com um botão central que abre a câmera para realizar uma nova análise. Para isso, é necessário que ao menos uma categoria esteja selecionada, caso contrário, uma mensagem de erro informa o usuário sobre essa obrigatoriedade. Com o mesmo intuito de facilitar a navegação, as categorias selecionadas permanecem ativas mesmo com a troca de telas, sendo reiniciadas apenas se o usuário sair do aplicativo ou desmarcá-las manualmente.

5. Resultados

Nesta seção, são apresentados os resultados obtidos com a aplicação do sistema proposto. Para isso, utilizou-se o mesmo conjunto de imagens adotado na avaliação das ferramentas de OCR, apresentado na Seção 4.2.1. Os tópicos a seguir descrevem os resultados relacionados ao desempenho do reconhecimento óptico de caracteres, da

detecção de alérgenos, do código-fonte Python desenvolvido para essa tarefa e do tempo de execução do sistema.

5.1. Resultados do Reconhecimento Óptico de Caracteres

Conforme apresentado na Seção 4.2.2, foram avaliadas duas ferramentas de OCR utilizando a acurácia como métrica de comparação. O Cloud Vision AI foi escolhido para a implementação do aplicativo devido ao seu desempenho superior em relação ao Pytesseract, apresentando acurácia média de 96,22%, mediana de 99,61% e desvio padrão de 8,41%. Esta seção aprofunda os resultados obtidos com essa ferramenta, apresentando dados mais detalhados sobre sua acurácia e desempenho na aplicação desenvolvida.

De modo geral, o Cloud Vision AI apresentou resultados satisfatórios, especialmente considerando a mediana elevada (99,61%), que indica que a maior parte dos rótulos processados atingiu uma acurácia próxima do ideal. Ademais, a ferramenta foi capaz de extrair corretamente os textos mesmo diante de desafios comuns no reconhecimento de caracteres em embalagens de produtos - desafios estes apontados por Rimdusit et al. (2024), Čakić et al. (2020) e Severo (2024), como textos curvos, diferentes tipos e tamanhos de fonte, reflexos e sombras.

Para uma análise mais detalhada do desempenho dessa ferramenta, a Tabela 6 apresenta a distribuição completa das acurácias obtidas, onde é possível observar que em 87,27% dos casos (56,36% + 30,91%) a acurácia foi igual ou superior a 90%.

Tabela 6. Distribuição da acurácia do Cloud Vision AI

Faixa de acurácia	Número de casos	Percentual (%)
0% - 50%	0	0,00
50% - 70%	2	3,64
70% - 80%	2	3,64
80% - 90%	3	5,45
90% - 100%	31	56,36
100%	17	30,91
Total	55	100

Fonte: Elaboração própria, 2025

Apesar do desempenho positivo, algumas dificuldades específicas foram observadas. A principal delas ocorreu em rótulos com justificação forçada, responsável por gerar espaços excessivamente grandes entre as palavras dos ingredientes. Esse tipo

de formatação de texto levou o Cloud Vision AI a interpretar palavras isoladas como blocos independentes, resultando na fragmentação de nomes compostos e na desorganização da sequência original dos ingredientes. Tal distorção reduziu significativamente a correspondência entre o texto extraído e o texto real do rótulo. Um exemplo dessa ocorrência pode ser observado no rótulo nº46 (Figura 11), cuja justificação forçada impactou negativamente a qualidade do reconhecimento, resultando em uma das menores acurácias registradas pela ferramenta: 66,53%. Em complemento, a Tabela 7 apresenta o texto real e o texto extraído pelo Cloud Vision AI nesse caso.

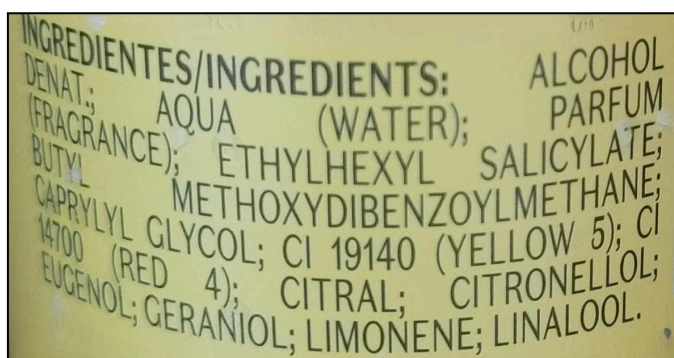


Figura 11. Imagem do rótulo nº46

Fonte: Elaboração própria, 2025

Tabela 7. Texto real e texto obtido pelo Cloud Vision AI no rótulo nº 46

Texto real	ingredientes/ingredients: alcohol denat.; aqua (water); parfum (fragrance); ethylhexyl salicylate; butyl methoxydibenzoylmethane; caprylyl glycol; ci 19140 (yellow 5); ci 14700 (red 4); citral; citronellol; eugenol; geraniol; limonene; linalool.
Texto OCR	ingredientes/ingredients: denat butyl alcohol fragrance); ethylhexyl salicylate; (water); parfum caprylyl glycol; ci 19140 (yellow 5); ci 14700 (red 4); citral; citronellol; methoxydibenzoylmethane; eugenol; geraniol; limonene; linalool.
Acurácia	66,53%

Fonte: Elaboração própria, 2025

Outra limitação observada foi a dificuldade em diferenciar caracteres visualmente semelhantes, como “i”, “I”, “l” e “L”. Erros recorrentes também ocorreram na identificação de sinais de pontuação que separam ingredientes, como vírgulas (,), pontos (.), dois-pontos (:) e ponto e vírgula (;).

Para identificar os caracteres mais propensos a erros de reconhecimento, foi realizada uma avaliação quantitativa dos níveis de confiança retornados pelo Cloud Vision AI para cada caractere extraído nas 55 imagens. Ao todo, foram processados

22.555 caracteres, dos quais apenas 1,94% apresentaram nível de confiança inferior a 70%. A Tabela 8 apresenta os dez caracteres com maior número de ocorrências nessa faixa de baixa confiança.

Tabela 8. Os primeiros dez caracteres com confiança abaixo de 70%

Caractere	Nome	Nº de ocorrências
;	Ponto e vírgula	104
,	Vírgula	61
I	I maiúsculo	61
:	Dois pontos	35
.	Ponto	19
l	L minúsculo	12
L	L maiúsculo	11
C	C maiúsculo	8
-	Hífen	8
)	Barra direita	7

Fonte: Elaboração própria, 2025

Contudo, os principais desafios enfrentados pelo Cloud Vision AI concentram-se no reconhecimento de sinais de pontuação e de caracteres com formas visuais semelhantes. Apesar da baixa proporção de caracteres com confiança inferior a 70%, em algumas das imagens analisadas, esses erros comprometeram a correta identificação dos ingredientes e, conseqüentemente, a precisão na detecção de alérgenos. Por outro lado, a mesma avaliação quantitativa indicou que 87,48% dos caracteres extraídos apresentaram confiança igual ou superior a 95%, o que evidencia a confiabilidade da ferramenta na maior parte dos casos.

5.2. Resultados da Detecção de Alérgenos

A fim de avaliar a capacidade do sistema em identificar corretamente os alérgenos presentes nos rótulos, foi adotada a sensibilidade como métrica de avaliação. Conforme Vilela Junior et al. (2022), a sensibilidade avalia a porcentagem de todos os verdadeiros positivos que foram corretamente previstos em um sistema, sendo expressa pela seguinte fórmula:

$$Sensibilidade = \frac{VP}{VP + FN}$$

Em que:

- *VP* (Verdadeiro positivo) = quando o sistema identifica corretamente um caso como positivo.
- *FN* (Falso negativo) = quando o sistema falha em detectar um caso positivo, classificando-o erroneamente como negativo.

Ainda de acordo com Vilela Junior et al. (2022), a sensibilidade é particularmente relevante em cenários onde a ocorrência de falsos negativos é mais preocupante do que a de falsos positivos. No caso da detecção de ingredientes alergênicos, um falso negativo, isto é, quando o sistema não identifica um alérgeno que de fato está presente no rótulo, pode representar riscos à saúde dos usuários. Já um falso positivo tende a gerar apenas uma precaução adicional, ao sinalizar um alérgeno que, na realidade, não consta na composição. Por esse motivo, a sensibilidade foi adotada como métrica de avaliação, com o objetivo de garantir que o sistema reconheça corretamente o maior número possível de alérgenos presentes nos cosméticos analisados.

Para o cálculo da sensibilidade neste trabalho, foi utilizado o código-fonte em Python apresentado na Figura 12, onde os verdadeiros positivos (*VP*) representam os alérgenos que realmente estavam presentes no rótulo do produto e foram corretamente identificados pelo sistema. Já os falsos negativos (*FN*) referem-se aos alérgenos que existiam no rótulo real, mas que o sistema falhou em detectar.

```
def calcular_sensibilidade(alergenos_reais, alergenos_detectados):  
    if not alergenos_reais:  
        return None  
  
    set_reais = set(alergenos_reais)  
    set_detectados = set(alergenos_detectados)  
  
    VP = len(set_reais.intersection(set_detectados))  
    FN = len(set_reais.difference(set_detectados))  
  
    sensibilidade = (VP / (VP + FN)) * 100  
  
    return sensibilidade
```

Figura 12. Código-fonte em Python para cálculo da sensibilidade

Fonte: Elaboração própria, 2025

Esta análise utilizou o mesmo conjunto de imagens descrito anteriormente na Seção 4.2.1, referente à avaliação das ferramentas OCR. Vale ressaltar que três imagens não continham alérgenos em suas composições e, portanto, não possuíam verdadeiros

positivos, sendo desconsideradas para o cálculo da sensibilidade. Com isso, os resultados foram obtidos a partir de 52 imagens.

O desempenho do aplicativo na identificação de alérgenos demonstrou bons resultados, com uma sensibilidade média de 92,93%. Tanto a mediana quanto a moda da sensibilidade atingiram 100% indicando que, na maioria dos casos, o sistema conseguiu identificar todos os alérgenos presentes. Por outro lado, a sensibilidade mínima registrada foi de 50%, sinalizando algumas detecções incompletas.

A Tabela 9 apresenta a distribuição dos valores de sensibilidade, revelando que em 57,69% das imagens o sistema obteve desempenho máximo. Ao considerar as faixas de 80% a 100%, percebe-se que em 92,3% dos casos (21,15% + 13,46% + 57,69%) o sistema teve sensibilidade igual ou superior a 80%. Adicionalmente, a Figura 13 apresenta essa distribuição de forma gráfica.

Tabela 9. Distribuição da sensibilidade

Faixa de sensibilidade	Número de casos	Percentual (%)
0% - 50%	0	0,00
50% - 70%	2	3,85
70% - 80%	2	3,85
80% - 90%	11	21,15
90% - 100%	7	13,46
100%	30	57,69
Total	52	100

Fonte: Elaboração própria, 2025

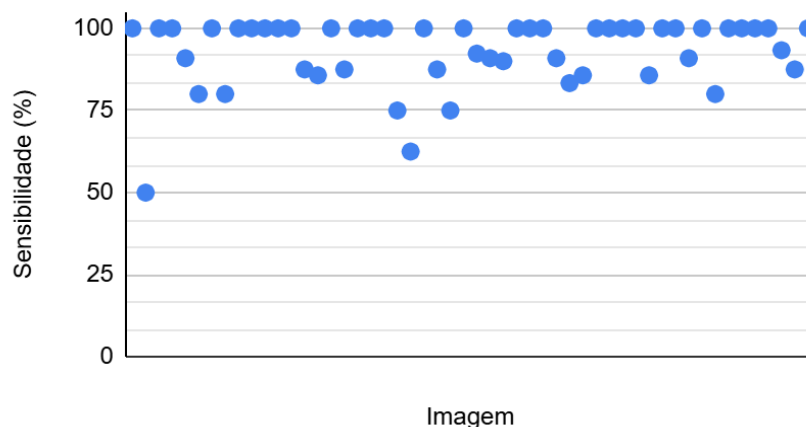


Figura 13. Gráfico de dispersão da sensibilidade

Fonte: Elaboração própria, 2025

5.3. Resultados da Codificação Python para Detecção de Alérgenos

O código-fonte em Python desenvolvido para a detecção de alérgenos utiliza uma lógica simples, baseada na busca por correspondências exatas de palavras. Ele recebe como entrada o texto normalizado e a categoria a ser analisada, em seguida acessa os ingredientes cadastrados no banco de dados para essa categoria e verifica se algum deles está presente no texto. Essa codificação foi utilizada em ambos *endpoints* da API, apresentados na Seção 4.2.4, tanto no que recebe uma imagem e utiliza OCR, quanto no que recebe diretamente um texto.

Conforme os resultados apresentados na Seção 5.2, o sistema obteve desempenho satisfatório na identificação de alérgenos nas 52 imagens analisadas, alcançando uma sensibilidade média de 92,93%. No entanto, por adotar uma abordagem literal, a codificação depende da integridade do conteúdo analisado, sendo sensível a qualquer variação, omissão ou erro no reconhecimento dos caracteres. Os poucos falsos negativos ocorreram, em sua maioria, na categoria de corantes. De acordo com a INCI, esses ingredientes são identificados pela sigla “CI” (Colour Index) seguida de números, por exemplo “CI 28440”. Devido a erros do Cloud Vision AI no reconhecimento correto de “CI”, ocorreram omissões ou variações dos caracteres desse prefixo que não correspondiam aos registros no banco de dados e, conseqüentemente, impediam a detecção desses corantes. Tais erros podem estar relacionados à baixa confiabilidade dos caracteres “C”, “I” e “L”, conforme evidenciado na Tabela 8 da Seção 5.1.

Embora o objetivo do trabalho resida em avaliar o uso de OCR na identificação de alérgenos em rótulos, foi feita, de forma complementar, uma análise do código-fonte Python para detecção de alérgenos a fim de avaliar seu desempenho isoladamente. Para isso, utilizou-se o *endpoint* de busca de alérgenos em texto para analisar os arquivos .txt correspondentes às mesmas 52 imagens utilizadas na Seção 5.2. Como esses arquivos continham o texto real dos rótulos e os ingredientes alergênicos presentes, o teste permitiu avaliar o código-fonte sem interferência do OCR. Os resultados dessa análise demonstraram que todos os alérgenos reais foram corretamente detectados, alcançando uma sensibilidade média de 100%. Dessa forma, é possível confirmar que os falsos negativos observados na Seção 5.2 decorreram de falhas na extração textual pelo Cloud Vision AI e não da lógica de detecção.

Além da análise de sensibilidade, também foram examinados os falsos positivos com o objetivo de identificar possíveis limitações dessa codificação. Um dos casos mais recorrentes envolveu o ingrediente álcool etílico, que segundo a nomenclatura INCI pode ser nomeado como “alcohol”, “alcohol denat” ou variações como “alcohol denat sd alcohol [número]” e “alcohol denat sd alcohol [número]-[letra]”. Como o código-fonte realiza a busca por termos exatos, sem considerar o contexto em que aparecem, qualquer ocorrência isolada de “alcohol” no texto era interpretada como

álcool etílico. Com isso, outros tipos de álcool, como “cetyl alcohol” (álcool cetílico), por exemplo, eram incorretamente identificados como alérgenos dessa categoria.

Outro ponto observado relacionado aos falsos positivos, envolveu ingredientes nomeados por múltiplas palavras. A codificação lidou adequadamente com alérgenos que contêm outros como *substrings*, isto é, quando o nome de um ingrediente está “incluso” em outro, por exemplo, na presença de “methylparaben” não houve identificação incorreta de “ethylparaben”. Por outro lado, essa lógica não funcionou da mesma forma para ingredientes nomeados por duas ou mais palavras. Um caso observado foi “amyl cinnamal”, onde o código-fonte identificou corretamente esse ingrediente, mas também detectou “cinnamal” que é um alérgeno distinto. Com isso, em determinados contextos, houve a identificação de alérgenos adicionais quando uma parte do nome composto coincidia com outro ingrediente do banco de dados.

5.4. Resultados do Tempo de Execução do Sistema

Com o objetivo de validar o requisito referente ao processamento das análises em tempo razoável, foram realizados testes de desempenho no aplicativo para estimar o tempo que o usuário levaria para obter os resultados durante o uso real da aplicação.

Os testes foram executados diretamente na interface, com a seleção das seis categorias disponíveis, e o tempo registrado corresponde ao intervalo entre a confirmação do recorte da imagem e a exibição dos resultados na tela. Entre as 55 imagens testadas, o tempo médio de execução foi de 2,01 segundos, com mínimo de 1,06 segundos e máximo de 3,55 segundos. Esses dados indicam que o sistema atende ao requisito, oferecendo respostas em poucos segundos. Vale ressaltar que as imagens utilizadas continham apenas a área do rótulo referente à composição do cosmético, o que contribuiu para maior agilidade. Por isso, o sistema disponibiliza uma ferramenta de recorte, que permite ao usuário selecionar apenas a região de interesse antes da análise. Naturalmente, imagens com maior volume de texto podem demandar um tempo ligeiramente superior.

6. Considerações Finais

Através deste trabalho, foi possível desenvolver um aplicativo *mobile* que faz uso da tecnologia de Reconhecimento Óptico de Caracteres (OCR) para analisar imagens de rótulos cosméticos e identificar a presença de ingredientes alergênicos. Ao longo do estudo, os objetivos específicos foram integralmente alcançados, demonstrando a viabilidade e o potencial da solução proposta.

Os testes comparativos entre as ferramentas Pytesseract e Cloud Vision AI evidenciaram a superioridade da segunda em termos de acurácia frente aos desafios impostos pelas características gráficas dos rótulos, como textos curvos, reflexos e variações tipográficas. O Cloud Vision AI apresentou um ótimo desempenho, com acurácia média de 96,22%, mediana de 99,61% e desvio padrão de 8,41%, motivo pelo qual foi utilizado no desenvolvimento do aplicativo. Apesar dos bons resultados, a ferramenta enfrentou dificuldades específicas no reconhecimento de caracteres, especialmente relacionadas à formatação do texto presente nos rótulos e à diferenciação de caracteres semelhantes.

A etapa de detecção de alérgenos também apresentou resultados bastante satisfatórios, evidenciando que a combinação entre uma ferramenta de OCR e uma lógica de busca exata é capaz de entregar um desempenho confiável ao usuário. Com sensibilidade média de 92,93% e mediana e moda de 100%, o aplicativo atingiu sensibilidade perfeita em 57,69% dos casos e sensibilidade igual ou superior a 80% em 92,3% dos testes. Entretanto, limitações foram identificadas, sobretudo devido à abordagem literal adotada, que exige correspondência exata dos ingredientes e não considera variações textuais causadas por erros de OCR. Tais limitações resultaram em falsos positivos e falhas na identificação de alguns alérgenos, principalmente corantes.

Em suma, o desenvolvimento do trabalho proposto apresentou resultados satisfatórios, comprovando a hipótese de que a tecnologia OCR pode ser aplicada de forma eficaz para aumentar a segurança na identificação de alérgenos em cosméticos. Embora o sistema apresente limitações que requerem aprimoramentos futuros, os resultados indicam um potencial significativo para aplicações práticas que beneficiem consumidores com alergias ou sensibilidades a ingredientes cosméticos. Além disso, a base de dados, por ter sido criada manualmente pela autora, carece de uma gama mais ampla de alérgenos, restringindo-se a seis categorias principais. Essa limitação, embora suficiente para a experimentação realizada neste trabalho de conclusão de curso, limita a aplicabilidade do aplicativo em cenários mais amplos.

Para trabalhos futuros, sugere-se a expansão da base de dados de alérgenos para incluir uma gama maior de ingredientes, tornando a utilização do aplicativo mais abrangente. A incorporação de técnicas de aprendizado de máquina através do treinamento em grandes bases de dados de imagens de rótulos cosméticos poderia aprimorar tanto o desempenho do OCR quanto a detecção de alérgenos. Adicionalmente, como uma solução mais imediata, a utilização de algoritmos de correspondência aproximada, como *fuzzy matching*, permitiria maior tolerância a erros do OCR, aumentando a confiabilidade da identificação de alérgenos mesmo diante de imprecisões na extração de texto. Por fim, destaca-se que a aplicação foi desenvolvida e testada apenas em ambiente local, sendo necessária sua adaptação para uma arquitetura de produção a fim de ser disponibilizada efetivamente aos usuários finais.

7. Referências

ABIHPEC. **Panorama do Setor.** 2025. Disponível em: https://abihpec.org.br/site2019/wp-content/uploads/2025/04/Panorama_do_-Setor-Atualizacao_16.05.25_Port.pdf. Acesso em: abr. 2025.

ANVISA. **Nota técnica INCI.** 2020. Disponível em: <https://www.gov.br/anvisa/pt-br/acessoainformacao/perguntasfrequentescosmeticos/nota-tecnica-inci>. Acesso em: set. 2024.

ANVISA. **Tradução INCI.** 2023. Disponível em: <https://www.gov.br/anvisa/pt-br/assuntos/cosmeticos/traducao>. Acesso em: jan. 2025.

ASBAI. **Cosméticos estão entre as maiores causas de alergia.** 2017. Disponível em: <https://asbai.org.br/cosmeticos-estao-entre-as-maiores-causas-de-alergia/>. Acesso em: nov. 2024.

BRASIL. Agência Nacional de Vigilância Sanitária. **Resolução da Diretoria Colegiada - RDC nº 898, de 28 de agosto de 2024.** Dispõe sobre a obrigatoriedade da tradução da composição dos ingredientes de produtos de higiene pessoal, cosméticos e perfumes para o português. Diário Oficial da União, Brasília, DF. 2024. Disponível em: <https://anvisa.gov.br/legis/ato-publico/ato-publico-00000898>. Acesso em: jan. 2025.

BRASIL. Agência Nacional de Vigilância Sanitária. **Resolução da Diretoria Colegiada - RDC nº 907, de 19 de setembro de 2024.** Dispõe sobre a definição, a classificação, os requisitos técnicos para rotulagem e embalagem, os parâmetros para controle microbiológico, bem como os requisitos técnicos e procedimentos para a regularização de produtos de higiene pessoal, cosméticos e perfumes. Diário Oficial da União, Brasília, DF. 2024. Disponível em: <https://anvisa.gov.br/legis/ato-publico/ato-publico-00000907>. Acesso em: jan. 2025.

ČAKIĆ, Stevan; POPOVIĆ, Tomo; LANDI, Stevan; KRČO, Srđan; GAZIVODA, Anita. The Use of Tesseract OCR Number Recognition for Food Tracking and Tracing. *In: INTERNATIONAL CONFERENCE ON INFORMATION TECHNOLOGY (IT)*, 24., 2020, Zabljak, Montenegro. **Anais eletrônicos [...]** . [S.L.]: IEEE, 2020. p. 1-4. Disponível em: <https://ieeexplore.ieee.org/document/9070558>. Acesso em: set. 2024.

CANHEDO, Leticia Alves Barbosa; REIS, Yara Prado Barolli. **Alergias Causadas por Cosmético**. 2016. Disponível em: <http://repositorio.unis.edu.br/handle/prefix/458>. Acesso em: jan. 2025.

DE AZAMBUJA, Rogério Xavier; MORAIS, A. Jorge; FILIPE, Vítor. X-Wines: A Wine Dataset for Recommender Systems and Machine Learning. **Big Data and Cognitive Computing**, v. 7, n. 1, p. 20, 22 jan. 2023. Disponível em: <https://doi.org/10.3390/bdcc7010020>. Acesso em: jan. 2025.

DJANGO. Disponível em: <https://www.djangoproject.com/>. Acesso em: nov. 2024.

DRAW.IO. Disponível em: <https://www.drawio.com/>. Acesso em: nov. 2024.

FLUTTER. Disponível em: <https://flutter.dev/>. Acesso em: nov. 2024.

GONZALEZ, Rafael; WOODS Richard. **Digital Image Processing**. 4. ed. 2017. Pearson International.

GONZÁLEZ-MUÑOZ, P.; CONDE-SALAZAR, L.; VAÑÓ-GALVÁN, S. Allergic Contact Dermatitis Caused by Cosmetic Products. **Actas Dermo-Sifiliográficas**, v. 105, n. 9, p. 822–832, nov. 2014. Disponível em: <https://doi.org/10.1016/j.adengl.2014.09.007>. Acesso em: jan. 2025.

GOOGLE CLOUD. Cloud Vision AI. Disponível em: <https://cloud.google.com/vision>. Acesso em: nov. 2024.

HEGGHAMMER, Thomas. OCR with Tesseract, Amazon Textract, and Google Document AI: a benchmarking experiment. **Journal of Computational Social Science**, v. 5, n. 1, p. 861–882, 2022. Disponível em: <https://doi.org/10.1007/s42001-021-00149-1>. Acesso em: set. 2024.

HEUSER, Carlos Alberto. **Projeto de banco de dados**. 2. ed. Porto Alegre, RS: Instituto de Informática da UFRGS, Sagra Luzzato, 1999.

IMPEDOVO, Sebastiano; OTTAVIANO, Lorenzo; OCCHINEGRO, Silvia. Optical Character Recognition - A Survey. **International Journal of Pattern Recognition and Artificial Intelligence**, Bari, Italy, 1991, p. 1-24. Disponível em: <https://www.worldscientific.com/doi/abs/10.1142/S0218001491000041>. Acesso em: set. 2024.

KUMAR, Naveen H. N.; KIRAN; KOTIYAL, Arnav; B, Jagadeesh; MAHADEVASWAMY; S, Guru Prasad M. Real-Time Automated Detection of Errors

in the Product Labels Using Image Processing and OCR. *In: INTERNATIONAL CONFERENCE ON AMBIENT INTELLIGENCE, KNOWLEDGE INFORMATICS AND INDUSTRIAL ELECTRONICS (AIKIIE)*. 2023, Ballari, India. **Anais eletrônicos [...]** . [S.L.]: IEEE, 2023. p. 1-6. Disponível em: <https://ieeexplore.ieee.org/document/10390320/>. Acesso em: set. 2024

KUMARI, Sweta; ATEM, Tambe Daniel; CHAUDHARY, Vaibhav; SAHU, Sanjeev Kumar; PAL, Biplab. Prevalence and risk factors of cosmetic-induced adverse events: A systematic review and meta-analysis. **Journal Of Applied Pharmaceutical Science**. India, v. 14, n. 11, p. 252-263. nov. 2024. Disponível em: <https://dx.doi.org/10.7324/JAPS.2024.193512>. Acesso em: jan. 2025.

MITHE, Ravina; INDALKAR, Supriya; DIVEKAR, Nilam. Optical character recognition. **International journal of recent technology and engineering (IJRTE)**, [s. l.] v. 2, n. 1, p. 72-75, 2013. Disponível em: <https://www.ijrte.org/wp-content/uploads/papers/v2i1/A0504032113.pdf>. Acesso em: set. 2024.

NEUDECKER, Clemens; BAIERER, Konstantin; GERBER, Mike; CLAUSSNER, Christian; ANTONACOPOULOS, Apostolos; PLETSCHACHER, Stefan. A survey of OCR evaluation tools and metrics. *In: INTERNATIONAL WORKSHOP ON HISTORICAL DOCUMENT IMAGING AND PROCESSING*, 6., 2021, Lausanne. **Anais eletrônicos [...]** New York: ACM, 2021. p. 13–18. Disponível em: <https://doi.org/10.1145/3476887.3476888>. Acesso em: abr. 2025.

PEREIRA, Suelen Marinho; SANTOS, Viviane Marinho dos. REAÇÕES ALÉRGICAS CAUSADAS POR COSMÉTICOS E DERMOCOSMÉTICOS: uma revisão de literatura. **Recima21 - Revista Científica Multidisciplinar - Issn 2675-6218**, [S.L.], v. 5, n. 11, p. 1-2, 18 nov. 2024. Editora RECIMA21 LTDA. Disponível em: <https://recima21.com.br/index.php/recima21/article/view/5952>. Acesso em: maio 2025.

PERSONAL CARE PRODUCTS COUNCIL. **INCI**. Disponível em: <https://www.personalcarecouncil.org/resources/inci/>. Acesso em: jan. 2025.

POSTGRESQL. Disponível em: <https://www.postgresql.org/>. Acesso em: nov. 2024.

PYTESSERACT. Disponível em: <https://pypi.org/project/pytesseract/>. Acesso em: nov. 2024.

PYTHON. Disponível em: <https://www.python.org/>. Acesso em: nov. 2024.

RICE, S. V. **Measuring the accuracy of page-reading systems**. 1996. 81 f. Tese (Doutorado em Ciência da Computação) - University of Nevada, Las Vegas, 1996. Disponível em: <http://dx.doi.org/10.25669/hfa8-0cqv>. Acesso em: abr. 2025.

RIMDUSIT, Peeraporn; APHIROMYARNONT, Phakhatorn; PHIMOLTAIRES, Suphakant; PANTHUWADEETHORN, Sasipa. Extracting Information from Drug Label Using Image and Text Processing. *In: INTERNATIONAL JOINT CONFERENCE ON COMPUTER SCIENCE AND SOFTWARE ENGINEERING (JCSSE)*, 21., 2024, Phuket, Thailand. **Anais eletrônicos [...]** [S.L.]: IEEE, 2024. p. 377-382, 2024. Disponível em: <https://ieeexplore.ieee.org/document/10613666/>. Acesso em: set. 2024

SAMANTA, Saptadip. Impacts of Chemical Beauty. **Indian Journal of Physiology and Allied Sciences**, v. 73, n. 1, 2022. Disponível em: <https://doi.org/10.55184/ijpas.v73i1.24>. Acesso em: maio 2025.

SEVERO, Deise Dall'Agnol. **Belezati: desenvolvimento de um módulo de escaneamento e descrição de ingredientes cosméticos baseado na web utilizando tecnologia OCR via API**. 2024. 60 f. Dissertação (Mestrado em Engenharia de Software) – Universidade Federal do Pampa, São Gabriel, 2024. Disponível em: <https://repositorio.unipampa.edu.br/jspui/handle/rii/9022>. Acesso em: set. 2024.

SOMMERVILLE, Ian. **Engenharia de software**. 10. ed. São Paulo, SP: Pearson, 2018

TEIXEIRA, Ana Paula Coelho Penna; ALMEIDA, Andreia Carla de; MELO, Danilo Feitoza; LEITÃO, Leonardo Oliveira; SILVA, Luis Henrique Calazans. Análise descritiva das notificações de eventos adversos de produtos cosméticos registradas no Notivisa, no período de 2006 a 2018. **Vigilância Sanitária em Debate**, Rio de Janeiro, v. 7, n. 4, p. 17–25, nov. 2019. Disponível em: <https://visaemdebate.incqs.fiocruz.br/index.php/visaemdebate/article/view/1384>. Acesso em: jan. 2025.

VILELA JUNIOR, Guanis de Barros; LIMA, Bráulio Nascimento; PEREIRA, Adriano Almeida; RODRIGUES, Marcelo Francisco; OLIVEIRA, José Ricardo Lourenço de; SILIO, Luís Felipe; CARVALHO, Anderson dos Santos; FERREIRA, Heros Ribeiro; PASSOS, Ricardo Pablo. MÉTRICAS UTILIZADAS PARA AVALIAR A EFICIÊNCIA DE CLASSIFICADORES EM ALGORITMOS INTELIGENTES. **Revista CPAQV - Centro de Pesquisas Avançadas em Qualidade de Vida**, [S. l.], v. 14, n. 2, 2022. Disponível em: <https://revista.cpaqv.org/index.php/CPAQV/article/view/933..> Acesso em: mai. 2025.

**APÊNDICE A - EXEMPLOS DE IMAGENS DE RÓTULOS COSMÉTICOS
UTILIZADAS NOS TESTES**

Ingredientes: Aqua, Homosalate, Octocrylene, C12-15 Alkyl Benzoate, Butyl Methoxydibenzoylmethane, Ethylhexyl Salicylate, Triaccontanyl PVP, Potassium Cetyl Phosphate, Cyclopentasiloxane, Glycerin, Triethanolamine, Diethylhexyl 2,6-Naphthalate, Dimethicone, Glyceryl Stearate, Acrylic Acid/VP Crosspolymer, Titanium Dioxide, Phenethyl Benzoate, Phenoxyethanol, Caprylyl Glycol, Isocetyl Stearoyl Stearate, Disodium Laurinodipropionate Tocopheryl Phosphates, Carnosine, Parfum, Laureth-23, Daucus Carota Sativa Seed Oil, BHT, Disodium EDTA, Hexyl Cinnamal, Linalool, d-Limonene.

COMPOSIÇÃO: aqua, dihydroxyacetone, cetearyl alcohol, ocyldodecanol, paraffinum liquidum, propylene glycol, cyclopentasiloxane, sorbitol, dimethicone, erythrose, cetearith-20, glyceryl stearate, polysorbate 60, caramel, phenoxyethanol, parfum, sodium citrate, sodium metabisulfite, hydroxyethyl acrylate/sodium acryloyldimethyl taurate copolymer, isohexadecane, citric acid, BHT, ethylhexylglycerin, PEG-4, sorbitan isostearate, alpha-isomethyl ionone, amyl cinnamal, citronellol, coumarin, d-limonene, eugenol, linalool.

INGREDIENTS: WATER (AQUA, EAU), CETEARYL ALCOHOL, DIMETHICONE, AMODIMETHICONE, GLYCERIN, ETHYLHEXYL PALMITATE, DISTEARYLDIMONIUM CHLORIDE, POLYSORBATE 60, TRIDECETH-12, PARFUM, BENZYL ALCOHOL, PHENOXYETHANOL, SODIUM BENZOATE, CETRIMONIUM CHLORIDE, TETRASODIUM EDTA, CITRIC ACID, ISOPROPYL ALCOHOL, ARGANIA SPINOSA KERNEL OIL, PANTHENOL, COCODIMONIUM HYDROXYPROPYL HYDROLYZED KERATIN, POTASSIUM SORBATE.

Ingredientes: Mica, Paraffinum Liquidum, Magnesium Stearate, Aluminum Starch Octenylsuccinate, Polyisobutene, Ethylhexyl Palmitate, Sorbitan Sesquiolate, Phenoxyethanol, Caprylyl Glycol, Pigment.
Pode Conter: Mica, Tin Oxide, CI 77891, CI 77491, CI 77499, CI 16035, CI 77266, CI 77510.

INGREDIENTES/INGREDIENTS: ALCOHOL DENAT., PARFUM (FRAGRANCE), AQUA (WATER), ETHYLHEXYL SALICYLATE, TRIS (TETRAMETHYLHYDROXYPYPERIDINOL) CITRATE, BUTYL METHOXYDIBENZOYL METHANE, CAPRYLYL GLYCOL, BENZYL BENZOATE, BENZYL SALICYLATE, CITRAL, CITRONELLOL, COUMARIN, GERANIOL, HEXYL CINNAMAL, HYDROXYCITRONELLAL, LIMONENE, LINALOOL.

INGREDIENTS: CYCLOPENTASILOXANE, DIMETHICONE, DIMETHICONOL, C12-15 ALKYL BENZOATE, SCLEROCARYA BIRREA SEED OIL, ROSA CANINA FRUIT OIL, SIMMONDSIA CHINENSIS (JOJOBA) SEED OIL, CHENOPODIUM QUINOA SEED EXTRACT, TOCOPHERYL ACETATE, PARFUM, ETHYLHEXYL SALICYLATE, CI 47000.

DIMETHYL SILYLATE, TRIBEHENIN, ACACIA SENEGAL GUM, XANTHAN GUM, POTASSIUM SORBATE, HEXYLENE GLYCOL, TETRASODIUM EDTA, PENTAERYTHRITYL TETRA-DI-T-BUTYL HYDROXYHYDROCINNAMATE, TOCOPHERYL ACETATE, CITRULLUS LANATUS (WATERMELON) SEED EXTRACT, ASTROCARYUM MURUMURU SEED BUTTER, ASTROCARYUM TUCUMA SEED BUTTER, BUTYROSPERMUM PARKII (SHEA) BUTTER, PUNICA GRANATUM STEROLS, TETRAHEXYLDECYL ASCORBATE, THEOBROMA GRANDIFLORUM SEED BUTTER, LENS ESCULENTA (LENTIL) FRUIT EXTRACT, PYRUS MALUS (APPLE) FRUIT EXTRACT, SODIUM LACTATE, SODIUM HYALURONATE, SODIUM PCA, SODIUM BENZOATE. MAY CONTAIN/PEUT CONTENIR: IRON OXIDES (CI 77491, CI 77492, CI 77499), TITANIUM DIOXIDE (CI 77891).

INGREDIENTS: AQUA/WATER/EAU, HAMAMELIS VIRGINIANA (WITCH HAZEL) WATER, ALCOHOL DENAT., GLYCERIN, ALCOHOL, OLETH-10, PARFUM (FRAGRANCE), CITRONELLOL, EUGENOL, LIMONENE, LINALOOL, PHENOXYETHANOL.

Ingredientes: Sodium Silicate, Ammonium Persulfate, Magnesium Trisilicate, Hydrated Silica, Paraffinum Liquidum, Cellulose Gum, Sorbitol, Zea Mays Starch, EDTA, Hydroxyethylcellulose, Sodium Lauryl Sulfate, Parfum, Prunus Persica Kernel Oil, CI 77007, Glycine Soja Oil, Helianthus Annuus Seed Oil, Linalool, Amyl Cinnamal, Hexyl Cinnamal, Chamomilla Recutita Flower Extract, Rosa Centifolia Flower Extract, BHT.

INGREDIENTS/INGRÉDIENTS/INGREDIENTES: WATER (AQUA/EAU), SODIUM LAURETH SULFATE, ACRYLATES COPOLYMER, FRAGRANCE (PARFUM), PROPYLENE GLYCOL, COCAMIDOPROPYL BETAINE, SODIUM HYDROXIDE, SODIUM CHLORIDE, ALOE BARBADENSIS LEAF JUICE, DISODIUM EDTA, BHT, METHYLCHLOROISO THIAZOLINONE, METHYLISOTHIAZOLINONE, LIMONENE, LINALOOL.

Ingredients: Propylene Glycol, Water, Sodium Stearate, Fragrance, Glyceryl Laurate, Glycol Distearate, Butyrospermum Parkii (Shea Butter), Sodium Chloride, Steareth-21, Butylated Hydroxy Toluene.