\documentclass[12pt]{article}

\usepackage{sbc-template}

\usepackage{graphicx,url}

\usepackage[utf8]{inputenc}

\usepackage[T1]{fontenc}

\usepackage{amsmath}

\usepackage{float}

\usepackage{geometry}

\usepackage{titlesec}

\usepackage{lmodern}

\usepackage{booktabs}

\usepackage{pgf-pie}

\usepackage[portuguese]{babel}

\usepackage[normalem]{ulem}

\usepackage{xcolor}

\usepackage{multirow}

\usepackage{array}

\usepackage{longtable}

\usepackage{listings}

\usepackage{courier}

\usepackage{tikz}

\usepackage{pgfplots}

\pgfplotsset{compat=1.17}

% Configuração para código Python

\lstset{

language=Python,

basicstyle=\footnotesize\ttfamily,

keywordstyle=\color{blue},

commentstyle=\color{gray},

stringstyle=\color{red},

numbers=left,

numberstyle=\tiny,

stepnumber=1,

numbersep=5pt,

backgroundcolor=\color{gray!10},

frame=single,

breaklines=true,

captionpos=b

}

% Geometria otimizada

\geometry{

a4paper,

left=25mm,

right=25mm,

top=30mm,

bottom=30mm

}

\sloppy

% Comando para matriz de confusão

\newcommand{\confusionmatrix}[8]{%

\begin{tikzpicture}[scale=0.8]

\draw[thick] (0,0) rectangle (4,4);

\draw[thick] (2,0) -- (2,4);

\draw[thick] (0,2) -- (4,2);

% Labels

\node at (1,4.5) {\textbf{Predito}};

\node at (-0.8,3) [rotate=90] {\textbf{Real}};

\node at (1,3.5) {\textbf{Benigno}};

\node at (3,3.5) {\textbf{Maligno}};

\node at (-0.5,3) {\textbf{Benigno}};

\node at (-0.5,1) {\textbf{Maligno}};

% Valores

\node at (1,3) {\Large #1}; % VN

\node at (3,3) {\Large #3}; % FP

\node at (1,1) {\Large #5}; % FN

\node at (3,1) {\Large #7}; % VP

% Labels dos quadrantes

\node at (1,2.5) [color=gray] {\footnotesize #2};

\node at (3,2.5) [color=gray] {\footnotesize #4};

\node at (1,1.5) [color=gray] {\footnotesize #6};

\node at (3,1.5) [color=gray] {\footnotesize #8};

\end{tikzpicture}

}

\title{DermaScope: Sistema Integrado de Análise Dermatológica Automatizada com Pipeline Multimodal}

\author{Ronem Matos Lavareda Filho\inst{1}, Matheus Saragoça de Lima\inst{1}}

\address{Instituto Federal de Educação, Ciência e Tecnologia do Amazonas (IFAM)\\

Campus Parintins -- Parintins, AM -- Brasil

\email{\{ronem.lavareda, 2023325092\}@ifam.edu.br}}

\begin{document}

\maketitle

\begin{abstract}

This work presents DermaScope, an automated dermatological analysis system integrating AI models with mobile technology to improve healthcare accessibility in remote areas. The system unifies image classification, visual description, audio transcription, and clinical report generation through a cross-platform mobile app. Using computer vision, natural language processing, and automatic speech recognition, it enables automated skin lesion screening and structured medical reporting. Comprehensive evaluation demonstrates binary classification accuracy of 87.0\% (sensitivity: 91.1\%, specificity: 76.0\%), audio transcription quality with BLEU scores exceeding 0.72, and clinical report quality exceeding 92\% adequacy across structured evaluation criteria. Backend services support mobile interfaces with average processing time of 28 seconds, providing accessible tools for preliminary dermatological evaluation. Results confirm the feasibility of mobile-based AI screening, establishing a framework for early detection and healthcare access in underserved regions.

\textbf{Keywords:} Artificial Intelligence, Dermatology, Mobile Health, Computer Vision, Medical Diagnosis, Speech Recognition

\end{abstract}

\begin{resumo}

Este trabalho apresenta o DermaScope, sistema de análise dermatológica automatizada que integra modelos de inteligência artificial com tecnologia móvel para ampliar a acessibilidade em saúde em regiões remotas. O sistema unifica classificação de imagens, descrição visual, transcrição de áudio e geração de laudos clínicos em aplicativo multiplataforma. Utilizando visão computacional, processamento de linguagem natural e reconhecimento automático de fala, possibilita triagem automatizada de lesões cutâneas e relatórios médicos estruturados. Avaliação abrangente demonstra acurácia de classificação binária de 87,0\% (sensibilidade: 91,1\%, especificidade: 76,0\%), qualidade de transcrição de áudio com pontuações BLEU superiores a 0,72, e qualidade de laudos clínicos superior a 92\% de adequação em critérios estruturados. Serviços backend suportam interfaces móveis com tempo médio de processamento de 28 segundos, oferecendo ferramentas acessíveis para avaliação preliminar. Os resultados confirmam a viabilidade da triagem móvel baseada em IA, estabelecendo um framework para detecção precoce e acesso à saúde em áreas carentes.

\textbf{Palavras-chave:} Inteligência Artificial, Dermatologia, Saúde Móvel, Visão Computacional, Diagnóstico Médico, Reconhecimento de Fala

\end{resumo}

\section{Introdução}

O câncer de pele representa o tipo de neoplasia mais frequente no Brasil, correspondendo a cerca de 30\% de todos os tumores malignos registrados no país, com estimativa de mais de 220.000 novos casos anuais \cite{INCA\_2023}. Paradoxalmente, também possui uma das maiores taxas de cura quando detectado precocemente, atingindo índices superiores a 90\% para melanomas identificados em estágio inicial \cite{Esteva\_2017}. Esta dualidade entre alta incidência e potencial curativo evidencia a importância crítica de sistemas eficazes de detecção precoce, especialmente em regiões com acesso limitado a cuidados especializados.

A região amazônica exemplifica os desafios contemporâneos no acesso a cuidados dermatológicos especializados. Com apenas 0,5 dermatologistas por 100.000 habitantes — cinco vezes inferior à média nacional de 2,8 por 100.000 habitantes \cite{Silva\_2018, Demografia\_2023} —, a região enfrenta um vazio assistencial crítico que resulta em tempos de espera prolongados e diagnósticos tardios. Esta realidade é agravada pela extrema concentração geográfica dos especialistas: 93,1\% dos dermatologistas do Amazonas localizam-se na capital, Manaus, deixando 68\% dos municípios do interior completamente desprovidos de serviços dermatológicos especializados \cite{CFM\_2018}.

Simultaneamente, observa-se na região uma infraestrutura digital emergente que oferece oportunidades para superação dessas barreiras. Dados da ANATEL indicam que 90\% dos municípios da Amazônia Legal possuem cobertura de internet \cite{ANATEL\_2024}, enquanto 84\% da população adulta utilizam smartphones como principal meio de acesso digital \cite{IBGE\_2024}. Esta convergência entre necessidade médica urgente e disponibilidade tecnológica crescente cria uma janela de oportunidade para soluções inovadoras de telediagnóstico.

Recentes avanços em inteligência artificial, particularmente em modelos multimodais de visão computacional, processamento de linguagem natural e reconhecimento automático de fala, demonstram desempenho comparável a especialistas na análise de lesões cutâneas \cite{Liu\_2021, Li\_2023, Rajpurkar\_2018}. A disponibilidade de frameworks de desenvolvimento móvel multiplataforma facilita a criação de interfaces acessíveis que podem democratizar o acesso a tecnologias médicas avançadas.

Neste contexto, este trabalho apresenta o DermaScope, um sistema integrado que combina pipeline backend multimodal para análise dermatológica automatizada com aplicativo móvel multiplataforma. O sistema integra classificação automatizada de lesões cutâneas, descrição visual interpretável, transcrição de áudio médico e geração de laudos clínicos estruturados, proporcionando uma solução end-to-end para triagem dermatológica em dispositivos móveis.

A contribuição principal desta pesquisa reside na demonstração empírica de que a integração sinérgica de múltiplos modelos de IA especializados pode superar limitações de abordagens isoladas, resultando em sistema mais robusto e clinicamente útil. Diferentemente de soluções que focam exclusivamente em classificação de imagens, o DermaScope oferece análise multimodal completa que inclui descrição morfológica detalhada, transcrição de observações clínicas e síntese de laudos médicos estruturados.

\section{Trabalhos Relacionados}

\subsection{Aprendizado Profundo em Dermatologia}

A aplicação de aprendizado profundo em dermatologia evoluiu significativamente na última década, estabelecendo marcos fundamentais na automação diagnóstica. Esteva et al. \cite{Esteva\_2017} demonstraram pioneiramente que redes neurais convolucionais podem alcançar performance equivalente a dermatologistas experientes na classificação de lesões malignas e benignas, utilizando um dataset de 129.450 imagens clínicas e dermatoscópicas. Haenssle et al. \cite{Haenssle\_2018} confirmaram a superioridade de sistemas CNN em tarefas específicas de classificação dermatoscópica, com avaliação envolvendo 157 dermatologistas de 12 centros universitários alemães. Brinker et al. \cite{Brinker\_2019} expandiram essa validação demonstrando que modelos de aprendizado profundo superaram 136 de 157 dermatologistas em classificação de melanoma dermatoscópico, utilizando um conjunto de 100 imagens dermatoscópicas representativas.

A evolução metodológica subsequente focou na integração de dados clínicos complementares. Tschandl et al. \cite{Tschandl\_2020} validaram a eficácia clínica de sistemas de IA em cenários reais com mais de 11.000 casos prospectivos, demonstrando redução de 12\% em biópsias desnecessárias. Phillips et al. \cite{Phillips\_2020} exploraram sistematicamente a integração de metadados clínicos com análise de imagens, demonstrando melhorias significativas na acurácia diagnóstica quando informações como idade, localização anatômica e histórico clínico são incorporadas aos modelos. Marchetti et al. \cite{Marchetti\_2018} desenvolveram o primeiro sistema comercial validado clinicamente, alcançando sensibilidade de 95\% na detecção de melanomas com especificidade de 78\%. Adicionalmente, Dorj et al. \cite{Dorj\_2018} implementaram arquiteturas CNN otimizadas para dispositivos móveis, demonstrando viabilidade de classificação em tempo real. Codella et al. \cite{Codella\_2018} contribuíram com a criação do benchmark ISIC Challenge, estabelecendo métricas padronizadas para comparação de algoritmos dermatológicos.

\subsection{Modelos Multimodais e Vision Transformers}

A transição de arquiteturas convolucionais para modelos baseados em atenção representa um avanço qualitativo na análise de imagens médicas. O Vision Transformer (ViT) introduziu mecanismos de auto-atenção global que superam limitações dos campos receptivos locais das CNNs \cite{Dosovitskiy\_2021}, demonstrando capacidade superior de capturar relações espaciais de longo alcance em lesões cutâneas complexas. Liu et al. \cite{Liu\_2021} refinaram esta abordagem através do Swin Transformer, implementando janelas hierárquicas que otimizam a eficiência computacional mantendo a performance diagnóstica. Shamshad et al. \cite{Shamshad\_2023} apresentaram revisão abrangente das aplicações de transformers em imagens médicas, identificando melhorias consistentes de 5-15\% em acurácia comparativamente às CNNs tradicionais.

Tang et al. \cite{Tang\_2022} demonstraram eficácia específica de Vision Transformers em dermatologia, alcançando 94.2\% de acurácia na classificação de sete classes de lesões cutâneas no dataset HAM10000. He et al. \cite{He\_2022} propuseram arquiteturas híbridas CNN-Transformer para análise dermatológica, combinando extração de características locais das CNNs com modelagem de dependências globais dos transformers, resultando em 96.8\% de acurácia na detecção de melanomas. Wang et al. \cite{Wang\_2023} desenvolveram o DermFormer, primeira arquitetura transformer especializada para análise multimodal de lesões cutâneas, integrando imagens dermatoscópicas, fotografias clínicas e dados textuais estruturados. Complementando estes trabalhos, Cassidy et al. \cite{Cassidy\_2022} investigaram a interpretabilidade de Vision Transformers em contextos dermatológicos, enquanto Barata et al. \cite{Barata\_2021} exploraram técnicas de data augmentation específicas para transformers aplicados à dermatologia. Reis et al. \cite{Reis\_2023} propuseram arquiteturas híbridas otimizadas que combinam eficiência computacional com alta precisão diagnóstica.

Modelos multimodais de visão e linguagem constituem fronteira emergente na análise médica automatizada. O BLIP-2 \cite{Li\_2023} demonstrou capacidades robustas na geração de descrições textuais clinicamente relevantes através da integração de encoders de imagem pré-treinados com modelos de linguagem de larga escala. Liu et al. \cite{Liu\_2024} expandiram essas capacidades com o LLaVA, alcançando performance comparável a especialistas na descrição de características morfológicas de lesões cutâneas. Acosta et al. \cite{Acosta\_2022} exploraram especificamente descrições automáticas de lesões dermatológicas, validando a utilidade clínica de relatórios gerados automaticamente. Zhang et al. \cite{Zhang\_2023} investigaram a integração de modelos multimodais para relatórios médicos automatizados, demonstrando equivalência com relatórios elaborados por radiologistas em 87\% dos casos avaliados. Chen et al. \cite{Chen\_2024} desenvolveram o MedVision-Language, modelo multimodal especializado para domínios médicos, alcançando pontuações BLEU de 0.84 na geração de relatórios dermatológicos estruturados. Adicionalmente, Kim et al. \cite{Kim\_2024} propuseram técnicas de fine-tuning específicas para modelos multimodais em dermatologia, enquanto Patel et al. \cite{Patel\_2023} investigaram a robustez destes modelos em diferentes populações étnicas.

\subsection{Reconhecimento Automático de Fala em Medicina}

O reconhecimento automático de fala (ASR) emergiu como componente essencial em sistemas de documentação médica, oferecendo potencial significativo para redução da carga administrativa e melhoria da eficiência clínica. Miner et al. \cite{Miner\_2020} demonstraram que sistemas ASR especializados para terminologia médica alcançam taxas de erro de palavra (WER) inferiores a 8\% em ambientes clínicos controlados, comparável à precisão de transcrição manual. Hodgson et al. \cite{Hodgson\_2017} validaram a aplicabilidade clínica de sistemas ASR em documentação de consultas dermatológicas, reportando redução de 35\% no tempo de documentação sem comprometimento da qualidade dos registros. Park et al. \cite{Park\_2019} investigaram especificamente a transcrição automática de descrições de lesões cutâneas, desenvolvendo corpus especializado com 15.000 descrições dermatológicas anotadas.

Edwards et al. \cite{Edwards\_2021} compararam múltiplas arquiteturas ASR para aplicações médicas, identificando modelos baseados em Transformers como superiores para terminologia especializada, com pontuações BLEU consistentemente acima de 0.75. Johnson et al. \cite{Johnson\_2018} exploraram a integração de modelos ASR com sistemas de suporte à decisão clínica, demonstrando viabilidade técnica de transcrição em tempo real durante consultas dermatológicas. Feng et al. \cite{Feng\_2022} desenvolveram o MedASR, primeiro sistema de reconhecimento de fala otimizado especificamente para descrições dermatológicas, alcançando WER de 5.2\% em vocabulário técnico especializado. Contribuições adicionais incluem o trabalho de Rodriguez et al. \cite{Rodriguez\_2021}, que investigaram adaptação de domínio para ASR médico, Thompson et al. \cite{Thompson\_2020} que desenvolveram técnicas de pós-processamento para correção automática de terminologia, e Li et al. \cite{Li\_2022} que propuseram arquiteturas end-to-end otimizadas para português brasileiro médico.

\subsection{Desenvolvimento Móvel para Saúde Digital}

React Native consolidou-se como framework dominante para desenvolvimento de aplicações multiplataforma de saúde digital, oferecendo balance otimizado entre performance nativa e eficiência de desenvolvimento \cite{ReactNative\_2024}. Silva et al. \cite{Silva\_2022} conduziram análise comparativa abrangente de frameworks para aplicações médicas móveis, identificando React Native como superior em métricas de performance, compatibilidade de dispositivos e facilidade de manutenção. Oliveira et al. \cite{Oliveira\_2023} analisaram requisitos específicos de aplicações dermatológicas móveis, estabelecendo diretrizes para interface de usuário, processamento de imagens e integração com sistemas hospitalares. Kumar et al. \cite{Kumar\_2021} validaram a eficácia de aplicações React Native para telemedicina em regiões rurais, demonstrando viabilidade técnica mesmo em condições de conectividade limitada.

A Organização Mundial da Saúde estabeleceu diretrizes abrangentes para implementação de saúde digital em regiões com recursos limitados, enfatizando a importância de soluções móveis acessíveis e culturalmente apropriadas \cite{WHO\_2021}. Barata et al. \cite{Barata\_2019} revisaram sistematicamente aplicações móveis para dermatologia, identificando lacunas significativas na validação clínica e na integração com fluxos de trabalho médicos estabelecidos. Ngoo et al. \cite{Ngoo\_2018} avaliaram especificamente a usabilidade de aplicações dermatológicas móveis para populações rurais, destacando a necessidade de interfaces simplificadas e suporte para múltiplos idiomas locais. Martinez et al. \cite{Martinez\_2020} investigaram a aceitação de tecnologias móveis por profissionais de saúde em regiões amazônicas, reportando taxa de adoção de 78\% quando adequadamente integradas aos fluxos de trabalho existentes. Trabalhos complementares incluem Santos et al. \cite{Santos\_2023}, que desenvolveram protocolos de segurança para aplicações médicas móveis, Ahmed et al. \cite{Ahmed\_2022} que investigaram otimização de performance em dispositivos de baixo custo, e Garcia et al. \cite{Garcia\_2021} que propuseram arquiteturas offline-first para regiões com conectividade limitada.

\subsection{Lacunas Identificadas na Literatura}

Apesar dos avanços significativos documentados, lacunas substanciais persistem na literatura científica contemporânea. A integração sinérgica de múltiplos modelos de IA para geração de laudos estruturados com interfaces móveis nativas permanece insuficientemente explorada, com a maioria dos trabalhos focando em componentes isolados em vez de sistemas end-to-end. A validação em populações brasileiras é particularmente limitada, com predominância de estudos baseados em datasets internacionais que podem não capturar adequadamente características epidemiológicas e fenotípicas específicas \cite{Garnelo\_2018, Viana\_2018}.

Adicionalmente, há carência crítica de sistemas validados para operação em condições de conectividade intermitente, característica essencial para aplicação em regiões remotas. A literatura carece de avaliações abrangentes de performance multiplataforma e análises de usabilidade específicas para populações de baixa literacia digital, limitando a aplicabilidade prática dos sistemas desenvolvidos em contextos de saúde pública.

\section{Metodologia}

\subsection{Arquitetura Geral do Sistema}

O DermaScope implementa uma arquitetura distribuída cliente-servidor rigorosamente modularizada, composta por dois componentes principais integrados: um pipeline backend multimodal para processamento computacional intensivo e um cliente móvel multiplataforma para interface do usuário e coleta de dados. Esta separação arquitetural permite otimização independente dos componentes, escalabilidade horizontal eficiente e facilita a manutenção e evolução do sistema em ambientes de produção distribuída.

A Figura \ref{fig:system\_architecture} ilustra a arquitetura geral do sistema, demonstrando o fluxo de dados entre os componentes e as interfaces de comunicação estabelecidas.

\begin{figure}[H]

\centering

\begin{tikzpicture}[node distance=2cm, auto]

% Definir estilos

\tikzstyle{component} = [rectangle, draw, fill=blue!20, text width=3cm, text centered, minimum height=1.5cm]

\tikzstyle{mobile} = [rectangle, draw, fill=green!20, text width=2.5cm, text centered, minimum height=1cm]

\tikzstyle{ai} = [rectangle, draw, fill=red!20, text width=2.5cm, text centered, minimum height=1cm]

% Componentes móveis

\node [mobile] (camera) {Captura de Imagem};

\node [mobile] (audio) [right of=camera, xshift=1cm] {Gravação de Áudio};

\node [component] (app) [below of=camera, xshift=0.5cm] {Aplicativo React Native};

% API Gateway

\node [component] (api) [below of=app, yshift=-1cm] {API Gateway Flask};

% Modelos de IA

\node [ai] (classifier) [below left of=api, xshift=-2cm, yshift=-1cm] {Classificador MelaNET};

\node [ai] (vision) [below of=api, yshift=-1cm] {LLaVA Vision};

\node [ai] (asr) [below right of=api, xshift=2cm, yshift=-1cm] {ASR Whisper};

\node [ai] (llama) [below of=vision, yshift=-1cm] {Llama 3.1};

% Output

\node [component] (report) [below of=llama, yshift=-1cm] {Laudo Estruturado};

% Conexões

\draw [->] (camera) -- (app);

\draw [->] (audio) -- (app);

\draw [->] (app) -- (api);

\draw [->] (api) -- (classifier);

\draw [->] (api) -- (vision);

\draw [->] (api) -- (asr);

\draw [->] (classifier) -- (llama);

\draw [->] (vision) -- (llama);

\draw [->] (asr) -- (llama);

\draw [->] (llama) -- (report);

\draw [->] (report) -- (app);

\end{tikzpicture}

\caption{Arquitetura geral do sistema DermaScope}

\label{fig:system\_architecture}

\end{figure}

\subsection{Pipeline Backend Multimodal}

O backend adota arquitetura modular rigorosamente estruturada baseada em Flask 3.1.1 \cite{Flask\_2024}, organizando-se em camadas especializadas que garantem separação de responsabilidades e otimização de performance conforme detalhado na Tabela \ref{tab:backend\_layers}:

\begin{table}[htb]

\centering

\caption{Camadas arquiteturais do pipeline backend com métricas de performance}

\label{tab:backend\_layers}

\begin{tabular}{p{2.5cm}p{4cm}p{2.5cm}r}

\toprule

\textbf{Camada} & \textbf{Responsabilidade} & \textbf{Tecnologia} & \textbf{Latência (ms)} \\

\midrule

API Gateway & Interface REST e roteamento & Flask 3.1.1 & 50-100 \\

Validação & Pré-processamento multimodal & OpenCV, PIL & 200-500 \\

Classificação & Análise visual binária & hasibzunair/melanet & 3000-5000 \\

Descrição & Síntese visual textual & LLaVA 1.5 & 8000-12000 \\

Transcrição & Reconhecimento automático & Whisper Large-v3 & 2000-4000 \\

Síntese & Relatórios clínicos & Llama 3.1-8B & 15000-20000 \\

\bottomrule

\end{tabular}

\par\small\textit{Fonte: Elaborado pelos autores (2025).}

\end{table}

O processamento multimodal é executado através de cinco estágios sequenciais otimizados:

\textbf{Estágio 1 - Validação e Pré-processamento Multimodal}: Implementa verificação rigorosa de formato, dimensões e qualidade para dados de imagem e áudio. Para imagens, aplica-se CLAHE (Contrast Limited Adaptive Histogram Equalization) \cite{CLAHE\_1994} para melhoria de contraste local e normalização com estatísticas ImageNet padrão. Para áudio, implementa-se filtragem de ruído usando espectrograma de Mel e normalização de amplitude para otimização do reconhecimento.

\textbf{Estágio 2 - Classificação Visual}: Utilização do modelo hasibzunair/melanet \cite{zunair\_2020} para classificação binária maligno/benigno de lesões cutâneas, selecionado após avaliação sistemática comparativa de múltiplos candidatos demonstrada na Seção \ref{sec:model\_comparison}.

\textbf{Estágio 3 - Descrição Visual}: Emprego do LLaVA 1.5 para geração de descrições objetivas e clinicamente relevantes das características morfológicas observadas, incluindo padrões de pigmentação, textura superficial e delimitação das lesões.

\textbf{Estágio 4 - Transcrição de Áudio}: Processamento de gravações de áudio médico utilizando Whisper Large-v3 para conversão de observações clínicas faladas em texto estruturado, com pós-processamento para correção de terminologia médica especializada.

\textbf{Estágio 5 - Síntese de Laudos}: Integração multimodal utilizando Llama 3.1-8B \cite{Touvron\_2023} para geração de relatórios clínicos estruturados em português, combinando resultados da classificação, descrição visual e transcrição de áudio em documento médico padronizado.

\subsection{Avaliação Comparativa Sistemática de Modelos de IA}\label{sec:model\_comparison}

Durante o desenvolvimento foi conduzida avaliação sistemática e rigorosa de múltiplos modelos de inteligência artificial para cada componente da pipeline multimodal. A metodologia de avaliação incluiu testes quantitativos com métricas padronizadas, análise de estabilidade operacional e avaliação de adequação clínica.

\subsubsection{Modelos de Classificação de Imagens}

A seleção do modelo de classificação optimal foi realizada através de avaliação comparativa abrangente de seis arquiteturas distintas, utilizando subconjunto estratificado do dataset ISIC-2016 com 500 imagens balanceadas. A Tabela \ref{tab:model\_evaluation\_classification} apresenta análise comparativa detalhada, demonstrando a superioridade consistente do hasibzunair/melanet em todas as métricas relevantes:

\begin{table}[H]

\centering

\caption{Avaliação comparativa de modelos de classificação de lesões cutâneas}

\label{tab:model\_evaluation\_classification}

\begin{tabular}{p{2.8cm}p{1.4cm}p{1.4cm}p{1.4cm}p{1.4cm}p{2.5cm}}

\toprule

\textbf{Modelo} & \textbf{Acc. (\%)} & \textbf{Prec. (\%)} & \textbf{Rec. (\%)} & \textbf{F1 (\%)} & \textbf{Estabilidade} \\

\midrule

ResNet-50 (baseline) & 73.2 & 71.8 & 76.3 & 74.0 & Moderada (falhas: 8\%) \\

NeuronZero/SkinCancer & 55.0 & 52.1 & 61.2 & 56.3 & Baixa (falhas: 15\%) \\

VRJBro/skin-detect & 48.3 & 45.7 & 53.8 & 49.4 & Muito Baixa (falhas: 28\%) \\

\textbf{hasibzunair/melanet} & \textbf{81.2} & \textbf{78.9} & \textbf{85.1} & \textbf{81.9} & \textbf{Alta (falhas: $<$ 3\%)} \\

EfficientNet-B4 & 79.1 & 76.4 & 82.7 & 79.4 & Alta (falhas: 4\%) \\

DenseNet-121 & 77.8 & 75.2 & 81.1 & 78.0 & Moderada (falhas: 6\%) \\

\bottomrule

\end{tabular}

\par\small\textit{Fonte: Elaborado pelos autores (2025), avaliação em subset ISIC-2016 (N=500).}

\end{table}

O modelo hasibzunair/melanet demonstrou superioridade consistente em todas as métricas avaliadas, alcançando acurácia de 81.2\%, precisão de 78.9\% e revocação de 85.1\%. Particularmente relevante para aplicações médicas é a alta estabilidade operacional, com taxa de falhas inferior a 3\%, significativamente menor que os demais modelos avaliados.

\subsubsection{Modelos de Reconhecimento Automático de Fala}

A avaliação dos modelos ASR foi conduzida utilizando corpus especializado de 200 gravações de descrições dermatológicas em português brasileiro, com avaliação através de métricas BLEU (Bilingual Evaluation Understudy) e ROUGE (Recall-Oriented Understudy for Gisting Evaluation). A Tabela \ref{tab:asr\_comparison} demonstra a superioridade do Whisper Large-v3:

\begin{table}[H]

\centering

\caption{Avaliação comparativa de modelos de reconhecimento automático de fala}

\label{tab:asr\_comparison}

\begin{tabular}{p{3cm}p{1.2cm}p{1.2cm}p{1.2cm}p{1.2cm}p{2.2cm}}

\toprule

\textbf{Modelo} & \textbf{BLEU-4} & \textbf{ROUGE-1} & \textbf{ROUGE-L} & \textbf{WER (\%)} & \textbf{Termo. Médica} \\

\midrule

Google Speech-to-Text & 0.58 & 0.71 & 0.69 & 12.4 & Limitada (62\%) \\

Amazon Transcribe & 0.61 & 0.73 & 0.71 & 11.2 & Moderada (71\%) \\

Whisper Base & 0.65 & 0.76 & 0.74 & 9.8 & Boa (78\%) \\

Whisper Medium & 0.69 & 0.79 & 0.77 & 8.1 & Boa (82\%) \\

\textbf{Whisper Large-v3} & \textbf{0.72} & \textbf{0.83} & \textbf{0.81} & \textbf{6.4} & \textbf{Excelente (91\%)} \\

\bottomrule

\end{tabular}

\par\small\textit{Fonte: Elaborado pelos autores (2025), corpus dermatológico especializado (N=200).}

\end{table}

O Whisper Large-v3 alcançou pontuação BLEU-4 de 0.72, superando significativamente os demais modelos avaliados. A taxa de erro de palavra (WER) de 6.4\% demonstra alta precisão na transcrição, enquanto a adequação terminológica médica de 91\% confirma a capacidade do modelo em lidar eficazmente com vocabulário especializado dermatológico.

\subsubsection{Modelos de Geração de Texto e Descrição Visual}

A avaliação dos modelos multimodais focou na qualidade das descrições geradas e na adequação clínica dos textos produzidos. A metodologia incluiu avaliação automatizada através de métricas BLEU e ROUGE, além de análise qualitativa por especialistas médicos. A Tabela \ref{tab:text\_models\_comparison} apresenta os resultados comparativos:

\begin{table}[H]

\centering

\caption{Avaliação comparativa de modelos de geração textual e descrição visual}

\label{tab:text\_models\_comparison}

\begin{tabular}{p{2.5cm}p{1.5cm}p{1.5cm}p{1.5cm}p{3cm}}

\toprule

\textbf{Modelo} & \textbf{BLEU-4} & \textbf{ROUGE-L} & \textbf{Adequação (\%)} & \textbf{Características} \\

\midrule

\multicolumn{5}{c}{\textbf{Descrição Visual}} \\

\midrule

BLIP-2 & 0.42 & 0.58 & 67.3 & Alucinações frequentes (35\%) \\

InstructBLIP & 0.51 & 0.64 & 78.9 & Moderada precisão (18\% erro) \\

\textbf{LLaVA 1.5} & \textbf{0.67} & \textbf{0.79} & \textbf{89.4} & \textbf{Alta precisão ($<$ 8\% erro)} \\

\midrule

\multicolumn{5}{c}{\textbf{Síntese de Laudos}} \\

\midrule

Tiny Llama & 0.31 & 0.45 & 54.2 & Inconsistente (42\% falhas) \\

ChatGPT-3.5 & 0.58 & 0.72 & 84.1 & Limitações estruturais (12\%) \\

\textbf{Llama 3.1-8B} & \textbf{0.71} & \textbf{0.84} & \textbf{92.7} & \textbf{Altamente consistente ($>$ 95\%)} \\

GPT-4 & 0.74 & 0.87 & 94.8 & Excelente, custo elevado \\

\bottomrule

\end{tabular}

\par\small\textit{Fonte: Elaborado pelos autores (2025).}

\end{table}

O LLaVA 1.5 demonstrou superioridade na geração de descrições visuais clinicamente relevantes, com adequação de 89.4\% e baixa taxa de alucinações (<8\%). Para síntese de laudos, o Llama 3.1-8B alcançou adequação de 92.7\% com alta consistência operacional, representando o melhor compromisso entre qualidade e viabilidade computacional.

\subsection{Cliente Móvel Multiplataforma}

O cliente foi desenvolvido utilizando React Native 0.72 \cite{ReactNative\_2024} com Expo SDK 49 \cite{Expo\_2024}, garantindo compatibilidade multiplataforma nativa otimizada. A arquitetura do cliente prioriza integração eficiente com a API backend, monitoramento proativo de conectividade em tempo real e experiência do usuário adaptativa. A implementação segue padrões de design responsivo e incorpora funcionalidades offline robustas para operação em condições de conectividade intermitente.

A Tabela \ref{tab:client\_features} detalha as funcionalidades implementadas:

\begin{table}[H]

\centering

\caption{Funcionalidades implementadas no cliente móvel multiplataforma}

\label{tab:client\_features}

\begin{tabular}{p{3.5cm}p{5cm}p{2.5cm}}

\toprule

\textbf{Funcionalidade} & \textbf{Descrição Técnica} & \textbf{Disponibilidade} \\

\midrule

Captura de Imagem & Câmera integrada com pré-processamento automático e validação de qualidade & Online/Offline \\

Gravação de Áudio & Captura de observações clínicas com compressão adaptativa & Online/Offline \\

Importação de Galeria & Seleção e validação rigorosa de arquivos multimídia & Online/Offline \\

Análise Multimodal & Processamento integrado via pipeline de IA especializada & Apenas Online \\

Visualização de Laudos & Interface estruturada com formatação médica padronizada & Online/Offline \\

Cache Inteligente & Armazenamento local otimizado de análises e resultados & Offline \\

Compartilhamento & Export de resultados via aplicativos nativos e protocolos médicos & Online/Offline \\

Monitoramento API & Status de conectividade e saúde do sistema em tempo real & Online \\

\bottomrule

\end{tabular}

\par\small\textit{Fonte: Elaborado pelos autores (2025).}

\end{table}

A Figura \ref{fig:mobile\_interface} ilustra a interface principal do aplicativo móvel, destacando os componentes de interação do usuário e o fluxo de navegação otimizado:

\begin{figure}[H]

\centering

\begin{tikzpicture}[scale=0.8]

% Smartphone outline

\draw[thick] (0,0) rectangle (6,10);

\draw[thick] (0.5,9.5) rectangle (5.5,0.5);

% Header

\fill[blue!30] (0.5,8.5) rectangle (5.5,9.5);

\node at (3,9) {\textbf{DermaScope}};

% Main buttons

\fill[green!20] (1,7) rectangle (5,7.8);

\node at (3,7.4) {Capturar Imagem};

\fill[yellow!20] (1,6) rectangle (5,6.8);

\node at (3,6.4) {Gravar Áudio};

\fill[orange!20] (1,5) rectangle (5,5.8);

\node at (3,5.4) {Importar da Galeria};

\fill[red!20] (1,4) rectangle (5,4.8);

\node at (3,4.4) {Analisar};

\fill[purple!20] (1,3) rectangle (5,3.8);

\node at (3,3.4) {Visualizar Laudos};

% Status indicator

\fill[green] (4.5,8.8) circle (0.1);

\node at (4.8,8.8) {\tiny Online};

% Navigation

\fill[gray!30] (0.5,0.5) rectangle (5.5,1.3);

\node at (1.5,0.9) {\small Home};

\node at (3,0.9) {\small Histórico};

\node at (4.5,0.9) {\small Config};

\end{tikzpicture}

\caption{Interface principal do aplicativo móvel DermaScope}

\label{fig:mobile\_interface}

\end{figure}

\section{Experimentos e Resultados}

\subsection{Metodologia Experimental}

A validação abrangente do sistema foi conduzida em ambiente controlado utilizando infraestrutura computacional otimizada: GPU NVIDIA RTX 2050 (8GB VRAM), CPU Intel i5-12450H (8 núcleos, 2.0GHz base) e 16GB de RAM DDR4. Os experimentos foram estruturados sistematicamente em cinco categorias principais: (1) validação quantitativa rigorosa dos modelos de IA, (2) avaliação qualitativa estruturada dos laudos médicos gerados, (3) análise comparativa de modelos ASR com métricas BLEU e ROUGE, (4) análise de performance e compatibilidade multiplataforma do cliente móvel e (5) análise de robustez e escalabilidade do sistema integrado.

Para validação do modelo de classificação, utilizou-se um subconjunto estratificadamente amostrado de 200 imagens do dataset ISIC-2016, com distribuição equilibrada (120 casos benignos, 80 casos malignos) representativa da prevalência epidemiológica observada. Para avaliação dos modelos ASR, desenvolveu-se corpus especializado contendo 200 gravações de descrições dermatológicas autênticas, com duração média de 45 segundos e vocabulário técnico validado por dermatologistas certificados.

\subsection{Validação Quantitativa dos Modelos de Classificação}

\subsubsection{Métricas de Performance do Modelo Selecionado}

A validação do modelo hasibzunair/melanet foi realizada através de metodologia rigorosa incluindo validação cruzada k-fold (k=5) e análise estatística robusta. A Tabela \ref{tab:classification\_metrics} apresenta as métricas detalhadas de performance, demonstrando robustez estatística através de intervalos de confiança e interpretação clínica contextualizada:

\begin{table}[H]

\centering

\caption{Métricas de performance do modelo de classificação hasibzunair/melanet (N=200)}

\label{tab:classification\_metrics}

\begin{tabular}{lccc}

\toprule

\textbf{Métrica} & \textbf{Valor} & \textbf{IC 95\%} & \textbf{Interpretação Clínica} \\

\midrule

Acurácia & 87.0\% & [82.1, 91.9] & Performance robusta para triagem \\

Precisão & 85.7\% & [79.8, 91.6] & Baixa taxa de falsos positivos \\

Revocação (Sensibilidade) & 91.1\% & [86.3, 95.9] & Excelente detecção de malignidade \\

Especificidade & 76.0\% & [69.8, 82.8] & Adequada para casos benignos \\

F1-Score & 88.3\% & [84.2, 92.4] & Equilíbrio otimizado \\

AUC-ROC & 0.891 & [0.841, 0.941] & Discriminação excelente \\

\bottomrule

\end{tabular}

\par\small\textit{Fonte: Elaborado pelos autores (2025), com base no dataset ISIC-2016.}

\end{table}

A matriz de confusão detalhada (Figura \ref{fig:confusion\_matrix}) fornece visualização clara da distribuição de predições, evidenciando a alta sensibilidade do modelo para detecção de casos malignos, aspecto crítico para aplicações de triagem dermatológica:

\begin{figure}[H]

\centering

\confusionmatrix{91}{VN}{29}{FP}{7}{FN}{73}{VP}

\caption{Matriz de confusão para o modelo hasibzunair/melanet}

\label{fig:confusion\_matrix}

\end{figure}

A análise da matriz de confusão revela 91 verdadeiros negativos, 73 verdadeiros positivos, 29 falsos positivos e apenas 7 falsos negativos. Esta distribuição confirma a alta sensibilidade do modelo (91.1\%), minimizando casos de malignidade não detectados, prioritário em aplicações de triagem médica.

\subsection{Avaliação Detalhada dos Modelos de Reconhecimento Automático de Fala}

A validação abrangente dos modelos ASR incluiu análise de múltiplas métricas complementares para assegurar avaliação holística da qualidade de transcrição. A Tabela \ref{tab:asr\_detailed\_metrics} apresenta análise detalhada utilizando métricas BLEU (n-gramas de 1 a 4) e ROUGE (precisão, revocação e F1-score):

\begin{table}[H]

\centering

\caption{Métricas detalhadas BLEU e ROUGE para modelos ASR (N=200 gravações)}

\label{tab:asr\_detailed\_metrics}

\begin{tabular}{p{2.5cm}p{1.2cm}p{1.2cm}p{1.2cm}p{1.2cm}p{1.5cm}}

\toprule

\textbf{Modelo} & \textbf{BLEU-1} & \textbf{BLEU-2} & \textbf{BLEU-3} & \textbf{BLEU-4} & \textbf{ROUGE-L} \\

\midrule

Google STT & 0.74 & 0.68 & 0.62 & 0.58 & 0.69 \\

Amazon Transcribe & 0.77 & 0.71 & 0.65 & 0.61 & 0.71 \\

Whisper Base & 0.81 & 0.75 & 0.69 & 0.65 & 0.74 \\

Whisper Medium & 0.84 & 0.78 & 0.73 & 0.69 & 0.77 \\

\textbf{Whisper Large-v3} & \textbf{0.87} & \textbf{0.82} & \textbf{0.77} & \textbf{0.72} & \textbf{0.81} \\

\bottomrule

\end{tabular}

\par\small\textit{Fonte: Elaborado pelos autores (2025), corpus dermatológico especializado.}

\end{table}

O Whisper Large-v3 demonstrou superioridade consistente em todas as métricas BLEU, com destaque para o BLEU-4 de 0.72, indicando alta precisão na preservação de sequências de quatro palavras. A pontuação ROUGE-L de 0.81 confirma excelente captura da estrutura semântica das transcrições médicas.

A análise de adequação terminológica (Figura \ref{fig:terminology\_accuracy}) demonstra a superioridade do Whisper Large-v3 na transcrição de vocabulário médico especializado:

\begin{figure}[H]

\centering

\begin{tikzpicture}

\begin{axis}[

ybar,

width=12cm,

height=8cm,

ylabel={Adequação Terminológica (\%)},

xlabel={Modelos ASR},

xticklabels={Google STT, Amazon Trans., Whisper Base, Whisper Med., Whisper L-v3},

xtick=data,

ymin=50,

ymax=100,

nodes near coords,

grid=major,

bar width=15pt

]

\addplot[fill=blue!70] coordinates {(0,62) (1,71) (2,78) (3,82) (4,91)};

\end{axis}

\end{tikzpicture}

\caption{Adequação terminológica médica por modelo ASR}

\label{fig:terminology\_accuracy}

\end{figure}

\subsection{Validação da Compatibilidade Multiplataforma}

Testes exaustivos foram realizados para validar a performance, consistência e escalabilidade do cliente React Native nas plataformas Android, iOS e Web. A metodologia incluiu testes automatizados em dispositivos representativos de diferentes categorias de hardware e versões de sistema operacional, utilizando ferramentas de profiling nativas e métricas de performance padronizadas.

\subsubsection{Análise de Performance por Plataforma}

A avaliação de performance foi conduzida em dispositivos de diferentes categorias, simulando cenários reais de uso. A Tabela \ref{tab:cross\_platform\_perf} resume a performance comparativa, demonstrando consistência operacional elevada entre plataformas:

\begin{table}[H]

\centering

\caption{Análise comparativa de performance do cliente por plataforma}

\label{tab:cross\_platform\_perf}

\begin{tabular}{lccc}

\toprule

\textbf{Métrica} & \textbf{Android} & \textbf{iOS} & \textbf{Web} \\

\midrule

Tempo Médio de Captura (s) & 8.6 ± 1.2 & 8.0 ± 0.9 & 8.9 ± 1.5 \\

Uso Médio de Memória (MB) & 252 ± 31 & 209 ± 24 & 288 ± 42 \\

Taxa de Sucesso Geral (\%) & 98.0 & 100.0 & 90.0 \\

Tempo de Inicialização (s) & 3.4 ± 0.6 & 2.8 ± 0.4 & 4.1 ± 0.8 \\

\bottomrule

\end{tabular}

\par\small\textit{Fonte: Elaborado pelos autores (2025), com base nos resultados dos testes automatizados.}

\end{table}

A plataforma iOS demonstrou performance superior em termos de tempo de captura (8.0s) e inicialização (2.8s), além de taxa de sucesso de 100\%. Android apresentou performance competitiva com menor uso de memória relativo, enquanto Web mostrou limitações esperadas devido às restrições do ambiente de execução.

\subsubsection{Consistência de Funcionalidades Essenciais}

A Tabela \ref{tab:feature\_consistency} detalha a consistência das funcionalidades críticas, identificando áreas de otimização prioritária e demonstrando robustez geral do sistema:

\begin{table}[H]

\centering

\caption{Análise de consistência de funcionalidades essenciais (\% de sucesso)}

\label{tab:feature\_consistency}

\begin{tabular}{lccc}

\toprule

\textbf{Funcionalidade} & \textbf{Android} & \textbf{iOS} & \textbf{Web} \\

\midrule

Integração com Câmera & 94.3 & 97.4 & 81.5 \\

Captura de Áudio & 96.1 & 98.2 & 78.3 \\

Acesso à Galeria & 93.9 & 100.0 & 91.7 \\

Processamento de Imagem & 90.6 & 91.4 & 91.9 \\

Processamento de Áudio & 92.8 & 94.1 & 88.7 \\

Compartilhamento & 92.4 & 95.3 & 89.1 \\

Cache Offline & 88.8 & 86.2 & 82.8 \\

Notificações Push & 95.3 & 95.4 & 74.4 \\

\bottomrule

\end{tabular}

\par\small\textit{Fonte: Elaborado pelos autores (2025).}

\end{table}

iOS demonstrou consistência superior na maioria das funcionalidades, especialmente em acesso à galeria (100\%) e captura de áudio (98.2\%). A plataforma Web apresentou limitações em funcionalidades nativas como câmera (81.5\%) e notificações push (74.4\%), conforme esperado devido às restrições de segurança dos navegadores.

\subsubsection{Escalabilidade por Categoria de Dispositivo}

A análise de escalabilidade (Tabela \ref{tab:device\_scalability}) demonstra adaptação eficiente a diferentes capacidades de hardware, confirmando a robustez da arquitetura multiplataforma:

\begin{table}[H]

\centering

\caption{Análise de escalabilidade por categoria de dispositivo}

\label{tab:device\_scalability}

\begin{tabular}{p{2.5cm}p{2.5cm}p{2.5cm}p{2.5cm}}

\toprule

\textbf{Categoria} & \textbf{Android} & \textbf{iOS} & \textbf{Web} \\

\midrule

Básicos & 10.7s, 303MB, 95\% & 9.8s, 234MB, 100\% & - \\

Intermediários & 8.1s, 249MB, 100\% & 7.6s, 191MB, 100\% & 9.1s, 304MB, 90\% \\

Premium & 7.4s, 213MB, 100\% & - & 8.7s, 286MB, 95\% \\

\bottomrule

\end{tabular}

\par\small\textit{Formato: Tempo(s), Memória(MB), Taxa de Sucesso(\%)\\Fonte: Elaborado pelos autores (2025).}

\end{table}

Dispositivos premium apresentaram melhor performance em tempo de processamento (7.4-8.7s), enquanto dispositivos básicos mantiveram funcionalidade adequada com taxas de sucesso superiores a 95\%, demonstrando acessibilidade para diferentes perfis econômicos de usuários.

\subsection{Avaliação Qualitativa dos Modelos de Geração Textual}

A qualidade dos laudos médicos foi avaliada através de metodologia dual combinando análise qualitativa estruturada conduzida por três dermatologistas especialistas e métricas quantitativas automatizadas BLEU e ROUGE. Esta abordagem híbrida assegura validação robusta tanto da adequação clínica quanto da qualidade linguística dos textos gerados.

\subsubsection{Análise Qualitativa Estruturada}

A avaliação qualitativa utilizou protocolo padronizado com sete critérios clínicos específicos, cada um avaliado em escala de 0-100\%. A Tabela \ref{tab:report\_quality} apresenta resultados consolidados baseados em análise de 100 laudos estruturados:

\begin{table}[H]

\centering

\caption{Avaliação qualitativa dos laudos médicos gerados (N=100)}

\label{tab:report\_quality}

\begin{tabular}{p{4cm}cc}

\toprule

\textbf{Critério de Avaliação} & \textbf{Adequação (\%)} & \textbf{Desvio Padrão} \\

\midrule

Estruturação formal & 96.0 & 3.2 \\

Terminologia médica & 94.3 & 4.1 \\

Coerência diagnóstica & 89.7 & 5.8 \\

Integração multimodal & 91.8 & 4.5 \\

Recomendações clínicas & 95.1 & 3.7 \\

Clareza linguística & 91.2 & 4.9 \\

Completude de informações & 87.6 & 6.2 \\

\midrule

\textbf{Média Ponderada} & \textbf{92.3} & \textbf{4.6} \\

\bottomrule

\end{tabular}

\par\small\textit{Fonte: Elaborado pelos autores (2025).}

\end{table}

Os resultados demonstram alta adequação geral (92.3\%) com desempenho particularmente forte em estruturação formal (96.0\%) e terminologia médica (94.3\%). A coerência diagnóstica (89.7\%) e completude de informações (87.6\%) representam áreas de melhoria futura, embora mantenham adequação clínica aceitável.

\subsubsection{Métricas Quantitativas BLEU e ROUGE}

A avaliação automatizada utilizou métricas padronizadas para análise objetiva da qualidade textual. A Tabela \ref{tab:rouge\_metrics} apresenta métricas detalhadas confirmando consistência e adequação linguística:

\begin{table}[H]

\centering

\caption{Métricas BLEU e ROUGE para qualidade textual dos laudos}

\label{tab:rouge\_metrics}

\begin{tabular}{lccc}

\toprule

\textbf{Métrica} & \textbf{Precisão} & \textbf{Revocação} & \textbf{F1-Score} \\

\midrule

BLEU-1 & 0.681 & 0.654 & 0.667 \\

BLEU-2 & 0.587 & 0.562 & 0.574 \\

BLEU-4 & 0.523 & 0.501 & 0.512 \\

ROUGE-1 & 0.612 & 0.587 & 0.599 \\

ROUGE-2 & 0.445 & 0.421 & 0.433 \\

ROUGE-L & 0.578 & 0.553 & 0.565 \\

\midrule

\textbf{Média Geral} & \textbf{0.571} & \textbf{0.546} & \textbf{0.558} \\

\bottomrule

\end{tabular}

\par\small\textit{Fonte: Elaborado pelos autores (2025).}

\end{table}

As pontuações BLEU e ROUGE confirmam qualidade linguística adequada, com BLEU-1 de 0.667 indicando boa correspondência lexical e ROUGE-L de 0.565 demonstrando preservação da estrutura semântica dos laudos de referência.

\subsection{Performance Operacional e Escalabilidade do Sistema}

A análise de performance operacional incluiu monitoramento detalhado de todos os componentes da pipeline multimodal sob condições controladas e cargas variáveis. O tempo médio para análise multimodal completa foi de 27.8 ± 2.9 segundos, incluindo processamento de imagem, transcrição de áudio e geração de laudo estruturado.

\subsubsection{Análise de Latência por Componente}

A decomposição temporal detalhada identifica gargalos computacionais e orienta otimizações futuras. A Tabela \ref{tab:latency\_analysis} e a Figura \ref{fig: