

**UNIVERSIDADE DO VALE DO RIO DOS SINOS  
UNIDADE ACADÊMICA DE GRADUAÇÃO  
CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO**

**JAQUELINE DAHMER STEFFENON**

**KATIE**

**Modelo de tecnologia assistiva aplicando visão computacional para auxiliar pessoas com  
deficiência visual na identificação de medicamentos**

São Leopoldo  
2020

JAQUELINE DAHMER STEFFENON

**KATIE**

**Modelo de tecnologia assistiva aplicando visão computacional para auxiliar pessoas com deficiência visual na identificação de medicamentos**

Artigo apresentado como requisito parcial para obtenção do título de Bacharel em Ciência da Computação, pelo Curso de Ciência da Computação da Universidade do Vale do Rio dos Sinos (UNISINOS)

Orientador: Prof. Dr. Mateus Raeder

São Leopoldo  
2020

## **KATIE: Modelo de tecnologia assistiva aplicando visão computacional para auxiliar pessoas com deficiência visual**

Jaqueline Dahmer Steffenon<sup>1</sup>

Mateus Raeder<sup>2</sup>

**Resumo:** A tecnologia assistiva visa promover a inclusão de pessoas com deficiência na participação de atividades cotidianas através do uso de inovações tecnológicas. O Brasil conta com aproximadamente 1 milhão de cegos, de modo que torna-se imprescindível o desenvolvimento de tecnologias para melhorar a vida dessa população. A partir do propósito de promover a autonomia dessas pessoas, chega-se à temática deste trabalho: a expansão do acesso a informações àqueles que dele são privados em virtude de cegueira. Tarefas rotineiras, tais como a identificação de uma cartela de medicamentos ou leitura da embalagem de um produto alimentício, transformam-se em desafios para aqueles que convivem com essa deficiência. Inspirado pela solução já existente do *Be my eyes*, este trabalho emerge com o intuito de desenvolver uma solução tecnológica para o problema e tornar a realização dessas tarefas corriqueiras totalmente independente de outras pessoas. Nesse sentido, apresenta-se o modelo KATIE, no qual o usuário pode realizar uma pergunta sobre o ambiente em que se encontra e ter uma resposta em tempo real. Os resultados obtidos com testes na identificação de medicamentos e cédulas de dinheiro, mostram que o protótipo foi capaz de cumprir o objetivo estimado, realizando o reconhecimento da pergunta e dados da imagem e retornando uma resposta adequada ao usuário.

**Palavras-chave:** Pessoas com deficiência visual. Visão Computacional. Reconhecimento de fala. Processamento de Linguagem Natural.

**Abstract:** Assistive technology aims to promote the inclusion of people with disabilities in daily activities through the use of technological innovations. Brazil has approximately 1 million blind people, so it is essential to develop technologies to improve the lives of this population. From the objective of promoting the autonomy of these people, the theme of this work is reached: the expansion of access to information to those who are deprived of it due to blindness. Routine tasks, such as identifying a drug pack or reading the packaging of a food product, become challenges for those who live with this disability. Routine tasks, such as identifying a drug pack or reading the packaging of a food product, become challenges for those who live with this disability. Inspired by existing solution *Be my eyes*, this work emerges in order to develop a technological solution to the problem and make these totally independent mundane tasks of others. In this sense, the KATIE model is presented, in which the user can ask a question about the scenario in which he finds himself and have an answer in real time. The results obtained with tests in the identification of medicines and cash notes, show that the prototype was able to fulfill the estimated objective, performing the recognition of the question and image data and returning an adequate answer to the user.

**Keywords:** Visually impaired people. Computer Vision. Speech Recognition. Natural Language Processing.

---

<sup>1</sup>Graduanda em Ciência da Computação pela Unisinos. Email: jaquelinesteffenon@gmail.com

<sup>2</sup>Orientador. Coordenador do curso de Ciência da Computação da UNISINOS. Email: mraeder@unisinos.br

## 1 INTRODUÇÃO

Conforme o portal do Governo Digital (2014), acessibilidade é conceituada como a inclusão de pessoas com deficiência na participação de atividades cotidianas, como o uso de produtos, serviços e informações. Assim, a tecnologia assistiva visa, justamente, promover acessibilidade através do uso de inovações tecnológicas para a criação de ferramentas que atendam às demandas de pessoas com necessidades específicas.

Dentre os 17 Desafios de Desenvolvimento do Milênio da Organização das Nações Unidas, destaca-se a intenção de tornar as cidades espaços mais acessíveis e inclusivos para todos e todas. (ONU BRASIL, 2015). Nesse sentido, visando a concretização desse objetivo, surge a temática deste trabalho: a expansão do acesso à informação àqueles que não dispõem dele em virtude da manifestação de deficiência visual.

Segundo uma pesquisa recente do Conselho Brasileiro de Oftalmologia (2019), há no Brasil cerca de 1 milhão de pessoas cegas, além dos inúmeros casos de pessoas com outras escalas de deficiência visual. No panorama global pelo menos 2,2 bilhões de pessoas são deficientes visuais ou cegas. (WORLD HEALTH ORGANIZATION, 2019).

Logo, tarefas rotineiras como, por exemplo, a leitura de rótulos, a escolha do sabor de refrigerante pelo o que está escrito na lata, a consulta de informações de produtos alimentícios e medicamentos ou a realização da leitura de correspondências, entre tantas outras, tornam-se barreiras para o contingente de indivíduos que convivem com problemas de visão.

Diante desse contexto, uma das soluções existentes no mercado para mitigar tal problema é o aplicativo *Be my eyes*, por meio do qual cegos são conectados a voluntários via vídeo chamada com a finalidade de que os olhos do voluntário sejam "emprestados" para auxiliar as pessoas com deficiência visual na realização de tarefas. Presente em mais de 150 países, o aplicativo conta com aproximadamente três milhões de voluntários cadastrados, além de 170 mil cegos e pessoas com baixa visão. (BE MY EYES, 2015).

Contudo, seja pelo tempo necessário para a efetuação da ajuda do voluntário, seja pela exposição da privacidade de quem precisa de ajuda ou até mesmo a dependência de outras pessoas, soluções como o *Be My Eyes* nem sempre são acessíveis e práticas. Partindo de tal princípio, este trabalho emerge com o intuito de desenvolver uma solução tecnológica para o problema, de modo a tornar a realização de tarefas corriqueiras totalmente independente de outras pessoas. Para tanto, o intuito é desenvolver um sistema que reconheça a pergunta do usuário, analise as características do ambiente pela imagem da câmera e, após uma comparação entre a dúvida e o cenário, resulte em uma resposta para a situação.

O objetivo geral deste trabalho consiste na criação de um modelo capaz de reconhecer e responder perguntas realizadas pelo usuário sobre o contexto em que ele está inserido, como por exemplo a identificação de medicamentos. Além disso, todo o desenvolvimento deverá ser projetado considerando o público alvo da proposta (pessoas com deficiência visual), tornando essa uma tecnologia assistiva. Os objetivos específicos foram elencados a partir da subdivisão

do objetivo geral em atividades específicas. Seguem relacionados abaixo:

- Estudar técnicas para reconhecimento de fala;
- Explorar métodos para obtenção de textos de imagem;
- Pesquisar ferramentas capazes de analisar dados e retornar uma resposta;
- Implementar um protótipo baseado nas tecnologias estudadas;
- Avaliar a aplicação proposta por meio de testes.

Este trabalho está estruturado em 7 seções. A seção 2 apresenta os conceitos principais encontrados na literatura que estão envolvidos na evolução do presente trabalho, detalhando os seguintes temas: tecnologias assistivas, inteligência artificial, visão computacional, reconhecimento de fala, processamento de linguagem natural e chatbot. Na seção 3, são apresentados os trabalhos relacionados por meio de seções em que cada um deles é descrito e comparado entre eles; por fim, são destacados os diferenciais dos trabalhos anteriores com o modelo atual apresentado neste artigo. Na seção 4, são apresentados os detalhes do modelo KATIE, elucidando características das etapas da modelagem computacional. Na seção 5, está presente a implementação do modelo, ao passo que a seção 6 conta com os testes e resultados obtidos. Por fim, a seção 7 apresenta as considerações finais e as perspectivas futuras a partir deste trabalho.

## 2 FUNDAMENTAÇÃO TEÓRICA

Com o intuito de adentrar nas explicações a respeito do modelo proposto, é necessário haver embasamento nos conceitos que o permeiam. Desse modo, na sequência, seguem apresentadas definições e pesquisas a respeito dos tópicos de tecnologia assistiva, visão computacional, e processamento de linguagem natural.

### 2.1 Tecnologia Assistiva

Frente a novas realidades e paradigmas na sociedade, é almejada uma sociedade mais permeável à diversidade, questionando os mecanismos de segregação e aspirando meios de inclusão social para pessoas com deficiência. Este contexto somado ao avanço tecnológico proporcionam um ambiente favorável para o surgimento das tecnologias assistivas.

Conforme Bersch (2008) o termo *Assistive Technology*, conhecido no Brasil como Tecnologia Assistiva, foi criado oficialmente em 1988 como importante elemento jurídico dentro da legislação norte americana, conhecida como *Public Law 100-407*, que constitui juntamente com outras leis o *ADA - American with Disabilities Act*. Este conjunto de regulamentações define os direitos civis de pessoas com deficiência nos Estados Unidos e também garantem um fundo público para obtenção de recursos necessários a esses usuários. Nesse sentido, a Tecnologia

Assistiva ou simplesmente TA, teve sua regulamentação legal juntamente a essa lei, proporcionando o benefício de serviços e recursos que promovem uma vida mais independente, produtiva e incluída com relação a sociedade.

No Brasil, o conceito oficial de tecnologia assistiva ficou consignado na Lei Brasileira de Inclusão (LBI) também conhecida como Estatuto das Pessoas com Deficiência, a qual reproduziu o conceito elaborado pelo Comitê de Ajudas Técnicas. O Artigo 3º da LBI, promulgada em janeiro de 2016, define Tecnologia Assistiva como:

III - Produtos, recursos, metodologias, estratégias, práticas e serviços que objetivam promover a funcionalidade, relacionada à atividade e participação, de pessoas com deficiência, incapacidades ou mobilidade reduzida, visando sua autonomia, independência, qualidade de vida e inclusão social. (BRASIL, 2015).

Em suma, TA é uma área de conhecimento que engloba recursos e serviços com o objetivo de proporcionar ou ampliar habilidades funcionais de uma pessoa com deficiência ou com incapacidades advindas do envelhecimento, promovendo independência e inclusão aos usuários. Além disso, pode ser definida como "uma ampla gama de equipamentos, serviços, estratégias e práticas concebidas e aplicadas para minorar os problemas encontrados pelos indivíduos com deficiências". (COOK; HUSSEY, 2002).

Logo, o foco da TA é proporcionar aos usuários independência, qualidade de vida e inclusão social através da ampliação de suas formas de comunicação, mobilidade, controle ambiental, habilidades de aprendizado, trabalho e integração com família, amigos e sociedade.

## 2.2 Visão computacional

Segundo Hayo (2017) a proposta da Visão Computacional é emular a visão humana usando imagens digitais através de três componentes principais de processamento, executados um após o outro aquisição, processamento, análise e compreensão de imagens. Dessa forma, tal área tem como objetivo fazer com que os computadores tenham uma compreensão visual do mundo que seja semelhante ou superior à dos humanos.

A Visão Computacional é um campo de estudo da Ciência da Computação e Inteligência Artificial, o qual teve seus primeiros experimentos já em 1950, com a construção de redes neurais para detecção de limites de objetos categorizando-os como círculos e quadrados. Em 1970, a visão computacional teve sua primeira aplicação a nível comercial, onde o avanço permitia a interpretação de textos manuscritos e digitados utilizando reconhecimento ótico de caracteres para pessoas com deficiência visual. Com a evolução da internet e o grande volume de imagens disponíveis, em 1990, o reconhecimento facial começou seus primeiros passos.

Atualmente, as aplicações são as mais diversas, alimentando um sistema com várias imagens de um mesmo objeto, o modelo tem a capacidade de compreender os padrões e consegue fazer

o reconhecimento com facilidade.

### 2.3 Processamento de Linguagem Natural

Com o reconhecimento de fala, surge a área de Processamento da Linguagem Natural (PLN), uma subárea da Inteligência Artificial a qual estuda a capacidade e as limitações de uma máquina em entender a linguagem dos seres humanos. O objetivo da PLN é desenvolver algoritmos capazes de entender e compor textos, isso significa, reconhecer o contexto, realizar as análises sintática, semântica, léxica e morfológica, criar resumos, extrair informações e analisar sentimentos, além de aprender com os dados processados.

A PLN atualmente lida com problemas relacionados à automação da interpretação e da geração da língua humana em aplicações como tradução automática, sumarização automática de textos, ferramentas de auxílio à escrita, perguntas e respostas, categorização textual, recuperação e extração de informação. (SOCIEDADE BRASILEIRA DE COMPUTAÇÃO, 2019).

Além de permitir que computadores se comuniquem na linguagem dos seres humanos, o PLN possibilita que computadores leiam textos, ouçam e interpretem falas, identificando sentimentos, significados e determinando quais trechos são importantes. Para o processo de desenvolvimento de recursos, ferramentas e aplicações, essa área possui raízes fortes na área de Inteligência Artificial, que torna-se base para os projetos.

## 3 TRABALHOS RELACIONADOS

Esta seção tem como objetivo citar e descrever trabalhos anteriormente publicados que se relacionam ao tema proposto, de modo a explicitar a busca por projetos de visão computacional e processamento de linguagem natural que se aplicam a contextos de apoio a pessoas com algum grau de deficiência visual. Para cada artigo relacionado, há um resumo compacto, no qual constam objetivos, metodologia, resultados, dificuldades e ideias de trabalhos futuros.

### 3.1 Uso de Visão Computacional em Dispositivos Móveis para o Reconhecimento de Faixa de Pedestres

O trabalho elaborado por Sousa e Marengoni (2012) propõe o desenvolvimento de um sistema para auxiliar pessoas com deficiência visual na detecção de pontos seguros para a travessia de ruas através de um aplicativo *mobile*. Inicialmente o sistema é ativado pelo usuário por meio do toque na tela ou por voz, então analisa a imagem da câmera para analisar a condição de travessia na faixa de pedestre e, por fim informa ao usuário se é a travessia é segura ou não, por áudio ou vibração do aparelho.

O algoritmo foi concebido na plataforma *Android* utilizando a biblioteca *OpenCV* para análise da imagem e para a transmissão de informação por voz foi utilizada a própria biblioteca

fornecida pelo SDK do Android. O pré-processamento da imagem é feito ajustando a resolução, enquadramento e escala de cores. Em seguida, a função do *OpenCV* para a detecção de contornos é aplicada e os resultados são comparados com as regras estipuladas de definição de faixas de pedestres. Essas regras estipulam essencialmente a necessidade de três contornos no mínimo para que sejam considerados válidos, uma vez que é comum algum dos componentes não estar de acordo com as regras em função de condições de desgaste ou iluminação.

Quanto aos testes do sistema, foram aplicados 41 cenários, dos quais 31 detectaram corretamente as faixas de segurança, resultando numa eficácia de 76%, afirmando a viabilidade do sistema. Nos cenários mal-sucedidos observou-se variações climáticas, de iluminação e também deformações na faixa que podem ser o motivo para tal fracasso. A importância da temática do trabalho é muito evidente e com os resultados apresentados, sugerem como proposta de produção futura, estender o reconhecimento para outros sinais que indiquem se é viável ou não a travessia, como reconhecimento de pessoas, carros e faróis.

### **3.2 Design of a Mobile Face Recognition System for Visually Impaired Persons**

A identificação de pessoas é considerada uma tarefa difícil para pessoas com determinados graus de deficiência visual. Embora o reconhecimento de voz seja um método alternativo, ainda é um processo intuitivo e impreciso para essa parcela da população. Com o avanço do poder computacional Chaudhry e Chandra (2015) optaram por desenvolver uma solução tecnológica para essa problemática.

A tecnologia assistiva proposta utiliza um dispositivo móvel com câmera, para a detecção e reconhecimento de faces; além disso, usa um fone de ouvido para interação usuário-sistema. Para o desenvolvimento da aplicação, a primeira etapa foi o algoritmo de detecção de faces, desenvolvido utilizando a biblioteca *OpenCV* do próprio *Android* e assim, foi treinado com padrões para encontrar rostos. Na sequência, o rosto identificado é comparado com os registrados no banco de dados do sistema, para enfim, retornar ao usuário o nome da pessoa identificada.

Os experimentos mostraram uma acurácia de 93% na detecção de rostos e 70% na precisão de reconhecimento de rostos em condições bem iluminadas. Além disso, os pesquisadores analisaram o desempenho do aplicativo, o qual teve pouco consumo de dados e memória. Por outro lado, as limitações mostram um grande uso da energia e processamento. As aplicações futuras desse trabalho envolvem um aprimoramento do algoritmo de detecção e reconhecimento, testes com usuários reais e estudo sobre uso de *wearables* para facilitar a interação do sistema com os usuários.

### **3.3 Real Time Multi Object Detection for Blind Using Single Shot Multibox Detector**

Com foco em conquistar a autonomia da mobilidade de cegos, Arora et al. (2019) propõe a construção de um protótipo que execute a detecção de objetos em tempo real, por meio do uso

de segmentação de imagens e redes neurais profundas, informando os objetos próximos e suas categorias.

Os itens necessários para o desenvolvimento do protótipo foram um *Raspberry pi*, fones de ouvido e uma câmera fixa em um boné. Com a coleta da imagem pela câmera, as informações seguem para o *Raspberry pi*, onde o algoritmo realiza o processamento da imagem. Com base no seu treinamento, identifica o objeto retornando para o usuário através de um áudio no fone de ouvido.

A acurácia apresentada pelo sistema é realmente alta, nos testes apresentados chega a 98,95% para detecção de pessoas e 98,75% para detecção de cães. Como trabalhos futuros, os pesquisadores pretendem promover um aumento na eficiência do dispositivo, utilizando algoritmos mais complexos para diminuir a latência, adicionar mais câmeras para aumentar a precisão e, inclusive, investir em reconhecimento facial, para o algoritmo aprender a identificar rostos familiares do usuário.

### 3.4 VQA: Visual Question Answering

Buscando disponibilizar respostas a perguntas cotidianas de pessoas com determinados graus de deficiência visual, o trabalho de Antol et al. (2015) analisa uma foto tirada pelo usuário e a pergunta sobre o contexto. A partir disso, gera a resposta correspondente.

Por meio de *datasets* de perguntas e imagens já existentes, como o *MS COCO (Microsoft Common Objects in Context)*, o grupo elaborou um modelo para responder as perguntas através da análise da imagem e texto. Além disso, também foram feitas diversas análises estatísticas como por exemplo, com relação as palavras mais comuns nas perguntas ou sobre a necessidade da imagem para responder determinada questão.

A precisão do melhor modelo obtido no padrão de teste de respostas para perguntas visuais foi de 54,06%. Segundo a avaliação dos autores, estima-se que o modelo tem um desempenho para responder perguntas tão bom quanto uma criança de 4 a 5 anos.

### 3.5 VizWiz Grand Challenge: Answering Visual Questions from Blind People

Com a finalidade de incentivar o desenvolvimento de projeto que atendam ao interesse de pessoas cegas, a proposta de Gurari et al. (2018) visa desenvolver um modelo de conjunto de dados originários de pessoas cegas.

Partindo de um trabalho anterior, no qual um aplicativo de celular coletou mais de 70.000 perguntas visuais, contendo uma foto e uma pergunta correspondente. Tal trabalho iniciou com um processo de filtragem para a remoção de dados que comprometessem a segurança ou privacidade dos usuários. Em seguida foram coletadas fontes para apoiar o treinamento e avaliação dos algoritmos e também experimentos para caracterizar as imagens, perguntas e, finalmente, as respostas. Para os testes, foram utilizados oito métodos diferentes para avaliar a

possibilidade de construção de respostas para as perguntas e imagens capturadas.

Por diversos motivos, como falta de foco ou má manuseio da câmera, 28% das imagens receberam o rótulo de "imagem inadequada", fazendo com que esse material não pudesse ser utilizado. Apesar das dificuldades encontradas no desenvolvimento de um algoritmo moderno, o grupo construiu um *dataset* originado de perguntas e imagens registradas por cegos e respostas sobre esses dados. Por fim, o trabalho desenvolvido pelo grupo é *open source*, o que facilita a elaboração de trabalhos futuros.

### 3.6 Google Lookout

No primeiro semestre de 2019, a renomeada multinacional Google lançou um aplicativo que promete fornecer informações sobre os arredores de pessoas com baixa visão, através da visão computacional. Conforme página online do Google (2019), o *Lookout* utiliza informações da câmera e dos sensores do dispositivo móvel para reconhecer objetos e textos e, posteriormente, fornecer as informações que considera importante por meio da fala.

O aplicativo possui quatro categorias para aplicação em diferentes atividades: explorar (modo como o programa inicia por padrão e serve para tarefas diárias); compras (serve para a leitura de códigos de barras de produtos e identificação de notas e moedas); leitura rápida (para leitura de emails, placas e etiquetas); e, por fim, descrição da cena (possibilita que o indivíduo ouça uma descrição instantânea de uma imagem).

Apesar da inovação tecnológica que o sistema proporciona, ainda há algumas limitações. Atualmente, apenas alguns modelos de celular com sistema operacional Android tem suporte ao aplicativo, que encontra-se disponível apenas nos Estados Unidos, ou seja, em língua inglesa.

### 3.7 Microsoft Seeing AI

Outra grande empresa que investiu no desenvolvimento de uma tecnologia assistiva para pessoas com deficiência visual, foi a gigante Microsoft. Descrito pela companhia como um aplicativo gratuito que narra o mundo ao seu redor (MICROSOFT, 2017), conta com multi funções que possibilitam a descrição do que está em frente à câmera do usuário.

Dentre as funcionalidades estão a leitura de pequenos textos, documentos e até mesmo escritas a mão, identificação de cores, notas e moedas, descrição de cenários e reconhecimento de pessoas e suas emoções através da expressão facial. Além disso, o sistema também auxilia na busca pelo código de barra de produtos através de escaneamento, avisando por um efeito sonoro quando ele é detectado, e então faz a identificação por meio do escaneamento,

Conforme a página online da Microsoft (2017), a tecnologia atualmente possui suporte para sete idiomas, sendo eles, inglês, holandês, francês, alemão, italiano, japonês e espanhol. O aplicativo está limitado a ser utilizado apenas em dispositivos com sistema a partir da versão iOS 10.

### 3.8 Comparação de trabalhos relacionados

Durante o trabalho desenvolvido neste artigo foram estudadas diversas tecnologias que possuem objetivos semelhantes ao proposto. Com o intuito de revisar o estado da arte, as pesquisas foram conduzidas relacionando três principais perspectivas que fazem referência ao tema deste trabalho: tecnologia assistiva para pessoas baixa visão ou cegos, o uso de visão computacional e processamento de linguagem natural.

Os principais artigos, estudados nesta seção, são um sistema capaz de informar cegos sobre pontos seguros para a travessia de ruas (SOUSA; MARENGONI, 2012), uma TA para identificar pessoas por meio de reconhecimento facial e repassar as informações aos usuários (CHAUDHRY; CHANDRA, 2015), um protótipo onde objetos são detectados em tempo real informando os objetos próximos (ARORA et al., 2019), um modelo capaz de responder a perguntas por meio da análise de imagens e texto baseado no dataset MS COCO (ANTOL et al., 2015), um modelo construído com base em dados gerados por cegos para a detecção de imagens e perguntas (GURARI et al., 2018), um aplicativo destinado a sistemas Android com o objetivo de fornecer informações sobre os arredores do usuário (GOOGLE, 2019) e, por fim, um aplicativo destinado a sistemas iOS cujo intuito é semelhante ao anterior. (MICROSOFT, 2017).

Com a finalidade de comparar os artigos analisados, um quadro de comparação entre eles foi construído, contrastando a proposta e o que há de tecnologia assistiva, visão computacional e processamento de linguagem natural (PLN) de cada trabalho estudado.

Quadro 1 – Comparação de trabalho relacionados

<b>Autores</b>	<b>Proposta</b>	<b>Tecnologia Assistiva</b>	<b>Visão Computacional</b>	<b>PLN</b>
Sousa e Maren-goni (2012)	Sistema para auxiliar pessoas com deficiência visual na detecção de pontos seguros para travessia de ruas.	Aplicativo mobile.	Sim. Ferramenta OpenCV.	Acionamento do sistema por voz ou toque na tela.
Chaudhry e Chandra (2015)	Solução para identificação de pessoas através de reconhecimento facial.	Aplicativo mobile.	Sim. Ferramenta OpenCV.	Não há.
Arora et al. (2019)	Detectar objetos em tempo real usando segmentação de imagens e redes neurais profundas, informando objetos próximos e suas categorias.	Raspberry Pi, fones de ouvido e câmera fixa em um boné.	Treinamento sobre dataset do MS COCO.	Não há. Apenas retorno de voz com informações ao usuário.
Antol et al. (2015)	Modelo para responder as perguntas através da análise da imagem e texto com base em um dataset.	Modelo base para futuras tecnologias.	Treinamento sobre dataset do MS COCO.	Análise sobre as perguntas do dataset.
Gurari et al. (2018)	Modelo de conjunto de dados para detecção de imagens e perguntas originários de dados gerados por pessoas cegas.	Modelo base para futuras tecnologias.	Análise em imagens coletadas pelo aplicativo VizWiz.	Análise das perguntas coletadas pelo aplicativo VizWiz.
Google (2019)	Utiliza informações da câmera e sensores do celular para informar ao usuário sobre os arredores.	Aplicativo Android.	Sim, ferramentas próprias da Google.	Não há. Apenas retorno de voz com informações ao usuário.
Microsoft (2017)	Com diversas funcionalidades, descreve o que está em frente à câmera do usuário.	Aplicativo iOS.	Sim, ferramentas próprias da Microsoft.	Não há. Apenas retorno de voz com informações ao usuário.
Steffenon (2020)	Por meio de uma pergunta e imagem captada, retorna uma resposta ao usuário, focado na identificação de medicamentos.	Aplicativo Android.	Sim, utilizando o Kit ML do Firebase.	Sim, por meio da ferramenta de <i>chatbot</i> Dialogflow.

Fonte: Elaborado pela autora.

### 3.9 Diferencial do trabalho proposto

Partindo dos aprendizados assimilados pelos trabalhos pesquisados, é possível realizar uma comparação com este trabalho. O artigo da seção 3.1 utiliza visão computacional e reconheci-

mento de fala, porém de uma forma muito restrita para faixas de segurança, enquanto o presente trabalho preconiza a aplicação nos mais diversos cenários, podendo reconhecer diversos objetos. No tópico 3.2 há uma situação parecida, onde a ideia é identificar e reconhecer pessoas, o modelo deste trabalho não pretende, primordialmente, denominar pessoas por meio de reconhecimento facial, apenas demarcar a presença delas. O artigo do item 3.3, por outro lado, desenvolve um modelo semelhante ao proposto, porém não há aplicação de PLN para reconhecimento de perguntas. Nos itens 3.4 e 3.5 o modelo apresentado é apenas construído aplicando visão computacional a um banco de dados, já o modelo KATIE pretende além dessas aplicações, construir algo que possa atingir o público alvo. Os itens 3.6 e 3.7 apresentam produtos já existentes no mercado, o diferencial do modelo KATIE é a aplicação de perguntas do usuário para extrair a resposta pela captura de texto da imagem, ao contrário desses dois itens, que apenas fornecem as informações, sem receber perguntas como entrada. Um outro diferencial comparado com todos os artigos e os dois últimos produtos é que o modelo do presente trabalho será desenvolvido com dados em português. Na seção 4 o modelo KATIE proposto nesse trabalho é apresentado.

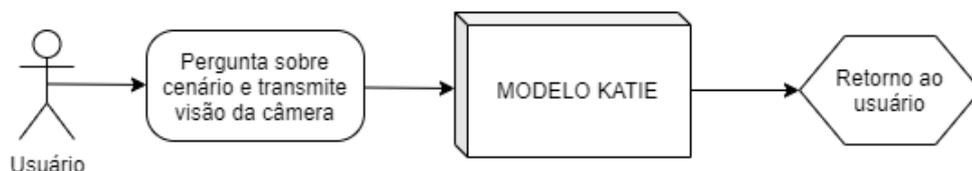
#### 4 MODELO KATIE

Esta seção tem como objetivo descrever a arquitetura do modelo proposto neste trabalho, denominado KATIE, em homenagem a pesquisadora especialista em visão computacional Katherine Louise Bouman, uma das responsáveis por mostrar ao mundo algo que ninguém nunca havia visto, a primeira imagem de um buraco negro obtido com a aplicação da visão computacional. O conteúdo deste tópico visa explicar uma visão geral do trabalho e um aprofundamento sobre as etapas de funcionamento.

##### 4.1 Visão geral

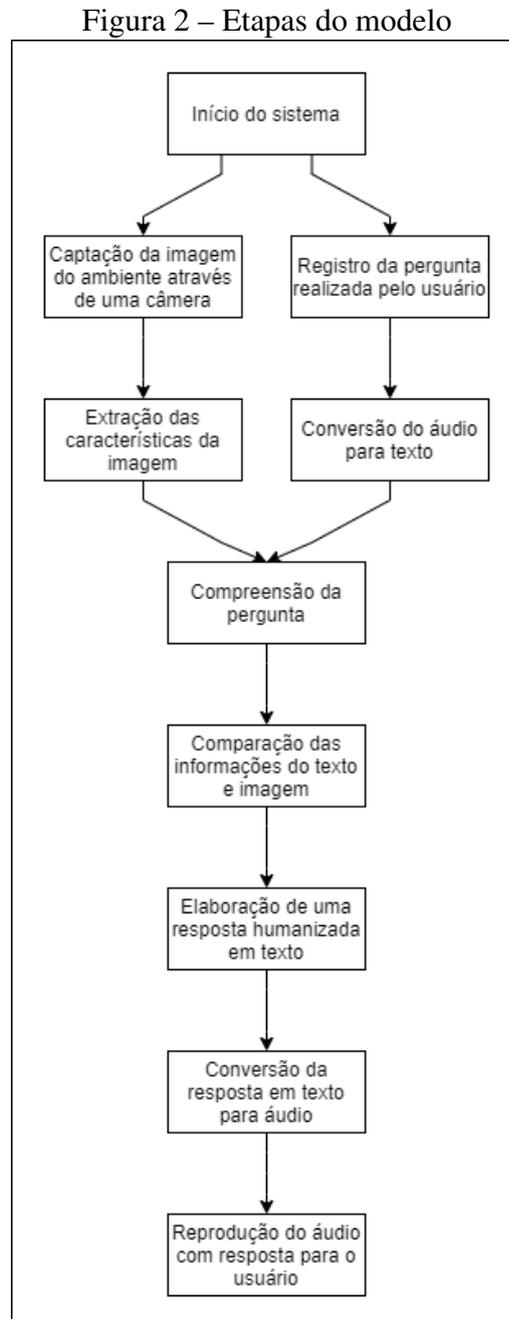
Inspirado no projeto já existente *Be my eyes*, este projeto visa o desenvolvimento de uma solução tecnológica para que pessoas com deficiência visual tenham maior independência na realização de tarefas cotidianas. Quanto ao funcionamento do projeto KATIE, na Figura 1 há uma visão geral de alto nível sobre o seu funcionamento.

Figura 1 – Visão geral do modelo



Fonte: Elaborado pela autora.

Para uma melhor compreensão do funcionamento do modelo, na Figura 2 adentra-se no detalhamento de cada uma das etapas do modelo KATIE, as quais são: o início do sistema, a identificação, o reconhecimento da pergunta, a extração das características da imagem, a análise dos dados e, por fim, a resposta para a pergunta do usuário.



Fonte: Elaborado pela autora.

A seguir são descritas as etapas supracitadas:

#### 1. Início do sistema

A inicialização do sistema se dá com o usuário abrindo a aplicação, direcionando a câmera

para a cena que deseja identificar e realizando sua pergunta.

## 2. Identificação e reconhecimento da pergunta

Após o comando de iniciar, a pergunta é captada e através de um algoritmo com base em técnicas de Processamento de Linguagem Natural, as principais etapas do PLN são percorridas, obtendo então uma identificação do objetivo da pergunta realizada.

## 3. Extração das características da imagem

Previamente, antes da utilização do sistema, o algoritmo precisa ser previamente treinado para identificar diversos objetos e suas características em imagens para que, ao usuário utilizar a solução, seus resultados sejam satisfatórios, também é possível a utilização de um algoritmo já treinado. A compreensão dos objetos da imagem torna possível a elaboração de uma comparação entre a pergunta do usuário e o que a câmera capturou. O não reconhecimento de objeto pode acontecer devido a fatores como: o mau posicionamento da câmera para obter a imagem e a tentativa de reconhecimento de objetos para os quais o algoritmo ainda não foi treinado.

## 4. Análise dos dados

Nessa etapa, são comparadas as informações compreendidas nas fases de análise da pergunta feita pelo usuário e as características obtidas nas imagens da câmera. Com essa comparação é possível encontrar um resultado coerente e, conseqüentemente, uma resposta para o usuário.

## 5. Resposta para a pergunta do usuário

Por fim, com o resultado da análise da imagem e o texto da pergunta, esses dados são analisados por um *chatbot*, para então a resposta ser dada ao usuário, preferencialmente no formato de um áudio.

Com todas as etapas do modelo projetadas, é possível, em um cenário onde o usuário necessita de auxílio, por exemplo, para identificar uma cartela de medicamento. Para tal situação, o sistema será inicializado pelo usuário; em seguida, a câmera será direcionada para os objetos e a dúvida será questionada para o servidor. Após a compreensão da pergunta e o reconhecimento das características da imagem passada pela câmera de forma paralela, o sistema compilará as informações. Dessa forma, por meio da sua inteligência, responderá o usuário. Esse cenário escrito é ilustrado na Figura 3.

Figura 3 – Exemplo prático do modelo



Fonte: Elaborado pela autora.

## 5 IMPLEMENTAÇÃO

Essa seção aborda temas referentes à implementação do modelo KATIE, dividida em três subseções: conversas com o público alvo, escolha das tecnologias e o protótipo. O objetivo foi colocar em prática o modelo teórico através do desenvolvimento de um aplicativo para dispositivos móveis, com a finalidade de avaliar os resultados do modelo.

### 5.1 Conversas com o público alvo

Com a finalidade de avaliar o modelo da solução e entender mais sobre a problemática, foram marcadas conversas com duas pessoas com baixa visão e uma pessoa cega. O ponto central do diálogo foi entender se havia algum aplicativo de apoio que utilizavam no dia-a-dia, no formato do *Be my eyes*, além do ato de elencar as maiores dificuldades em tarefas cotidianas, indagar o sistema operacional que utilizam em seus celulares e falar sobre a proposta do protótipo do modelo KATIE.

Quanto ao uso de outros aplicativos, nenhuma delas conhecia o *Be my eyes* e uma delas afirmou ter utilizado um aplicativo semelhante onde apenas apontava para os arredores e a ferramenta fazia a identificação e leitura, sem o reconhecimento de uma pergunta como entrada. Atualmente, a forma como as pessoas entrevistadas resolvem essas dúvidas é por meio do contato com amigos e familiares. A estratégia utilizada é o envio de uma foto por meio de aplicativo de conversa.

As dificuldades citadas foram várias, sendo passível de citar: o uso de equipamentos eletrodomésticos que não possuem acessibilidade, como micro-ondas, forno e máquina de lavar; a identificação de descrição, modo de uso, preço e validade de produtos; a detecção de objetos

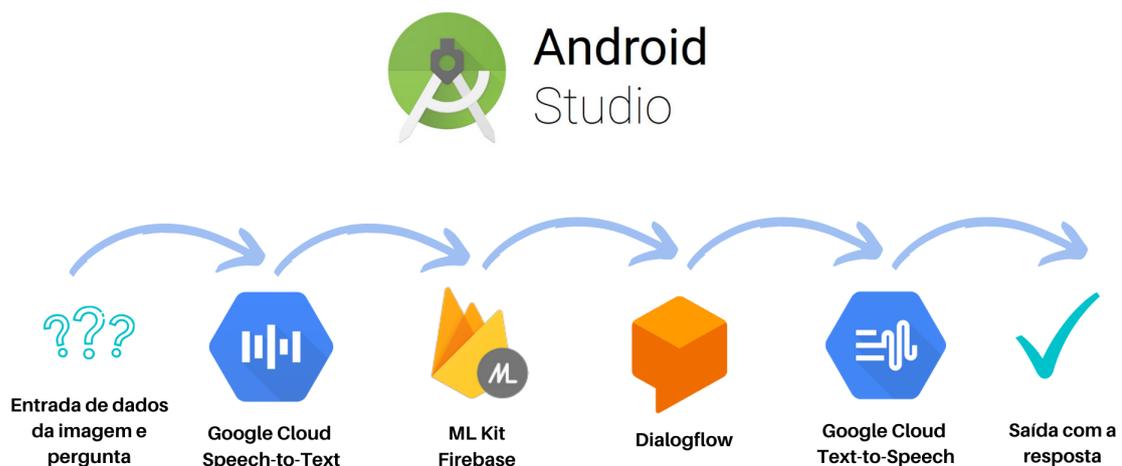
transparentes, como copos sobre a mesa; a escolha de tamanho de roupas e calçados em lojas; a identificação de cores; a identificação de objetos pequenos, como agulha, alfinete e tarrachas de brincos; a diferenciação de capas de livros e CDs; a identificação de cartelas e caixas de medicamentos; e a percepção a respeito de insetos, para conseguir diferenciar, por exemplo, a presença de uma abelha ou de uma mosca, plantas e chás.

Quando apresentada a proposta do protótipo KATIE, a reação foi extremamente positiva. O diferencial de poder realizar perguntas para obter respostas mais precisas e específicas foi um fator que despertou interesse nos entrevistados. É relevante ressaltar que um deles utiliza sistema operacional iOS, ao passo que os outros dois utilizam Android.

## 5.2 Escolha das tecnologias

Mais de 70% dos celulares vendidos em escala global são modelos com o sistema operacional Android. (STATCOUNTER GLOBAL STATS, 2020). Por essa razão, o protótipo deste trabalho foi desenvolvido para o sistema operacional Android. As ferramentas utilizadas para o desenvolvimento foram as oficiais para a plataforma, que são o Android Studio, Google Cloud, Firebase e Dialogflow, todos desenvolvidos pela Google e com uma boa capacidade de integração. Na Figura 4 o fluxo de funcionamento entre elas é demonstrado e, nos próximos tópicos, cada uma dessas tecnologias é detalhada.

Figura 4 – Fluxo com as tecnologias escolhidas



Fonte: Elaborado pela autora.

### 5.2.1 Android Studio

Apesar de existirem diversas ferramentas para desenvolvimento de aplicativos móveis, o Android Studio, lançado em 2013, é o ambiente oficial e criado especialmente para o desenvolvimento integrado de aplicativos no sistema operacional Android. (ANDROID DEVELOPERS, 2020). Em conjunto com as APIs do Android, tem sua programação em Java ou Kotlin (que foi agregado em 2017), além elementos visuais em XML. Para este trabalho foi escolhida a linguagem de programação Java devido aos conhecimentos prévios da autora e pela maior disponibilidade e suporte de recursos. O Android Studio ainda possui uma fácil integração com Firebase e Google Cloud, ferramentas que são mencionadas a seguir e que também foram integradas no desenvolvimento do modelo.

### 5.2.2 Google Cloud Speech-to-Text e Text-to-Speech

A ferramenta *Speech-to-Text* é uma API poderosa e de fácil utilização, que possibilita a tradução de áudio para texto usando a tecnologia de inteligência artificial. (GOOGLE, 2020a). É uma solução flexível para criar experiências naturais em diversos casos de uso. Para esse trabalho é utilizada a conversão da fala do usuário para texto com o objetivo de reconhecer a pergunta realizada. A tecnologia *Text-to-Speech* tem uma finalidade muito semelhante, porém, com um fluxo inverso ao converter texto em áudio e é utilizada para realizar uma resposta final ao usuário, respondendo a pergunta inicial. (GOOGLE, 2020b).

### 5.2.3 ML Kit Firebase

Composto por um eficiente conjunto de ferramentas de desenvolvimento de software, o Kit de ML do Firebase proporciona uma experiência com *machine learning* do Google para aplicativos Android e também iOS. (FIREBASE, 2020). Esse conjunto é formado por diversas APIs de aplicações no uso comum de dispositivos móveis, como leitura de códigos de barras, detecção facial, reconhecimento de imagens e identificação de textos em imagens, como é o caso da aplicação que este trabalho englobará.

### 5.2.4 Dialogflow

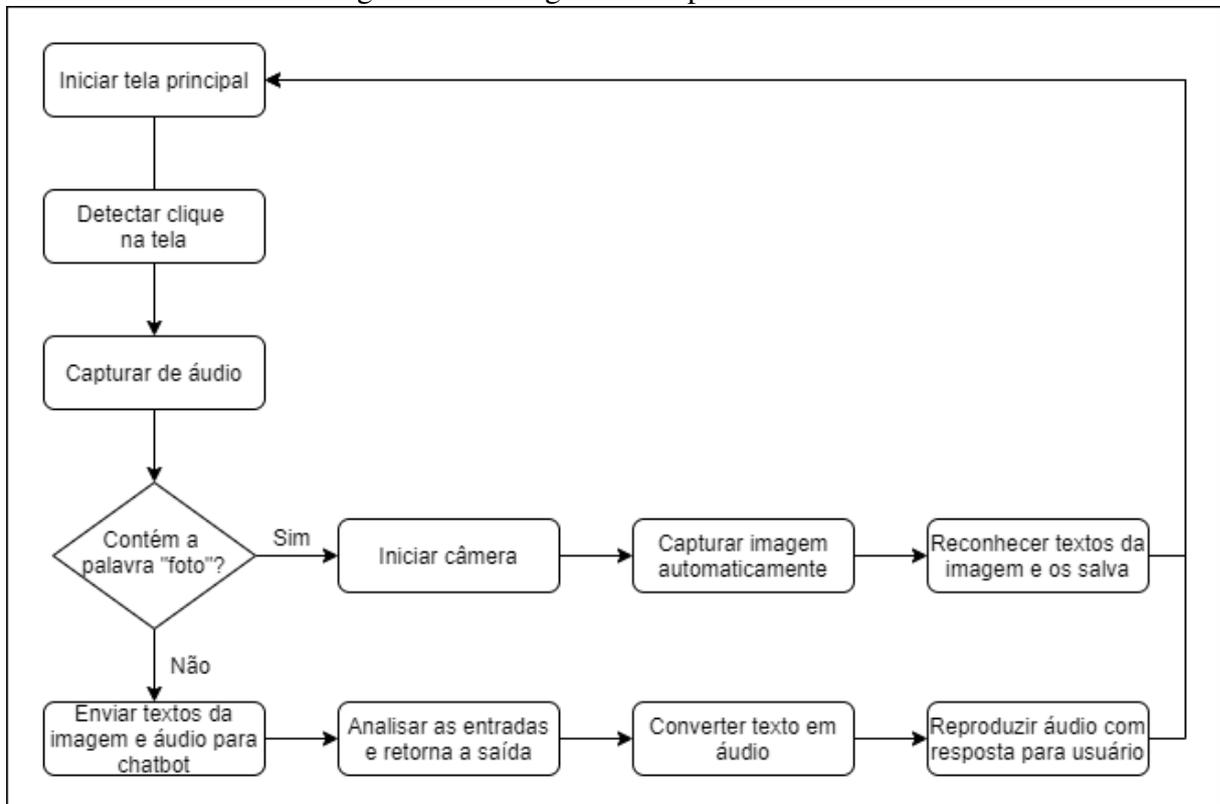
Ser uma plataforma de processamento de linguagem natural para facilitar a integração de usuários com dispositivos é a proposta dessa ferramenta. (DIALOGFLOW, 2020). É possível analisar diversos tipos de entrada, como texto ou áudio, assim como a saída pode ser por meio de um texto ou fala sintética. Para o caso dessa aplicação, o Dialogflow é utilizado para receber e analisar a pergunta do usuário e o texto captado da imagem, para então devolver a resposta adequada, em forma de áudio.

### 5.3 Protótipo

Com as tecnologias definidas, um escopo para o protótipo também foi delimitado. Analisando as conversas com o público alvo e as dificuldades citadas por eles, optou-se por focar este protótipo na identificação de medicamentos, por meio de cartelas, caixas e outras embalagens que contenham textos, concentrando no reconhecimento dos textos presentes nesse material. Com isso, o desenvolvimento do aplicativo foi iniciado e segue o fluxo descrito na sequência.

Ao abrir o aplicativo, uma mensagem de áudio dá as boas-vindas e introduz o funcionamento do aplicativo. Na sequência, ao clique do usuário em qualquer espaço da tela, um pop-up de captação de áudio se abre e capta o que vier a ser dito. O sistema então valida se a palavra "foto" está presente no texto do áudio. Caso positivo, a câmera do dispositivo é iniciada e a captura é feita automaticamente. No caso contrário, o sistema considera o áudio como uma pergunta sobre a imagem já captada, então salva e segue para a próxima etapa, na qual o texto da imagem é reconhecido. Os dois textos são enviados a um *chatbot* responsável por fazer a análise do conteúdo e retornar a resposta, a qual será convertida em um áudio e reproduzida ao usuário. Na Figura 5 é possível visualizar o fluxo do sistema e a maneira como os módulos se integram. Na sequência, cada um dos blocos é detalhado.

Figura 5 – Fluxograma do aplicativo KATIE



Fonte: Elaborado pela autora.

### 5.3.1 Tela principal

Em função da demanda do público-alvo, o aplicativo possui apenas uma tela, na qual todos os processos acontecem. Na criação da tela, as permissões para uso da câmera e áudio são solicitadas e todos os componentes são instanciados. Uma mensagem de áudio introdutória dá as boas-vindas ao usuário e solicita que toque na tela e pede para o dispositivo registrar uma foto. A tela fica no aguardo desse toque para então ativar o reconhecimento de fala.

### 5.3.2 Captura de áudio

Nesse módulo, a biblioteca *Speech Recognizer* nativa do Android, configurada para detectar no idioma português brasileiro, é utilizada, criando um *pop-up* na tela e iniciando a captura do áudio (fala do usuário). Com a conclusão dessa funcionalidade, o áudio é convertido para texto e salvo para ser utilizado na sequência. Se o texto contiver a palavra "foto", a função de inicialização da câmera é iniciada, caso contrário, a funcionalidade de enviar mensagem ao *chatbot* é acionada.

### 5.3.3 Captura de imagem

A captura automática da imagem foi um grande desafio para o desenvolvimento do protótipo, visto que a biblioteca de câmera nativa do Android não realiza capturas automaticamente e ainda há um outro obstáculo, pois ela solicita uma confirmação sobre a foto captada. Com isso, optou-se por utilizar uma outra biblioteca, realizando adaptações, denominada EZCam. Essa funcionalidade inicia com a abertura da câmera. Em seguida, é definido um intervalo de 1.5 segundo, importante para que a velocidade de ISO da câmera aumente para garantir que a imagem não fique muito escura. Por fim, a captura é realizada. Após a captura, a imagem é incorporada à tela do aplicativo e inicia a função para o reconhecimento dos textos presentes na imagem.

### 5.3.4 Reconhecimento de textos da imagem

Para essa etapa, foi utilizada a tecnologia ML Kit Firebase, em específico a biblioteca *Text Recognizer*. Assim que a função é chamada, a imagem da tela é analisada e são localizados blocos de textos. Caso nenhum bloco seja encontrado, uma mensagem é retornada ao usuário. Com os blocos de textos localizados, cada um deles é separado em linhas e, na sequência, em palavras. Com isso, todo o texto encontrado na imagem é salvo em uma variável.

### 5.3.5 Processamento dos dados

Com as informações de pergunta e imagem, os textos são encaminhados ao Dialogflow, onde o processamento desses dados é realizado. De acordo com essa entrada de dados, a saída é enviada de volta. Para o caso de uma dúvida sobre qual medicamento encontra-se diante do usuário, há a pergunta e a imagem (na qual consta o título do fármaco). Ao enviar esses dados, o *chatbot* retorna uma frase com a resposta, indicando o nome. O mesmo pode acontecer para outras perguntas, como o tempo de intervalo entre um medicamento e outro ou qual a sua aplicação. O Dialogflow tem uma inteligência artificial que faz com que o aplicativo consiga identificar que perguntas com pequenas variações tem a mesma intenção, como é o caso das perguntas "Que remédio é esse aqui?" e "Qual o nome desse medicamento?". Como complemento, também há uma funcionalidade para informar sinônimos de palavras escritas, a qual foi utilizada preenchendo nomes derivados dos fármacos. É possível aplicar essas funcionalidades em outros objetos, como é o caso do reconhecimento de cédulas de dinheiro.

### 5.3.6 Resposta ao usuário

Como última etapa, ocorre a resposta ao usuário, que é elaborada por meio da entrada dos dados de imagem e da pergunta. Por meio de dados já cadastrados no chatbot, é feita a identificação da intenção da pergunta realizada. Com isso, o *chatbot* consegue realizar uma saída com a resposta adequada, que é reproduzida em um áudio para o usuário. Por fim, o sistema pode voltar ao começo do ciclo, aguardando um toque na tela para o próximo passo.

## 6 TESTES E RESULTADOS OBTIDOS

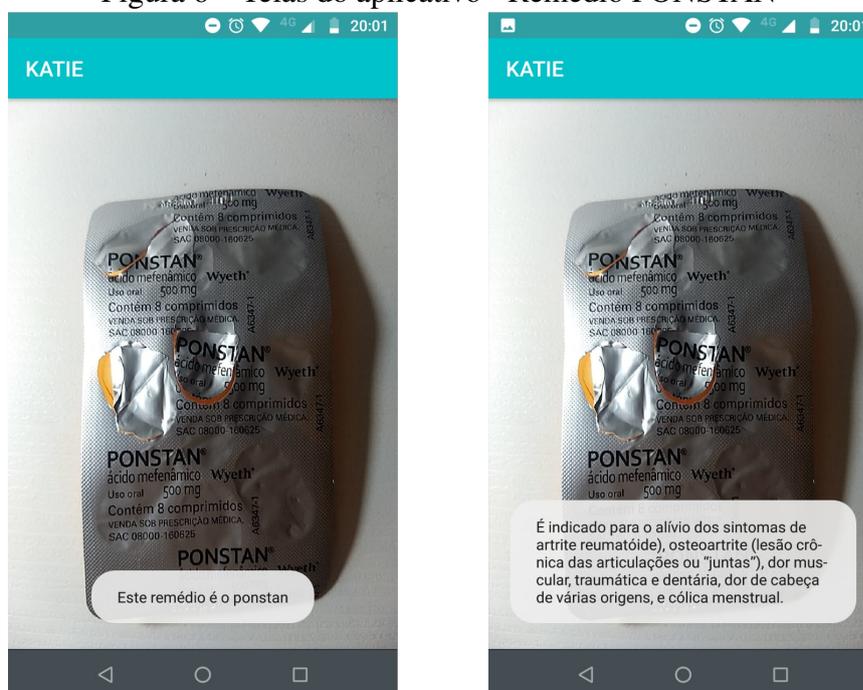
A partir do protótipo apresentado do aplicativo KATIE, buscou-se avaliar se o modelo proposto soluciona o problema apresentado. A intenção inicial era de realizar testes com possíveis usuários, como pessoas cegas ou de baixa visão. Porém, em função da pandemia em curso no período que compreende a realização deste trabalho, os planos precisaram ser modificados. Com isso, os testes foram realizados apenas pela desenvolvedora do aplicativo, que também é a autora deste trabalho.

Para a realização destes testes, foi utilizado um celular modelo Motorola Moto G5 Plus com o aplicativo KATIE instalado. Além de materiais de amostra para as fotos, como cartelas, frascos e caixas de medicamentos e também notas de dinheiro. O objetivo era avaliar o desempenho do aplicativo KATIE em diversos cenários, com variações de luminosidade, contraste, plano de fundo, fontes e tamanhos de letras distintas. Os resultados são apresentados em quatro tópicos: identificação do remédio PONSTAN, identificação de tipos variados de embalagens de fármacos, reconhecimento de notas de dinheiro e por fim, as operações em que houve falhas.

## 6.1 Identificação remédio PONSTAN

Na Figura 6 é possível ver o resultado de um dos testes, onde foi utilizada uma cartela do remédio PONSTAN. A captura do áudio armazenou as perguntas corretamente, que foram "Que remédio é esse?" para a primeira captura da tela, e "Para que serve?" para a segunda. A análise dos textos da imagem foi bem sucedida, assim como o entendimento da intenção de cada pergunta, e portanto, obteve-se o resultado esperado do *chatbot*, de acordo com o que está sendo exibido nas telas apresentadas na Figura 6 (as respostas foram reproduzidas em áudio).

Figura 6 – Telas do aplicativo - Remédio PONSTAN

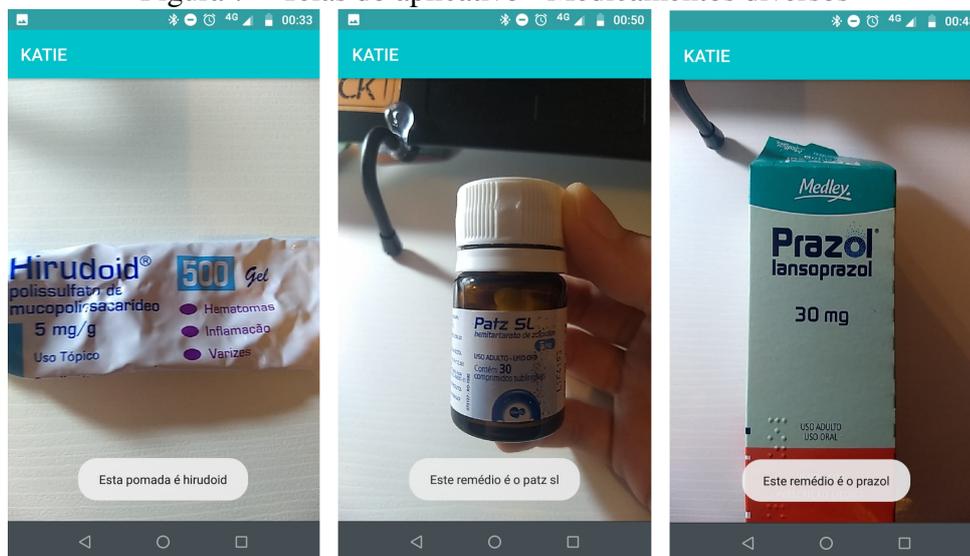


Fonte: Elaborado pela autora.

## 6.2 Identificação de remédios com embalagens variadas

Seguindo o mesmo procedimento de uso do aplicativo, na Figura 7 são apresentadas capturas de três telas. Na primeira, é identificada uma pomada (vale ressaltar as ranhuras do produto que poderiam ter dificultado a leitura); na segunda há um frasco de remédio; e por fim, na terceira imagem uma caixa de remédio. A pergunta realizada para todos os casos foi a mesma "Qual remédio é?" e as respostas foram compatíveis com o esperado.

Figura 7 – Telas do aplicativo - Medicamentos diversos

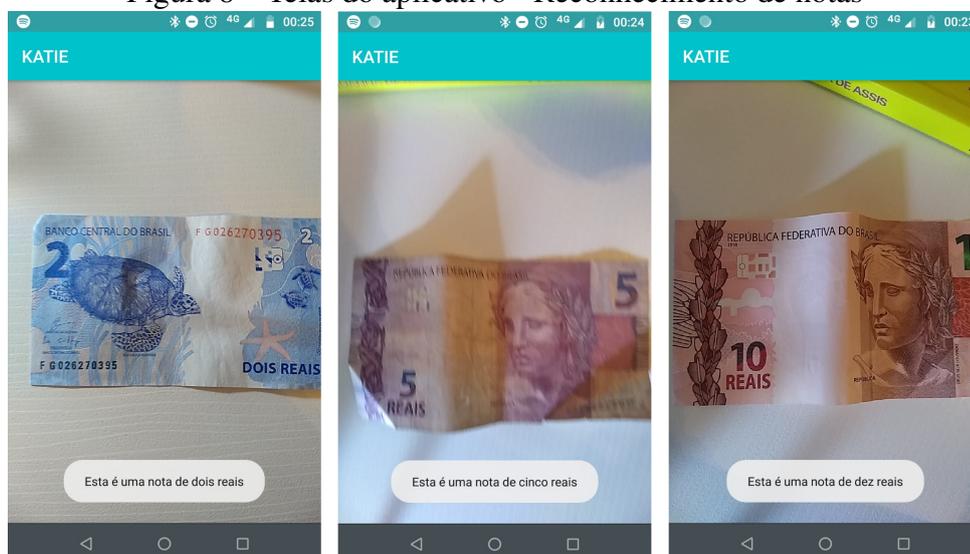


Fonte: Elaborado pela autora.

### 6.3 Reconhecimento de notas

Com a finalidade de demonstrar a capacidade do aplicativo KATIE de ir além do reconhecimento de fármacos, foram realizados testes com notas de dinheiro. Na Figura 8 é notável que o reconhecimento do valor das cédulas é compatível com a foto registrada. Para os três casos demonstrados, a pergunta realizada foi a mesma, "Que nota é essa?".

Figura 8 – Telas do aplicativo - Reconhecimento de notas

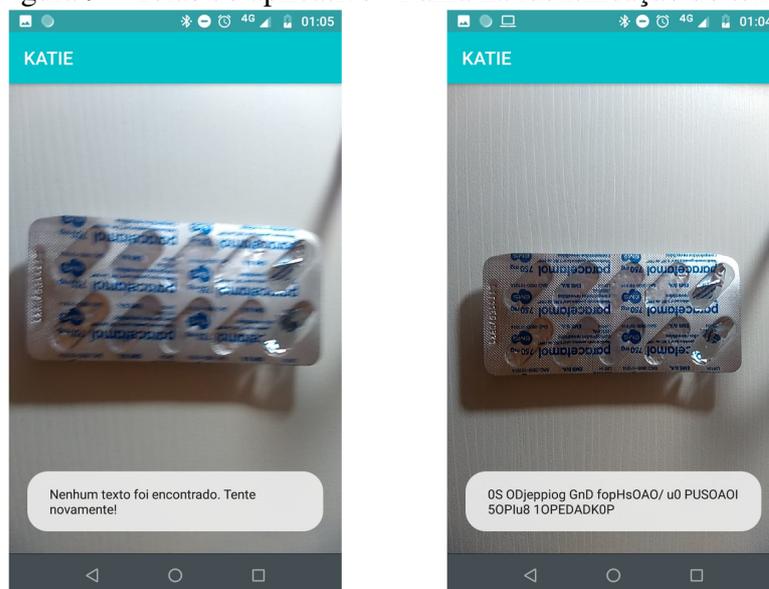


Fonte: Elaborado pela autora.

## 6.4 Aplicações mal sucedidas

Para este último tópico dos resultados, foram agrupados dois exemplos de casos onde o aplicativo apresentou falha e não atingiu a finalidade esperada. Na Figura 9 estão presentes duas capturas de tela, na primeira a foto capturada está tremida o que dificultou o reconhecimento do texto da imagem; já na segunda nota-se que a cartela de remédio está virada, o que impossibilitou uma leitura adequada das informações, o texto em destaque exibe o que o aplicativo reconheceu na imagem. Esses dois casos são alguns exemplos do que pode acontecer quando o aplicativo for utilizado por seu público alvo, pessoas com baixa visão ou cegas, visto que elas não tem a possibilidade de ajustar o objeto na câmera.

Figura 9 – Telas do aplicativo - Falha na identificação do texto



Fonte: Elaborado pela autora.

## 7 CONCLUSÕES E TRABALHOS FUTUROS

Por meio de uma contextualização a respeito do tema de acessibilidade, é possível afirmar que o objetivo deste trabalho foi alcançado, uma vez que foi possível desenvolver uma solução que reconheça a pergunta do usuário, analise as características do ambiente, mas especificamente o objeto, pela câmera e, após uma comparação das informações de imagem e áudio, resulte em uma resposta. A concretização do objetivo partiu de uma contextualização teórica, na qual foram apresentados os conceitos sobre tecnologias assistivas, visão computacional e processamento de linguagem natural. Além disso, a construção de um modelo foi precedida pela

busca por trabalhos relacionados, os quais foram estudados, descritos, caracterizados e comparados entre si. Na sequência, apresentou-se a construção desse modelo, denominado KATIE, os resultados com o protótipo foram compatíveis com o esperado, sendo possível identificar a pergunta e imagem, retornando uma resposta ao usuário.

Em aspectos teóricos, o modelo apresenta-se válido e há uma real necessidade do desenvolvimento de tecnologias assistivas num âmbito geral. Quanto à implementação e testes realizados em relação ao protótipo, pode-se dizer que todos os módulos da implementação foram desenvolvidos e os testes apresentaram os resultados esperados. Sendo possível a captura do áudio, captura automática da imagem, análise da presença de textos na imagem e retorno do *chatbot*.

Entretanto, ainda existem trabalhos futuros há serem elaborados: a ampliação da base de medicamentos cadastrados, a aplicação do modelo em outros cenários, como reconhecimento de notas e moedas, identificação de cores, leitura de embalagens de produtos, entre outras tantas aplicações. Outros testes mais elaborados devem ser realizados, principalmente com pessoas acometidas por cegueira ou baixa visão, com a finalidade de avaliar a viabilidade de uso do protótipo e modelo.

## REFERÊNCIAS

ANDROID DEVELOPERS. **Android Studio**. 2020. Disponível em: <<https://developer.android.com/studio/>>.

ANTOL, S. et al. Vqa: Visual question answering. In: **Proceedings of the IEEE international conference on computer vision**. Washington, DC, USA: The IEEE International Conference on Computer Vision (ICCV), 2015. p. 2425–2433.

ARORA, A. et al. Real time multi object detection for blind using single shot multibox detector. **Wireless Personal Communications**, Springer, p. 1–11, 2019.

BE MY EYES. **Bringing sight to blind and low-vision people**. 2015. Disponível em: <<https://www.bemyeyes.com/>>.

BERSCH, R. Introdução à tecnologia assistiva. **Porto Alegre: CEDI**, v. 21, 2008.

BRASIL. **Lei Brasileira de Inclusão da Pessoa com Deficiência**. 2015. Disponível em: <[http://www.planalto.gov.br/ccivil/\\_03/\\_ato2015-2018/2015/lei/113146.htm](http://www.planalto.gov.br/ccivil/_03/_ato2015-2018/2015/lei/113146.htm)>.

CHAUDHRY, S.; CHANDRA, R. Design of a mobile face recognition system for visually impaired persons. **arXiv preprint arXiv:1502.00756**, 2015.

CONSELHO BRASILEIRO DE OFTALMOLOGIA. **As Condições de Saúde Ocular no Brasil**. 2019. Disponível em: <[http://www.cbo.com.br/novo/publicacoes/condicoes\\_saude\\_ocular\\_brasil2019.pdf](http://www.cbo.com.br/novo/publicacoes/condicoes_saude_ocular_brasil2019.pdf)>.

COOK, A. M.; HUSSEY, S. Assistive technologies. **St. Louis: Mosby**, Elsevier, 2002.

DIALOGFLOW. **Documentação do Dialogflow**. 2020. Disponível em: <<https://cloud.google.com/dialogflow/docs/>>.

- FIREBASE. **Kit de ML para Firebase**. 2020. Disponível em: <<https://firebase.google.com/docs/ml-kit?hl=pt-br>>.
- GOOGLE. **Use Lookout to explore your surroundings**. 2019. Disponível em: <<https://support.google.com/accessibility/android/answer/9031274>>.
- GOOGLE. **Cloud Speech-to-Text**. 2020. Disponível em: <<https://cloud.google.com/speech-to-text>>.
- GOOGLE. **Cloud Text-to-Speech**. 2020. Disponível em: <<https://cloud.google.com/text-to-speech>>.
- GOVERNO DIGITAL. **Acessibilidade**. 2014. Disponível em: <<https://www.governodigital.gov.br/acessibilidade>>.
- GURARI, D. et al. Vizwiz grand challenge: Answering visual questions from blind people. In: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition**. Salt Lake City: Computer Vision Fundation, 2018. p. 3608–3617.
- HAYO. **What is Computer Vision?** 2017. Disponível em: <<https://hayo.io/computer-vision/>>.
- MICROSOFT. **Seeing AI**. 2017. Disponível em: <<https://www.microsoft.com/en-us/ai/seeing-ai>>.
- ONU BRASIL. **Agenda 2030**. 2015. Disponível em: <<https://nacoesunidas.org/pos2015/agenda2030/>>.
- SOCIEDADE BRASILEIRA DE COMPUTAÇÃO. **Processamento de Linguagem Natural**. 2019. Disponível em: <<https://www.sbc.org.br/14-comissoes/394-processamento-de-linguagem-natural>>.
- SOUSA, K.; MARENGONI, M. Uso de visão computacional em dispositivos moveis para o reconhecimento de faixa de pedestres. In: **Anais do VIII Workshop de Visão Computacional**. Brasil: Universidade de São Paulo, 2012. p. 1–85.
- STATCOUNTER GLOBAL STATS. **Mobile Operating System Market Share Worldwide**. 2020. Disponível em: <<https://gs.statcounter.com/os-market-share/mobile/worldwide>>.
- WORLD HEALTH ORGANIZATION. **World report on vision**. 2019. Disponível em: <<https://www.who.int/publications-detail/world-report-on-vision>>.