UNIVERSIDADE FEDERAL DE GOIÁS (UFG)

INSTITUTO DE INFORMÁTICA (INF)

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO (PPGCC)

ANDRÉ PIRES CORRÊA

Análise de um Fluxo Completo Automatizado de Etapas Voltado ao Reconhecimento de Texto em Imagens de Prescrições Médicas Manuscritas

Goiânia
2024

UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE INFORMÁTICA

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO (TECA) PARA DISPONIBILIZAR VERSÕES ELETRÔNICAS DE TESES E DISSERTAÇÕES NA BIBLIOTECA DIGITAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a Lei 9.610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou download, a título de divulgação da produção científica brasileira, a partir desta data.

O conteúdo das Teses e Dissertações disponibilizado na BDTD/UFG é de responsabilidade exclusiva do autor. Ao encaminhar o produto final, o autor(a) e o(a) orientador(a) firmam o compromisso de que o trabalho não contém nenhuma violação de quaisquer direitos autorais ou outro direito de terceiros.

1. Identificação do material bibliográfico

[X] Dissertação [ ] Tese [ ] Outro*:_____

*No caso de mestrado/doutorado profissional, indique o formato do Trabalho de Conclusão de Curso, permitido no documento de área, correspondente ao programa de pós-graduação, orientado pela legislação vigente da CAPES.

Exemplos: Estudo de caso ou Revisão sistemática ou outros formatos.

2. Nome completo do autor

André Pires Corrêa

3. Título do trabalho

Análise de um Fluxo Completo Automatizado de Etapas Voltado ao Reconhecimento de Texto em Imagens de Prescrições Médicas Manuscritas

4. Informações de acesso ao documento (este campo deve ser preenchido pelo orientador)

Concorda com a liberação total do documento [X] SIM [ ] NÃO[1]

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. Após esse período, a possível disponibilização ocorrerá apenas mediante:
a) consulta ao(à) autor(a) e ao(à) orientador(a);
b) novo Termo de Ciência e de Autorização (TECA) assinado e inserido no arquivo da tese ou dissertação.
O documento não será disponibilizado durante o período de embargo.
Casos de embargo:
- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro;
- Publicação da dissertação/tese em livro.

Obs. Este termo deverá ser assinado no SEI pelo orientador e pelo autor.

Documento assinado eletronicamente por **Hugo Alexandre Dantas Do Nascimento**, **Professor do Magistério Superior**, em 16/02/2024, às 19:01, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **André Pires Corrêa**, **Discente**, em 16/02/2024, às 19:06, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **4384013** e o código CRC **6B5AE2E6**.

ANDRÉ PIRES CORRÊA

Análise de um Fluxo Completo Automatizado de Etapas Voltado ao Reconhecimento de Texto em Imagens de Prescrições Médicas Manuscritas

Dissertação apresentada ao Programa de Pós–Graduação em Ciência da Computação do Instituto de Informática (INF) da Universidade Federal de Goiás (UFG), como requisito parcial para obtenção do título de Mestre em Ciência da Computação.

Área de concentração: Ciência da Computação.
Linha de Pesquisa: Sistemas Inteligentes e Aplicações.
Orientador: Prof. Hugo Alexandre Dantas do Nascimento

Co-Orientador: Prof. Eliomar Araújo de Lima

Goiânia
2024

Ficha de identificação da obra elaborada pelo autor, através do
Programa de Geração Automática do Sistema de Bibliotecas da UFG.

Corrêa, André Pires
Análise de um Fluxo Completo Automatizado de Etapas Voltado ao Reconhecimento de Texto em Imagens de Prescrições Médicas Manuscritas [manuscrito] / André Pires Corrêa. - 2024.
78 f.: il.

Orientador: Prof. Dr. Hugo Alexandre Dantas do Nascimento; co orientador Dr. Eliomar Araújo de Lima.
Dissertação (Mestrado) - Universidade Federal de Goiás, Instituto de Informática (INF), Programa de Pós-Graduação em Ciência da Computação, Goiânia, 2024.
Bibliografia. Anexos. Apêndice.
Inclui tabelas, algoritmos, lista de figuras, lista de tabelas.

1. Prescrições médicas. 2. Reconhecimento de texto manuscrito. 3. Aprendizado de máquina. I. do Nascimento, Hugo Alexandre Dantas, orient. II. Título.

CDU 004

UNIVERSIDADE FEDERAL DE GOIÁS

INSTITUTO DE INFORMÁTICA

ATA DE DEFESA DE DISSERTAÇÃO

Ata nº **01/2024** da sessão de Defesa de Dissertação de **André Pires Corrêa**, que confere o título de Mestre em **Ciência da Computação**, na área de concentração em Ciência da Computação.

Aos dez dias do mês de janeiro de dois mil e vinte e quatro, a partir das oito horas e trinta minutos, via sistema de webconferência da RNP, realizou-se a sessão pública de Defesa de Dissertação intitulada **"Sistema para Extração Automatizada de Dados de Prescrições Médicas"**. Os trabalhos foram instalados pelo Orientador, Professor Doutor Hugo Alexandre Dantas do Nascimento (INF/UFG) com a participação dos demais membros da Banca Examinadora: Professor Doutor Eliomar Araújo de Lima (INF/UFG), Coorientador; Professor Doutor Hélio Pedrini (IC/Unicamp), membro titular externo; e Professor Doutor Ronaldo Martins da Costa (INF/UFG), membro titular interno. A realização da banca ocorreu por meio de videoconferência. Durante a arguição os membros da banca fizeram sugestão de alteração do título do trabalho. A Banca Examinadora reuniu-se em sessão secreta a fim de concluir o julgamento da Dissertação, tendo sido o candidato **aprovado** pelos seus membros. Proclamados os resultados pelo Professor Doutor Hugo Alexandre Dantas do Nascimento, Presidente da Banca Examinadora, foram encerrados os trabalhos e, para constar, lavrou-se a presente ata que é assinada pelos Membros da Banca Examinadora, aos dez dias do mês de janeiro de dois mil e vinte e quatro.

TÍTULO SUGERIDO PELA BANCA

Análise de um Fluxo Completo Automatizado de Etapas Voltado ao Reconhecimento de Texto em Imagens de Prescrições Médicas Manuscritas

Documento assinado eletronicamente por **Eliomar Araujo De Lima**, **Professor do Magistério Superior**, em 10/01/2024, às 12:13, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Hugo Alexandre Dantas Do Nascimento**, **Professor do Magistério Superior**, em 10/01/2024, às 12:13, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Ronaldo Martins Da Costa**, **Professor do Magistério Superior**, em 10/01/2024, às 12:13, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Helio Pedrini**, **Usuário Externo**, em 10/01/2024, às 12:13, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **André Pires Corrêa**, **Discente**, em 10/01/2024, às 12:21, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **4290229** e o código CRC **9488CCAE**.

Agradecimentos

Gostaria de agradecer minha mãe e meu irmão por sempre estarem ao meu lado e me apoiarem ao longo de toda minha vida.

Agradeço também aos professores Hugo Alexandre Dantas do Nascimento e Eliomar Araújo de Lima por me orientarem ao longo do projeto de mestrado e me ajudarem a superar as diversas turbulências por quais passei durante esses últimos dois anos.

Agradeço aos meus amigos do curso de Ciência da Computação Gabriel e Marcos, por me darem o empurrão que eu precisava para seguir em frente com o mestrado acadêmico.

Agradeço ao Luiz Carlos Batistel Vieira, membro da equipe deste projeto de pesquisa, por suas diversas contribuições ao longo do projeto.

Agradeço ao Instituto Gyntec de Inovação e a Farmácia Artesanal por acreditarem no projeto e oferecerem o suporte que foi necessário para o seu desenvolvimento.

Por fim, agradeço ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo apoio financeiro, e ao Laboratório Multiusuário de Computação de Alto Desempenho da Universidade Federal de Goiás (LaMCAD/UFG) pelos recursos computacionais que possibilitaram o projeto.

Resumo

Corrêa, André Pires. **Análise de um Fluxo Completo Automatizado de Etapas Voltado ao Reconhecimento de Texto em Imagens de Prescrições Médicas Manuscritas**. Goiânia, 2024. 78p. Dissertação de Mestrado. Programa de Pós-Graduação em Ciência da Computação (PPGCC), Instituto de Informática (INF), Universidade Federal de Goiás (UFG).

Farmácias de manipulação lidam diariamente com grandes volumes de prescrições médicas, cujos dados precisam ser manualmente introduzidos em sistemas de gestão de informação para o acompanhamento dos pedidos de seus clientes. Uma parcela considerável dessas prescrições tende a ser escrita por médicos com caligrafia de baixa legibilidade, o que contribui para tornar a decodificação árdua e demorada. Trabalhos anteriores investigaram o uso de aprendizado de máquina para reconhecimento de prescrições médicas. No entanto, as taxas de acerto nesses trabalhos ainda são muito baixas e as abordagens empregadas comumente são limitadas, pois envolveram uma quantidade pequena de casos, focaram em poucas etapas do processo de automação da análise de prescrições médicas ou utilizaram ferramentas proprietárias, o que dificulta a replicação e a análise científica dos resultados. O presente trabalho contribui para preencher essa lacuna, apresentando um processo completo de segmentação, reconhecimento e processamento de prescrições médicas que foi construído com base em uma avaliação e adaptação de múltiplos métodos existentes na literatura para cada etapa do processo. Os métodos foram avaliados usando uma base de 993 imagens de prescrições médicas com 27.933 palavras anotadas, produzida com o apoio de uma farmácia de manipulação parceira do projeto. Os resultados obtidos pelos melhores métodos indicam que a abordagem é razoavelmente eficaz, atingindo uma acurácia de 68% no processo de segmentação, e uma taxa de acurácia de caractere de 86,8% no processo de reconhecimento de texto.

Palavras–chave

Prescrições médicas, reconhecimento de texto manuscrito, aprendizado de máquina.

Abstract

Corrêa, André Pires. **Analysis of an automated pipeline for text recognition in images of handwritten medical prescriptions**. Goiânia, 2024. 78p. MSc. Dissertation. Programa de Pós-Graduação em Ciência da Computação (PPGCC), Instituto de Informática (INF), Universidade Federal de Goiás (UFG).

Compounding pharmacies deal with large volumes of medical prescriptions on a daily basis, whose data needs to be manually inputted into information management systems to properly process their customers' orders. A considerable portion of these prescriptions tend to be written by doctors with poorly legible handwriting, which can make decoding them an arduous and time-consuming process. Previous works have investigated the use of machine learning for medical prescription recognition. However, the accuracy rates in these works are still fairly low and their approaches tend to be rather limited, as they typically utilize small datasets, focus only on specific steps of the automated analysis pipeline or use proprietary tools, which makes it difficult to replicate and analyse their results. The present work contributes towards filling this gap by presenting an end-to-end process for automated data extraction from handwritten medical prescriptions, from text segmentation, to recognition and post-processing. The approach was built based on an evaluation and adaptation of multiple existing methods for each step of the pipeline. The methods were evaluated on a dataset of 993 images of medical prescriptions with 27,933 annotated words, produced with the support of a compounding pharmacy that participated in the project. The results obtained by the best performing methods indicate that the developed approach is reasonably effective, reaching an accuracy of 68% in the segmentation step, and a character accuracy rate of 86.8% in the text recognition step.

Keywords

Medical prescriptions, handwritten text recognition, machine learning

Sumário

Lista de Figuras

Lista de Tabelas

CAPÍTULO 1

Introdução

Farmácias de manipulação trabalham diariamente com grandes volumes de pedidos de medicações a serem manipuladas. Para facilitar o gerenciamento destes pedidos, é comum a utilização de sistemas computacionais de gestão de informação, nos quais os dados das prescrições médicas de cada pedido são registrados e utilizados para guiar o processo de manipulação até a entrega do medicamento ao cliente. Essa tarefa manual de introduzir dados nos sistemas requer grande tempo e esforço por parte dos farmacêuticos pois necessita que seja feita a interpretação de prescrições médicas, sendo que uma considerável parcela delas costuma ser escrita a mão por médicos com caligrafia de baixa legibilidade. Além disso, as prescrições devem passar por uma etapa de verificação a fim de garantir que não existam irregularidades nos compostos ou nas doses das fórmulas prescritas, algo que também requer tempo e conhecimento altamente especializado.

Uma possível solução para tornar esse processo mais eficiente é a adoção de métodos automatizados de extração de dados a partir de fotos ou imagens escaneadas de prescrições médicas. Em um levantamento bibliográfico realizado como parte deste projeto, foram identificados múltiplos trabalhos que abordam tal problema (NAJAFI-RAGHEB; HATAM; HARIFI, 2017; CHUMUANG; KETCHAM, 2018; FAJARDO et al., 2019; BUTALA et al., 2020; KULATHUNGA et al., 2020; HASSAN et al., 2021; GUPTA; SOENY, 2021). No entanto, as taxas de acerto dos métodos utilizados nesses trabalhos ainda são relativamente baixas para prescrições manuscritas (na faixa de 30% a 80%) e, em geral, eles foram testados com uma quantidade pequena de casos (entre algumas dezenas a poucos milhares). Outro aspecto observado é que os trabalhos frequentemente focam em poucas etapas do processo de automação da análise de prescrições médicas; em alguns casos, eles utilizam ferramentas proprietárias, o que dificulta a replicação e a análise científica dos resultados.

Como exemplo de trabalho nessa área, Hassan et al. (2021) apresentam um sistema que extrai os nomes de medicamentos a partir de fotos de prescrições médicas tiradas por *smartphones*. O sistema aplica uma Rede Neural Convolucional para realizar o reconhecimento de texto manuscrito, e obteve uma acurácia de teste de 50% quando

avaliado em um conjunto de dados construído pelos autores, cujo tamanho não foi especificado. Em outro exemplo, Gupta e Soeny (2021) descrevem uma abordagem para detectar o nome de medicações em imagens de prescrições médicas impressas e manuscritas com base no uso da API proprietária de visão computacional da Google, denominada Google Vision. A abordagem foi avaliada usando um conjunto de 5.000 imagens de prescrições manuscritas, obtendo um F-Score de 26,2%.

O presente trabalho apresenta uma abordagem computacional desenvolvida para o reconhecimento e o processamento de prescrições médicas manuscritas para farmácias de manipulação. Ela vem a contribuir para preencher uma lacuna na área por tratar de um fluxo com etapas de processamento que, apesar de similares àquelas existentes em outros contextos de aplicação, possuem características particulares. A abordagem foi gerada dentro do escopo do Programa de Mestrado e Doutorado Acadêmico para Inovação (MAI/DAI) do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) em parceria com o Instituto Gyntec de Inovação e com uma farmácia de manipulação. Por ter o envolvimento da farmácia, interessada no desenvolvimento do projeto, os pesquisadores tiveram acesso a uma grande base de prescrições médicas digitalizadas e a farmacêuticos experientes nas suas interpretações. Esses recursos foram fundamentais para a condução da pesquisa.

1.1 Objetivos

O objetivo principal deste trabalho é apresentar uma abordagem que realize um processo completo de extração e processamento automatizado de informações a partir de imagens de prescrições médicas no contexto de uma farmácia de manipulação. São objetivos específicos:

- identificar etapas de processamento necessárias para a abordagem;
- apresentar, por meio de estudo comparativo, métodos que sejam efetivos para cada uma das etapas de processamento; e
- apresentar uma ferramenta computacional protótipo para anotação e reconhecimento de texto manuscrito em imagens de prescrições médicas.

1.2 Metodologia

Visando atingir os objetivos supracitados, foi realizada uma revisão bibliográfica para identificar etapas e métodos comumente aplicados no processamento de prescrições médicas manuscritas. Com base nos trabalhos revisados, foi estabelecido um fluxo de processamento de prescrições médicas em etapas.

Realizou-se então uma análise comparativa de soluções computacionais existentes na literatura para cada etapa do fluxo, com o intuito de escolher os métodos mais efetivos. Essa atividade foi apoiada pelo desenvolvimento e uso de uma ferramenta de anotação de imagens de prescrições médicas, e pela construção de um *dataset* estruturado, empregado no treinamento e na avaliação de métodos de aprendizado de máquina.

Um fluxo final com os melhores métodos identificados em cada etapa foi definido e avaliado. Por fim, esse fluxo foi incorporado na ferramenta de anotação, constituindo assim um protótipo de um futuro sistema para reconhecimento de prescrições médicas manuscritas no contexto de uma farmácia de manipulação.

1.3 Organização do documento

O restante deste trabalho está organizado como segue: o Capítulo 2 apresenta uma revisão da literatura; o Capítulo 3 apresenta uma visão geral da abordagem usada no processamento de prescrições médicas, os dados que foram empregados e as métricas aplicadas na avaliação de etapas especificas da abordagem desenvolvida; os Capítulos 4, 5 e 6 descrevem cada uma das etapas da abordagem, com os métodos que foram aplicados, os experimentos realizados e seus resultados; O Capítulo 7 apresenta os resultados obtidos em um experimento realizado com a abordagem completa; por fim, o Capítulo 8 apresenta as discussões finais e traz sugestões para trabalhos futuros. Adicionalmente, o Apêndice A mostra uma visão geral da arquitetura do sistema protótipo desenvolvido para apoiar a anotação de prescrições médicas e também para o uso pela farmácia parceira. O Anexo I contém o parecer consubstanciado emitido pelo Comitê de Ética em Pesquisa da Universidade Federal de Goiás que analisou os aspectos éticos do projeto.

CAPÍTULO 2

Revisão Bibliográfica

Este capítulo apresenta o processo de revisão bibliográfica realizado visando identificar trabalhos que abordam o assunto de análise e reconhecimento de texto em prescrições médicas, com um foco em artigos que descrevem métodos para prescrições manuscritas. A Seção 2.1 detalha as questões de pesquisa que guiaram a revisão, a Seção 2.2 descreve o protocolo de pesquisa usado para a busca e a seleção dos trabalhos, a Seção 2.3 apresenta um resumo dos trabalhos selecionados e a Seção 2.4 contém discussões acerca dos trabalhos revisados visando responder às questões de pesquisa levantadas.

2.1 Questões de Pesquisa

A revisão realizada para este trabalho buscou responder as seguintes questões de pesquisa:

QP1 Quais etapas de processamento são realizadas no reconhecimento de prescrições médicas?

QP2 Quais métodos de Aprendizado de Máquina têm sido mais frequentemente aplicados para o reconhecimento do texto manuscrito?

QP3 Em média, qual é o tamanho das bases de dados usadas para avaliar esses métodos de aprendizado?

QP4 Quais são as taxas de sucesso obtidas com tais métodos?

2.2 Protocolo de pesquisa e detalhes da condução

Para este estudo, foi desenhado um protocolo simplificado de pesquisa baseado em uma busca no *Google Scholar* por uma expressão formadas por palavras-chave e na análise dos trabalhos encontrados visando sua aceitação ou rejeição. Em seguida, os artigos aceitos foram lidos e sumarizados a fim de responder às perguntas de pesquisa.

O levantamento bibliográfico usando o motor do *Google Scholar* empregou a seguinte *string* de busca:

> (“OCR” OR “optical character recognition”) AND “prescription” AND (“machine learning” OR “ML” OR “natural language processing” OR “NLP”).

Foi então realizada a leitura dos títulos e *abstracts* dos trabalhos retornados nos primeiros 100 resultados da busca, ordenados por relevância, e selecionados aqueles que fizeram menção a um processo de reconhecimento automatizado de texto a partir de imagens de prescrições médicas. Neste passo, foram selecionados um total de doze trabalhos. Desses, foram excluídos da etapa de leitura e extração de informações três trabalhos que apresentavam métodos somente para prescrições impressas e um trabalho que não apresentou qualquer informação numérica sobre os resultados obtidos nem os dados ou métodos usados para avaliar sua abordagem. Os métodos de reconhecimento de dados de prescrições manuscritas dos oito trabalhos restantes estão resumidos na próxima seção.

2.3 Resumo dos artigos selecionados

O trabalho de Najafiragheb, Hatam e Harifi (2017) descreve um método para o reconhecimento de nomes de medicações em imagens que contenham somente uma lista de palavras manuscritas, onde cada linha na imagem tem o nome de uma medicação e uma palavra que indica o seu tipo. O método consiste de quatro etapas:

- Pre-processamento – As imagens são convertidas para escala de cinza e é aplicado um processo de limiarização para reduzir ruídos. Por fim, é aplicada uma operação de dilatação para reconectar componentes conexos que tenham sido separados pela limiarização.
- Extração de linhas e palavras – Cada uma das linhas de texto das imagens são segmentadas. Em seguida, as palavras em cada linha são segmentadas por uma análise de componentes conexos.
- Extração de características – É aplicado um filtro de Gabor para gerar matrizes de características para as palavras segmentadas.
- Classificação – É aplicado um classificador do tipo k-Nearest Neighbors para classificar as palavras, retornando o texto (classe) ao qual a palavra corresponde.

O método foi avaliado usando um conjunto de 259 imagens para treino e 60 imagens para teste. Foi obtida uma acurácia de 83%, provavelmente de palavras corretamente identificadas, embora a natureza da acurácia não tenha sido explicitamente

apresentada[1]. Todas as imagens continham nomes que pertenciam a um conjunto de 10 medicações.

Já o método descrito por Chumuang e Ketcham (2018) faz o reconhecimento do nome de medicações em imagens que contenham uma única linha de texto manuscrito. O artigo descreve um processo que inclui a identificação de regiões de interesse e a segmentação de linhas de texto em imagens de prescrições completas, mas menciona que essas tarefas foram realizadas de forma externa ao sistema proposto, sendo então a entrada esperada para o sistema uma imagem com uma única linha de texto. O método consiste das seguintes etapas:

- Pre-processamento – A imagem é convertida para escala de cinza e depois binarizada[2].
- Segmentação de caracteres – Os caracteres individuais são segmentados usando um processo de duas etapas, que aplica uma análise de projeção seguida de uma análise de componentes conexos.
- Classificação – Os caracteres segmentados são classificados usando um Multilayer Perceptron (MLP). O MLP foi treinado com 5200 imagens de caracteres manuscritos.
- Análise sintática – É feita uma análise sintática do posicionamento dos caracteres reconhecidos usando um banco com os nomes de 520 medicações populares. O primeiro caractere reconhecido delimita o conjunto de palavras a serem consideradas, e é então selecionada a palavra que tem o maior número de caracteres em comum na ordem em que foram reconhecidos.

O método foi testado com 35 linhas de texto, segmentadas a partir de 5 imagens de prescrições médicas. Foi obtida uma acurácia de palavra de 74.13%.

O trabalho de Fajardo et al. (2019) apresenta um método para o reconhecimento de texto cursivo manuscrito por médicos. O método recebe como entrada uma imagem de texto manuscrito pré-segmentada, que contém o nome de uma medicação e as instruções de uso correspondentes. A base de dados usada foi composta por 1,800 imagens, onde cada imagem contém o nome de uma medicação entre 12 medicações possíveis. O texto manuscrito contido nas imagens foi escrito por médicos de múltiplos hospitais e clínicas. As imagens da base foram pré-processadas aplicando binarização seguida do uso do algoritmo de afinamento de Zhang Suen e foram manualmente marcadas para indicar a qual classe pertenciam. Foram usadas 1,260 imagens da base para treinar

[1] Alguns trabalhos na área não explicitam o tipo de medida de acurácia utilizada, podendo ser de caracteres, de palavras ou de uma linha ou frase inteira de texto. Quando essa informação estiver clara ou for possível inferi-la, ela será apresentada no texto.

[2] Binarização é o processo de converter uma imagem colorida ou em escala de cinza numa imagem puramente preta e branca.

uma Rede Neural Convolucional Recorrente (CRNN). Essa rede foi então aplicada para classificar as 540 imagens restantes, atingindo uma acurácia de 72% para este conjunto. O modelo foi implementado em uma aplicação para dispositivos móveis e foi feita uma segunda validação com um conjunto de 48 imagens contendo texto escrito pelos próprios pesquisadores. A acurácia obtida neste caso foi de 35%.

O método descrito por Kulathunga et al. (2020) faz o reconhecimento do nome de medicações, suas dosagens e instruções de uso, a partir de imagens pré-cortadas de prescrições médicas que incluam apenas tais informações. O método segue as etapas abaixo:

- Segmentação – Cada palavra contida na imagem é segmentada e salva como uma imagem separada. Esse processo de segmentação envolve uma sub-etapa de pré-processamento onde as imagens são convertidas para escala de cinza e é aplicado o método de binarização de Otsu. Em seguida, é aplicada uma operação de dilatação na imagem. Por fim, é feita uma análise de contornos para segmentar cada palavra.
- Pré-processamento – As imagens das palavras segmentadas na etapa anterior são ajustadas para terem um tamanho de 128x32 pixels e são submetidas a filtros de aumento de contraste e erosão.
- Classificação – As imagens são classificadas por um modelo de Rede Neural Convolucional Recorrente denominado *SimpleHTR* (SCHEIDL, 2018a), o qual é uma versão simplificada do modelo previamente descrito por Scheidl (2018b). O modelo foi retreinado pelos autores do método usando imagens que contêm nomes de medicações, valores de dosagem e palavras/abreviações geralmente usadas como instruções de uso.
- Processamento de texto – As palavras reconhecidas na etapa anterior são categorizadas entre nome de medicação, dosagem e instrução de uso por um modelo de Reconhecimento de Entidades Nomeadas (NER) que foi treinado com um conjunto de 400 exemplos de dados médicos.

Para treinar e avaliar o método, os autores construíram inicialmente um *dataset* composto de 5700 imagens de palavras, do qual 5% foram alocadas para teste, e 95% para treino. No teste com esse *dataset*, foi obtida uma acurácia de palavra de 55%. Posteriormente, os autores mencionam que foi realizada uma expansão desse *dataset*, mas o tamanho final não foi especificado de forma clara. No teste com o *dataset* expandido, foi obtida acurácia de palavra de 64,07% no processo de reconhecimento do texto manuscrito, e 95% de acurácia no processo de categorização das palavras reconhecidas.

Butala et al. (2020) descrevem um método de reconhecimento de texto manuscrito *online*. Foi utilizado um equipamento especial na forma de uma *smartpen* que transmite automaticamente imagens de escrita para dispositivos *Android* ou *iOS*. As imagens adquiridas por esse método são segmentadas a nível de caracteres usando uma

análise de componentes conexos, e as imagens segmentadas de caracteres são normalizadas para um tamanho fixo NxN e, por fim, binarizadas. Essas imagens pré-processadas são então classificadas por um modelo de redes neurais, mas os detalhes específicos da composição dessa rede não foram explicitados. Em seguida, o texto reconhecido é categorizado usando um *pipeline* de técnicas de Processamento de Linguagem Natural, consistindo de Tokenização, *Tagging* de partes-do-discurso, extração de características e classes de entidades, e reconhecimento de entidades nomeadas. A abordagem foi testada com um conjunto de 24 palavras e foi obtida uma acurácia de palavra de 87,5%.

O trabalho de Hassan et al. (2021), já comentado no Capítulo 1, apresenta um sistema que extrai os nomes de medicamentos a partir de fotos de prescrições médicas tiradas por *smartphones*. As fotos são pre-processadas e segmentadas em três partes: um cabeçalho que contém o nome e a especialização do médico, o meio da prescrição que contém as medicações prescritas, e o rodapé que traz o endereço e dados de contato da clínica ou hospital. O rodapé é descartado e o cabeçalho é usado para identificar a especialização do médico que será usada para melhorar a classificação das medicações. As palavras manuscritas no meio da prescrição são segmentadas e então classificadas por uma Rede Neural Convolucional (CNN). O banco de dados usado foi composto de imagens de prescrições de médicos de múltiplos hospitais e várias especializações, mas o número de imagens que compõem este banco não foi especificado. Foi obtido 73% de acurácia de treino e apenas 50% de acurácia de teste.

Gupta e Soeny (2021), também citados no capítulo anterior, descrevem uma abordagem para detectar o nome de medicações em imagens de prescrições médicas impressas e manuscritas baseada no uso das ferramentas de reconhecimento óptico de caracteres (OCR) da API proprietária de visão computacional da Google, denominada Google Vision. As imagens são submetidas ao processo de OCR usando a API, a qual retorna um conjunto de palavras reconhecidas junto às coordenadas que indicam suas posições na imagem. Esses dados são usados para identificar agrupamentos de palavras que estejam próximas entre si na imagem e que contenham termos que são comumente vistos acompanhando nomes de medicações (ex: comprimidos, gotas). Após identificados estes agrupamentos, palavras que existam em dicionários de inglês são descartadas. No caso de prescrições impressas, as palavras remanescentes são apresentadas como sendo os nomes das medicações. Já no caso de prescrições manuscritas, é feita uma comparação aproximada das palavras com um banco de nomes de medicamentos usando distância de Demerau-Levenshtein e as palavras mais parecidas são apresentadas. A abordagem foi avaliada usando 5,176 prescrições impressas e 5,000 manuscritas, e foi obtido um F-Score de 79,4% para prescrições impressas e apenas 26,2% para prescrições manuscritas.

A abordagem de Wijewardena (2021) é essencialmente uma versão simplificada do trabalho de Kulathunga et al. (2020). Nela, é aplicado o mesmo *pipeline* de pré-

processamento, menos os filtros de contraste e erosão. Técnicas de segmentação automatizada são mencionadas, mas é realizada uma segmentação manual de palavras em imagens de prescrições médicas. O modelo *SimpleHTR* é usado para reconhecer o texto, mas não é feito um processo de retreino para melhorar os resultados. Para correção do texto, é aplicado um método de *matching* de *strings* que calcula um valor de similaridade usando o comprimento da maior *substring* comum dividido pela soma do comprimento das duas *strings*. É selecionada a *string* de maior similaridade dentro de um conjunto de *strings* definido pelo autor, que contém os nomes de todas as medicações nas imagens testadas. A base de imagens usada para avaliação consiste de 176 imagens de prescrições e 412 imagens de linhas de texto segmentadas. Foi obtida uma acurácia de palavra de 63,10% na avaliação com as linhas segmentadas.

2.4 Comentários

A Tabela 2.1 mostra algumas das informações desses oito trabalhos revisados, relacionados à temática de extração automatizada de dados a partir de imagens de prescrições médicas manuscritas. As colunas da tabela compreendem as etapas de processamento realizadas, o método de reconhecimento de texto implementado, os resultados de avaliação do método e os conjuntos de dados utilizados.

Tabela 2.1: *Informações dos trabalhos revisados. O Termo "imagem" é utilizado no geral para referenciar uma imagem de prescrição médica, a menos quando está especializado.*

Trabalho	Etapas de processamento	Método de reconhecimento de texto	Avaliação	Base de dados
Najafiragheb, Hatam e Harifi (2017)	Pré-processamento, Segmentação e Reconhecimento	k-Nearest Neighbors	Acurácia de palavra: 83%	259 imagens. 60 usadas para teste. 10 nomes de medicações
Chumuang e Ketcham (2018)	Pré-processamento, Reconhecimento e Correção	Multi-layer Perceptron	Acurácia de palavra: 74,13%	35 imagens de linhas de texto.
Fajardo et al. (2019)	Pré-processamento e Reconhecimento	Convolutional Recurrent Neural Network	Acurácia: 72%	1800 imagens de linhas de texto. 540 usadas para teste.
Kulathunga et al. (2020)	Pré-processamento, Segmentação, Reconhecimento e Categorização	Convolutional Recurrent Neural Network	Acurácia de palavra: 64%	Inicialmente 5700 imagens de palavras, extraídas de imagens com listas de linhas de texto. O *dataset* foi expandido, mas o tamanho final não foi claramente especificado.
Butala et al. (2020)	Pré-processamento, Reconhecimento e Categorização	Artificial Neural Network	Acurácia de palavra: 87,5%	24 imagens de palavras.
Hassan et al. (2021)	Pré-processamento, Segmentação e Reconhecimento	Convolutional Neural Network	Acurácia: 50%	Não especificado
Gupta e Soeny (2021)	Segmentação, Reconhecimento e Correção	Google Vision	F-Score: 26,2%	5000 imagens, todas usadas para teste.
Wijewardena (2021)	Pré-processamento, Reconhecimento e Correção	Convolutional Recurrent Neural Network	Acurácia de palavra: 63%	412 imagens de linhas de texto, todas usadas para teste.

Entre os trabalhos revisados, no que diz respeito à questão de pequisa sobre as etapas de processamento (**QP1**), foi observado que nem todos os estudos abordaram as mesmas etapas e seus subproblemas. Apesar disso, é possível identificar etapas de processamento comuns. Somente o trabalho de Kulathunga et al. (2020) envolveu a maior quantidade dessas etapas, sendo elas as de pré-processamento, segmentação de texto, reconhecimento de texto, correção de texto e categorização de palavras. Observou-se também que apenas dois trabalhos apresentaram abordagens que processam imagens completas de prescrições médicas reais (HASSAN et al., 2021; GUPTA; SOENY, 2021), com dois trabalhos (NAJAFIRAGHEB; HATAM; HARIFI, 2017; KULATHUNGA et al., 2020) usando imagens pré-cortadas que continham somente listas de linhas de texto manuscrito, três trabalhos usando imagens que linhas de texto individuais (CHUMUANG; KETCHAM, 2018; FAJARDO et al., 2019; WIJEWARDENA, 2021) e um trabalho que usou imagens de palavras individuais (BUTALA et al., 2020).

Relativo a questão de pesquisa **QP2**, foi observada uma tendência para o uso de Redes Neurais Convolucionais no processo de reconhecimento de texto manuscrito. Este método foi empregado em 5 dos trabalhos (FAJARDO et al., 2019; KULATHUNGA et al., 2020; BUTALA et al., 2020; HASSAN et al., 2021; WIJEWARDENA, 2021).

Em termos da questão de pesquisa **QP3**, apenas 2 trabalhos foram avaliados com bases de dados que continham mais de 1.000 imagens (FAJARDO et al., 2019; GUPTA; SOENY, 2021), enquanto os outros trabalhos apresentaram bases de dados pequenas, incluindo o trabalho de Butala et al. (2020) que foi testado com apenas 24 imagens de palavras individuais. Dois trabalhos omitiram ou apresentaram de forma ambígua as informações sobre suas bases de dados (KULATHUNGA et al., 2020; HASSAN et al., 2021).

Sobre a questão de pesquisa **QP4**, de forma geral, os trabalhos não obtiveram níveis de sucesso elevados no reconhecimento de texto manuscrito, com o estudo de Butala et al. (2020) alcançando a maior acurácia, de 87.5%. Mesmo assim, o seu método foi testado com apenas 24 imagens de palavras e foi feito o uso de um equipamento especializado na forma de uma *smartpen* para a aquisição das imagens durante o processo de escrita, algo que pode não ser viável para adoção em maior escala. O trabalho de Gupta e Soeny (2021) obteve o menor nível de sucesso, com um F-Score de 26.2% para prescrições manuscritas, mas o seu método foi avaliado com 5.000 imagens, o maior conjunto de imagens entre todos os trabalhos, e o método foi construído para trabalhar com imagens completas de prescrições médicas reais, diferente dos trabalhos de Najafiragheb, Hatam e Harifi (2017), Chumuang e Ketcham (2018), Fajardo et al. (2019), Kulathunga et al. (2020), Butala et al. (2020), Wijewardena (2021) que trabalharam com imagens pré-segmentadas ou em formatos que facilitassem o processo de reconhecimento de texto.

Os resultados obtidos por estes trabalhos, a maioria publicados nos últimos 4 anos, indicam que os métodos atuais para extração de dados de prescrições manuscritas ainda não apresentam níveis de sucesso satisfatórios. Percebe-se também uma tendência de trabalhos desenvolvidos com bases de dados pequenas, possivelmente indicando dificuldades por parte dos pesquisadores na obtenção destes dados; outra possível causa seria o esforço necessário para a preparação destas bases de dados, considerando que as imagens das bases precisam ser anotadas com as suas informações correspondentes para possibilitar o treinamento dos modelos de aprendizado de máquina.

CAPÍTULO 3

Visão Geral da Abordagem

Este capitulo apresenta uma visão geral da abordagem de processamento de prescrições médicas manuscritas desenvolvida no presente trabalho. A Seção 3.1 detalha as etapas do fluxo de processamento de prescrições adotado. A Seção 3.2 descreve os métodos aplicados na etapa de pre-processamento de imagens. A Seção 3.3 apresenta o processo de anotação de imagens de prescrições que foi realizado e o *dataset* estruturado resultante. Por fim, a Seção 3.4 mostra as métricas de avaliação que foram usadas nos capítulos seguintes.

3.1 Fluxo de etapas

Adotamos, no presente trabalho, com base nos achados da revisão da literatura descrita no capítulo anterior, quatro grandes etapas para a extração automatizada de informações a partir de fotos ou imagens digitalizadas de prescrições médicas de uma farmácia de manipulação:

1. **Pré-processamento** – Caracteriza-se pela aplicação de técnicas de processamento de imagens digitais para facilitar as próximas etapas do processo.
2. **Segmentação** – Consiste na identificação e segmentação das regiões de interesse da imagem, no caso, de palavras manuscritas. Essas regiões contêm informações relevantes, como o nome do paciente, do médico, as fórmulas prescritas com seus compostos e dosagens e as instruções de uso dos medicamentos.
3. **Reconhecimento de texto** – Essa etapa envolve a extração do texto digital a partir das regiões segmentadas usando técnicas de Aprendizado de Máquina (AM). Métodos de AM devem ser treinados e depois utilizados para reconhecer a imagem manuscrita e produzir o texto digital correspondente.
4. **Processamento do texto** – Nesta última etapa, o texto obtido na etapa anterior é corrigido e categorizado, de forma que seja possível identificar dados como nomes de medicações/compostos, dosagem, instruções de uso, etc. Em função do tempo disponível para a pesquisa de mestrado, apenas o processo de correção de texto

foi abordado no presente trabalho, com a tarefa de categorização de texto sendo deixada como atividade futura. Uma parte final do processamento é a construção do texto completo, formado pela união, em linhas consecutivas de texto contínuo, das palavras ou termos reconhecidos para cada segmento da imagem.

As Figuras 3.1 e 3.2 exemplificam as etapas de segmentação, reconhecimento e processamento de texto.

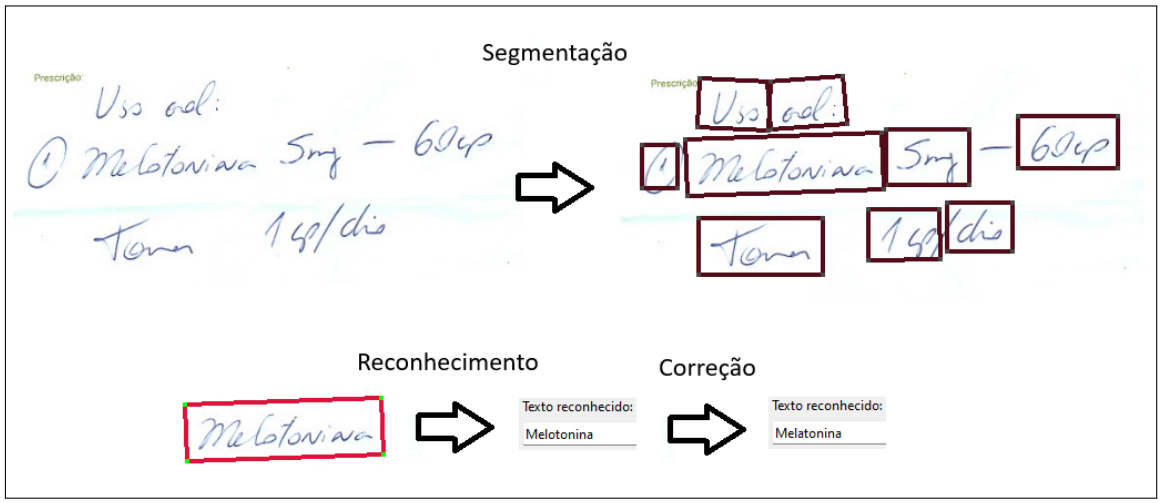


Figura 3.1: *Exemplo das etapas de segmentação, reconhecimento e correção de texto*

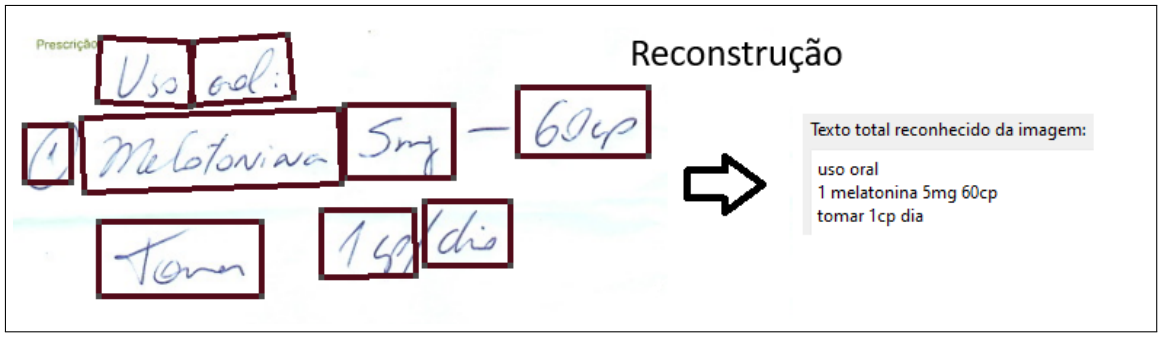


Figura 3.2: *Exemplo da reconstrução do texto completo*

Para atingir os objetivos deste trabalho, foi realizada uma investigação e avaliação de métodos existentes na literatura para cada uma das etapas previamente descritas. Os métodos que apresentaram o melhor desempenho em suas respectivas etapas foram escolhidos e integrados em um fluxo final que implementa a abordagem completa de extração de dados de prescrições médicas manuscritas.

As investigações sobre as etapas de segmentação, reconhecimento e correção de texto são apresentadas nos Capítulos 4, 5 e 6, respectivamente. Por ser uma tarefa relativamente mais simples, a etapa de pré-processamento é apresentada na Seção 3.2 a seguir. A composição do texto final para uma prescrição a partir da união do texto reconhecido para as regiões segmentadas é feita utilizando um algoritmo desenvolvido no escopo deste projeto e está descrito no Apêndice A.

3.2 Pré-processamento das imagens

Em experimentos preliminares com os métodos de reconhecimento de texto descritos neste trabalho (abordados no Capítulo 5), foi observado que eles são consideravelmente sensíveis à inclinação do texto nas imagens, produzindo resultados inferiores para segmentos de texto oblíquos. A fim de tratar esse problema, foi implementado, em uma etapa de pré-processamento, o método *Projection Profile*, inicialmente proposto por Postl (1986). Esse método consiste em definir um intervalo de possíveis ângulos para a imagem e, para cada um, calcular histogramas do número de *pixels* pretos nas linhas horizontais. É então escolhido como ângulo ideal o que tiver a maior variação entre os picos e as quedas do seu respectivo histograma, que correspondem às linhas de texto e espaços brancos entre as linhas em uma imagem não inclinada. A Figura 3.3 ilustra um exemplo do resultado da aplicação deste método.

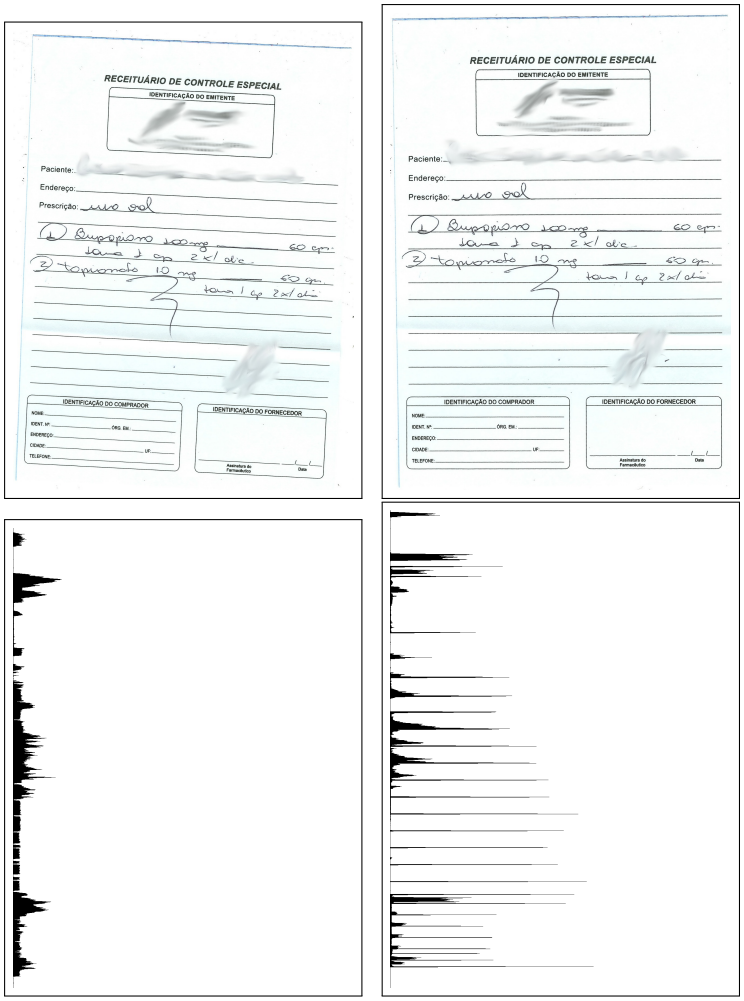


Figura 3.3: *Exemplo de aplicação do método Projection Profile para correção de inclinação da imagem. A figura da esquerda apresenta a imagem original, e a figura da direita mostra o resultado da correção de inclinação. Abaixo das imagens são apresentados os seus respectivos histogramas.*

Além do método de correção de inclinação, também foram aplicadas uma série de técnicas de processamento de imagem especificamente nos segmentos que são usados na etapa de reconhecimento de texto. Como essas técnicas não foram aplicadas nas imagens completas das prescrições, e apenas nas imagens cortadas de palavras individuais, elas estão descritas na Seção 5.3.

3.3 Anotação e Bases de Dados Utilizadas

Para possibilitar o treinamento e a avaliação de modelos de aprendizado de máquina em algumas etapas da abordagem de processamento de prescrições médicas manuscritas, é preciso ter um *dataset* estruturado de imagens de prescrições com seu texto manuscrito propriamente anotado e rotulado.

A princípio, o problema de reconhecimento de texto manuscrito em prescrições médicas não é fundamentalmente diferente do problema de reconhecimento de texto manuscrito geral, para o qual existem múltiplos *datasets* clássicos amplamente utilizados, como o *dataset IAM Handwriting Database* (MARTI; BUNKE, 2002), o *dataset Saint Gall* (FISCHER et al., 2011), e o *dataset Bentham* (SáNCHEZ et al., 2014).

No entanto, as imagens destes *datasets* costumam apresentar características que podem não ser representativas das imagens de prescrições a serem analisadas, como leiautes uniformes, ausência de ruídos associados à captura da imagem, texto escrito de forma retilínea e linguagem específica em que o texto foi escrito. Outro aspecto importante é a alta variabilidade da legibilidade do texto manuscrito em prescrições médicas, um problema constatado pelos farmacêuticos da farmácia parceira e que também se mostrou relevante mesmo em outros países, como visto nos trabalhos de Hartel et al. (2011), Murray et al. (2012), Cerio, Mallare e Tolentino (2015) e Vigneshwaran, Sadiq e Prathima (2016). Esses autores apontaram a baixa legibilidade como sendo uma característica comum em prescrições manuscritas. Um aspecto adicional observado na presente pesquisa foi que prescrições médicas para farmácias de manipulação podem possuir outros tipos de informações como nomes de compostos e suas dosagens para a manipulação de fórmulas, o que podem torná-las mais extensas ou detalhadas.

Tendo isso em vista, se mostrou necessária a construção de um *dataset* de prescrições anotadas para melhor avaliar a eficácia dos modelos de aprendizado de máquina quando aplicados aos problemas específicos abordados neste trabalho.

Para esse fim, a farmácia parceira do projeto cedeu, aos pesquisadores, acesso a uma cópia do banco de dados do sistema de gestão de informações utilizado em suas lojas, conhecido como *Fórmula Certa*. Esse banco contém dados de pedidos atendidos pelas lojas da farmácia no período de janeiro de 2020 até janeiro de 2022, totalizando 256.417

imagens de prescrições e as informações textuais dos pedidos que foram cadastradas pelos funcionários das lojas para cada imagem.

As imagens de prescrição na base têm origens variadas, incluindo imagens escaneadas pelos funcionários e fotos de celular submetidas pelos clientes, com diferentes níveis de ruído (exemplos de ruído incluem a presença de objetos não relevantes, fundos de imagem não uniformes e fotos de baixa qualidade, entre outros).

Para efetivamente utilizar tais imagens com métodos de aprendizado de máquina na presente pesquisa, foi necessário realizar um processo de anotação manual. Esse processo consistiu em, primariamente, especificar nas imagens o quadrilátero onde se encontra cada palavra ou termo manuscrito. Em seguida, essas imagens menores foram rotuladas com o texto equivalente.

Para apoiar todo o processo de anotação, foi implementada uma ferramenta para esse fim. Os detalhes de implementação e as funcionalidades desse ferramenta estão descritos no Apêndice A.

O processo de anotação foi realizado primariamente por duas farmacêuticas experientes na leitura de prescrições manuscritas e por um dos pesquisadores associados ao projeto durante o período de setembro de 2022 até julho de 2023. A anotação seguiu um protocolo composto de um conjunto de especificações:

- Somente texto manuscrito pode ser anotado.
- A anotação deve ser feita ao nível de palavras. Cada palavra das imagens deve ser manualmente segmentada e rotulada.
- Para cada palavra anotada, é preciso gerar dois rótulos, um referente ao texto exatamente como ele está escrito, incluindo erros ortográficos, e o outro referente à palavra como ela deveria estar escrita (corretamente).
- Palavras consideradas ilegíveis ou ambíguas pelos anotadores não devem ser rotuladas.
- A cada palavra deve ser associada uma categoria entre as seguintes: Nome Humano, Medicação, Dosagem, Instruções de uso e Outros.

Como resultado deste processo de anotação, foi construído um *dataset* de texto manuscrito apropriado para o treinamento de modelos de aprendizado de máquina, consistindo em 993 imagens de prescrições anotadas envolvendo 27.933 palavras manuscritas. A Tabela 3.1 apresenta os números de palavras para cada uma das categorias previamente definidas.

Tabela 3.1: *Quantidade de palavras por categoria*

Nome Humano	Medicação	Dosagem	Instruções de uso	Outros
3.523	3.661	4.249	7.295	9.205

Além desse *dataset* de imagens anotadas, também foi construído um dicionário de frequências a partir das informações cadastradas dos pedidos existentes no banco de dados do *Fórmula Certa* para ser aplicado durante a etapa de processamento de texto, especificamente na tarefa de correção. Esse dicionário consiste em uma lista de todas as palavras contidas na tabela que armazena os dados de prescrições já cadastradas no sistema da farmácia parceira, com cada palavra associada a um número que representa a sua quantidade de ocorrências. O dicionário construído contém 12.830 palavras.

Para complementar o dicionário, este foi combinado com um dicionário de frequência geral da língua portuguesa brasileira disponível no repositório *hermitdave/FrequencyWords*[1] do *GitHub*, composto de 848.043 palavras extraídas a partir de legendas de filmes disponíveis na página *OpenSubtitles*[2]. O dicionário combinado resultante contém 853.742 palavras, sendo composto, portanto, por palavras de domínio geral e termos específicos do domínio de prescrições médicas extraídos da base da farmácia parceira.

3.4 Métricas de avaliação

Foram adotadas duas métricas para avaliar a eficácia dos modelos de reconhecimento de texto manuscrito e de correção de texto: a *Character Accuracy Rate* (CAR) e a *Word Accuracy Rate* (WAR). Essas métricas são, respectivamente, os complementos (100 - valor) das métricas *Character Error Rate* (CER) e *Word Error Rate* (WER), comumente utilizadas na área.

A CER representa a taxa de erro ao nível de caractere e é calculada com base na medida de distância de Levenshtein (LEVENSHTEIN et al., 1966), que representa o menor número de operações de remoção, substituição e inserção de caracteres necessárias para transformar uma *string* em outra. A Equação 3-1 define a métrica CER:

$$CER(S_o, S_r) = 100 * \frac{LD(S_o, S_r)}{|S_o|} \tag{3-1}$$

em que S_o é uma *string* que contém a palavra objetivo que se deseja reconhecer, S_r contém a palavra que foi efetivamente reconhecida, $LD(S_o, S_r)$ é a distância de Levenshtein entre as strings S_o e S_r, e $|S_o|$ é o comprimento de S_o.

A métrica *WER* observa o erro ao nível de palavras, então seu cálculo consiste apenas em conferir se a palavra reconhecida é exatamente igual à palavra objetivo.

Para o cálculo dessas métricas, letras maiúsculas e minusculas foram consideradas como iguais.

[1] https://github.com/hermitdave/FrequencyWords
[2] www.opensubtitles.org

CAPÍTULO 4

Segmentação de texto

O processo de segmentação consiste em identificar, em uma imagem, regiões ou objetos de interesse e particioná-los em segmentos menores, visando reduzir a complexidade do seu processamento e análise. No processamento de imagens de prescrições médicas no escopo do presente trabalho, esta etapa envolve identificar apenas o texto manuscrito existente nas imagens e separá-lo em imagens menores contendo palavras individuais. A investigação dessa atividade é o alvo do presente capítulo. A Seção 4.1 apresenta os modelos de aprendizado de maquina que foram avaliados para esta etapa. A Seção 4.2 detalha os experimentos que foram realizados para avaliar os modelos e a Seção 4.3 descreve os resultado desses experimentos.

4.1 Modelos de segmentação

Foram avaliados dois modelos de aprendizado de máquina para realizar o processo de segmentação de texto: o modelo *Character-Region Awareness For Text detection* (CRAFT) (BAEK et al., 2019), desenvolvido pelo grupo de pesquisa em Inteligência Artificial *CLOVA*[1] do conglomerado sul-coreano *NAVER*[2], e o modelo DBNet++ desenvolvido por Liao et al. (2022). Ambos são modelos destinados ao problema de detecção de texto de cena, que consiste em detectar texto em imagens ou vídeos de ambientes complexos e não-controlados. Um exemplo deste problema pode ser visto na Figura 4.1.

[1] https://clova.ai/en/research/research-areas.html
[2] https://www.navercorp.com/en

Fonte: NAVER Corp, CRAFT: Character-Region Awareness For Text detection

Disponível em: https://github.com/clovaai/CRAFT-pytorch

Figura 4.1: *Exemplo de detecção de texto de cena*

A escolha destes modelos foi feito com base nos resultados positivos obtidos por Nguyen, Nguyen e Le (2021), que aplicou o modelo CRAFT para detecção de texto digital em imagens de prescrições médicas, e pelos resultados obtidos por Olejniczak e Šulc (2023), que avaliou o desempenho de ambos os modelos e sete outros métodos de detecção de texto de cena quando aplicados ao problema de detecção de texto digital em imagens de documentos estruturados.

Tendo em vista os bons resultados obtidos por esses trabalhos, foi feita uma avaliação para verificar se, quando devidamente treinados, esses modelos também seriam capazes de detectar texto manuscrito em imagens de prescrições médicas com um bom desempenho. Os resultados dessa avaliação podem ser vistos na Seção 4.3.

4.1.1 CRAFT

O modelo *CRAFT* realiza uma detecção de texto a nível de caracteres, utilizando uma rede neural convolucional profunda baseada na arquitetura VGG16 (SIMONYAN; ZISSERMAN, 2014). Ele calcula as regiões de cada caractere de texto em uma imagem e valores de afinidade entre os caracteres, de forma a agrupá-los em palavras. Por realizar uma detecção a nível de caractere, o modelo tem vantagens na detecção de texto curvo, distorcido ou muito longo, o que pode ser complexo de fazer corretamente com modelos treinados para calcular regiões delimitadoras a nível de palavras ou linhas de texto.

Devido à escassez de bases de dados de imagens reais com anotação a nível de caracteres, os autores do modelo desenvolveram uma abordagem de aprendizado fracamente supervisionado, onde é treinado inicialmente um modelo intermediário em uma base de imagens sintéticas conhecida como *SynthText* (GUPTA; VEDALDI; ZISSERMAN, 2016). Este modelo intermediário é então usado para gerar anotações a nível de caractere aproximadas para bases de dados com anotação a nível de palavra, que são usadas para realizar o treinamento do modelo final.

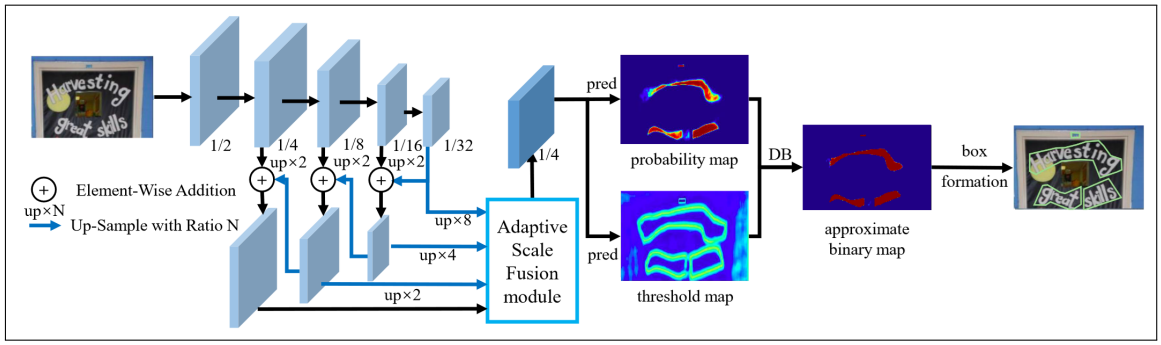
Nos testes dos autores usando o *dataset* do problema de reconhecimento de texto de cena *ICDAR2015* (KARATZAS et al., 2015), o modelo *CRAFT* atingiu um F1-Score

de 86,9%.

O modelo retorna como resultado final de sua execução uma lista de regiões delimitadoras na forma de retângulos ou polígonos que indicam onde cada palavra detectada está localizada na imagem.

4.1.2 DBNet++

O modelo DBNet++ realiza uma detecção a nível de palavras, utilizando uma rede neural convolucional profunda baseada na arquitetura ResNet (HE et al., 2016) para gerar um mapa das probabilidades de cada pixel da imagem de entrada representar texto, e um mapa de limiares que é aplicado conjuntamente ao de probabilidades num processo de binarização diferenciável para produzir um mapa binário, onde cada pixel que tem o valor 1 é indicativo de texto. Este mapa binário é então usado para gerar as regiões delimitadoras finais que são retornadas pelo modelo. A Figura 4.2 ilustra a arquitetura do modelo.



Fonte: Liao et al. (2022)

Figura 4.2: *Arquitetura do modelo DBNet++*

Nos testes dos autores usando o *dataset ICDAR2015*, o modelo *DBNet++* atingiu um F1-Score de 87,3%.

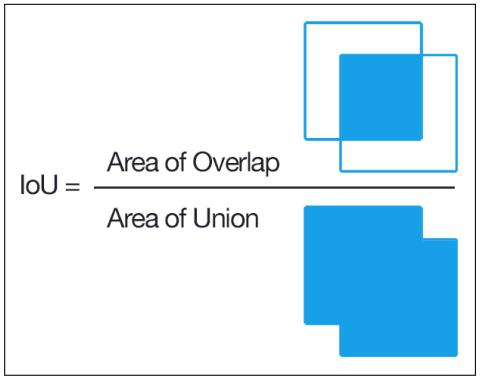
Assim como o *CRAFT*, o *DBNet++* retorna uma lista de regiões delimitadoras na forma de retângulos ou polígonos para cada palavra detectada na imagem.

4.2 Descrição dos experimentos

Para avaliar o desempenho dos modelos no problema de detectar texto manuscrito em prescrições médicas, foi utilizado o novo *dataset* de imagens anotadas descrito na Seção 3.3. O conjunto total de 993 imagens anotadas foi dividido em um conjunto de treino composto por 800 imagens, com 22.295 palavras rotuladas, e um conjunto de teste de 193 imagens, com 5.638 palavras. A separação entre os conjuntos foi feita de forma cronológica, com o conjunto de teste sendo composto pelas 193 imagens de prescrições mais recentes entre as anotações. Foram então conduzidos dois experimentos:

1. Os dois modelos foram avaliados no conjunto de teste usando as versões pré-treinadas e publicamente disponíveis nos seus repositórios do *GitHub*. Neste caso, ambos os modelos haviam sido treinados no *dataset* do problema de reconhecimento de texto de cena *ICDAR2015* (KARATZAS et al., 2015).
2. Os modelos foram treinados do zero no conjunto de treino desta pesquisa e depois avaliados no conjunto de teste. No caso do *CRAFT*, o modelo pré-treinado usado no experimento anterior foi empregado como o modelo intermediário para o processo de aprendizado fracamente supervisionado.

Considerando que o resultado da inferência desses modelos é uma lista de regiões delimitadoras na forma de retângulos ou polígonos, existe a necessidade da aplicação de alguma métrica para determinar se um par de regiões pode ser considerado igual, a fim de que seja possível identificar quantas palavras manualmente anotadas do conjunto de teste foram detectadas corretamente. Para tal, foi usada a métrica *Intersection-over-Union* (IoU), que consiste em calcular a razão entre a área da interseção das duas regiões com a área de sua união. Desta forma, quanto maior for a sobreposição entre as duas regiões, maior será o valor da métrica *IoU* resultante. A Figura 4.3 exemplifica o cálculo desta métrica.



Fonte: Rosebrock (2016)

Figura 4.3: *Cálculo da métrica IoU*

Para os experimentos realizados, foi considerado que um par de regiões com valor de *IoU* maior que 0,5 são iguais.

Os experimentos foram executados na máquina de um dos pesquisadores, com as seguintes especificações:

- CPU Intel Core i5-13600K,
- GPU NVIDIA RTX 4080 16GB,
- RAM de 32GB,
- Sistema Operacional Microsoft Windows 11 Pro.

Para o modelo *CRAFT*, foi usada uma reimplementação disponível no repositório *gmuffiness/CRAFT-train*[3] do *GitHub*, pois a implementação oficial[4] não disponibilizou o código necessário para realizar o processo de aprendizado fracamente supervisionado, por se tratar de propriedade intelectual do conglomerado *NAVER*. Para o modelo DBNet++, foi usada a implementação disponibilizada pelos autores originais no repositório *MhLiao/DB*[5] do *GitHub*. Os resultados destes experimentos são apresentados a seguir.

4.3 Resultados

As Tabelas 4.1 e 4.2 apresentam os resultados dos dois experimentos realizados.

Tabela 4.1: *Resultados do Experimento 1 de segmentação*

	CRAFT	**DBNet++**
Palavras corretamente detectadas	2.774 (49%)	1.027 (18%)
Regiões detectadas, mas sem correspondentes	8.624	20.759

No Experimento 1, o modelo *CRAFT* detectou corretamente 49% das palavras manuscritas do conjunto de imagens de teste, enquanto o modelo *DBNet++* detectou apenas 18%. Ambos os modelos detectaram uma elevada quantidade de regiões para as quais não existem correspondentes nas anotações do conjunto de teste. Isso é esperado, pois as imagens de prescrições médicas usadas para teste apresentam, em grande maioria, quantidades moderadas de texto digital impresso no papel das prescrições que foram detectadas por ambos os modelos. No entanto, existe uma grande discrepância na quantidade de regiões erroneamente detectadas entre os dois modelos, com o modelo *DBNet++* retornando aproximadamente 2,4 vezes a quantidade dessas regiões em comparação ao modelo *CRAFT*.

Tabela 4.2: *Resultados do Experimento 2 de segmentação*

	CRAFT	**DBNet++**
Palavras corretamente detectadas	3.842 (68%)	3.443 (61%)
Regiões detectadas, mas sem correspondentes	1.200	4.917

No Experimento 2, ambos os modelos apresentaram um aumento significativo de desempenho na tarefa de detecção de texto manuscrito em relação ao Experimento 1. O modelo *CRAFT* detectou corretamente 68% das palavras manuscritas do conjunto de

[3] https://github.com/gmuffiness/CRAFT-train
[4] https://github.com/clovaai/CRAFT-pytorch
[5] https://github.com/MhLiao/DB

imagens de teste, enquanto o modelo *DBNet++* detectou 61%. Os dois modelos também demonstraram uma redução expressiva na detecção de regiões sem correspondentes, embora perceba-se ainda uma discrepância entre eles, com o *DBNet++* detectando 4 vezes mais regiões sem correspondentes. Isso pode indicar que o modelo *DBNet++* é mais sensível à presença de texto digital ou outros tipos de ruído, mesmo quando treinado com um *dataset* específico para texto manuscrito.

Os resultados obtidos indicam que, quando propriamente treinados, os modelos testados podem apresentar potencial para a detecção de palavras de texto manuscrito. No entanto, o seu desempenho ainda é baixo, com o melhor resultado de palavras corretamente detectadas o de 68% (alcançado pelo *CRAFT*), e uma taxa não desprezível de detecção de elementos potencialmente indesejados, como texto digital ou outros tipos de ruídos presentes nas imagens.

CAPÍTULO 5

Reconhecimento de texto

Este capítulo avalia modelos para o reconhecimento de texto nas regiões segmentadas da imagem de prescrição médica manuscrita. Os modelos escolhidos para estudo são descritos na Seção 5.1. Em seguida, na Seção 5.2, são apresentados os detalhes dos experimentos organizados. A Seção 5.3 apresenta as técnicas de pré-processamento que foram aplicadas nas imagens cortadas de palavras individuais antes das mesmas serem submetidas ao reconhecimento pelos modelos selecionados. Os resultados dos experimentos são descritos e discutidos na Seção 5.4.

5.1 Modelos de reconhecimento

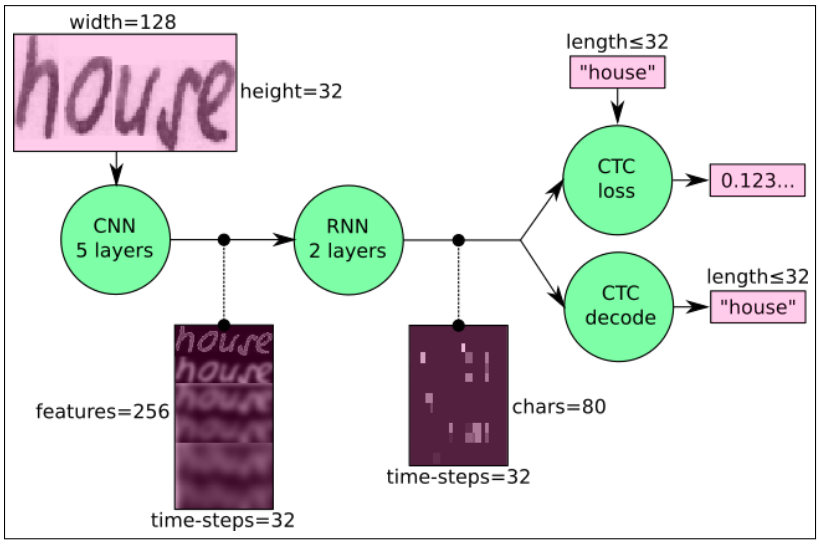
Foram avaliados quatro modelos de aprendizado de máquina para o processo de reconhecimento de texto manuscrito: o modelo *SimpleHTR*, que foi escolhido por ter sido aplicado nos trabalhos de Kulathunga et al. (2020) e Wijewardena (2021) para o reconhecimento de texto manuscrito em prescrições médicas, e os modelos *HTRFlor* (NETO et al., 2020), *AttentionHTR* (KASS; VATS, 2022) e *TrOCR* (LI et al., 2022), selecionados por serem modelos para reconhecimento de texto manuscrito geral publicados nos últimos 3 anos e que apresentaram alto desempenho quando avaliados por seus autores nos *datasets* clássicos de texto manuscrito.

5.1.1 SimpleHTR

O *SimpleHTR*[1] é uma versão simplificada do modelo desenvolvido por Scheidl (2018b) para reconhecimento de texto manuscrito. O modelo consiste em uma rede neural implementada usando o *framework TensorFlow*[2] composta de uma CNN de 5 camadas, uma RNN *Long Short-Term Memory* (LTSM) de 2 camadas, e uma camada *Connectionist Temporal Classification* (CTC) final. A rede trabalha com imagens de palavras de tamanho

[1] https://github.com/githubharald/SimpleHTR
[2] https://www.tensorflow.org/

128x32 e utiliza 32 *time-steps* para suas funções de perda e decodificação CTC, limitando o tamanho máximo das palavras reconhecidas para 32 caracteres. A Figura 5.1 ilustra a arquitetura deste modelo. As camadas de CNN foram treinadas para extrair características da imagem de entrada, gerando um mapa de 256 características para cada um dos 32 *time-steps*. As camadas RNN propagam informações relevantes do mapa de características e as mapeiam em uma matriz de probabilidades, onde cada elemento representa a probabilidade de um caractere específico ocorrer no *time-step* correspondente, sendo então o tamanho desta matriz dependente do número de caracteres possíveis do *dataset* que foi usado para treinar a rede. Por fim, durante o treinamento, a camada CTC recebe a matriz de probabilidades e o texto verdadeiro da imagem para calcular o valor de perda. Já durante o processo de inferência, a camada CTC recebe apenas a matriz de probabilidades e usa uma de três funções de decodificação para gerar o texto final: as funções *Best Path* e *Beam Search* existentes do *TensorFlow*, e uma terceira função chamada *Word Beam Search* (SCHEIDL; FIEL; SABLATNIG, 2018) que combina o método *Beam Search* com o uso de um dicionário gerado durante o processo de treinamento da rede.



Fonte: Scheidl (2018a)

Figura 5.1: *Diagrama do modelo SimpleHTR*

No trabalho de Kulathunga et al. (2020), o *SimpleHTR* foi treinado e avaliado usando um *dataset* construído pelos autores, composto de 5.700 imagens de palavras individuais de nomes de medicações, nomes de doenças, dosagens e instruções de uso. As palavras coletadas para construir o *dataset* foram limitadas a um conjunto de 6 doenças comuns, e 20 tipos de medicações usadas para o tratamento dessas doenças. 95% do conjunto de palavras foi usado para treino do modelo, e os outros 5% foram destinados ao seu teste. O modelo atingiu uma taxa de acurácia de caractere de 84,4% e uma acurácia de palavra de 55,59%. Os autores mencionam que realizaram um segundo experimento com um *dataset* expandido, mas o seu tamanho não foi claramente especificado no artigo.

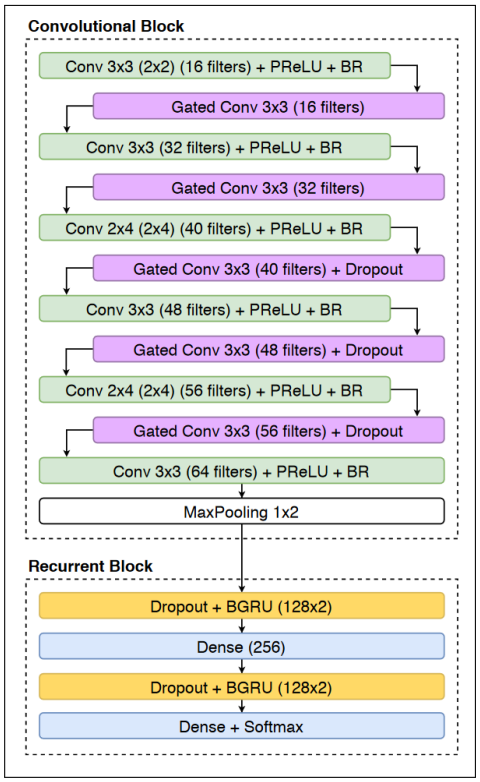
Nesse segundo experimento, o modelo atingiu uma taxa de acurácia de caractere de 85,9%, e uma acurácia de palavra de 64,07%.

Já no trabalho de Wijewardena (2021), o *SimpleHTR* foi treinado no *dataset* clássico *IAM*[3] e testado em um *dataset* construído pelos próprios autores, composto de 412 imagens de linhas de texto manuscrito manualmente segmentadas a partir de 176 imagens de prescrições médicas. Os autores não reportaram métricas para o desempenho imediato do modelo, mas sim para o resultado de um processo de correção de palavras realizado em seguida, sobre a saída do *SimpleHTR*. A correção empregou um dicionário composto de 414 nomes de medicações usadas para o tratamento de 3 doenças comuns. O processo completo foi capaz de identificar nomes de medicações com 63,10% de acurácia de palavra.

5.1.2 HTRFlor

O modelo *HTRFlor*, desenvolvido por Neto et al. (2020), consiste de uma rede neural convolucional recorrente fechada (Gated-CRNN) com unidades recorrentes bidirecionais (BGRU). A rede é composta de um bloco convolucional com 11 camadas, e um bloco recorrente com duas *BGRUs* e duas camadas densas. A Figura 5.2 ilustra a arquitetura dessa rede.

[3] O *IAM Handwriting Database* é um *dataset* de imagens de texto manuscrito de domínio geral na língua inglesa. Ele é composto de 1.539 imagens de formulários escaneados, com 13.353 linhas de texto e 115.320 palavras, possuindo anotações tanto em nível de linhas de texto quanto de palavras.



Fonte: Neto et al. (2020)

Figura 5.2: *Arquitetura do modelo HTRFlor*

Essa arquitetura apresenta, como principal vantagem em comparação às tradicionais baseadas em *LSTM*, a redução dos parâmetros treináveis. A quantidade de parâmetros no *HTRFlor* é na ordem de milhares, contra os milhões vistos em arquiteturas *CNN-BLSTM*. Isso torna o modelo mais leve computacionalmente e dá a ele a habilidade de trabalhar bem com sentenças de texto mais longas, mesmo quando treinado com um baixo volume de dados.

O *HTRFlor* foi avaliado pelos seus autores em cinco *datasets* clássicos de texto manuscrito, incluindo o *IAM*. Para esse *dataset*, o modelo foi treinado e avaliado usando as imagens de linhas de texto, atingindo uma taxa de acurácia de caractere de 96,28% e acurácia de palavra de 88,82%.

5.1.3 AttentionHTR

O modelo *AttentionHTR*, desenvolvido por Kass e Vats (2022), consiste de uma arquitetura composta de quatro estágios: transformação, extração de características, modelagem de sequência e predição.

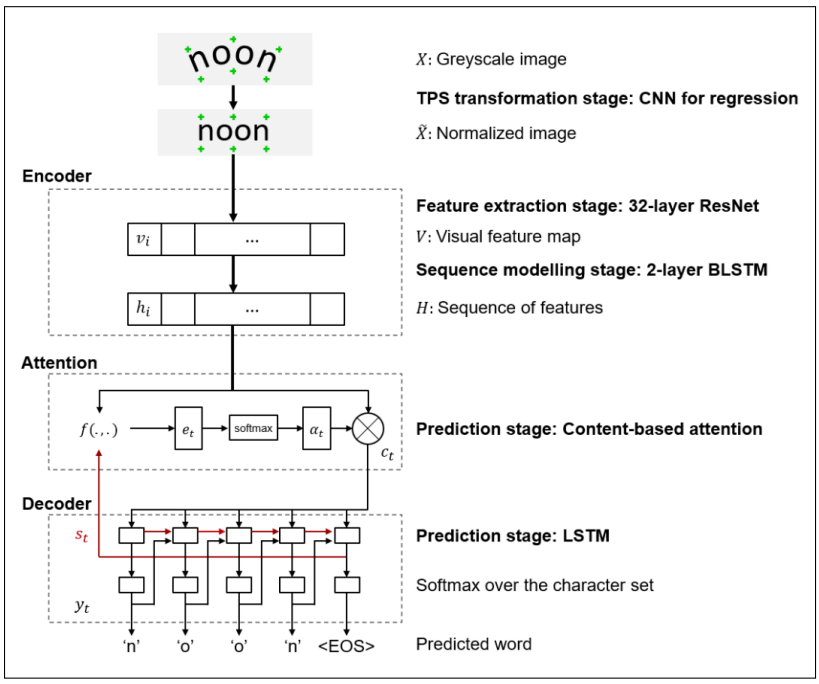
No estágio de transformação, a imagem de entrada é normalizada usando uma transformação *thin-plate spline* (TPS) (BOOKSTEIN, 1989), para normalizar possíveis curvaturas ou angulações comuns em texto manuscrito.

No estágio de extração de características, é aplicada uma rede *ResNet* de 32 camadas para calcular um mapa de características de tamanho 512x26, a partir de uma imagem de entrada normalizada.

No estágio de modelagem de sequência, são aplicadas duas camadas *BLSTM* para remodelar o mapa de características do estágio anterior em uma sequência de características de tamanho 256x26.

Por fim, no estágio de predição, é aplicada uma camada *LSTM* unidirecional com um mecanismo de atenção (BAHDANAU; CHO; BENGIO, 2016), para decodificar a sequência de caracteres da palavra reconhecida.

A Figura 5.3 ilustra a arquitetura desse modelo.



Fonte: Kass e Vats (2022)

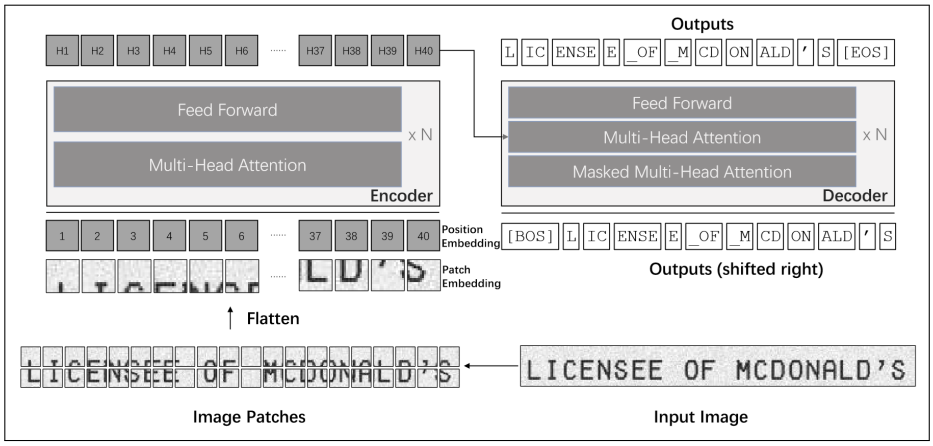
Figura 5.3: *Arquitetura do modelo AttentionHTR*

O *AttentionHTR* foi avaliado pelos autores no *dataset IAM*. Diferente do *HTR-Flor*, contudo, utilizaram-se as imagens de palavras individuais para o estudo, atingindo uma acurácia de caractere de 95,7% e acurácia de palavra de 87,18%.

5.1.4 TrOCR

Por último, o modelo *TrOCR*, desenvolvido por Li et al. (2022), adota uma arquitetura baseada no uso de *Transformers* (VASWANI et al., 2017), modelos de redes neurais simplificadas que aplicam somente mecanismos de atenção, sem convolução nem recorrência. O modelo inclui um *Transformer* pré-treinado para compreensão de imagens como seu codificador, e um *Transformer* pre-treinado para modelagem de linguagens como seu decodificador.

O modelo recebe como entrada uma imagem de texto a ser reconhecida, que é reescalonada para o tamanho 384x384 e depois dividida numa sequência de pedaços de tamanho 16x16. Esses pedaços servem como os dados de entrada do codificador, que realiza o processo de extração de características da imagem. O decodificador então gera segmentos de texto com base nas características extraídas e nos segmentos previamente produzidos. A Figura 5.4 ilustra a arquitetura do modelo.



Fonte: Li et al. (2022)

Figura 5.4: *Arquitetura do modelo TrOCR*

Como previamente mencionado, o codificador e o decodificador usam modelos pre-treinados. O codificador é inicializado com o modelo para tarefas de visão computacional *BEiT* (BAO et al., 2022), e o decodificador com o modelo para tarefas de processamento de linguagem natural *RoBERTa* (LIU et al., 2019). Após a inicialização, deve ser feito um *finetuning* dos modelos usando *datasets* para as tarefas de reconhecimento de texto manuscrito ou digital.

O *TrOCR* foi avaliado pelos autores no *dataset* clássico *IAM* usando as imagens de linhas de texto, atingindo uma taxa de acurácia de caractere de 97,11%.

5.2 Descrição dos experimentos

Assim como nos experimentos da Seção 4.2, foi empregado o *dataset* de imagens de prescrições médicas anotadas para avaliar o desempenho dos modelos no reconhecimento de texto manuscrito. A mesma divisão de conjuntos de treino e teste foi mantida. No entanto, como nesse caso o foco é no reconhecimento de texto, não se utilizou as imagens completas das prescrições, mas sim as imagens cortadas das palavras individuais, produzidas manualmente pelos anotadores, com seus devidos rótulos. Essas imagens de palavras foram organizadas por prescrição de origem, sendo assim possível manter a mesma divisão de treino e teste adotada nos experimentos de segmentação.

Foram realizados dois experimentos:

1. Todos os quatro modelos foram treinados no *dataset* clássico *IAM* e avaliados usando o subconjunto de teste da presente pesquisa (formado por 193 prescrições anotadas).
2. O treinamento realizado no Experimento 1 foi mantido para todos os modelos, com o intuito de implementar um processo de *Transfer Learning*. Esse processo foi feito no escopo de uma validação cruzada com 4 *folds* usando as 800 prescrições do subconjunto de treino do novo *dataset*. Isso significa que, a cada passo da validação, os modelos de aprendizado de máquina originais resultantes do Experimento 1 foram submetidos a um *finetuning* usando dados de 600 prescrições (3 folds) e a uma validação com dados de 200 prescrições (1 fold). Por fim, foi feita mais uma avaliação em que os modelos originalmente treinados no Experimento 1 foram submetidos a um *finetuning* usando o subconjunto completo de treino (dados de 800 prescrições) e testado com o conjunto de teste (dados de 193 prescrições).

Para o modelo *TrOCR*, foi usada uma implementação não-oficial disponível no repositório *rsommerfeld/trocr*[4] do *GitHub*. Para os modelos *SimpleHTR, HTRFlor* e *AttentionHTR*, foram usadas as implementações disponibilizadas pelos autores originais nos repositórios *githubharald/SimpleHTR*[5], *arthurflor23/handwritten-text-recognition*[6] e *dmitrijsk/AttentionHTR*[7], respectivamente.

Os experimentos foram executados na mesma máquina descrita na Seção 4.2. Os resultados desses experimentos, expressos em termos de CAR e WAR, são apresentados na Seção 5.4.

5.3 Pré-processamento

Antes de realizar o reconhecimento com os modelos previamente descritos, as imagens cortadas de palavras individuais foram pré-processadas com as seguintes técnicas:

- Redução de ruído *Non-local Means*
- Limiarização adaptativa gaussiana
- Erosão
- Dilatação

A Figura 5.5 ilustra um exemplo da aplicação dessas técnicas em um segmento de imagem contendo uma palavra manuscrita.

[4] https://github.com/rsommerfeld/trocr
[5] https://github.com/githubharald/SimpleHTR
[6] https://github.com/arthurflor23/handwritten-text-recognition
[7] https://github.com/dmitrijsk/AttentionHTR

Figura 5.5: *Exemplo da abordagem de pre-processamento usada para a etapa de reconhecimento de texto*

5.4 Resultados

As Tabelas 5.1 e 5.2 apresentam os resultados dos dois experimentos realizados. Os quatro modelos de aprendizado de máquina estão referenciados nas colunas, enquanto as linhas dizem respeito às métricas CAR e WAR.

Tabela 5.1: *Resultados do Experimento 1 de reconhecimento*

	SimpleHTR	**HTRFlor**	**AttentionHTR**	**TrOCR**
CAR	28,5%	20,2%	22,0%	39,0%
WAR	6,0%	3,7%	4,4%	11,1%

No Experimento 1, o modelo *TrOCR* apresentou o melhor desempenho entre os quatro modelos, com uma taxa de acurácia de caractere de 39% e acurácia de palavra de 11,1%, enquanto o modelo *HTRFlor* teve o pior desempenho, com uma acurácia de caractere de apenas 20,2% e acurácia de palavra de 3,7%. No geral, todos os modelos apresentaram acurácias extremamente baixas neste experimento, indicando que o treinamento no *dataset IAM* foi insuficiente para atingir um desempenho adequado neste problema. Existem múltiplos fatores que podem ter influenciado esse resultado, como a diferença de linguagem entre os *datasets*, com o *dataset IAM* contendo somente texto na língua inglesa, as diferenças de qualidade das imagens de treino e teste (o *IAM* é composto de imagens de formulários escaneados com resolução e leiautes uniformes), e a falta de palavras do domínio médico no *dataset IAM*.

Tabela 5.2: *Resultados do Experimento 2 de reconhecimento*

	SimpleHTR	**HTRFlor**	**AttentionHTR**	**TrOCR**
Fold 1	62,7% CAR 31,1% WAR	69,5% CAR 38,2% WAR	72,6% CAR 44,7% WAR	78,9% CAR 57,4% WAR
Fold 2	67,2% CAR 41,9% WAR	74,6% CAR 50,1% WAR	77,1% CAR 55,4% WAR	83,8% CAR 69,7% WAR
Fold 3	71,2% CAR 44,7% WAR	77,4% CAR 52,6% WAR	79,8% CAR 57,6% WAR	86,5% CAR 72,4% WAR
Fold 4	71,9% CAR 45,9% WAR	78,7% CAR 54,9% WAR	82,0% CAR 61,6% WAR	86,9% CAR 74,4% WAR
Média	68,3% CAR 40,9% WAR	75,1% CAR 49,0% WAR	77,9% CAR 54,9% WAR	84,1% CAR 68,5% WAR
Teste	72,3% CAR 45,2% WAR	77,1% CAR 51,1% WAR	81,1% CAR 58,6% WAR	86,8% CAR 73,0% WAR

No Experimento 2, o modelo *TrOCR* novamente apresentou o melhor desempenho entre os quatro. Na validação cruzada, ele atingiu um *CAR* médio de 84,1%, e *WAR* médio de 68,5%. Já quando treinado no conjunto de treino inteiro e avaliado no conjunto de teste, apresentou um *CAR* de 86,8% e WAR de 73,0%. Em contraste, o modelo *SimpleHTR* teve o pior desempenho, atingindo apenas 68,3% CAR e 40,9% WAR na validação cruzada.

Houve, em geral, um aumento significativo do desempenho de todos os modelos em comparação aos resultados obtidos no Experimento 1, o que confirma a necessidade de treinar os modelos em um conjunto de dados com características similares aos dados para os quais serão aplicados. Isso se confirma mesmo que o conjunto de treino tenha um volume de dados relativamente pequeno quando comparado a *datasets* como o *IAM*, o qual possui 115.320 palavras rotuladas, aproximadamente 5 vezes a quantidade de palavras presentes no conjunto de treino do *dataset* construído.

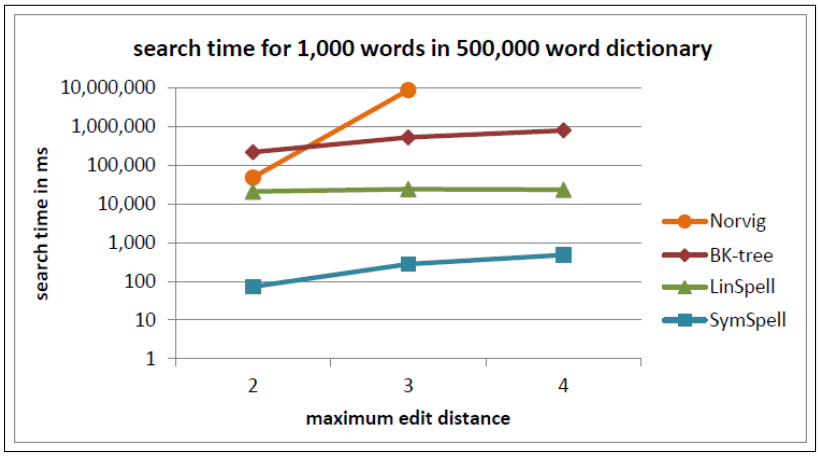
CAPÍTULO 6

Correção de texto

É comum que o texto gerado na etapa de reconhecimento contenha pequenos erros, mesmo quando o modelo utilizado apresenta um bom desempenho. O presente capítulo trata, portanto, da última etapa do fluxo de processamento, responsável pela correção de texto. O capítulo foi organizado de forma similar aos dois capítulos anteriores, com a Seção 6.1 descrevendo o método de correção investigado, a Seção 6.2 detalhando os experimentos e a Seção 6.3 apresentando e discutindo os resultados.

6.1 Método de correção

Foi escolhido o método *SymSpell* (GARBE, 2012) para realizar o processo de correção de texto. Essa escolha foi baseada na sua simplicidade de implementação e no baixo custo computacional quando comparado a métodos tradicionais de natureza equivalente. O *SymSpell* trabalha com a correção de palavras usando um dicionário de frequência, que consiste em um dicionário onde cada palavra tem um número associado que representa a sua frequência de uso. A Figura 6.1 ilustra uma comparação do tempo de busca do *SymSpell* relativo a outros métodos baseados em dicionário.



Fonte: Wolf Garbe, SymSpell

Disponível em: https://github.com/wolfgarbe/SymSpell

Figura 6.1: *Comparação de tempo de busca do método de correção SymSpell*

O funcionamento do *SymSpell* envolve duas fases. A primeira fase é um pré-processamento realizado uma única vez e cujo resultado é um novo dicionário salvo em memória para ser empregado posteriormente na correção de palavras. Tal pré-processamento consiste em gerar todas as possíveis variações de cada palavra do dicionário em uma distância de Levenshtein *N* (especificada pelo usuário) e usando somente a operação de deleção de caracteres. Essas variações são então adicionadas ao dicionário como entradas auxiliares, fazendo referência às palavras originais.

A segunda fase é realizada sempre que há uma palavra a ser corrigida. O mesmo processo de geração de variações é aplicado à palavra e ela e suas variações são buscadas no dicionário estendido. O algoritmo retorna uma lista de palavras candidatas originais para correção com uma distância de Levenshtein máxima *N*, ordenadas de acordo com seu valor de frequência.

6.2 Descrição dos experimentos

Para os experimentos de correção, foram utilizados os resultados de reconhecimento de texto advindos do *TrOCR* como base, uma vez que ele apresentou o melhor desempenho entre os modelos avaliados. O conjunto de dados para esses experimentos consistiu, assim, de 5.638 palavras reconhecidas pelo *TrOCR* e as suas respectivas versões corretas, manualmente anotadas. Além disso, foi empregado o dicionário de frequência descrito na Seção 3.3, o qual contempla palavras do banco de dados do sistema Fórmula Certa da farmácia parceira e expressões da língua portuguesa, totalizando 853.742 entradas.

Foram conduzidos dois experimentos:

1. Correção usando um dicionário parcial, composto apenas por 12.830 palavras extraídas do banco de dados do *Fórmula Certa*.
2. Correção usando o dicionário completo, com 853.742 palavras.

Em ambos os experimentos, somente foram submetidos à correção palavras com comprimento maior do que 2 caracteres e que não continham caracteres numéricos. A distância de Levenshtein máxima usada foi 4. Foi usada uma implementação não-oficial para *Python* do *SymSpell* disponível no repositório *mammothb/symspellpy*[1] do *GitHub*.

[1] https://github.com/mammothb/symspellpy

6.3 Resultados

As Tabelas 6.1 e 6.2 apresentam os resultados dos dois experimentos realizados. Elas mostram o resultado base do *TrOCR* e do texto após a correção em termos de taxa de acurácia de palavra (WAR), tanto para o conjunto total de palavras de teste quanto para cada categoria de palavra separadamente.

Tabela 6.1: *Resultados do Experimento 1 de correção*

	Total	**Nome Humano**	**Medicação**	**Dosagem**	**Instruções de uso**	**Outros**
Base	73,0%	60,9%	60,7%	86,1%	77,7%	73,2%
Corrigido	64,8%	27,2%	66,4%	85,1%	65,3%	69,3%

Observa-se, pelo Experimento 1, que houve uma redução da acurácia em todas as categorias de palavras, menos na categoria de Medicações, a qual teve um aumento de 5,7%. Esses resultados são consistentes com a composição do dicionário usado, o qual possui primariamente termos do domínio médico, carecendo de palavras do domínio geral da língua portuguesa. O ganho de 5,7% em Medicações é significativo, já que pode-se considerá-la uma das categorias mais importantes. Por outro lado, não é possível desprezar a perda de acurácia em todas as demais categorias, causada por 'correções' indevidas, já que têm grandes impactos negativos no entendimento das prescrições médicas.

Tabela 6.2: *Resultados do Experimento 2 de correção*

	Total	**Nome Humano**	**Medicação**	**Dosagem**	**Instruções de uso**	**Outros**
Base	73,0%	60,9%	60,7%	86,1%	77,7%	73,2%
Corrigido	73,5%	61,1%	64,3%	86,1%	78,3%	72,9%

No Experimento 2, houve um aumento de acurácia em todas as categorias, menos na de Dosagem, que se manteve igual, e na categoria Outros, que sofreu uma pequena redução, de 0,3%. O ganho mais expressivo ocorreu na categoria Medicação, com um aumento de 3,6%. O aumento de acurácia na categoria Medicação foi levemente menor do que aquele obtido no Experimento 1, mas, em contra-partida, foi evitada a deterioração geral dos resultados nas demais categorias.

Esses resultados indicam que a composição do dicionário usado afeta significativamente a qualidade do processo de correção. Como consequência, se for possível determinar previamente a categoria da palavra a ser corrigida, um dicionário mais apropriado pode ser escolhido e utilizado na etapa.

CAPÍTULO 7

Experimentação com o fluxo completo

Neste capítulo, fazemos uma avaliação do fluxo completo da abordagem, composto por uma integração em sequência da melhor solução computacional identificada nos capítulos anteriores. O capítulo está organizado como segue: a Seção 7.1 lista os métodos que foram escolhidos para compor a abordagem final, a Seção 7.2 detalha o experimento que foi realizado, e a Seção 7.3 apresenta os resultados obtidos.

7.1 Métodos e Modelos Escolhidos

Tendo em vista os resultados obtidos nos experimentos descritos nos Capítulos 4, 5 e 6, foram selecionados os seguintes métodos para compor a abordagem final:

- Para a etapa de pré-processamento, foram aplicadas as técnicas descritas nas Seções 3.2 e 5.3.
- Para a etapa de segmentação, foi escolhido o modelo *CRAFT*.
- Para a etapa de reconhecimento, foi usado o modelo *TrOCR*.
- Para a correção de texto, foi usado o método *SymSpell* em conjunto com o dicionario completo de 853.742 palavras descrito na Seção 3.3.

7.2 Sequência de experimentação

Para efetivamente avaliar o desempenho desta abordagem completa, foi realizado um único experimento consistindo em uma sequência de passos e usando as 193 imagens de prescrições manuscritas do conjunto de teste definido na Seção 3.3, a saber:

1. As 193 imagens de prescrições foram submetidas ao processo de segmentação usando o modelo *CRAFT*, que foi capaz de corretamente detectar 3.842 das 5.638 palavras contidas nessas imagens. As 1.200 segmentações incorretas foram descartadas.

2. As 3.842 palavras corretamente detectadas foram recortadas usando a segmentação do *CRAFT* e pré-processadas de acordo com o processo descrito na seção 5.3. Essas imagens de palavras foram então submetidas ao processo de reconhecimento usando o modelo *TrOCR*.
3. Por fim, os resultados de reconhecimento do *TrOCR* foram corrigidos pelo método *SymSpell*, usando o dicionario completo de 853.742 palavras descrito na seção 3.3.

Os resultados do experimento são apresentados a seguir.

7.3 Resultados

A Tabela 7.1 apresenta os resultados do experimento realizado. Ela mostra, em termos de acurácia de caractere (CAR) e acurácia de palavra (WAR), os resultados de reconhecimento base do *TrOCR* para o conjunto de 3.842 palavras corretamente detectadas pelo *CRAFT*, e os resultados após a correção de texto com o método *SymSpell*. Também são apresentados os resultados para o conjunto completo de 5.638 palavras de teste, onde as 1.796 palavras que não foram detectadas no processo de segmentação são contabilizadas como tendo sido reconhecidas completamente erradas (0% CAR). Em todos os casos, os resultados são apresentados tanto para o conjunto total de palavras quanto para cada categoria de palavra separadamente.

Tabela 7.1: *Resultados do experimento*

	Total	**Nome Humano**	**Medicação**	**Dosagem**	**Instruções de uso**	**Outros**
Parcial (3.842 palavras)	85,2% CAR 67,5% WAR	81,3% CAR 56,7% WAR	83,1% CAR 55,7% WAR	89,4% CAR 79,4% WAR	87,4% CAR 73,7% WAR	84,5% CAR 68,7% WAR
Parcial - Corrigido	84,6% CAR 68,8% WAR	79,6% CAR 58,3% WAR	81,8% CAR 60,5% WAR	89,4% CAR 79,4% WAR	87,4% CAR 74,7% WAR	84,1% CAR 68,8% WAR
Completo (5.638 palavras)	58,0% CAR 45,9% WAR	65,9% CAR 45,9% WAR	68,3% CAR 46,0% WAR	53,6% CAR 47,6% WAR	62,5% CAR 52,7% WAR	50,1% CAR 40,7% WAR
Completo - Corrigido	57,6% CAR 46,8% WAR	64,5% CAR 47,2% WAR	67,3% CAR 49,7% WAR	53,6% CAR 47,6% WAR	62,5% CAR 53,4% WAR	49,8% CAR 40,8% WAR

Quando se observa o conjunto de 3.842 palavras que foram corretamente segmentadas, a abordagem atingiu um *CAR* de 85,2% e *WAR* de 67,5% antes da correção, com uma perda de 0,6% *CAR* e um ganho de 1,3% *WAR* após a correção. Percebe-se que em termos de *WAR*, todas as categorias de palavras tiveram um ganho com o processo de correção ou se mantiveram iguais, com a categoria Medicação apresentando o aumento mais expressivo de 4,8%. Já em termos de *CAR*, observa-se uma pequena redução de acurácia em todas as categorias menos as de Dosagem e Instruções de uso, que permaneceram iguais. Isso indica que apesar do processo de correção resultar em um ganho global em termos da quantidade de palavras que são reconhecidas de forma completamente correta, ele também incorre uma leve deterioração dos resultados em nível de caractere em casos que são feitas ”correções”para as palavras erradas.

No entanto, temos 1.796 palavras que o processo de segmentação foi incapaz de detectar e que, por consequência, não foram passadas para as etapas seguintes do fluxo. Se contabilizarmos essas palavras como tendo sido reconhecidas de forma completamente errada (0% CAR), a abordagem completa terá apresentado um *CAR* de 57,6% e *WAR* de 46,8% para o conjunto total de 5.638 palavras que existem nas 193 imagens de teste. Isso representa uma redução de acurácia significativa comparado aos resultados obtidos nos Capítulos 5 e 6, onde os métodos foram avaliados usando as segmentações manuais que

foram feitas durante o processo de anotação de imagens.

Isso aponta que, apesar do modelo de reconhecimento de texto aplicado apresentar resultados relativamente eficazes, a baixa acurácia do processo de segmentação serve como um grande gargalo para o desempenho da abordagem como um todo, pois uma quantidade significativa de palavras são perdidas antes mesmo da etapa de reconhecimento. Tendo isso em vista, a exploração de possíveis procedimentos para melhorar os resultados do processo de segmentação se mostra como uma atividade futura de alta prioridade.

CAPÍTULO 8

Conclusões

A análise manual de prescrições médicas para inserção de dados em sistemas de gestão de informações de farmácias é um processo que pode requerer tempo e esforço significativo por parte dos farmacêuticos, principalmente pela prevalência de prescrições manuscritas, que ainda compõem uma parcela considerável das prescrições emitidas por médicos. Alguns trabalhos na literatura científica abordam a construção de métodos para a automatização tal processo, mas ainda apresentam limitações no escopo do seu processamento ou utilizam conjuntos de dados relativamente pequenos e fortemente controlados.

O presente trabalho abordou esse problema por meio da elaboração de uma abordagem de extração automatizada de dados de imagens de prescrições médicas que abrangesse as etapas típicas de tais processos, visando a aplicação em condições reais de uma farmácia de manipulação. Para tanto, foi construído um *dataset* de imagens de prescrições médicas manuscritas anotadas com a assistência de farmacêuticas experientes na interpretação da escrita médica. Esse *dataset* foi utilizado para avaliar múltiplos métodos existentes na literatura para os processos de segmentação, reconhecimento e correção de texto.

Para a segmentação, foram treinados e avaliados os modelos de aprendizado de máquina *CRAFT* (BAEK et al., 2019) e *DBNet++*(LIAO et al., 2022). Após treinados no *dataset* construído, o modelo *CRAFT* apresentou os melhores resultados, atingindo uma acurácia de segmentação de 68%.

Para o reconhecimento de texto, foram treinados e avaliados os modelos de aprendizado de máquina *SimpleHTR* (SCHEIDL, 2018b), *HTRFlor* (NETO et al., 2020), *AttentionHTR* (KASS; VATS, 2022) e *TrOCR* (LI et al., 2022). Após ser treinado no *dataset* clássico de texto manuscrito *IAM* (MARTI; BUNKE, 2002) e submetido a um *finetuning* no *dataset* construído, o modelo *TrOCR* atingiu um desempenho significativamente superior aos demais, com uma taxa de acurácia de caractere de 86,8% e acurácia de palavra de 73%.

Para correção de texto, aplicou-se o método *SymSpell* (GARBE, 2012) sobre os resultados do *TrOCR*. Foi construído um dicionário de frequência para utilização desse

método, o qual resultou em uma melhora na taxa total de acurácia de palavra de 0,5% e, mais especificamente na categoria de Medicações, de 3,6%.

Os resultados obtidos nas avaliações supracitadas indicam que a abordagem desenvolvida é razoavelmente eficaz, mas que ainda necessita de refinamentos. A acurácia de segmentação de 68% obtida pelo modelo *CRAFT* aponta que o modelo falha em detectar corretamente uma porção significativa do texto manuscrito presente nas imagens, além de introduzir ruídos nas próximas etapas do processo por realizar detecções que podem não corresponder a texto manuscrito. A acurácia de caractere de 86,8% obtida pelo modelo *TrOCR* indica que ele é substancialmente eficaz no reconhecimento de texto manuscrito, mas que ainda apresenta uma taxa de erro não insignificante, o que pode ser problemático quando se considera que existem múltiplas medicações com nomes extremamente similares, onde uma única letra pode indicar uma substância completamente diferente.

O método que apresentou o melhor desempenho em cada etapa foi integrado em um sistema protótipo destinado ao uso como ferramenta de anotação e de apoio aos funcionários da farmácia de manipulação parceira do projeto. Uma avaliação desse fluxo final levou a uma acurácia de caractere de 57,6% e uma acurácia de palavra de 46,8%. Esses valores diferem daqueles apresentados no Capitulo 6 porque consideram resultados próprios de uma etapa automatizada de segmentação, que foi capaz de segmentar corretamente apenas 3.842 palavras entre as 5.638 que compõem o conjunto de teste. As 1.796 palavras não detectadas foram contabilizadas como tendo sido reconhecidas de forma totalmente errada, o que resultou em acurácias médias significativamente mais baixas, considerando que para as 3.842 palavras corretamente detectadas, foi obtida uma acurácia de caractere de 84,6% e uma acurácia de palavra de 68,8%.

Tendo em vista as limitações observadas nos resultados obtidos, considera-se relevante os seguintes trabalhos futuros, para avaliar possibilidades de melhorias no processo de extração automatizada de dados de prescrições médicas:

- pesquisar o impacto de técnicas de *augmentation* para sinteticamente expandir o conjunto de dados usado no treinamento dos modelos de segmentação e de reconhecimento de texto;
- avaliar o impacto do uso de técnicas de pré-processamento de imagem no desempenho dos modelos de segmentação;
- construir um *dataset* de prescrições médicas com um volume de dados comparável ao de *datasets* clássicos como o *IAM*;
- modificar o modelo *TrOCR* para usar um decodificador treinado na língua portuguesa (o *TrOCR* inicializa seu decodificador com o modelo *RoBERTa* (LIU et al., 2019), que é um *Transformer* treinado em uma enorme base de dados da língua

inglesa. Usar um modelo treinado em texto em português pode vir a melhorar o seu desempenho);

- comparar o desempenho de modelos como o *TrOCR* quando treinados em imagens de linhas de texto e imagens de palavras individuais, para determinar se existem ganhos de desempenho provindos de informações contextuais;
- investigar o uso de técnicas de processamento de linguagem natural para categorizar automaticamente as palavras reconhecidas, como, por exemplo, modelos de reconhecimento de entidades nomeadas;
- realizar um treinamento especializado nos modelos de reconhecimento de texto usando um *dataset* de texto de um único médico com caligrafia de baixa legibilidade para avaliar se existem ganhos comparados ao de um treinamento generalizado, como o feito neste trabalho; e
- explorar o uso de métodos de correção de texto capazes de avaliar informação contextual em um nível de linhas de texto em vez de apenas palavras individuais.

Referências Bibliográficas

BAEK, Y. et al. Character region awareness for text detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2019. p. 9365–9374.

BAHDANAU, D.; CHO, K.; BENGIO, Y. *Neural Machine Translation by Jointly Learning to Align and Translate*. 2016.

BAO, H. et al. *BEiT: BERT Pre-Training of Image Transformers*. 2022.

BOOKSTEIN, F. L. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Trans. Pattern Anal. Mach. Intell.*, v. 11, p. 567–585, 1989. Disponível em: ⟨https://api.semanticscholar.org/CorpusID:47302⟩.

BUTALA, S. et al. Natural Language Parser for Physician's Handwritten Prescription. In: *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)*. [S.l.: s.n.], 2020. p. 1–7.

CERIO, A. A. P.; MALLARE, N. A. L. B.; TOLENTINO, R. M. S. Assessment of the legibility of the handwriting in medical prescriptions of doctors from public and private hospitals in quezon city, philippines. *Procedia Manufacturing*, v. 3, p. 90–97, 2015. ISSN 2351-9789. 6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences, AHFE 2015. Disponível em: ⟨https://www.sciencedirect.com/science/article/pii/S2351978915001134⟩.

CHUMUANG, N.; KETCHAM, M. Handwritten Character Strings on Medical Prescription Reading by Using Lexicon-Driven. In: THEERAMUNKONG, T.; KONGKACHANDRA, R.; SUPNITHI, T. (Ed.). *Advances in Natural Language Processing, Intelligent Informatics and Smart Technology*. Cham: Springer International Publishing, 2018. (Advances in Intelligent Systems and Computing), p. 137–147. ISBN 978-3-319-70016-8.

FAJARDO, L. J. et al. Doctor's Cursive Handwriting Recognition System Using Deep Learning. In: *2019 IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)*. [S.l.: s.n.], 2019. p. 1–6.

FISCHER, A. et al. Transcription alignment of latin manuscripts using hidden markov models. In: *Proceedings of the 2011 Workshop on Historical Document Imaging and Processing*. New York, NY, USA: Association for Computing Machinery, 2011. (HIP '11), p. 29–36. ISBN 9781450309165. Disponível em: ⟨https://doi.org/10.1145/2037342.2037348⟩.

GARBE, W. *1000x Faster Spelling Correction algorithm*. 2012. Disponível em: ⟨https://seekstorm.com/blog/1000x-spelling-correction/⟩.

GUPTA, A.; VEDALDI, A.; ZISSERMAN, A. Synthetic data for text localisation in natural images. In: *IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2016.

GUPTA, M.; SOENY, K. Algorithms for rapid digitalization of prescriptions. *Visual Informatics*, v. 5, n. 3, p. 54–69, set. 2021. ISSN 2468-502X. Disponível em: ⟨https://www.sciencedirect.com/science/article/pii/S2468502X21000334⟩.

HARTEL, M. et al. High incidence of medication documentation errors in a swiss university hospital due to the handwritten prescription process. *BMC health services research*, v. 11, p. 199, 08 2011.

HASSAN, E. et al. Medical Prescription Recognition using Machine Learning. In: *2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC)*. [S.l.: s.n.], 2021. p. 0973–0979.

HE, K. et al. Deep residual learning for image recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2016. p. 770–778.

KARATZAS, D. et al. Icdar 2015 competition on robust reading. In: *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*. [S.l.: s.n.], 2015. p. 1156–1160.

KASS, D.; VATS, E. Attentionhtr: Handwritten text recognition based on attention encoder-decoder networks. In: UCHIDA, S.; BARNEY, E.; EGLIN, V. (Ed.). *Document Analysis Systems*. Cham: Springer International Publishing, 2022. p. 507–522. ISBN 978-3-031-06555-2.

KULATHUNGA, D. et al. PatientCare: Patient Assistive Tool with Automatic Hand-written Prescription Reader. In: *2020 2nd International Conference on Advancements in Computing (ICAC)*. [S.l.: s.n.], 2020. v. 1, p. 275–280.

LEVENSHTEIN, V. I. et al. Binary codes capable of correcting deletions, insertions, and reversals. In: SOVIET UNION. *Soviet physics doklady*. [S.l.], 1966. v. 10, n. 8, p. 707–710.

LI, M. et al. *TrOCR: Transformer-based Optical Character Recognition with Pre-trained Models*. 2022.

LIAO, M. et al. Real-time scene text detection with differentiable binarization and adaptive scale fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE, 2022.

LIU, Y. et al. *RoBERTa: A Robustly Optimized BERT Pretraining Approach*. 2019.

MARTI, U.-V.; BUNKE, H. The IAM-database: an English sentence database for offline handwriting recognition. *International Journal on Document Analysis and Recognition*, v. 5, n. 1, p. 39–46, nov. 2002. ISSN 1433-2833. Disponível em: ⟨https://doi.org/10.1007/s100320200071⟩.

MURRAY, S. et al. Can you read this? legibility and hospital records: A multi-stakeholder analysis. *Clinical Risk*, v. 18, n. 3, p. 95–98, 2012. Disponível em: ⟨https://doi.org/10.1258/cr.2012.011065⟩.

NAJAFIRAGHEB, N.; HATAM, A.; HARIFI, A. An approach for handwritten prescribed medications detection using KNN. In: *6th International Conference on Electrical, Computer, Mechanical and Mechatronics Engineering (ICE2017)*. [S.l.: s.n.], 2017.

NETO, A. F. S. et al. HTR-Flor: a deep learning system for offline handwritten text recognition. In: *2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. Los Alamitos, CA, USA: IEEE Computer Society, 2020. (SIBGRAPI' 33), p. 54–61. Disponível em: ⟨https://doi.org/10.1109/SIBGRAPI51738.2020.00016⟩.

NGUYEN, T.-T.; NGUYEN, D.-V. V.; LE, T. Developing a Prescription Recognition System Based on CRAFT and Tesseract. In: NGUYEN, N. T. et al. (Ed.). *Computational Collective Intelligence*. Cham: Springer International Publishing, 2021. (Lecture Notes in Computer Science), p. 443–455. ISBN 978-3-030-88081-1.

OLEJNICZAK, K.; ŠULC, M. *Text Detection Forgot About Document OCR*. 2023.

POSTL, W. Detection of linear oblique structures and skew scan in digitized documents. In: *Proc. Int. Conf. on Pattern Recognition*. [S.l.: s.n.], 1986. p. 687–689.

ROSEBROCK, A. *Intersection over Union (IoU) for object detection*. nov. 2016. Disponível em: ⟨https://pyimagesearch.com/2016/11/07/intersection-over-union-iou-for-object-detection/⟩.

SCHEIDL, H. *Build a Handwritten Text Recognition System using TensorFlow*. jun. 2018. Disponível em: ⟨https://towardsdatascience.com/2326a3487cd5⟩.

SCHEIDL, H. *Handwritten Text Recognition in Historical Documents*. Dissertação (Mestrado) — TU Wien Faculty of Informatics, 2018.

SCHEIDL, H.; FIEL, S.; SABLATNIG, R. Word beam search: A connectionist temporal classification decoding algorithm. In: *2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR)*. [S.l.: s.n.], 2018. p. 253–258.

SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *arXiv 1409.1556*, 09 2014.

SáNCHEZ, J.-A. et al. Icfhr2014 competition on handwritten text recognition on transcriptorium datasets (htrts). In: . [S.l.: s.n.], 2014. v. 2014, p. 785–790.

VASWANI, A. et al. Attention is all you need. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates Inc., 2017. (NIPS'17), p. 6000–6010. ISBN 9781510860964.

VIGNESHWARAN, E.; SADIQ, M.; PRATHIMA, V. Assessment of completeness and legibility of prescriptions received at community pharmacies. *Journal of Health Research and Reviews*, v. 3, n. 2, p. 72–76, 2016. Disponível em: ⟨https://www.jhrr.org/article.asp?issn=2394-2010;year=2016;volume=3;issue=2;spage=72;epage=76;aulast=Vigneshwaran;t=6⟩.

WIJEWARDENA, W. R. a. D. Thesis, *Medical Prescription Identification Solution*. jul. 2021. Accepted: 2021-07-26T07:13:21Z. Disponível em: ⟨https://dl.ucsc.cmb.ac.lk/jspui/handle/123456789/4211⟩.

APÊNDICE A

Arquitetura do Sistema

Está em desenvolvimento um sistema protótipo que implementa a abordagem descrita nesta dissertação para uso como uma ferramenta de anotação de prescrições e como apoio a funcionários da farmácia parceira. Neste último caso, o objetivo maior é auxiliar no trabalho de reconhecimento de texto para facilitar a entrada de dados de prescrições médicas no sistema de gestão de informação utilizado pela farmácia, conhecido como *Fórmula Certa*[1]. A versão atual do protótipo, ilustrado na Figura A.1, tem interação com o banco de dados do *Fórmula Certa* para a consulta às imagens e aos dados das prescrições de pedidos já atendidos pela empresa. Ela também emprega um banco de dados auxiliar próprio para o armazenamento de dados de anotação das imagens de prescrições. Esses dados são usados em seguida para o treinamento dos modelos de aprendizado de máquina aplicados no próprio sistema.

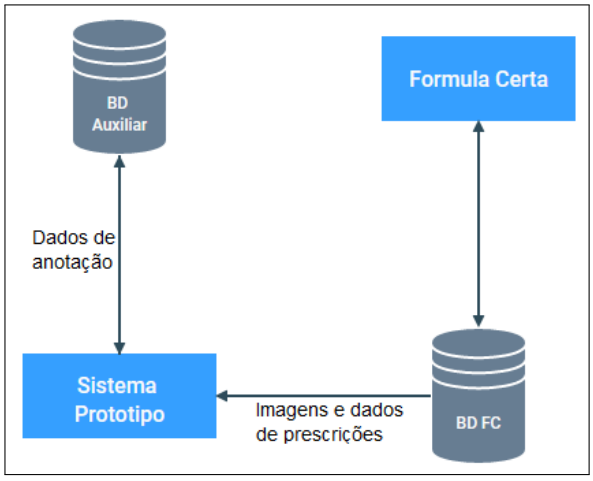


Figura A.1: *Diagrama de interação entre os sistemas envolvidos.*

As próximas seções descrevem elementos desse sistema.

[1] https://www.fagrontech.com.br/solucoes/formulacerta

A.1 Banco de dados

A versão do *Fórmula Certa* implantado nas lojas da farmácia parceira possui um banco de dados implementado usando o Sistema Gerenciador de Banco de Dados (SGBD) *Firebird* 2.5[2]. Considerando que o sistema protótipo é destinado ao uso pela farmácia, foi empregado o mesmo SGBD para o seu banco de dados auxiliar, com o intuito de reduzir o esforço de implantação. Este banco de dados auxiliar tem a finalidade de armazenar dados de anotações de imagens de prescrições médicas, assim como outros dados utilizados em interações com o banco de dados do *Fórmula Certa*. A Figura A.2 apresenta o Modelo Entidade-Relacionamento do banco de dados auxiliar.

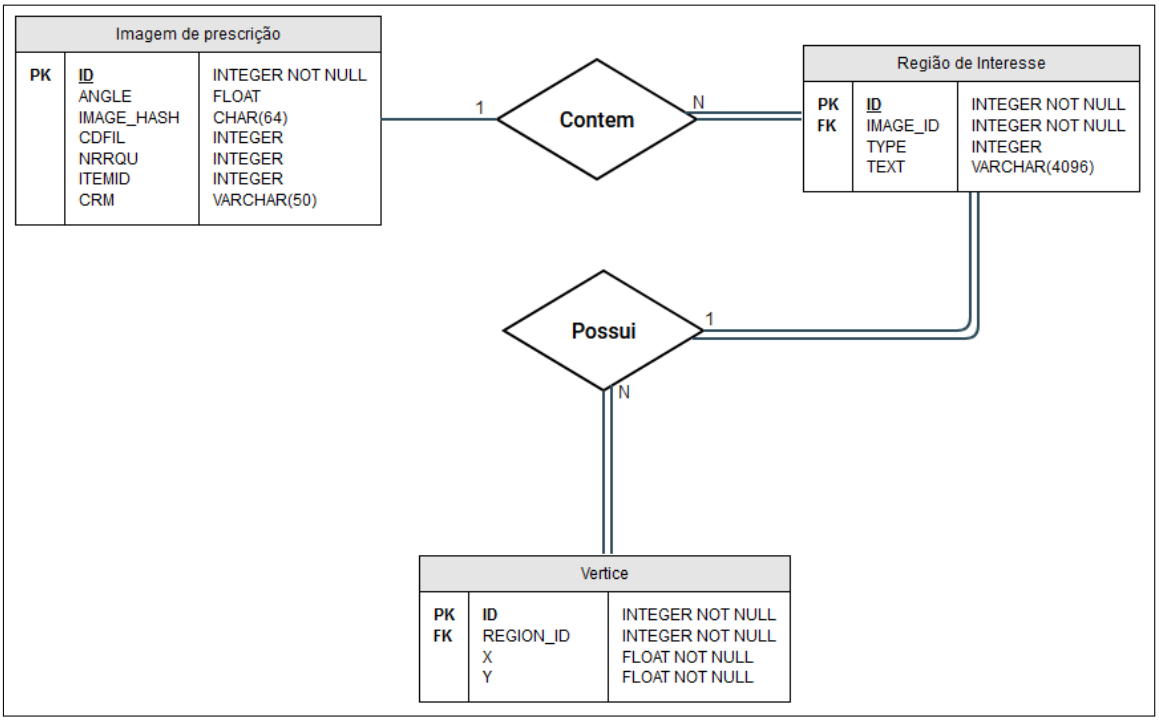


Figura A.2: *Modelo Entidade-Relacionamento do banco de dados auxiliar*

São descritos nas Tabelas A.1, A.2, A.3 os atributos de cada uma das entidades mostradas na Figura A.2.

[2]https://firebirdsql.org/

A entidade “Imagem de Prescrição” contém dados de imagens processadas pelo sistema protótipo. Ela inclui dados para identificar a imagem no banco de dados do *Fórmula Certa* e outros dados definidos pelo usuário na interface gráfica do protótipo.

Tabela A.1: *Atributos da entidade “Imagem de Prescrição”*

Nome	**Descrição**	**Tipo de dado**
ID	Chave primária. Identificador numérico da imagem	Número inteiro não nulo
ANGLE	Valor do ângulo de rotação da imagem definido pelo usuário na interface gráfica do protótipo	Número real
IMAGE_HASH	Hash SHA-256 da imagem. Usado primariamente para identificar imagens de prescrições que não estão contidas no BD do *Fórmula Certa*.	*String* de 64 caracteres
CDFIL	Código de filial da rede de lojas da farmácia. Usado para identificar imagens contidas no BD do *Fórmula Certa*.	Número inteiro
NRRQU	Número da requisição que contém a prescrição. Usado para identificar imagens contidas no BD do *Fórmula Certa*.	Número inteiro
ITEMID	Identificador numérico de uma prescrição na requisição. Usado para identificar imagens contidas no BD do *Fórmula Certa*.	Número inteiro
CRM	Número e sigla do estado do CRM do médico que escreveu a prescrição	*String* de tamanho variável

A entidade “Região de Interesse” representa as regiões detectadas pelo protótipo que contêm texto, especificamente palavras ou caracteres individuais. Entre os seus atributos, estão o texto equivalente contido na imagem, especificado pelo usuário ou reconhecido pelo protótipo, e o tipo de palavra que a região contém.

Tabela A.2: *Atributos da entidade “Região de Interesse”*

Nome	**Descrição**	**Tipo de dado**
ID	Chave primária. Identificador numérico da região de interesse.	Número inteiro não nulo
IMAGE_ID	Chave estrangeria da entidade “Imagem de Prescrição”	Número inteiro não nulo
TYPE	Numero que representa o tipo da palavra contida na região	Número inteiro
TEXT	Texto contido na região que foi anotado pelo usuário	*String* de tamanho variável

A entidade “Vértice” contém as coordenadas dos vértices dos polígonos que representam as regiões de texto detectadas pelo protótipo.

Tabela A.3: *Atributos da entidade “Vértice”*

Nome	**Descrição**	**Tipo de dado**
ID	Chave primária. Identificador numérico do vértice.	Número inteiro não nulo
REGION_ID	Chave estrangeria da entidade “Região de Interesse”	Número inteiro não nulo
X	Número com valor entre 0 e 1 que representa a coordenada X do vértice	Número real não nulo
Y	Número com valor entre 0 e 1 que representa a coordenada Y do vértice	Número real não nulo

Como mencionado anteriormente, o protótipo também tem interação com o banco de dados do *Fórmula Certa*. Como o protótipo continua em fase de desenvolvimento, está sendo usada uma cópia desse banco disponibilizada pela farmácia parceira e em operação no Laboratório Multiusuário de Computação de Alto Desempenho (LaMCAD) da UFG. A cópia do banco de dados foi disponibilizada mediante a assinatura de um termo de confidencialidade entre as partes envolvidas no desenvolvimento do projeto e, portanto, sua estrutura interna não será descrita neste documento.

A.2 Detalhes de implementação

O sistema protótipo é uma aplicação *desktop* com interface gráfica, denominada *Sistema de Reconhecimento de Prescrições Médicas* (SisRPM). Esta aplicação está sendo

desenvolvida utilizando a linguagem de programação *Python* 3.10[3], com uma interface gráfica baseada no *framework Qt 6*[4]. A aplicação tem suporte aos sistemas operacionais *Windows* e *Linux*. O diagrama na Figura A.3 ilustra os módulos que compõem a aplicação, cada um responsável por uma de suas funcionalidades.

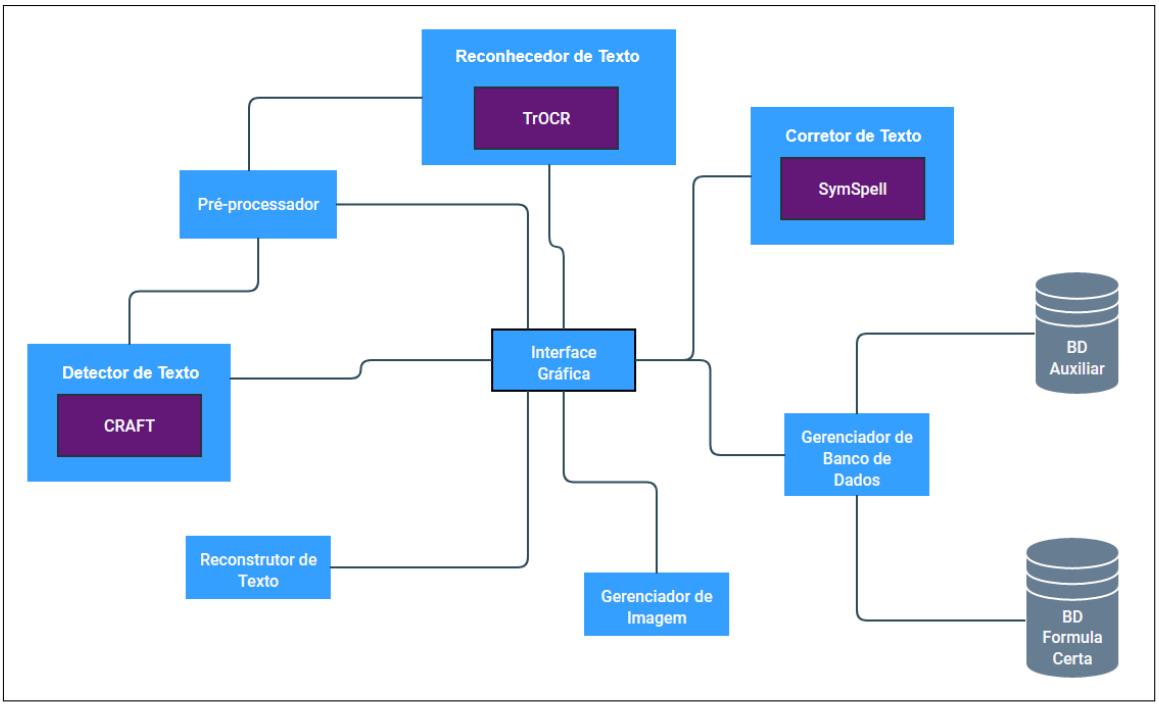


Figura A.3: *Diagrama de módulos do sistema protótipo (SiSRPM).*

A.2.1 Interface Gráfica

Este é o modulo central, responsável por exibir todos os elementos interativos da aplicação e tratar as ações do usuário, geralmente por meio de solicitações ou envios de dados e comandos aos outros módulos. A interface consiste em uma janela principal, divida em uma área esquerda dedicada ao módulo responsável por exibir as imagens de prescrições e gerenciar as interações com as ferramentas de anotação, descrito na Subseção A.2.3, e uma área direita que exibe múltiplos campos de texto interativos ao usuário. A janela principal também dispõe de uma barra de menu, a qual permite o acesso às sub-janelas de abertura de imagens e de conexão aos bancos de dados, assim como uma barra de ferramentas de anotação e de manipulação de imagens. Essa janela principal pode ser vista na Figura A.4.

[3]https://www.python.org/
[4]https://www.qt.io/

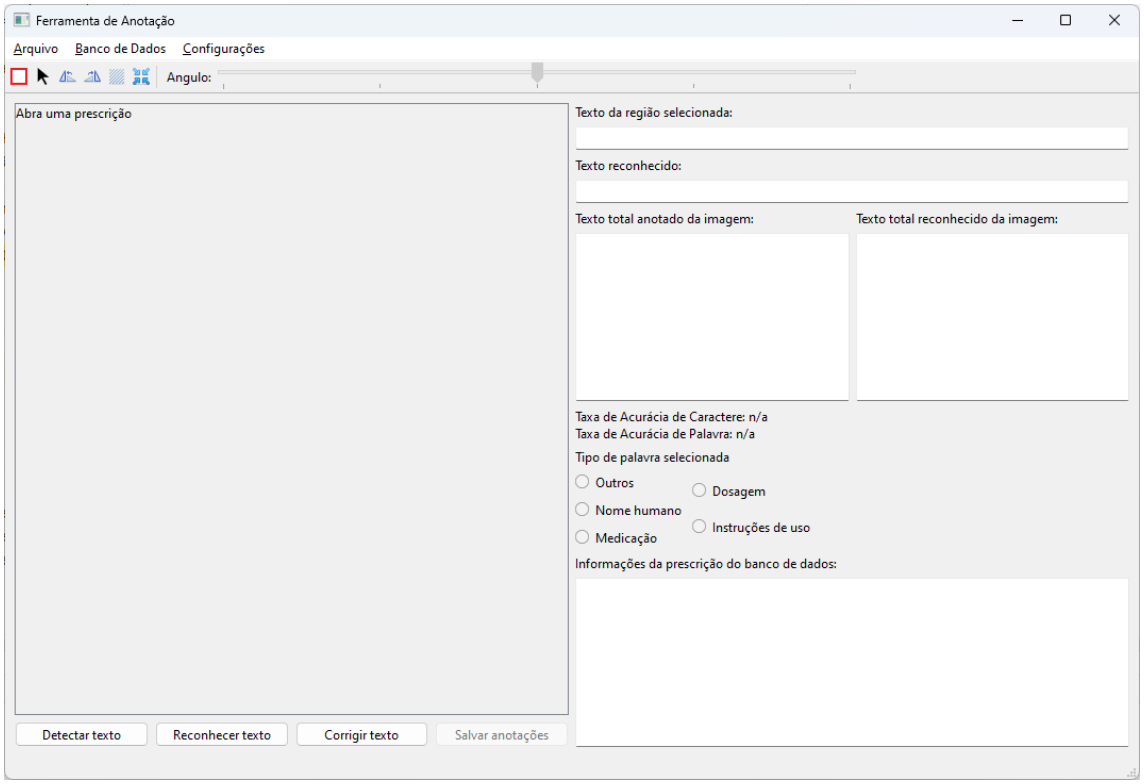


Figura A.4: *Janela principal do sistema protótipo.*

Este modulo é responsável por toda a comunicação entre os outros módulos, com exceção do módulo pré-processador, que também interage diretamente com os módulos de detecção e reconhecimento de texto.

A.2.2 Gerenciador de Banco de Dados

Este módulo gerencia a conexão com o banco de dados auxiliar (BD-A) e o banco de dados do *Fórmula Certa* (BD-FC). Ele atende solicitações da interface gráfica buscando imagens de prescrições e dados correspondentes diretamente do BD-FC, e também envia e busca dados de anotação ao BD-A. A interface gráfica dispõe de duas sub-janelas que interagem diretamente com este módulo, mostradas na Figura A.5.

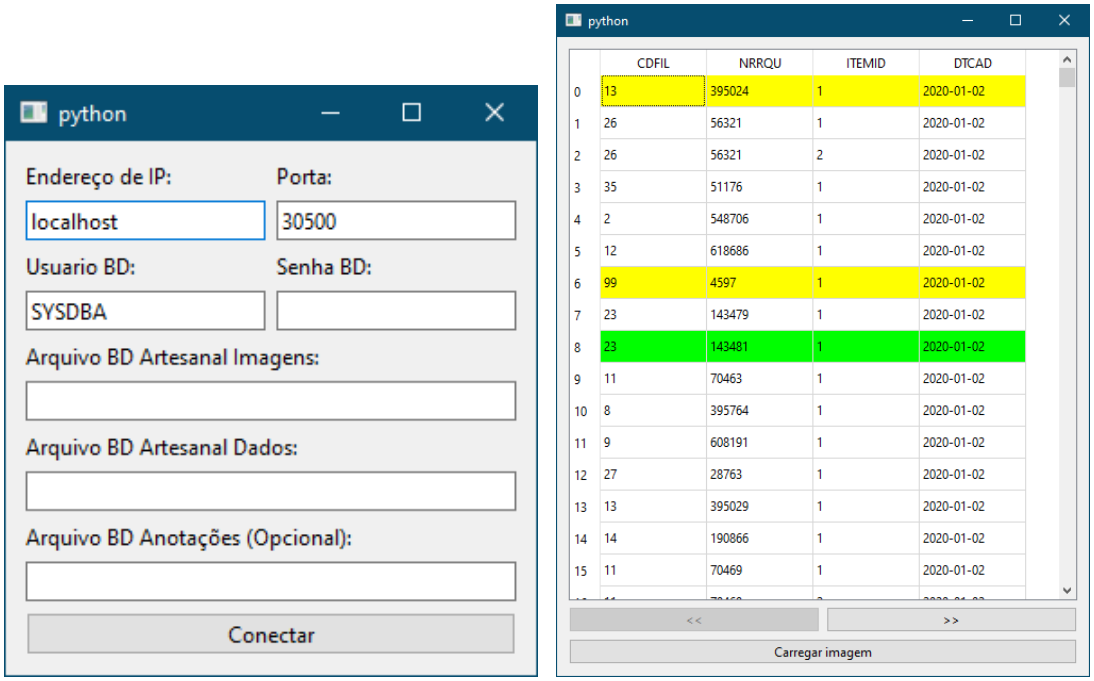


Figura A.5: *Janelas de conexão com os bancos de dados e de carregamento de imagens de prescrições do BD-FC.*

A.2.3 Gerenciador de Imagem

Este modulo é responsável por exibir as imagens de prescrições carregadas pela aplicação. Ele gerencia a imagem aberta e todos os dados relacionados a ela, incluindo os polígonos que representam regiões de texto detectadas pelo módulo detector ou definidas pelo usuário com a ferramenta de criação. Este módulo implementa uma série de ferramentas de anotação, as quais podem ser vistas na Figura A.6.

Figura A.6: *Barra de ferramentas de anotação e manipulação de imagens*

1. **Ferramenta de criação -** Permite que o usuário crie manualmente polígonos de regiões de texto.
2. **Ferramenta de seleção -** Permite que o usuário selecione e manipule regiões de texto existentes. O usuário pode selecionar regiões, apagá-las, movê-las e deslocar seus vértices para mudar seus formatos.
3. **Rotacionar à esquerda -** Rotaciona a imagem em 90 graus para a esquerda.
4. **Rotacionar à direita -** Rotaciona a imagem em 90 graus para a direita.
5. **Esconder regiões -** Esconde os polígonos de regiões de texto existentes para facilitar a visualização da imagem

6. **Ajustar imagem a tela -** O usuário pode aplicar um *zoom-in* e *zoom-out* na imagem usando o *scroll* do mouse. Esta ferramenta redefine o nível atual de *zoom* para que toda a imagem fique visível dentro do campo de visualização.
7. **Ajuste de ângulo -** Permite que o usuário ajuste manualmente o ângulo de rotação da imagem.

A.2.4 Detector de Texto

Este módulo recebe solicitações da interface gráfica para detectar automaticamente regiões de texto na imagem aberta. Ele encaminha a imagem ao modelo *CRAFT*, descrito na Seção 4.1, e retorna à interface gráfica uma lista de polígonos de regiões de texto detectadas, que são então encaminhadas ao módulo gerenciador de imagem para que sejam exibidas na tela.

A.2.5 Reconhecedor de Texto

Este módulo recebe comandos da interface gráfica para reconhecer o texto da região atualmente selecionada, ou de todas as regiões existentes se nenhuma estiver selecionada. Ele encaminha imagens recortadas das regiões ao modelo de reconhecimento de texto *TrOCR*, descrito na Seção 5.1, e retorna o texto reconhecido para que este possa ser exibido nos campos de texto interativos da janela principal.

A.2.6 Corretor de Texto

Similarmente ao reconhecedor de texto, este módulo recebe solicitações da interface gráfica para corrigir o texto reconhecido da região atualmente selecionada, ou de todas as regiões existentes se nenhuma estiver selecionada. Ele invoca o método *SymSpell*, descrito na Seção 6.1, que corrige palavras utilizando o dicionário de frequência descrito na Seção 3.3, e retorna para a interface gráfica as palavras corrigidas.

A.2.7 Reconstrutor de Texto

Este módulo é invocado pela interface gráfica sempre que é detectada uma mudança no texto reconhecido ou manualmente anotado das regiões de texto existentes. Ele recebe os dados de todas as regiões de texto e aplica o algoritmo descrito a seguir para reconstruir o texto total reconhecido ou anotado da imagem, que é retornado a interface para ser exibido ao usuário.

As abordagens de segmentação e de reconhecimento de texto implementadas no protótipo estão operando a nível de segmentos de palavras individuais, de forma que se

mostra necessário um método para reconstruir as linhas de texto a partir das regiões identificadas para cada palavra detectada. Assim, foi desenhado e implementado um método para realizar tal tarefa, o qual é descrito no Algoritmo 1, com um Algoritmo 2 complementar, a seguir. A entrada do Algoritmo 1 é uma lista V de polígonos representando regiões de texto segmentados da prescrição. Considera-se aqui que os polígonos são retângulos e que seus lados são paralelos aos eixos X,Y da imagem. Caso isso não seja verdadeiro para um dado polígono, então ele é substituído pelo menor retângulo com lados paralelos aos eixos X,Y que o envolve. Para cada retângulo $u \in V$, assume-se ainda que é possível obter as suas menores e maiores coordenadas x e y usando as notação $u.minX$, $u.maxX$, $u.minY$ e $u.maxY$. O sistema de coordenadas tem origem no canto superior esquerdo da imagem e cresce em Y para baixo e em X para direita.

Algoritmo 1 Algoritmo de reconstrução de texto

Entrada: ***V*** - lista de polígonos de regiões de texto

Saída: *resultado* - Texto total reconstruído

início

```
E ← {}
G ← (V, E)
para cada par u, v em V faça
    se u não é um retângulo então
        u ← retangulo_delimitador(u)
    se v não é um retângulo então
        v ← retangulo_delimitador(v)
    se u.minY ≤ v.maxY e v.minY ≤ u.maxY então
        se u.minY ≤ v.minY então
            p_y ← u.maxY - v.minY
        senão
            p_y ← v.maxY - u.minY
        se u.minX ≤ v.minX então
            crie aresta horizontal u → v em E com peso p_y
        senão
            crie aresta horizontal v → u em E com peso p_y
    se u.minX ≤ v.maxX e v.minX ≤ u.maxX então
        se u.minX ≤ v.minX então
            p_x ← u.maxX - v.minX
        senão
            p_x ← v.maxX - u.minX
        se u.minY ≤ v.minY então
            crie aresta vertical u → v em E com peso p_x
        senão
            crie aresta vertical v → u em E com peso p_x
lista_linhas ← []
repita
    Seja v o vértice em V que não possui arestas de entrada e que tem menor minY
    L ← construir_L(G, v)
    Adicione L ao final de lista_linhas
    Remova de V os vértices em L
    Remova de E as arestas conectadas aos vértices em L
até V ser vazio;
Seja resultado uma String vazia
para cada L em lista_linhas em ordem faça
    para cada v em L faça
        resultado ← resultado+ "    " +v.texto
    resultado ← resultado+ "\n"
retorne resultado
```

Algoritmo 2 construir_L

Entrada: ***G*** - Grafo ***G***(***V***, ***E***), **v** - Vértice em ***V***

Saída: ***L*** - Lista de vértices em ***V*** que formam uma linha

início

Seja ***S*** uma lista com todos os vértices em ***V*** acessíveis seguindo um caminho direcionado a partir de v percorrendo somente arestas horizontais, onde, para cada aresta $u \rightarrow w$ do caminho, $peso(u \rightarrow w)/min(u.maxY - u.minY, w.maxY - w.minY) \geq 0.6$

Adicione v à S

Ordene S em ordem crescente usando minX como critério

Seja ***i*** a posição de v em S

repita

Adicione S$[i]$ ao final de L

Remova o elemento na posição ***i*** de S

até $i \geq len(S)$;

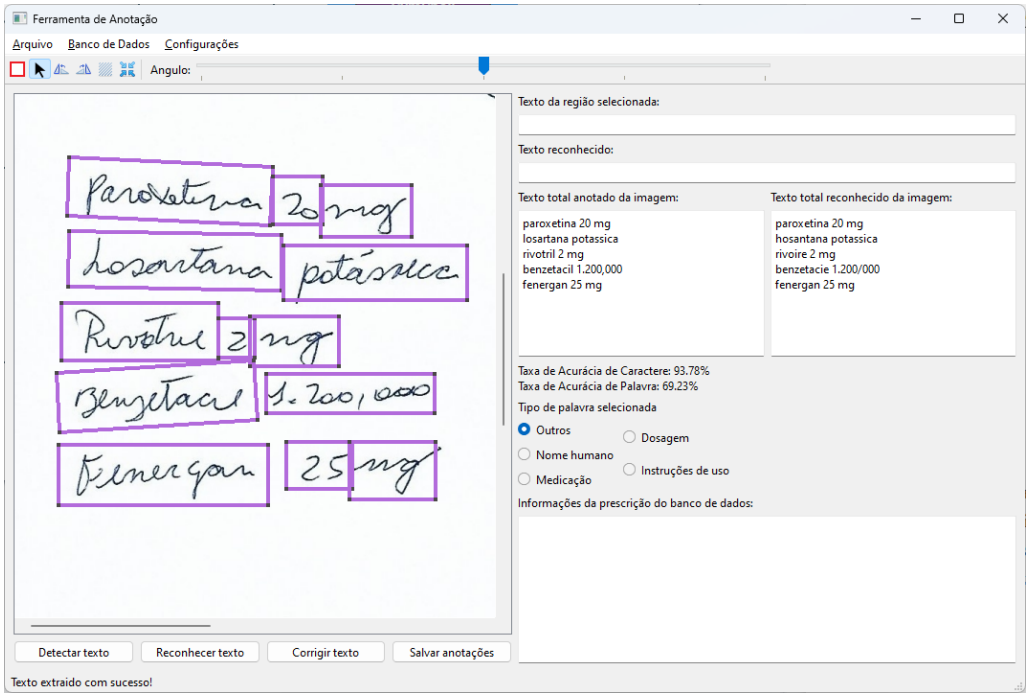


Figura A.7: *Exemplo de reconstrução de texto bem sucedida*

Tendo em vista as limitações do método atualmente implementado, pretende-se explorar outras abordagens para realizar a reconstrução do texto. Uma possível abordagem sendo investigada é uma modificação do algoritmo atual, que cria as arestas entre as regiões de texto utilizando segmentos de reta projetados a partir dos polígonos originais em vez de retângulos delimitadores.

Parecer do Comitê de Ética

UFG - UNIVERSIDADE FEDERAL DE GOIÁS

PARECER CONSUBSTANCIADO DO CEP

DADOS DA EMENDA

Título da Pesquisa: SISTEMA PARA EXTRAÇÃO AUTOMATIZADA DE DADOS DE PRESCRIÇÕES MÉDICAS
Pesquisador: ANDRE PIRES CORREA
Área Temática:
Versão: 2
CAAE: 54808021.3.0000.5083
Instituição Proponente: Instituto de Informática
Patrocinador Principal: CONS NAC DE DESENVOLVIMENTO CIENTIFICO E TECNOLOGICO

DADOS DO PARECER

Número do Parecer: 5.395.556

Apresentação do Projeto:
Essa versão se trata de um pedido de emenda para extensão de cronograma.

Título da Pesquisa: SISTEMA PARA EXTRAÇÃO AUTOMATIZADA DE DADOS DE PRESCRIÇÕES MÉDICAS
Pesquisador Responsável: ANDRE PIRES CORREA
Área Temática:
Versão: 2
CAAE: 54808021.3.0000.5083
Submetido em: 04/05/2022
Instituição Proponente: Instituto de Informática
Situação da Versão do Projeto: Em relatoria
Localização atual da Versão do Projeto: UFG - Universidade Federal de Goiás
Patrocinador Principal: CONS NAC DE DESENVOLVIMENTO CIENTIFICO E TECNOLOGICO

Farmácias de manipulação trabalham diariamente com grandes volumes de pedidos de medicações a serem manipuladas. Para facilitar o gerenciamento destes pedidos, é comum a utilização de sistemas computacionais de gestão de informação, nos quais os dados das prescrições médicas de cada pedido são registrados e utilizados para guiar o processo de manipulação até a entrega do

Endereço: Alameda Flamboyant, Qd. K, Edifício K2, sala 110
Bairro: Campus Samambaia, UFG **CEP:** 74.690-970
UF: GO **Município:** GOIANIA
Telefone: (62)3521-1215 **E-mail:** cep.prpi@ufg.br

medicamento ao cliente. Essa tarefa manual de introduzir dados nos sistemas requer grande tempo e esforço por parte dos farmacêuticos pois necessita que seja feita a interpretação de prescrições médicas, sendo que uma considerável parcela delas costuma ser escrita a mão por médicos com caligrafia de baixa legibilidade. Além disso, as prescrições devem passar por uma etapa de verificação a fim de garantir que não existam irregularidades nos compostos ou nas doses das formulas prescritas, algo que também requer tempo e conhecimento altamente especializado. Uma possível solução para tornar esse processo mais eficiente é a adoção de métodos automatizados de extração de dados a partir de fotos ou imagens escaneadas de prescrições médicas. Em um levantamento bibliográfico preliminar realizado como parte deste projeto, foram identificados múltiplos trabalhos que abordam este problema (NAJAFIRAGHEB;HATAM; HARIFI, 2017; CHUMUANG; KETCHAM, 2018; FAJARDO et al., 2019; BUTALA et al., 2020; KULATHUNGA et al., 2020; HASSAN et al., 2021; GUPTA; SOENY, 2021). No entanto, as taxas de acerto dos métodos utilizados nesses trabalhos ainda são muito baixas para prescrições manuscritas (na faixa de ~30% a ~80%) e, em geral, eles foram testados com
uma quantidade pequena de casos (entre algumas dezenas a poucos milhares). Outro aspecto observado é que os trabalhos frequentemente focam em poucas etapas do processo de automação da análise de prescrições médicas; em alguns casos, eles utilizam ferramentas proprietárias, o que dificulta a replicação e a análise científica dos resultados.

Como exemplo de trabalho nessa área, Hassan et al. (2021) apresentam um sistema que extrai os nomes de medicamentos a partir de fotos de prescrições médicas tiradas por smartphones. As fotos são pré processadas e segmentadas em três partes: um cabeçalho que contém o nome e a especialização do médico, o meio da prescrição que contém as medicações prescritas, e o rodapé que traz o endereço e dados de contato da clínica ou hospital. O rodapé é descartado e o cabeçalho é usado para identificar a especialização do médico que será usada para melhorar a classificação das medicações. As palavras manuscritas no meio da prescrição são segmentadas, e são então classificadas por uma Rede Neural Convolucional.

O banco de dados usado foi composto de imagens de prescrições de médicos de múltiplos hospitais e várias especializações, mas o numero de imagens que compõem este banco não foi especificado. Foi obtido 73% de acurácia de treino e apenas 50% de acurácia de teste.

Em outro exemplo, Gupta e Soeny (2021) descrevem uma abordagem para detectar o nome de medicações em imagens de prescrições médicas impressas e manuscritas baseada no uso das ferramentas de reconhecimento óptico de caracteres (OCR) da API proprietária de visão computacional da Google, denominada Google Vision. As imagens são submetidas ao processo de

OCR usando a API, a qual retorna um conjunto de palavras reconhecidas junto às coordenadas que indicam suas posições na imagem. Esses dados são usados para identificar agrupamentos de palavras que estejam próximas entre si na imagem e que contenham termos que são comumente vistos acompanhando nomes de medicações (ex: comprimidos, gotas). Após identificados estes agrupamentos, palavras que existam em dicionários de inglês são descartadas. No caso de prescrições impressas, as palavras remanescentes são apresentadas como sendo os nomes das medicações. Já no caso de prescrições manuscritas, é feita uma comparação aproximada das palavras com um banco de nomes de medicamentos usando distância de DemerauLevenshtein e as palavras mais parecidas são apresentadas. A abordagem foi avaliada usando 5,176 prescrições impressas e 5,000 manuscritas, e foi obtido um F-Score de 79.4% para prescrições impressas e apenas 26.2% para prescrições manuscritas.

Metodologia

O processo para extração automatizada de dados a partir de fotos ou imagens de prescrições médicas pode ser descrito em quatro grandes etapas:

1. Pré-processamento – aplicação de filtros e técnicas de processamento de imagens digitais para reduzir ruídos e facilitar as próximas etapas do processo.
2. Segmentação – identificação e segmentação das regiões de interesse da imagem. As regiões da imagem que contêm informações relevantes, como o nome do paciente, do médico e os medicamentos prescritos são identificadas e as linhas de texto a serem reconhecidas são segmentadas.
3. Reconhecimento de texto – extração de texto digital a partir das regiões segmentadas usando técnicas de Aprendizado de Máquina (AM). As palavras segmentadas na etapa anterior são classificadas por métodos de AM treinados para reconhecer a imagem manuscrita e o texto digital correspondente é gerado.
4. Processamento do texto – processar o texto obtido na etapa anterior usando técnicas de Processamento de Linguagem Natural para categorizar as palavras reconhecidas, de forma que seja possível identificar dados como nomes de medicações/compostos, dosagem, instruções de uso, etc. No caso específico das farmácias de manipulação, há também a necessidade de se avaliar se os compostos prescritos para manipulação podem ser combinados e nas dosagens solicitadas.

Para o presente projeto, pretende-se investigar algoritmos e desenvolver um protótipo funcional de uma ferramenta que implemente todas as etapas descritas anteriormente, do processo de

análise automática de prescrições médicas para uma farmácia de manipulação. O projeto será executado na forma de série de tarefas, as quais estão descritas a seguir:

• Expansão e atualização da revisão bibliográfica. Será realizada uma expansão inicial do levantamento bibliográfico já feito, para melhor mapear o estado atual da pesquisa na área. A revisão também será atualizada periodicamente a cada seis meses para acompanhar os últimos desenvolvimentos no tema.

• Organização e preparação dos dados cedidos pela empresa parceira. Os dados necessários para treinar e validar os métodos de Aprendizado de Máquina a serem aplicados no processo de reconhecimento de texto serão providos pela 9 farmácia. Estes dados consistem de imagens de prescrições médicas associadas às informações previamente extraídas, como os nomes do paciente e do médico e as medicações prescritas, entre outros. Tais dados serão organizados em um banco de dados próprio para o projeto e ampliados com as posições das regiões nas imagens onde as informações estão, a serem anotadas manualmente.

• Desenvolvimento da primeira versão do protótipo. Será desenvolvida uma versão do sistema protótipo com uma arquitetura de software que servirá de base para o restante do projeto. Essa versão terá a funcionalidade de importar imagens de prescrições e aplicar uma solução de reconhecimento óptico de caracteres existente para gerar um resultado textual bruto.

• Estudo e implementação da detecção automática das regiões de interesse.

Serão investigados e implementados dentro do protótipo métodos para reconhecer automaticamente as regiões de interesse das imagens prescrições. Também será implementada uma funcionalidade para que o usuário possa manualmente ajustar essas regiões. Experimentos serão realizados e os resultados utilizados para avaliar e, se possível, melhor calibrar os métodos.

• Refinamento do reconhecimento e implementação de métodos de categorização de texto. Serão escolhidos e implementados no protótipo dois métodos especializados em reconhecimento de texto identificados na literatura científica sobre análise automática de prescrições médicas. Também será desenvolvido um método para categorização do texto. Experimentos serão realizados para identificar as vantagens e desvantagens desses métodos para o contexto em foco.

• Módulo de regras de compatibilidade de compostos e dosagens. Será implementado um módulo de regras para detecção de incompatibilidades entre compostos médicos e verificação de suas doses. Este modulo será estruturado de forma que os usuários do sistema possam inserir regras a serem verificadas automaticamente após o processo de extração de dados das imagens.

• Desenvolvimento da versão final do protótipo. Serão investigadas possibilidades de combinação dos métodos de reconhecimento de texto previamente implementados, com o objetivo de

aumentar a qualidade do processo. Essa combinação será implementada e testada dentro do protótipo, permanecendo no sistema como uma versão final, caso apresente resultados positivos.

• Escrita de artigos, relatórios e monografias. Serão elaborados relatórios técnicos e artigos científicos para documentar os resultados obtidos ao longo do projeto.

Também serão escritos documentos destinados às atividades avaliativas da Universidade, incluindo um texto para o processo de qualificação e a dissertação para a defesa de mestrado, bem como relatórios de acompanhamento de pesquisa.

• Etapas avaliativas da instituição. Serão apresentados os resultados obtidos pelo projeto durante duas etapas avaliativas, de qualificação e de defesa de dissertação de mestrado.

O projeto será realizado principalmente por um aluno de mestrado e por um aluno de graduação (este, ainda a ser selecionado), ambos com bolsa do Programa MAI/DAI, sob orientação dos demais pesquisadores e com apoio de profissionais da Gyntec e da empresa farmacêutica.

Objetivo da Pesquisa:

O objetivo principal deste projeto de pesquisa é o desenvolvimento de uma abordagem que realize um processo completo de extração e processamento automatizados de informações a partir de imagens de prescrições médicas no contexto de uma farmácia de manipulação. Para tal, serão estudadas, implementadas e comparadas pelo menos duas abordagens existentes na literatura para etapas específicas desse processo e investigadas possibilidades de melhorias que possam ser realizadas para atingir um nível de desempenho superior ao atual. O desenvolvimento de um sistema protótipo totalmente funcional é um objetivo específico do projeto, dada a sua natureza de contribuição para o setor produtivo.

Avaliação dos Riscos e Benefícios:

Riscos:

Por se tratar de um projeto que utilizará uma base de dados já existente, consistindo de imagens de prescrições médicas de pedidos passados de uma farmácia de manipulação, a pesquisa não apresenta riscos físicos aos seus participantes. O único risco considerado no projeto seria o possível vazamento de informações pessoais dos pacientes e médicos que estão nas prescrições. No entanto, essas informações servem exclusivamente como entrada para o treinamento dos modelos preditivos, sendo reportadas apenas informações estatísticas gerais sobre o desempenho desses modelos. Além disso, mecanismos de criptografia e firewall de rede serão implantados para preservar os dados armazenados no banco de dados durante o desenvolvimento da pesquisa.

Benefícios:

A abordagem a ser desenvolvida durante o projeto poderá ser aplicada de forma a auxiliar farmacêuticos no processo de interpretação e extração de dados de prescrições médicas, permitindo maior segurança e celeridade no atendimento dos pedidos de seus clientes.

Comentários e Considerações sobre a Pesquisa:

Os pesquisadores propõem dispensa do TCLE, justificando que a pesquisa será desenvolvida utilizando uma base de dados de uma farmácia de manipulação, composta por imagens escaneadas de prescrições médicas e de informações textuais extraídas delas. Esses dados foram coletados e armazenados pela própria farmácia no decorrer dos anos para viabilizar o atendimento de seus pedidos. Os dados serão disponibilizados aos pesquisadores, pela própria farmácia, única e exclusivamente para o desenvolvimento dos algoritmos de aprendizado de máquina visando a extração automatizada de informações textuais a partir das imagens das prescrições. Os pesquisadores estarão sujeitos a um termo de confidencialidade como condição para o acesso a esses dados.

A base de dados ficará armazenada em uma máquina servidora do Laboratório Multiusuário de Computação de Alto Desempenho (LaMCAD) da UFG. Serão empregados, durante a pesquisa, mecanismos de proteção como firewall de rede e criptografia para garantir o acesso restrito aos dados. Em termos de publicação dos resultados da pesquisa, serão apresentados apenas dados estatísticos sobre a base utilizada, como o número de prescrições que a compõem e os métodos de aprendizado de máquina treinados que não permitam recuperação da informação original por engenharia reversa. A base de dados será mantida por um prazo de cinco anos após a publicação dos trabalhos e será posteriormente apagada. Considerando que a pesquisa não envolve interação direta com seres humanos, que a base de dados a ser utilizada já foi coletada e é de responsabilidade da própria farmácia parceira, que o projeto de pesquisa será voltado ao beneficio dessa farmácia, que as informações da base serão mantidas em sigilo pelos pesquisadores e que os resultados de pesquisa científica a serem gerados não possuem informações de identificação de pacientes ou médicos, sendo apenas algoritmos, modelos de aprendizado de máquina treinados e dados estatísticos.

Consideramos os riscos de perda de sigilo e confidencialidade dos dados de pacientes e médicos, mas que os pesquisadores estão atentos e apresentam medidas para mitigar esses riscos.

A justificativa apresentada para a emenda foi: "Esta solicitação de emenda se dá pela necessidade

de estender o prazo do projeto de pesquisa, devido a inclusão do aluno bolsista de IT previsto para iniciar suas atividades em setembro de 2022, com conclusão em agosto de 2023 e uma possível extensão por outros 12 meses. Tendo isto em vista, o prazo do projeto
foi estendido até setembro de 2024, e novas atividades que serão desempenhadas pelo bolsista de IT neste período foram incluídas no cronograma."

O cronograma foi atualizado.

Considerações sobre os Termos de apresentação obrigatória:

Foram adequadamente apresentados: folha de rosto datada e assinada pelo diretor do Instituto de Informática da UFG, Anuência da Farmácia artesanal.

Conclusões ou Pendências e Lista de Inadequações:

Este estudo não apresenta óbices éticos, assim como o presente pedido de emeda.

O pesquisador responsável, após seleção do aluno bolsista de TI, deverá solicitar outra emenda para inclusão do mesmo como membro de equipe e apresentar o respectivo Termo de Compromisso antes que ele inicie suas atividades na pesquisa.

Considerações Finais a critério do CEP:

Informamos que o Comitê de Ética em Pesquisa/CEP-UFG considera a presente solicitação de Emenda APROVADA, pois a mesma foi considerada em acordo com os princípios éticos vigentes. Reiteramos a importância deste Parecer Consubstanciado, e lembramos que o(a) pesquisador(a) responsável deverá encaminhar ao CEP-UFG o Relatório Final baseado na conclusão do estudo e na incidência de publicações decorrentes deste, de acordo com o disposto na Resolução CNS n. 466/12 e Resolução CNS n. 510/16. O prazo para entrega do Relatório é de até 30 dias após o encerramento da pesquisa previsto para setembro de 2024.

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_1942174_E1.pdf	04/05/2022 19:44:54		Aceito
Outros	carta_justificativa.pdf	04/05/2022 19:42:51	ANDRE PIRES CORREA	Aceito
Outros	Relatorio_parcial.pdf	04/05/2022 19:41:35	ANDRE PIRES CORREA	Aceito

Projeto Detalhado / Brochura Investigador	Projeto_de_Pesquisa__Farmacia___Comite_de_Etica___Cronograma_Atualizado___Estendido_.pdf	04/05/2022 19:37:14	ANDRE PIRES CORREA	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	Dispensa_TCLE.pdf	04/01/2022 10:44:10	ANDRE PIRES CORREA	Aceito
Outros	termo_de_anuencia.jpg	31/12/2021 16:23:32	ANDRE PIRES CORREA	Aceito
Outros	Termo_Compromisso.pdf	31/12/2021 16:22:53	ANDRE PIRES CORREA	Aceito
Folha de Rosto	folhaDeRosto.pdf	31/12/2021 16:19:04	ANDRE PIRES CORREA	Aceito

Situação do Parecer:
Aprovado

Necessita Apreciação da CONEP:
Não

GOIANIA, 09 de Maio de 2022

Assinado por:
Rosana de Morais Borges Marques
(Coordenador(a))