

Universidade FUMEC Faculdade de Ciências Empresariais  
Programa de Pós-Graduação em Sistemas de Informação e Gestão do  
Conhecimento

**Reconhecimento de Patologias em Guias de  
Solicitações Médicas Utilizando Aprendizado de  
Máquina**

**Lucas Moura Belo**

Belo Horizonte

2021

**Lucas Moura Belo**

**Reconhecimento de Patologias em Guias de Solicitações Médicas Utilizando Aprendizado de Máquina**

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Sistemas de Informação e Gestão do Conhecimento da Universidade FUMEC como parte dos requisitos para a obtenção do título de Mestre em Sistemas de Informação e Gestão do Conhecimento.

**Área de concentração:** Gestão de Sistemas de Informação e do Conhecimento.

**Linha de pesquisa:** Tecnologia e Sistemas de Informação.

**Orientador:** Prof. Dr. Fernando Silva Parreiras

Belo Horizonte

2021

### **Dados Internacionais de Catalogação na Publicação (CIP)**

B452r Belo, Lucas Moura, 1991-  
Reconhecimento de patologias em guias de solicitações  
médicas utilizando aprendizado de máquina / Lucas Moura Belo.  
- Belo Horizonte, 2021.  
87 f. : il.

Orientador: Fernando Silva Parreiras  
Dissertação (Mestrado em Sistemas de Informação e  
Gestão do Conhecimento), Universidade FUMEC, Faculdade de  
Ciências Empresariais, Belo Horizonte, 2021.

1. Operadoras de planos de saúde. 2. Inteligência artificial.  
3. Redes neurais (Computação). 4. Aprendizado do computador.  
I. Título. II. Parreiras, Fernando Silva. III. Universidade FUMEC,  
Faculdade de Ciências Empresariais.

CDU: 004.8

Dissertação intitulada “**Reconhecimento de patologias em guias de solicitações médicas utilizando aprendizado de máquina**” de autoria de Lucas Moura Belo, aprovada pela banca examinadora constituída pelos seguintes professores:

Prof. Dr. Fernando Silva Parreiras – Universidade FUMEC  
(Orientador)

Prof. Dr. João Victor Boechat Gomide – Universidade FUMEC  
(Examinador Interno)

Profª. Dra. Zilma Silveira Nogueira Reis – UFMG  
(Examinador Externo)

Prof. Dr. Fernando Silva Parreiras  
Coordenador do Programa de Pós-Graduação em Sistemas de Informação e Gestão do  
Conhecimento da Universidade FUMEC

Belo Horizonte, 26 de março de 2021.

*Professora Zilma Silveira Nogueira Reis*

*Fernanda Silva Parreiras*

*Professor João Victor Boechat Gomide*

 REQUESTED	TITLE	<b>Assinatura de ata e contra-capas Universidade</b>
	FILE NAME	<b>e1acfded-7f8c-4fc9-b375-6905e0e680fc.pdf</b>
	REQUEST ID	<b>signature_request_51c6c94e-608e-404b-93c4-d63c7</b>
	REQUESTED BY	<b>Karem Estefani Oliveira De Paula</b>
	STATUS	<b>● Completed</b>

Professor (jvictor@fumec.br)

 SENDED	16/09/2021 20:26:51UTC±0	 SIGNED	16/09/2021 20:27:15UTC±0 186.214.218.32
---	-----------------------------	---	---

Professor (zilma.medicina@gmail.com)

 SENDED	16/09/2021 20:35:29UTC±0	 SIGNED	16/09/2021 20:35:52UTC±0 177.134.100.187
---	-----------------------------	---	--

Professor (fernando.parreiras@fumec.br)

 SENDED	05/10/2021 20:57:51UTC±0	 SIGNED	05/10/2021 20:58:00UTC±0 187.111.30.10
---	-----------------------------	---	--

 COMPLETED	05/10/2021 20:58:00 UTC±0 The document has been completed.
--	--

**Assinado por:FERNANDO SILVA PARREIRAS:03073186646**

Data do Certificado:09/12/2020 15:03:25

Data de Validade:08/12/2025 15:03:25

Data da Assinatura:25/11/2021 16:37:42

**REQUERENTE:**  
**CN=FERNANDO SILVA PARREIRAS:03073186646**  
**OU=(em branco)**  
**OU=RFB e-CPF A3**  
**OU=Secretaria da Receita Federal do Brasil - RFB**  
**OU=01160628000117**  
**OU=Presencial**  
**O=ICP-Brasil**  
**C=BR**

**EMISSOR:**  
**CN=AC Certisign RFB G5**  
**OU=Secretaria da Receita Federal do Brasil - RFB**  
**O=ICP-Brasil**  
**C=BR**

**Assinatura Válida**

## Resumo

As operadoras de planos de saúde contam com o processo de regulação médica para garantir a qualidade na liberação de guias de solicitação de procedimentos (guias de serviços). Normalmente a liberação dos procedimentos é feita pela classificação da patologia contida na indicação clínica, por haver ligação entre as patologias e o rol de procedimentos da Agência Nacional de Saúde Suplementar (ANS). Para realizar este processo, o setor de regulação médica conta com profissionais da área de saúde, como médicos e enfermeiros. Esse trabalho é moroso e redundante, tornando o processo oneroso para as operadoras, em termos de custo e tempo de realização. O reconhecimento automático das patologias na indicação clínica poderá auxiliar o trabalho manual dos profissionais responsáveis, diminuindo o custo monetário e o tempo gasto na liberação dos procedimentos, além de tornar possível a aplicação da IA para automatizar o processo da regulação médica. Foi encontrada na literatura uma abordagem similar à proposta neste trabalho utilizando técnicas de Handwritten Text Recognition (HTR) em guias de serviços médicos, porém, para extrair o nome de remédios das prescrições médicas (E.KAMALANABAN et al., 2018), estando, além disso, em língua estrangeira. Neste trabalho, é proposto um modelo de reconhecimento de patologias descritas na indicação clínica de guias de solicitações médicas, escritas a mão por médicos, aplicando técnicas de Convolutional Neural Networks (CNN) em uma base de dados contendo guias de serviços digitalizadas do tipo SP/SADT, além de estarem utilizando a língua portuguesa (Brasil). A baixa ocorrência e a alta variação das patologias exigiu o retrabalho do dataset, onde novas imagens foram criadas através de técnicas de "zoom" e "rotation". O dataset final consiste em 3.480 imagens, distribuídas igualmente entre 4 classes. As imagens originais foram recortadas de guias de solicitações médicas, na seção de indicação clínica. Foi estabelecida a utilização de uma Deep Convolutional Neural Network (DCNN), utilizando a arquitetura VGG-16. Foram criados 7 modelos para chegar às configurações do modelo final. O modelo final alcançou 99% de acurácia no teste e uma taxa de perda de 0,05% no treino.

**Palavras-chave:** Operadoras de plano de saúde, Inteligência Artificial, Reconhecimento de palavras escritas à mão, Redes Neurais Convolucionais, VGG-16.

## **Abstract**

Health plan operators rely on the medical regulation process to ensure the quality in the release of procedure request guides (service guides). Normally, the release of procedures is done by the classification of the pathology contained in the clinical indication, as there is a link between the pathologies and the list of procedures of the Agência Nacional de Saúde Suplementar (ANS). To carry out this process, the medical regulation sector relies on healthcare professionals, such as doctors and nurses. This work is time-consuming and redundant, making the process onerous for operators, in terms of cost and time to carry out. The automatic recognition of pathologies in the clinical indication may help the professionals responsible for manual work, reducing the monetary cost and the time spent in releasing the procedures, in addition to making it possible to apply IA to automate the process of medical regulation. An approach similar to the one proposed in this work using Handwritten Text Recognition (HTR) techniques in medical service guides was found in the literature, however, to extract the name of medicines from medical prescriptions (E.Kamalanaban et al., 2018), and in a foreign language. In this work, a model for the recognition of pathologies described in the clinical indication of medical request guides, handwritten by doctors, is proposed, applying Convolutional Neural Networks (CNN) techniques in a database containing digitized service guides of the SP/SADT type, in addition to using the Portuguese language (Brazil). The low occurrence and high variation of pathologies required the rework of the dataset, where new images were created through "zoom" and "rotation" techniques. The final data-set consists of 3,480 images, distributed evenly among 4 classes. The original images were cut from medical requests guides in the clinical indication section. The use of a Deep Convolutional Neural Network (DCNN) was established, using the VGG-16 architecture. 7 models were created to reach the final model configurations. The final model achieved 99% accuracy in the test and a loss rate of 0.05% in training.

**Keywords:** Health plan operators, Artificial intelligence, Recognition of handwritten words, Convolutional Neural Networks, VGG-16.

## Lista de Figuras

Figura 1 – Segmentação das linhas de um documento escrito à mão .....	18
Figura 2 – Segmentação dos caracteres de um documento escrito à mão .....	19
Figura 3 – Ruídos na escrita .....	20
Figura 4 – Segmentação de linha, segmentação de palavras, segmentação de caracteres e texto reconhecido .....	20
Figura 5 – Ciclo de percepção da Inteligência Artificial - Interação do agente com o ambiente.....	22
Figura 6 – Representação simplificada do neurônio biológico.....	24
Figura 7 – Representação simplificada do neurônio matemático .....	25
Figura 8 – Utilização de um diagrama de Veen.....	26
Figura 9 – Percentual com os métodos de Deep Learning mais utilizados na informática para a saúde .....	27
Figura 10 – Ilustração de um modelo de Deep Learning.....	28
Figura 11 – Arquitetura de uma CNN para reconhecimento de dígitos .....	29
Figura 12 – Camada convolucional recebendo uma imagem 32x32x3 e aplicando um filtro de 5x5x3 .....	31
Figura 13 – Camada de "pooling". O aspecto mais importante na criação de um modelo CNN é a escolha entre diferentes camadas de "pooling" .....	31
Figura 14 – Função matemática da representação de uma RNN.....	32
Figura 15 – Equação de atualização LSTM.....	33
Figura 16 – A célula padrão de uma LSTM têm uma unidade linear com uma conexão recorrente em si com peso 1.0 .....	34
Figura 17 – Detalhe esquemático de uma RNN (esquerda) e uma LSTM (direita) utilizadas nas camadas ocultas de uma RNN .....	35
Figura 18 – Estratégia de pesquisa .....	39
Figura 19 – Exemplo de guia SP/SADT .....	44
Figura 20 – Campo indicação clínica contido na guia SP/SADT .....	44
Figura 21 – Representação visual da metodologia.....	45
Figura 22 – Software para rotulação de imagens .....	51
Figura 23 – Gráfico de variação entre a ocorrência de patologias .....	53

Figura 24 – Recortes de patologias rotuladas.....	53
Figura 25 – Arquitetura de uma VGG16.....	55
Figura 26 – Arquitetura simplificada de uma VGG16 .....	56
Figura 27 – Comparação entre VGG-16 e VGG-19.....	57
Figura 28 – Estrutura de pastas passadas para o gerador de dados .....	58
Figura 29 – Sumário do modelo criado.....	61
Figura 30 – Sumário do segundo modelo criado contendo camadas de "Dropout" ..	73
Figura 31 – Gráfico de acurácia do modelo V .....	80
Figura 32 – Gráfico de perda do modelo V.....	80

## Lista de Tabelas

Tabela 1 – Distribuição dos estudos selecionados nas bases de dados e depois da aplicação do processo de seleção.....	40
Tabela 2 – Matriz de confusão utilizada para avaliar os resultados .....	46
Tabela 3 – Relatório de classificação .....	47
Tabela 4 – Quadro de trabalho dos rotuladores. ....	50
Tabela 5 – Status do dataset do processo de rotulação.....	52
Tabela 6 – Distribuição de patologias entre as guias elegíveis .....	52
Tabela 7 – Disposição dos datasets.....	62
Tabela 8 – Configuração do modelo "A" .....	63
Tabela 9 – Acurácia do modelo A.....	64
Tabela 10 – Matriz de confusão do modelo A .....	64
Tabela 11 – Relatório de classificação do modelo A .....	65
Tabela 12 – Configuração do modelo "B" .....	66
Tabela 13 – Acurácia do modelo B.....	66
Tabela 14 – Matriz de confusão do modelo B .....	66
Tabela 15 – Relatório de classificação do modelo B .....	67
Tabela 16 – Configuração do modelo "C" .....	68
Tabela 17 – Acurácia do modelo C .....	68
Tabela 18 – Matriz de confusão do modelo C .....	69
Tabela 19 – Relatório de classificação do modelo C.....	69
Tabela 20 – Configuração do modelo "D" .....	70
Tabela 21 – Acurácia do modelo D .....	71
Tabela 22 – Matriz de confusão do modelo D .....	71
Tabela 23 – Relatório de classificação do modelo D .....	71
Tabela 24 – Configuração do modelo "E" .....	74
Tabela 25 – Acurácia do modelo E.....	74
Tabela 26 – Matriz de confusão do modelo E .....	74
Tabela 27 – Relatório de classificação do modelo E .....	75
Tabela 28 – Configuração do modelo "F" .....	76
Tabela 29 – Acurácia do modelo F.....	76
Tabela 30 – Matriz de confusão do modelo F.....	77

Tabela 31 – Relatório de classificação do modelo F .....	77
Tabela 32 – Configuração do modelo "G" .....	78
Tabela 33 – Acurácia do modelo G .....	78
Tabela 34 – Matriz de confusão do modelo G .....	79
Tabela 35 – Relatório de classificação do modelo G.....	79
Tabela 36 – Acurácia do modelo V .....	81
Tabela 37 – Matriz de confusão do modelo V .....	81
Tabela 38 – Relatório de classificação do modelo V.....	81

## Lista de Abreviaturas e Siglas

- AIS** Agentes Inteligentes. 25
- ANN** Artificial Neural Network. 26, 29, 41, 51
- ANS** Agência Nacional de Saúde Suplementar. 3, 5, 17
- ASR** Automatic Speech Recognition. 21
- 
- CNN** Convolutional Neural Networks. 3, 5, 7, 15, 24, 26, 28–32, 38, 41, 46, 51, 52, 54,73
- CTC** Connectionist Temporal Classification. 24
- 
- DCNN** Deep Convolutional Neural Network. 3, 5, 51
- DL** *Deep Learning*. 7, 15, 26–30
- DML** *Deep Machine Learning*. 26, 38
- 
- GPU** Unidade de Processamento Gráfico Dedicada. 31
- 
- HCR** Handwritten Character Recognition. 51, 53
- HTR** Handwritten Text Recognition. 3, 5, 15–17, 19, 21–24, 37, 38, 43, 47, 50, 51, 54,71, 73
- 
- IA** Inteligência Artificial. 3, 5, 7, 17–19, 24, 25, 28, 37
- IAM** IAM Handwriting Database. 38
- ILSVRC** ImageNet Large Scale Visual Recognition Competition. 52
- 
- LGPD** Lei Geral de Proteção de Dados. 47
- LSTM** Long Short-Term Memory. 8, 24, 33–36, 38
- 
- ML** *Machine Learning*. 7, 17–19, 25–28, 38
- MNIST** MNIST Handwritten Database. 30
- NN** Neural Networks. 24,26,52,53, 56
- OCR** Optical Character Recognition. 37–41
- OPS** Operadoras de Planos de Saúde. 15–17, 47, 73, 74

**QAC** Critério de Avaliação de Qualidade. 40

**RBM** Restricted Boltzmann Machine. 29

**RENN** Recursive Neural Networks. 27

**RNN** Recurrent Neural Network. 7, 8, 24, 26, 33–36, 38, 41

**RSL** Revisão Sistemática de Literatura. 38, 40

**RTEM** Reconhecimento de Textos Escritos à Mão. 26, 37

**TF** TensorFlow. 38

**UPN** Unsupervised Pretrained Networks. 26

**VGG** Visual Geometry Group. 54

**VGG Net** Visual Geometry Group Net. 51

## SUMÁRIO

<b>1</b>	<b>Introdução</b>	<b>12</b>
1.1.	Problema	12
1.2.	Objetivos	13
1.2.1.	Geral	13
1.2.2.	Específicos	13
1.3.	Lacuna a Ser Explorada	13
1.4.	Contribuição da Pesquisa	14
1.5.	Justificativa	14
1.6.	Aderência ao Programa de Mestrado da Uuniversidade FUMEC	15
1.7.	Estrutura do Documento	16
<b>2.</b>	<b>Fundamentação Teórica</b>	<b>17</b>
2.1.	<i>Handwritten text Recognition</i>	17
2.2.	Inteligência Artificial	21
2.3.	<i>Machine Learning</i>	23
2.4.	<i>Deep Learning</i>	23
2.5.	Redes Neurais Convolucionais	27
2.6.	Redes Neurais Recorrentes	31
2.7.	<i>Long Short-Term Memory</i>	32
<b>3.</b>	<b>Trabalhos Relacionados</b>	<b>36</b>
<b>4.</b>	<b>Metodologia</b>	<b>42</b>
4.1.	Caracterização da pesquisa	42
4.2.	Descrição dos dados	43
4.3.	Processo de Validação dos Dados	44
4.4.	Validação dos Resultados	46
<b>5.</b>	<b>Execução do experimento</b>	<b>48</b>
5.1.	Preparação dos dados	48
5.1.1.	Captação e Separação das Guias	48
5.1.2.	Revisão das Guias	49
5.1.3.	Rotulação das Guias	49
5.2.	Caracterização da Técnica de HTR	53
5.3.	Preparação do Modelo	57
5.3.1.	Modelo A	63
5.3.2.	Modelo B	65
5.3.3.	Modelo C	67
5.3.4.	Modelo D	70
5.3.5.	Modelo E	72

<b>5.3.6.</b>	<b>Modelo F .....</b>	<b>75</b>
<b>5.3.7.</b>	<b>Modelo G .....</b>	<b>77</b>
<b>5.4.</b>	<b>Resultado do Modelo.....</b>	<b>79</b>
<b>6.</b>	<b>Considerações finais.....</b>	<b>83</b>
<b>6.1.</b>	<b>Contribuições da Pesquisa .....</b>	<b>83</b>
<b>6.2.</b>	<b>Limitações da Pesquisa .....</b>	<b>84</b>
<b>6.3.</b>	<b>Trabalhos Futuros .....</b>	<b>84</b>
	<b>Referências .....</b>	<b>85</b>

## 1 INTRODUÇÃO

A correta interpretação da indicação clínica é crucial para a regulação médica na saúde suplementar. A indicação clínica traz as evidências para a liberação dos serviços solicitados pelo médico. A discriminação da patologia, descrita na indicação clínica, determina a correta avaliação para a liberação dos serviços solicitados.

Porém as guias de serviços médicos são encaminhadas para as Operadoras de Planos de Saúde em formato de documento digitalizado e a indicação clínica está escrita à mão pelo médico.

Atualmente as operadoras de planos de saúde contam com uma equipe de especialistas para analisar a indicação clínica, uma vez que existe dificuldade para a leitura e compreensão por partes das pessoas que não possuem conhecimento médico ou não estão familiarizadas com os termos utilizados.

Para facilitar este processo, foi utilizada a técnica de Handwritten Text Recognition (HTR) no reconhecimento da patologia contida na indicação clínica. Para isso, foi utilizada uma abordagem de *Deep Learning*, com uma Convolutional Neural Networks (CNN) a fim de executar tal tarefa.

Para construir a CNN, foi utilizado o dataset de imagens de documentos de guias de serviços, porém com o campo de indicação clínica recortado. O dataset criado com guias de serviços e o recorte e a rotulação da indicação clínica foi realizada pela equipe de especialistas responsáveis pela regulação médica de uma Operadoras de Planos de Saúde.

### 1.1. Problema

A correta interpretação da indicação clínica é crucial para a regulação médica na saúde suplementar. A indicação clínica trás as evidências para a liberação dos procedimentos solicitados pelo médico.

O reconhecimento e análise da patologia descrita na indicação clínica determina a correta avaliação para a liberação dos serviços solicitados. Porém as guias são encaminhadas para as Operadoras de Planos de Saúde em formato de documento digitalizado, e a indicação clínica está escrita à mão pelo médico.

Com isso chegamos ao seguinte questionamento: **Qual a acurácia das técnicas de HTR no reconhecimento de patologias descritas na indicação clínica**

## **em guias de serviços médicos escritas à mão?**

### **1.2. Objetivos**

Nesta seção serão apresentados o objetivo geral e os objetivos específicos deste trabalho.

#### **1.2.1. Geral**

Esta pesquisa tem como objetivo geral mensurar a acurácia da técnica de Hand-written Text Recognition no reconhecimento de textos escritos à mão por médicos, reconhecendo patologias contidas em guias de solicitação de serviços.

#### **1.2.2. Específicos**

Para chegar ao objetivo geral, delineiam-se os seguintes objetivos específicos:

- a) Identificar técnicas de HTR capazes de processar escritas complexas;
- b) Criar o dataset com imagens das guias de serviços médicos e prepará-lo para aplicação das técnicas de HTR;
- c) Evidenciar o desempenho do método selecionado na tarefa de HTR, incluindo performance e acurácia;
- d) Avaliar os resultados e identificar a possibilidade de uso e implantação da técnica.

### **1.3. Lacuna a Ser Explorada**

Os estudos atuais, relacionados ao reconhecimento da escrita do médico, propõem o uso de HTR em prontuários médicos, e estão em língua estrangeira. Apesar de apresentarem problemas estruturais similares aos encontrados no Brasil, a caligrafia médica pode variar, além da formação estrutural das palavras. Esta pesquisa se propõe a aplicar técnicas de HTR na indicação clínica de guias de serviços médicos

na língua Portuguesa/Brasil. O reconhecimento da patologia abrirá o caminho para aplicações como, por exemplo, no uso de Inteligência Artificial na regulação médica realizada por Operadoras de Planos de Saúde.

#### **1.4. Contribuição da Pesquisa**

A contribuição desta pesquisa encontra-se na apresentação de técnicas capazes de reconhecer a patologia descrita na indicação clínica em guias de solicitações médicas escrita à mão por médicos, com complexidade de interpretação. Outra contribuição está na facilitação para o desenvolvimento de técnicas de ML no auxílio da regulação médica em Operadoras de Planos de Saúde.

#### **1.5. Justificativa**

Independente da sua posição no mercado, pequena, média ou grande, as Operadoras de Planos de Saúde contam com o processo de regulação médica para garantir a qualidade na prestação dos seus serviços, seu enquadramento nas exigências da Agência Nacional de Saúde Suplementar (ANS) e também para a sua manutenção no mercado.

O mercado de saúde suplementar no Brasil conta com mais de 1.500 Operadoras de Planos de Saúde, assistindo mais de 47 milhões de pessoas, tornando-se um importante pilar na sustentabilidade do setor.

A regulação médica garante que as solicitações feitas pelos prestadores ou profissionais da área de saúde sejam avaliadas em termos de adequação e custo dos serviços médicos prestados. É executada por meio de normas técnicas e administrativas, regras de utilização e prestação dos serviços, tabelas de procedimentos e honorários, além de protocolos técnico-científicos que orientam às boas práticas médicas e profissionais.

A regulação médica é feita com base na guia de solicitação de procedimentos que o profissional solicitante julga necessário, e o motivo da realização destes procedimentos deve ser fundamentado pela indicação clínica do paciente. Sendo assim, a indicação clínica sustenta a necessidade da realização dos procedimentos solicitados. O correto reconhecimento das patologias é necessário, logo que para determinadas patologias a ANS torna obrigatório a liberação de determinado

procedimento. Mas da mesma forma, determinada patologia pode tornar dispensável a realização de um procedimento solicitado.

Com a evolução dos algoritmos de Inteligência Artificial e da capacidade computacional é possível trazer soluções anteriormente impossíveis, que nos trazem a possibilidade de melhorar processos ou promover novas opções de uso.

Tais fatos despertam o interesse na utilização da Inteligência Artificial e *Machine Learning* para auxiliarem no processo de reconhecimento das patologias, contidas na indicação clínica das guias de serviços médicos escritas à mão.

A partir das condições apresentadas, o projeto propõe avaliar as técnicas de reconhecimento de textos escritos à mão para transcrever a patologia contida na indicação clínica para o formato digital.

### **1.6. Aderência ao Programa de Mestrado da Uuniversidade FUMEC**

O Programa de Pós-Graduação em Sistemas de Informação e Gestão do Conhecimento da Universidade FUMEC está estruturado em uma área de concentração:

- Gestão de sistemas de Informação e do Conhecimento

O programa abrange duas linhas de pesquisa, sendo:

- Gestão da Informação e do Conhecimento e
- Tecnologia e Sistemas de Informação

Este trabalho posiciona-se na linha de pesquisa "Sistemas e Tecnologia de Informação", sendo que o objetivo desse é a análise das técnicas de HTR na indicação clínica de guias de solicitação de procedimentos e no reconhecimento da patologia. Estas técnicas atuam no contexto da Inteligência Artificial, mais especificamente em *Machine Learning*, suas derivações, ferramentas de implementação e metodologias adequadas. Como há a proposta de modelos que possibilitem o reconhecimento de textos escritos a mão e extração de palavras de interesse em documentos digitalizados, entende-se que o trabalho se enquadra ao Programa de Mestrado da

FUMEC.

### 1.7. Estrutura do Documento

Este estudo foi dividido em 6 capítulos, com a intenção de apresentar, de forma organizada, toda temática relacionada ao trabalho.

O **Capítulo 1** apresenta o que foi tratado neste projeto, demonstrando a importância do trabalho desenvolvido, além de situá-lo sobre o assunto abordado nos próximos capítulos. O **Capítulo 2** apresenta os tópicos abordados durante a pesquisa, dando informação suficiente para que se consiga compreender os itens da pesquisa. Neste capítulo, serão apresentados os seguintes itens:

- Handwritten Text Recognition;
- Inteligência Artificial;
- Machine Learning;
- Deep Learning;
- Redes Neurais Convolucionais
- Redes Neurais Recorrentes e;
- Long Short-Term Memory.

Já o **Capítulo 3** apresenta os trabalhos relacionados a esta pesquisa. O **Capítulo 4** apresenta os procedimentos metodológicos utilizados para a realização desta pesquisa. O **Capítulo 5** apresenta a execução do experimento. Em seguida, o capítulo 6 apresenta as considerações finais. Para finalizar o projeto, foi apresentado o capítulo de referências que embasaram a pesquisa.

## 2. FUNDAMENTAÇÃO TEÓRICA

### 2.1. *Handwritten text Recognition*

Reconhecido como a técnica moderna para reconhecimento de textos, o Handwritten Text Recognition (HTR) é uma técnica importante para o acesso ao conhecimento. É muito explorado para o reconhecimento de documentos digitalizados. Tem como objetivo transcrever estes documentos, deixando seu conteúdo acessível e fácil de pesquisar.

Para (SÁNCHEZ et al., 2019), os sistemas HTR progrediram nas últimas duas décadas, por duas razões: 1) o uso de conceitos e técnicas holísticas de treinamento e reconhecimento, previamente desenvolvidas no campo do Automatic Speech Recognition (ASR) e; 2) por conta da existência de um número de datasets públicos disponíveis para treinar e testar os sistemas de HTR.

A aplicação destes sistemas consiste no armazenamento de conteúdo escrito em formato digitalizado para então convertê-lo ao formato em que seja possível sua leitura por um computador. Para (MANIK NIJHAWAN, 2019) a detecção de textos escritos a mão pode ser dividido em 4 passos:

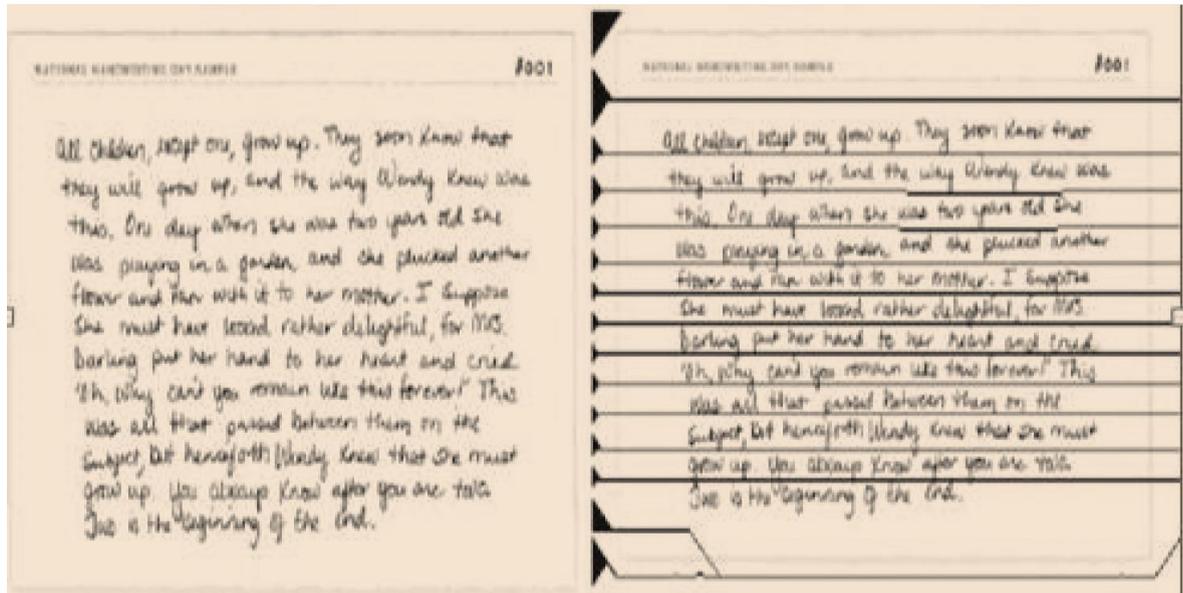
- a) segmentação de linha;
- b) segmentação de caracteres;
- c) pré-processamento da imagem para remover ruídos e;
- d) aplicar técnicas adequadas para a detecção da escrita.

A segmentação de linhas consiste na separação das linhas de um texto. A primeira tarefa é rotular com enumerações. Uma vez enumeradas, cada linha é processada para prepará-la para o processo de reconhecimento.

Uma vez que as pessoas possuem estilos diferentes de caligrafia, existem variações na formação das linhas. Algumas pessoas têm caligrafia inclinada e as linhas não são espaçadas uniformemente, ou então quando as linhas estão muito próximas umas das outras, tudo isso se torna um desafio na segmentação de linhas. A imagem 1 apresenta o exemplo de um documento com texto escrito a mão antes e depois do processo de segmentação de linha.

Uma vez que as linhas estão segmentadas, agora é possível separar as imagem em partes, e tratar o conjunto de caracteres presente em cada linha, separando-os um a um. A segmentação de caracteres é uma operação que busca decompor uma imagem com a sequência de caracteres em sub imagens.

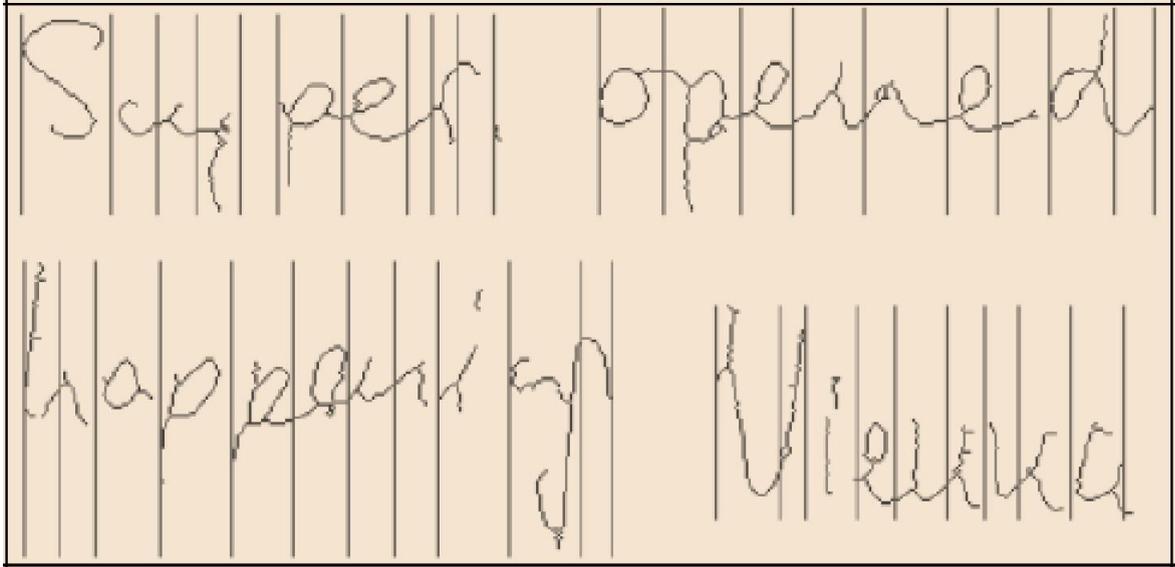
**Figura 1** – Segmentação das linhas de um documento escrito à mão.



Fonte: Adaptado de Manik Nijhawan, (2019).

Esta etapa também possui algumas peculiaridades, variando de acordo com a caligrafia de cada pessoa. Existem diferentes casos, onde os caracteres podem estar muito próximos ou muito distantes, então a formação estrutural do texto depende de estilos individuais de escrita. Assim, a correspondência dos modelos é obtida pela extração dos atributos que distinguem os caracteres da imagem. A imagem 2 apresenta a segmentação de caracteres em uma imagem.

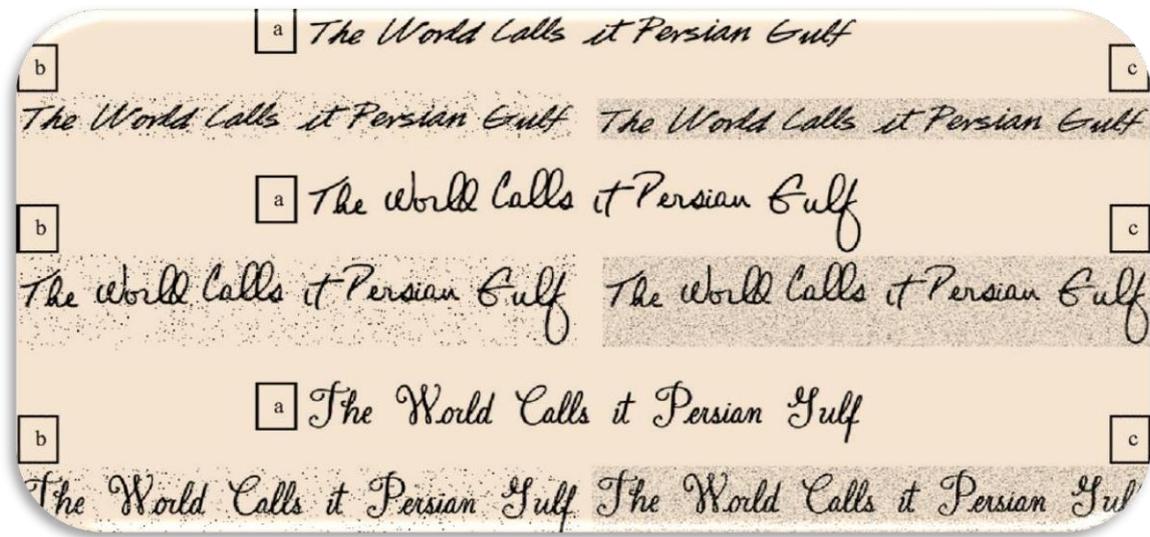
**Figura 2** – Segmentação dos caracteres de um documento escrito à mão.



Fonte: Adaptado de Manik Nijhawan, (2019).

Para (MANIK NIJHAWAN, 2019), a primeira técnica de pré-processamento necessária para o HTR é a remoção de ruídos. Um dos ruídos mais comuns é o ruído "sal e pimenta", observado na figura 3. Esse ruído pode ser observado como pontos pretos e brancos em uma imagem. Ou pela legenda: a) imagem original de um texto escrito à mão digitalizado (b) imagem corrompida por ruído sal e pimenta (c) imagem corrompida por ruído Gaussian Ao capturar uma imagem, esses ruídos podem ocorrer por conta dos distúrbios do ambiente, por partículas de sujeira no ambiente e várias outras razões. A redução efetiva para este tipo de ruído é o filtro de "média". Essa é a técnica mais comum para remoção desses ruídos. Este filtro consiste na escolha mediana de um grupo de pixels para então atribuí-los aos pixels vizinhos.

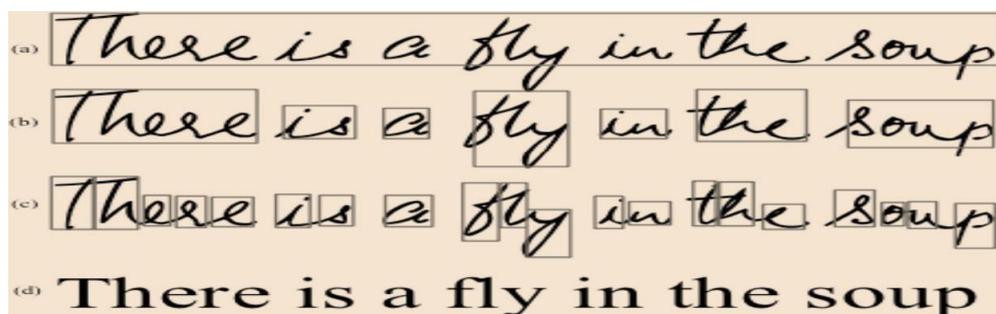
**Figura 3 – Ruídos na escrita.**



Fonte: Adaptado de Manik Nijhawan, (2019).

Para Rao e Aditya, (2014) existem mais algumas etapas no processamento das imagens, como o tresholding, extração da matéria textual em primeiro plano, através da remoção do plano de fundo texturizado e segmentação de palavras. Corroboram ainda ao acrescentarem a separação de linhas individuais após a segmentação de linha, a segmentação de palavras, seguida da separação de palavras textuais e a segmentação de caracteres, seguido da separação de caracteres individuais.

**Figura 4 – Segmentação de linha, segmentação de palavras, segmentação de caracteres e texto reconhecido.**



Fonte: Adaptado de Rao & Aditya,(2014).

Ao relatar os problemas desafiadores relacionados ao HTR, (SÁNCHEZ et al.,

2019) destaca-se que o HTR requer combinações cuidadosas de várias técnicas avançadas de reconhecimento de padrões, entre elas o processamento de imagens, análise de imagens de documentos, extração de recursos, abordagens de redes neurais e modelagem linguística.

Os textos escritos à mão são encontrados em diversos formatos e tipos diferentes de imagens e uma solução completa no reconhecimento de textos precisa oferecer suporte à leitura de textos em imagens independente de sua estrutura (INGLE et al., 2019). Vários estudos relacionados ao HTR têm adotado estratégias de reconhecimento em nível de linha, utilizando uma combinação de *Convolutional Neural Networks* (CNN) e *Recurrent Neural Network* (RNN) do *Long Short-Term Memory* (LSTM) para a extração de recursos, treinados com *Connectionist Temporal Classification* (CTC) ou classificação temporal conexionista.

Os sistemas HTR são baseados em *Neural Networks*, demonstrando sua relação/dependência com a Inteligência Artificial.

## 2.2. Inteligência Artificial

Para resumir a Inteligência Artificial (IA) em uma breve história, (NILSSON, 1997) volta ao desenvolvimento dos primeiros computadores digitais, entre 1940 e 1950, onde vários pesquisadores escreveram programas que podiam resolver tarefas de raciocínio elementar. Destaque entre os artigos da época estava o primeiro programa de computador que podia jogar xadrez, jogo de damas e provar teoremas em geometria plana. Intitulado de Automata Studies, o trabalho de (SHANNON & McCARTHY, 1956) tratava a teoria matemática do autômato, porém, McCarthy decidiu utilizar a frase Inteligência Artificial como título para a conferência de Dartmouth que seria realizada na época. Vários outros artigos foram submetidos para a conferência, levando várias sugestões de nome para o campo, tais como: processamento de informação complexa, inteligência de máquina, programação heurística e cognologia, porém o nome Inteligência Artificial persistiu.

Para Barr e Feigenbaum, (1981) a Inteligência Artificial é um ramo da ciência da computação – é o estudo da relação entre computação e cognição. Basicamente, as pesquisas em IA envolvem a escrita de programas que buscam alcançar algum tipo de comportamento inteligente.

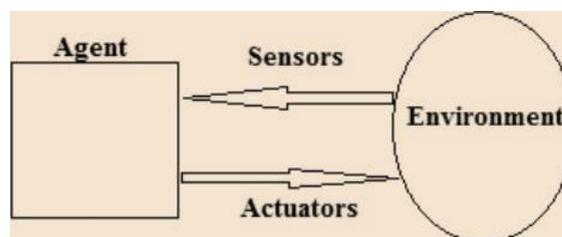
Técnicas e métodos de Inteligência Artificial têm sido aplicados em problemas

de várias áreas diferentes. Para demonstrar sua diversidade, (Nilsson, 1982) descreve algumas de suas aplicações:

- a) Processamento de Linguagem Natural;
- b) Recuperação de Informações de Bases de Dados;
- c) Sistemas Experts de Consultoria;
- d) Prova de Teoremas;
- e) Robótica;
- f) Programação Automática;
- g) Problemas Combinatórios e de Programação e;
- h) Problemas de Percepção.

Também chamados de Agentes Inteligentes (AIS), os programas de Inteligência Artificial são desenvolvidos para interagir com o ambiente. O agente pode interagir com o ambiente por meio de seus sensores e então alterar o seu estado por meio dos atuadores. A figura 5 ilustra o ciclo de percepção da Inteligência Artificial.

**Figura 5** – Ciclo de percepção da Inteligência Artificial - Interação do agente com o ambiente.



Fonte: Adaptado de Das et al., (2015).

O principal objetivo da IA é desenvolver inteligência similar à humana em máquinas. Este feito pode ser realizado por meio de algoritmos de aprendizagem que tentam imitar como o cérebro humano aprende (DAS et al., 2015). O *Machine Learning*, que é um mecanismo de construção de sistemas de Inteligência Artificial, é importante por permitir que as máquinas obtenham inteligência humana sem programação explícita.

### **2.3. Machine Learning**

Para (DAS et al., 2015), o *Machine Learning* (ML) é definido como o campo de estudo que dá aos computadores a habilidade de aprenderem sem serem explicitamente programados. Em uma definição formal, são programas de computador que aprendem a partir da experiência (E) em relação a alguma tarefa (T) e medidas de desempenho (D), sendo que, se seu desempenho em T, mensurado por D, melhora a experiência E o programa é chamado de Programa de *Machine Learning*. Ao descrever os tipos de algoritmos de aprendizagem de máquina, (Das et al., 2015) apresentam quatro tipos diferentes:

- a) aprendizado supervisionado;
- b) aprendizado não supervisionado;
- c) aprendizado por reforço e;
- d) sistemas de recomendação.

Além disso, os autores trazem da literatura as principais aplicações do *Machine Learning* relacionando-as aos diferentes tipos de algoritmos. Entre elas está o Reconhecimento de Textos Escritos à Mão, tarefa associada aos algoritmos de aprendizagem supervisionada. Em relação a esta tarefa, os autores descrevem o seguinte:

Acontece que um dos motivos pelos quais hoje é tão barato rotear um correio pelos países é que, quando um endereço é escrito em um envelope, ocorre que um algoritmo de aprendizado de máquina aprendeu a ler a caligrafia para que seja possível rotear automaticamente esse envelope ao seu destino, portanto, custa menos (DAS et al., 2015).

### **2.4. Deep Learning**

Recentemente, o *Deep Learning* (DL) trouxe uma nova tendência para o Machine Learning. Os fundamentos teóricos do *Deep Learning* estão enraizados na *Artificial Neural Network* (ANN) clássica da literatura. Porém o *Deep Learning* se diferencia do uso mais tradicional de *Neural Networks* por utilizar muitos neurônios ocultos e camadas ocultas, geralmente mais de duas, apresentando uma vantagem

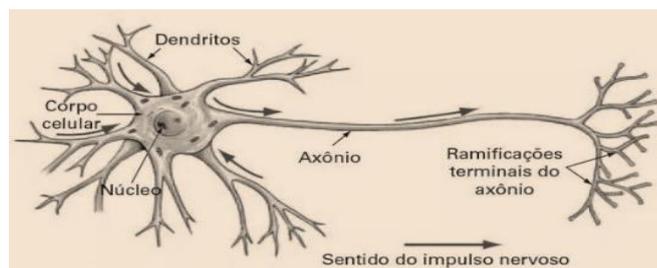
em sua arquitetura, combinada com novos paradigmas de treinamento.

Para se compreender melhor quando se fala que os fundamentos de *Deep Learning* estão enraizados na *Artificial Neural Network*, vale a pena explicar melhor seu fundamento. *Artificial Neural Network* são modelos computacionais inspirados no sistema nervoso central humano (SETTI & WANTO, 2019). Para Goldberg, (2017), as *Artificial Neural Networks* são uma família de modelos para aprendizagem de máquina. Ao falar sobre as *Neural Networks*, Kumar, (2016) diz que a *Artificial Neural Network* é uma poderosa ferramenta para modelagem de dados capaz de capturar e representar entradas/saídas complexas e relacioná-las. A semelhança entre as *Artificial Neural Network* e o cérebro humano pode ser analisada nas figuras 6 e 7.

Na verdade, existe um desafio ao definir o *Deep Machine Learning* por conta das alterações que o mesmo sofreu em sua forma. Para Josh Patterson, (2017), uma definição usual é que o *Deep Learning* lida com uma *Neural Network* com mais de duas camadas. Podendo ser melhor definida como uma *Neural Network* com grande número de parâmetros e camadas e projetada em uma das quatro arquiteturas fundamentais:

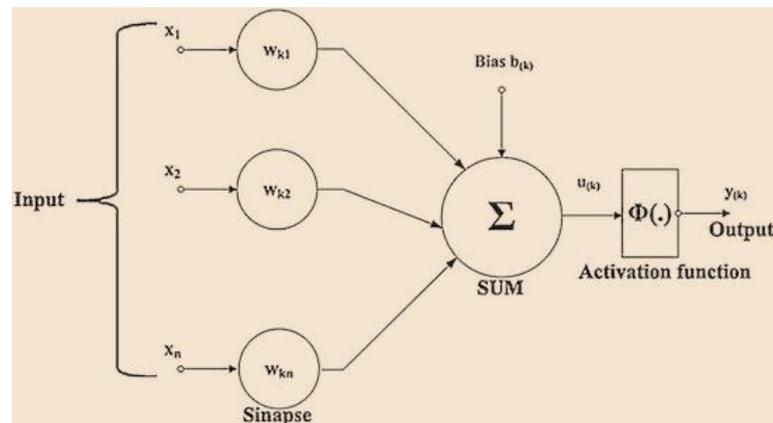
- a) *Unsupervised Pretrained Networks*,
- b) *Convolutional Neural Networks*,
- c) *Recurrent Neural Network* e
- d) *Recursive Neural Networks*.

**Figura 6** – Representação simplificada do neurônio biológico



Fonte: Adaptado de DSA, (2019).

**Figura 7** – Representação simplificada do neurônio matemático



Fonte: Adaptado de DSA, (2019).

A extração automática de características também é outra grande vantagem que o *Deep Learning* apresenta sobre os algoritmos tradicionais de *Machine Learning*.

Ao explicar as características da arquitetura do *Deep Learning*, Ravi et al., (2017) dizem que, embora o recurso de muitos neurônios permita cobrir uma extensão mais ampla de dados, o pipeline camada a camada de combinação não linear de suas saídas gera uma projeção dimensional mais baixa do espaço de entrada. Toda projeção de menor dimensão corresponde a um maior nível perceptivo. Desde que a rede seja idealmente ponderada, ela resulta em uma abstração eficaz do alto nível dos dados ou imagens brutas. Esse alto nível de abstração renderiza um conjunto de recursos automáticos que, caso contrário, exigiriam recursos feitos à mão ou sob medida.

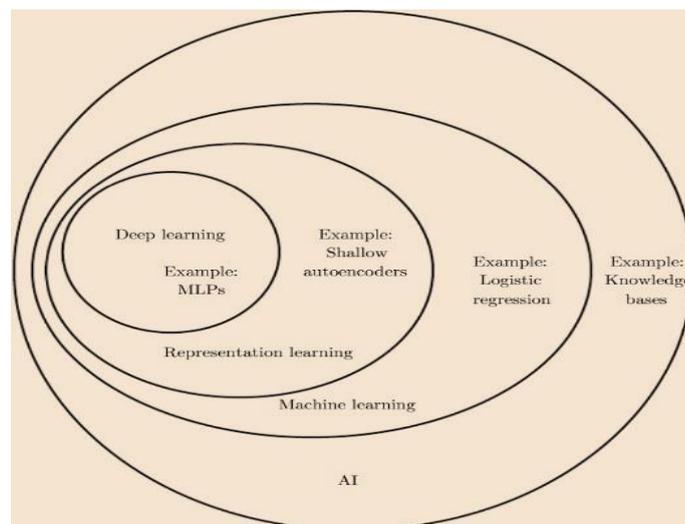
O histórico de tendências relacionado ao *Deep Learning* torna mais fácil entender seu contexto. Seguindo este caminho, Goodfellow et al., (2016) apresentaram as principais tendências:

- a. O *Deep Learning* possui uma tendência história longa e rica, mas existem muitos nomes que refletem pontos de vista filosóficos diferentes, e cresceu e diminui em popularidade.
- b. O *Deep Learning* se tornou mais usual assim que a quantidade de dados para treinamento cresceu.
- c. Os modelos de *Deep Learning* cresceram ao longo do tempo assim que a infraestrutura computacional (*hardware* e *software*) para *Deep Learning* foi aperfeiçoada.
- d. O *Deep Learning* resolveu aplicações mais complexas com aumento

da acurácia ao longo do tempo.

Para se conseguir identificar a relação do *Deep Learning* com o *Machine Learning*, Goodfellow et al., (2016) criaram uma representação em forma de diagrama de Venn, disponível na figura 8, exibindo como o *Deep Learning* é um tipo de representação de *Machine Learning*, o qual é usado por várias mas não todas abordagens de Inteligência Artificial. Cada seção do diagrama inclui um exemplo de uma tecnologia de Inteligência Artificial.

**Figura 8** – Utilização de um diagrama de Venn.



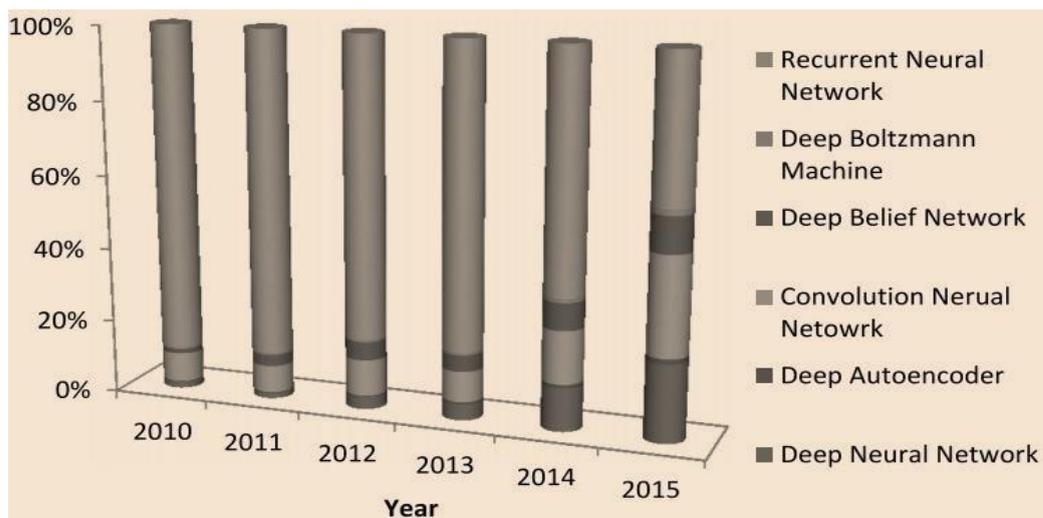
Fonte: Adaptado de Goodfellow et al., (2016).

Existem várias variantes metodológicas do *Deep Learning*, e várias arquiteturas se destacam. Ao analisarem as publicações referentes às variantes metodológicas do *Deep Learning*, Ravì et al., (2017) trazem as variações das ocorrências de utilização encontradas em publicações relacionadas ao método de *Deep Learning* desde 2010. Entre as analisadas, a *Convolutional Neural Networks* (CNN) apresentou maior impacto no campo da informática relacionada à saúde.

Outras arquiteturas plausíveis para o *Deep Learning* incluem aquelas baseadas em composições de *Restricted Boltzmann Machine* (RBM), além de redes de crenças profundas (DBN), funções de pilhas de “*autoencoders*” como “*autoencoders*” profundos estendendo a *Artificial Neural Network* com muitas camadas como *Deep Learning* ou com ciclos direcionados como CNN. A figura 9 apresenta os métodos de *Deep Learning* mais utilizados na informática para a saúde. As estatísticas dos

métodos de aprendizagem podem ser obtidas no Google Scholar, usando o nome do método com pelo menos um dos itens medicina ou saúde como frase de pesquisa.

**Figura 9** – Percentual com os métodos de *Deep Learning* mais utilizados na informática para a saúde.



Fonte: Adaptado de Ravi et al., (2017).

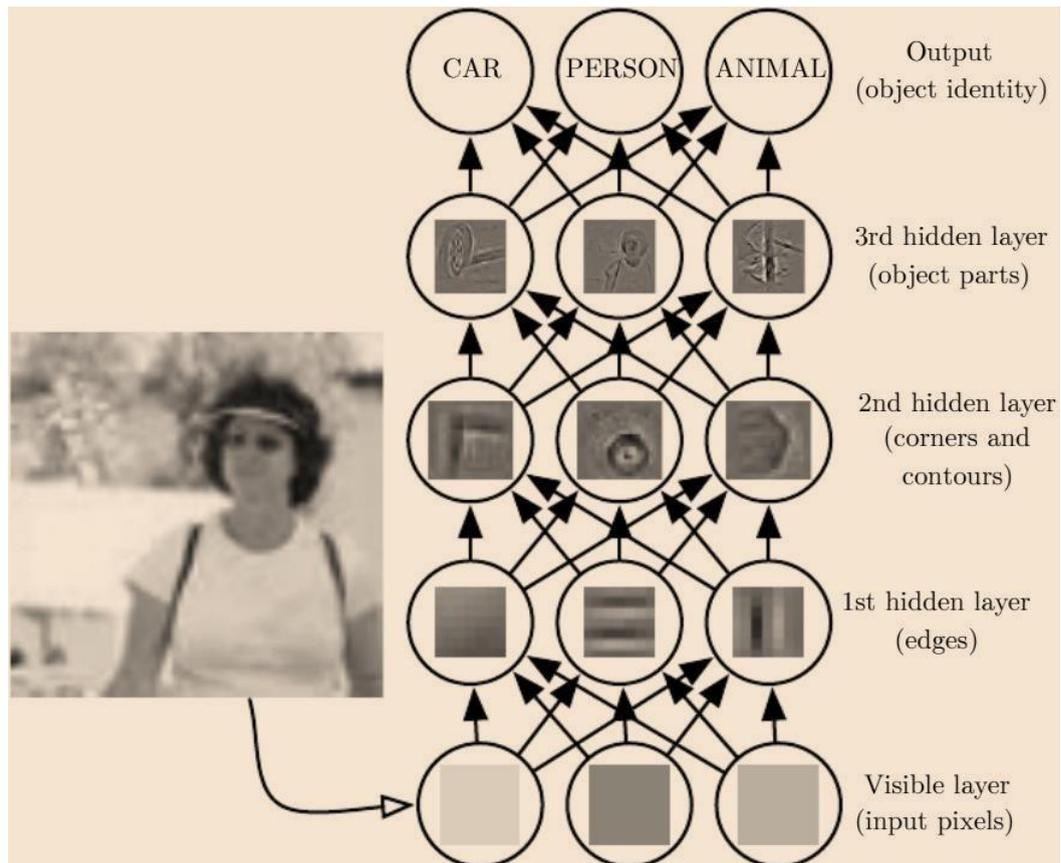
O *Deep Learning* pode ser apresentado por meio de suas perspectivas. Para Goodfellow et al., (2016) a ideia de aprender a correta representação para os dados fornece uma perspectiva sobre *Deep Learning*. Outra perspectiva do *Deep Learning* é que a profundidade permite ao computador aprender por meio de um programa de vários passos. Cada camada da representação pode ser pensada como um estado da memória do computador após executar outro conjunto de instruções em paralelo. As *Artificial Neural Networks* mais profundas podem executar mais instruções sequenciais. Geralmente as instruções sequenciais oferecem melhores resultados, uma vez que as instruções posteriores podem se referir aos resultados das instruções anteriores.

## 2.5. Redes Neurais Convolucionais

As *Convolutional Neural Networks* (CNN) é um modelo expandido de Artificial Neural Network (ANN) tradicionais. É utilizada para reconhecimento e detecção de imagem, desenvolvidas para processar dados bidimensionais (Harjoseputro et al., 2019). A figura 10 ilustra um modelo de *Deep Learning* onde existe dificuldade para

um computador entender o significado de um dado de entrada cru, assim como esta imagem representada por uma coleção de pixels.

**Figura 10** – Ilustração de um modelo de *Deep Learning*.



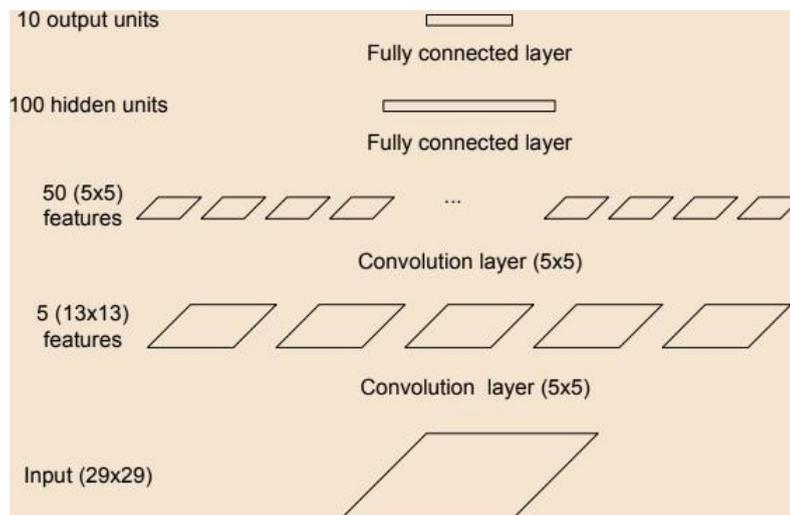
Fonte: Adaptado de Goodfellow et al., (2016).

Ao explicar as CNN, Ming Liang & Xiaolin Hu, (2015) voltam à proposta inicial do Neocognitron, proposto por Fukushima, (1980). Ao relatar suas diferenças, os autores destacam a incorporação do Back-propagation como o algoritmo para a aprendizagem de campos receptivos ou unidades simples. Desde seu nascimento, as CNN têm sido caracterizadas por suas conexões locais, compartilhamento de pesos e pool local. Muito utilizadas no reconhecimento de imagens, as CNN foram utilizadas por (SIMARD et al., 2003) em sua pesquisa para a Microsoft com a intenção de analisar documentos visuais. Na pesquisa, eles apresentam métodos fáceis de depuração. O objetivo do trabalho era reconhecer dígitos escritos à mão, e para isso, foi utilizada a base de dados do MNIST Handwritten Database (MNIST). A arquitetura da CNN utilizada pode ser vista na figura 11.

Ao analisarem a CNN, Ravi et al., (2017) referem-se a arquitetura dessas

redes como um conjunto intercalado de camadas "*feed-forward*" implementando filtros convolucionais, seguidos de camadas de redução, retificação ou "*pool*". Cada camada da rede origina um recurso abstrato de alto nível. Essa arquitetura de inspiração biológica se assemelha ao procedimento pelo qual o córtex visual assimila as informações visuais na forma de campos receptivos.

**Figura 11** – Arquitetura de uma CNN para reconhecimento de dígitos.



Fonte: Adaptado de Simard et al., (2003).

Ao aprofundarem nos detalhes da arquitetura das CNN, Harjoseputro et al., (2019) perceberam que a fraqueza destas redes está na complexidade do número de camadas construídas, o que tem implicações para o tempo de treinamento de dados; portanto, há necessidade de recursos computacionais adicionais como, por exemplo, uma Unidade de Processamento Gráfico Dedicada (GPU).

O treinamento de uma CNN possui três estágios, abrangendo as camadas de convolução, "*pool*" e as camadas conectadas.

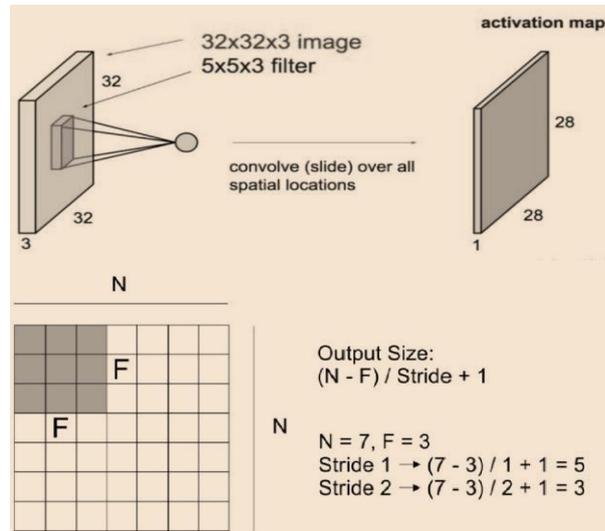
A camada de convolução tende a executar operações baseadas na saída da camada anterior, sendo o principal processo subjacente a uma CNN, que neste caso é um termo matemático que significa aplicar uma função na saída de outras repetidamente. Portanto, uma camada de convolução tende a experimentar a complexidade de um modelo por meio da otimização de sua saída. A figura 12 demonstra a camada convolucional.

A camada de "*pool*" tende a usar funções com mapa de recursos como entrada, processando posteriormente com várias operações estatísticas disponíveis,

usando valores de pixels próximos. Além disso, a aplicação requer o uso de (1) “max pooling” com tamanho  $2 \times 2$  e passo 2, com um valor obtido para cada troca desse filtro e o maior na área  $2 \times 2$ , (2) “average pooling”, que aceita o valor médio. O aspecto mais importante na criação de um modelo de CNN é a escolha entre diferentes camadas de “pooling”, que melhoram a *performance* do modelo. A figura 13 exemplifica a camada de “pooling”.

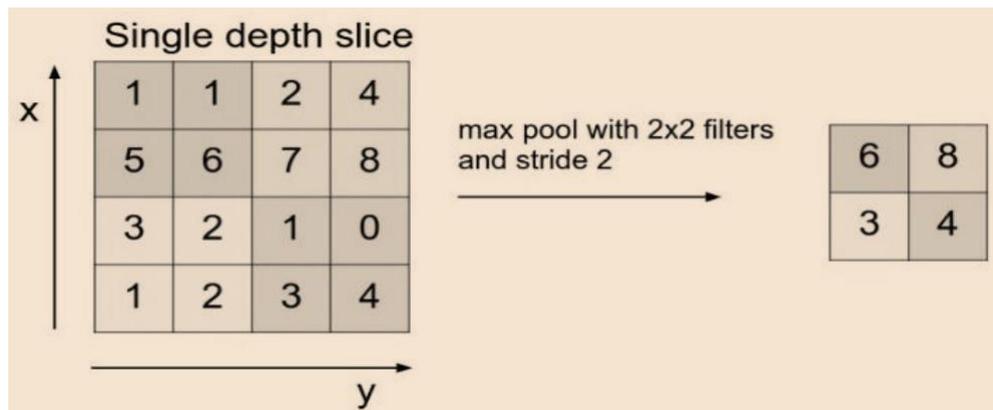
No sistema conectado, todos os neurônios de ativação da camada anterior são ligados aos neurônios do próximo, bem como o conceito de neurônios artificiais comuns (HARJOSEPUTRO et al., 2019). Portanto, a segurança na ativação da camada anterior é convertida em dados de uma dimensão antes da conexão com todos os neurônios conectados à camada conectada, e a diferença entre ambos é observada no aspecto em que o tipo de convolução está associado apenas a determinadas regiões, enquanto o último possui toda a conexão. Além disso, eles também tendem a ter um núcleo  $1 \times 1$ , que desempenha funções semelhantes, mantendo o caráter espacial dos dados. E também, o tipo conectado tende a ter várias camadas ocultas e de saída, além de funções de ação e perda.

**Figura 12** – Camada convolucional recebendo uma imagem 32x32x3 e aplicando um filtro de 5x5x3



Fonte: Adaptado de Harjoseputro et al., (2019).

**Figura 13** – Camada de "pooling".



Fonte: Adaptado de Harjoseputro et al., (2019).

O aspecto mais importante na criação de um modelo CNN é a escolha entre diferentes camadas de "pooling".

## 2.6. Redes Neurais Recorrentes

Para Gers et al., (2000), as Recurrent Neural Network (RNN) constituem uma classe muito poderosa de modelos computacionais, capazes de instanciar dinâmicas

quase arbitrárias. O limite desse potencial pode ser explorado, contudo é limitado pela eficácia do procedimento de treinamento aplicado.

A RNN é capaz de processar uma sequência de tamanho arbitrário aplicando de forma recursiva uma função de transição no vetor de estado oculto interno  $h_t$  da sequência de entrada. A ativação do estado oculto  $h_t$  no momento  $t$  é computado como uma função  $f$  do símbolo de entrada corrente  $x_t$  e do estado oculto anterior  $h_{t-1}$ , como visto na figura 14.

**Figura 14** – Função matemática da representação de uma RNN

$$h_t = \begin{cases} 0 & t = 0 \\ f(h_{t-1}, x_t) & \text{otherwise} \end{cases}$$

Fonte: Adaptado de Liu et al., (2016).

É comum utilizar funções de transição  $f$  estado-a-estado como a composição de uma não linearidade por elementos com uma transformação afim de  $x_t$  e  $h_{t-1}$ .

Ao apontarem uma estratégia simples para modelar sequências, Liu et al., (2016) dizem que é utilizado mapear a sequência de entrada para um vetor de tamanho fixo utilizando uma RNN, e então alimentar um vetor de uma camada softmax para classificação ou outra tarefa desejada.

Contudo, um problema com RNN com funções de transição dessa forma é que, durante o treino, os componentes do vetor gradiente podem crescer ou decrescer exponencialmente em longas sequências. Esse problema com gradientes que explodem ou desaparecem torna difícil para o modelo RNN aprender correlações de longa distância em uma sequência. Contudo, as LSTM trazem uma proposta que adiciona características que potencializam a RNN comum (LIU et al., 2016).

## **2.7. Long Short-Term Memory**

Ao lembrar as origens da *Long Short-Term Memory* (LSTM), Sherstinsky, (2018) falou sobre o artigo original publicado em 1997 que retratava a LSTM. Desde então, inúmeros trabalhos teóricos e experimentais foram publicados com o assunto deste tipo de RNN, e vários destes relatando os surpreendentes resultados alcançados em uma ampla variedade de domínios de aplicativos em que os dados

são sequenciais. A modelagem de linguagem sofreu forte impacto das LSTM, fato também percebido na transcrição de fala para escrita, máquina de tradução, além de outras aplicações (SHERSTINSKY, 2018).

Ao explicar o LSTM, Cheng et al., (2016) referem-se ao processamento de uma sequência de tamanho variável  $x = (x_1, x_2, \dots, x_n)$  que é adicionado novos conteúdos em um único *slot* de memória, com portas controlando até que ponto o novo conteúdo deve ser memorizado, o conteúdo antigo deve ser apagado e o conteúdo atual deve ser exposto. No passo  $t$ , a memória  $c_t$  e o estado oculto  $h_t$  são atualizados com as seguintes equações apresentadas na figura 15.

**Figura 15** – Equação de atualização LSTM.

$$\begin{bmatrix} i_t \\ f_t \\ o_t \\ \hat{c}_t \end{bmatrix} = \begin{bmatrix} \sigma \\ \sigma \\ \sigma \\ \tanh \end{bmatrix} W \cdot [h_{t-1}, x_t] \quad (1)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \hat{c}_t \quad (2)$$

$$h_t = o_t \odot \tanh(c_t) \quad (3)$$

Fonte: Adaptado de Cheng et al., (2016).

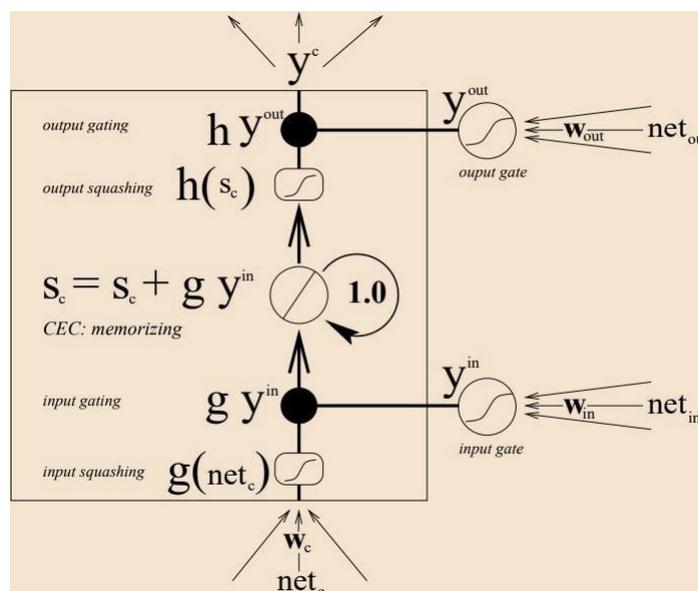
Comparado com a RNN padrão, a LSTM utiliza atualização de memória adicional e separa a memória  $c$  do estado oculto  $h$ , o qual interage com o ambiente ao fazer previsões. Onde  $i$ ,  $f$  e  $o$  são ativações de porta. Comparado ao RNN padrão, o LSTM usa atualizações aditivas de memória e separa a memória  $c$  do estado oculto  $h$ , que interage com o ambiente ao fazer previsões.

Ao lembrar as limitações das RNN, (Sherstinsky, 2018) falam que as RNN padrão não conseguem aprender na presença de atrasos de tempo superiores a 5-10 intervalos entre eventos de entrada relevantes e sinais de destino. O problema do erro de fuga põe em dúvida se uma RNN padrão pode de fato apresentar vantagens práticas significativas ao longo do tempo em redes do tipo feed-forward baseadas em janelas. Por fim, Sherstinsky, (2018) relatam como os modelos LSTM não são afetados por estes problemas, já que eles podem aprender a reduzir atraso mínimos de tempo em mais de 1.000 etapas discretas, aplicando erro constante por meio de

"carrosséis de erro constante" em unidades especiais, chamadas células. Os portões de unidades multiplicadoras aprendem a abrir e fechar o acesso às células. O algoritmo de aprendizado da LSTM é local no espaço e tempo; sua complexidade computacional por etapa de tempo e peso é  $O(1)$ . Ele resolve tarefas complexas de longo prazo que nunca foram resolvidas pelos algoritmos RNN anteriores.

A célula de uma LSTM padrão possui uma unidade linear com uma conexão recorrente em si própria com peso 1.0. O portão de entrada e o de saída regulam o acesso de leitura e escrita a célula na qual o estado é denotado  $s_c$ . A função  $g$  esmaga a entrada da célula;  $h$  esmaga a saída da célula. Os detalhes podem ser vistos na imagem 16.

**Figura 16** – A célula padrão de uma LSTM



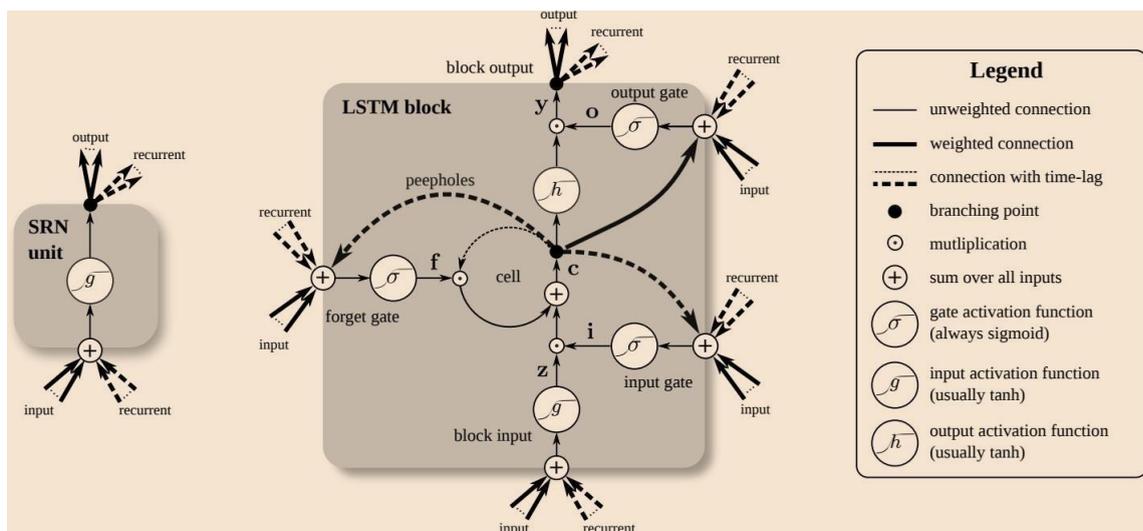
Fonte: Adaptado de Gers et al., (2000).

Uma célula padrão de uma LSTM tem uma unidade linear com uma conexão recorrente em si com peso 1.0. Ao buscarem as configurações mais comuns de LSTM, Greff et al., (2015) encontram na literatura a "vanilla LSTM", descrita como a mais utilizada na literatura. Este modelo de LSTM incorpora algumas mudanças antes acrescentadas na LSTM padrão e utiliza o treinamento com gradiente completo. O modelo esquemático desta LSTM pode ser visto na figura 17. Ela contém três portões (entrada, esquecimento e saída), bloco de entrada, uma única célula (Carrossel de erro constante), uma função de ativação de saída e conexões de "olho mágico". A

saída do bloco é recorrentemente conectada de volta ao bloco de entrada e em todos os portões.

Apesar de apresentarem grande efetividade, é necessário relatar as limitações da LSTM. Assim como descrito por Gers et al., (2000), a LSTM permite que as informações sejam armazenadas em intervalos de tempo arbitrários e que os sinais de erro sejam transmitidos no tempo. Essa força potencial, no entanto, pode contribuir para uma fraqueza em algumas situações: os estados das células tendem a crescer linearmente durante a apresentação de uma série temporal (os aspectos não lineares do processamento de sequência são deixados para as funções de esmagamento e os portões altamente não lineares). Quando apresentado um fluxo de entrada contínuo, os estados da célula podem crescer de forma ilimitada, causando saturação da função de esmagamento da saída,  $h$ .

**Figura 17** – Detalhe esquemático de uma RNN (esquerda) e uma LSTM (direita)



Fonte: Adaptado de Greff et al., (2015).

### 3. TRABALHOS RELACIONADOS

Para auxiliar esta pesquisa, foram identificados trabalhos relacionados ao Reconhecimento de Textos Escritos à Mão por meio do uso de Inteligência Artificial. Foram encontrados os termos mais comuns para esta prática, os quais serão retratados a seguir.

1) *Optical Character Recognition* (OCR) o termo foi bastante observado nos trabalhos analisados para denominar a técnica de obter o texto escrito ou impresso e convertê-lo para o formato digital. 2) *Handwritten Text Recognition* (HTR) foi observado em estudos mais recentes e é reconhecida como a técnica moderna para reconhecimento de textos, por focar em reconhecer todos os caracteres em uma linha de texto, diferente das técnicas tradicionais que segmentam caracteres individualmente a fim de reconhecê-los.

Foram identificadas duas formas de Reconhecimento de Textos Escritos à Mão, referenciados como: 1) reconhecimento *On-line* e 2) reconhecimento *Off-line*. A diferença está no fato de que para o reconhecimento *On-line* são utilizados equipamentos especiais, como canetas e quadros, capazes de identificar os caracteres no momento da escrita. Já no reconhecimento *Off-line* o texto é reconhecido após a escrita, sem a necessidade de equipamentos especiais, reconhecendo documentos digitalizados.

Foram identificados na literatura trabalhos relacionados a este projeto e as técnicas propostas por eles foram fundamentadas por uma revisão sistemática de literatura relacionada ao HTR. Estes estudos serão descritos a seguir.

Um grupo de estudantes indianos publicou em 2016 um trabalho de reconhecimento de caracteres escritos à mão HTR utilizando o reconhecimento *Off-line*. O objetivo do trabalho era utilizar a técnica de OCR para o uso médico, extraíndo a prescrição de receitas médicas. Neste trabalho, Kumar, (2016) afirma que o reconhecimento de escritas à mão significa extrair as palavras escritas no papel. Ao explicarem a técnica de OCR, os mesmos autores dizem que esta técnica é utilizada para obter o texto impresso ou escrito à mão de um documento digitalizado para enfim convertê-lo ao formato digital. O autor afirma que o processo principal do reconhecimento *Off-line* é medir a distância, criar protótipos, extrair recursos, processar, classificar e obter conhecimento dos dados de treinamento. Além do mais, os autores apontam três técnicas de classificação que são geralmente utilizadas.

Estas técnicas de classificação utilizam abordagem estatística, cujos recursos extraídos são rotulados pelo uso de médias estatísticas. As técnicas abordadas foram **Neural Networks, Backpropagation Algorithms** e o **K-nearest Neighbour**.

A publicação de 2018 “Medicine Box: Doctor’s Prescription Recognition Using Deep *Machine Learning*” relata um experimento voltado para o reconhecimento da prescrição médica escrita à mão contendo um formulário de instruções que descrevem os remédios indicados ao paciente pelo médico. Segundo E.Kamalanaban et al., (2018), ao receber uma nova prescrição médica, não é possível saber qual o nome do remédio contido na mesma; algumas vezes por utilizar abreviações em Latim, outrora por utilizar terminologias médicas na prescrição, tornando-a irreconhecível para as pessoas em geral que não estão habituadas às terminologias médicas. Para que qualquer pessoa conseguisse entender a prescrição médica escrita à mão e obter o nome dos medicamentos, E.Kamalanaban et al., (2018) utilizaram uma abordagem com Deep *Machine Learning* (DML), por meio do TensorFlow (TF).

O *TensorFlow* foi utilizado para criar uma Convolutional Neural Networks (CNN), com o auxílio de um dataset de textos escritos à mão, disponibilizado pelo IAM Handwriting Database (IAM). Após criar a CNN com o dataset, uma Recurrent Neural Network (RNN) do tipo Long Short-Term Memory (LSTM) foi utilizada para prever o caractere com base numa lista de medicamentos. As técnicas e ferramentas utilizadas foram **Deep Machine Learning, Convolutional Neural Network, Recurrent Neural Network, Long Short-Term Memory, TensorFlow** e **IAM Handwriting Database**.

Os métodos utilizados por E.Kamalanaban et al., (2018) vão ao encontro da Revisão Sistemática de Literatura (RSL) feita por Memon et al., (2019), que evidenciam o uso de RNN do tipo LSTM para a leitura de caracteres escritos à mão. Publicada em 2019, a RSL de Memon et al., (2019) providencia a direção para estudos relacionados ao HTR, por meio da identificação e apresentação da literatura relacionada ao OCR utilizando a formulação de questões de pesquisa e selecionados os estudos relevantes. Foi seguido pela proposta estratégica de Kitchenham et al., (2010), utilizando os critérios de inclusão e exclusão, processo de estratégia de busca, processo de seleção e processo de extração e síntese de dados. Resumindo os métodos para a revisão:

- a) Resumir o trabalho de pesquisa existente (técnicas de *Machine*

*Learning* e base de dados) em sistemas de reconhecimento de caracteres escritos à mão em diferentes línguas;

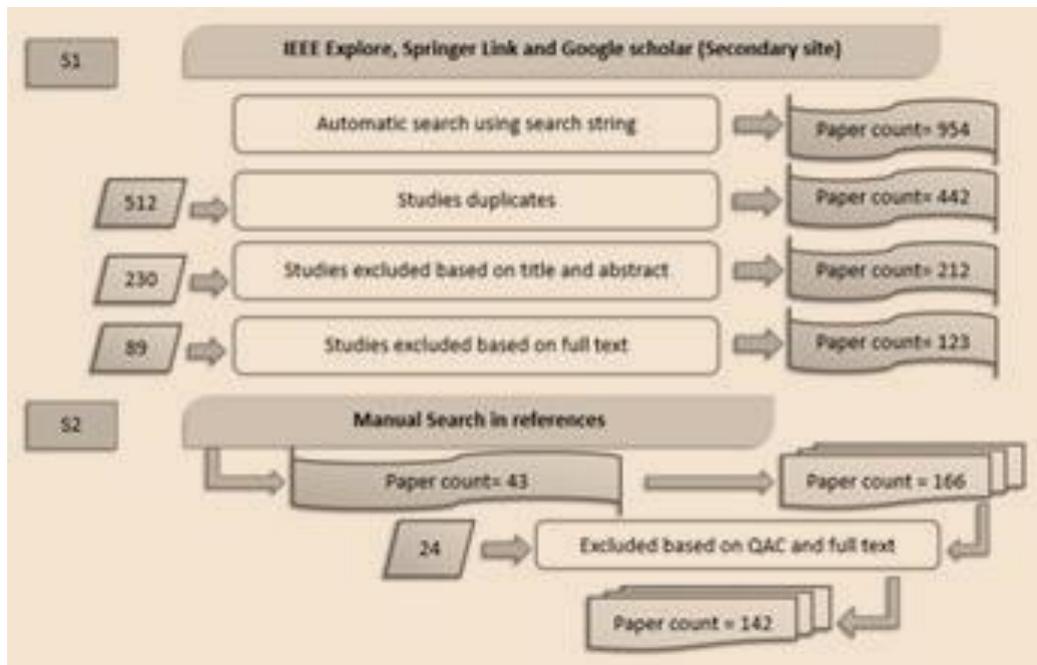
- b) Destacar as fraquezas da pesquisa para eliminá-las por meio de pesquisas adicionais;
- c) Identificar novas áreas de pesquisas no domínio do OCR.

Para garantir que apenas os artigos relevantes para o estudo fossem incluídos, foi usado o critério de inclusão e exclusão. Foram consideradas as pesquisas de revistas, conferências, simpósios e workshops em OCR, nas seguintes línguas:

- a) Inglês;
- b) Urdo;
- c) Árabe;
- d) Persa;
- e) Indiano e;
- f) Línguas Chinesas.

A estratégia de pesquisa foi dividida em pesquisa automática e manual, como mostra o resumo completo do processo de seleção de estudos na figura 18.

**Figura 18 – Estratégia de Pesquisa**



Fonte: Adaptado de Memon et al., (2019).

A pesquisa automática ajudou a encontrar os estudos primários. Mas, a revisão foi estendida com a inclusão de estudos adicionais, por meio da identificação de referências usadas nos trabalhos encontrados pela pesquisa automática. As bases incluídas na pesquisa foram:

- a) IEEE Explore;
- b) ISI Web of Knowledge;
- c) Elsevier e;
- d) Springer.

Foram consideradas como *keywords* derivadas das possibilidades de permutação do conceito de Optical Character Recognition. Foi utilizada uma ferramenta para gerenciamento de bibliografias - Mendeley - para organizar e referenciar os artigos encontrados na pesquisa. O Mendeley também ajudou na identificação de estudos duplicados, porque o mesmo artigo pode ser encontrado em várias bases de dados. Foram encontradas 954 pesquisas por meio da pesquisa automática, a maioria foi eliminada por estar duplicada, somando 512. O critério de inclusão e exclusão baseado no título, resumo, palavras-chave e tipo de publicação

eliminaram mais 230 estudos, restando 212. Então, foi aplicado o critério de seleção, eliminando mais 89 estudos, restando 123. A checagem de artigos relevantes foi feita por meio das referências usadas nos 123 artigos. Com isso, mais 43 artigos foram adicionados. A lista pré-final de estudos primários foi de 166. O último passo foi aplicar o Critério de Avaliação de Qualidade (QAC) na lista de 166 estudos. Com a isso, a lista final de estudos para a RSL foi obtida. Por fim, mais 24 estudos foram removidos, restando 142 estudos primários. A tabela 6 apresenta a distribuição dos dados selecionados nas bases de dados antes e depois do processo de seleção:

**Tabela 1** – Distribuição dos estudos selecionados nas bases de dados e depois da aplicação do processo de seleção.

<b>Fonte</b>	<b>Contagem antes de aplicar o processo de seleção</b>	<b>Contagem após aplicar o processo de seleção</b>
Elsevier	207	51
IEEE Explore	293	46
Springer	273	19
Others	182	26
Total	955	142

Fonte: Dados da própria pesquisa (2021).

O critério de avaliação de qualidade ajudou na identificação da força das inferências e na seleção dos estudos de pesquisa mais relevantes para a pesquisa desenvolvido por Memon et al., (2019). As perguntas para o QAC, foram:

- a) Os tópicos apresentados no artigo de pesquisa são relevantes aos objetivos da RSL?
- b) O estudo de pesquisa descreve o contexto da pesquisa?
- c) O artigo de pesquisa explica a abordagem e a metodologia da pesquisa com clareza?
- d) O processo de coleta de dados foi explicado, caso a coleta tenha sido feita no estudo?
- e) O processo de análise de dados explicou com exemplos apropriados?

Seguindo este critério, 142 estudos foram selecionados para o artigo de revisão. A RSL desenvolvida por Memon et al., (2019) analisou os métodos de classificação de escrito à mão, descrevendo seu processo de aplicação. Os autores categorizam os algoritmos de OCR escritos à mão como sendo algoritmos treinados em uma coleção de dados conhecida. Desta forma, o algoritmo descobre como categorizar/classificar com precisão os dígitos e as letras do alfabeto. Com isso, foram discutidas as técnicas de classificação de OCR mais prevalentes nos estudos entre os anos 2000 e 2018, sendo:

- a) Redes Neurais Artificiais;
- b) Métodos kernel;
- c) Métodos estatísticos;
- d) Técnicas de correspondência de modelos e;
- e) Reconhecimento de padrões estruturais.

A revisão sistemática de literatura desenvolvida por Memon et al., (2019) fortaleceu as evidências encontradas nos estudos desenvolvidos por Kumar, (2016) e E.Kamalanaban et al., (2018), que utilizaram *Artificial Neural Network* para o processo de reconhecimento de textos escritos à mão. Além do mais, Memon et al., (2019) evidenciam o uso das RNN em vários trabalhos. Em especial, as CNN foram apontadas como portadoras de resultados na tarefa de reconhecimento de caracteres. As CNN foram utilizadas para classificação e reconhecimento em quase todas as línguas que foram revisadas pela revisão sistemática de literatura. Comprovadas as evidências da eficiência das CNN na tarefa de reconhecimento de textos escritos à mão, e por apresentar semelhança aos objetivos, serão utilizadas como base para o desenvolvimento deste trabalho as técnicas propostas por E.Kamalanaban et al., (2018).

## 4. METODOLOGIA

Para Lana et al., (2018), a metodologia de pesquisa deve ser feita com calma e após muita reflexão e estudo. É muito importante definir a metodologia durante o desenvolvimento da pesquisa, logo que toda a análise dos resultados pode ser comprometida. A mesma deve ser adequada ao objetivo de pesquisa, à área de colaboração, às características dos dados, entre outros fatores. A metodologia tem a finalidade de orientar o pesquisador, garantindo que os objetivos propostos em sua pesquisa sejam cumpridos.

Para Trevisol Neto, (2018), todas as ciências são caracterizadas pelos métodos científicos, já que não existe ciência sem o emprego de métodos científicos. O método escolhido apresenta a posição teórica que foi explicada para evidenciar a abordagem do estudo.

### 4.1. Caracterização da Pesquisa

Por ter como objetivo estudar as técnicas e métodos de HTR para o reconhecimento de patologias descritas na indicação clínica contida em guias de serviços pode-se dizer que esta pesquisa utilizou o método de **pesquisa descritiva**, sendo caracterizada como de **natureza aplicada**. Segundo Trevisol Neto, (2018), o método de pesquisa descritiva tem como característica a utilização de técnicas padronizadas para a coleta de dados. Nessa modalidade não há manipulação dos fenômenos estudados, sendo os fatos registrados, analisados, classificados e interpretados sem interferência do pesquisador. Ao explicar a pesquisa aplicada, Trevisol Neto, (2018) diz que essa tem o propósito de buscar solução para problemas concretos e com fins práticos, por meio da aplicação do conhecimento.

É de **abordagem quantitativa**, por utilizar métodos estatísticos e matemáticos para avaliar a acurácia e precisão das técnicas e métodos sugeridos para o reconhecimento de textos escritos à mão. Para Trevisol Neto, (2018), a pesquisa quantitativa possui como característica a objetividade, pressupondo que tudo pode ser quantificado, traduzindo resultados em números. Tem a intenção de obter a precisão dos resultados, evitando distorção na interpretação da análise.

A técnica de **pesquisa é experimental**, logo que será avaliada uma técnica

e método para reconhecimento de textos escritos à mão. A classificação da pesquisa acontece conforme os procedimentos técnicos empregados no estudo. É definida após a resolução das questões conceituais da pesquisa, delineando a mesma. De acordo com Trevisol Neto, (2018), a pesquisa experimental consiste em determinar um objeto de estudo, identificar as variáveis capazes de influenciá-la e definir o controle e a observação dos efeitos das variáveis no objeto.

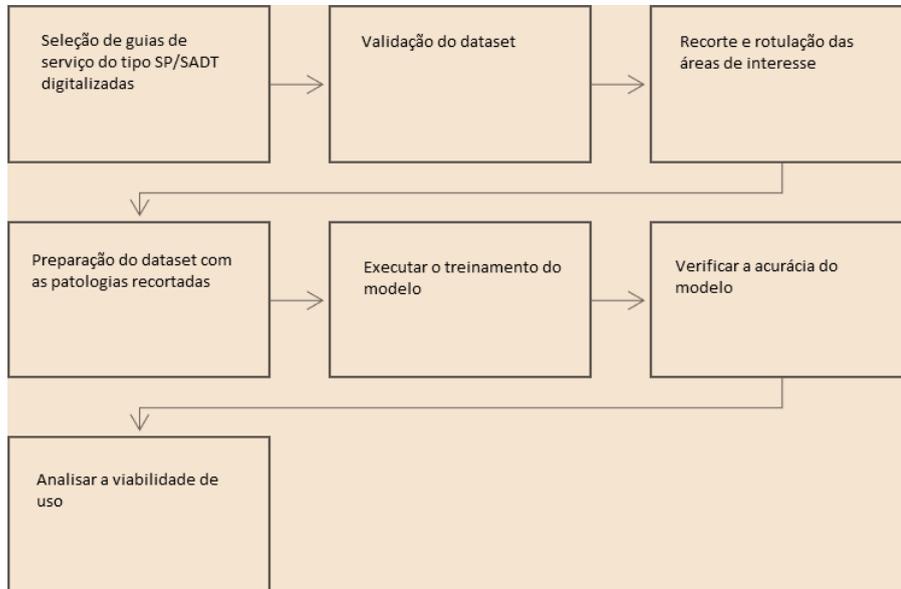
#### **4.2. Descrição dos Dados**

Os dados que foram utilizados no processo de reconhecimento da escrita à mão são os contidos em guias de serviços reais, disponibilizadas por uma empresa de auditoria responsável por uma operadora de planos de saúde. Os médicos, ao solicitarem a liberação de serviços, fornecem a guia escrita à mão em imagem digitalizada.

As guias consideradas para o experimento foram apenas guias do tipo SP/SADT. O modelo de guia SP/SADT pode ser observado na figura 19, com destaque no campo indicação clínica na imagem 20. Além do mais, apenas o campo “Indicação Clínica” foi analisado. As guias foram coletadas no momento da análise para liberação dos procedimentos. As guias são enviadas, normalmente, como anexo pelos profissionais solicitantes através do portal de solicitação de procedimentos. Em seguida elas entraram para fila de análise e aguardam pela avaliação de um profissional. A seleção das guias foi realizada com o auxílio dos enfermeiros e médicos responsáveis pela análise. Estes também foram responsáveis pela extração do campo de interesse, neste caso, a indicação clínica. Em seguida, a imagem foi rotulada com a indicação clínica e a descrição da patologia.



**Figura 21** – Representação visual da metodologia Fonte: Autor



Fonte: dados da pesquisa (2021).

- a) O primeiro passo foi a coleta e seleção das guias de serviços do tipo SP/SADT, em formato de documento digitalizado, que foram enviadas pelos profissionais solicitantes. O objetivo foi avaliar as guias de serviços em um modelo específico, para extração dos documentos, e seus respectivos metadados, que atenderem ao requisito.
- b) Com o dataset de guias de serviços já criado, foi necessário analisar os documentos recuperados, para garantir que as guias tivessem o campo "indicação clínica" no formato esperado. Como não há regra para a utilização deste campo, alguns textos poderiam não condizer com o objetivo esperado, além de poderem conter informações irrelevantes, gerando ruídos desnecessários.
- c) Nesta etapa a região de interesse, ou seja, o campo "indicação clínica", foi analisado, recortado e rotulado pelo especialista responsável pela análise da guia. Sendo assim, garantimos a correta classificação do texto e da patologia.

- d) O dataset de patologias consiste na descrição da patologia recortada do campo "indicação clínica".
- e) Neste momento já estava de posse de todos os dados necessários para aplicar as técnicas de inteligência artificial para o reconhecimento da patologia, sendo utilizado uma CNN para a realização de tal tarefa.
- f) Após o treinamento serão realizados testes de acurácia para avaliar a capacidade do modelo em reconhecer a patologia.
- g) Com base nos resultados será realizada a análise de viabilidade de utilização prática do modelo proposto neste trabalho, levando em conta a taxa de acurácia.

#### 4.4. Validação dos Resultados

Para avaliar os resultados alcançados pelo experimento foi utilizada uma matriz de confusão. A estrutura da matriz foi definida por 4 classes, onde foram apresentados os valores de entrada e os valores alcançados pelo algoritmo 2.

**Tabela 2** – Matriz de confusão utilizada para avaliar os resultados.

	DRGE	Dor Abdom.	HAS	Osteoporose
DRGE	VP	FP	FP	FP
Dor Abdominal	FP	VP	FP	FP
HAS	FP	FP	VP	FP
Osteoporose	FP	FP	FP	VP

Fonte: Elaborada pelo próprio autor (2021).

Por meio da matriz de confusão foi possível extrair alguns indicadores necessários para a avaliação dos resultados. Assim, foram extraídos os indicadores reportados no relatório de classificação, conforme tabela 3.

**Tabela 3** – Relatório de classificação.

	Precisão	Recall	F1-score	Nº ocorrênc.
<b>DRGE</b>	%	%	%	X
<b>Dor Abdominal</b>	%	%	%	X
<b>HAS</b>	%	%	%	X
<b>Osteoporose</b>	%	%	%	X
<b>Acurácia</b>			%	X
<b>Macro avg</b>	%	%	%	X
<b>Weighted avg</b>	%	%	%	X

Fonte: Elaborada pelo próprio autor (2021).

As métricas avaliadas foram fornecidas diretamente pela ferramenta utilizada na criação do modelo, sendo extraídas em forma de relatório.

## **5. EXECUÇÃO DO EXPERIMENTO**

Este capítulo apresentou os resultados da preparação, execução e análise dos dados. Primeiramente, abordou-se a preparação dos dados a fim de esclarecer os critérios utilizados na seleção das guias do tipo SP/SADT. Em seguida, foi executada a técnica de HTR, que era a etapa responsável pelo reconhecimento das patologias escritas à mão na indicação clínica de guias do tipo SP/SADT. Por fim, a análise da acurácia da técnica de HTR para o reconhecimento de patologias na indicação clínica é apresentada.

### **5.1. Preparação dos dados**

O processo de captação e análise de guias aconteceu de acordo com o seguinte processo: 1) Captação e separação das guias; 2) Avaliação das guias selecionadas e; 3) Rotulação e recorte da área de interesse. A distribuição das imagens geradas foi realizada na execução da técnica de HTR.

#### **5.1.1. Captação e Separação das Guias**

Inicialmente, era esperado o fornecimento das guias do tipo SP/SADT pela empresa responsável pela auditoria das OPS. Porém, com a recém-chegada da LGPD, houve dificuldade por parte da empresa auditora em fornecer as guias médicas, uma vez que há informações sensíveis sobre os beneficiários. Com isso, foi levado ao setor jurídico da empresa o pedido de liberação legal das guias médicas para uso em pesquisa acadêmica. Este processo, por ser burocrático, retardou a coleta das guias médicas, atrasando o cronograma previsto para realização deste experimento. Por fim, foi aconselhado o contato direto com as OPS para requerer a liberação das guias médicas; a coleção de guias de três OPS foram colocadas à disposição para realização deste experimento.

As operadoras que se dispuseram a fornecer as coleções de guias médicas perceberam a importância desta pesquisa, demonstrando interesse e prontidão para colaborar com o que fosse necessário.

Cada operadora forneceu aproximadamente 5 anos de guias médicas

armazenadas em formato digital, sendo que não havia separação por tipo de guia, nem mesmo de guias digitadas em computador ou escritas à mão. Para diminuir a variação entre o padrão da escrita, foi selecionada a coleção de guias de apenas uma das três operadoras, logo que há recorrência entre os médicos/prestadores solicitantes por participarem de contrato ativo com a operadora. Não houve separação por período, podendo haver guias dentre os 5 anos participando do experimento.

Após selecionar a coleção de guias, foram necessários os seguintes passos para eleger as guias que seriam postas para rotulação:

- a) Identificar e separar guias do tipo SP/SADT;
- b) Identificar e separar guias digitadas em computador das guias escritas à mão;
- c) Pré-selecionar guias com a indicação clínica preenchida;
- d) Eliminar guias ilegíveis.

Ao terminar esta etapa, percebeu-se a necessidade de revisão das guias selecionadas para rotulação. O motivo e o processo serão descrito a seguir.

#### 5.1.2. Revisão das Guias

Durante a etapa de separação das guias que seriam colocadas para rotulação, percebeu-se a necessidade de revisá-las e padronizá-las a fim de melhorar a qualidade do processo de rotulação.

O processo consistiu em repassar todas as guias selecionadas, 1) eliminando aquelas que deveriam ter sido descartadas na etapa anterior; 2) configurando-as na mesma orientação e rotação e 3) analisando superficialmente as patologias mais recorrentes.

#### 5.1.3. Rotulação das Guias

Para a rotulação das guias, foram selecionados três profissionais da área da saúde, sendo um enfermeiro chefe e outros dois enfermeiros, sob a coordenação do primeiro. Os três possuem experiência com análise de guias, sendo o enfermeiro

chefe coordenador em uma empresa de auditoria de guias médicas. A experiência dos enfermeiros em analisar guias foi um fator de seleção, já que a escrita médica é de difícil compreensão e mesmo os enfermeiros que não possuem experiência com análise de guias podem ter dificuldade em interpretá-la.

Foram disponibilizadas 5 mil guias para rotulação, não sendo distribuídas em montantes iguais para cada rotulador, ficando livre a quantidade a rotular entre cada um. Vale informar que, para este processo, foi estimado um valor a ser pago pelo montante de 5 mil guias para o enfermeiro chefe, que ficou a cargo de coordenar o processo de rotulação, sendo que este montante seria dividido entre os participantes. Foi acordado um valor total de R\$1.200,00 pelo montante de 5 mil guias. Ficou estabelecido um prazo de duas semanas para a realização das tarefas, logo que este trabalho seria um "extra" para os profissionais, não havendo compromisso em horário demarcado, por todos estarem empregados e ocupando seus cargos em tempo integral. A tabela a seguir ilustra o quadro de trabalho dos rotuladores, ressaltando que o coordenador não atuou no processo de rotulação.

**Tabela 4 – Quadro de trabalho dos rotuladores.**

<b>Rotulador</b>	<b>Montante de guias</b>	<b>Prazo limite</b>	<b>Pagamento</b>
Rotulador 1	até 5.000	2 semanas	até R\$ 1.200,00
Rotulador 2	até 5.000	2 semanas	até R\$1.200,00
Coordenador	0	0	0

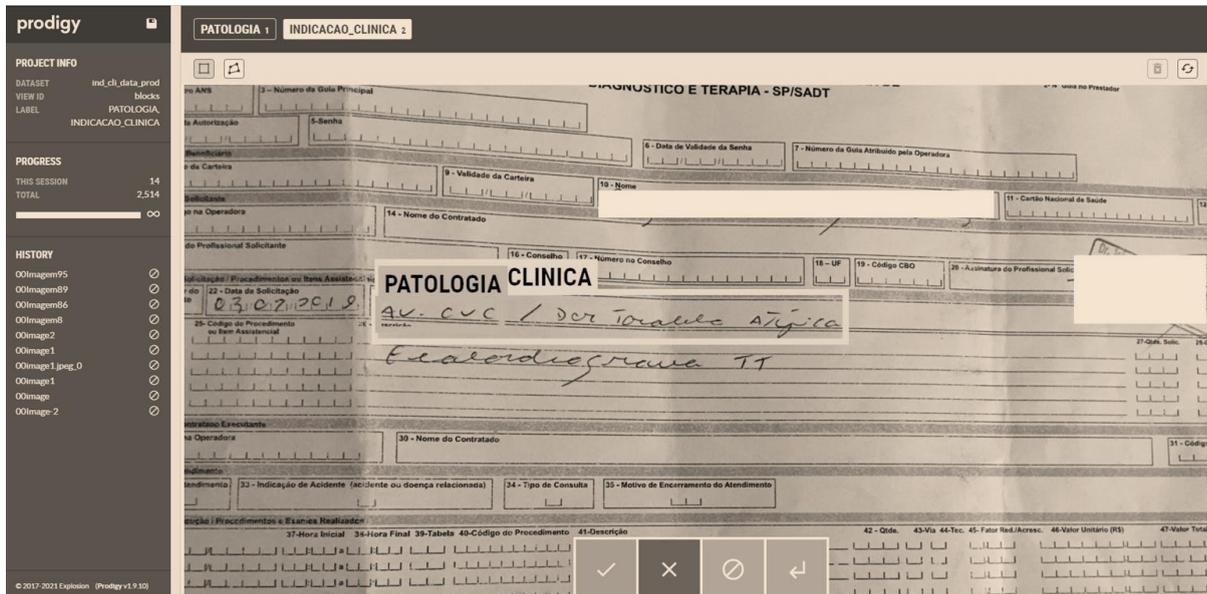
Fonte: Elaborado pelo pesquisador (2021).

O processo de rotulação foi realizado através de um *software* com acesso online via navegador. As regras de rotulação foram:

- a) demarcar a região com a indicação clínica, e;
- b) demarcar a região com a descrição da patologia.

Foram inseridas as opções de aceitar, recusar, ignorar ou desfazer uma guia. É possível ter uma visão do funcionamento do *software* na figura a seguir:

**Figura 22 – Software para rotulação de imagens.**



**Fonte:** Dados da pesquisa (2021).

Cada rotulador possui um usuário, ao qual ficou vinculado uma sessão. Assim foi possível analisar o trabalho individual. Vale ressaltar que foi avaliado o montante de guias analisadas por cada rotulador, não só as que foram marcadas como "aceitas".

No início, houve certa dificuldade ao utilizar a ferramenta, uma vez que o carregamento das imagens estava lento, o que gerou frustração nos rotuladores. Com isso, o ambiente de hospedagem da ferramenta foi melhorado, diminuindo a lentidão no carregamento das imagens. Apesar da frustração dos rotuladores, a qualidade do processo não foi comprometida.

Apesar da expectativa de conclusão de 5 mil guias em 3 semanas, após 2 semanas de trabalho apenas 2.500 guias foram rotuladas. Isto implicou em outro atraso no cronograma, levando ao limite do prazo disposto para o processo de rotulação. Com isso, foi necessário parar o processo de rotulação com apenas 2.500 guias disponíveis para a execução da técnica de HTR, o que poderia comprometer o resultado do experimento devido à baixa quantidade de dados. O resultado do processo de rotulação pode ser visto na tabela a seguir.

**Tabela 5** – Status do *dataset* do processo de rotulação.

<b>Anotações</b>	<b>Aceitas</b>	<b>Rejeitadas</b>	<b>Ignoradas</b>
2.500	1.854	621	25

Fonte: Dados da pesquisa (2021).

No montante de 2.500 guias entregues pelos rotuladores, apenas 1.854 estão aptas a serem incluídas na seleção de guias para a realização da prática de HTR. Com isso, já é possível analisar a distribuição de patologias entre as guias. A tabela a seguir disponibiliza o montante de guias por patologias, ordenado do maior para o menor.

**Tabela 6** – Distribuição de patologias entre as guias elegíveis.

<b>Patologia</b>	<b>Ocorrências</b>
HAS	110
Osteoporose	38
DRGE	31
Dor abdominal	29
Climatério	25
...	...
Dispineia aos pequenos esforços	1
Laceração de lóbulo de orelha	1
Lesão na face	1
Hipotireoidismo	1
Lesão manguito ombro D	1

Fonte: Dados da própria pesquisa (2021).

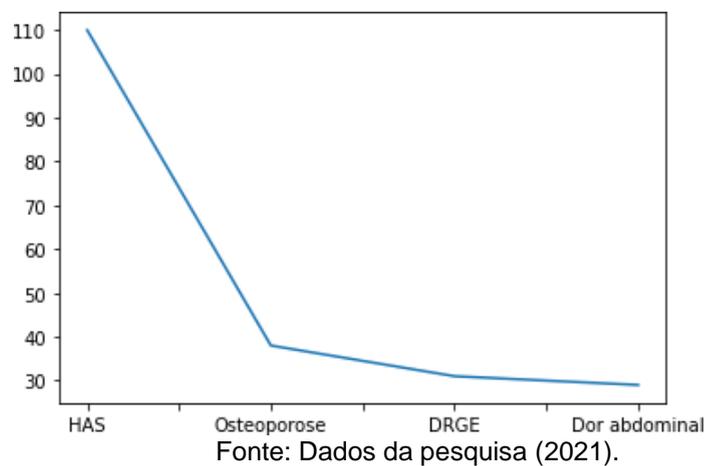
É perceptível o desbalanceamento entre a ocorrência das patologias nas guias, o que foi contra o que se esperava inicialmente, ficando evidente a maior ocorrência de determinadas patologias. É possível analisar a variação entre a ocorrência das patologias no gráfico representado pela figura 23.

Notou-se qualidade na demarcação das áreas de interesse no processo de rotulação, mas a quantidade de guias elegíveis para o processo de execução da

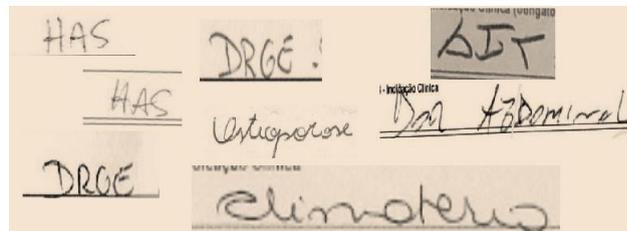
técnica de HTR pode comprometer o resultado esperado. A figura 24 demonstra alguns recortes de patologias rotuladas.

O processo de rotulação também foi responsável por atrasar o cronograma de realização deste experimento, além de não obter o resultado esperado. O montante de guias elegíveis para o próximo processo é baixo, e a variação entre a ocorrência de patologias foi acima do esperado. Todos estes fatores inferiram risco na execução da próxima etapa, podendo comprometer o resultado final.

**Figura 23** – Gráfico de variação entre a ocorrência de patologias.



**Figura 24** – Recortes de patologias rotuladas.



Fonte: Dados da pesquisa (2021).

## 5.2. Caracterização da Técnica de HTR

As práticas de HTR não estão limitadas em ANN, mas estas apresentam resultados promissores em termos de velocidade e acurácia. Métodos baseados em *Deep Learning* têm chamado atenção no reconhecimento de caracteres escritos à mão (ALOM et al., 2018).

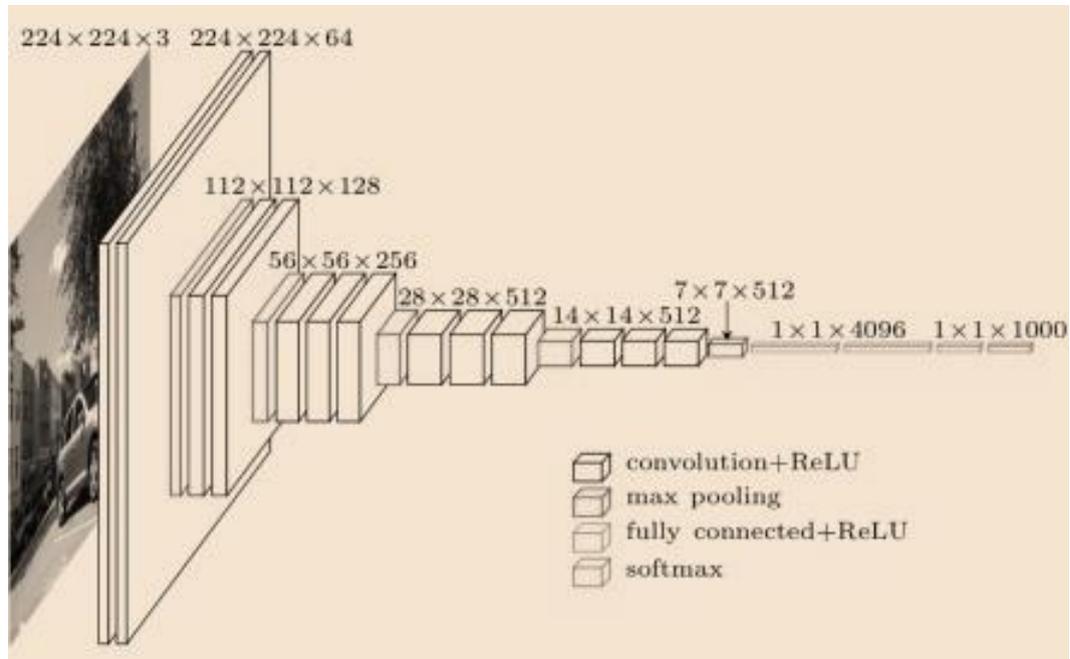
O mesmo autor cita vários artigos relacionados ao assunto, trazendo referências a utilização de CNN como método de reconhecimento. Além disso, ao proporem um modelo HCR para o reconhecimento de caracteres da língua bangali, Alom et al., (2018) introduzem alguns modelos de *Deep Convolutional Neural Network* (DCNN) na prática de *Handwritten Character Recognition* (HCR), incluindo as redes VGG.

Existem três tipos de *Visual Geometry Group Nets* (VGG Net) baseadas na arquitetura. Estas três arquiteturas possuem 11, 16 e 19 camadas, e são chamadas de VGG-11, VGG-16 e VGG-19, respectivamente. A estrutura básica de uma arquitetura VGG-11 possui 8 camadas de convolução, uma camada de "*max-pooling*" e 3 camadas totalmente conectadas (FC) seguidas de uma única camada "*Softmax*".

A configuração de uma VGG- 16 é a seguinte: o número de camadas de convolução são 13 e de "*max-pooling*" 1, 3 camadas FC, sendo 1 camada "*Softmax*". O número total de pesos é de 138 milhões. A VGG-19 consiste em 16 camadas de convolução, 1 camada de "*max-pooling*" e 3 camadas FC seguida por uma camada "*Softmax*".

A VGG16 é uma arquitetura de CNN utilizada no reconhecimento de imagens há alguns anos. Em 2014 esta arquitetura se destacou na competição *Image Net Large Scale Visual Recognition Competition* (ILSVRC), tornando-se muito popular na comunidade de pesquisa por conta de sua abordagem simples e por conta de seus pesos pré-treinados disponíveis gratuitamente *online*, facilitando o refinamento do modelo para realizar novas tarefas (NASH et al., 2018). É possível verificar a arquitetura desta NN através da figura 25.

**Figura 25** – Arquitetura de uma VGG16.

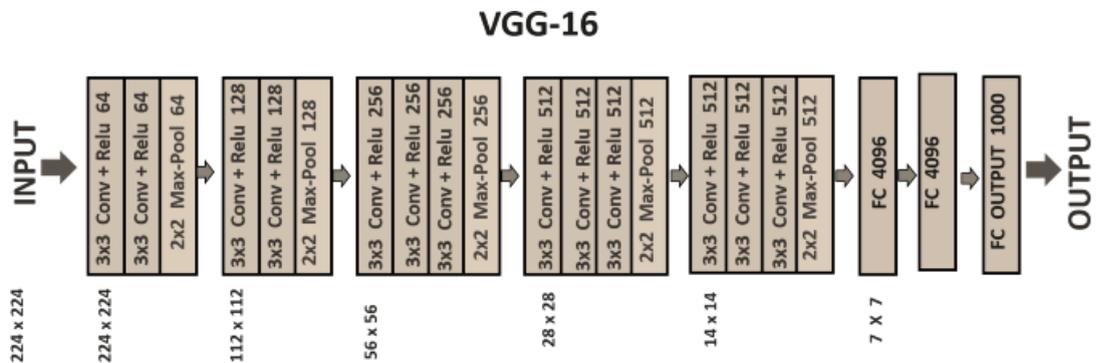


Fonte: Adaptado de Nash et al., (2018).

Ao detalharem a estrutura da arquitetura VGG-16, Deore e Pravin, (2020) descrevem a consistência da arquitetura contendo 13 camadas, 5 camadas de "*max-pooling*" e 3 camadas densas. Conv1 possui 64 filtros de 3x3 núcleos, enquanto Conv2 possui 128 filtros de 3x3 núcleos, já Conv3 possui 256 filtros de 3x3 núcleos, Conv4 e Conv5 possuem 512 filtros de 3x3 núcleos. Todas as camadas de "*max-pooling*" possuem 2x2 núcleos. A função de ativação do tipo "ReLU" é adicionada para cada camada de convolução e nas duas camadas densas, assim valores negativos não serão enviados para as camadas subsequentes. A função de ativação "*Softmax*" é usada pela última camada densa de saída como camada de predição.

Hoje em dia, a arquitetura VGG-16 é considerada como uma excelente arquitetura para modelos de visão computacional. Uma das características da VGG16 é que ao invés de possuir um enorme número de hiper-parâmetros ela conta com mais camadas convolucionais. É possível verificar a arquitetura VGG-16 em sua forma simplificada na figura 26 a seguir.

**Figura 26** – Arquitetura simplificada de uma VGG16.



Detalhes de uma NN do tipo VGG-16:

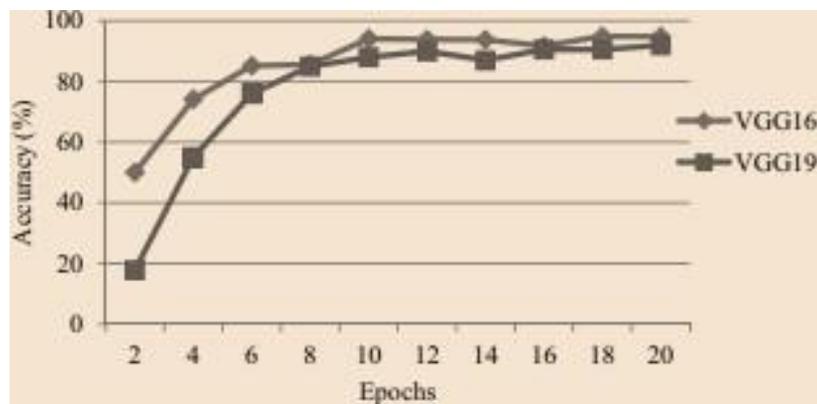
- a) **Camada de entrada:** aceita imagens coloridas e com tamanho de 224x224 e três canais como entrada;
- b) **Camada convolucional:** as imagens passam por uma pilha de camadas convolucionais onde cada filtro convolucional tem um pequeno campo receptivo de 3x3 e "stride" 1. Cada núcleo convolucional utiliza o preenchimento de uma linha e uma coluna para que o tamanho da entrada permaneça o mesmo que os mapas de recursos de saída ou em outras palavras, a resolução após a convolução permanece a mesma;
- c) **Max pooling:** é executada em uma janela de "max-pool" com tamanho 2x2 e "stride" de 2, o que significa que as janelas de "max-pool" não são janelas sobrepostas;
- d) Nem todas as camadas convolucionais são seguidas por uma camada "max-pool", já que em alguns locais as camadas convolucionais são seguidas por outras camadas convolucionais sem a camada "max-pool" entre elas.
- e) As duas primeiras camadas totalmente conectadas possuem 4.096 canais cada e a terceira camada totalmente conectada, que também é a camada de saída, possui 1.000 canais, um para cada categoria de imagens na base de dados imagenet.

- f) As camadas ocultas possuem uma ReLU como função de ativação.

Ao realizarem o processo de HCR no alfabeto devanágari, Deore e Pravin, (2020) mencionam o uso da arquitetura VGG-16 em um *dataset* com baixa taxa de amostragem.

Outro detalhe apontado pelos autores é a simplicidade da arquitetura VGG e sua adequação ao "*fine tuning*" para melhorar os resultados em datasets pequenos. Neste mesmo trabalho, os autores fizeram um teste comparativo entre as variações da arquitetura VGG- 16 e VGG-19, que concluem que o modelo criado por meio da arquitetura VGG-16 obteve melhor performance, requerendo menos espaço de armazenagem e melhor acurácia 27, alcançando 96,55% de acurácia.

**Figura 27** – Comparação entre VGG-16 e VGG-19. Fonte:



Fonte: Deore ePravin, (2020).

Esta arquitetura emoldurou o processo utilizado na configuração da CNN para a realização do HTR proposto neste trabalho, que se deu pelos passos seguintes.

### 5.3. Preparação do Modelo

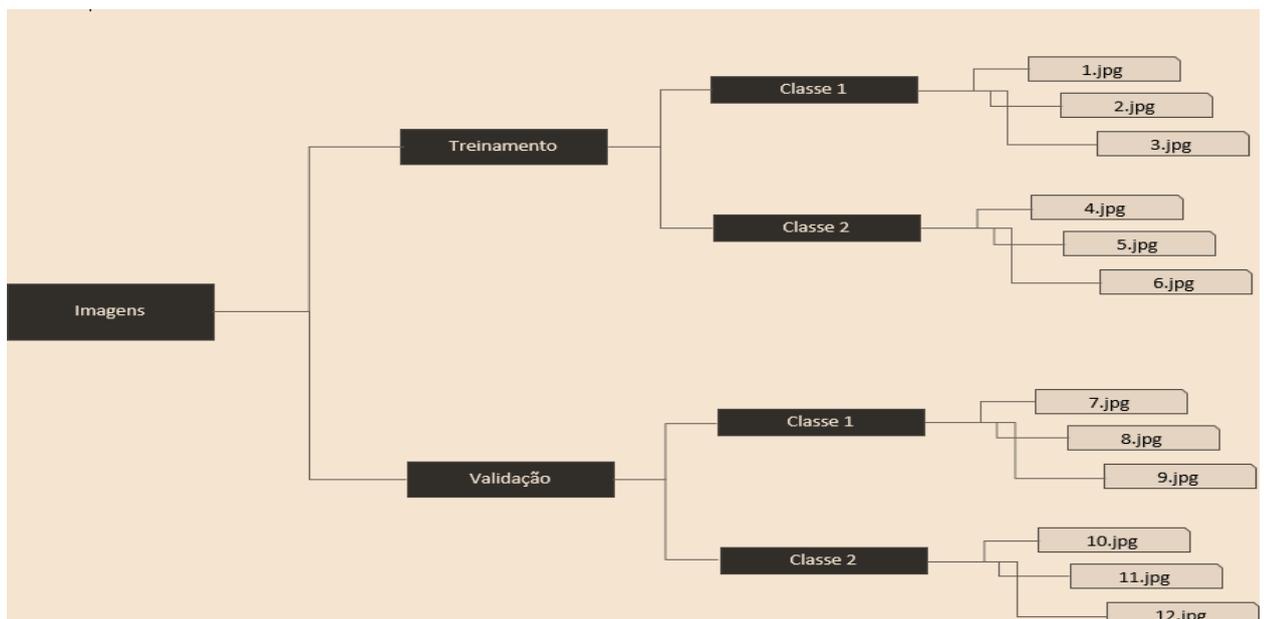
O ambiente de desenvolvimento adotado para a execução do experimento foi o Google Colab. O dataset de imagens foi carregado para o Google Drive para facilitar o acesso. Foi utilizado o *framework* Keras em conjunto com a biblioteca *TensorFlow*. Inicialmente a CNN foi configurada seguindo o modelo padrão da arquitetura VGG-16, contendo o mesmo número de "*features maps*" nas camadas convolucionais, porém foram aplicados diferentes ajustes durante a avaliação do modelo, originando 7

modelos para avaliação: A, B, C, D, E, F e G. Cada modelo surgiu a partir da avaliação do modelo antecedente, na tentativa de melhorar os resultados obtidos durante o treinamento. A configuração inicial do modelo padrão será descrita a seguir.

Foi utilizado o método sequencial, conforme é proposto pela arquitetura. O modelo sequencial determina que todas as camadas do modelo sejam arranjadas em sequencia. Para simplificar a entrada de imagens, foi utilizado o gerador de dados de imagens disponibilizado pelo Keras, que permite importar dados rotulados facilmente através de subdiretórios. A estrutura de diretórios pode ser vista na figura 28.

Foram criados dois geradores de dados: 1) "*TrainDataGen*"- responsável por fornecer dados de treinamento para o modelo e 2) "*TestDataGen*"- responsável por fornecer dados de avaliação, cada um apontado para seu devido diretório. Em ambos foi configurado o atributo "*target\_size*" para 224x224, configurando o gerador de dados para alterar a resolução das imagens sem alterar o arquivo original. Em seguida, foram definidas as camadas convolucionais e as camadas de "*max-pool*", seguindo a proposta inicial da arquitetura, totalizando 13 camadas convolucionais e 5 camadas "*maxpool*", conforme se vê abaixo.

**Figura 28** – Estrutura de pastas passadas para o gerador de dados.



Fonte: Dados da pesquisa (2021)

- a) 2x camadas convolucionais de 64 canais com núcleo de 3x3 e com o mesmo "padding";

- b) 1x camada "maxpool" de tamanho 2x2 e "stride"2x2;
- c) 2x camadas convolucionais de 128 canais com núcleo de 3x3 e com o mesmo "padding";
- d) 1x camada "maxpool" de tamanho 2x2 e "stride"2x2;
- e) 3x camadas convolucionais de 256 canais com núcleo de 3x3 e com o mesmo "padding";
- f) 1x camada "maxpool" de tamanho 2x2 e "stride"2x2;
- g) 3x camadas convolucionais de 512 canais com núcleo de 3x3 e com o mesmo "padding";
- h) 1x camada "maxpool" de tamanho 2x2 e "stride"2x2;
- i) 3x camadas convolucionais de 512 canais com núcleo de 3x3 e com o mesmo "padding";
- j) 1x camada "maxpool" de tamanho 2x2 e "stride"2x2.

Também foi adicionada uma função de ativação do tipo "ReLU" em cada camada, assim todos os valores negativos não são encaminhados para a próxima camada. Além das camadas convolucionais e "maxpool", foram criadas camadas densas, totalizando 3 camadas densas, a saber:

- a) 1x camada densa de 4096 unidades;
- b) 1x camada densa de 4096 unidades;
- c) 1x camada densa com função de ativação Softmax de (x classes) unidades.

Foram utilizadas funções de ativação do tipo "ReLU" na duas camadas densas de 4.096 unidades, fazendo com que valores negativos não fossem passados para a próxima camada da NN. Por fim, foi utilizada uma camada densa de 4 unidades com uma função de ativação do tipo "Softmax", utilizada como camada de predição.

Após a configuração das camadas convolucionais, "maxpool" e das camadas densas, foi possível compilar o modelo. Foi utilizado o otimizador "Adam" para alcançar o mínimo global enquanto o modelo era treinado. Caso o modelo ficasse preso em um mínimo local no momento do treinamento, o otimizador ajudaria a sair do mínimo local e alcançaria o mínimo global. A especificação da taxa de aprendizado

do otimizador foi configurada para 0,001. Caso o treinamento saltasse muito nas epochs, seria necessário diminuir a taxa de aprendizado para atingir os mínimos globais. O sumário do modelo criado está disponível na figura 29.

A métrica avaliada pelo modelo foi a "*accuracy*" e o método de classificação foi o "*categorical\_crossentropy*". Após criar o modelo, foram importados os métodos "*ModelCheckpoint*" e "*EarlyStopping*" do Keras. O "*ModelCheckpoint*" salva os pesos por meio do monitoramento de um parâmetro específico do modelo. Neste caso, foi monitorada a validação da acurácia, fazendo com que os pesos fossem salvos apenas se a validação da acurácia na "*epoch*" atual aumentasse. Já o "*EarlyStopping*" interromperia o treinamento do modelo caso não houvesse ganho nos parâmetros monitorados. Neste contexto, foi monitorado o parâmetro de validação da acurácia. A configuração determinou que fossem executadas no máximo 20 "*epochs*", conforme propõe a arquitetura, sem que houvesse melhora na validação da acurácia. Foi utilizada a função de treinamento "*fit\_generator*", com os atributos "*steps\_per\_epoch*" configurado para 100, "*validation\_step*" para 10 e "*epochs*" para 100.

**Figura 29** – Sumário do modelo criado

Layer (type)	Output Shape	Param #
conv2d_14 (Conv2D)	(None, 224, 224, 64)	1792
conv2d_15 (Conv2D)	(None, 224, 224, 64)	36928
max_pooling2d_5 (MaxPooling2D)	(None, 112, 112, 64)	0
conv2d_16 (Conv2D)	(None, 112, 112, 128)	73856
conv2d_17 (Conv2D)	(None, 112, 112, 128)	147584
max_pooling2d_6 (MaxPooling2D)	(None, 56, 56, 128)	0
conv2d_18 (Conv2D)	(None, 56, 56, 256)	295168
conv2d_19 (Conv2D)	(None, 56, 56, 256)	590080
conv2d_20 (Conv2D)	(None, 56, 56, 256)	590080
max_pooling2d_7 (MaxPooling2D)	(None, 28, 28, 256)	0
conv2d_21 (Conv2D)	(None, 28, 28, 512)	1180160
conv2d_22 (Conv2D)	(None, 28, 28, 512)	2359808
conv2d_23 (Conv2D)	(None, 28, 28, 512)	2359808
max_pooling2d_8 (MaxPooling2D)	(None, 14, 14, 512)	0
conv2d_24 (Conv2D)	(None, 14, 14, 512)	2359808
conv2d_25 (Conv2D)	(None, 14, 14, 512)	2359808
conv2d_26 (Conv2D)	(None, 14, 14, 512)	2359808
max_pooling2d_9 (MaxPooling2D)	(None, 7, 7, 512)	0
flatten_1 (Flatten)	(None, 25088)	0
dense_3 (Dense)	(None, 4096)	102764544
dense_4 (Dense)	(None, 4096)	16781312
dense_5 (Dense)	(None, 3)	12291
Total params: 134,272,835		
Trainable params: 134,272,835		
Non-trainable params: 0		

Fonte: Dados da própria pesquisa (2021).

A separação e seleção das imagens deram-se por conta da quantidade de ocorrências das patologias. Foi realizada a ordenação do dataset pela quantidade de ocorrências em ordem decrescente. Em seguida, foram selecionadas as 4 primeiras patologias, determinando as classes do modelo. Com isso, foram consideradas apenas 4 classes: 1) HAS, 2) Osteoporose, 3) DRGE e 4) Dor abdominal. A baixa ocorrência de imagens levou a necessidade de retrabalhar o *dataset*, com o objetivo de aumentar a amostragem, assim foram gerados três novos datasets: 1) DS1, 2) DS2 e 3) DS3. Para isso foram utilizadas técnicas de "zoom" e "rotação" nas imagens. O DS1 foi originado a partir da cópia do dataset original aplicando as proporções de "zoom" para (0.9, 0.95, 1, 1.05 e 1.1) e "rotação" para (-5 e 5) em cada imagem, acrescentando 3.536 novas imagens. Ao DS2, foram adicionadas mais 4 proporções de "zoom" de (0.93, 0.97, 1.03 e 1.07), acrescentando 2.506 novas imagens.

Já DS3 foi criado com base em DS2, porém a quantidade de imagens disponíveis em cada classe foi reduzida ao mesmo número da classe com menor quantidade, balanceando o dataset. Foram analisadas aproximadamente 10% das imagens com "zoom" de (0.9 e 1.1) para avaliar cortes no conteúdo, mas não foram encontrados problemas. Sendo assim, DS1 totaliza 3.744 e DS2 6250 imagens referentes às 4 classes selecionadas. Vale ressaltar que foram adicionadas novas imagens para cada uma das classes, mantendo o desbalanceamento dos datasets. Já DS3 totaliza 3.480 imagens com a mesma quantidade de ocorrências entre cada uma das 4 classes.

A tabela 7 apresenta os detalhes de cada *dataset*. Por padrão a divisão das imagens entre treinamento e avaliação foi realizada através da separação de 10% do total de imagens do dataset para avaliação e 90% para treinamento. Inicialmente não houve ordenação ou embaralhamento antes da separação das imagens.

**Tabela 7** – Disposição dos *datasets*.

<b>Dataset</b>	<b>Total de amostras</b>	<b>Balanceado</b>
DS1	3.744	Não
DS2	6.250	Não
DS3	3.480	Sim

Fonte: Dados da pesquisa (2021).

Com os datasets disponíveis para o treinamento, o primeiro modelo foi

executado, seguindo as configurações iniciais indicadas anteriormente. Na sequência, foram executadas outros 6 modelos na tentativa de melhorar os resultados alcançados pelo modelo inicial. As alterações consistem em ajustes de parâmetros, inclusão de novas camadas, balanceamento do dataset e embaralhamento de amostras. Para enfatizar as alterações aplicadas nos modelos seguintes, foi elaborado uma tabela de configurações, demonstrando as condições de execução do modelo. As avaliações dos modelos serão apresentadas nas subseções seguintes.

### 5.3.1. Modelo A

O modelo "A" foi configurado conforme o padrão descrito anteriormente. Foi o modelo base para as melhorias dos modelos seguintes. As configurações iniciais estão descritas na tabela 8. Após configurado, foi executado o treinamento do modelo "A".

**Tabela 8** – Configuração do modelo "A"

<b>Configuração</b>	<b>Valor</b>
<i>Bach size</i>	padrão
Embaralhamento dos dados de teste	não
Taxa de aprendizagem	0.0001
Balanceado	não
Embaralhamento na separação entre dados de treinamento/validação	não
<i>Dataset</i>	DS1
<i>Early stop</i>	20
Camadas de dropout	0
<i>Steps per epoch</i>	100

Fonte: Dados da Pesquisa (2021).

A execução do modelo "A" foi a avaliação inicial das configurações. Os ajustes para o modelo seguinte foram feitos com base nos resultados alcançados através das configurações atuais. Para isto, foram avaliadas as métricas "acurácia" e "perda", geradas pelo treinamento do modelo. Ao fim do treinamento, o modelo foi

avaliado através de uma matriz de confusão e um relatório de classificação, através do qual é possível analisar as métricas "precisão", "recall" e "f1-score", além de demonstrar o "nº de ocorrências".

A tabela 9 apresenta os valores alcançados pelas métricas "acurácia" e "perda", decorrente do treinamento do modelo. É possível analisar alta acurácia, mas também alta perda.

**Tabela 9 – Acurácia do modelo A.**

<b>Acurácia</b>	<b>Perda</b>
0.8954423666000366	1.0906866788864136

Fonte: Dados da pesquisa (2021).

A matriz de confusão foi gerada por meio da predição de uma série de imagens. Estas imagens foram passadas como parâmetro de entrada para o modelo na expectativa da predição da classe. A tabela 10 demonstra a predição alcançada pelo modelo "A" em cada classe treinada. É possível analisar a tendência de predição à classe "HAS", devido a quantidade superior de amostras em relação às outras classes.

**Tabela 10 – Matriz de confusão do modelo A.**

	<b>DRGE</b>	<b>Dor Abdom.</b>	<b>HAS</b>	<b>Osteoporose</b>
<b>DRGE</b>	7	4	33	11
<b>Dor Abdominal</b>	8	9	31	4
<b>HAS</b>	17	39	112	30
<b>Osteoporose</b>	5	17	40	6

Fonte: Dados da pesquisa (2021).

Por fim, a tabela 11 exibe o relatório de classificação alcançado pelo modelo "A". É possível analisar baixa precisão na predição das classes, também é possível analisar baixa acurácia na etapa de predição quando comparado com a etapa de treinamento, esta diferença está ligada às diferentes amostras utilizadas entre as duas

etapas. Vale ressaltar que o embaralhamento das amostras no modelo "A" acontece somente ao testar a capacidade de predição, porém é possível que sejam utilizadas amostras diferentes da etapa de treinamento.

**Tabela 11** – Relatório de classificação do modelo A.

	<b>Precisão</b>	<b>Recall</b>	<b>F1-score</b>	<b>Nº ocorrênc.</b>
DRGE	0.19	0.13	0.15	55
Dor Abdominal	0.13	0.17	0.15	52
HAS	0.52	0.57	0.54	198
Osteoporose	0.12	0.09	0.10	68
Acurácia	0.24	0.24	0.36	373
Macro avg			0.24	373
Weighted avg	0.34	0.36	0.35	373

Fonte: Dados da Pesquisa (2021).

A execução do modelo "A" levou à validação de novas configurações, na tentativa de melhorar a acurácia, a precisão e diminuir a perda. Um novo teste foi avaliado no modelo seguinte.

### 5.3.2. Modelo B

Com base nos resultados do modelo "A", o modelo "B" buscou melhorar os indicadores "acurácia" e "perda". Sendo assim, o parâmetro "*batch size*" foi modificado, além da configuração do parâmetro "*shuffle*" no gerador "*TestDataGen*". A nova configuração pode ser vista na tabela 12.

A tabela 13 exibe os valores alcançados pelas métricas "acurácia" e "perda". É possível analisar melhora na acurácia e na perda quando comparado ao modelo "A", mas a alta perda se manteve.

Para a avaliação da capacidade de predição do modelo "B" foi gerada novamente a matriz de confusão. Quando comparada com os resultados alcançados pelo modelo "A", é visto melhora nos resultados alcançados na tabela 14. Novamente é possível analisar a tendência de predição à classe "HAS", devido à quantidade superior de amostras em relação às outras classes.

**Tabela 12 – Configuração do modelo "B"**

<b>Configuração</b>	<b>Valor</b>
<i>Bach size</i>	9
Embaralhamento dos dados de teste	sim
Taxa de aprendizagem	0.0001
Balanceado	não
Embaralhamento na separação entre dados de treinamento/validação	não
<i>Dataset</i>	DS1
<i>Early stop</i>	20
Camadas de <i>dropout</i>	0
<i>Steps per epoch</i>	100

Fonte: Dados da Pesquisa (2021).

**Tabela 13 – Acurácia do modelo B.**

<b>Acurácia</b>	<b>Perda</b>
0.919571042060852	1.0228327512741089

Fonte: Dados da pesquisa (2021).

**Tabela 14 – Matriz de confusão do modelo B.**

	<b>DRGE</b>	<b>Dor Abdom.</b>	<b>HAS</b>	<b>Osteoporose</b>
DRGE	15	6	26	8
Dor Abdominal	4	12	31	5
HAS	23	35	109	31
Osteoporose	7	13	38	10

Fonte: Dados da pesquisa (2021).

Por fim, a tabela 15 exibe o relatório de classificação alcançado pelo modelo "B". É possível analisar aumento na precisão na predição das classes, também é possível analisar aumento na acurácia na etapa de predição quando comparado ao modelo "A" .

**Tabela 15** – Relatório de classificação do modelo B.

	Precisão	Recall	F1-score	Nº ocorrênc.
DRGE	0.31	0.27	0.29	55
Dor Abdominal	0.18	0.23	0.20	52
HAS	0.53	0.55	0.54	198
Osteoporose	0.19	0.15	0.16	68
Acurácia	0.30	0.30	0.39	373
Macro avg			0.30	373
Weighted avg	0.39	0.39	0.39	373

Fonte: Dados da pesquisa (2021).

Os resultados alcançados pelo modelo "B" demonstram a necessidade de melhora no indicador "perda". Também são necessários ajustes na tentativa de melhorar o indicador "precisão". Novas configurações foram aplicadas, e serão avaliadas no modelo seguinte.

### 5.3.3. Modelo C

O modelo "C" teve o parâmetro "*learning rate*" ajustado, na tentativa de diminuir a métrica "perda", conforme tabela 16. A expectativa é que, aumentando a taxa de aprendizagem, o modelo tenderia a diminuir a métrica "perda".

**Tabela 16 – Configuração do modelo "C"**

<b>Configuração</b>	<b>Valor</b>
Bach size	9
Embaralhamento dos dados de teste	sim
Taxa de aprendizagem	0.001
Balanceado	não
Embaralhamento na separação entre dados de treinamento/validação	não
Dataset	DS1
Early stop	20
Camadas de dropout	0
teps per epoch	100

Fonte: Dados da pesquisa (2021).

A expectativa foi confirmada, havendo significativa melhora na métrica "perda", mas como consequência a acurácia foi comprometida, como se vê na tabela 17.

**Tabela 17 – Acurácia do modelo C.**

<b>Acurácia</b>	<b>Perda</b>
0.8766756057739258	0.27917030453681946

Fonte: Dados da pesquisa (2021).

O modelo "B" se mostrou menos capaz na tarefa de predição, quando comparado aos modelos "A" e "B" (tabela 18). Novamente é possível analisar a tendência de predição à classe "HAS", devido à quantidade superior de amostras em relação às outras classes.

**Tabela 18** – Matriz de confusão do modelo C.

	<b>DRGE</b>	<b>Dor Abdom.</b>	<b>HAS</b>	<b>Osteoporose</b>
<b>DRGE</b>	5	6	37	7
<b>Dor Abdominal</b>	3	13	28	8
<b>HAS</b>	30	27	104	37
<b>Osteoporose</b>	8	5	38	17

Fonte: Dados da pesquisa (2021).

Por fim, a tabela 19 exibe o relatório de classificação alcançado pelo modelo "C", confirmando queda nas métricas de avaliação quando comparado ao modelo "B".

**Tabela 19** – Relatório de classificação do modelo C.

	<b>Precisão</b>	<b>Recall</b>	<b>F1-score</b>	<b>Nº ocorrênc.</b>
<b>DRGE</b>	0.11	0.09	0.10	55
<b>Dor Abdominal</b>	0.25	0.25	0.25	52
<b>HAS</b>	0.50	0.53	0.51	198
<b>Osteoporose</b>	0.25	0.25	0.25	68
<b>Acurácia</b>	0.28	0.28	0.37	373
<b>Macro avg</b>			0.28	373
<b>Weighted avg</b>	0.36	0.37	0.37	373

Fonte: Dados da pesquisa (2021).

Mesmo conseguindo melhorar a métrica "perda", não houve melhora na capacidade de predição do modelo "C" quando comparado ao modelo "B". Além do mais, a métrica "acurácia" foi comprometida. Foram estabelecidas novas configurações na tentativa de melhorar as métricas de avaliação, e serão avaliadas no modelo seguinte.

#### 5.3.4. Modelo D

Os resultados anteriores demonstram a possibilidade de melhora nas métricas através da adição de novas imagens ao *dataset*. Esta conclusão foi alcançada ao avaliar as métricas da classe "HAS", que obteve as melhores métricas de precisão nos três modelos anteriores. Este questionamento foi levantado uma vez que esta é a classe com maior quantidade de amostras. Sendo assim, o modelo "D" foi configurado para utilizar o dataset DS2 20.

**Tabela 20** – Configuração do modelo "D"

<b>Configuração</b>	<b>Valor</b>
<i>Bach size</i>	9
Embaralhamento dos dados de teste	sim
Taxa de aprendizagem	0.0001
Balanceado	não
Embaralhamento na separação entre dados de treinamento/validação	não
<i>Dataset</i>	DS2
<i>Early stop</i>	20
Camadas de <i>dropout</i>	0
<i>Steps per epoch</i>	100

Fonte: Dados da pesquisa (2021).

Aumentar a quantidade de amostras gerou melhora nas métricas "acurácia" e "perda", conforme se vê na tabela 21, gerando os melhores resultados para estas métricas dentre os 4 modelos já avaliados.

Apesar da melhora nas métricas "acurácia" e "perda", o modelo "D" não apresentou melhora na capacidade de predição quando comparado ao modelo "B" (tabela 14).

**Tabela 21** – Acurácia do modelo D.

<b>Acurácia</b>	<b>Perda</b>
0.932692289352417	0.30696767568588257

Fonte: Dados da pesquisa (2021).

**Tabela 22** – Matriz de confusão do modelo D.

	<b>DRGE</b>	<b>Dor Abdom.</b>	<b>HAS</b>	<b>Osteoporose</b>
<b>DRGE</b>	16	24	45	8
<b>Dor Abdominal</b>	12	14	46	15
<b>HAS</b>	57	58	166	49
<b>Osteoporose</b>	22	17	59	16

Fonte: Dados da pesquisa (2021).

É possível analisar melhora na métrica "precisão" quando comparado ao modelo "A", mas no geral é notado piora quando comparado aos modelos "B" e "C" 23.

**Tabela 23** – Relatório de classificação do modelo D.

	<b>Precisão</b>	<b>Recall</b>	<b>F1-score</b>	<b>Nº ocorrênc.</b>
<b>DRGE</b>	0.15	0.17	0.16	93
<b>Dor Abdominal</b>	0.12	0.16	0.14	87
<b>HAS</b>	0.53	0.50	0.51	330
<b>Osteoporose</b>	0.18	0.14	0.16	114
<b>Acurácia</b>	0.25	0.24	0.34	624
<b>Macro avg</b>			0.24	624
<b>Weighted avg</b>	0.35	0.34	0.34	624

Fonte: Dados da pesquisa (2021).

A análise realizada no modelo "D" leva à busca por melhoria nas métricas "acurácia" e "perda" na expectativa de melhorar a capacidade de predição do modelo. Sendo assim, novas configurações foram aplicadas, e serão avaliadas no modelo

seguinte.

#### 5.3.5. Modelo E

A expectativa para o modelo "E" é alcançar melhores resultados para as métricas "acurácia" e "perda", buscando também melhorar os resultados da métrica "precisão". Com base nos resultados dos modelos anteriores, foram aplicadas novas configurações. O parâmetro "*Early stop*" foi reduzido e também foram adicionadas duas novas camadas de "*Dropout*", conforme pode ser visto na figura 30. Além do mais, as amostras foram embaralhadas antes de serem separadas para treinamento e validação. As novas configurações podem ser vistas na tabela 24.

O modelo "E", tabela 25, alcançou os melhores resultados para as métricas "acurácia" e "perda" até então. É possível analisar o efeito positivo proveniente das novas configurações.

**Figura 30** – Sumário do segundo modelo criado contendo camadas de "Dropout".

```

Model: "sequential_1"
Layer (type)                Output Shape                Param #
=====
conv2d_13 (Conv2D)          (None, 224, 224, 64)      1792
conv2d_14 (Conv2D)          (None, 224, 224, 64)      36928
max_pooling2d_5 (MaxPooling2 (None, 112, 112, 64)      0
conv2d_15 (Conv2D)          (None, 112, 112, 128)     73856
conv2d_16 (Conv2D)          (None, 112, 112, 128)     147584
max_pooling2d_6 (MaxPooling2 (None, 56, 56, 128)      0
conv2d_17 (Conv2D)          (None, 56, 56, 256)       295168
conv2d_18 (Conv2D)          (None, 56, 56, 256)       590080
conv2d_19 (Conv2D)          (None, 56, 56, 256)       590080
max_pooling2d_7 (MaxPooling2 (None, 28, 28, 256)      0
conv2d_20 (Conv2D)          (None, 28, 28, 512)       1180160
conv2d_21 (Conv2D)          (None, 28, 28, 512)       2359808
conv2d_22 (Conv2D)          (None, 28, 28, 512)       2359808
max_pooling2d_8 (MaxPooling2 (None, 14, 14, 512)      0
conv2d_23 (Conv2D)          (None, 14, 14, 512)       2359808
conv2d_24 (Conv2D)          (None, 14, 14, 512)       2359808
conv2d_25 (Conv2D)          (None, 14, 14, 512)       2359808
max_pooling2d_9 (MaxPooling2 (None, 7, 7, 512)        0
flatten_1 (Flatten)         (None, 25088)              0
dense_3 (Dense)             (None, 4096)               102764544
dropout_2 (Dropout)         (None, 4096)               0
dense_4 (Dense)             (None, 4096)               16781312
dropout_3 (Dropout)         (None, 4096)               0
dense_5 (Dense)             (None, 4)                   16388
=====
Total params: 134,276,932
Trainable params: 134,276,932
Non-trainable params: 0

```

Fonte: Dados da pesquisa

**Tabela 24** – Configuração do modelo E

<b>Configuração</b>	<b>Valor</b>
<i>Bach size</i>	9
Embaralhamento dos dados de teste	sim
Taxa de aprendizagem	0.0001
Balanceado	não
Embaralhamento na separação entre dados de treinamento/validação	sim
<i>Dataset</i>	DS2
<i>Early stop</i>	5
Camadas de <i>dropout</i>	2
<i>Steps per epoch</i>	100

Fonte: Dados da pesquisa (2021).

**Tabela 25** – Acurácia do modelo E.

<b>Acurácia</b>	<b>Perda</b>
0.9791666865348816	0.06536780297756195

Fonte: Dados da pesquisa (2021).

Apesar da melhora alcançada na etapa de treinamento do modelo, a capacidade de predição do modelo não melhorou como esperado (tabela 26).

**Tabela 26** – Matriz de confusão do modelo E.

	<b>DRGE</b>	<b>Dor Abdom.</b>	<b>HAS</b>	<b>Osteoporose</b>
DRGE	17	15	47	14
Dor Abdominal	11	16	46	14
HAS	47	46	178	59
Osteoporose	19	10	58	27

Fonte: Dados da pesquisa (2021).

É possível analisar melhora na métrica "precisão"<sup>19</sup>. Ainda sim é notório a diferença entre a classe "HAS" e as demais.

**Tabela 27** – Relatório de classificação do modelo E.

	<b>Precisão</b>	<b>Recall</b>	<b>F1-score</b>	<b>Nº ocorrênc.</b>
DRGE	0.18	0.18	0.18	93
Dor Abdominal	0.18	0.18	0.18	87
HAS	0.54	0.54	0.54	330
Osteoporose	0.24	0.24	0.24	114
Acurácia	0.29	0.29	0.38	624
<i>Macro avg</i>			0.29	624
<i>Weighted avg</i>	0.38	0.38	0.38	624

Fonte: Dados da pesquisa (2021).

O desequilíbrio entre a ocorrência das amostras pode estar interferindo nos resultados. Com isso, são realizados novos ajustes na tentativa de balancear o dataset na expectativa de alcançar melhores resultados. As novas configurações foram aplicadas, e serão avaliadas no modelo seguinte.

#### 5.3.6. Modelo F

Os resultados anteriores demonstram que a variação na ocorrência das amostras pode estar afetando as métricas de avaliação. Novamente este questionamento se deu por conta da variação nos resultados da classe "HAS" quando comparada às outras. Sendo assim, o modelo "F" foi configurado para utilizar o dataset DS3 (tabela 28).

**Tabela 28** – Configuração do modelo F

<b>Configuração</b>	<b>Valor</b>
Bach size	9
Embaralhamento dos dados de teste	sim
Taxa de aprendizagem	0.0001
Balanceado	sim
Embaralhamento na separação entre dados de treinamento/validação	sim
Dataset	DS3
Early stop	5
Camadas de dropout	2
Steps per epoch	100

Fonte: Dados da pesquisa (2021).

O modelo "F" (tabela 29) conseguiu resultados ainda melhores para as métricas "acurácia" e "perda", quando comparado com o modelo "E". Com isso, é possível avaliar que o balanceamento do dataset interferiu nas métricas "acurácia" e "perda".

**Tabela 29** – Acurácia do modelo F.

<b>Acurácia</b>	<b>Perda</b>
0.9885057210922241	0.025087924674153328

Fonte: Dados da pesquisa (2021).

Apesar da melhora alcançada na etapa de treinamento do modelo, a capacidade de predição do modelo não melhorou, conforme se vê na tabela 30.

**Tabela 30** – Matriz de confusão do modelo F.

	<b>DRGE</b>	<b>Dor Abdom.</b>	<b>HAS</b>	<b>Osteoporose</b>
<b>DRGE</b>	22	25	21	19
<b>Dor Abdominal</b>	26	19	22	20
<b>HAS</b>	24	20	26	17
<b>Osteoporose</b>	18	21	16	32

Fonte: Dados da pesquisa (2021).

É possível analisar piora na métrica "precisão" quando comparado ao modelo "E".

**Tabela 31** – Relatório de classificação do modelo F.

	<b>Precisão</b>	<b>Recall</b>	<b>F1-score</b>	<b>Nº ocorrênc.</b>
<b>DRGE</b>	0.24	0.25	0.25	87
<b>Dor Abdominal</b>	0.22	0.22	0.22	87
<b>HAS</b>	0.31	0.30	0.30	87
<b>Osteoporose</b>	0.36	0.37	0.37	87
<b>Acurácia</b>	0.28	0.28	0.28	348
<b>Macro avg</b>			0.28	348
<b>Weighted avg</b>	0.28	0.28	0.28	348

Fonte: Dados da pesquisa (2021).

### 5.3.7. Modelo G

Os resultados dos modelos anteriores contribuíram para a aplicação do "fine tuning" ao modelo final. A observação no comportamento do modelo ao ajustar determinados parâmetros resultou na construção de um modelo promissor. Com isso, neste último modelo, foram aplicadas alterações nos parâmetros "*batch\_size*", "*learning\_rate*", "*early\_stop*" e "*steps\_per\_epoch*". Além disso, o parâmetro "*shuffle*" do gerador "TestDataGen" foi alterado. Estes ajustes foram feitos considerando a

utilização do dataset DS3. Em seguida foi apresentada a tabela 32 com as configurações utilizadas neste modelo.

**Tabela 32 – Configuração do modelo "G"**

<b>Configuração</b>	<b>Valor</b>
Bach size	32
Embaralhamento dos dados de teste	não
Taxa de aprendizagem	0.000001
Balanceado	sim
Embaralhamento na separação entre dados de treinamento/validação	sim
Dataset	DS3
Early stop	20
Camadas de dropout	2
Steps per epoch	70

Fonte: Dados da pesquisa (2021).

O modelo "G" alcançou as melhorias pretendidas nos modelos anteriores. As métricas "acurácia" e "perda" obtiveram melhor resultado quando comparadas aos modelos anteriores, conforme se vê na tabela 33, abaixo.

**Tabela 33 – Acurácia do modelo G.**

<b>Acurácia</b>	<b>Perda</b>
0.994252860546112110	0.05301027372479439

Fonte: Dados da pesquisa (2021).

É possível notar melhora na capacidade de predição do modelo ao observar os resultados da matriz de confusão (tabela 34). Ao compará-la com os modelos anteriores, nenhum obteve resultado semelhante.

**Tabela 34** – Matriz de confusão do modelo G.

	<b>DRGE</b>	<b>Dor Abdom.</b>	<b>HAS</b>	<b>Osteoporose</b>
<b>DRGE</b>	87	0	0	0
<b>Dor Abdominal</b>	0	87	0	0
<b>HAS</b>	1	0	86	0
<b>Osteoporose</b>	0	1	0	86

Fonte: Dados da pesquisa (2021).

Concluindo, a tabela 35 exibe o relatório de classificação alcançado pelo modelo "G". É possível analisar a evolução das métricas ao compará-las aos resultados anteriores. Neste modelo, é possível comparar a métrica "acurácia" do resultado do treinamento com a métrica "acurácia" do resultado do teste de predição.

**Tabela 35** – Relatório de classificação do modelo G.

	<b>Precisão</b>	<b>Recall</b>	<b>F1-score</b>	<b>Nº ocorrênc.</b>
<b>DRGE</b>	0.99	1.00	0.99	87
<b>Dor Abdominal</b>	0.99	1.00	0.99	87
<b>HAS</b>	1.00	0.99	0.99	87
<b>Osteoporose</b>	1.00	0.99	0.99	87
<b>Acurácia</b>	0.99	0.99	0.99	348
<b>Macro avg</b>			0.99	348
<b>Weighted avg</b>	0.99	0.99	0.99	348

Fonte: Dados da pesquisa (2021).

Vale ressaltar que os modelos anteriores serviram como base para as configurações do modelo "G", podendo este ser considerado uma evolução dos modelos "A", "B", "C", "D", "E" e "F".

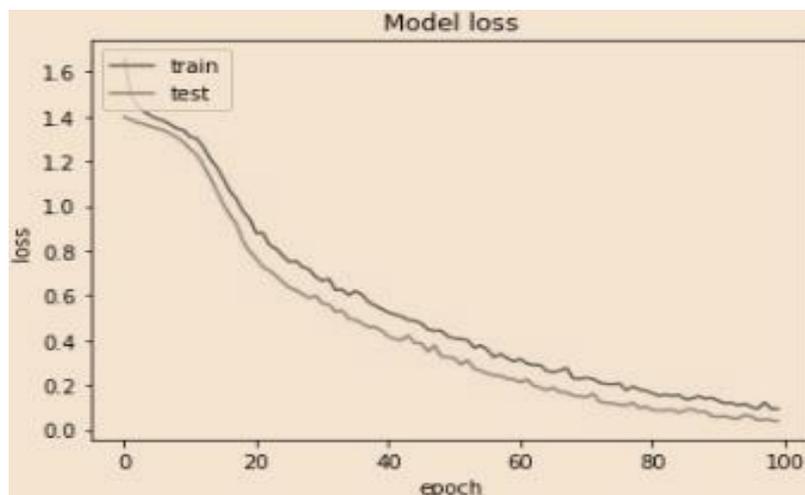
#### **5.4. Resultado do Modelo**

A evolução realizada no modelo padrão, inicialmente proposto, resultou em alta "acurácia" e "precisão". Com o objetivo de validar as configurações propostas pelo

modelo "G" um novo modelo "Modelo V" foi treinado e avaliado com base os mesmos parâmetros, a fim de avaliar certa "previsibilidade"no algoritmo proposto. Os resultados podem ser verificados a seguir.

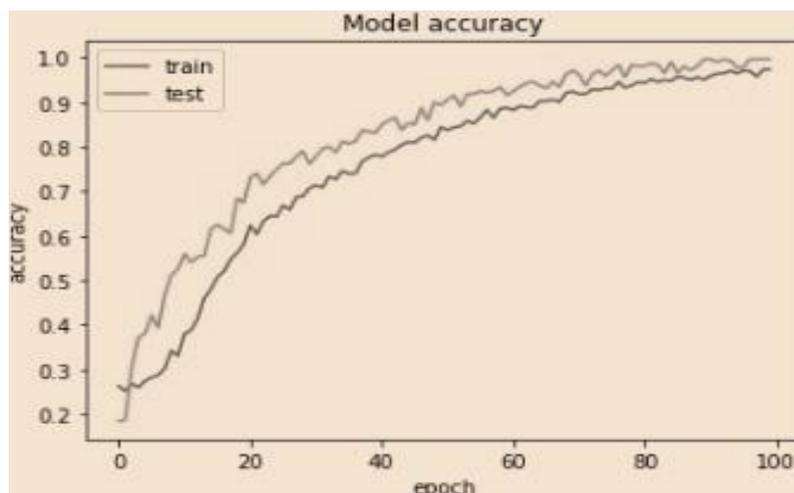
As métricas "acurácia" e "perda"alcançadas pelo modelo "V" foram extremamente semelhantes ao modelo "G". É possível notar uma pequena variação nas métricas "acurácia"e "perda"quando comparadas ao modelo "G". Também é possível analisar o gráfico de evolução das métricas "acurácia" e "perda" nas figuras 31 e 32

**Figura 31** – Gráfico de acurácia do modelo V.



Fonte: Dados da pesquisa (2021).

**Figura 32** – Gráfico de perda do modelo V.



Fonte: Dados da pesquisa

**Tabela 36** – Acurácia do modelo V.

<b>Acurácia</b>	<b>Perda</b>
0.9971264600753784	0.05649818107485771

Fonte: Dados da pesquisa (2021).

A capacidade de predição do modelo "V" também foi semelhante ao modelo "G", alcançando alta capacidade de predição 37.

**Tabela 37** – Matriz de confusão do modelo V.

	<b>DRGE</b>	<b>Dor Abdom.</b>	<b>HAS</b>	<b>Osteoporose</b>
<b>DRGE</b>	87	0	0	0
<b>Dor Abdominal</b>	0	87	0	0
<b>HAS</b>	1	0	86	0
<b>Osteoporose</b>	0	0	0	87

Fonte: Dados da pesquisa (2021).

Finalizando, a tabela 38 exibe o relatório de classificação alcançado pelo modelo "V". Também é possível verificar resultados semelhantes quando comparado ao modelo "G".

**Tabela 38** – Relatório de classificação do modelo V.

	<b>Precisão</b>	<b>Recall</b>	<b>F1-score</b>	<b>Nº ocorrências</b>
<b>DRGE</b>	0.99	1.00	0.99	87
<b>Dor Abdominal</b>	1.00	1.00	1.00	87
<b>HAS</b>	1.00	0.99	0.99	87
<b>Osteoporose</b>	1.00	1.00	1.00	87
<b>Acurácia</b>	1.00	1.00	1.00	348
<b>Macro avg</b>			1.00	348
<b>Weighted avg</b>	1.00	1.00	1.00	348

Fonte: Dados da pesquisa (2021).

Os resultados obtidos no modelo "V" confirmam os resultados alcançados

pelo modelo "G" ao conseguir manter certa "previsibilidade" no algoritmo. Também é possível validar a pertinência da arquitetura VGG-16 na aplicação do HTR.

É notória a eficácia do modelo proposto na extração de patologias em guias médicas, do tipo SP/SADT, escritas à mão por médicos.

Os resultados apresentados neste trabalho podem ser facilmente replicados. Os *datasets* <sup>1</sup> e o algoritmo <sup>2</sup> serão disponibilizados na internet. O acesso será concedido mediante solicitação.

---

<sup>1</sup> Datasets: [https://drive.google.com/drive/folders/1NyA\\_lsgz0Fs4LcVm806xackeel8dOyq4?usp=sharing](https://drive.google.com/drive/folders/1NyA_lsgz0Fs4LcVm806xackeel8dOyq4?usp=sharing)

<sup>2</sup> Algoritmo: [https://colab.research.google.com/drive/13nOC5a8I3mABe1\\_r2-WHyvVvx4lwGjCn?usp=sharing](https://colab.research.google.com/drive/13nOC5a8I3mABe1_r2-WHyvVvx4lwGjCn?usp=sharing)

## 6. CONSIDERAÇÕES FINAIS

Esta dissertação teve como objetivo apresentar a acurácia de um modelo de HTR no reconhecimento de patologias escritas a mão em guias médicas. No desenvolvimento da pesquisa foi realizado um experimento para avaliar a acurácia do modelo proposto. Para a realização do HTR foi proposta uma CNN utilizando a arquitetura VGG-16 por ser uma arquitetura comumente utilizada em problemas de classificação e detecção. Foi proposto responder a seguinte pergunta neste trabalho: **Qual a acurácia das técnicas de HTR no reconhecimento de patologias descritas na indicação clínica em guias de serviços médicos escritas à mão?**

Conforme se verifica na literatura, outras pesquisas utilizaram técnicas semelhantes para resolver problemas relacionados ao reconhecimento de palavras escritas à mão por médicos.

Os resultados apresentados nesta pesquisa corroboram com os dados apresentados por pesquisas semelhantes realizadas em outros países. Para as operadoras de planos de saúde, os resultados vão de acordo com o esperado, sendo viável avaliar a aplicação de tal técnica no auxílio da análise de guias médicas.

Nota-se também a necessidade das operadoras de planos de saúde de organizar o processo de coleta e arquivamento das imagens de guias médicas. Apesar da grande quantidade de guias disponíveis, não há nenhum tipo de classificação ou ordenação na coleção de dados.

Por fim, o retrabalho realizado no dataset para aumentar a amostragem foi essencial para alcançar tais resultados.

### 6.1. Contribuições da Pesquisa

Como contribuição, este trabalho se destaca em dois pontos. O primeiro é na averiguação da capacidade de utilização de um sistema de reconhecimento de escritas à mão para reconhecer patologias em guias médicas, o que poderia auxiliar no processo de auditoria de guias médicas, atualmente realizada por profissionais da saúde, como médicos e enfermeiros.

Outro ponto de relevância deste trabalho foi a utilização de conceitos e modelos de práticas de reconhecimento de escritas à mão por médicos utilizados em outros países. Foram encontrados na literatura casos semelhantes ao proposto por

este trabalho. No caso desta pesquisa, o modelo foi avaliado e ajustado para que o seu resultado seja mais condizente com as características encontradas nas OPS do nosso país.

## **6.2. Limitações da Pesquisa**

A principal limitação da pesquisa foi a reduzida variedade de patologias nas guias entregues pelas OPS e utilizadas no experimento, resultando num total de apenas 4 classes, sendo elas: "HAS", "Dor abdominal", "DRGE" e "Osteoporose".

Esta pesquisa também limitou-se ao reconhecimento de patologias, não classificando-as com o C.I.D.

## **6.3. Trabalhos Futuros**

A principal sugestão para uma pesquisa futura é a criação de modelo capaz de reconhecer e classificar, de acordo com o C.I.D., uma patologia contida na descrição clínica de uma guia médica do tipo SP/SADT escrita à mão por médicos.

## REFERÊNCIAS

- Alom, M. Z., Sidike, P., Hasan, M., Taha, T. M., Asari, V. K., and Schwenker, F. (2018). Handwritten bangla character recognition using the state-of-the-art deep convolutional neural networks. *Intell. Neuroscience*, 2018.
- Barr, A. and Feigenbaum, E. A. (1981). *The Handbook of Artificial Intelligence*, volume 1.
- Cheng, J., Dong, L., and Lapata, M. (2016). Long short-term memory-networks for machine reading. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 551–561, Austin, Texas. Association for Computational Linguistics.
- Das, S., Dey, A., Pal, A., and Roy, N. (2015). Article: Applications of artificial intelligence in *Machine Learning: Review and prospect. International Journal of Computer Applications*, 115(9):31–41. Full text available.
- Deore, S. P. and Pravin, A. (2020). Devanagari handwritten character recognition using fine-tuned deep convolutional neural network on trivial dataset. *Sadhana*, 45(1):243.
- DSA, D. S. A. (2019). *Deep Learning Book*.
- E.Kamalanaban, D., Gopinath, M., and Premkumar, S. (2018). Medicine box: Doctor's prescription recognition using deep *Machine Learning. International Journal of Engineering and Technology*, 7(3.34):114–117.
- Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36:193–202.
- Gers, F., Schmidhuber, J., and Cummins, F. (2000). Learning to forget: Continual prediction with lstm. *Neural computation*, 12:2451–71.
- Goldberg, Y. (2017). Neural network methods for natural language processing. *Synthesis Lectures on Human Language Technologies*, 10:1–309.
- Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. Adaptive Computation and *Machine Learning* series. MIT Press.
- Greff, K., Srivastava, R., Koutník, J., Steunebrink, B., and Schmidhuber, J. (2015). Lstm: A search space odyssey. *IEEE transactions on neural networks and learning systems*, 28.

- Harjoseputro, Y., dri handarkha, Y., and Tresy Renata Adie, H. (2019). The javanese letters classifier with mobile client-server architecture and convolution neural network method. *International Journal of Interactive Mobile Technologies (iJIM)*, 13:67.
- Ingle, R. R., Fujii, Y., Deselaers, T., Baccash, J., and Popat, A. C. (2019). A scalable handwritten text recognition system. In *2019 International Conference on Document Analysis and Recognition (ICDAR)*, pages 17–24.
- Josh Patterson, A. G. (2017). *Deep Learning*. O'Reilly Media, Inc.
- Kitchenham, B., Pretorius, R., Budgen, D., Brereton, P., Turner, M., Niazi, M., and Linkman, S. (2010). Systematic literature reviews in software engineering – a tertiary study. *Information and Software Technology*, 52:792–805.
- Kumar, S. (2016). Offline handwriting character recognition (for use of medical purpose) using neural network. *International Journal Of Engineering And Computer Science*.
- Lana, J., Partyka, R. B., Alberton, A., and Marcon, R. (2018). Caso para ensino: o processo de escolhas metodológicas em uma abordagem quantitativa. *Revista de Contabilidade e Organizações*, 12:e148286.
- Liu, P., Qiu, X., and Huang, X. (2016). Recurrent neural network for text classification with multi-task learning.
- Manik Nijhawan, P. S. (2019). Handwriting detection using neural network. Technical report, Department of Computer Science and Engineering and Information Technology, Jaypee University of Information Technology Waknaghat.
- Memon, J., Sami, M., and Khan, R. (2019). Handwritten optical character recognition (ocr): A comprehensive systematic literature review (slr).
- Ming Liang and Xiaolin Hu (2015). Recurrent convolutional neural network for object recognition. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3367–3375.
- Nash, W., Drummond, T., and Birbilis, N. (2018). A review of *Deep Learning* in the study of materials degradation. *npj Materials Degradation*, 2.
- Nilsson, N. (1982). *Principles of Artificial Intelligence*, volume (Nilsson, 1997) Nilsson, N. (1997). *Artificial Intelligence: A New Synthesis*, volume
- Rao, P. S. and Aditya, J. (2014). Handwriting recognition – “ offline ” approach.

- Ravi, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., Lo, B., and Yang, G. (2017). *Deep Learning* for health informatics. *IEEE Journal of Biomedical and Health Informatics*, 21(1):4–21.
- Setti, S. and Wanto, A. (2019). Analysis of backpropagation algorithm in predicting the most number of internet users in the world. *Jurnal Online Informatika*, 3:110.
- Shannon, C. E. and McCarthy, J. (1956). *Automata Studies. (AM-34) (Annals of Mathematics Studies)*. Princeton University Press, USA.
- Sherstinsky, A. (2018). Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network.
- Simard, P. Y., Steinkraus, D., and Platt, J. C. (2003). Best practices for convolutional neural networks applied to visual document analysis. In *Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings.*, pages 958–963.
- Sánchez, J. A., Romero, V., Toselli, A., Villegas, M., and Vidal, E. (2019). A set of benchmarks for handwritten text recognition on historical documents. *Pattern Recognition*, 94.
- Trevisol Neto, O. (2018). *Métodos e Técnicas de Pesquisa*.