



**Use of Computer Vision for Claims Adjustment Automation:  
A Case Study with a Large Health Insurance Provider**



# Introduction

## Health insurance provider

- **Beneficiary**
  - **Signs a contract with the provider**
- **Physician**
  - **Is affiliated with the provider**
- **Provider**
  - **Must comply with the agreed contract terms**
  - **Should not pay for non-covered services**





# Objectives

## ▣ **General objective**

Develop an application to automate the medical claims auditing process for a health insurance provider. The application must be capable of extracting information from scanned medical forms and comparing it with data recorded in the health insurance provider's system. It is expected that services performed by contracted hospitals that were not prescribed by physicians in the scanned forms will be flagged and denied.

# Objetives

## □ **Specific objectives**

- Acquire a dataset of medical form images;
- Label the medical forms dataset with support from a specialist;
- Compare the main OCR tools;
- Train a neural network to detect components within the forms;
- Extract content from the medical forms;
- Compare the extracted information with the data stored in the health insurance provider's system.



# Theoretical framework

## OCR (Optical Character Recognition)

- Computer vision
- Image processing
- Deep learning
- CNN

### Sistema

Detalhe da Nota

Nº NOTA:  
INDICAÇÃO CLÍNICA:  
DATA ATENDIMENTO:  
PRESTADOR:  
BENEFICIÁRIO:  
REQUISITANTE:

TOSSE 4  
21/05/2022 08:22:53 3  
HOSP GERAL DO VALE DO JAGUARIBE LTDA  
1  
2

OBSERVAÇÃO OPERADOR:  
NÃO CONTEM OBSERVAÇÃO

CÓDIGO	SERVIÇO	STATUS
4030436	HEMOGRAMA COM CONTAGEM DE PLAQUETAS 12	L
4030251	TRANSAMINASE PIRUVICA (AMINO TRANSFERASE DE ALANINA) 10	L
4030250	TRANSAMINASE OXALACETICA (AMINO TRANSFERASE ASPARTATO) 9	L
4030839	Proteina C reativa, quantitativa 13	L
4030110	ACIDO LACTICO 14	L
4030258	UREIA 11	L
4030163	CREATININA 8	L

Quantidade: 7  
Valor Total: R\$ 52,53

ELINAIVA C CHAVES.pdf

### Guia

GUIA DE SERVIÇO PROFISSIONAL / SERVIÇO AU  
DIAGNÓSTICO E TERAPIA - SP/SADT

1-Registro ANS	5-Numero da Guia Principal	10-Nome
4-Data de Autorização	6-Senha	11-Numero no CRM
8-Numero da Carteira	9-Validade da Carteira	14-Nome do Contratado
13-Código na Operadora / CNPJ / CPF	14-Nome do Contratado	17-Numero no CRM
16-Numero do Profissional Solicitante	19-Indicação Clínica	20-Indicação Clínica
21-Caracter de Atendimento	22-Data da Solicitação	23-Indicação Clínica
URGÊNCIA/EMERGÊNCIA	21/05/2022 - 08:22 3	TOSSE 4
24-Tabela	25-Código do Procedimento	26-Descrição
22	40301630	CREATININA - (LABORATORIO EXAMES URGENCIA E EMERGENCIA) 8
22	40302504	TGO-AST - (LABORATORIO EXAMES URGENCIA E EMERGENCIA) 9
22	40302512	TGP-ALT - (LABORATORIO EXAMES URGENCIA E EMERGENCIA) 10
22	40302580	UREIA - (LABORATORIO EXAMES URGENCIA E EMERGENCIA) 11
22	40304361	HEMOGRAMA COMPLETO - (LABORATORIO EXAMES URGENCIA E EMERGENCIA) 12
22	40308391	PCR-PROTEINA C REATIVA - (LABORATORIO EXAMES URGENCIA E EMERGENCIA) 13
22	40301109	ACIDO LACTICO (LACTATO) - (LABORATORIO EXAMES URGENCIA E EMERGENCIA) 14
29-Código na Operadora / CNPJ / CPF	30-Nome do Contratado	

# Referencial Teórico

## String similarity

- Based on char
- Based on vectorial space

## Uses

- Plagiarism Detection
- Text Comparison
- OCR Post-Processing

## Sistema

### Detalhe da Nota

Nº NOTA:  
INDICAÇÃO CLÍNICA:  
DATA ATENDIMENTO:  
PRESTADOR:  
BENEFICIÁRIO:  
REQUISITANTE:

4  
3  
1  
2

OBSERVAÇÃO OPERADOR:  
NÃO CONTEM OBSERVAÇÃO

COEIGO	SERVICO	STATUS
4030436	HEMOGRAMA COM CONTADEM DE PLAQUETAS	L
4030251	TRANSAMINASE PIRUVICA (AMINO TRANSFERASE DE ALANINA)	L
4030250	TRANSAMINASE CIGALACTICA (AMINO TRANSFERASE ASPARTATO)	L
4030839	Proteina C reativa, quantitativa	L
4030110	ACIDO LACTICO	L
4030258	UREIA	L
4030163	CREATININA	L

Quantidade: 7  
Valor Total: R\$ 52,53

EIINALVA C CHAVES.pdf

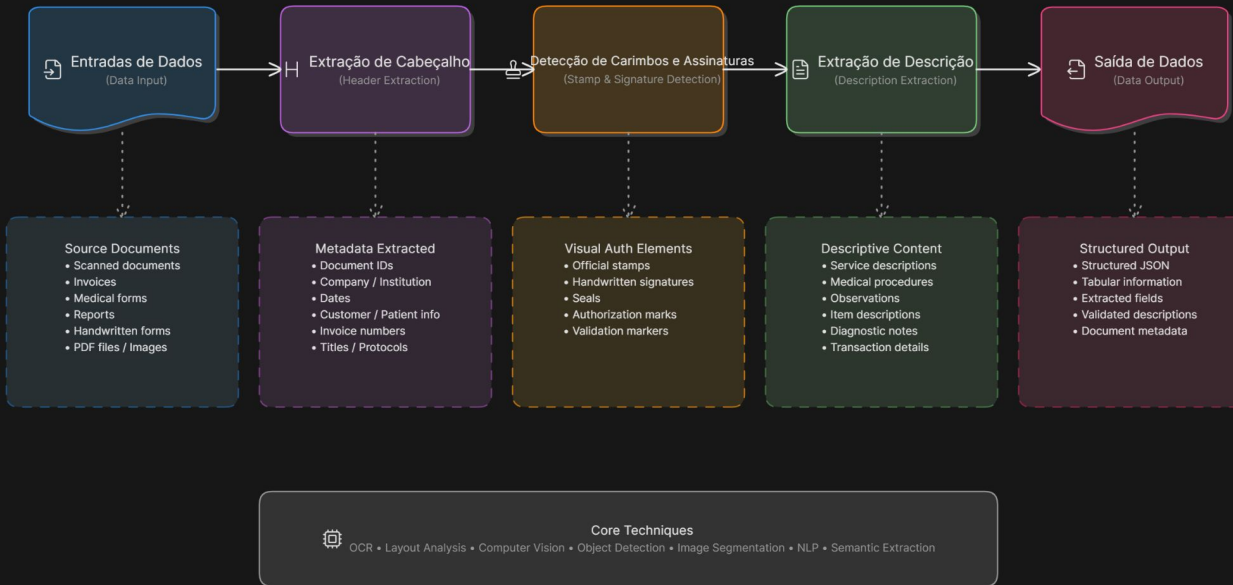
## Guia

GUIA DE SERVIÇO PROFISSIONAL / SERVIÇO AU  
DIAGNÓSTICO E TERAPIA - SP/SADT

1-Registro ANS	2-Numero da Guia Principal	3-Data de Autorização	4-Semina	5-Data de Validade da Semina
6-Dados do Beneficiário	7-Numero da Carteira	8-Validade da Carteira	9-Nome	10-Endereço
11-Dados do Solicitante	12-Código na Operadora / CNPJ / CPF	13-Nome do Contratado	14-HOSPITAL GERAL DO VALE DO JAGUARIBE LTDA	15-Consócio Profissional
16-Nome do Profissional Solicitante	17-Numero no CRM	18-Dados de Solicitação / Procedimentos e Exames	19-Data da Solicitação	20-Indicação Clínica
21-Código do Atendimento	22-URGÊNCIA/EMERGÊNCIA	23-21/05/2022 - 08:22	24-TOSSE	
25-Tabela	26-Código do Procedimento	27-Descrição	28-8	29-12
22	40301630	CREATININA - (LABORATORIO EXAMES URGENCIA E EMERGENCIA)	8	12
22	40302504	TGO-AST - (LABORATORIO EXAMES URGENCIA E EMERGENCIA)	9	13
22	40302512	TGP-ALT - (LABORATORIO EXAMES URGENCIA E EMERGENCIA)	10	14
22	40302580	UREIA - (LABORATORIO EXAMES URGENCIA E EMERGENCIA)	11	
22	40304361	HEMOGRAMA COMPLETO - (LABORATORIO EXAMES URGENCIA E EMERGENCIA)	12	
22	40308391	PCR-PROTEINA C REATIVA - (LABORATORIO EXAMES URGENCIA E EMERGENCIA)	13	
22	40301109	ACIDO LACTICO (LACTATO) - (LABORATORIO EXAMES URGENCIA E EMERGENCIA)	14	
29-Dados do Contrato Executante		30-Nome do Contratado		

# Methodology

## Document Processing Pipeline



**1** Nome: [Redacted]

**2** Nome do Profissional: [Redacted]

**3** Data da Solicitação: 21/05/2022 - 08:22

**4** Indicação Clínica: TOSSE

**6** [Redacted Signature]

22	20-Código do Procedimento	26-Descrição	27-Qtd. Solic.	28-Qtd. Aut.
	40301630	CREATININA - (LABORATORIO EXAMES URGENCIA E EMERGENCIA)	1	
	40302504	TGO-AST - (LABORATORIO EXAMES URGENCIA E EMERGENCIA)	1	
	40302512	TGP-ALT - (LABORATORIO EXAMES URGENCIA E EMERGENCIA)	1	
	40302580	UREIA - (LABORATORIO EXAMES URGENCIA E EMERGENCIA)	1	
	40304361	HEMOGRAMA COMPLETO - (LABORATORIO EXAMES URGENCIA E EMERGENCIA)	1	
	40308391	PCR-PROTEINA C REATIVA - (LABORATORIO EXAMES URGENCIA E EMERGENCIA)	1	
	40301109	ACIDO LACTICO (LACTATO) - (LABORATORIO EXAMES URGENCIA E EMERGENCIA)	1	

**17** [Redacted]



# Methodology

## Comparação de OCRs

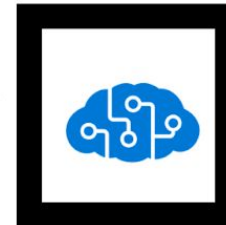
- Google Vision
- Azure Form Recognizer
- **Amazon Textract**
- Tesseract

## Métricas

- Pricing
- Response Quality
- Particularities

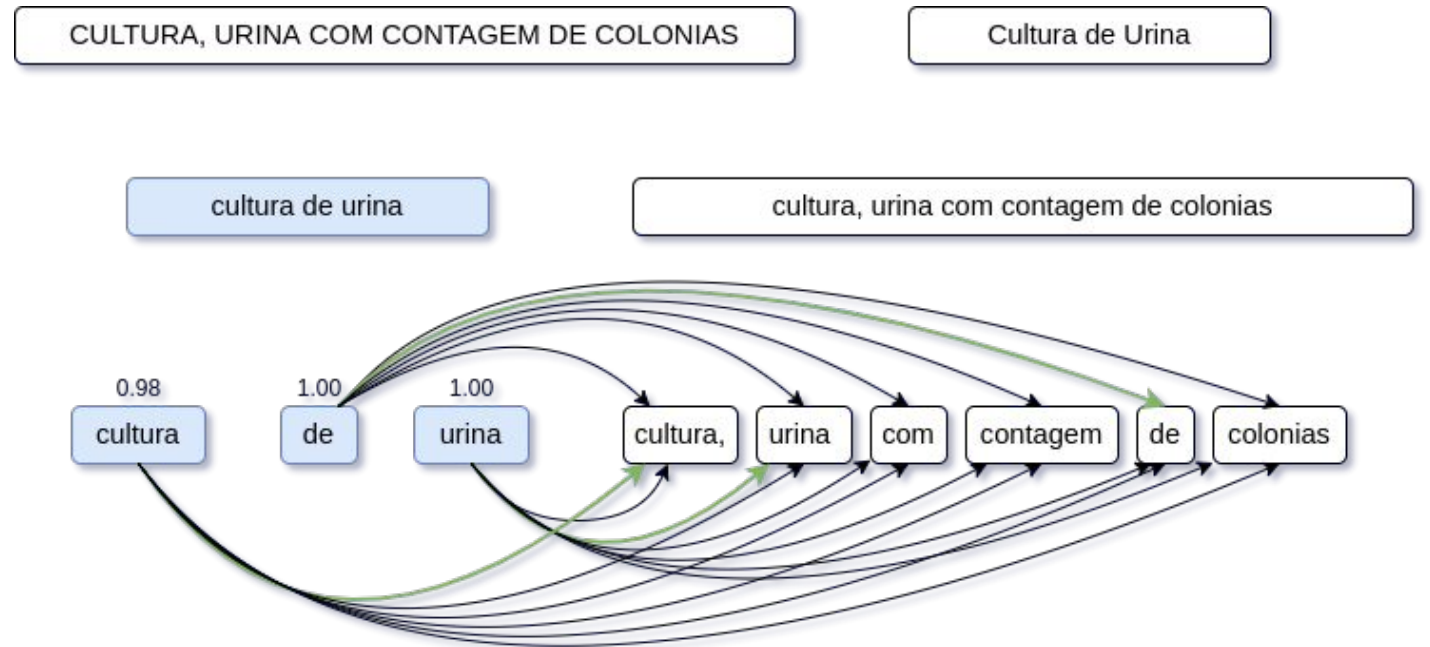


Azure Form Recognizer  
API



# Methodology

- Similaridade de string
- Levenshtein
  - word splitter



$$\text{Semelhança} = \frac{(0.98 + 1.00 + 1.00) * 100}{3}$$

$$\text{Semelhança} = 99,333 \%$$

# Methodology

## Módulos da ferramenta desenvolvida

- stamps and signature detection
  - Yolov5
- head information extraction
  - OCR
  - Levenshtein

YOLOv5



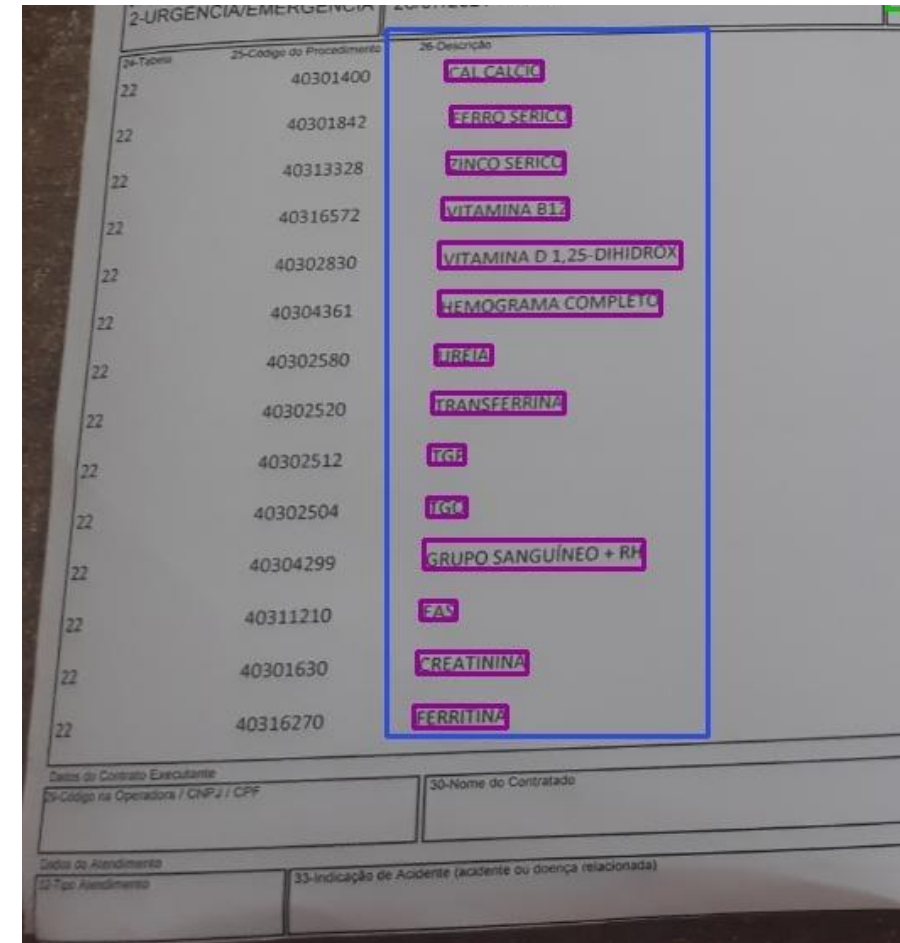
# Methodology

## Modules of the Developed Tool

- Service Description Identification
  - OCR (Amazon Textract)
  - YOLOv5
  - Description Dictionary
  - Levenshtein Distance Algorithm

Tabela 1 – Tabela de comparação de descrições

ACIDO ASCORBICO	ACIDO LACTICO	Anti-TPO
VITAMICA C	ÁC. LÁTICO	PEROXIDASE
VIT C	LACTATO	PEROXIDASE TPO
ÁC. ASCORBICO	Ácido láctico (lactato) - pesquisa e/ou dosagem	Anti-TPO - pes- quisa e/ou dosa- gem
Ácido ascórbico (vita- mina C) - pesquisa e/ou dosagem	ACIDO LACTICO	Anti-TPO
ACIDO ASCORBICO		

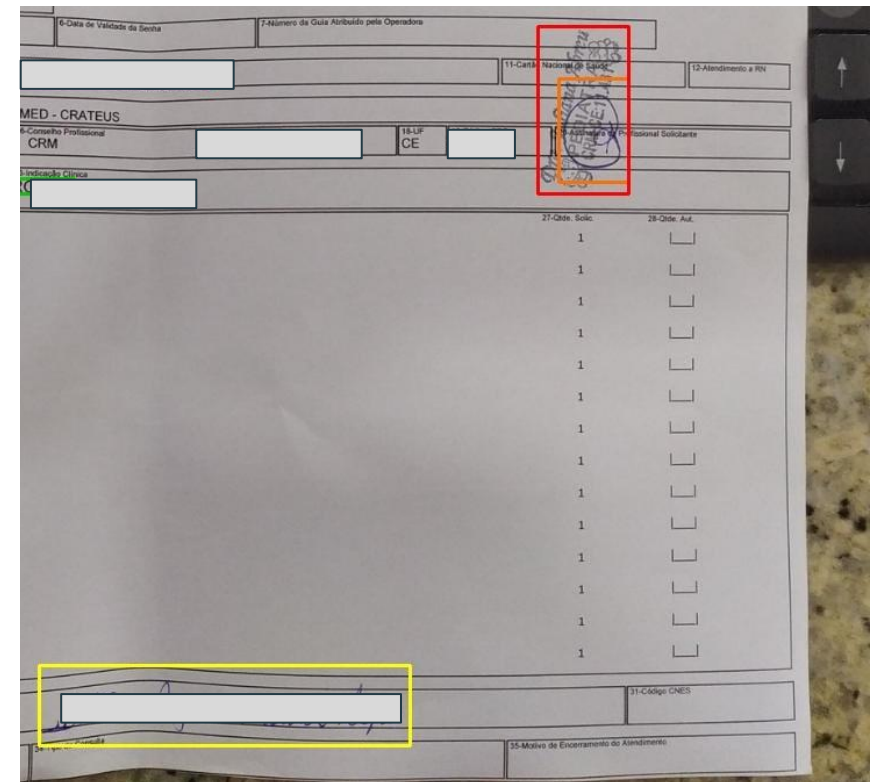


# Result

## stamps and signature detection

**Tabela 5 – Tabela de detecção de carimbos e assinaturas**

Dado	Positivo	Falso Positivo	Total
Carimbo Medico	100	0	100
Assinatura Medico	80	0	100
Assinatura Beneficiário	73	0	90



# Resultados

## head information extraction

**Tabela 6 – Tabela de detecção de cabeçalho**

Dado	Positivo	Falso Positivo	Total
Data	97	0	97
Requisitante	90	0	97
Beneficiário	86	0	97
Indicação clinica	90	0	97

8-Numero da Carteira [redacted] 92		9-Validade da Carteira 31/12/2022	10-Nome [redacted]			
Dados do Solicitante		14-Nome do Contratado PRONTO ATENDIMENTO UNIMED - CRATEUS				
13-Código na Operadora / CNPJ / CPF [redacted]		15-Nome do Profissional Solicitante [redacted] C		16-Conselho Profissional CRM	17-Número no Conselho [redacted]	18-UF CE
Dados da Solicitação / Procedimentos e Exames Solicitados		22-Data da Solicitação 20/08/2021 - 14:48		23-Indicação Clínica [redacted]		
21-Caráter do Atendimento 2-URGÊNCIA/EMERGÊNCIA						

# Results

## Service Description Extraction

- 100 invoices/documents used for testing;
- 669 service descriptions identified in the dataset;
- 569 successfully extracted, with 5 false positives.

The image shows a screenshot of a medical invoice table. The table has three columns: '24-Tipos', '25-Código do Procedimento', and '26-Descrição'. The '26-Descrição' column contains various medical services, each highlighted with a purple box. A blue rectangle highlights the entire '26-Descrição' column. Below the table, there are fields for 'Dados do Contrato Executante' and 'Dados do Atendimento'.

24-Tipos	25-Código do Procedimento	26-Descrição
22	40301400	CALCULIC
22	40301842	FERRO SERICO
22	40313328	ZINCO SERICO
22	40316572	VITAMINA B12
22	40302830	VITAMINA D 1,25-DIHIIDROX
22	40304361	HEMOGRAMA COMPLETO
22	40302580	TREIA
22	40302520	TRANSFERRINA
22	40302512	TCG
22	40302504	TCG
22	40304299	GRUPO SANGUÍNEO + RH
22	40311210	FA
22	40301630	CREATININA
22	40316270	FERRITINA

Dados do Contrato Executante  
29-Código na Operadora / CNPJ / CPF

Dados do Atendimento  
33-Indicação de Acidente (acidente ou doença relacionada)

# Results

## Complete solution

- 200 medical forms used for testing;
- 1,307 service descriptions, 580 signatures and stamps, and 772 header information fields, totaling 2,659 pieces of information that needed to be detected and extracted;
- 2,517 correctly extracted items, achieving an accuracy rate of 94.56%;
- 52% of the medical forms had 100% of the information successfully detected and extracted.

**Tabela 7 – Tabela Erros por Guias**

Numero de Erros	Numero de Guias
0	104
1	65
2	23
3	5
4	2
>4	1

# Results

**GUIA DE SERVIÇO PROFISSIONAL / SERVIÇO AUXILIAR DE DIAGNÓSTICO E TERAPIA - SP/SADT**

2 - Nº Guia no Prestador: [REDACTED]  
RG: [REDACTED]  
D/N: 05/05/1997

1-Região ANS [REDACTED] 3-Número da Guia Principal [REDACTED] 7-Número da Guia Atribuída pelo Operadora [REDACTED]

4-Data de Autorização [REDACTED] 5-Data [REDACTED] 6-Data de Validade da Guia [REDACTED] 11-Certidão Nacional de Saúde [REDACTED] 12-Assinatura a PIS [REDACTED]

8-Data do Beneficiário [REDACTED] 9-Validade da Carteira 31/12/2022 [REDACTED] 13-Assinatura a PIS [REDACTED]

10-Número da Carteira [REDACTED] 14-Nome do Contratado PRONTO ATENDIMENTO CRATEUS [REDACTED] 15-Código Nacional de Saúde [REDACTED] 16-Assinatura a PIS [REDACTED]

17-Código de Operadora / CNPJ / CNP [REDACTED] 18-Comércio Profissional CRM [REDACTED] 19-Código CBO [REDACTED] 20-Assinatura a PIS [REDACTED]

21-Código de Atendimento 2-URGÊNCIA/EMERGÊNCIA 05/09/2021 - 16:54 22-Descrição Rotina Médica [REDACTED]

24-Ítem	25-Código de Procedimento	26-Descrição	27-Quantidade	28-Status
22	40302423	HEMOGRAMA	1	<input type="checkbox"/>
22	40304361	HEMOGRAMA COMPLETO	1	<input type="checkbox"/>
22	40307760	PARA	1	<input type="checkbox"/>
22	40302580	UREIA	1	<input type="checkbox"/>
22	40302547	TRIGLICÉRIDOS	1	<input type="checkbox"/>
22	40302512	UREIA	1	<input type="checkbox"/>
22	40302504	UREIA	1	<input type="checkbox"/>
22	40302318	SODASSIA	1	<input type="checkbox"/>
22	40302237	MAGNÉSIO	1	<input type="checkbox"/>
22	40301591	LDL	1	<input type="checkbox"/>
22	40302733	HEMOGLOBINA GLICOSADA	1	<input type="checkbox"/>
22	40301583	UREIA	1	<input type="checkbox"/>
22	40302040	UREIA	1	<input type="checkbox"/>
22	40301630	UREIA	1	<input type="checkbox"/>
22	40301605	COLESTEROL TOTAL	1	<input type="checkbox"/>
22	40316114	HBSAG	1	<input type="checkbox"/>

29-Data de Assinatura [REDACTED]

# Conclusão

- Most of the information could be successfully extracted.
- The solution can reduce the workload of auditors.

## -> Limitations

- Low image quality;
- Handwritten text.

## -> Future Work

- Increase dataset diversity;
- Apply NLP techniques for key-value extraction;
- Use OCR specialized in medical handwriting recognition.

## Referências Bibliográficas

**FORSYTH, D.; PONCE, J. Computer Vision: A Modern Approach. (Second edition). Pren-tice Hall, 2011. 792 p. Disponível em: <<https://hal.inria.fr/hal-01063327>>.**

**SILVA, I. N. D.; SPATTI, D. H.; FLAUZINO, R. A. Redes neurais artificiais para engenharia e ciências aplicadas-curso prático. São Paulo: Artliber, 2016.**

**YOUNAS, J.; AFZAL, M. Z.; MALIK, M. I.; SHAFAIT, F.; LUKOWICZ, P.; AHMED, S.D-StaR: A generic method for stamp segmentation from document images. In: 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). IEEE, 2017. Disponível em: <<https://doi.org/10.1109/icdar.2017.49>>.**