

Detecção e Análise Automática do Deslocamento Vertical do Osso Hioide em sua Amplitude Máxima em Exames de Ultrassonografia da Deglutição

Arthur José Lopes da Silva



CENTRO DE INFORMÁTICA
UNIVERSIDADE FEDERAL DA PARAÍBA

João Pessoa - PB

2023

Arthur José Lopes da Silva

Detecção e Análise Automática do Deslocamento
Vertical do Osso Hioide em sua Amplitude Máxima
em Exames de Ultrassonografia da Deglutição

Dissertação apresentada ao programa Pós-Graduação em Informática
do Centro de Informática, da Universidade Federal da Paraíba,
como requisito para a obtenção do grau de Mestre em Informática

Orientadora: Profa. Dra. Thaís Gaudencio Rêgo

Julho de 2023

Catálogo na publicação
Seção de Catalogação e Classificação

S586d Silva, Arthur José Lopes da.

Detecção e análise automática do deslocamento vertical do osso hioide em sua amplitude máxima em exames de ultrassonografia da deglutição / Arthur José Lopes da Silva. - João Pessoa, 2023.

56 f. : il.

Orientação: Thaís Gaudencio do Rêgo.
Dissertação (Mestrado) - UFPB/CI.

1. Inteligência artificial. 2. Ultrassonografia. 3. Detecção de objetos. 4. Osso hioide. 5. Aprendizado profundo. 6. YOLO. I. Rêgo, Thaís Gaudencio do. II. Título.

UFPB/BC

CDU 004.8(043)



UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA



Ata da Sessão Pública de Defesa de Dissertação de Mestrado de **Arthur José Lopes Da Silva**, candidato ao título de Mestre em Informática na área de Sistemas de Computação, realizada em 31 de julho de 2023.

Aos trinta e um dias do mês de julho do ano de dois mil e vinte e três, às 10h, no Centro de Informática da Universidade Federal da Paraíba, reuniram-se os membros da Banca Examinadora constituída para julgar o Trabalho Final do discente David Pereira Galvão Júnior, vinculado a esta Universidade sob a matrícula nº 20211023510, candidato ao grau de Mestre em Informática, na área de “*Sistemas de Computação*”, na linha de pesquisa “*Computação Distribuída*”, do Programa de Pós-Graduação em Informática. A comissão examinadora foi composta pelos professores: Thais Gaudencio do Rego, Orientadora e Presidente; Leonardo Vidal Batista, Examinador Interno ao Programa; Yuri de Almeida Malheiros Barbosa, Examinador Externo ao Programa. Dando início aos trabalhos, a Presidente da Banca cumprimentou os presentes, comunicou a finalidade da reunião e passou a palavra ao candidato para que ele fizesse a exposição oral do trabalho de dissertação intitulado “**Detecção e Análise Automática do Deslocamento Vertical do Osso Hioide em sua Amplitude Máxima em Exames de Ultrassonografia da Deglutição**”. Concluída a exposição, o candidato foi arguido pela Banca Examinadora que emitiu o seguinte parecer: “**aprovado**”. Do ocorrido, eu, Fernando Menezes Matos, Coordenador do Programa de Pós-Graduação em Informática, lavrei a presente ata que vai assinada por mim e pelos membros da Banca Examinadora. João Pessoa, 31 de julho de 2023.

Prof. Dr. Fernando Menezes Matos

Prof. Dr^a. Thais Gaudencio do Rego
Orientadora (PPGI-UFPB)

Prof. Dr. Leonardo Vidal Batista
Examinador Interno (PPGI-UFPB)

Prof. Dr. Yuri de Almeida Malheiros Barbosa
Examinador Externo à Instituição (UFPB)

RESUMO

Localizado abaixo da mandíbula, o osso hioide está presente na parte superior do pescoço e participa de funções essenciais para o ser humano, como a fala, deglutição e respiração. Na deglutição, o osso hioide realiza uma elevação e um deslocamento para frente, simultaneamente. O estudo do seu movimento é de grande importância, pois, caso esteja comprometido e não seja realizado de maneira correta, pode gerar limitações ao indivíduo e auxiliar no aparecimento de doenças que comprometem a sua rotina. Tal estudo, geralmente, é feito a partir de uma análise do deslocamento do osso, onde os profissionais da saúde executam a análise sem ferramentas que automatizam esse processo. Por ser um osso de tamanho pequeno (com de cerca de 42,04 milímetros) e em movimento, a qualidade do resultado dos exames ou diagnósticos para o paciente podem ser comprometidos. Pensando nisso, o presente estudo tem como objetivo avaliar a possibilidade de detectar e calcular o deslocamento vertical do osso hioide com auxílio de métodos de detecção de objetos por meio de aprendizado profundo, através de informações provenientes dos exames de ultrassonografia de pacientes jovens adultos deglutindo água, saliva e líquido pastoso, e, partir disso, desenvolver uma ferramenta capaz de auxiliar os profissionais da saúde. Foi realizado o treinamento e, posteriormente, a detecção utilizando YOLO na versão 7-tiny. Uma vez detectado, foi feito o cálculo do deslocamento vertical do osso hioide através de dados do exame de ultrassonografia e informações resultantes da detecção, tal cálculo teve o foco na amplitude máxima, ou seja, do repouso ao pico máximo do deslocamento. Após isso, todos resultados foram mostrados e avaliados segundo precisão e revocação, obtendo valores de 0,95 e 0,93 respectivamente, e também segundo ao cálculo de deslocamento, onde foram comparados a outros resultados da literatura, de três comparações feitas, duas se mostraram mais próximas, nas deglutições de líquido pastoso ($12,22\text{mm} \pm 3,63$) x Chi-Fishman e Sonies (2002) ($8,50\text{mm} \pm 3,31$) e líquido ($10,33\text{mm} \pm 5,77$) x Yamazaki et al. (2017) ($10,88\text{mm} \pm 1,47$). O presente estudo mostrou uma boa precisão na detecção do movimento do osso hioide e uma aproximação em relação ao cálculo do deslocamento vertical comparado à literatura.

Palavras-chave: Osso Hioide, Ultrassonografia, Detecção de Objetos, Inteligência Artificial, Aprendizado Profundo, YOLO

ABSTRACT

Located below the mandible, the hyoid bone is present in the upper part of the neck and participates in essential human functions such as speech, swallowing and breathing. When swallowing, the hyoid bone simultaneously lifts and moves forward. The study of movement is of great importance, because if it is compromised and not performed correctly, it can lead to limitations for the individual and help to cause illnesses that compromise their routine. This study is usually based on an analysis of bone displacement, where healthcare professionals carry out the analysis without tools to automate this process. Because it is a small bone (about 42.04 millimeters) and it is moving, the quality of test results or diagnoses for the patient can be compromised. With this in mind, this study aims to evaluate the possibility of detecting and calculating the vertical displacement of the hyoid bone with the aid of deep learning object detection methods, using information from ultrasound scans of young adult patients swallowing water, saliva and pasty liquid, and, from this, develop a tool capable of assisting healthcare professionals. Training and then detection were carried out using YOLO version 7-tiny. Once detected, the vertical displacement of the hyoid bone was calculated using data from the ultrasound scan and information resulting from the detection, focusing on the maximum amplitude, i.e. from rest to the maximum peak of the displacement. After that, all the results were shown and evaluated according to precision and recall, obtaining values of 0.95 and 0.93 respectively, and also according to the displacement calculation, where they were compared to other results in the literature, of three comparisons made, two proved to be closer, in the swallowing of pasty liquid ($12,22\text{mm} \pm 3,63$) x Chi-Fishman e Sonies (2002) ($8,50\text{mm} \pm 3,31$) and liquid ($10,33\text{mm} \pm 5,77$) x Yamazaki et al. (2017) ($10,88\text{mm} \pm 1,47$). This study showed good accuracy in detecting the movement of the hyoid bone and an approximation in relation to the calculation of vertical displacement compared to the literature.

Key-words: Hyoid Bone, Ultrasonography, Object Detection, Artificial Intelligence, Deep Learning, YOLO

LISTA DE FIGURAS

1	Variações de Formato e Tamanho do Osso Hioide	18
2	Osso Hioide e suas partes	19
3	Imagem do Exame de Ultrassonografia, onde está destacado o osso hioide com uma seta vermelha	20
4	Imagem do Exame de Videolaringoscopia	21
5	Imagem do Exame de Videofluoroscopia	21
6	Modelo descritivo da rede neural convolucional - a passagem da imagem de entrada a convolução, extraindo suas características ou pixels mais importantes; a camada de pooling, diminuindo a sua dimensionalidade para aumentar a eficiência computacional e preservando as característica; passando para o flattening, transformando a matriz resultante das camadas anteriores, para uma lista, que assim servirá de entrada para a rede neural, na camada densa.	23
7	Arquitetura da YOLO	25
8	Grade $S \times S$ na imagem	26
9	O Y corresponde ao vetor de informações correspondentes a bounding box, confiança (p_c), posição (b_x, b_y), dimensões (b_h e p_w) e a classe do objeto (c)	26
10	IoU - Intersecção sobre União, representada graficamente na fórmula e na imagem detectada. A caixa verde caracteriza a rotulação real, onde realmente o objeto está localizado, já a caixa vermelha caracteriza o que foi detectada pelo modelo. Nesse caso, o IoU apresentaria um valor próximo de 1, pois as duas caixas estão bem sobrepostas.	27
11	Funcionamento da YOLO - As cores do mapa de probabilidade representam que há possibilidades para a existência de classes nas determinadas regiões, podemos confirmar na detecção final, onde as caixas delimitadoras estão com bordas das respectivas cores.	29
12	Posições do Exame de Ultrassom: (1) Posição Original do Exame, (2) Imagem rotacionada do exame para melhor entendimento e realização do cálculo do deslocamento	37
13	Exame de Ultrassonografia utilizado no presente estudo	38
14	Ilustração da Proposta de Cálculo do deslocamento vertical do osso hioide (do seu repouso ao pico máximo)	40

15	Fluxo da Página WEB	41
16	Detecção do osso hioide com o paciente deglutindo saliva	42
17	Detecção do osso hioide com o paciente deglutindo consistência pastosa . .	43
18	Detecção do osso hioide com o paciente deglutindo água	43
19	IoU - Osso Hioide - A caixa delimitadora da cor verde representa a rotulação real do osso, já a de cor rosa, a detecção realizada pelo modelo. O IoU do exemplo é retratado na expressão.	44
20	Erro na detecção, no caso, duas caixas delimitadores estão sendo detecta- das.	45
21	Página inicial do ambiente de visualização	49
22	Página de resultado do ambiente de visualização	49
23	Ambiente de Visualização - Arquivo inserido com sucesso	58
24	Ambiente de Visualização - Nenhum arquivo carregado	59
25	Ambiente de Visualização - Número de arquivos excedidos	59

LISTA DE TABELAS

1	Tabela de comparação entre os trabalhos relacionados e o presente estudo .	33
2	Valores do deslocamento vertical com pacientes deglutindo Saliva	46
3	Valores do deslocamento vertical com pacientes deglutindo uma consistência Pastosa	46
4	Valores do deslocamento vertical com pacientes deglutindo Água	47
5	Tabela de Comparações com a Literatura (mm)	48

LISTA DE ABREVIATURAS

- API - *Application Programming Interface* (Interface de Programação de Aplicação)
- ATOM - *Accurate Tracking by Overlap Maximization* (Rastreo preciso por maximização de sobreposição)
- BiFPN - Rede de Pirâmide Bidirecional
- CNN - *Convolutional Neural Network* (Rede Neural Convolutacional)
- CPN - *Cascaded Pyramid Network* (Rede de Pirâmides em Cascata)
- DiMP - *Discriminative Model Prediction* (Preditor de modelo descritivo)
- DL - *Deep Learning* (Aprendizado Profundo)
- FP - Falsos Positivos
- FN - Falsos Negativos
- GPU - *Graphics Processing Unit* (Unidade de Processamento Gráfico)
- IA - Inteligência Artificial
- IOU - *Intersection-Over-Union* (Intersecção sobre União)
- JSX - *JavaScript XML*
- KNN - *K-Nearest Neighbors* (*K*-Vizinhos mais próximos)
- MAP - *Mean Average Precision* (Média de Precisão Média)
- MDNet - *Multi-Domain Networks* (Redes Multi-Domínios)
- ML - *Machine Learning* (Aprendizagem de Máquina)
- MVC - *Model-View-Controller* (Modelo-Visão-Controle)
- RNA - Rede Neural Artificial
- SiamFC - *Fully-Convolutional Siamese Networks* (Redes Siamesas Totalmente Convolutacionais)
- SVM - *Support Vector Machine* (Máquina de Vetores de Suporte)
- UFPB - Universidade Federal da Paraíba
- USG - Ultrassografia
- VDG - Videodeglutograma
- VFF - Videofluoroscopia
- VLD - Videolaringoscopia da Deglutição

VP - Verdadeiros Positivos

YOLO - *You Only Look Once* (Você só olha uma vez)

Sumário

1	INTRODUÇÃO	15
1.1	Objetivo geral	16
1.2	Objetivos específicos	16
1.3	Estrutura da monografia	17
2	CONCEITOS GERAIS	18
2.1	Osso Hioide	18
2.2	Ultrassonografia	19
2.3	Inteligência Artificial e Aprendizado Profundo	21
2.4	Detecção de objetos	22
2.5	Métricas de Avaliação	24
2.6	YOLO	24
3	TRABALHOS RELACIONADOS	30
4	METODOLOGIA	34
4.1	Base de dados	34
4.2	Ambiente de Execução	35
4.3	YOLO e Treinamento	35
4.4	Métrica de Avaliação dos Resultados	35
4.5	Cálculo de Deslocamento	36
4.6	Cálculo do deslocamento vertical do osso hioide	38
4.7	Ambiente de Visualização	40
5	APRESENTAÇÃO E ANÁLISE DOS RESULTADOS	42
5.1	Erros encontrados	44
5.2	Apresentação quantitativa do deslocamento vertical do osso hioide em sua amplitude máxima	45
5.3	Ambiente de Visualização	48
6	CONSIDERAÇÕES FINAIS	50

REFERÊNCIAS	51
APÊNDICE A - Detalhes do uso da darknet - YOLO	56
APÊNDICE B - Instruções para detecção em vídeo na darknet - YOLO	57
APÊNDICE C - Feedbacks para o usuário - Ambiente de Visualização	58

1 INTRODUÇÃO

Os nutrientes são componentes indispensáveis para o pleno funcionamento do corpo humano. Eles são compostos por proteínas, carboidratos, vitaminas e minerais que estão presentes nos alimentos ingeridos, com isso, através do mecanismo chamado de deglutição, que é realizado de forma que o alimento saia da boca e seja encaminhado para o estômago, a absorção dos nutrientes acontece. Tal mecanismo atua junto com músculos, ossos e outras estruturas do corpo humano. Um dos ossos presente nessa ação de transportar o alimento da boca ao estômago, é o osso hioide, nome derivado do grego “hyodeides”, que tem o significado de “com a forma da letra Upsilon”, baseado, justamente, no seu formato em U ou ferradura. Fica localizado na parte anterior do pescoço, e não faz conexão com nenhum outro osso. Ele ocupa uma posição importante e participa de funções essenciais, além da deglutição, como na fala e respiração (AUVENSHINE et al., 2018).

Quando o mecanismo da deglutição é acionado, o osso hioide realiza um movimento vertical e horizontal, simultaneamente, contribuindo para o êxito do processo, levando o alimento da boca ao estômago. Tal deslocamento ósseo é importante, e caso ele esteja comprometido e não seja realizado de maneira correta, pode gerar limitações ao indivíduo e auxiliar no aparecimento de doenças que comprometem toda a absorção dos nutrientes (ANDRADE, 2021). Com isso, é importante uma análise para compreender o que pode estar acarretando o impedimento do funcionamento pleno do movimento desse osso.

Para tal análise, exames de imagem devem ser realizados primordialmente, e, para a verificação da estrutura e do processo da deglutição, podemos destacar três: a videofluoroscopia (VFF) ou videodeglutograma (VDG), que se trata do padrão ouro para a avaliação da deglutição, pois promove uma mensuração objetiva e imediata do processo deglutivo em diversos ângulos, porém expõe pacientes a radiação (ANEAS; DANTAS, 2014). Em segundo, existe a videolaringoscopia da deglutição (VLD), um exame útil e eficaz, que fornece imagens claras e nítidas do processo de deglutição, porém é invasiva (YANAGISAWA et al., 1983). Por fim, a ultrassonografia (USG) é um exame não invasivo, capaz de fornecer imagens dinâmicas que auxiliam na análise do processo de deglutição (LEITE et al., 2014). O último deles foi o utilizado neste presente trabalho.

Após a realização dos exames, geralmente os profissionais da saúde executam a análise sem ferramentas que automatizam esse processo. O movimento do osso hioide, por exemplo, é detectado e analisado de maneira visual e manual, podendo gerar dificuldade, uma vez que se trata de um osso com tamanho de cerca de 42,04 milímetros, segundo (CHAVAN; JOSHI, 2019). Isso pode retardar e comprometer a qualidade de exames ou diagnósticos para o paciente. Com uma ferramenta que auxilia os profissionais na análise e detecção do osso e seu movimento, é possível aumentar a velocidade de diagnóstico dos pacientes, por exemplo, evitando assim, possíveis complicações e o início precoce dos

tratamentos específicos para cada caso. Além disso, ferramentas desse tipo podem facilitar o trabalho do profissional diminuindo, também, a sua fadiga física.

A detecção automática de objetos, que teve seu início realizado através de técnicas de processamento de imagens, vem sendo impactada pela inteligência artificial (IA), que tem contribuído com a área da saúde, principalmente com toda a possibilidade de processamento de dados, com o uso de aprendizado de máquina - ML (do inglês, *machine learning*), junto com o aprendizado profundo - DL (do inglês, *deep learning*). Este último método, parte da ML, vem contribuindo enormemente na área de detecção e será utilizada no presente estudo, com o intuito de auxiliar na análise quantitativa do movimento osso e promover precisão e velocidade a esse processo (YU; BEAM; KOHANE, 2018). Além disso, o foco no exame de ultrassonografia, por ser não invasivo, de baixo custo, e eficiente na visão do mecanismo de deglutição, é outro ponto importante, fortalecendo as vantagens entre os pacientes e profissionais de saúde.

Pensando nisso, o presente trabalho tem como princípio objetivo desenvolver uma ferramenta para detectar e analisar quantitativamente o osso hioide e seu movimento, utilizando DL para auxiliar o trabalho dos profissionais de saúde.

1.1 Objetivo geral

O objetivo do presente estudo é avaliar a possibilidade de detectar e calcular o deslocamento vertical do osso hioide com auxílio de métodos de detecção de objetos por meio de aprendizado profundo, através de informações provenientes dos exames de ultrassonografia de pacientes jovens adultos deglutindo água, saliva e líquido pastoso, e, partir disso, desenvolver uma ferramenta capaz de auxiliar os profissionais da saúde.

1.2 Objetivos específicos

- Treinar e avaliar modelos de detecção de objetos para detectar o osso hioide;
- Desenvolver proposta de cálculo para medição do deslocamento vertical do osso;
- Calcular a distância em milímetros do deslocamento vertical do osso hioide na deglutição com água, saliva e líquido pastoso;
- Comparar os resultados calculados com os valores encontrados na literatura;
- Desenvolver uma ferramenta web para manipulação do processo de detecção e visualização dos resultados.

1.3 Estrutura da monografia

Este documento está organizado em 6 capítulos. No Capítulo 2 serão apresentados os conceitos gerais e fundamentais para entendimento do trabalho, assuntos como Osso Hioide, Exame de Ultrassom, Detecção de Objetos e YOLO serão abordados. No Capítulo 3, a revisão da literatura é detalhada, destacando os trabalhos relacionados para a comparação com o presente estudo.

No Capítulo 4 será apresentada toda a metodologia do trabalho, incluindo a forma aplicada e as ferramentas utilizadas. Já no Capítulo 5, os resultados da pesquisa são expostos. Para finalizar, serão apresentadas as considerações finais no Capítulo 6.

2 CONCEITOS GERAIS

Nesta seção será abordado o conteúdo teórico, importante para o entendimento do presente trabalho.

2.1 Osso Hioide

Localizado abaixo da mandíbula, que fica na parte superior do pescoço, o osso hioide é o único que não se articula com nenhum dos outros 205 que existem no corpo humano, fazendo apenas conexões com músculos ao seu redor. O osso hioide se encontra firme à base do crânio, mandíbula e laringe, e participa de funções vitais para o ser humano, atuando na fonação (em ligação com a laringe), consequentemente, fazendo parte da ação da fala, respiração, mastigação, deglutição, sendo a última, o foco do presente estudo. Seu nome tem uma derivação do grego “hyodeides”, que significa “formato da letra Upsilon”, relacionado a sua forma em “U”, ou de uma ferradura, existindo algumas variações em relação a ela e a seu tamanho, como podemos ver na Figura 1 (AUVENSHINE et al., 2018).



Figura 1: Variações de Formato e Tamanho do Osso Hioide

Fonte: Auvenshine et al. (2018)

Em relação a sua anatomia, o osso hioide é dividido em 3 partes principais: o corpo, cornos maiores e cornos menores (Figura 2). O corno maior tem um formato irregular e alongado nas laterais, além de da convexidade em “U”, e também pode ser chamado de arco hioide. Os cornos maiores têm uma extensão nas laterais do corpo e possuem um achatamento horizontal. Com a união do corpo com os cornos maiores é formado um

relevo ósseo modelado com a aparência de um cone, também chamado de corno menor. Eles se conectam ao corpo do osso hioide através de um tecido de fibras, segundo Fakhry et al. (2013).

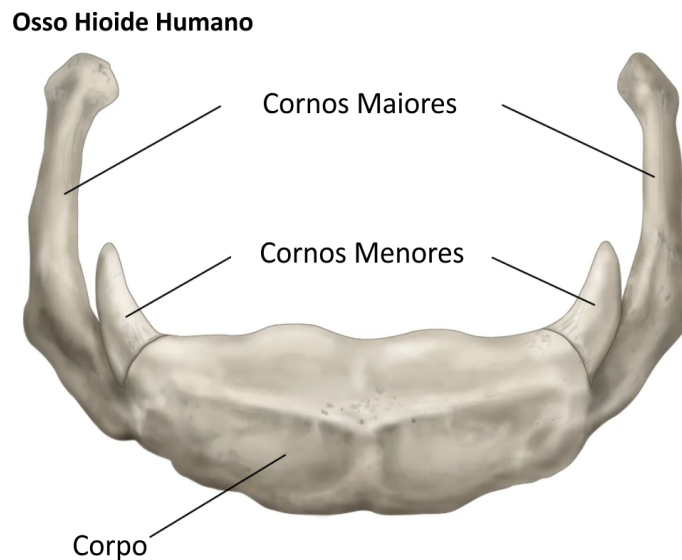


Figura 2: Osso Hioide e suas partes

Fonte: Auvenshine et al. (2018)

Na deglutição, o osso hioide realiza um movimento, ou mais precisamente, uma elevação para cima e para frente, em decorrência da contração dos músculos supra-hióideos - que são músculos que se estendem da mandíbula até o osso hioide - , causando a abertura da esfíncter superior do esôfago e que também auxilia na proteção das vias aéreas no momento da ingestão dos alimentos (ANDRADE, 2021). Para a visão detalhada e possibilidade de estudo do osso e seu movimento, exames de imagem devem ser feitos, onde, o presente estudo terá o foco no exame ultrassonográfico.

2.2 Ultrassonografia

Ultrassom, ou Ultrassonografia (USG), é um exame de imagem que permite a visualização de órgãos e estruturas internas do corpo humano através de ondas sonoras (Figura 3). Junto com um transdutor, o profissional de saúde aplica um gel sob a região e movimenta o equipamento na pele. A partir daí são emitidas ondas sonoras, e as imagens resultantes são provenientes do eco formado pelas ondas (CDC, 2019).

Tal exame foi escolhido neste trabalho como foco para a análise do movimento do osso hioide na deglutição, visto que é não invasivo, com baixo custo, seguro, rápido e bastante utilizado nos estudos das estruturas envolvidas na deglutição. A ultrassonografia

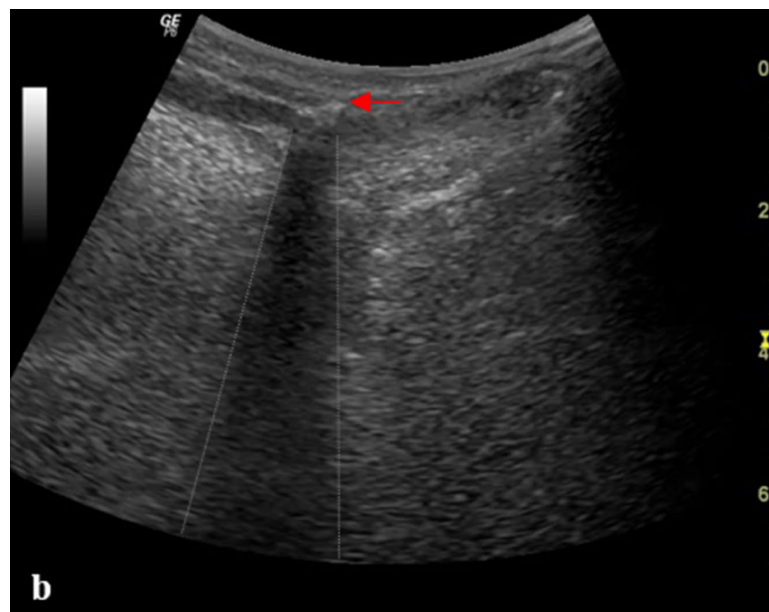


Figura 3: Imagem do Exame de Ultrassonografia, onde está destacado o osso hioide com uma seta vermelha

Fonte: Costa et al. (2021)

possibilita o paciente ingerir alimento na situação habitual, sem a presença de contrastes, e a realização do exame em lugares cômodos, como por exemplo na beira do leito, e, além disso, não o expõe a radiação, diferente de outros exames já mencionados e comuns, também no estudo da deglutição, como a videolaringoscopia e o videodeglutograma, ou videofluoroscopia (LEITE et al., 2014).

A videolaringoscopia, ou videolaringoscopia da deglutição, é dita como uma técnica invasiva e desconfortável ao paciente (Figura 4), assim como a videofluoroscopia, ou videodeglutograma (Figura 5). Mesmo esta última sendo considerada o melhor recurso para avaliação da deglutição, é um exame que expõe o paciente à radiação (ANEAS; DANTAS, 2014; FREITAS, 2017). Com isso, é reforçado o intuito do uso da USG no presente estudo, pois além de ser uma técnica que é utilizada no estudo da deglutição, promove vantagens aos pacientes.

Profissionais da saúde analisam exames diariamente, porém, a possibilidade do uso da inteligência artificial para auxiliar, ou automatizar a análise para diagnósticos com mais velocidade e precisão, é de grande valia ao profissional e a área da saúde.



Figura 4: Imagem do Exame de Videolaringoscopia

Fonte: Otomedica (2020)



Figura 5: Imagem do Exame de Videofluoroscopia

Fonte: Marques (2015)

2.3 Inteligência Artificial e Aprendizado Profundo

Uma das áreas mais impactadas pelo surgimento e crescimento da inteligência artificial é a da saúde. Hoje são desenvolvidos sistemas para apoio à decisão, com o intuito de escolher tratamentos adequados para os pacientes e auxiliar na geração de hipóteses diagnósticas de doenças. No entanto, os primeiros sistemas desenvolvidos e que ainda não faziam uso dessa tecnologia, eram limitados e possivelmente pouco eficazes, uma vez que requerem expressões nítidas de regras de decisão e também de atualização humana rotineira. Com o uso de IA, o processo foi facilitado, tornando os sistemas capazes de tomar decisões mais complexas, e recentemente, sendo capazes de diagnosticar doenças, inclusive através de imagens médicas, aumentando, em algumas situações, drasticamente sua precisão (FREITAS, 2017).

O período vivido hoje é o de informações em grandes quantidades, para vários cenários. Com isso, dados de pacientes podem ser acessados para processamento, consequentemente, garantindo uma maior experiência para a construção de um modelo em IA. Um profissional de saúde, com anos de carreira, examina um número expressivo de exames, mas uma IA tem a capacidade de examinar um número ainda maior em menos tempo, agilizando, dessa forma, os diagnósticos de doenças em pacientes por especialistas, o que acarretará em um tratamento precoce, aumentando as chances de cura (MINTZ; BRODIE, 2019).

A inteligência artificial (IA) é um ramo da ciência da computação, que tem como objetivo a mecanização de tarefas complexas como reconhecer rostos, fazer diagnósticos de doenças, compor música, escrever histórias, entre outras tarefas, presentes em diversas áreas do cotidiano. (SINGH, 2021)

A IA é uma grande área que abrange a aprendizagem de máquina - AM, do inglês Machine Learning (ML), e sua subárea, aprendizado profundo, do inglês, Deep Learning (DL), entre outras que, juntas contribuem para sua evolução. Dentre os campos citados, a AM consiste no aprendizado automático dos computadores por relações e paradigmas que permitem gerar sistemas inteligentes, apresentando previsões, respostas, recomendações, entre outras (JANIESCH; ZSCHECH; HEINRICH, 2021).

Existem vários algoritmos dentro do aprendizado de máquina que são responsáveis por diversas tarefas diferentes, dependendo da forma que for aplicado. Regressão Linear, Árvores de decisão, Naive Bayes, K -vizinhos mais próximos (do inglês, *K-Nearest neighborhood*, K -NN, Máquina de Vetor de Suporte (do inglês, *Support Vector Machine* - SVM e Redes Neurais Artificiais (RNAs) são alguns dos métodos presentes no aprendizado de máquina, onde o último, tem uma aplicação ainda mais vasta no aprendizado profundo, que consiste no ramo da ML e consequentemente da IA, que tem como objetivo melhorar como o modelo faz o reconhecimento visual, da fala, a detecção de objetos, entre outros, detectando também, automaticamente, os atributos dos dados de entrada relevantes para essas tarefas (LECUN; BENGIO; HINTON, 2015).

2.4 Detecção de objetos

A arquitetura de rede neural profunda utilizada neste trabalho são as redes neurais convolucionais, do inglês, *Convolutional Neural Network* - CNN, ou ConvNets, que podem ser utilizadas para classificar imagens, reconhecer objetos dentro de cenas, identificar rostos, indivíduos e muitos outros dados visuais. A rede neural convolucional é capaz de capturar as dependências de espaço de uma imagem através da aplicação de filtros, ou seja, a rede pode ser treinada para captar características da imagem. A arquitetura de uma CNN é um tipo de RNA, com alguns diferenciais. O principal diferencial é a

aplicação da convolução, uma operação básica realizada no processamento de imagens, onde filtros realizam esse processo com alguma máscara ou *kernel*, como por exemplo, para detectar bordas em uma imagem, o resultado da convolução é um mapa (matriz) de características mais importantes da imagem que usualmente é passada para a camada de *pooling*, onde será reduzido a sua dimensionalidade e consequentemente diminuindo o custo de memória e de processamento. Após toda a redução, a matriz será convertida em uma lista na camada *Flattening* para assim estar apta como entrada para a camada densa, onde fica a rede neural totalmente conectada. (DESHPANDE, 2016) Na Figura 6 podemos ver o modelo que descreve esse processo

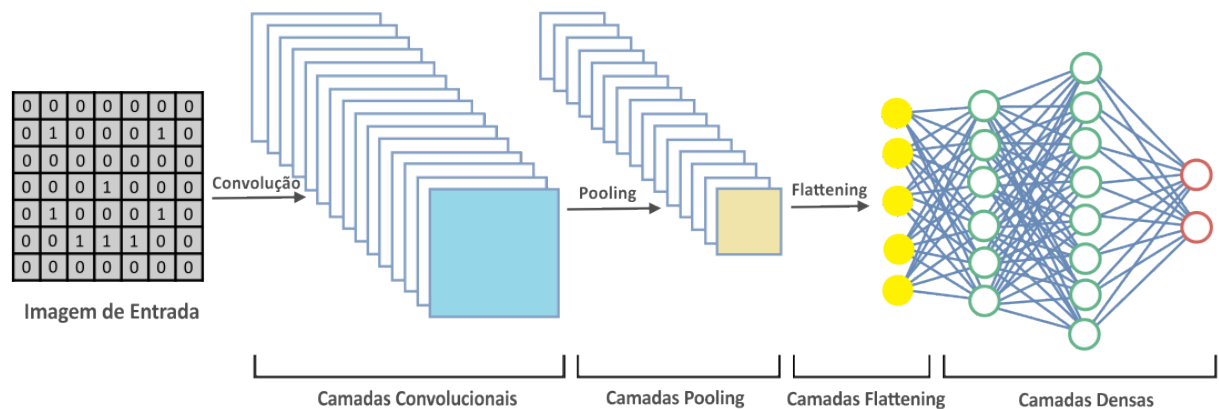


Figura 6: Modelo descritivo da rede neural convolucional - a passagem da imagem de entrada a convolução, extraindo suas características ou pixels mais importantes; a camada de pooling, diminuindo a sua dimensionalidade para aumentar a eficiência computacional e preservando as característica; passando para o flattening, transformando a matriz resultante das camadas anteriores, para uma lista, que assim servirá de entrada para a rede neural, na camada densa.

Fonte: Elaborada pelo autor e adaptada de Staff (2020)

A detecção está relacionada a outras técnicas como reconhecimento e a segmentação de imagens. O reconhecimento de imagens provê apenas o rótulo da classe de um objeto e na segmentação é feita a identificação dos pixels dos objetos de interesse presentes em uma cena. A diferença da detecção de objetos em relação às duas mencionadas, é o fato de ser capaz de localizar objetos na imagem ou vídeo, permitindo o rastreamento ou contagem desses objetos. Com isso, ela pode ser aplicada de várias maneiras, como por exemplo na área de segurança, em vídeos de vigilância, na detecção de faces, de anomalias, demonstrando assim, que essa tarefa está cada vez mais presente no nosso dia a dia (PIEMONTEZ, 2022).

2.5 Métricas de Avaliação

Os detectores de objetos são capazes de oferecer a localização e o rótulo dos objetos, e para determinar o desempenho do modelo, a métrica de avaliação que é mais utilizada é chamada de Interseção sobre a União, do inglês, *Intersection-over-Union* (IoU), onde, a partir de duas caixas delimitadoras, é possível fazer essa operação de divisão, descrita na equação 1. A variação do valor resultante é de 0 a 1, onde 0 seria sem sobreposição e 1 perfeitamente sobreposto (KARIMI, 2021).

$$IoU = \frac{\text{Área de Intersecção}}{\text{Área de União}} \quad (1)$$

Além do IoU, é possível destacar outras métricas para analisar o desempenho da YOLO, como a precisão e o *revocação*, que tem como fórmulas, respectivamente:

$$\text{Precisão} = \frac{\text{Verdadeiros Positivos}}{\text{Verdadeiros Positivos} + \text{Falsos Positivos}} \quad (2)$$

$$\text{Revocação} = \frac{\text{Verdadeiros Positivos}}{\text{Verdadeiros Positivos} + \text{Falsos Negativos}} \quad (3)$$

Também pode-se destacar a Média de Precisão Média (do inglês, *mean Average Precision* - mAP). A precisão média (do inglês, *average precision* - AP, é uma métrica de avaliação da precisão média dos detectores de objeto, e o mAP é a média da precisão média, ou seja, enquanto o AP é calculado para cada classe, no mAP, é feito a média do AP. (HUI, 2018).

Existem vários modelos que pertencem aos conceitos destacados, e, a seguir, será apresentado o modelo de detecção de objetos utilizado no presente trabalho.

2.6 YOLO

A abreviação YOLO advém do termo “Só se olha uma vez”, do inglês, *You Only Look Once*. É um algoritmo capaz de detectar e reconhecer um ou vários objetos em imagens, destacando-se por ser em tempo real. A YOLO fornece probabilidades das classes detectadas presentes nas imagens, chamadas de confianças, e também utiliza CNNs para realizar as detecções. Só necessita percorrer a imagem apenas uma vez, assim como o nome sugere. Toda previsão na imagem é realizada em uma execução do algoritmo (LIMA, 2020).

Comparada a outros modelos, a YOLO se destaca em alguns pontos, como por exemplo, na velocidade, por não possuir pipelines complexos. Além disso, também possui uma detecção com precisão em várias situações, permitindo o desenvolvimento de

aplicações na detecção dos objetos. É um algoritmo de código aberto, facilitando assim a participação da comunidade para melhorias constantes do modelo (KEITA, 2022).

A YOLO utiliza uma rede neural profunda, que possui uma arquitetura chamada Darknet, que têm o mesmo nome do *framework* que implementa o detector de objetos. Segundo Redmon et al. (2015), a arquitetura foi implementada com camadas convolucionais no início da rede, que são responsáveis pela extração de características da imagem interessada, e, as camadas que são totalmente conectadas são encarregadas da previsão de probabilidades e das coordenadas que são geradas na saída. Tal arquitetura tem uma inspiração da GoogLeNet (ALAKE, 2020), que possui 24 camadas de convolução, junto as camadas *maxpool* responsáveis por agrupar os pixels de maior intensidade da imagem e 2 camadas que são totalmente conectadas, além disso utiliza camadas de redução 1x1, acompanhadas de camadas convolucionais 3x3 (Figura 7).

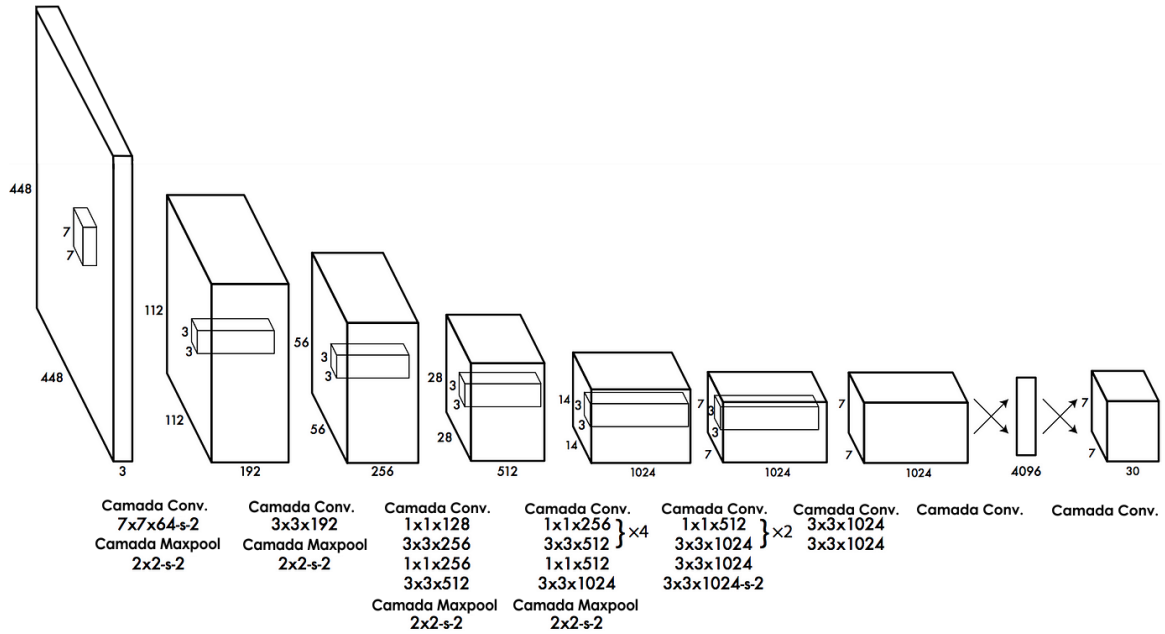


Figura 7: Arquitetura da YOLO

Fonte: Redmon et al. (2015)

Em relação ao seu funcionamento, primeiro, a YOLO divide a imagem em $S \times S$, onde S^2 é o número de células de grade, e cada célula é responsável pela localização e previsão da classe do objeto que sobrepõe, acompanhado do seu valor de confiança (probabilidade). Por exemplo, caso o centro do objeto se localize em uma determinada célula, ela vai ser responsável pela sua detecção (Figura 8).

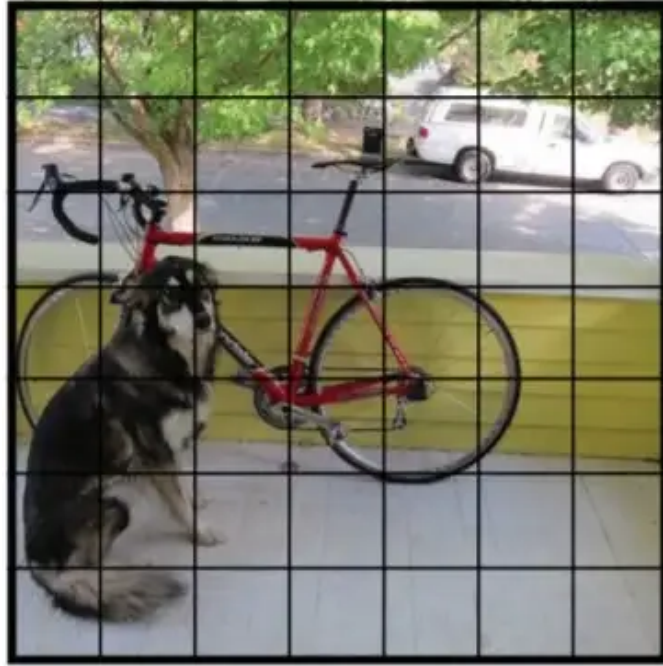


Figura 8: Grade $S \times S$ na imagem

Fonte: Khurana (2019)

Após isso, temos as caixas delimitadoras, do inglês *bounding boxes*, que são as caixas que fazem destaque aos objetos detectados na imagem. O número de *bounding boxes* pode ser maior do que o número de objetos esperados presentes, e elas possuem atributos que as representam vetorialmente após a execução do modelo (Figura 9).



$$y = (p_c, b_x, b_y, b_h, b_w, c)$$

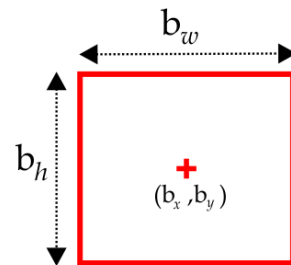


Figura 9: O Y corresponde ao vetor de informações correspondentes a bounding box, confiança(p_c), posição (b_x, b_y), dimensões (b_h e b_w) e a classe do objeto (c)

Fonte: Elaborada pelo autor e adaptada de Thatte (2020)

- p_c : corresponde a probabilidade da grade conter um objeto;

- b_x e b_y : são as coordenadas x e y , representando o centro da caixa delimitadora;
- b_h e b_w : altura e largura da caixa, respectivamente, em relação a célula da grade;
- c : representa a classe do objeto presente na imagem, que pode ser, por exemplo, carro, pessoa, gato ou cachorro. É possível ter quantas classes forem necessárias, de acordo com o problema (KARIMI, 2021).

Na maioria dos casos, um objeto em uma imagem pode ter várias opções de caixa na grade que podem fazer sua previsão, mesmo que nem todas tenham uma confiança que confirme a sua presença. Por esse motivo, utiliza-se a métrica IoU (Figura 10), responsável por descrever como as caixas delimitadoras estão sobrepostas, mantendo apenas as que envolvem de forma correta os objetos presentes na imagem (KEITA, 2022).

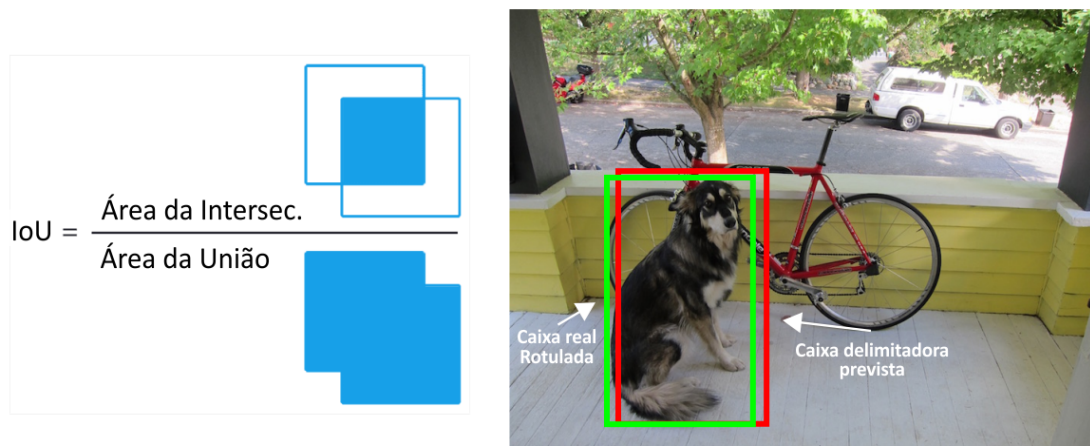


Figura 10: IoU - Intersecção sobre União, representada graficamente na fórmula e na imagem detectada. A caixa verde caracteriza a rotulação real, onde realmente o objeto está localizado, já a caixa vermelha caracteriza o que foi detectada pelo modelo. Nesse caso, o IoU apresentaria um valor próximo de 1, pois as duas caixas estão bem sobrepostas.

Fonte: Elaborada pelo autor e adaptada de Thatte (2020)

Na Figura 11 é mostrado todo processo de funcionamento da YOLO, onde a imagem é dividida em células da grade com dimensão $S \times S$ que, na imagem, seria 7×7 , e cada célula tem a sua previsão da caixa delimitadora e tem definida a sua pontuação de confiança, que na imagem, é definida pela espessura das bordas das caixas, quanto maior a espessura, maior a pontuação de confiança. A partir disso, na detecção final, a métrica IoU mantém apenas as caixas delimitadoras que são semelhantes às caixas reais (padrão-ouro) com os objetos de interesse, e assim, todos estarão com as suas respectivas *bindings boxes*.

Para a aplicação da YOLO, é necessário executar um treinamento da rede neural, e nesse processo é possível se deparar com alguns termos como *batch* ou Lote, *subdivisions* ou subdivisões e o mAP, onde lote se define como o número de imagens carregadas por iteração, e as subdivisões são também conhecidas como mini-lotes, onde, por exemplo, se o lote tiver o valor de 64 e a subdivisão, 8, significa que por 8 vezes, 8 imagens (64/8) serão enviadas para o processo de treinamento, até que o lote seja concluído, e assim, uma nova interação será iniciada com 64 novas imagens.

Com toda a sua capacidade, a YOLO é aplicada em várias áreas, como saúde, agricultura, segurança, entre outras, contribuindo assim, para o seu desenvolvimento. Porém, mesmo com muitas vantagens e aplicações, a YOLO também possui limitações, sendo uma delas a dificuldade para detectar pequenos objetos que aparecem em grupos, visto que, cada célula de grade é responsável por detectar apenas um objeto.

Além disso, existem várias versões da YOLO, onde cada uma delas oferece recursos diferentes, com o objetivo de melhorar suas características, corrigindo situações das versões passadas e/ou alterando sua arquitetura. O presente trabalho utilizou a versão 7 minúscula, do inglês, *tiny*. A YOLO v7 possui diversas melhorias, como a possibilidade de detectar um conjunto mais amplo de tamanhos e formatos de objetos, resultando em uma redução dos falsos positivos. Outra melhoria que é importante destacar é sua resolução que foi ampliada, em relação as versões anteriores, antes 416 pixels por 416 pixels, agora 608 pixels por 608 pixels, permitindo assim a detecção de objetos menores e também uma precisão mais alta. (KUNDU, 2023)

As versões possuem algumas variações de arquitetura, e a versão (v7-tiny) utilizada no trabalho consiste em um menor número de camadas, e também é mais rápida comparada a versão YOLO-v7. (KUKIL; RATH, 2022)

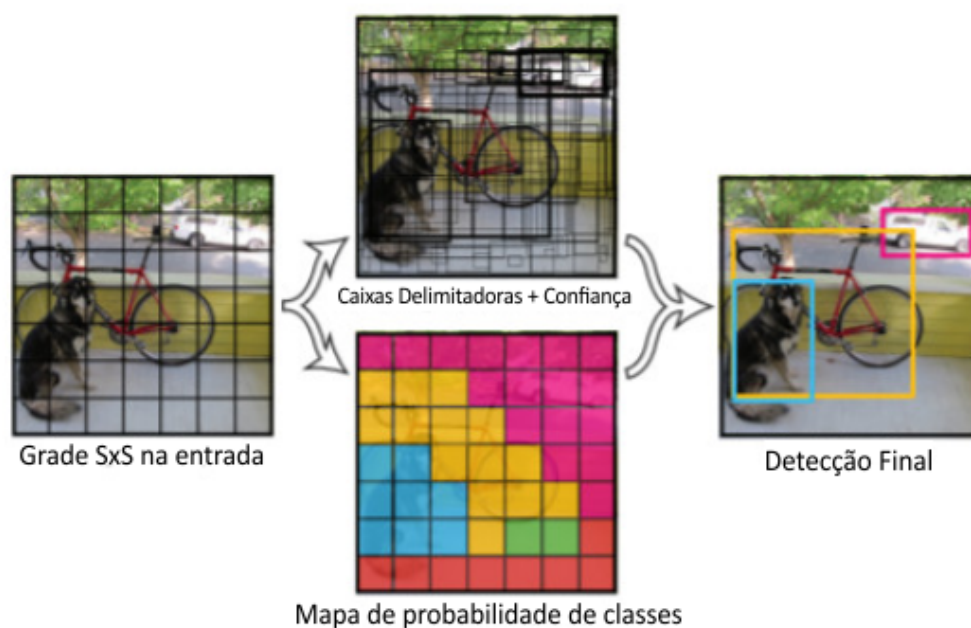


Figura 11: Funcionamento da YOLO - As cores do mapa de probabilidade representam que há possibilidades para a existência de classes nas determinadas regiões, podemos confirmar na detecção final, onde as caixas delimitadoras estão com bordas das respectivas cores.

Fonte: Thatte (2020)

3 TRABALHOS RELACIONADOS

Neste capítulo será apresentada uma revisão de publicações que são relevantes na literatura, considerando o tema apresentado no presente trabalho e a semelhança com seus objetivos: detecção do osso hioide com soluções em inteligência artificial e análise de deslocamento após a sua detecção.

Foram escolhidas publicações dos últimos dez anos (2012-2022). As bases de dados escolhidas para o levantamento da literatura foram a Pubmed, IEEE, Springer e Taylor. Palavras chave combinadas como: “Osso hioide”, “Deglutição”, “Deslocamento”, “Movimento”, “Posição”, “Hyoid Bone”, “Displacement”, “Movement”, “Swallowing”, “Artificial Intelligence”, “Deep Learning”, foram utilizadas com maior frequência.

O trabalho de Hossain et al. (2014) teve como objetivo propor um método semi-automático que rastreia o osso hioide e quantifica o seu deslocamento no movimento. Foi utilizado o exame de videofluoroscopia, com a deglutição de alimento não especificado junto ao sulfato de bário, que é responsável por tornar a comida visível nesse tipo de exame, e além da identificação do osso hioide, também identificava as vértebras cervicais, com o intuito de estabelecer uma relação de referência com o paciente. Como metodologia de estudo, foram utilizadas 5.000 amostras, com o objetivo de classificar imagens que possuem vértebras cervicais e imagens que não possuem vértebras cervicais, a primeira, amostras positivas, e a segunda, amostras negativas, respectivamente. Após essa classificação, foi possível identificar a região de interesse, que seriam as vértebras cervicais, e, através dos estudos videofluoroscópicos, o osso hioide, sempre localizado do lado esquerdo da região de interesse, as vértebras cervicais, era encontrado.

Com a localização das vértebras e, por consequência, do osso hioide, a parte avaliativa do seu movimento se inicia através de um sistema de referência ao paciente, derivado do alinhamento das vértebras cervicais, assumindo o sistema de coordenadas da imagem, onde o eixo X representa a direção horizontal em relação as mesmas, e eixo Y, referencia a direção vertical. Concluiu-se que o método apresentado identifica e rastreia com sucesso o osso hioide e pode fornecer *feedback* para os profissionais da área, porém, também foi exposto que se trata de um método semiautomático, ou seja, requer identificação manual do usuário das regiões de interesse, onde o osso hioide está localizado, e expõe que um método totalmente automático seria preferível, ou seja, sem necessidade de intervenção manual do profissional de saúde.

Já o trabalho de Hsiao et al. (2022) trata de um algoritmo que é baseado em CNN, o *Cascaded Pyramid Network* (CPN). O objetivo deste estudo é localizar e rastrear automaticamente o osso hioide, além disso, calcular várias grandezas físicas, como velocidade média e o deslocamento no eixo horizontal ou vertical. Para o treinamento, foi utilizado um conjunto de dados com 365 vídeos de deglutição obtidos através do exame de vide-

ofluoroscopia de 189 pacientes, onde deglutiram um alimento não identificado na forma de sulfato de bário fino, espesso e pastoso. Para teste, foram utilizados 44 vídeos de 44 pacientes com diferentes causas de disfagia. O CPN pode prever mapas de calor e coordenadas, de acordo com marcos anatômicos, realizando uma transformação no sistema de coordenadas aplicadas ao osso hioide para analisar a sua trajetória. Nesta etapa, foi escolhido um sistema de coordenadas que se alinha ao eixo da coluna cervical do paciente para garantir um cálculo preciso do deslocamento do osso. Dessa forma, foi concluído que o trabalho apresentou resultados com confiabilidade entre os avaliadores da detecção do osso hioide e o algoritmo e três profissionais que avaliavam manualmente a trajetória do osso. Essa confiabilidade foi encontrada no cálculo de deslocamento máximo e velocidade média, tanto horizontal, quanto verticalmente. Porém, também foi exposto que o estudo apresenta limitações, como a falta de uma análise cinemática detalhada da excursão do osso hioide, ou seja, incluir velocidade instantânea, aceleração e padrão de trajetória.

O trabalho de Lee et al. (2020) propôs um algoritmo de rastreamento do osso hioide, o *Multi-Domain Networks* (MDNet), baseado em CNN, além de uma U-Net para segmentação do osso e das vértebras cervicais. Foram utilizados dados de exames de videofluoroscopia para o estudo, adquiridos de um repositório anônimo aproveitado em outros estudos clínicos. Os dados são de 77 indivíduos, onde 26 são de pacientes saudáveis, 32 de pacientes com doença de Parkinson e 19 de pacientes com acidente vascular cerebral (AVC), com bolus líquido administrado de 2mL de solução de bário diluída. O estudo explica que para estimar a trajetória do osso hioide é necessária a sua segmentação e das vértebras cervicais. Obtendo a localização de c2 e c4 (vértebras específicas, que atuam como referência para posição do osso hioide), é possível aplicar a transformação dos eixos e obter e calibrar a trajetória do osso hioide. Foi concluído que o algoritmo de segmentação teve um alto desempenho em sua tarefa, junto com o algoritmo de rastreamento do osso, possibilitando trajetórias precisas e localizando coordenadas exatas. O estudo também evidencia a comparação entre os experimentos com o algoritmo, em relação a “verdade absoluta” sobre a trajetória e foi visto uma variação ligeiramente para cima ou abaixo.

Desenvolver um modelo de DL que possa detectar o osso hioide e o seu movimento com precisão sem intervenção humana é o objetivo do trabalho de Kim et al. (2021), onde foram utilizados os dados de exames de videofluoroscopia (197 arquivos de vídeo), com pacientes deglutindo 3ml de líquido espesso, uma espécie de mingau de arroz, iogurte e sulfato de bário dissolvido em água. Três médicos especialistas no estudo videofluoroscópico da deglutição participaram do processo de rotulagem do osso hioide (alvo de rastreamento), da coluna cervical, e de uma moeda fixada ao pescoço, como base para correção da profundidade do osso. Todo processo de rastreamento do osso hioide proposto pelo estudo é composto pela detecção dos objetos de interesse, no caso, o osso, a coluna cervical e a moeda. Ou seja, o movimento do osso hioide é detectado com base na coluna

cervical, e para executar a detecção foi utilizada a U-Net. No entanto, observou-se que era preciso uma rede capaz de segmentar objetos de vários tamanhos, então propuseram a utilização da rede de pirâmide bidirecional, do inglês, Bidirectional Feature Pyramid Network (BiFPN). Pode-se concluir que o método proposto teve um bom desempenho de detecção, junto à distância estimada do movimento do osso hioide, que mostrou um valor próximo às medidas reais. Também foi explicado que a diferença entre condições normais e de pacientes com disfagia não foi examinada, visto que estudos são necessários nesse sentido para aumentar a confiabilidade do sistema em relação a disfagia.

Utilizando exame de ultrassonografia, o trabalho de Feng et al. (2021) tem como proposta a aplicação de algoritmos de rastreamento baseados em DL e filtros de correlação para a detecção da posição do osso hioide durante a deglutição em vídeos de ultrassom. O conjunto de dados apresentado foi construído a partir do exame de 10 pacientes adultos jovens e saudáveis, que realizaram cinco tipos de deglutições com líquidos pastoso e ralo e saliva. A rotulagem da localização do osso hioide foi feita por profissionais treinados. Redes Siamesas Totalmente Convolucionais (do inglês, *Fully-Convolutional Siamese Networks* - SiamFC), Rastreo preciso por maximização de sobreposição (do inglês, *Accurate Tracking by Overlap Maximization* - ATOM) e Preditor de modelo descritivo (do inglês, *Discriminative Model Prediction* - DiMP) foram utilizados, junto aos filtros de correlação, que fizeram a previsão das caixas delimitadoras em cada quadro para indicar a localização do osso hioide. Concluiu-se que as SiamFC tiveram um desempenho quantitativo e qualitativo de rastreamento superior, mas foi explicado que o modelo apresentou erros quando a velocidade do osso hioide era maior.

Após descrever os trabalhos relacionados, é possível destacar que o exame de videofluoroscopia prevaleceu, sendo utilizado em quatro dos cinco apresentados. Outro destaque é o fato da cervical fazer parte do estudo, como referência para movimento do osso hioide, como podemos observar nos trabalhos de Hossain et al. (2014), Lee et al. (2020), Kim et al. (2021). Em termos de bolus ingerido, o trabalho de Feng et al. (2021) se assemelha ao presente estudo, onde os exames presentes no conjunto de dados foram obtidos durante a ingestão de líquido pastoso, ralo e saliva, além disso, sugere como futuro potencial o treinamento com YOLO, método utilizado no presente trabalho. Na Tabela 1 podemos ver de forma resumida as características dos trabalhos encontrados na literatura comparados ao do presente estudo.

Tabela 1: Tabela de comparação entre os trabalhos relacionados e o presente estudo

Autor	Método	Exame	Objetivo
Hossain et al. (2014)	Segmentação	Videofluoroscopia	Segmentação e medição do movimento do osso hioide.
Hsiao et al. (2022)	<i>Cascaded Pyramid Network</i> (CPN)	Videofluoroscopia	Análise cinemática automática completa da excursão do osso hioide.
Lee et al. (2020)	<i>Multi-Domain Networks</i> (MDNet)	Videofluoroscopia	Rastrear automaticamente o osso hioide quadro a quadro.
Kim et al. (2021)	BiFPN network	Videofluoroscopia	Detectar o osso hioide e rastrear seu movimento.
Feng et al. (2021)	Algoritmos de rastreamento baseados em deep learning e filtros de correlação	Ultrassonografia	Detectar o osso hioide.
Presente Estudo	YOLO	Ultrassonografia	Deteccão e análise quantitativa do movimento do osso hioide.

4 METODOLOGIA

Neste capítulo serão descritos todos os métodos e dados utilizados no desenvolvimento do presente estudo, como também o ambiente de execução e as ferramentas adotadas.

4.1 Base de dados

A primeira etapa para a realização do trabalho foi toda captação da base de dados que seria utilizada para o treinamento da rede neural convolucional e YOLO. Tal base de dados é oriunda do estudo de Lopes (2019), onde 19 pacientes adultos foram submetidos ao exame de ultrassonografia para estudo da deglutição, realizado por um grupo de pesquisa da área de Fonoaudiologia da Universidade Federal da Paraíba (UFPB).

Todo estudo foi dividido em 4 situações de deglutição, resultando em 10 vídeos por paciente, 1 vídeo tratava do paciente em repouso, ou seja, sem realizar nenhum movimento de deglutição. Já no restante dos 9 vídeos, acontecia a deglutição, em 3 grupos de 3 vídeos, o primeiro grupo com o paciente realizando uma deglutição com saliva, o segundo grupo com o paciente ingerindo um líquido de consistência pastosa, e por último, o grupo de vídeos onde foi realizada a deglutição com água.

Porém, o presente estudo focou apenas nos grupos de vídeo que apresentaram deglutição, ou seja, nos grupos onde o paciente fez a deglutição com água, saliva e um líquido pastoso, visto que é o momento onde o osso hioide realiza seu movimento, que este será detectado. O estudo utilizou, no total, 171 vídeos (57 vídeos de cada tipo de deglutição). O cálculo do deslocamento do osso, uma vez detectado, será explicado nas próximas seções.

A partir dos vídeos de ultrassom, foram separadas 590 imagens, extraídas dos vídeos já mencionados e divididas de maneira aleatória em 422 imagens para treinamento, 85 para teste e 83 para validação. As imagens utilizadas no treinamento e na validação possuíam o osso hioide rotulado por profissionais da área de fonoaudiologia, através do software *LabelImg*¹, onde após a rotulação é gerado o arquivo que contém as coordenadas do osso na imagem de ultrassom, no seguinte formato:

[classe do objeto, coordenada x, coordenada y, largura w, altura h]

- Classe do objeto: refere-se à presença do osso hioide;
- Coordenadas x e y : posição do objeto no eixo x e y na imagem;

¹Disponível em: <https://github.com/heartexlabs/labelImg>

- Largura w e altura h : largura e altura da caixa delimitadora onde o objeto encontra-se detectado

Além dos 171 vídeos, o trabalho dispõe também de mais 62 vídeos onde os pacientes realizavam as deglutições com os três tipos de bolus já mencionados. Tais vídeos foram utilizados para fazer o cálculo do deslocamento, visto que são diferentes da base de dados citada anteriormente, importante para um teste mais amplo, buscando uma precisão de detecção e cálculo, pois são vídeos que não foram visto pelo modelo no treinamento.

Os 62 vídeos utilizados para detecção e cálculo não estavam distribuídos igualmente entre os tipos de deglutição, por isso, com a finalidade de promover essa igualdade, foram escolhidos quinze vídeos aleatórios de cada tipo de deglutição.

4.2 Ambiente de Execução

Foram utilizadas duas configurações para execução do treinamento, validação e teste. Para o treinamento, o presente trabalho utilizou o Google Colab com uma GPU NVIDIA Tesla T4 16GB, e para teste, uma GPU NVIDIA GTX 1050 TI 4GB, CPU Intel(R) Core (TM) i7-7700HQ CPU@2.80GHz e um sistema operacional Linux Ubuntu v20.04.

4.3 YOLO e Treinamento

Para o treinamento foi utilizado o repositório darknet do AlexeyAB (ALEXEYAB, 2021), e, em relação a YOLO, foi escolhida sua versão 7. O “batch” ou lotes, e as “subdivision” ou subdivisões, foram alterados para 64 e 8, respectivamente, o que significa que foram 64 imagens por iterações, e 8 imagens por mini-lotes (64/8).

Como ponto de partida foram definidas 50.000 iterações para análise, e, o critério de parada utilizado foi a função de perda, ou seja, o treinamento era interrompido assim que esta se estabilizasse. Junto a isso, a métrica mAP foi responsável por medir a média da precisão média da detecção. Cerca de 15.000 iterações foram necessárias até a interrupção.

4.4 Métrica de Avaliação dos Resultados

É importante destacar que os resultados que serão apresentados na próxima seção, estarão utilizando apenas um limiar de IoU de 0,5, ou seja, são consideradas detecções corretas àquelas que possuem um valor de IoU acima de 0,5. Todas detecções abaixo ou igual esse valor, serão consideradas incorretas.

Essa relação auxilia no entendimento dos verdadeiros positivos (VP), falsos positivos (FP) e falsos negativos (FN). Uma detecção correta ($\text{IoU}_i \geq 0,5$) é considerada um verdadeiro positivo, já as incorretas ($\text{IoU}_i = 0,5$) e/ou se o objeto (osso hioide) for detectado mais de uma vez são considerados falsos positivos. Já os falsos negativos são considerados como tal, quando o objeto não for detectado na imagem/vídeo ou caso tenha um valor de $\text{IoU}_i = 0,5$.

4.5 Cálculo de Deslocamento

Nesta seção será apresentada a proposta de cálculo do deslocamento do osso hioide, e, para obter êxito na execução da proposta, foram utilizadas funcionalidades extras do repositório do AlexeyAB², que permite a extração de um arquivo no formato .txt, com as coordenadas da caixa delimitadora, ao longo do vídeo.

Antes de apresentar a proposta de cálculo do deslocamento do osso hioide em movimento, algumas questões devem ser explicadas para melhor entendimento. A primeira é em relação ao exame de ultrassonografia, onde é importante destacar duas situações:

- Posição do Exame: A ultrassonografia da deglutição possui a aparência original (Figura 12) (1), porém, esta posição representa o paciente de perfil e deitado. Como a finalidade da proposta de cálculo é o deslocamento vertical do osso hioide (do repouso ao pico máximo), foi necessário rotacionar a imagem (apenas para realizar o cálculo), invertendo assim os eixos x e y . Dessa forma, garante-se o curso legítimo do osso, acompanhando seu deslocamento no sentido vertical, representado pelo paciente de perfil e sentado (Figura 12) (2).

²Disponível em: <https://github.com/AlexeyAB/darknet>

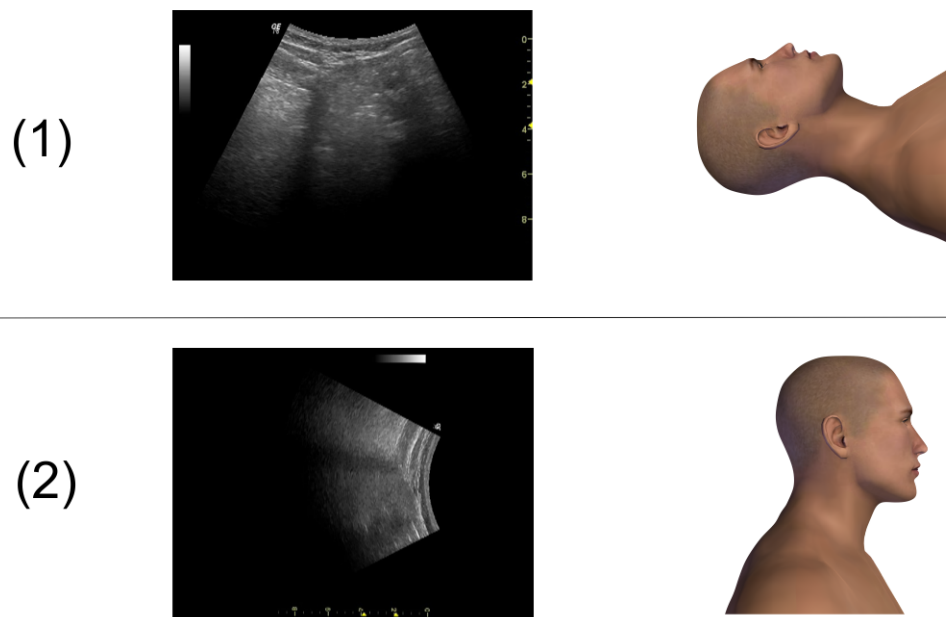


Figura 12: Posições do Exame de Ultrassom: (1) Posição Original do Exame, (2) Imagem rotacionada do exame para melhor entendimento e realização do cálculo do deslocamento

Fonte: Autoria Própria

- Quantidade de pixels/centímetros: Uma informação importante para o cálculo, em centímetros ou milímetros, do deslocamento do osso hioide em movimento, é proveniente do exame de ultrassonografia e trata da quantidade de pixels por essas unidades de distância na imagem obtida do exame. É possível visualizar uma espécie de régua no seu canto direito (Figura 13), que retrata a quantidade de centímetros na imagem. A partir disso, foi possível estimar a quantidade de pixels existentes em 1 centímetro, nesse caso, 40 pixels.

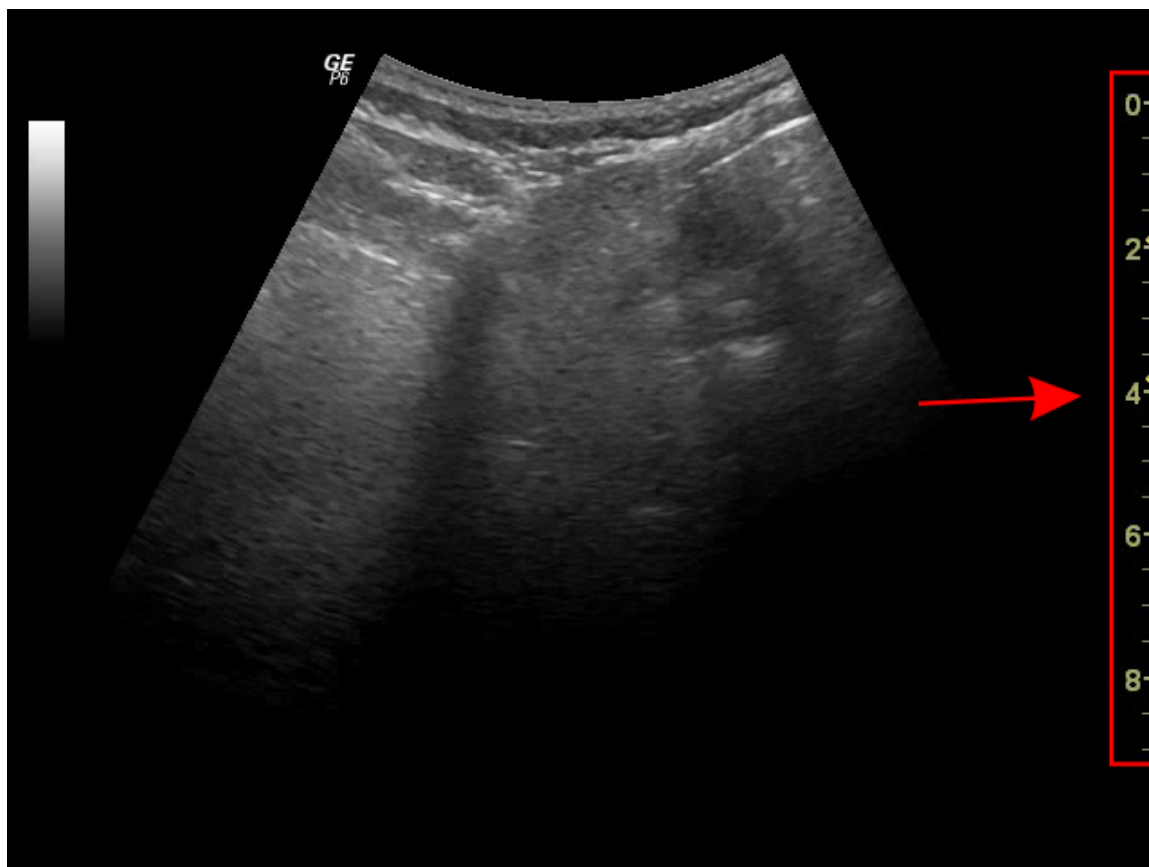


Figura 13: Exame de Ultrassonografia utilizado no presente estudo

Fonte: Lopes (2019)

O segundo ponto importante que deve ser destacado trata do deslocamento da *bounding box*. Em relação ao seu deslocamento no eixo X , assim que a caixa delimitadora se desloca para direita, seu valor em X aumenta, e ao contrário, diminui. Já em relação ao eixo Y , foi notado um comportamento diferente, quando a *bounding box* desloca-se verticalmente, seu valor em Y diminui, e ao contrário, aumenta.

Com os detalhes necessários para seu entendimento explicitados, a proposta do cálculo do deslocamento será mostrada a seguir.

4.6 Cálculo do deslocamento vertical do osso hioide

O cálculo do deslocamento vertical do osso hioide se dá a partir dos valores obtidos através da coordenada X , porém, assumimos esta como Y , após a rotação da imagem, como explicado anteriormente. Essa coordenada corresponde a elevação do osso hioide na detecção, do repouso ao pico máximo do deslocamento vertical. Após essa identificação, é feita uma subtração entre o valor da posição inicial e a posição do pico máximo da caixa delimitadora, ou seja, a posição mais elevada que ela alcançou, representando assim o repouso e o pico máximo do deslocamento do osso hioide, obtendo, a partir disso, o valor do deslocamento em pixels.

É realizada, a partir do valor obtido anteriormente, sua conversão de pixels para milímetros, visto que os valores em pixel não são usuais para profissionais da área de saúde, e a título de comparação com a literatura. A conversão foi realizada a partir das dimensões originais do exame, no caso, 640 de largura x 480 de altura e 96dpi e também da quantidade pixels/centímetros estimados a partir do mesmo, como foi explicado anteriormente. Fazendo um ajuste com os parâmetros do caso deste estudo, foi obtida a equivalência de 1 pixel para 0,250 mm. Após a conversão, podemos obter o valor em milímetros do deslocamento vertical, ou amplitude vertical máxima do osso hioide.

Apresenta-se a seguir a sequência de passos necessários para o cálculo do deslocamento do osso hioide, que pode ser observada na Figura 14.

- Passo 1 - Identificação dos valores do deslocamento da *bounding box* no eixo *X*, porém, assumido como *Y* após a rotação da imagem, do repouso ao ponto máximo, onde assumimos que a *bounding box* é o osso hioide.
- Passo 2 - Subtração dos valores de repouso (posição inicial) e pico máximo (posição mais elevada da *bounding box*), para assim obter o valor do deslocamento em pixel.
- Passo 3 - Conversão de pixel (px) em milímetro (mm) dos valores obtidos no Passo 2. A taxa de conversão é obtida através das dimensões originais do exame e também da quantidade pixels/centímetros estimados a partir do mesmo.
- Passo 4 - Após a conversão, resulta-se no valor final do deslocamento vertical, ou amplitude vertical máxima do osso hioide, em milímetros.

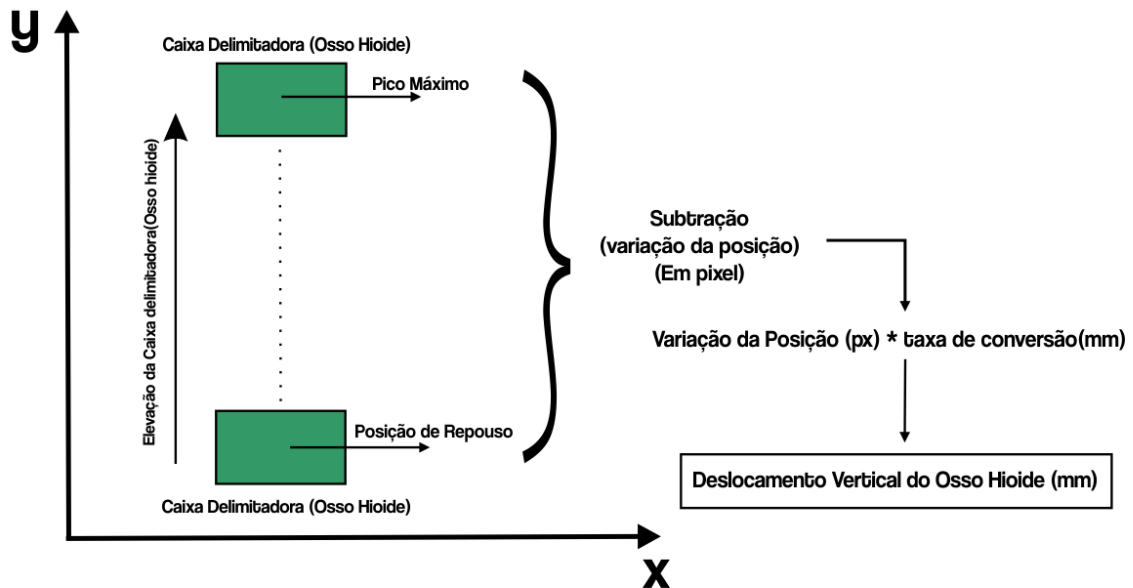


Figura 14: Ilustração da Proposta de Cálculo do deslocamento vertical do osso hioide (do seu repouso ao pico máximo)

Fonte: Própria

Na próxima seção será apresentado o ambiente responsável por fornecer a visualização da detecção e do cálculo do deslocamento do osso hioide.

4.7 Ambiente de Visualização

Para facilitar a interação entre os profissionais da saúde e o sistema aqui proposto para a detecção e análise do movimento do osso hioide, foi criado um ambiente de visualização, em forma de uma página web, onde os profissionais poderão fazer o *upload* do vídeo gerado no exame de ultrassonografia e obter a localização do osso detectado ao longo das imagens e com o seu deslocamento calculado.

A construção do ambiente se inicia no front-end com o uso do ReactJS, que é uma biblioteca *JavaScript* (JS), participante da camada View, do modelo MVC (*Model View Controller*). Para o complemento do fluxo da ferramenta, foi desenvolvida uma API (*Application Programming Interface*), ou Interface de Programação de Aplicação, para manipular os vídeos e encaminhá-los para a Darknet, junto com a YOLO, para a detecção do osso hioide. Foi escolhido o *NodeJS* para essa construção, que se trata de uma plataforma também *JS*, que trabalha do lado do servidor, e promove o desenvolvimento de aplicativos escalonáveis e de baixa latência (LOW et al., 2014).

O Fluxo da aplicação se dá da seguinte forma: o *upload* do vídeo é feito e enviado ao servidor, onde está a Darknet que, junto com a YOLO, farão a detecção, e também

o cálculo do deslocamento, realizado por um script em python. O vídeo rotulado (com a *bounding boxes* delimitando o osso hioide detectado), junto ao valor do deslocamento, serão devolvidos como resultado, promovendo assim facilidade, agilidade e interação com o profissional da saúde que fará toda interação com a ferramenta (Figura 15).

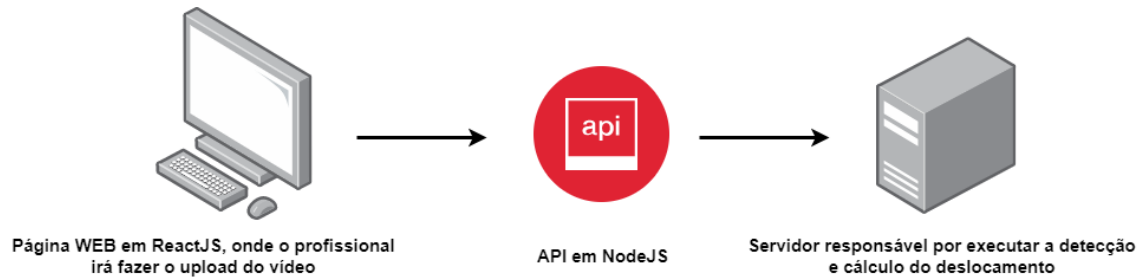


Figura 15: Fluxo da Página WEB

Fonte: Autoria Própria

5 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS

Os resultados estão divididos em três partes, onde a primeira representa a detecção do osso hioide no exame de ultrassom nas deglutições, e as duas últimas correspondem a uma proposta de cálculo do seu deslocamento vertical durante a realização do movimento.

A seguir serão mostradas imagens que retratam a detecção no exame de ultrassom e logo após, será apresentado quantitativamente o deslocamento vertical do osso hioide em tabelas, divididas nas três situações de deglutição e nas duas propostas de cálculos descritas anteriormente.

Para o conhecimento do resultado da detecção, serão mostrados alguns *frames* da detecção nas três situações de deglutições: paciente deglutindo a própria saliva (Figura 16), líquido de consistência pastosa (Figura 17) e água (Figura 18).

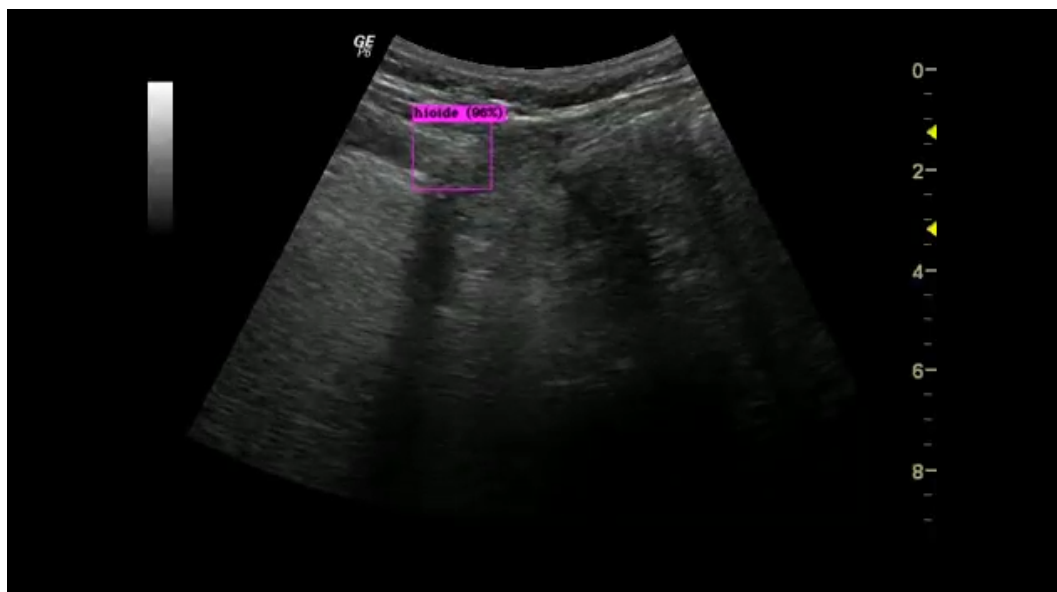


Figura 16: Detecção do osso hioide com o paciente deglutindo saliva

Fonte: Autoria Própria

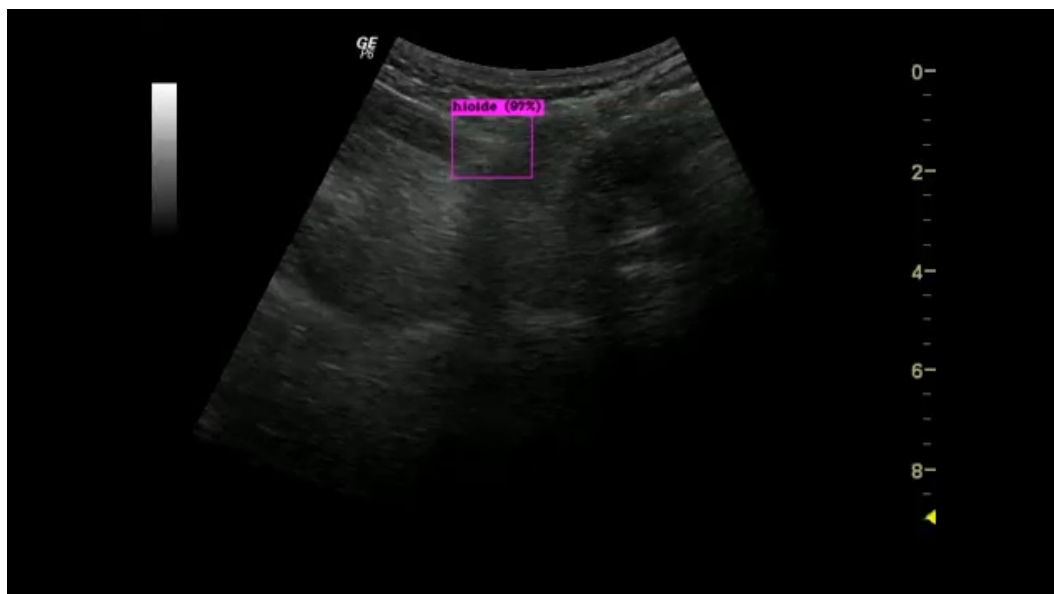


Figura 17: Detecção do osso hioide com o paciente deglutindo consistência pastosa

Fonte: Autoria Própria

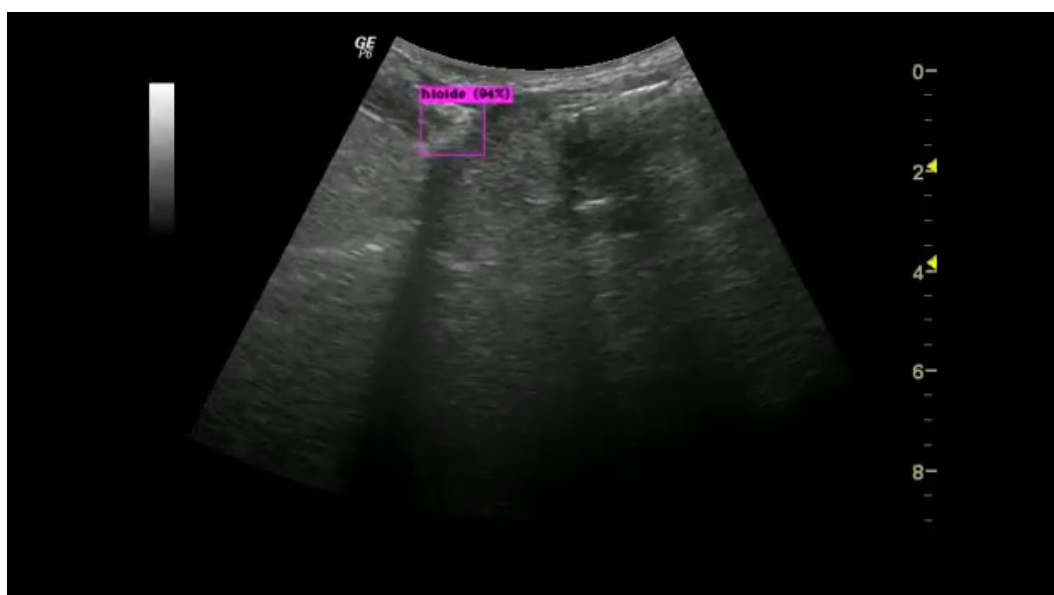


Figura 18: Detecção do osso hioide com o paciente deglutindo água

Fonte: Autoria Própria

As detecções apresentadas mostraram o bom desempenho do método proposto, como indicam os resultados a seguir, utilizando um limiar de IoU de 0,5. Como valor médio o modelo apresentou uma média de 0,778 ou 77,80%, ou seja, o mesmo apresentou uma boa sobreposição. A Figura 19 representa essa média obtida.

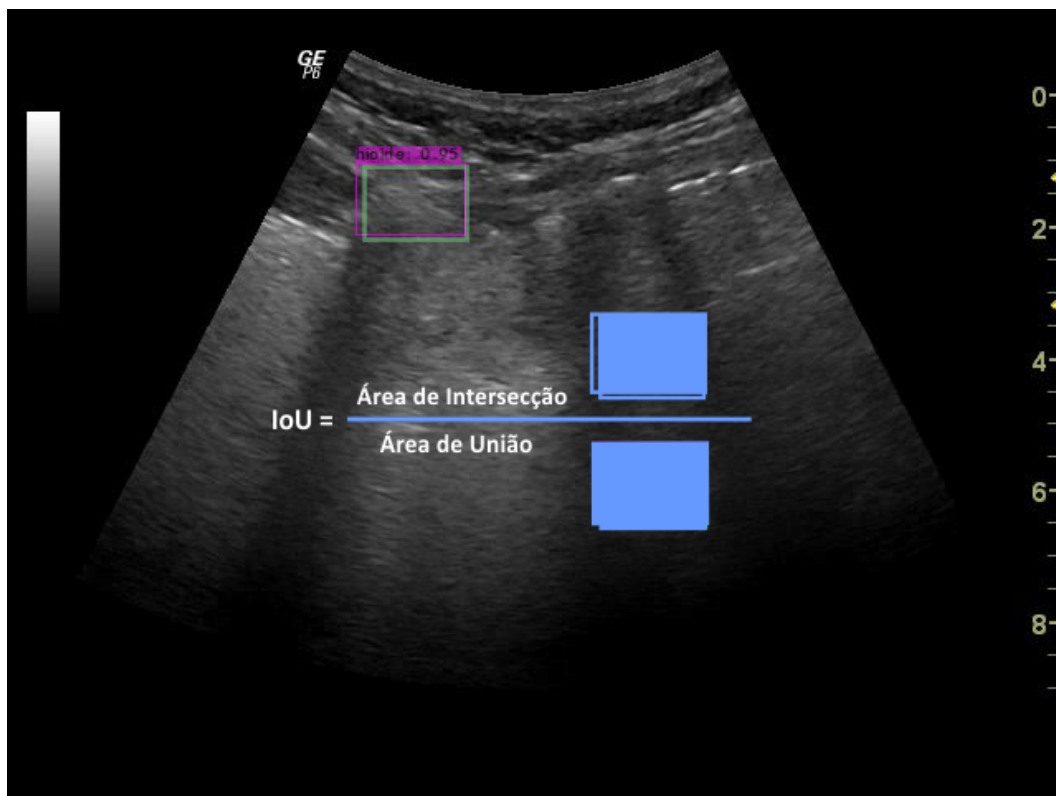


Figura 19: IoU - Osso Hioide - A caixa delimitadora da cor verde representa a rotulação real do osso, já a de cor rosa, a detecção realizada pelo modelo. O IoU do exemplo é retratado na expressão.

Fonte: Autoria Própria

Dentre as imagens em que o modelo detectou o osso hioide, obteve 95% (precisão). Por outro lado, dentre aqueles que realmente são osso hioide, a taxa de acerto do modelo foi de 93% (revocação). Ou seja, quando realmente existia um osso hioide no vídeo, o modelo acertou em 93% dos casos.

5.1 Erros encontrados

Nas detecções, alguns frames apresentaram erros bem perceptíveis (Figura 20), mais de uma bounding box estavam sendo detectadas, podendo assim, contribuir para erros no cálculo de deslocamento. Um novo treinamento, ou apenas uma alteração no limiar de confiança pode ser a solução.

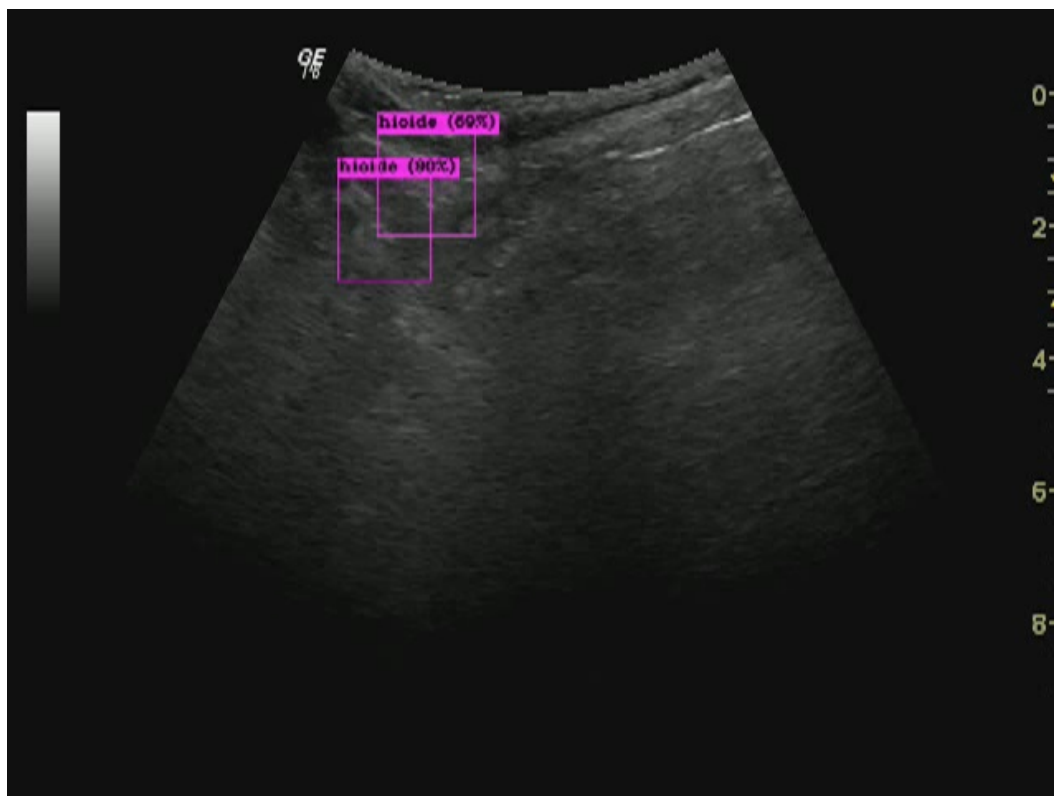


Figura 20: Erro na detecção, no caso, duas caixas delimitadores estão sendo detectadas.

Fonte: Autoria Própria

5.2 Apresentação quantitativa do deslocamento vertical do osso hioide em sua amplitude máxima

A seguir serão apresentadas as Tabelas 2 (deglutição com saliva), 3 (deglutição com líquido pastoso) e 4 (deglutição com água), com os valores do deslocamento vertical do osso hioide, a partir da proposta do cálculo explicado anteriormente.

Tabela 2: Valores do deslocamento vertical com pacientes deglutindo Saliva

Vídeo	Deslocamento Vertical (mm)
Vídeo 1	12,00 mm
Vídeo 2	9,75 mm
Vídeo 3	15,25 mm
Vídeo 4	11,75 mm
Vídeo 5	13,50 mm
Vídeo 6	10,75 mm
Vídeo 7	5,50 mm
Vídeo 8	11,00 mm
Vídeo 9	9,25 mm
Vídeo 10	7,50 mm
Vídeo 11	6,75 mm
Vídeo 12	24,25 mm
Vídeo 13	15,50 mm
Vídeo 14	30,00 mm
Vídeo 15	11,50 mm
Média $\pm$ desvio padrão	12,95 $\pm$ 6,49

Tabela 3: Valores do deslocamento vertical com pacientes deglutindo uma consistência Pastosa

Vídeo	Deslocamento Vertical (mm)
Vídeo 1	15,25 mm
Vídeo 2	6,50 mm
Vídeo 3	12,00 mm
Vídeo 4	11,75 mm
Vídeo 5	19,50 mm
Vídeo 6	13,00 mm
Vídeo 7	9,50 mm
Vídeo 8	12,25 mm
Vídeo 9	9,75 mm
Vídeo 10	14,00 mm
Vídeo 11	18,75 mm
Vídeo 12	8,25 mm
Vídeo 13	11,75 mm
Vídeo 14	12,50 mm
Vídeo 15	8,50 mm
Média $\pm$ desvio padrão	12,22 $\pm$ 3,63

Tabela 4: Valores do deslocamento vertical com pacientes deglutindo Água

Vídeo	Deslocamento Vertical (mm)
Vídeo 1	6,25 mm
Vídeo 2	8,75 mm
Vídeo 3	8,50 mm
Vídeo 4	26,75 mm
Vídeo 5	12,00 mm
Vídeo 6	10,75 mm
Vídeo 7	11,25 mm
Vídeo 8	13,50 mm
Vídeo 9	11,00 mm
Vídeo 10	11,50 mm
Vídeo 11	0,00 mm
Vídeo 12	5,75 mm
Vídeo 13	12,50 mm
Vídeo 14	11,25 mm
Vídeo 15	5,25 mm
Média $\pm$ desvio padrão	10,33 $\pm$ 5,77

Após o cálculo do deslocamento vertical do osso hioide, foi realizada uma comparação com alguns estudos existentes na literatura, visto que, não há um padrão-ouro.

Hsiao et al. (2017), assim como neste trabalho, utilizou ultrassonografia em sua análise e obteve o resultado médio de 22,10 mm no deslocamento vertical na deglutição de água e 20,50 mm de deslocamento na deglutição de líquido pastoso, sendo os valores de deslocamento encontrados na literatura que mais se distanciam do resultado do presente estudo. Porém, Hsiao et al. (2017) não especificam se o deslocamento do osso hioide foi total (ida e volta) ou a amplitude máxima, utilizado no presente estudo, onde é considerada apenas a ida, ou seja, do repouso ao pico.

Yamazaki et al. (2017) utilizou o exame de videofluoroscopia, e calculou em sua pesquisa apenas o valor de deslocamento vertical na deglutição com líquido, no caso, com sulfato de bário diluído, obtendo um valor médio de deslocamento vertical de aproximadamente 10,88 mm com uma variação de 1,47 mm, para mais ou para menos. Esse valor aproxima-se da proposta do presente estudo, onde, obteve-se uma média de deslocamento vertical (amplitude máxima) de 10,33 mm na deglutição com água.

No trabalho desenvolvido por Chi-Fishman e Sonies (2002), assim como Hsiao et al. (2017), utilizou-se como exame a ultrassonografia, e foi obtida informação apenas do líquido pastoso, com um valor médio de deslocamento vertical de aproximadamente 8,50 mm, com uma variação de 3,37 mm para mais ou para menos. Dessa forma, observou-se uma aproximação (caso sua variação positiva seja aplicada) com a proposta do presente estudo, com uma média de deslocamento vertical (amplitude máxima) de 12,21 mm na

deglutição com líquido pastoso. Chi-Fishman e Sonies (2002) explicita que seu resultado equivale a amplitude máxima, diferente de Hsiao et al. (2017) e Yamazaki et al. (2017).

Na Tabela 5 podemos observar, de maneira mais clara, a diferença entre a proposta desenvolvida nesse trabalho e os trabalhos encontrados na literatura. Da terceira coluna em diante, são expostos os valores do deslocamento vertical do osso hioide em cada situação de deglutição (Saliva, Líquido Pastoso e Líquido, respectivamente). É importante destacar que, na última coluna foi utilizado apenas o termo "Líquido" para generalizar a situação de deglutição, visto que, a proposta de Yamazaki et al. (2017) utiliza líquido como sulfato de bário diluído, ao invés de água (líquido utilizado no presente estudo).

Tabela 5: Tabela de Comparações com a Literatura (mm)

Autor	Saliva	Líquido Pastoso	Líquido
Hsiao et al. (2017)	x	20,50	22,10
Yamazaki et al. (2017)	x	x	$10,88 \pm 1,47$
Chi-Fishman e Sonies (2002)	x	$8,50 \pm 3,37$	x
Este Estudo	$12,95 \pm 6,49$	$12,22 \pm 3,63$	$10,33 \pm 5,77$

Em relação aos valores individuais de alguns vídeos, podemos destacar dois, o Vídeo 14 da Tabela 2 e o Vídeo 11 da Tabela 4. O primeiro trata-se de uma situação em que o valor de deslocamento foi mais elevado do que o restante da tabela, nesse caso, 30mm de deslocamento. A explicação para essa diferença, se dá pelo fato de que há um *frame* que possui duas detecções, uma delas, com confiança mais baixa, considerada errada, pois a caixa delimitadora encontrava-se em uma posição que não condiz com o osso hioide, afetando assim, o cálculo de deslocamento.

Já no segundo caso, o valor de deslocamento foi 0,00mm, e o motivo se dá pelo próprio vídeo do exame, por um erro de detecção, teve como início o pico de deslocamento do osso hioide. A partir disso, o cálculo partiu do pressuposto que o valor de repouso foi igual ao valor de pico, gerando um resultado com valor 0 para a subtração utilizada na variação da posição.

5.3 Ambiente de Visualização

Nas Figuras 21 e 22 podemos observar as páginas principais do desenvolvimento do ambiente de visualização do deslocamento do osso hioide em sua amplitude máxima.

A página inicial (Figura 21) é composta por um campo de *upload*, onde o usuário vai carregar o vídeo do exame de ultrassonografia da deglutição e, logo após, clicar no botão "Iniciar detecção" para inicializar o processamento de detecção e cálculo de deslocamento vertical.



Figura 21: Página inicial do ambiente de visualização

Fonte: Própria

Na página de resultado (Figura 22) apresentam-se as informações de detecção e cálculo. Do lado esquerdo da página é disponibilizado o vídeo com osso hioide detectado no exame de ultrassonografia da deglutição que foi carregado pelo usuário na página inicial (Figura 21). Do lado direito, é possível obter a informação do deslocamento vertical em sua amplitude máxima (mm). Logo abaixo, a ferramenta disponibiliza a opção de realizar o download do vídeo com o osso detectado, e também de realizar uma nova detecção com um vídeo diferente.

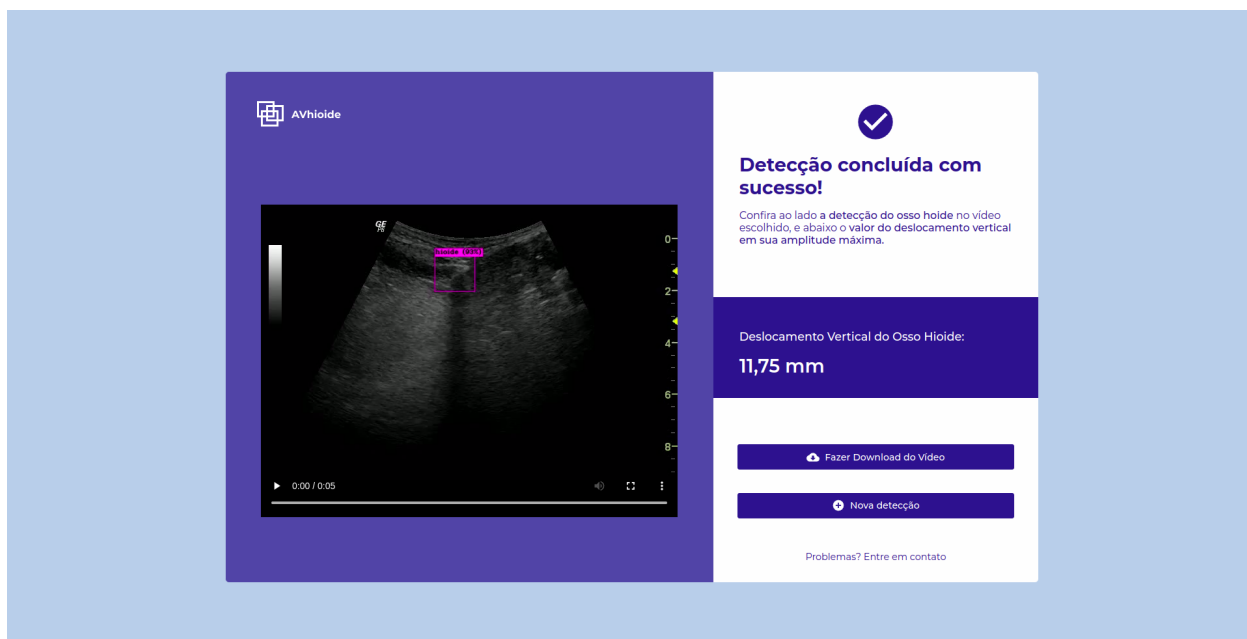


Figura 22: Página de resultado do ambiente de visualização

Fonte: Autoria Própria

6 CONSIDERAÇÕES FINAIS

Neste trabalho foi apresentada uma proposta de detecção automática do osso hioide em exames de ultrassonografia, utilizando IA, além do cálculo do seu deslocamento vertical durante o movimento. O osso hioide está localizado na parte anterior do pescoço e está presente na ação da deglutição, que é responsável pela absorção de nutrientes para o organismo e também participa da fala e respiração, tornando o seu estudo relevante na área de saúde.

O movimento vertical investigado no presente trabalho, foca na sua amplitude máxima, ou seja, na ida do seu repouso ao pico, utilizando informações da detecção do osso com um modelo de *deep learning*, mais precisamente a YOLOv7. Além disso, foi desenvolvido um ambiente de visualização com uma interface de fácil uso, voltado, principalmente, para profissionais da área. É importante destacar que os resultados gerados utilizam como base exames de ultrassonografia da deglutição com dimensões de 640px de largura x 480px de altura. Visto que profissionais da saúde executam esse processo de forma manual, esse sistema surge como uma ferramenta de auxílio para o profissional, automatizando a detecção do osso hioide e o cálculo de seu deslocamento no movimento, podendo tornar o processo mais rápido e efetivo.

Em relação a detecção, o modelo obteve resultados considerados satisfatórios, onde foram utilizados 3 tipos de deglutições (saliva, líquido pastoso e água) no treinamento, teste e validação. De acordo com 83 imagens utilizadas para validação, o mesmo apresentou uma precisão de 95%, revocação de 93%, com IoU de 77,87 %.

Em relação a proposta de cálculo, o início é dado a partir da detecção do osso hioide no vídeo, que é resultante do exame de ultrassonografia da deglutição. É possível ter acesso a posição do centro da caixa delimitadora que demarca a detecção do osso, onde os valores da posição são utilizados para a primeira parte do cálculo, que trata da variação da posição do osso hioide em pixels (repouso - pico). A segunda parte do cálculo é responsável por uma taxa de conversão resultante de dados presentes no exame.

De acordo com dados encontrados na literatura, a proposta de cálculo se aproxima de dois trabalhos, nas deglutições de líquido pastoso e de água. Não foram encontrados trabalhos que apresentassem valores de deslocamento vertical com saliva, em termos de distância, grandeza utilizada no presente estudo. Não foi levado em consideração: idade, sexo e quantidade de bolus diluído, apenas o que foi deglutido e o estado de saúde dos pacientes, que estavam saudáveis.

Com isso, os resultados mostraram que a pesquisa realizada pode contribuir para a detecção e cálculo do deslocamento vertical do osso hioide em exames de ultrassonografia, observando uma boa precisão na detecção e uma aproximação dos valores médios de

deslocamento da literatura.

Como trabalhos futuros, uma base com mais dados poderia ser utilizada para o treinamento e ampliação dos testes, além disso, separá-los por idade, sexo e o que foi deglutido, para assim, explorar mais resultados e diferenças no deslocamento com esses parâmetros.

REFERÊNCIAS

- ALAKE, R. *Deep Learning: GoogLeNet Explained — towardsdatascience.com*. 2020. [⟨https://towardsdatascience.com/deep-learning-googlenet-explained-de8861c82765⟩](https://towardsdatascience.com/deep-learning-googlenet-explained-de8861c82765). [Accessed 28-Feb-2023].
- ALEXEYAB. *Alexeyab/Darknet: Yolov4 / scaled-yolov4 / yolo - neural networks for object detection (windows and linux version of darknet)*. 2021. Disponível em: [⟨https://github.com/AlexeyAB/darknet⟩](https://github.com/AlexeyAB/darknet).
- ANDRADE, R. A. d. *Ri Ufpe: Confiabilidade da Medida Ultrassonográfica do deslocamento do osso hioide durante a deglutição : Uma Revisão Sistemática E metanálise*. Universidade Federal de Pernambuco, 2021. Disponível em: [⟨https://repositorio.ufpe.br/handle/123456789/41440⟩](https://repositorio.ufpe.br/handle/123456789/41440).
- ANEAS, G. C. G.; DANTAS, R. O. A videofluoroscopia da deglutição na investigação da disfagia oral e faringea. *GE Jornal Português de Gastrenterologia*, v. 21, n. 1, p. 21–25, 2014. ISSN 0872-8178. Disponível em: [⟨https://www.sciencedirect.com/science/article/pii/S0872817813001379⟩](https://www.sciencedirect.com/science/article/pii/S0872817813001379).
- AUVENSHINE et al. The hyoid bone: an overview. *Cranio*, England, v. 38, n. 1, p. 6–14, out. 2018.
- CDC. *Ultrassonografia: O que É, para que serve e para quem É indicado*. 2019. Disponível em: [⟨https://cliniacadacidade.com.br/ultrassonografia-o-que-e-para-que-serve-e-para-quem-e-indicado/⟩](https://cliniacadacidade.com.br/ultrassonografia-o-que-e-para-que-serve-e-para-quem-e-indicado/).
- CHAVAN, S. K.; JOSHI, M. M. New perspective of hyoid bone morphology. In: . [S.l.: s.n.], 2019.
- CHI-FISHMAN, G.; SONIES, B. C. Effects of systematic bolus viscosity and volume changes on hyoid movement kinematics. *Dysphagia*, v. 17, n. 4, p. 278–287, Dec 2002. ISSN 1432-0460. Disponível em: [⟨https://doi.org/10.1007/s00455-002-0070-7⟩](https://doi.org/10.1007/s00455-002-0070-7).
- COSTA, B. O. I. d. et al. Quantitative ultrasound assessment of hyoid bone displacement during swallowing following thyroidectomy. *Dysphagia*, v. 36, n. 4, p. 659–669, Aug 2021. ISSN 1432-0460. Disponível em: [⟨https://doi.org/10.1007/s00455-020-10180-5⟩](https://doi.org/10.1007/s00455-020-10180-5).
- DESHPANDE, A. *A Beginner's Guide To Understanding Convolutional Neural Networks — adeshpande3.github.io*. 2016. [⟨https://adeshpande3.github.io/A-Beginner's-Guide-To-Understanding-Convolutional-Neural-Networks/⟩](https://adeshpande3.github.io/A-Beginner's-Guide-To-Understanding-Convolutional-Neural-Networks/). [Accessed 10-Jan-2023].
- FAKHRY, N. et al. Analysis of hyoid bone using 3D geometric morphometrics: An anatomical study and discussion of potential clinical implications. *Dysphagia*, v. 28, n. 3, p. 435–445, set. 2013.
- FENG, S. et al. Automatic hyoid bone tracking in Real-Time ultrasound swallowing videos using deep learning based and correlation filter based trackers. *Sensors (Basel)*, Switzerland, v. 21, n. 11, maio 2021.

FREITAS, A. B. d. *Relações Entre Deglutição e qualidade vocal: Análise Ultrassonográfica E Perceptiva*. Pontifícia Universidade Católica de São Paulo, 2017. Disponível em: <https://repositorio.pucsp.br/jspui/handle/handle/19790>.

HOSSAIN, I. et al. Semi-automatic assessment of hyoid bone motion in digital videofluoroscopic images. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, Taylor Francis, v. 2, n. 1, p. 25–37, 2014. Disponível em: <https://doi.org/10.1080/21681163.2013.833859>.

HSIAO, M.-Y. et al. Effects of bolus consistency and volume on hyoid bone displacement during swallowing. *Ultrasound in Medicine Biology*, v. 43, p. S200, 2017. ISSN 0301-5629. WFUMB 2017 Taipei. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0301562917320422>.

HSIAO, M.-Y. et al. Deep learning for automatic hyoid tracking in videofluoroscopic swallow studies. *Dysphagia*, abr. 2022.

HUI, J. *mAP (mean Average Precision) for Object Detection* — *jonathan-hui.medium.com*. 2018. <https://jonathan-hui.medium.com/map-mean-average-precision-for-object-detection-45c121a31173>. [Accessed 22-Mar-2023].

JANIESCH, C.; ZSCHECH, P.; HEINRICH, K. Machine learning and deep learning. *Electronic Markets*, v. 31, n. 3, p. 685–695, set. 2021.

KARIMI, G. *Introduction to YOLO Algorithm for Object Detection* — *section.io*. 2021. <https://www.section.io/engineering-education/introduction-to-yolo-algorithm-for-object-detection/>. [Accessed 06-Jan-2023].

KEITA, Z. *Yolo Object Detection explained: A beginner's guide*. DataCamp, 2022. Disponível em: <https://www.datacamp.com/blog/yolo-object-detection-explained>.

KHURANA, P. *Object Detection—YOLO(You Look Only Once)* — *prvnk10.medium.com*. 2019. <https://prvnk10.medium.com/object-detection-yolo-you-look-only-once-32c47cf9eea1>. [Accessed 06-Jan-2023].

KIM, H.-I. et al. Hyoid bone tracking in a videofluoroscopic swallowing study using a Deep-Learning-Based segmentation network. *Diagnostics (Basel)*, Switzerland, v. 11, n. 7, jun. 2021.

KUKIL; RATH, S. *YOLOv7 Object Detection Paper Explanation Inference* — *learnopencv.com*. 2022. <https://learnopencv.com/yolov7-object-detection-paper-explanation-and-inference/>.

KUNDU, R. *YOLO Algorithm for Object Detection Explained [+Examples]* — *v7labs.com*. 2023. <https://www.v7labs.com/blog/yolo-object-detection>. [Accessed 29-Abr-2023].

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *Nature*, v. 521, n. 7553, p. 436–444, maio 2015.

LEE, D. et al. Online learning for the hyoid bone tracking during swallowing with neck movement adjustment using semantic segmentation. *IEEE Access*, v. 8, p. 157451–157461, 2020.

LEITE, K. K. de A. et al. Ultrassonografia e deglutição: revisão crítica da literatura. *Audiology - Communication Research*, FapUNIFESP (SciELO), v. 19, n. 4, p. 412–420, dez. 2014. Disponível em: <https://doi.org/10.1590/s2317-64312014000300001378>.

LIMA, A. *YOLO: Você olha uma vez - Detecção de objetos em tempo real 2013; Acervo Lima — acervolima.com*. 2020. <https://acervolima.com/yolo-voce-so-olha-uma-vez-deteccao-de-objetos-em-tempo-real/>. [Accessed 05-Apr-2023].

LOPES, M. K. S. *Detecção Automática de Osso Hióide para Estudo de Deglutição no Exame de Ultrassonografia Usando Deep Learning*. 2019. [Accessed 06-Jan-2023].

LOW, A. et al. Introduction to node.js. In: *Proceedings of 24th Annual International Conference on Computer Science and Software Engineering*. USA: IBM Corp., 2014. (CASCON '14), p. 283–284.

MARQUES, V. *Exame de videofluoroscopia de Uma Deglutição normal - fonoaudiologia - Viviane Marques*. YouTube, 2015. Disponível em: <https://www.youtube.com/watch?v=yTdZN46LE9s>.

MINTZ, Y.; BRODIE, R. Introduction to artificial intelligence in medicine. *Minimally Invasive Therapy & Allied Technologies*, Taylor Francis, v. 28, n. 2, p. 73–81, 2019. PMID: 30810430. Disponível em: <https://doi.org/10.1080/13645706.2019.1575882>.

OTOMEDICA. *Videolaringoscopia*. YouTube, 2020. Disponível em: <https://www.youtube.com/watch?v=eNZuNdFX8BI>.

PIEMONTEZ, P. *Identificação, Detecção, reconhecimento e segmentação de imagem E objetos*. 2022. Disponível em: <https://visaocomputacional.com.br/identificacao-deteccao-reconhecimento-e-segmentacao-de-imagem-e-objetos/>.

REDMON, J. et al. *You Only Look Once: Unified, Real-Time Object Detection*. arXiv, 2015. Disponível em: <https://arxiv.org/abs/1506.02640>.

SINGH, G. *Introduction to Artificial Intelligence for Beginners — analyticsvidhya.com*. 2021. <https://www.analyticsvidhya.com/blog/2021/09/introduction-to-artificial-intelligence-for-beginners/>. [Accessed 14-Mar-2023].

STAFF. *Arquiteturas de Redes Neurais Convolucionais para reconhecimento de imagens - Viceri*. 2020. <https://viceri.com.br/insights/arquiteturas-de-redes-neurais-convolucionais-para-reconhecimento-de-imagens/>. (Accessed on 03/15/2023).

THATTE, A. V. *Evolution of YOLO—YOLO version 1 — towardsdatascience.com*. 2020. <https://towardsdatascience.com/evolution-of-yolo-yolo-version-1-afb8af302bd2>. [Accessed 06-Jan-2023].

YAMAZAKI, Y. et al. Excessive anterior cervical muscle tone affects hyoid bone kinetics during swallowing in adults. *Clin Interv Aging*, New Zealand, v. 12, p. 1903–1910, nov. 2017.

YANAGISAWA, E. et al. Videolaryngoscopy: A comparison of fiberscopic and telescopic documentation. *Annals of Otolaryngology, Rhinology & Laryngology*, v. 92, n. 5, p. 430–436, 1983. PMID: 6625439. Disponível em: <https://doi.org/10.1177/000348948309200504>.

YU, K.-H.; BEAM, A. L.; KOHANE, I. S. Artificial intelligence in healthcare. *Nature Biomedical Engineering*, v. 2, n. 10, p. 719–731, out. 2018.

APÊNDICE A - Detalhes do uso da darknet - YOLO

Para obter êxito na execução da YOLO, é importante se ater a algumas informações, como o arquivo de extensão *.name*, *.data*, e os arquivos responsáveis pelo caminho dos arquivos que se encontram na base de dados. O *nomedoarquivo.name* trata do nome da classe que será detectada, no modelo, o osso hioide. Esta classe poderá ser visualizada assim que a detecção acontecer.

O *nomedoarquivo.data* é um arquivo direcional, ou seja, ele é responsável por identificar algumas informações e caminhos importantes para o treinamento, teste e validação, nele, se encontra o número de classes presentes no modelo, o caminho do arquivo *.name*, dos arquivos que contém a rota para a base de dados e do backup, onde será salvo informações após o término do treinamento.

Os arquivos responsáveis pelo caminhos dos arquivos que se encontram na base de dados, são divididos em treino, teste e validação, e podem ser gerados através de script, como foi feito no presente trabalho, arquivo esse que tem o seu caminho identificado no *.data* já mencionado.

APÊNDICE B - Instruções para detecção em vídeo na darknet - YOLO

Além disso, para realizar corretamente a detecção em vídeos, é necessário habilitar no arquivo makefile, o OpenCV de “0” para “1”, e também a GPU, que foi mencionada anteriormente, junto à instalação do CUDA. A maneira de executar o teste em vídeo, tem suas diferenças, a darknet possibilita este ato, da seguinte maneira:

```
./darknet detector demo <arquivo.data> <arquivo.cfg> <arquivo.weights> <arquivo de vídeo>
```

Onde, o *<arquivo.weights>*, é o arquivo de pesos já treinados pela YOLO, e por último o *<arquivo de vídeo>* que consiste no vídeo do exame de ultrassom que terá o osso hioide detectado.

APÊNDICE C - Feedbacks para o usuário - Ambiente de Visualização

Na página inicial o ambiente de visualização oferece alguns *feedbacks*:

- Quando um arquivo foi carregado/inserido com sucesso (Figura 23);
- Quando nenhum arquivo foi carregado, ou seja, quando o usuário inicia a detecção sem carregar nenhum arquivo (Figura 24);
- Quando o usuário tenta carrega mais de um arquivo por vez, visto que o limite é um arquivo para cada carregamento e detecção (Figura 25).



Figura 23: Ambiente de Visualização - Arquivo inserido com sucesso

Fonte: Própria

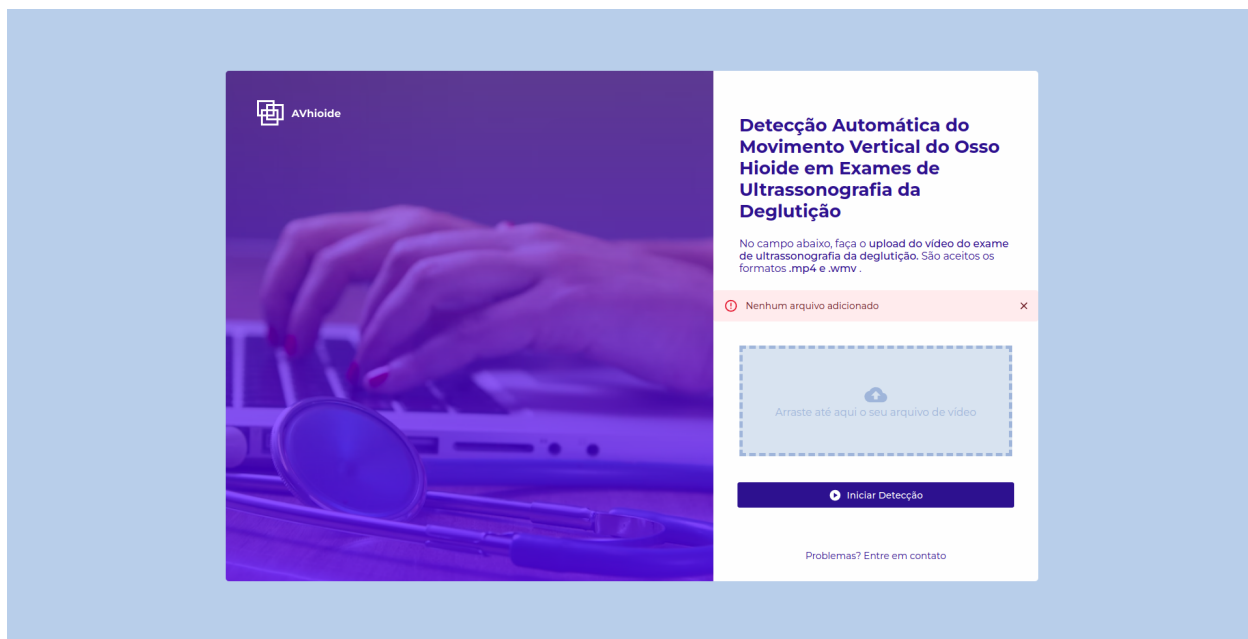


Figura 24: Ambiente de Visualização - Nenhum arquivo carregado
Fonte: Própria

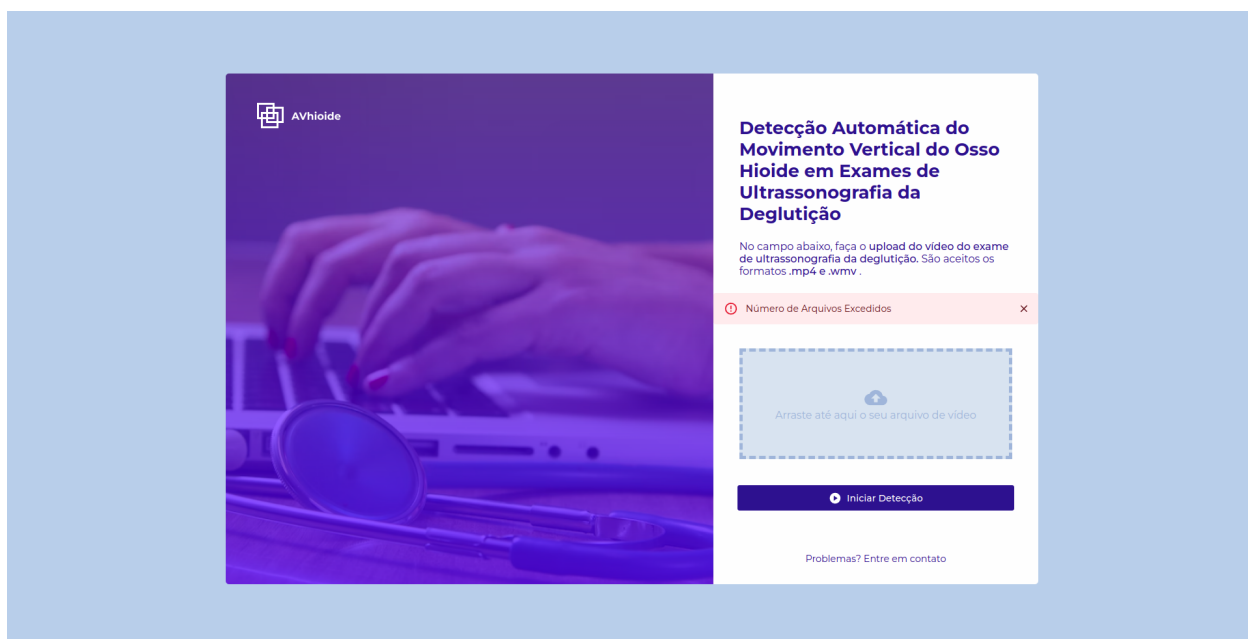


Figura 25: Ambiente de Visualização - Número de arquivos excedidos
Fonte: Própria