UNIVERSIDADE ESTADUAL DO MARANHÃO
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DA COMPUTAÇÃO E SISTEMAS

YOLOV11 DE BAIXA LATÊNCIA PARA CITOLOGIA CERVICAL: VIABILIZANDO DETECÇÃO EM TEMPO REAL ATRAVÉS DE MICROSCOPIA APOIADA POR COMPUTAÇÃO EM BORDA

FRANCISCO CARLOS SILVA PIMENTEL

SÃO LUÍS – MA
2026

UNIVERSIDADE ESTADUAL DO MARANHÃO
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DA COMPUTAÇÃO E SISTEMAS

FRANCISCO CARLOS SILVA PIMENTEL

YOLOV11 DE BAIXA LATÊNCIA PARA CITOLOGIA CERVICAL: VIABILIZANDO DETECÇÃO EM TEMPO REAL ATRAVÉS DE MICROSCOPIA APOIADA POR COMPUTAÇÃO EM BORDA

Dissertação de mestrado apresentada como requisito para obtenção do grau de mestre no Programa de Pós-Graduação em Engenharia da Computação e Sistemas – Mestrado em Engenharia da Computação e Sistemas.

Orientador: Prof.º PhD. Omar Carmona Andres Cortes.

SÃO LUÍS - MA
2026

Pimentel, Francisco Carlos Silva.
Yolov11 de baixa latência para citologia cervical: viabilizando detecção em tempo real através de microscopia apoiada por computação em borda. / Francisco Carlos Silva Pimentel. - São Luís - MA, 2026.
100 f.

Dissertação (Mestrado em Engenharia de Computação e Sistemas) - Universidade Estadual do Maranhão, São Luís, 2026.

Orientador: Prof. Dr. Omar Andres Carmona Cortes.

1. Câncer Cervical. 2. Papanicolau. 3. IA. 4. Aprendizado Profundo. 5. YOLO. I. Título.

CDU: 616-006.6:004

Elaborado por Luciana de Araújo - CRB 13/445

FRANCISCO CARLOS SILVA PIMENTEL

YOLOV11 DE BAIXA LATÊNCIA PARA CITOLOGIA CERVICAL: VIABILIZANDO DETECÇÃO EM TEMPO REAL ATRAVÉS DE MICROSCOPIA APOIADA POR COMPUTAÇÃO EM BORDA

Dissertação apresentada junto ao curso de Mestrado Profissional do Programa de Pós-graduação em Engenharia da Computação e Sistemas da Universidade Estadual do Maranhão (UEMA) para obtenção do título de Mestre em Engenharia da Computação e Sistemas.

Aprovado, 6 de março de 2026.

gov.br Documento assinado digitalmente
RICARDO MARCONDES MARCACINI
Data: 29/05/2026 03:37:13-0300
Verifique em https://validar.iti.gov.br

Prof.º Dr. Ricardo Marcondes Marcacini
Examinador Externo
ICM/USP

gov.br Documento assinado digitalmente
EWALDO EDER CARVALHO SANTANA
Data: 26/05/2026 07:16:13-0300
Verifique em https://validar.iti.gov.br

Prof.º Dr. Ewaldo Eder Carvalho Santana
Examinador Interno
PECS/UEMA

gov.br Documento assinado digitalmente
OMAR ANDRES CARMONA CORTES
Data: 29/05/2026 11:59:44-0300
Verifique em https://validar.iti.gov.br

Prof.º Dr. Omar Andres Carmona Cortes
Orientador
UEMA

Ao Deus toda honra e toda gloria em todo seu esplendor, ao qual eu consagro todos os meus passos e decisões.

AGRADECIMENTOS

O biênio deste Mestrado foi uma opção e decisão de vida que exigiu muito esforço, dedicação e sacrifícios pessoais. Por isso quero deixar registro o meu profundo agradecimento às pessoas que tiveram uma parcela de contribuição ou mesmo participaram junto comigo de todo este processo.

A Deus pelas multiformes surpresas que Ele tem me proporcionado, mesmo em tantos momentos de dor e sofrimento, os quais forjaram uma nova criatura em mim, mais resiliente e convicto de que "todas as coisas cooperam para o bem daqueles que O amam e seguem os Seus propósitos", pois, dias melhores virão sob a lente do olhar do coração: muito obrigado!

Aos meus pais, Nocrato e Geny, por estarem comigo, incondicionalmente, principalmente por toda compreensão e incentivos: muito obrigado!

À minha irmã, Cláudia Pimentel, pelas palavras de incentivo, pelos momentos que deixou seu descanso para me ajudar com revisões e sugestões: muito obrigado!

A Universidade Estadual do Maranhão (UEMA), por apoiar e financiar este trabalho por meio do qual foi possível aquisição de equipamentos que possibilitaram o desenvolvimento desta pesquisa: muito obrigado!

Ao meu Orientador, Prof.º PhD. Omar Carmona Andres Cortes, por ter me concedido a oportunidade de ser seu orientando, pela sua paciência, capacidade, inteligência, suas sábias sugestões e por ter provido todos os meios necessários para a conclusão deste trabalho: muito obrigado!

Ao Carlos Funasa, que me concedeu umas horas de seu tempo para uma reunião na qual me apresentou o Laboratório Central e me sugeriu o tema deste trabalho: muito obrigado!

Ao meu amigo Fernando Mendes, que é citologista e me ajudou com todo o conhecimento técnico e prático, tão imprescindível para o desenvolvimento deste trabalho: muito obrigado!

Aos demais professores e colegas que sempre contribuem com uma parte do aprendizado: muito obrigado!

A todos aqueles que contribuíram direta ou indiretamente com esta retomada aos estudos: muito obrigado!

“A tarefa não é tanto ver aquilo que ninguém viu, mas pensar o que ninguém ainda pensou sobre aquilo que todo mundo vê.”

Arthur Schopenhauer

RESUMO

Atualmente o câncer cervical representa o quarto tipo de neoplasia maligna mais comum entre as mulheres, estando predominantemente associado a infecção pelo papiloma vírus humano (HPV). Em se tratando de sua alta incidência, o diagnóstico precoce é fundamental, uma vez que os sintomas costumam ser silenciosos ou pouco perceptíveis, neste caso, o exame Papanicolau convencional continua sendo utilizado amplamente, sobretudo em regiões menos desenvolvidas, contudo, apresenta significativas limitações como a propensão a resultados falsos negativos. Neste contexto, com o objetivo de aprimorar a acurácia e a eficiência diagnóstica, este estudo propõe o desenvolvimento de um sistema automatizado para análise citológica cervical, integrando uma câmera de microscópio CMOS e algoritmos de detecção de imagem baseados em inteligência artificial. Para tal, foram empregadas as variantes *nano* (n) e *small* (s) do modelo YOLOv11, visando à detecção de células cervicais em condições reais de uso. Assim, o treinamento dos modelos foram realizados com base em conjuntos de dados públicos, incluindo SIPaKMeD, CRIC Cervix e RepoMedUNM. O modelo YOLOv11n alcançou um desempenho satisfatório, com precisão em mAP de 99,30%, *Recall* de 99,20% e taxa de processamento de 21,36 quadros por segundo. Adicionalmente, foram conduzidos experimentos comparativos utilizando dois bancos de dados com as mesmas imagens, em resoluções de 1280 x 1280 e 2016 x 2016 pixels. O objetivo foi avaliar o impacto do uso de resoluções superiores a 640 x 640 na performance das variantes (n) e (s), o que otimizou sua eficiência por meio da análise de imagens de maior definição.

Palavras-chave: *Câncer Cervical, Papanicolau, IA, Aprendizado Profundo, CNN, YOLO.*

ABSTRACT

Cervical cancer is the fourth most common malignant neoplasm among women worldwide and is predominantly associated with infection by the human papillomavirus (HPV). Given its high incidence, early diagnosis is essential, as symptoms are often silent or minimally perceptible. Although the conventional Papanicolaou test remains widely used, particularly in less developed regions, it presents important limitations, including a notable susceptibility to false-negative results. To improve diagnostic accuracy and efficiency, this study proposes the development of an automated system for cervical cytology analysis, integrating a CMOS microscope camera with artificial intelligence–based image detection algorithms. For this purpose, the nano (n) and small (s) variants of the YOLOv11 model were employed to detect cervical cells under real-world operating conditions. Model training was performed using publicly available datasets, including SIPaKMeD, CRIC Cervix, and RepoMedUNM. The YOLOv11n model demonstrated robust performance, achieving a mean Average Precision (mAP) of 99.30%, a recall of 99.20%, and a processing rate of 21.36 frames per second. Furthermore, comparative experiments were conducted using two datasets containing identical images with resolutions of 1280 × 1280 and 2016 × 2016 pixels. The objective was to evaluate the impact of employing resolutions higher than 640 × 640 on the performance of the (n) and (s) variants, enabling improved efficiency through the analysis of higher-definition images.

Keywords: *Cervical Cancer, Pap-smear, AI, Deep Learning, CNN, YOLO.*

LISTA DE FIGURAS

LISTA DE QUADROS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

ASC-H	*Atypical Squamous Cells, Cannot Exclude HSIL* (Células Escamosas Atípicas, Não Podem Excluir HSIL)
ASCUS	*Atypical Squamous Cells of Undetermined Significance* (Células Escamosas Atípicas, Não Podem Excluir HSIL)
AP	*Average Precision* (Precisão Média)
BB	*Bounding Boxes* (Caixas Delimitadoras)
CC	Câncer Cervical
CCC	Citologia Cervical Convencional
CML	Citologia em Meio Líquido
CNN	*Convolutional Neural Network* (Rede Neural Convolucional)
CSAIL	*MIT Computer Science and Artificial Intelligence Laboratory* (Laboratório de Ciência da Computação e Inteligência Artificial do MIT)
DL	*Deep Learning* (Aprendizado Profundo)
FOV	*Field of View* (Campos de Visão)
FPS	*Frames per Second* (Quadros por segundo)
FC	*Fully Connected* (Totalmente Conectada)
IA	Inteligência Artificial
IoU	*Intersection over Union* (interseção pela união)
HSIL	*High-grade Squamous Intraepithelial Lesion* (Lesão Intraepitelial Escamosa de Alto Grau)
HPV	*Papilomavirus humano*
LSIL	*Low-grade Squamous Intraepithelial Lesion* (A Lesão Intraepitelial Escamosa de Baixo Grau)
ILSVRC	*ImageNet Large Scale Visual Recognition Challenge*
ML	*Machine Learning* (Aprendizado de Máquina)
NILM	*Negative for Intraepithelial Lesion or Malignancy* (Negativo para Lesão Intraepitelial ou Malignidade)
NMS	*Non-Max Suppression* (Supressão Não Máxima)
mAP	*Mean Average Precision* (Média da Precisão Média)
OMS	Organização Mundial de Saúde
ROI	*Region of Interest* (Região de Interesse)
RPN	*Region Proposal Network* (Rede de Proposta de Regiões)
SIFT	*Scale-Invariant Feature Transform* (Transformação de características invariantes de escala)
SUS	Sistema Único de Saúde

SVM	*Support Vector Machines* (Máquina de Vetores de Suporte)
TBS	*The Bethesda System* (Sistema Bethesda)
WSI	*Whole Slide Imaging* (Imagem de lâmina inteira)
YOLO	*You Only Look Once* (Você olha somente uma vez)
VOC	*Visual Object Classes* (Classes de Objetos Visuais)

SUMÁRIO

1 INTRODUÇÃO

O câncer cervical (CC) que afeta o colo do útero, é o quarto tipo de neoplasia maligna mais comum entre as mulheres (RAHAMAN, LI, *et al.*, 2020, p. 61687). Segundo estatísticas da Organização Mundial de Saúde (OMS), em 2020 foram registrados aproximadamente 604.127 novos casos e cerca de 341.831 mortes (SINGH, VIGNAT, *et al.*, 2023, p. 197). Dentre estes, 85% ocorrem em países menos desenvolvidos, refletindo a falta de conscientização da população sobre os riscos da doença e acesso limitado a exames e serviços de saúde (NAKISIGE, SCHWARTZ e NDIRA, 2017, p. 37).

No Brasil, anualmente, estima-se que surjam cerca de 17.010 novos casos, com aproximadamente 7.000 óbitos (SANTOS, LIMA, *et al.*, 2023, p. 10). A incidência é maior em regiões carentes distantes dos centros de triagem e diagnóstico. O CC está fortemente associado à infecção pelo *papilomavírus humano* (HPV), sendo que cerca de 95% dos casos decorrem de infecções por cepas de alto risco, entre a população feminina, especialmente os tipos 16 e 18. A doença possui um longo estágio pré-canceroso, durante o qual alterações nucleares nas células podem ser observadas, como o aumento da proporção núcleo-citoplasma (SCHILITHZ, DE LIMA, *et al.*, 2015, p. 38-39).

Nas fases iniciais, o CC pode ser assintomático. Com a progressão, sintomas como sangramento vaginal anormal, corrimento e dor pélvica tornam-se mais evidentes. Por isso, o diagnóstico precoce é essencial para o tratamento eficaz e prognóstico positivo (COHEN, 2019 *Apud* (JIANG, LI, *et al.*, 2023, p. 2688).

A OMS recomenda três métodos de triagem para o CC: teste de HPV para tipos de alto risco, citologia cervical e inspeção visual com ácido acético. Dentre eles, a citologia cervical é amplamente utilizada por sua capacidade de indicar alterações celulares significativas (JIANG, LI, *et al.*, 2023, p. 2688). O exame citopatológico convencional, conhecido como teste de Papanicolau, foi introduzido por George Papanicolau e Aurel Babes em 1928, consolidando-se como método de prevenção a partir de 1943 (STABILE, EVANGELISTA, *et al.*, 2012, p. 467). Sua relevância reside na análise morfológica das células cervicais. Classificadas como normais, pré-neoplásicas ou malignas (JIANG, LI, *et al.*, 2023, p. 2691).

Apesar de ser um exame rápido, acessível, eficaz e de baixo custo, o Papanicolau apresenta limitações, como falhas na coleta e preparação das lâminas, além da escassez de

profissionais qualificados e diagnósticos com resultados falsos negativos (CAETANO, VIANNA, *et al.*, 2006, p. 100; FILHO, PEREIRA, *et al.*, 2005, p. 33). Embora existam metodologias de triagem mais avançadas, como a Citologia em Meio Líquido (CML), o exame convencional ainda é o único ofertado pelo Sistema Único de Saúde (SUS) no Brasil por mais de 30 anos (GIRIANELLI, THULER, *et al.*, 2004, p. 101; CAETANO, VIANNA, *et al.*, 2006, p. 101).

Diante deste cenário, o presente estudo propõe automatizar o exame de Papanicolau convencional empregando Inteligência Artificial (IA), utilizando redes neurais convolucionais (*Convolutional Neural Network,* CNN) para detectar lesões precursoras de câncer em células cervicais. A abordagem desta pesquisa envolve o emprego de câmeras CMOS adaptáveis a microscópios ópticos capazes de captura de imagens de alta resolução e a exibição de vídeo em tempo real na tela de um computador.

O objetivo é avaliar a capacidade e a velocidade de inferência de modelos de visão computacional com arquiteturas reduzidas, adequadas para detecção de vídeo em tempo real. Para tal feito, foi adotado o modelo YOLOv11 (*You Only Look Once*), conhecido por sua eficiência na detecção de objetos em tempo real, mesmo sem o uso de unidades de processamento gráfico (GPU) (REDMON, DIVVALA, *et al.*, 2016, p. 1-3).

Embora a versão YOLOv12 apresente maior precisão, seu custo computacional não o qualificou para os experimentos. Por essa razão, optou-se pela versão YOLOv11 que apresentou desempenho similar em *hardware* com ou sem GPUs. Entretanto, as variantes *nano* (n) e *small* (s), que oferecem menor complexidade e maior velocidade de inferência foram uma opção mais viável para esta pesquisa (KHANAM e HUSSAIN, 2024, p. 2).

Além disso, considerando que a percepção de movimento em vídeo requer uma taxa mínima de 24 quadros por segundo (FPS), por isso, é fundamental que o modelo de detecção opere em milissegundos para garantir fluidez visual. A maioria das CNNs apresenta tempo de inferência de 2 a 3 segundos por imagem (SZEGEDY, LIU, *et al.*, 2015, p. 9), o que compromete a exibição em tempo real. O YOLOv11, por sua arquitetura otimizada, supera essa limitação (REDMON, DIVVALA, *et al.*, 2016, p. 2-3).

E por fim, como propósito deste trabalho, destacam-se as seguintes contribuições:

- Produzir um estudo que possa estabelecer um *framework* de detecção de lesões

de células cervicais assistido por IA por meio de câmeras para microscópio em tempo real;

- Destacar a utilização exclusivamente bases de dados públicas de imagens de células cervicais, como SIPaKMeD, CRIC *Cervix* e RepoMedUNM, promovendo reprodutibilidade e acessibilidade para pesquisadores em países em desenvolvimento;

- Promover uma pesquisa direcionada a exploração do aumento de precisão e da vantagem da velocidade de inferência das variantes reduzidas do modelo YOLOv11;

- Desenvolver uma ferramenta que possa contribuir na luta do câncer de colo de útero, bem como, dispor mais uma pesquisa para exames por vídeo em tempo real com detecção por IA.

1.1 PROBLEMA

Apesar da ampla utilização do exame Papanicolau como método de rastreio do CC, sua análise manual apresenta limitações críticas que podem comprometer a eficácia diagnóstica. Entre os principais problemas estão a possibilidade de erros na coleta e preparação das lâminas, além da escassez de profissionais especializados, o que contribui para sobrecarga de trabalho e aumento da incidência de resultados falso negativos. Esses fatores dificultam a detecção precoce da doença, comprometendo o tratamento e a segurança das pacientes.

Diante deste cenário, torna-se necessário o desenvolvimento de soluções tecnológicas que ampliem a capacidade diagnóstica com maior precisão e agilidade. A IA, especialmente por meio de CNNs, como o modelo YOLO, apresenta-se como uma alternativa promissora para automatizar a detecção de lesões precursoras de CC em tempo real.

Portanto, considerando a relevância da questão, ora introduzida, o presente estudo traz como pergunta orientadora o seguinte problema: ***uma solução baseada em IA utilizando uma CNN YOLOv11 pode auxiliar na detecção automatizada de lesões percursoras de câncer em tempo real, visando mitigar a sobrecarga dos citologistas e aumentar a precisão e velocidade diagnóstica dos exames de Papanicolau?***

1.2 JUSTIFICATIVA

Apesar do exame de Papanicolau convencional ser eficaz e apresentar bom custo-benefício, essa metodologia possui limitações inerentes aos procedimentos manuais de coleta, preparação e análise. Tais limitações comprometem a acurácia diagnóstica, principalmente em contextos de alta demanda e escassez de profissionais especializados (CAETANO, VIANNA, *et al.*, 2006, p. 100).

Desde a década de 1950, há tentativas de automatizar a citologia cervical. Contudo, apenas a partir da década de 1990, com o surgimento da Citologia em Meio Líquido (CML), houve avanços significativos no desenvolvimento de ferramentas automatizadas para triagem do CC (GUPTA, KUMAR, *et al.*, 2023, p. 1644). A CML, embora mais precisa e padronizada, ainda não está amplamente disseminada no Brasil, sobretudo em regiões menos desenvolvidas (FILHO, PEREIRA, *et al.*, 2005, p. 33).

Desde 2018, estudos envolvendo IA têm registrado melhoria na precisão, especialmente após a transição do uso de algoritmos de Aprendizado de Máquina (*Machine Learning*, ML) para as CNNs. No entanto, apesar de todos os avanços significativos com o emprego de IA, sua utilização na triagem do câncer cervical ainda não é amplamente difundida (VARGAS-CARDONA, RODRIGUEZ-LOPEZ, *et al.*, 2024, p. 568;570; BADI, AGARWAL, *et al.*, 1991, p. 37).

Posto isto, entende-se que é necessário propor alternativas acessíveis e reprodutíveis, alinhadas à realidade de países em desenvolvimento que ainda dependem da metodologia convencional, uma vez que o exame de Papanicolau é utilizado mundialmente para triagem de câncer cervical (FILHO, PEREIRA, *et al.*, 2005, p. 33).

Neste contexto, que envolve limitações da citologia convencional e o histórico de aplicação de IA na área médica, esta pesquisa propõe a aplicação de modelos de visão computacional para automatizar e aumentar tanto a precisão quanto a celeridade dos diagnósticos de exames citopatológicos.

Em consonância do caráter profissional deste mestrado, optou-se por uma abordagem simplificada e acessível, adotando o uso exclusivo de bases de dados públicas de imagens de células cervicais. Tal iniciativa visa incentivar outros pesquisadores a reproduzirem os resultados e fomentar novas contribuições voltadas à triagem automatizada do câncer

cervical.

Além disso, nos experimentos, pretende-se empregar câmeras digitais USB CMOS, adaptáveis a microscópios, por serem facilmente conectáveis a computadores. Essa escolha se justifica pela possibilidade de aproveitar toda a infraestrutura já existente nos laboratórios tradicionais.

Para que o movimento em um vídeo seja percebido como real e fluido, é necessário a exibição de uma sequência de imagens. Segundo Bordweel e Tompson (2018, p. 6), "*para que o movimento no cinema pareça suave e contínuo, pelo menos 24 imagens distintas (ou quadros/frames) devem ser projetados a cada segundo"*.

Essa quantidade de imagens se tornou a base técnica da indústria, que padronizou a captura e exibição na taxa de 24 a 25 quadros (*frames*) por segundo (FPS) (MURCH, 2001, p. 142). Quanto menor a taxa de FPS, mais lento e menos continuo o vídeo é exibido (ZETTL, 2016, p. 58). Portanto, destaca-se a importância de manter a exibição de pelo menos 24 FPS para garantir a percepção de movimento (SALT, 2009, p. 31).

Posto isto, a maioria das CNNs apresenta um tempo de inferência de 2 a 3 segundos por imagem, o que as torna inviáveis para detecção em vídeo em tempo real (REN, HE, *et al.*, 2017, p. 91). A consequência disso é simples: se, para cada uma das 25/24 imagens em um segundo de vídeo, uma CNN leva 2 segundos para inferir a uma única imagem, isso reduz drasticamente a taxa de FPS (ZETTL, 2016, p. 58).

Consequentemente, o vídeo ficará muito lento, com uma média de apenas 2 FPS. Por essa razão, é fundamental que o modelo de detecção seja rápido o suficiente para realizar inferências em milissegundos, aproximando-se de taxas de inferência que permitam a operação em tempo real (LIU, ANGUELOV, *et al.*, 2016, p. 5).

Ademais, considerando o fornecimento de vídeos em tempo real pelas referidas câmeras CMOS mencionadas, o algoritmo CNN YOLO foi selecionado com base em sua ampla utilização na área médica, especialmente na detecção de lesões e tumores. Trata-se de uma solução com uma arquitetura eficiente para a detecção em vídeo, capaz de atingir boa velocidade em tempo real com ou sem o uso de GPUs (RAGAB, ABDULKADIR, *et al.*, 2024, p. 57816-57817).

Por fim, embora a versão 12 do YOLO tenha demonstrado maior precisão em

diversos experimento com diferentes bases de dados, seus avanços são atingidos a um certo custo computacional. Portanto, foi adotada a versão 11 do referido algoritmo, que apresenta desempenho similar com menor complexidade, com foco especial, nas suas variantes reduzidas, *nano* (n) e *small* (s), capazes de alcançar maior velocidade de inferência e eficiência (NIDHAL , KOH, *et al.*, 2024, p. 26-28).

1.3 QUESTÕES DE PESQUISA

Norteando-se pelo problema, três questões foram levantadas relacionadas a solução de automatização do exame de Papanicolau:

a) Q01: É possível utilizar apenas bases públicas para produzir um modelo treinado de IA em visão computacional para detecção de células cervicais?
b) Q02: A partir de apenas recursos de domínio público, é possível propor uma solução de automatização de citologia cervical alternativa que possa ser totalmente reproduzível por outros pesquisadores?
c) Q03: Arquiteturas de CNNs com arquiteturas reduzidas podem melhorar o desempenho de treinamento aumentando a precisão com implementação de imagens com dimensões superiores a 640 x 640 pixels?

1.4 HIPÓTESES DA PESQUISA

Considerando as questões de pesquisa elencados anteriormente a proposta de pesquisa pretende investigar as hipóteses formuladas que nortearão e serão testadas ao longo do estudo, como seguem:

a) H1: o uso de câmeras CMOS acopladas a microscópios ópticos permite a captura de imagens com qualidade suficiente para inferência em tempo real por modelos de visão computacional.
b) H2: a aplicação de modelos de arquitetura reduzida, como YOLOv11 nas variantes *nano* e *small*, proporciona desempenho satisfatório (mAP >= 60%) na detecção de células cervicais em vídeo em tempo real.
c) H3: o uso de imagens com resoluções superiores a 640 x 640 pixels melhora a precisão e o *recall* dos modelos de detecção, sem comprometer significativamente a velocidade de inferência.

d) H4: a automação do exame de Papanicolau convencional com inteligência artificial reduz a incidência de falsos negativos em comparação ao método CML, contribuindo para maior confiabilidade diagnóstica.
e) H5: as bases de dados públicas juntamente com a aplicação de técnicas de aumento de dados (*data augmentation*), fornecem amostras suficientes de lesões celulares capazes de gerar bom desempenho no treinamento de modelos CNN do tipo YOLO.

Quadro 1 - Relação entre questões de pesquisa e hipóteses

Questões de Pesquisa	**Hipótese Correspondente**
Q01: É possível utilizar apenas bases publicas para produzir um modelo treinado de IA em visão computacional para detecção de células cervicais?	H2, H3 e H5
Q02: A partir de apenas recursos de domínio público, é possível propor uma solução de automatização de citologia cervical alternativa que possa ser totalmente reproduzivel por outros pesquisadores?	H1, H4 e H5
Q03: Arquiteturas de CNNs com variantes reduzidas podem melhorar o desempenho de treinamento aumentanto a precisão com implementação de imagens com dimensões superior a 640 x 640 pixels?	H2 e H3

Fonte: Elaborado pelo Autor (2025).

Com o intuito mais didático, elaborou-se o Quadro 1 para apresentar a correspondência direta entre as questões de pesquisa e as hipóteses formuladas, evidenciando o alinhamento lógico entre os elementos estruturantes da investigação.

1.5 OBJETIVOS

Para responder às perguntas e averiguar se há uma solução ou mitigação do problema comentado na justificativa, os objetivos gerais e específicos são descritos nos subitens a seguir.

1.5.1 Objetivo Geral

Este presente trabalho tem como objetivo geral, propor um sistema automatizado para auxiliar o diagnóstico de câncer cervical em vídeo em tempo real. Tal sistema visa empregar algoritmos de visão computacional reduzidos a fim de explorar a velocidade e eficiência destes modelos aumentando sua precisão.

1.5.2 Objetivos Específicos

a) Eleger apenas bases de dados de células cervicais de domínio público para garantir a reprodutibilidade desta pesquisa;
b) Conduzir experimentações de câmeras digitais CMOS para microscópios, a fim de deduzir o melhor equipamento para inferência de vídeo;
c) Definir qual o melhor modelo de visão computacional com arquiteturas reduzidas que proporcione detecção de vídeo em tempo real com melhor velocidade e precisão;
d) Realizar teste de inferência em vídeo em tempo real com CNN YOLOv11;
e) Realizar testes de desempenho das variantes YOLO (n) e (s) em tarefas de detecção por caixas delimitadoras e segmentação semântica;
f) Realizar testes de inferência em vídeo em tempo real provido tanto por câmeras CMOS montadas no microscópio quanto em vídeos de domínio público;

1.6 METODOLOGIA

Esta seção descreve a abordagem metodológica adotada para o desenvolvimento e avaliação desta pesquisa, utilizando algoritmos de visão computacional com arquiteturas reduzidas. No contexto científico, com base em Gil (2017) e Wazlawick (2021), a pesquisa é classificada como aplicada, voltada a mitigação das limitações do exame de citologia cervical convencional por meio de modelos de IA com rápida inferência em vídeo. A natureza do estudo é experimental, utilizando técnicas de aprendizado profundo (*Deep Learning,* DL), adotando as CNNs para detectar células cervicais anormais em imagens capturadas por câmeras CMOS acopladas diretamente a microscópios.

A abordagem metodológica combina características descritivas e exploratórias. O aspecto descritivo está relacionado ao levantamento e análise de dados, enquanto o exploratório busca identificar padrões e associações entre diferentes sistemas de classificações celulares em relação às imagens (WAZLAWICK, 2021, p. 43). Embora a pesquisa incorpore elementos quantitativos e qualitativos (GIL, 2017, p. 113), predomina a vertente quantitativa, com uso de métricas e técnicas de DP para responder à questão central do estudo, complementadas pela validação dos resultados por citologistas.

Todo o desenvolvimento deste trabalho segue a metodologia CRISP-DM, constituída como padrão de referência em projetos de Ciência de Dados. Essa metodologia está

dividida em seis fases: Compreensão de Negócios, Compreensão de Dados, Preparação de Dados, Modelagem, Avaliação e Implantação (SEKAR, 2022, p. 44; CHAPMAN, CLINTON, *et al.*, 2000). Todas as fases dessa metodologia são descritas em detalhes na seção METODOLOGIA CRISP-DM.

1.6.1 Procedimentos Metodológicos

Para a condução deste trabalho, adotou-se a metodologia *Cross Industry Standard Process for Data Mining*, traduzida como "Processo Padrão Inter-Indústrias para Mineração de Dados" (CRISP-DM). Essa metodologia é considerada referência em ciência de dados, conforme Schröer, Kruse e Gómez (2021), e amplamente utilizada na prática (SEKAR, 2022, p. 43).

Os resultados apresentados foram obtidos a partir de experimentos descritos no Quadro 4, constante na seção 3.1.1, os quais se relacionam diretamente às três questões de pesquisa delineadas na seção 1.3. A associação entre essas questões e os experimentos constitui a base dos procedimentos metodológicos aplicados. A seguir, apresenta-se a síntese dos experimentos realizados e descritos no tópico 3:

- **Experimento 01**: Testes de desempenho das variantes menores do YOLOv11, (*Nano* e *Small)* em tarefas de detecção por caixas delimitadoras e segmentação de instâncias. O objetivo foi identificar qual variante oferece melhor equilíbrio entre precisão e velocidade de detecção em tempo real.

- **Experimento 02**: Avaliação dos modelos treinados nas variantes do YOLOv11 utilizando o banco de imagens cervicais APACC, não incluído no treinamento. Esse experimento permitiu verificar a capacidade de generalização dos modelos em situações inéditas.

- **Experimento 03**: Análise do desempenho das variantes em vídeos em tempo real capturados por câmeras USB CMOS acopladas a microscópios, com foco na tarefa de segmentação.

- **Experimento 04**: Seleção das bases de dados por meio de revisão de publicações sobre imagens de células cervicais. Foram utilizadas apenas bases públicas, sem envolvimento de dados sensíveis ou materiais genéticos de pacientes, garantindo

a replicabilidade do estudo.

- **Experimento 05**: Verificação da distribuição de classes nas bases selecionadas antes da unificação, com o objetivo de evitar enviesamentos no *dataset* final.

- **Experimento 06**: Unificação das bases em um único *dataset* de treinamento, harmonizando divergências de classificação celular em um sistema único de categorias de lesões.

- **Experimento 07**: Aplicação da técnica de *data augmentation* (aumento de dados) para balancear as classes após a unificação, fundamentada nos resultados do experimento 05, visando evitar possíveis enviesamentos de classes após a unificação dos bancos de dados.

- **Experimento 08**: Padronização das dimensões das imagens no *dataset* final, gerando dois conjuntos: um com resolução de 2016 x 2016 *pixels* e outro com 1280 x 1280 *pixels*. O objetivo, ligado à questão de pesquisa 03, foi investigar se resoluções superiores ao padrão de 640 pixels do YOLO influenciam na precisão dos modelos. Registrando que os tamanhos escolhidos correspondem ao dobro (1280) e aproximadamente o triplo (2016) de 640 *pixels*.

- **Experimento 09**: Treinamento das variantes YOLOv11 (*Nano* e *Small*) utilizando os dois *datasets* com as resoluções de 1280 x 1280 e 2016 x 2016 *pixels*, respectivamente. Essa etapa contemplou todas as combinações possíveis, incluindo ajustes de hiperparâmetros e resolução de problemas durante o treinamento.

1.7 TRABALHOS CORRELATOS

Esta pesquisa visa desenvolver uma ferramenta assistida por IA, em tempo real, para o diagnóstico do câncer cervical. Para tanto, investigaram-se os trabalhos similares mais relevantes. No entanto, as pesquisas neste domínio são bastante recentes, resultando em um número limitado de publicações. Os trabalhos correlatos discutidos nesta seção atendem a critérios específicos, incluindo: abordar algoritmos de detecção em tempo real e oferecer soluções aplicáveis tanto às metodologias CML quanto à citopatologia convencional.

Também foi conduzida a seleção de pesquisas que consideram CNNs capazes de

fornecer inferência rápida e adequada ao processamento de vídeo em tempo real, com auxílio de câmeras adaptadas a microscópios.

Tang et al. (2021) apresentaram um estudo sobre microscópios inteligentes assistidos por IA, projetados para fornecer diagnósticos citopatológicos em tempo real. Diferentemente dos sistemas baseados em *Whole Slide Imaging* (Microscopia Virtual ou Digitalização de Lâmina Inteira, WSI), sua abordagem utilizou um microscópio convencional equipado com um *display* de realidade aumentada na unidade ocular e uma unidade de computação de IA integrada.

Uma câmera montada no microscópio capturou imagens de alta resolução do campo de visão (*Field of View*, FOV), que foram então enviadas para um sistema de detecção utilizando um modelo *RetinaNet*. Os resultados foram posteriormente projetados na unidade ocular (TANG, CAI, *et al.*, 2021).

O FOV do microscópio pode ser alterado por meio da seleção manual de objetivas, permitindo ampliações de 10x, 20x e 40x. Para cada ampliação, três algoritmos de detecção distintos foram treinados para processar a respectiva resolução da objetiva.

O conjunto de dados foi coletado no Hospital de Saúde Materno-Infantil de Shenzhen, China, compreendendo um total de 2.167 lâminas de preparação em base líquida, resultando em um *dataset* de 84.156 amostras de células cervicais anotadas. Neste estudo preliminar, as células foram classificadas e anotadas de acordo com o TBS, com foco em cinco classes: NILM, ASCUS, ASCH, LSIL e HSIL. A avaliação do algoritmo para cada FOV foi conduzida com base na precisão em nível de célula.

Para o FOV de baixa ampliação (10x), a precisão foi de 0,254 e a sensibilidade, de 0,996. Para FOVs com ampliações de 20x e 40x, a sensibilidade foi de 0,90 e a precisão, de 0,80. No entanto, o tempo de inferência do modelo de detecção de células por FOV foi de aproximadamente 200 milissegundos, o que não é propício ao desempenho em tempo real.

Os autores também destacaram dois paradigmas existentes para o diagnóstico assistido por IA em patologia e citopatologia. O primeiro paradigma envolve a pré-varredura de uma lâmina inteira para uma análise WSI subsequente. O segundo paradigma envolve a análise do FOV do microscópio do observador e a apresentação dos resultados da detecção por IA em tempo real.

Aproveitando modelos de IA e tecnologia de *Internet* das Coisas (IoT), Sampaio e outros autores (2021) propuseram uma solução automatizada, portátil e baseada em IoT para triagem de CC. Eles desenvolveram um dispositivo de microscopia de baixo custo, baseado em dispositivos móveis, chamado *µSmartScope*, para analisar amostras LBC.

Este microscópio compacto é controlado por um smartphone. No entanto, reconhecendo a dificuldade de implantar modelos computacionais de detecção de imagens localmente em arquiteturas de dispositivos móveis, o sistema adota uma abordagem baseada em nuvem.

A pesquisa utilizou o conjunto de dados SIPaKMeD, um banco de dados de domínio público que compreende 966 imagens, totalizando 4049 amostras de células de citologia cervical categorizadas em cinco tipos distintos de células. O treinamento neste conjunto de dados produziu uma precisão média de 0,377 e uma *Recall* médio de 0,63. O estudo selecionou meticulosamente o modelo de IA para detecção de imagens, testando-o em resoluções de 320 x 320 e 640 x 640 *pixels*.

Além disso, os modelos *RetinaNet*, SSD, *Faster* R-CNN, *MobileNetV2* e *ResNet50* foram avaliados por meio de experimentos de ajuste de hiperparâmetros e validação cruzada, sendo a arquitetura Faster R-CNN identificada como a mais promissora. O tempo de inferência para a região de interesse (*Region of Interest*, ROI) dentro de cada imagem (que foi dividida em vários *patches*) foi em média de 2,768 segundos por imagem.

O autor considerou isso promissor, atribuindo o desempenho às conexões de rede com servidores remotos. No entanto, essa estimativa não levou em conta o tempo de aquisição de imagens, que é de aproximadamente 30 a 50 minutos na versão atual do sistema. Embora o desempenho atual, baseado em nuvem, apresente limitações, o autor prevê versões futuras incorporando detecção local em vez de processamento remoto. Essa abordagem local parece promissora para o desenvolvimento de uma verdadeira ferramenta de detecção em tempo real.

Alcaraz-Chavez et al. (2024) propuseram um método inovador para detectar e rastrear células precursoras de CC, utilizando um algoritmo de detecção em vez de um modelo de aprendizado profundo. Embora sua abordagem não empregasse modelos de aprendizado profundo, a pesquisa demonstrou detecção de células em tempo real.

Os sistemas de detecção atuais geralmente operam com algoritmos de duas etapas:

primeiro, a segmentação do citoplasma e do núcleo; em seguida, a detecção e classificação de células cervicais. Em contraste, os autores empregaram um algoritmo para detecção morfológica. Especificamente, foi utilizado o SIFT (*Scale-Invariant Feature Transform*), um algoritmo de visão computacional, capaz de reconhecer células por meio de uma única amostra pré-processada.

O algoritmo opera gerando pontos-chave a partir de características distintas dentro de uma célula. Esses pontos formam descritores de características celulares, que são submetidos a um cálculo de distância. Esse processo permite a identificação de pontos descritores consistentes, independentemente do tamanho da célula na imagem, destacando a invariância de escala do SIFT. As imagens foram fornecidas pelo Laboratório Estadual de Saúde Pública de Mochoácan (SPHLM), México, capturadas por uma câmera de microscópio de 5 *megapixels*, totalizando 15.000 fotos e 20 horas de vídeo de espécimes.

No entanto, o TBS não foi adotado para a classificação. Em vez disso, as lesões e os cânceres foram classificados por tipos celulares. A coleta inicial de dados foi realizada utilizando uma câmera de microscópio de 5 *megapixels*, a inferência de vídeo em tempo real foi realizada usando um dispositivo móvel *Samsung Note* 10, cuja câmera de 12 *megapixels* foi adaptada à lente do microscópio.

Durante a fase de testes, foram utilizadas 100 imagens contendo 1800 instâncias de células precursoras do câncer cervical. O sistema alcançou desempenho de detecção em tempo real com as seguintes métricas: precisão de 98,30%; *recall* de 98,20%; e uma medida F de 98,05%. A inferência foi realizada em lotes de 100 imagens e teve um tempo médio de 0,57 segundos, o que resultou em 0,0057 segundos por imagem.

Terasaki et al. (2024) apresentaram um sistema focado em exames de Papanicolau convencionais, no qual destacaram as dificuldades diagnósticas decorrentes de aglomerados de células sobrepostas. O objetivo de sua proposta foi propor uma alternativa à elevada demanda computacional dos sistemas atuais baseados em WSI que dependem do monitoramento de IA. A proposta consistiu no desenvolvimento de uma aplicação prática para microscopia tradicional, utilizando modelos de detecção baseados em aprendizado profundo para diagnóstico assistido por IA, com foco específico em citopatologia.

Para detecção em tempo real, o algoritmo YOLOv5 foi empregado em sua variante X (extragrande). As imagens usadas para treinar o modelo foram fornecidas pelo *Nippon*

Medical School Hospital, Japão. Elas foram capturadas com um *smartphone iPhone* SE conectado por meio de um adaptador em um microscópio *Olympus* BX53. O conjunto de dados resultante consistiu em 3.151 imagens de células cervicais, originalmente com uma resolução de 4.032 x 4.024 *pixels* e posteriormente redimensionadas para 640 x 640 *pixels*.

No entanto, foi utilizado um sistema simplificado para classificar as células em duas classes distintas: "benignas" e "malignas". Para inferência, empregou-se uma câmera Canon CCD JCS-HR5U conectada a um microscópio *Nikon ECLIPSE Ci,* fornecendo imagens de vídeo a uma taxa de 30 FPS. Nos experimentos de detecção, conduzidos com o modelo YOLOv5x e utilizando caixas delimitadoras (*Bounding Boxes*, BB), obteve-se uma precisão de 85% e um *recall* de 90%. Notavelmente, o modelo superou 600 patologistas em casos comparáveis. Além disso, o diagnóstico assistido por IA reduziu o tempo médio de análise do exame da lâmina de 4,458 minutos para 2,460 segundos.

Até onde se sabe, pesquisas que abordam especificamente a detecção automatizada em tempo real de células cervicais diretamente a partir de câmeras de microscópio ainda constituem um campo pouco explorado. Grande parte dos estudos existentes depende de imagens em super resolução obtidas por WSI, que são computacionalmente intensivas para análise pós-aquisição, ou recorre a processamento baseado em nuvem, o que introduz latência incompatível com a necessidade de *feedback* imediato.

1.7.1 Pesquisa Prévia

A escolha por utilizar as variantes reduzidas do YOLOv11 e buscar maior precisão desses modelos foi motivada por uma pesquisa anterior (PIMENTEL, REIS, *et al.*, 2023). Esse estudo tinha como objetivo detectar, em tempo real, o momento exato da abertura ou fechamento da tampa do porão de um navio. Para isso, comparou-se o desempenho das variantes menores do YOLOv4 em relação ao modelo padrão.

Neste contexto, o *dataset* utilizado continha 27,213 imagens de porões de navio em diferentes situações, treinadas tanto no YOLOv4 quanto em sua versão reduzida (*Fast/Tiny*). O treinamento foi realizado em um *notebook Acer Gamer*, RAM de 8G e GPU NVIDIA GTX 1650.

Os modelos produziram resultados distintos, sendo o modelo com melhor desempenho, o YOLOv4 com uma acurácia de 92%, *Recall* de 90% e um F1 de 80%. Já a

variante reduzida do YOLO, o *Fast* YOLO, obteve uma acurácia de 59% um *Recall* de 53% e um F1 de 62%. Em resumo os resultados estão listados no Quadro 2,em condições iguais de treinamento nos dois modelos.

Quadro 2 – Resultados da Pesquisa Prévia

Modelo	**Acurácia**	***Recall***	***F1-Score***	**T. Inferência**	**FPS**
YOLOv4	0,92	0,90	0,80	0,35 s	2
Fast YOLOv4	0,59	0,53	0,62	0,04 ms	24

Fonte: Elaborado pelo Autor (2025).

Os resultados do Quadro 2 mostraram diferenças significativas entre os modelos. A variante normal do YOLOv4 rendeu um melhor desempenho quanto à precisão. No entanto, seu tempo médio de inferência foi de 0,35 segundos por imagem, resultando em apenas 2 FPS. Essa taxa é insuficiente para aplicações em vídeo em tempo real, pois compromete a fluidez da exibição.

Em contraste, o *Fast* YOLOv4 apresentou desempenho inferior em termos de acurácia (59%), *recall* (53%) e F1-*score* (62%). Contudo, seu tempo de inferência foi de apenas 0,04 milissegundos por imagem, alcançando 24 FPS. Essa taxa garante fluidez na exibição dos quadros e torna o modelo adequado para aplicações em tempo real, ainda que com menor acurácia.

A pesquisa concluiu que as variantes reduzidas oferecem grande vantagem em velocidade, mas apresentam limitação de precisão, que não ultrapassava 60%, mesmo com aumento do número de amostras no treinamento. Essa constatação motivou novos experimentos, buscando superar a barreira de precisão sem perder a rapidez de inferência.

Diante do exposto, e considerando o estudo de Liang, Feng et al. (2023, p. 6), que sugere que a baixa precisão do YOLOv3 em modelos mais rápidos está associada à resolução das imagens de entrada, assim, este trabalho buscou avançar nessa direção.

Com base nessa premissa, foram conduzidos experimentos utilizando imagens em resoluções superiores a 640 x 640 *pixels*, de modo a dar continuidade à pesquisa prévia. Deste modo, o objetivo foi aplicar resoluções maiores no treinamento das variantes reduzidas do YOLOv11, na tentativa de superar a barreira de precisão observada em modelos anteriores e, ao mesmo tempo, preservar a vantagem da alta velocidade de inferência característica dessas arquiteturas.

1.7.2 Análise Comparativa dos Trabalhos Correlatos

Este estudo se diferencia dos trabalhos correlatos principalmente pelo uso exclusivo de bases públicas. Enquanto a maioria das pesquisas optou por *datasets* privados, apenas Sampaio et al. (2021) utilizou uma base pública. Ainda assim, esta pesquisa ampliou esse aspecto ao empregar três *datasets* distintos de repositórios abertos, garantindo maior diversidade e reprodutibilidade.

O Quadro 3 resume os trabalhos correlatos e, entre eles, como citado, somente a pesquisa de Sampaio et al. usou realmente uma base de dados pública e todas as outras optaram por *datasets* próprios e privados. O ponto que difere essa pesquisa que priorizou apenas bases públicas e mesmo comparada a única pesquisa que usou uma base de imagens pública, esta pesquisa utilizou três *datasets* de repositórios de domínio público.

Quadro 3 - Resumo dos trabalhos correlatos

Autor	***Dataset***	**Imagens**	**Modelo**	**Resolução**	**Precisão**	**Inferência**
Tang et al (2021)	Proprietário	84.156	RetinaNet	---	90,00%	200 ms
Sampaio et al. (2024)	Público	4.049	Faster R-CNN	320x320 640x640	37,70%	2,768 s
Alcaraz et al. (2024)	Proprietário	15.000	Algoritmo	---	98,30%	0,0057 s
Teresaki et al. (2024)	Proprietário	3.151	YOLOv5x	640x640	90%	---

Fonte: Elaborado pelo Autor (2025).

Em relação ao total de imagens, entre os trabalhos correlatos se destacou a pesquisa de Terasaki et al. (2024), com a maior quantidade de imagens celulares (84.156), sendo superado com o *dataset* final, contemplando 155.318 amostras de células cervicais. Além disso, a maioria dos *datasets* empregaram resoluções de imagens em tamanho padrão de 640 x 640 *pixels*, mas duas publicações não informaram a resolução no treinamento de seus modelos.

Em se tratando de resoluções, foram empregados nesta pesquisa imagens nas dimensões de 1280 x 1280 *pixels* e de 2016 x 2016 *pixels*, em *datasets* distintos, cujo tamanho é o dobro e um pouco mais do triplo de 640 *pixels*, o que proporcionou novos experimentos para investigar resultados em resoluções maiores que o padrão empregado nos trabalhos correlatos listados no Quadro 3.

Em relação aos modelos empregados nos trabalhos correlatos, excluindo o algoritmo da pesquisa Alcaraz et al. (2024), que apesar de ter obtido bons resultados, não empregou um CNN na detecção de imagens cervicais. Desta forma, destacou-se a publicação de Terasaki et al. (2024) empregando um modelo mais atual, o YOLOv5x, que atingiu uma

precisão de 90%, mas sem informar o tempo de inferência. Entre os modelos de CNN, o estudo de Tang et al. (2021) obteve um tempo de 200 milissegundo, superior aos tempos em segundos de outros estudos, mas ainda distante da *performance* necessária para aplicações em tempo real. Já nesta pesquisa, o YOLOv11 com variantes menores alcançou mAP de 99,30%, com tempo mínimo de 35 ms, superando os resultados anteriores tanto em precisão quanto em velocidade.

Diante do exposto, esta pesquisa não apenas se diferencia pelo uso de bases públicas e resoluções superiores, mas também demonstra avanços significativos em desempenho, consolidando o YOLOv11 como uma alternativa viável para aplicações de citologia cervical em tempo real.

1.8 ESTRUTURA DO DOCUMENTO

O presente documento está estruturado em quatro partes, incluindo esta introdução. No Capítulo 2, apresenta-se a fundamentação teórica, com o intuito de embasar os principais temas envolvidos na solução do problema. O Capítulo 3 descreve a execução do trabalho, por meio do relato dos experimentos conduzidos. Por fim, o Capítulo 4 reúne os resultados obtidos e as considerações finais acerca desta pesquisa.

2 FUNDAMENTAÇÃO TEÓRICA

Nesta seção propõe-se discutir os principais conceitos necessários para a compreensão da proposta deste trabalho, cujo foco é solucionar um problema da área de citologia cervical por meio de IA com visão computacional. Em consequência, será necessário introduzir termos e conceitos tanto da área médica quanto da detecção de imagens.

Portanto, aborda-se primeiro o universo dos exames de citologia cervical, com seus princípios e conceitos básicos para contextualizar a problemática. Em seguida, discorre-se sobre a evolução das CNNs, desde as arquiteturas de dois estágios até as mais modernas de estágio único. Essas duas áreas de conhecimento, que são abordadas nas subseções seguintes, esclarecem tanto os desafios relacionados ao exame de Papanicolau convencional quanto a justificativa para a adoção de uma CNN YOLO na detecção de imagens em tempo real.

2.1 O UNIVERSO DO DIAGNÓSTICO DE CANCER CERVICAL

Com vista a entender as metodologias envolvidas no exame de células cervicais, segue a explicação do exame de Papanicolau convencional (ou teste de Papanicolau), formalmente denominado citologia cervical. O nome foi atribuído em homenagem ao seu criador, o patologista grego George Nicholas Papanicolau (1883-1962). Esse exame foi desenvolvido na década de 1920, demonstrando que células coletadas do colo do útero e da vagina poderiam identificar lesões celulares por meio de análise microscópicas. Durante muitos anos, essa técnica se tornou padrão para o rastreamento do câncer cervical (GIRIANELLI, THULER, *et al.*, 2004, p. 1).

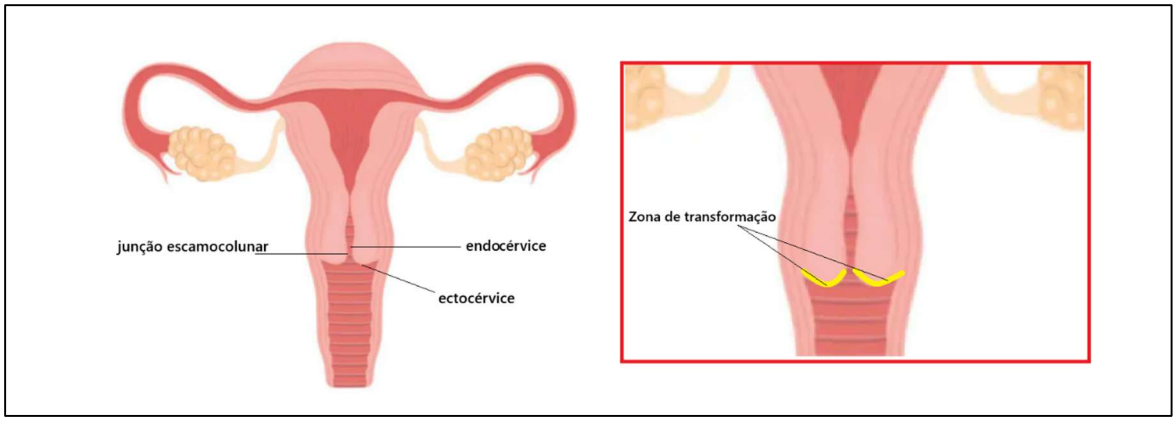
O objetivo da citologia é identificar alterações na morfologia celular, a partir de amostras obtidas do colo do útero, com foco na análise do citoplasma e do núcleo (CONSOLARO e MARIA-ENGLER, 2012, p. 4). A avaliação do citoplasma e, sobretudo, do núcleo das células é essencial para determinar se estão normais e saudáveis ou se apresentam alterações inflamatórias, pré-cancerosas e/ou cancerosas (CONSOLARO e MARIA-ENGLER, 2012, p. 7).

2.1.1 Metodologias de Citologia Cervicais

A estrutura das paredes do colo do útero, conforme ilustrado na Figura 1, na qual encontram-se duas camadas: ectocérvice a endocérvice. A ligação entre essas camadas é denominada junção escamocolunar (BARROS, LIMA, *et al.*, 2012, p. 16; KOSS e GOMPEL ,

2006, p. 35). Nessa região se encontra a Zona de Transformação (ZT), considerada a mais favorável as alterações genéticas e ao desenvolvimento de lesões celulares (CONSOLARO e MARIA-ENGLER, 2012, p. 65).

Figura 1 – Localização da Zona de Transformação no útero



Fonte: Adaptado de (SANTOS, 2025).

Na metodologia da citologia cervical convencional, o processo de coleta de células é realizado por meio de raspagem (esfregaços) de material genético retirado diretamente do colo do útero, contemplando amostras representativas da ectocérvice, endocérvice e a ZT. Por fim, as amostras obtidas pelos esfregaços são espalhadas em uma lâmina de vidro para análise microscópica (CONSOLARO e MARIA-ENGLER, 2012, p. 32; KOSS e GOMPEL , 2006, p. 16-34).

2.1.2 Limitações da Citologia Cervical Convencional

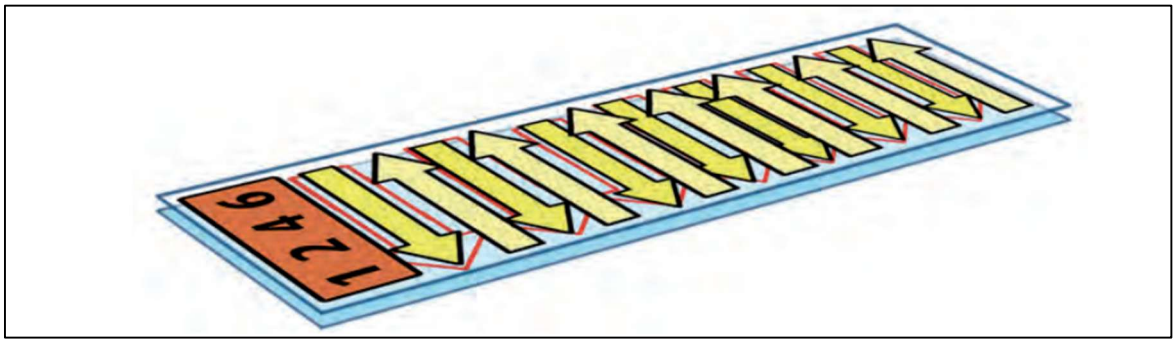
O exame de Papanicolau pode identificar diversas anormalidades, não apenas câncer, mas lesões pré-cancerosas o que possibilita um tratamento profilático para evitar sua progressão. Neste sentido, a eficácia na redução da taxa de mortalidade por CC é de 70%, demonstrando a relevância desse método (LANDY et al. 1974 *Apud* (LEE, LEE, *et al.*, 2023, p. 1-2). Contudo, apesar de sua eficiência, existem certas limitações, tais como:

- **demorado e trabalhoso**: exige que citotecnologistas ou patologistas treinados revisem manualmente um grande número de lâminas;
- **subjetivo e inconsistente**: diferentes especialistas podem ter interpretações diferentes sobre a mesma lâmina;
- **propenso a erros humanos**: como classificações incorretas, falsos negativos ou lesões não detectadas;

- **baixa sensibilidade e especificidade**: algumas anormalidades sutis ou raras podem não ser detectadas, sendo confundidas com condições benignas ou malignas;
- **outros fatores**: coleta inadequada da amostra ou presença de elementos obscurecedores, como sangue, muco, detritos e outros.

Além disso, o exame manual de lâminas oriundas de citologia cervical convencional é um processo trabalhoso, tedioso, subjetivo e sujeito a falhas (SOLOMON e NAYAR, 2024, p. 4086). Geralmente, para examinar uma única lâmina, são necessários de 5 a 10 minutos (Traut e Papanicolaou 1943 *Apud* (RAHAMAN, LI, *et al.*, 2020, p. 2697). Cada lâmina pode conter cerca de 300.000 células distintas (JANTZEN, NORUP e GEORGIOS , 2025, p. 1), com padrões de orientação e sobreposição diferentes (GENÇTAV et al., 2012 *Apud* (RAHAMAN, LI, *et al.*, 2020, p. 61688).

Figura 2- Análise de lâmina em movimento em zigue-zague



Fonte: (BARROS, LIMA, et al., 2012).

Neste contexto, considerando a dificuldade de orientação das células, recomenda-se que um citologista não analise mais de 70 amostras por dia (ELSHEIKH et al., 2013 *Apud* (RAHAMAN, LI, *et al.*, 2020, p. 61688). As células são examinadas sob microscópio para a detecção de anormalidades. Para evitar a fadiga visual, recomenda-se tecnicamente que a varredura da lâmina seja iniciada pela parte superior esquerda da lâmina, seguindo em sentido vertical com movimentos alternados (BARROS, LIMA, *et al.*, 2012, p. 32), conforme ilustrado na Figura 2.

2.1.3 O Fenômeno da Sobreposição de Células

Um dos desafios para a aplicação de IA na automação da detecção de células cervicais, no âmbito deste trabalho, refere-se à sobreposição celular. Nesta subseção, torna-se necessária uma breve descrição para esclarecer as razões que orientaram as decisões

relacionadas à definição das classes celulares nas bases de dados de imagens utilizadas para o treinamento dos algoritmos.

O fenômeno da sobreposição celular ocorre frequentemente nas metodologias de exames de Papanicolau convencional, dificultando a análise quantitativa de parâmetros como a morfologia e a densidade nuclear (SULAIMAN, ISA, *et al.*, 2010, p. 1228). Isso se deve ao fato de que o procedimento de preparação do esfregaço não pode ser padronizado, resultando na distribuição desigual das células sobre a lâmina (COX e BARBARA, 2004, p. 604).

Neste contexto, a lâmina pode conter uma única célula cervical ou um aglomerado delas em um conjunto de citoplasmas translúcidos, que podem estar dispostos aleatoriamente de forma sobreposta (BENGTSSON et al. 1981 *Apud* (JIANG, LI, *et al.*, 2023, p. 2744). Durante a análise da lâmina, é necessário observar as proporções celulares, bem como outras características relevantes, como o tamanho do núcleo e do citoplasma (SULAIMAN, ISA, *et al.*, 2010, p. 1219). Entre outros agravantes, as lesões menos extensas que sejam de baixo ou de alto grau, podem ser camufladas e não identificadas em decorrência da sobreposição (COLONELLI, 2014, p. 63).

Sulaiman (2010, p. 1218) se refere ao tema como o "problema da sobreposição" e apresenta diversos estudos de algoritmos voltados à sua solução. Atualmente, existe um concurso em formato de desafio (*Overlapping Cervical Cytology Image Segmentation Challenge*), que tem incentivado vários pesquisadores a abordar essa questão e a produzir trabalhos relevantes na área (JIANG, LI, *et al.*, 2023).

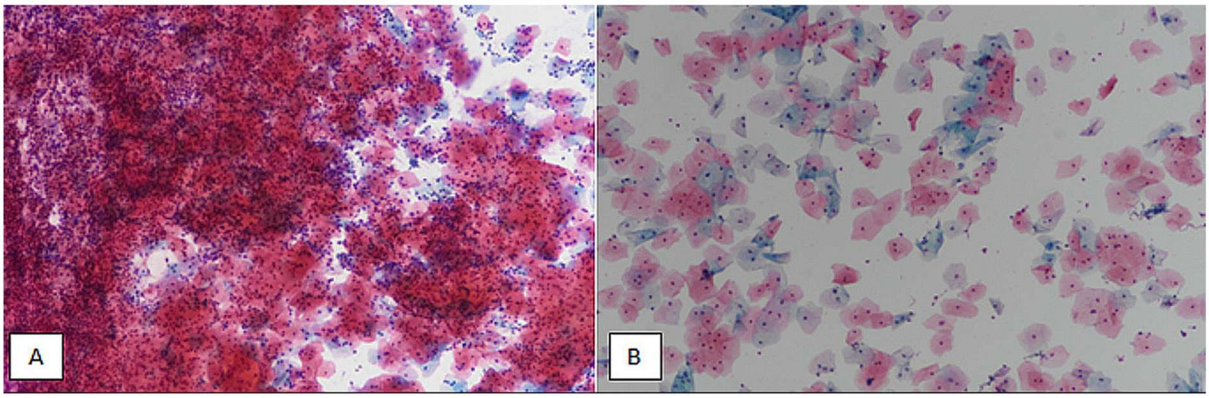
2.1.4 Outras Metodologias de Coleta de Espécimes

Na tentativa de aumentar a sensibilidade e especificidade do exame de Papanicolau convencional, patologistas dedicaram anos de trabalho para alcançar tal feito. Na década de 1950, surgiram diversas iniciativas de dispositivos de triagem cervical automatizada, com o objetivo de reduzir erros e a fadiga dos citologistas. Já na década de 1990, foi desenvolvida a Citologia em Meio Liquido (CML), que inicialmente buscava atender às demandas dos exames automatizados, minimizando os falsos negativos relacionados à análise dos espécimes. Essa metodologia obteve aprovação do *Food and Drug Administration* (FDA) em 1996 (MAKDE e SATHAWANE, 2022).

O exame de Papanicolau convencional apresenta 20% de aproveitamento do

material coletado, sendo que cerca de 80% se perde nos materiais de coleta. Em comparação, a CML alcança 100% de aproveitamento (CONSOLARO e MARIA-ENGLER, 2012). Além disso, a adoção dessa metodologia, em relação à técnica tradicional, aumenta a sensibilidade do exame citológico e diminui o grau de incidência de amostras inadequadas (CONSOLARO e MARIA-ENGLER, 2012, p. 45).

Figura 3 - (A) Papanicolau convencional demonstrando células sobrepostas com muito fundo inflamatório e hemorrágico. (B) LBC com ampliação de 100X representando células bem distribuídas



Fonte: (HASHMI, AAMIR , et al., 2020).

Apesar da CML ser realizada por meio de esfregaços, no ato de coleta do material genético a transferência para a lâmina é padronizada e auxiliada por processos automatizados. Segundo Consolaro e Maria-Engler (2012), as lâminas em CML passam por um procedimento de homogeneização celular (*Vortex*) e por um processo de enriquecimento que permite maior concentração celular, obtida por meio de centrifugação. Nesse método, a tradicional transferência do espécime por esfregaço é substituída pela fixação por carga elétrica. Como resultado, essa técnica permite a retirada de materiais indesejados nas amostras, como restos celulares, fundo hemorrágico, detritos e muco. Além de todos os benefícios, a sobreposição de células é um dos problemas eliminados pela CML. A Figura 3-A ilustra uma amostra de Papanicolau convencional com bastante sobreposição celular, comparada a uma distribuição mais homogênea das células em CML (Figura 3-B) (HASHMI, AAMIR , *et al.*, 2020, p. 7). Dessa forma, observa-se que a análise das células em CML é mais fácil e precisa.

2.1.5 Sistemas de Classificação de Lesões Celulares

Para compreender as classificações das lesões celulares presentes nos *datasets* públicos, é necessário conhecer os sistemas terminológicos empregados para classificar células normais e anormais. Isso se deve ao fato de que existem diferenças nas interpretações desses

sistemas, refletindo diretamente nas fontes de dados e nas classificações das imagens de células cervicais.

Segundo Colonelli (2014, p. 27-28), ao citar Papanicolaou e Traut (1941),

> "a classificação para o diagnóstico citológico sofreu modificações ao longo dos anos. A primeira foi proposta por Papanicolaou e Traut, porém não se generalizou como prática diagnóstica devido à baixa correlação entre os resultados citológicos e histológicos."

A interpretação de lesões celulares, lesões precursoras e carcinoma em exames citológicos cervicais tradicionalmente se baseia em diferentes terminologias derivadas de tipos celulares específicos, sobretudo aquelas presentes bancos de dados mais antigos. No entanto, para superar a inconsistência diagnóstica e clínica inerente a essas classificações, foi estabelecido em 2001 o Sistema Bethesda (*The Bethesda System*, TBS), como uma nomenclatura padronizada internacionalmente para relatar resultados de citologia cervical. Consequentemente, esse sistema é mais empregado em banco de dados mais recentes (SOLOMON e NAYAR, 2024, p. Prefácio).

O TBS classifica os diagnósticos de citologia cervical em categorias distintas, cada uma com implicações clínicas específicas. Essas categorias incluem:

- **Negativo para Lesão Intraepitelial ou Malignidade (Negative for Intraepithelial Lesion or Malignancy, NILM):** indica um resultado normal**.**
- **Células Escamosas Atípicas de Significância Indeterminada (*Atypical Squamous Cells of Undetermined Significance*, ASCUS):** denotam alterações celulares ambíguas, não claramente benignas ou pré-malignas, exigindo acompanhamento.
- **Células Escamosas Atípicas, Não Podem Excluir HSIL (*Atypical Squamous Cells, Cannot Exclude HSIL*, ASC-H)**: sugerem maior probabilidade de lesões subjacentes de alto grau.
- **Lesão Intraepitelial Escamosa de Baixo Grau (*Low Grade Squamous Intraepithelial Lesion*, LSIL):** geralmente associada à infecção por HPV, pode regredir espontaneamente, mas requer monitoramento.

- **Lesão Intraepitelial Escamosa de Alto Grau (*High-Grade Squamous Intraepithelial Lesion*, HSIL)**: apresenta risco significativo de progressão para câncer invasivo, necessitando de investigação e tratamento imediato.

- **Carcinoma de Células Escamosas (*Squamous Cell Carcinoma*, SCC):** representa um diagnóstico de câncer invasivo, exigindo intervenção terapêutica urgente (SOLOMON e NAYAR, 2024):INTRODUÇÃO).

2.2 REDES NEURAIS CONVOLUCIONAIS

A proposta deste trabalho visa à automatização da citologia cervical e do diagnóstico de CC em tempo real. Portanto, para atingir esse objetivo, faz-se necessário entender a razão da escolha e do uso da CNN YOLO. Assim, nesta seção, pretende-se explorar o funcionamento das arquiteturas da CNN de dois estágios e acompanhar os avanços que culminaram no desenvolvimento das redes convolucionais mais avançadas de um estágio, capazes de oferecer alto desempenho e rapidez na inferência em tempo real.

2.3 CNNS de DOIS ESTÁGIOS

As primeiras CNNs basicamente eram especializadas em tarefas de classificação de imagens. No entanto, a área da detecção de objetos evoluiu significativamente nas últimas décadas, consolidando-se como uma tarefa fundamental em visão computacional.

No processo de classificação em CNNs, o fluxo de processamento ocorre da seguinte forma: após a entrada dos dados, inicia-se a extração de características por meio das camadas convolucionais e de *pooling*. Em seguida, essas características são mapeadas por uma camada totalmente conectada. A partir daí, são produzidas pontuações de classificação para cada classe. A última camada, frequentemente, utiliza uma função de ativação *softmax*, que gera essas pontuações permitindo que a imagem seja classificada em sua totalidade (KRIZHEVSKY, SUTSKEVER e HINTON, 2012, p. 4).

Além da classificação, atualmente as CNNs podem ser aplicadas em tarefas de localização espacial, identificando instâncias de objetos em uma imagem e associando-as a caixas delimitadoras (*bounding boxes,* BB) (UIJLINGS, SANDE, *et al.*, 2013, p. 2). Neste estágio da evolução das CNNs, elas não apenas mantiveram a função de classificação, como também aprimoraram a localização precisa de objetos dentro da imagem, dividindo-se em duas abordagens principais:

- **Detectores de duas etapas (*two-stage*),** como R-CNN (GIRSHICK, DONAHUE, *et al.*, 2014, p. 580-587), Fast R-CNN (REN, HE, *et al.*, 2017, p. 1137-1149); e

- **Detectores de uma etapa (*one-stage*),** como YOLO (REDMON, DIVVALA, *et al.*, 2016, p. 779-788).

As CNNs de dois estágios funcionam basicamente em duas fases: inicialmente, a arquitetura identifica regiões promissoras da imagem e, em seguida, executa uma segunda etapa para decidir quais dessas regiões correspondem a objetos. Entre as arquiteturas que implementaram esse modelo de detecção em duas etapas, destacam-se a R-CNN e *Fast* R-CNN, ambas surgiram em 2015, que estabeleceram um padrão posteriormente aprimorado pelo *Faster* RCNN, aumentando a precisão geral em diversos cenários.

Além disso, uma outra reviravolta interessante ocorreu em 2017, com o surgimento da *Mask* R-CNN, que não apenas seguiu a abordagem consolidada, mas também integrou a segmentação de instâncias à tarefa de detecção de objetos. Tal combinação tornou possível lidar com ambas as tarefas simultaneamente, tornando a *Mask* R-CNN como uma ferramenta útil em aplicações avançadas, como robótica e veículos autônomos (MOHAMMED, 2025, p. 1).

No contexto da evolução das CNNs, destacam-se as tentativas que propuseram uma abordagem mais simples para a previsão de caixas delimitadoras em redes de dois estágios. Essa estratégia consistia em varrer toda a imagem por meio de técnicas como *slide window,* realizando recortes sequenciais na imagem do mesmo tamanho das caixas delimitadoras.

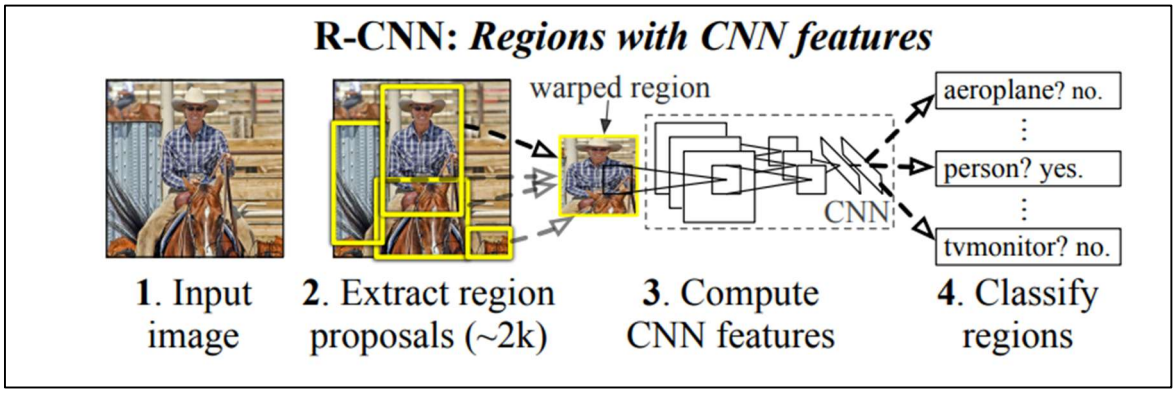
Contudo, tal abordagem implicava elevado custo computacional para uma CNN. Em contrapartida, surgiu a alternativa proposta pela R-CNN (*Regions with CNN Features*), baseada na geração de regiões específicas da imagem, de forma totalmente independente e sem relações diretas com classes predefinidas (MOHAMMED, 2025, p. 1-3).

Nessa abordagem, em vez de utilizar todos os recortes de uma imagem, a tarefa foi transferida para um método específico, responsável por identificar as regiões da imagem as quais contém um objeto. Uma vez propostas essas regiões, aplicam-se procedimentos de detecção sobre elas para efetivar a detecção dos objetos. Ou seja, o sistema combina um módulo de previsão de regiões associadas a outro módulo contendo uma CNN dedicada à classificação. Esse método foi denominado Regiões de características (GIRSHICK, DONAHUE, *et al.*, 2014,

p. 1-2).

A arquitetura da R-CNN é composta por três módulos. (Figura 4-2). O primeiro é responsável pela geração de um conjunto com cerca de 2.000 propostas de regiões (*region proposals*) de uma imagem, que servirão como possíveis candidatos disponíveis para os módulos subsequentes. (Figura 4-3). O segundo módulo corresponde a uma CNN tradicional, que recebe e extrai características de cada proposta de região individualmente. Por fim, o terceiro módulo, que consiste em um conjunto de *Suport Vector Machines* (SVMs) (Figura 4-4) lineares específicos para cada classe (GIRSHICK, DONAHUE, *et al.*, 2014, p. 2).

Figura 4 - Visão geral do sistema de detecção de objetos. (1) Primeiro, o sistema obtém uma imagem de entrada, (2) extrai cerca de 2.000 propostas de regiões (3), calcula características para cada proposta usando uma grande rede neural convolucional. (4) classificação



Fonte: (GIRSHICK, DONAHUE, et al., 2014).

Apesar de representar um marco na evolução das CNNs de dois estágios, o processamento da R-CNN no hardware disponível à época, apresentou um tempo de inferência de aproximadamente 47 segundos por imagem, o que a tornava inviável para aplicações em tempo real.

O Fast R-CNN introduziu avanços em relação ao seu predecessor, reduzindo o tempo de processamento em 47% e sendo cerca de 200 vezes mais rápida que o R-CNN original. Esse progresso foi alcançado por meio da adoção de uma abordagem unificada, na qual a imagem é processada integralmente de uma única vez por uma CNN.

Além disso, a eficiência também pode ser atribuída a implementação do mecanismo de *Region of Interest* (RoI) *Pooling*, que converte características de regiões de diferentes tamanhos em representações fixas partir do mapa global de características. Essa camada elimina a necessidade de processamento individual de cada proposta de região, permitindo que

múltiplas regiões sejam avaliadas simultaneamente (GIRSHICK, 2015, p. 3-5).

O *Faster* R-CNN representa o modelo mais avançado da família R-CNN, introduzindo inovações significativas para superar limitações de velocidade e eficiência em detecção de objetos. Esse modelo aborda especificamente o gargalo computacional associado à geração de propostas de regiões, que constituía um obstáculo crítico para aplicações em tempo real (REN, HE, *et al.*, 2017, p. 1).

Em comparação com seus predecessores, a R-CNN e a *Fast* R-CNN, que dependiam do algoritmo de busca seletiva para geração de propostas (consumindo aproximadamente 1,5–2 segundos por imagem), o *Faster* R-CNN integra uma *Region Proposal Network* (RPN) diretamente em sua arquitetura da rede neural.

A RPN compartilha características convolucionais com a rede de detecção principal, permitindo a geração de propostas de regiões de forma instantânea, com "*quase sem custo computacional*" (REN, HE, *et al.*, 2017, p. 2). Essa inovação reduziu drasticamente o tempo total de detecção, eliminando a dependência de métodos externos.

2.4 YOLO - CNN DE UM ESTÁGIO

Em 2015 foi lançada a primeira versão do algoritmo YOLO, que trata a detecção de objetos como um problema de regressão de estágio único. Nesse modelo, a rede neural prevê diretamente as coordenadas das caixas delimitadoras e as probabilidades das classes em uma única passagem pela rede. Diferentemente da R-CNN e da *Faster* R-CNN, que primeiro geram propostas de regiões para depois classifica-las, o YOLO realiza todo o processo de forma integrada, garantindo maior rapidez e eficiência (REDMON, DIVVALA, *et al.*, 2016).

No processo de detecção, a imagem de entrada é passada diretamente como uma única CNN, que atua como *backbone* para extrair características hierárquicas da imagem, capturando desde padrões de baixo nível (como bordas e texturas) até características de alto nível (como formas mais complexas e contextos semânticos). Após a extração de características, a rede realiza duas tarefas principais:

1) prever as coordenadas das caixas delimitadoras para os objetos detectados; e

2) atribuir probabilidades de classes a cada uma dessas caixas.

Essa arquitetura unificada permite que o YOLO alcance alta velocidade de inferência, processando imagens em tempo real (até 45 *frames* por segundo em configurações padrão) (REDMON, DIVVALA, *et al.*, 2016, p. 2-3).

Um mecanismo fundamental do YOLO se baseia na divisão da imagem em uma grade de células. Na implementação original, os autores utilizam uma grade de dimensões $S \times S$, onde $S = 7$. Cada célula dessa grade é responsável por destacar objetos cujos centros se encontram dentro dela. Por exemplo, em uma imagem contendo dois objetos, uma pessoa e um carro, a célula que contém o centro da pessoa (definido pela caixa delimitadora da verdade absoluta) será designada para prever esse objeto, enquanto a célula contendo o centro do carro será responsável por sua detecção. Essa atribuição direta simplifica o treinamento e reduz ambiguidades (REDMON, DIVVALA, *et al.*, 2016, p. 2).

Para cada célula da grade, o YOLO prevê B caixas delimitadoras (sendo $B = 2$ na implementação original). Cada caixa é representada por cinco valores: coordenadas (x, y) do centro da caixa (relativas aos limites da célula), largura(w) e altura (h) (normalizadas em relação as dimensões da imagem), além de uma pontuação de confiança que reflete tanto a probabilidade de a caixa conter um objeto quanto a precisão de sua localização. Adicionalmente, cada célula prevê probabilidades de classe para os objetos detectados, independentemente do número de caixas delimitadoras geradas (REDMON, DIVVALA, *et al.*, 2016, p. 2-3).

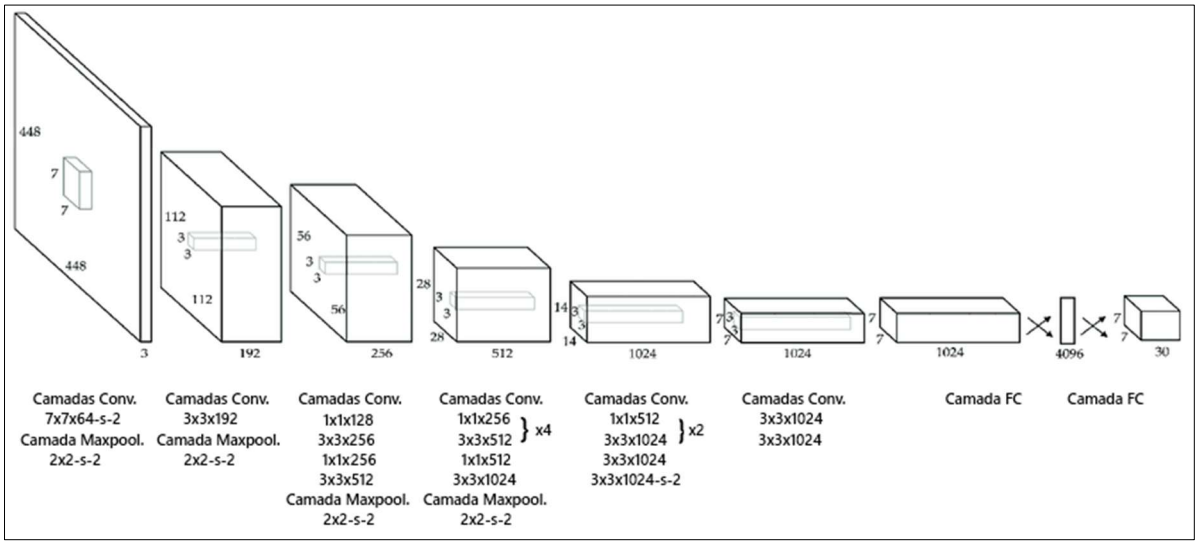
Durante o treinamento, o YOLO aprende a ajustar suas previsões de modo que as caixas das células designadas se aproximem ao máximo das caixas de verdade absoluta. Esse processo ocorre por meio da minimização de uma função de perda composta, que combina:

- **Erro de localização:** calculado com base no erro quadrático para coordenadas das caixas.
- **Erro de classificação:** baseado na entropia cruzada das probabilidades de classe).
- **Termos adicionais:** penalizam falsos positivos e reforçam a confiança em previsões precisas.

Após o treinamento, na fase de inferência, as previsões são refinadas por meio das

técnicas de Supressão Não-Máxima (*Non-Max Suppression*, NMS), que elimina caixas redundantes ou sobrepostas, resultando nas detecções finais (GIRSHICK, 2015, p. 4).

Figura 5 – A arquitetura da rede YOLO



Fonte: Adaptada pelo autor de (REDMON, DIVVALA, et al., 2016).

A arquitetura do YOLO, em sua primeira versão (YOLOv1), consistia em 24 camadas convolucionais seguidas por 2 camadas totalmente conectadas, utilizando funções de ativação *Leak* ReLU e técnicas regularização, como *batch normalization e dropout* (REN, HE, *et al.*, 2017, p. 3). Sua eficiência e velocidade revolucionaram o campo da detecção de objetos em tempo real, influenciando gerações subsequentes de modelos (como YOLOv2 até YOLOv12) (What is YOLO? The Ultimate Guide, 2025; TIAN, YE e DOERMANN, 2025).

A implementação do modelo como uma rede CNN ocorre de modo que as camadas convolucionais iniciais extraem características da imagem, enquanto as camadas totalmente conectadas são responsáveis por prever as probabilidades e coordenadas de saída. O *backbone* original é uma adaptação da *GoogleNet* (*Inceptio* v1), composto por 24 camadas convolucionais seguidas por 2 camadas totalmente conectadas. No entanto, em vez dos módulos iniciais utilizados pela *GoogLeNet*, foram empregadas camadas de redução 1 x 1 seguidas por três camadas convolucionais 3 x 3. A arquitetura completa pode ser observada na Figura 5 (REN, HE, *et al.*, 2015, p. 3).

O YOLO introduziu uma abordagem distinta para a predição de caixas delimitadoras na detecção de objetos, em contraste com métodos baseados em propostas de regiões, como a R-CNN. Em sua arquitetura, o algoritmo realiza todas as previsões em uma única passagem pela rede neural, otimizando simultaneamente a localização e a classificação

(REDMON, DIVVALA, *et al.*, 2016, p. 1).

Para cada caixa, o YOLO prevê cinco parâmetros fundamentais:

1. Coordenadas do centro (*offset* de *x* e *y*), relativas à posição dentro da célula.
2. Largura (*w*), normalizada em relação às dimensões da imagem.
3. Altura (*h*), também normalizada em relação às dimensões da imagem.
4. Pontuação de confiança (*confidence score*), que reflete a probabilidade de a caixa conter um objeto e a precisão de sua localização.

Esses parâmetros são calculados de forma normalizada em relação à imagem inteira, garantindo invariância escalar (GIRSHICK, 2015, p. 2).

As coordenadas do centro da caixa delimitadora são representadas como *offsets* em relação ao canto superior esquerdo da célula da grade cujo a caixa pertence. Especificamente, para uma grade de dimensão *S x S*, o centro (b_x, b_y) é calculado conforme Equação 1 e 2.

Se os valores de *offset* forem (0, 0), o centro da caixa delimitadora estará localizado no canto superior esquerdo da célula; se forem (1, 1), estará no canto inferior direito. Essa parametrização permite uma localização precisa mesmo quando o objeto se estende por múltiplas células (GIRSHICK, 2015, p. 3).

$$b_x = \sigma(t_x) + c_x \tag{1}$$

$$b_y = \sigma(t_y) + c_y \tag{2}$$

Onde:

- c_x, c_y representam as coordenadas do canto superior esquerdo da célula na grade;
- t_x, t_y são os *offsets* previstos pela rede, normalizados entre 0 e 1 por meio da função sigmoide (σ).

Outro complemento das coordenadas da caixa delimitadora contempla a largura (b_w) e a altura (b_h), que são previstas de forma normalizada em relação às dimensões da imagem original, variando entre 0 e 1. Por exemplo, um valor 1 para a largura indica que a caixa possui a mesma largura que a imagem inteira.

Já a pontuação de confiança em modelos de detecção de objetos encapsula duas

informações fundamentais para a avaliação da qualidade da predição:

1. A **probabilidade da presença de um objeto** dentro da caixa delimitadora proposta;
2. A **precisão da localização** do objeto, quantificada por meio da métrica IoU (*Intersection over Union*), que mede a sobreposição entre a caixa prevista e a caixa de verdade absoluta.

Matematicamente, o valor *target* do escore de confiança durante o treinamento é definido pela expressão:

$$C = \Pr(Objeto) \ . IoU_{pred}^{verdade} \tag{3}$$

Onde:

- $\Pr(Objeto)$ =1 se a célula da grade em análise contém um objeto, e 0 caso contrário;
- $IoU_{pred}^{verdade}$ representa o valor de IoU entre a caixa prevista e a anotação verdadeira.

Dessa forma, *C* assume valores no intervalo [0,1], sendo igual a zero quando não há objeto presente ou quando a IoU é baixa, e aproximando-se de 1 quando um objeto está presente e a sua localização é precisa (GIRSHICK, 2015, p. 3). Essa pontuação desempenha um papel crítico na etapa de pós-processamento, particularmente no algoritmo de NMS, onde é utilizada para eliminar predições redundantes ou de baixa confiança, priorizando caixas com maior sobreposição e maior certeza de detecção (GIRSHICK, 2015, p. 3).

2.5 YOLOv11

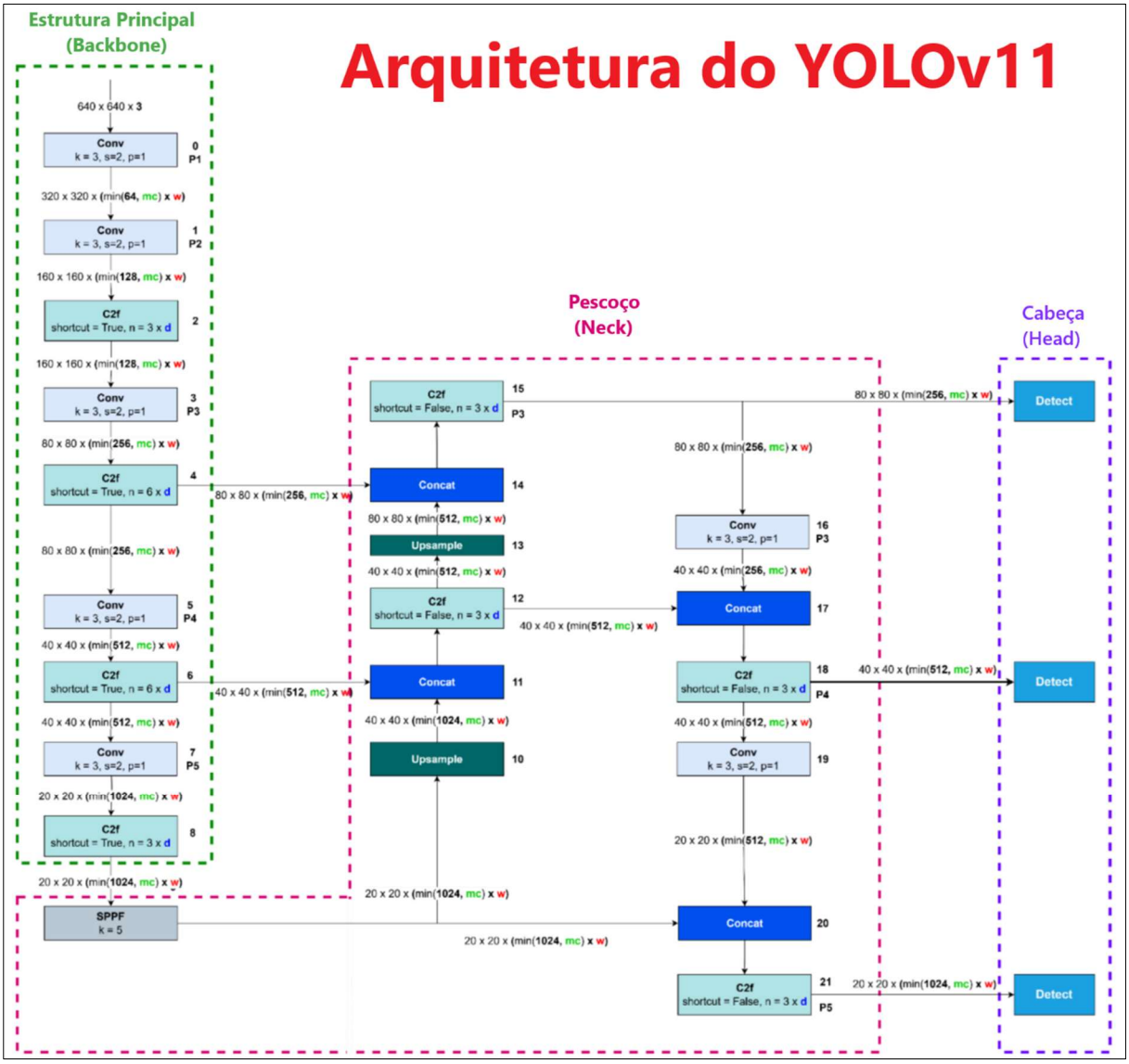
O YOLOv12 é a mais recente evolução da série YOLO, introduzindo um módulo de Atenção de Área (A2). Além disso, ele apresenta Redes de Agregação de Camadas Residuais Eficientes (R-ELAN), que melhoram a estabilidade do treinamento. Entre outras novidades, o YOLO12 incorpora também o *FlashAttention,* que minimiza a sobrecarga de acesso à memória e elimina a lacuna de velocidade em relação às CNNs (NIDHAL , KOH, *et al.*, 2024, p. 12).

Apesar do YOLOv12 ser a versão mais atual, a versão anterior, YOLOv11, apresenta desempenho semelhante à nova versão. Segundo a comparação realizada por Nidhal, Koh, Abdelatti e Hendawi (2024, p. 16), entre todas as versões do YOLO submetidas a diversos *datasets* em diferentes situações, o YOLOv11 mostrou-se um modelo eficiente, mesmo em comparação com a versão mais nova, apresentando menor custo computacional (NIDHAL ,

KOH, *et al.*, 2024, p. 20).

O *design* do YOLOv11 incorpora técnicas avançadas de extração de características, permitindo a captura de mais detalhes e a redução do número de parâmetros. Isso resulta em maior precisão, suporte a um conjunto mais amplo de tarefas de visão computacional e ganhos significativos em velocidade e processamento, aprimorando substancialmente as capacidades de desempenho em tempo real (KHANAM e HUSSAIN, 2024, p. 2). Em sua essência, a arquitetura do YOLOv11 consiste em três componentes fundamentais (Figura 6): estrutura principal (*backbone*), o pescoço (*neck*) e a cabeça (*head*).

Figura 6 – A arquitetura do YOLOv11



Fonte: Adaptada pelo autor de (ULTRALYTICS, 2024).

O primeiro componente, a estrutura principal atua como o extrator de características, utilizando redes neurais convolucionais para transformar dados brutos das imagens. Trata-se de um elemento crucial da arquitetura YOLO, responsável por extrair informações da imagem de entrada em múltiplas escalas. Esse processo envolve o empilhamento de camadas convolucionais e blocos especializados, que geraram mapas de características em diferentes resoluções (KHANAM e HUSSAIN, 2024, p. 3).

O segundo componente, o pescoço, atua como um estágio intermediário de processamento, empregando camadas especializadas para agregar e aprimorar representações de características em diferentes escalas. O pescoço combina essas representações multiescala e as transmite à cabeça para a etapa de previsão. Esse processo normalmente envolve operações de aumento das amostras (*upsampling*) e concatenação de mapas de características de diferentes níveis, permitindo que o modelo capture informações multiescala de forma eficaz (KHANAM e HUSSAIN, 2024, p. 3).

O Terceiro componente, a cabeça, funciona como o mecanismo de predição, gerando as saídas finais para localização e classificação de objetos com base nos mapas de características refinados. A cabeça do YOLOv11 é responsável por produzir as previsões finais em termos de detecção e classificação de objetos. Ela processa os mapas de características transmitidos pelo pescoço, gerando caixas delimitadoras e rótulos de classe para os objetos presentes na imagem (KHANAM e HUSSAIN, 2024, p. 3).

Uma melhoria significativa no YOLOv11 é a introdução do bloco C3k2, que substitui o bloco C2f utilizando em versões anteriores. O bloco C3k2 é uma implementação computacionalmente mais eficiente do gargalo *Cross Stage Partial* (CSP). Ele emprega duas convoluções menores em vez de uma convolução maior. O sufixo "k2" em C3k2 indica o uso de um *kernel* reduzido, o que contribui para um processamento mais rápido, mantendo o desempenho (KHANAM e HUSSAIN, 2024, p. 3).

O YOLOv11 mantém o bloco *Spatial Pyramid Pooling* - Fast (SPPF) das versões anteriores, mas introduz um novo bloco CSP com Atenção Espacial (C2PSA) após dele. O bloco C2PSA é uma adição relevante que aprimora a atenção espacial nos mapas de características. Esse mecanismo de atenção espacial nos mapas de características, o C2PSA possibilita que o YOLLOv11 foque em áreas específicas de interesse, potencialmente melhorando a precisão da detecção para objetos de diferentes tamanhos e posições (KHANAM

e HUSSAIN, 2024, p. 4).

Além disso, uma adição relevante ao YOLOv11 é o maior foco na atenção espacial por meio do módulo C2PSA. Esse mecanismo de atenção permite que o modelo se concentre em regiões-chave da imagem, potencialmente resultando em uma detecção mais precisa, especialmente para objetos menores ou parcialmente ocluídos. A inclusão do C2PSA diferencia o YOLOv8, que não possui esse mecanismo específico de atenção especial (KHANAM e HUSSAIN, 2024, p. 4).

Uma característica interessante do YOLO surgiu a partir do TOLOv5, marcada pelo fato de cada versão incluir variações da sua própria arquitetura em diferentes escalas, como por exemplo: YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l e YOLOv5x. Estes sufixos indicam o tamanho e a complexidade do modelo, representando a escala da estrutura, como: “n” para *nano*, “s” para *small*, “m" para médio (*medium*), “l” para grande (*large*) e “x” para extragrande (*extra large*) (NIDHAL , KOH, *et al.*, 2024, p. 9). Assim como as outras versões, o YOLOv11 também oferece essas variantes em diferentes escalas.

3 EXPERIMENTOS COMPUTACIONAIS

Nesta seção, foram descritos os processos realizados para alcançar os objetivos deste trabalho. Para isso, empregaram-se modelos de visão computacional integrados a câmeras digitais acopladas a microscópios, permitindo a análise automatizada de imagens citológicas em tempo real.

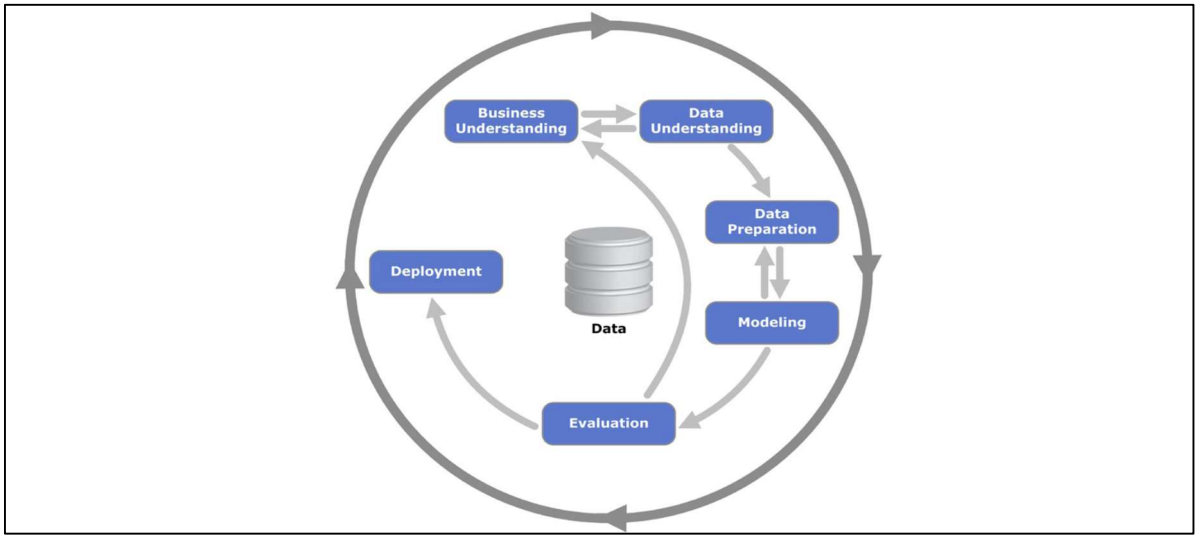
O escopo deste capítulo abrange a descrição das atividades planejadas, desde a compreensão do problema e dos dados até a seleção dos equipamentos, algoritmos IA e construção dos *datasets*.

É importante destacar que toda a proposta foi elaborada com base na primeira iteração de um protótipo, cujo propósito foi viabilizar o principal objetivo deste trabalho: avaliar a viabilidade e a reprodutibilidade da abordagem proposta.

3.1 METODOLOGIA CRISP-DM

Desde o desenvolvimento até a avaliação do sistema de detecção de células cervicais seguem a metodologia CRISP-DM, constituída como padrão de referência em projetos de Ciência de Dados. Essa metodologia assegura uma abordagem estruturada e cíclica, em seis fases ilustradas na Figura 7: Compreensão de Negócios, Compreensão de Dados, Preparação de Dados, Modelagem, Avaliação e Implantação (SEKAR, 2022, p. 44; CHAPMAN, CLINTON, *et al.*, 2000).

Figura 7 – Diagrama da estrutura do CRISP-DM



Fonte: Sekar (2022, p. 44) e Chapman e outros autores (2000, p. 13).

A literatura padroniza as etapas de cada fase por meio de um *roadmap*, o qual as organiza em tarefas cujos artefatos genéricos estão listados na Figura 8. A aplicação prática dessas fases foi detalha nesta Seção, adaptada ao contexto da citologia cervical convencional.

Figura 8 – CRISP-DM Tarefas genéricas e seus artefatos

Fonte: Adaptado de (CHAPMAN, CLINTON, *et al.*, 2000, p. 15) (tradução nossa).

3.1.1 Primeira Fase de Entendimento do Negócio

Seguindo as etapas do *roadmap* proposto pelo CRISP-DM, a primeira fase será dedicada à conversão dos objetivos de pesquisa em um plano de projeto de *Deep Learning*. O foco consistiu na proposição de uma solução acessível e de baixo custo, baseada na arquitetura YOLLOv11 de baixa latência, com o objetivo de mitigar resultados de falsos negativos no exame Papanicolau.

Quanto aos critérios de sucesso foram definidos considerando o alcance de alta precisão (mAP >= 60%) e elevada velocidade de inferência (FPS), ambos essenciais para viabilizar a detecção em tempo real em ambientes de computação em borda.

Para a avaliação da situação e do inventário de recursos, foi necessária a aquisição de equipamentos auxiliares, como câmeras digitais USB CMOS adaptadas a um microscópio *Olympus* BX53. Em seguida, os experimentos foram conduzidos em um computador equipado com processador *Intel(R) Core(™)* i5-11400H CPU @ 2.7GHz, com placa de vídeo *GeForce* GTX 1650 de 4 GB, memória RAM de 64GB e disco SSD M.2 de 1 TB, utilizando o sistema operacional *Microsoft Windows* 11, linguagem *Python* versão 3.9 e biblioteca gráfica OpenCV.

Assimilando o conhecimento adquirido e com o intuito de responder às questões de pesquisa (seção 1.3), esta fase planejou os experimentos listados no Quadro 6 – Listagem dos experimentos de treinamento.

Quadro 4 – Listagem dos nove experimentos do plano de projeto inicial

Questão de pesquisa	Experimento
Q01:É possível utilizar apenas bases publicas para produzir um modelo treinado de IA em visão computacional para detecção de células cervicais?	**Exp. 01**: Realizar testes de desempenho das variantes YOLO (n) e (s) em tarefas de detecção por caixas delimitadores e segmentação semântica.
	Exp. 02: Realizar teste de inferência em um banco de dados externo em ambos as variantes do YOLO (n) e (s), tanto em detecção por caixas delimitadoras quanto segmentação semântica.
	Exp. 03: Realizar teste de inferência em vídeo em tempo real provido tanto por câmeras CMOS montadas no microscópio quanto em vídeos de domínio público.
Q02:A partir de apenas recursos de domínio público, é possível propor uma solução de automatização de citologia cervical alternativa que possa ser totalmente reproduzível por outros pesquisadores?	**Exp. 04**: Eleger apenas bases de dados de células cervicais de domínio público.
	Exp. 05: Contabilizar as classes em diferentes bases de dados e mitigar o enviesamento.
	Exp. 06: Unificar os diferentes tipos de classificação celulares em diferentes bases de dados antes de unifica-las em um único *dataset*.
	Exp. 07: Aplicar técnicas de *Data Augmentation* para balancear possíveis enviesamento nas amostras de imagens e diversificar o total de imagens final.
Q03:Arquiteturas de CNNs com variantes reduzidas podem melhorar o desempenho de treinamento aumentanto a precisão com implementação de imagens com dimensões superior a 640 x 640 *pixels*?	**Exp. 08**: Converter as imagens unificadas em dois *dataset* com resoluções de 2016 x 2016 e 1280 x 1280 pixels respectivamente.
	Exp. 09: Treinamento e inferência das variantes do YOLOv11 com arquiteturas reduzidas, *Nano* (n) e *Small* (s) submetidos a dois *datasets* com resoluções diferentes.

Fonte: Elaborado pela Autor (2025).

Quanto à última tarefa desta fase, referente à avaliação inicial e às técnicas, existem várias ferramentas disponíveis no mercado para essas atividades (SCHRÖER, KRUSE e GÓMEZ, 2021, p. 529), incluindo aquelas voltadas à documentação do processo CRISP-DM (MARTÍNEZ-PLUMED, CONTRERAS-OCHANDO, *et al.*, 2021, p. 3050). Contudo, foram priorizadas ferramentas sem custo, como a linguagem *Python* na distribuição *Anaconda* e YOLOv11 disponibilizado pela *Ultralytics*.

Os experimentos culminam em um objetivo comum, que é treinar modelos de visão computacional usando uma CNN YOLO, cujo resultado possa identificar lesões cervicais. Os critérios desejáveis de identificação dessas lesões por parte desses modelos de IA, foram associados aos mesmos utilizados em citologia cervical convencional, que por sua vez, são baseados na identificação de padrões por meio de análise minuciosa das alterações morfológicas celulares (KOSS e GOMPEL , 2006, p. 61-75; SOLOMON e NAYAR, 2024, p. 68-90), tais como:

- modificação nucleares evidenciadas por coloração;
- hipercromasia (escurecimento);
- aumento da relação núcleo/citoplasma;
- irregularidade da membrana nuclear; cromatina grosseira;
- e presença de núcleos anormais.

3.1.2 Segunda Fase de Entendimento dos dados

Esta fase está diretamente correlaciona a "Coleta dos dados", que será iniciada com a aquisição das bases de dados contendo imagens de células cervicais de domínio públicos para *download*. Para as tarefas de "Descrever dados" e "Explorar dados" (CHAPMAN, CLINTON, *et al.*, 2000, p. 19) foram unificadas em um único processo, uma vez que os dados se encontram em arquivos de imagens. Por fim, a atividade de "Verificar a qualidade dos dados" consiste na preparação de um Relatório de qualidade, confeccionado a partir da listagem dos nomes dos arquivos após a checagem manual de cada um deles.

Nesta fase, direcionou-se a pesquisa e seleção das imagens iniciais com o objetivo de desenvolver uma base de dados composta por células normais e atípicas. Para isso, tornou-se necessária a consulta a publicações acadêmica, a fim de avaliar as principais bases de dados públicas e seus respectivos repositórios. Após essa análise, foram selecionadas três bases públicas para compor um *dataset* único de imagens: *SipakMed*, CRIC *Cervix* e RepoMedUNM.

Tabela 1 - SIPaKMeD *dataset*, distribuição das classes

Tipo de células	**Imagens**	**Qtd. Células**
Superficial/Intermediária	126	813
Parabasal	108	787
Coilocitótico	238	825
Metaplástica	271	793
Disqueratótico	223	813
Total	**966**	**4049**

Fonte: Elaborado pela Autor (2025).

SiPaKMeD *dataset*: Os dados que compõe estes *dataset* foram coletados a partir de lâminas convencionais de Papanicolau, utilizando uma câmera digital CCD (*Infinity 1 Lumenera*) acoplada a um microscópio *Olympus* BX53F. O conjunto de dados reúne 4.049 amostras de células isoladas em 966 arquivos de imagens no formato BMP, com a resolução de

2048 x 1536 *pixels*.

Este *dataset* é notavelmente classificado por tipos celulares. Portanto, as células foram distribuídas em cinco classes distintas: duas classes de células anormais não malignas (Coilocitóticas e Disqueratóticas) e uma classe de célula benigna (Metaplásicas). A distribuição dessas células podem ser observada na Tabela 1 (PLISSITI, DIMITRAKOPOULOS, *et al.*, 2018, p. 3144; RAHAMAN, LI, *et al.*, 2021, p. 11-14).

CRIC Cervix: Trata-se de um conjunto de dados do colo do útero que reúne uma coleção de imagens convencionais de lâminas de Papanicolau, desenvolvido como iniciativa da plataforma do Centro de Reconhecimento e Inspeção de Células. O *dataset* totaliza 400 imagens, com resolução de 1.376 x 1.020 *pixels*, contendo coletivamente 11.534 células. As células são classificadas de acordo com o TBS em seis classes distintas: NILM, ASC-US, LSIL, HSIL, ASC-H e SCC. A Tabela 2 apresenta a distribuição dessas células no conjunto de dados. A principal contribuição deste conjunto de dados concentra-se em duas áreas:

1. Fornecer um recurso projetado especificamente para o treinamento de modelos de detecção de células cervicais e
2. Disponibilizar amostras de imagens de citologia cervical convencional que contemplam situações desafiadoras do mundo real, como aglomerados de células sobrepostas (REZENDE, BIANCHI, *et al.*, 2021, p. 1-6; FERREIRA, RAMALHO, *et al.*, 2019, p. 2-3).

Tabela 2 - CRIC Cervix *dataset*, distribuição das classes

Classes	**Qtd. Células**
NILM	6,779
ASC-US	606
ASC-H	925
LSIL	1,360
HSIL	1,703
SCC	161
Total	**11,534**

Fonte: Elaborado pelo Autor (2025).

RepoMedUNM: Este conjunto de imagens foi desenvolvido especificamente para avaliar separação de células sobrepostas. Todas as imagens são hospedadas no Repositório de Processamento de Imagens Médicas da Universidade Nusa Mandiri (RepoMedUNM). O conjunto de dados consiste em pares de imagens, cada uma imagem armazenada em formato JPG, com resolução de 512 x 512 *pixels*.

Tabela 3 - RepoMedUNM *dataset*, distribuição das classes

Classes	**Qtd. Células**
NILM	210
KOILOCYTOTIC	210
LSIL	210
HSIL	210
Total	**840**

Fonte: Elaborado pela Autor (2025).

As imagens são categorizadas em quatro classes: NILM, LSIL, HSIL e KOILOCYTOTIC. Essas quatro classes totalizam 420 arquivos de imagem (Tabela 3), cada um contendo duas células sobrepostas, resultando em 840 amostras de células individuais.

Além das imagens originais, o repositório também disponibiliza versões em escala de cinza que funcionam como máscaras de referência (*ground truth*) para experimentos de separação de células (RIANA, JAMIL, *et al.*, p. 814-817). Para os propósitos desta pesquisa, a classe KOILOCYTOTIC deste conjunto de dados foi excluída da composição do *dataset* final.

APACC: Foi selecionado e separado um quarto conjunto de dados distinto, utilizado exclusivamente nesta pesquisa para testes de inferência. Esse conjunto de dados, denominado Imagens de Células Papanicolau Anotadas e Fatias de Esfregaço para Classificação Celular, é público e originou-se de um projeto de pesquisa colaborativa entre o Departamento de Patologia do Centro Clínico e a Faculdade de Informática da Universidade de Debrecen.

As imagens do *dataset* APACC foram adquiridas com o *scanner 3DHistech Pannoramic* 1000. A partir desse equipamento, uma única imagem de alta resolução foi capturada (100.000 x 220.000 *pixels*) e posteriormente dividida em 107 arquivos menores, cada um com resolução de 2000 x 2000 *pixels*, em um nível de ampliação de 200x.

O conjunto totalizaou 103.675 células cervicais, classificadas em quatro classes: saudável (*health*), não saudável (*unhealth*), ambas as células (*bothcells*) e lixo (*rubbish*). Esse conjunto de dados foi projetado exclusivamente para o treinamento de modelos de IA. Uma vantagem fundamental é o fornecimento de imagens originais acompanhadas das caixas delimitadoras de verdade básica (BB), essenciais para o treinamento e avaliação de algoritmos de detecção de objetos em citologia cervical (KUPAS, HAJDU, *et al.*, 2023, p. 2-6).

3.1.3 Terceira Fase de Preparação dos dados

Considerando que a fase de "Preparação de dados" é a mais crítica e longa da metodologia, o registro das tarefas-chave inclui:

1. **Pré-processamento das imagens**: visa a unificação das dimensões das diferentes bases de dados públicas. O objetivo principal dessa etapa é padronizar todas as imagens com única resolução, compondo os *datasets* de treinamento;
2. **Anotações de imagens:** tem como finalidade rotular os dados visuais para o treinamento em tarefa de segmentação de instâncias;
3. **Aplicação de técnicas de *Data Augmentation***: busca balancear e diversificar o conjunto de imagens no *dataset* final.

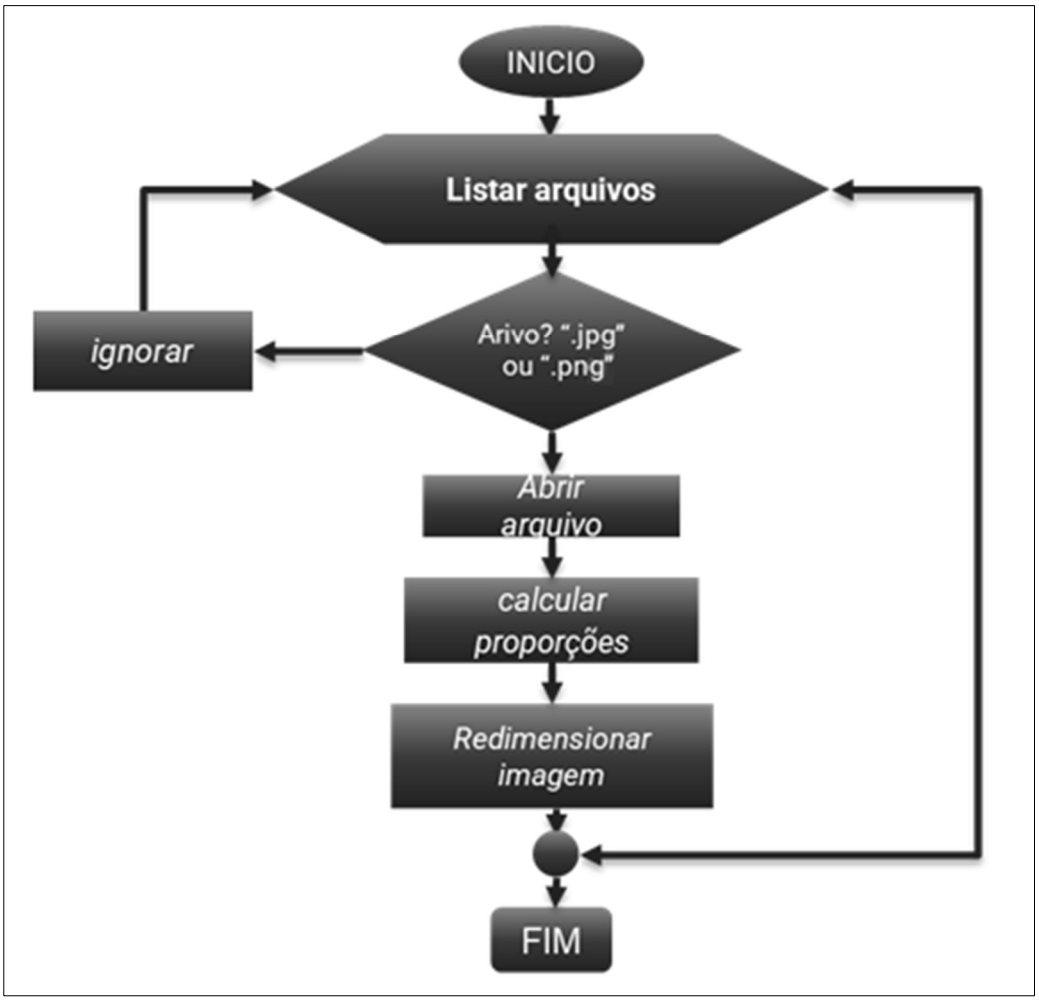
A proposta é seguir a mesma ordem das atividades do CRISP-DM (CHAPMAN, CLINTON, *et al.*, 2000, p. 25 e 65), utilizando as ferramentas *LabelMe*, *Roboflow* e linguagem *Python*. Ademais, em conformidade com as orientações da atividade de "Limpar dados" (CHAPMAN, CLINTON, *et al.*, 2000, p. 49), foram executados procedimentos adicionais após a unificação do *dataset* final, consistindo na exclusão de imagem com baixa resolução ou qualidade.

3.1.3.1 Pré-processamento

Durante a fase de unificação das imagens, diversas questões críticas demandaram cuidadosa análise. O principal desafio esteve relacionado às diferentes dimensões de imagem presentes nos conjuntos de dados. Para superar essa limitação, as imagens foram redimensionadas de forma uniforme em duas resoluções distintas: 1280 x 1280 e 2016 x 2016 *pixels*, formando assim dois conjuntos de dados de treinamento independentes.

Ainda neste contexto, destaca-se a razão pela qual foi estabelecida a resolução de 1280 x 1280 *pixels* como dimensão das imagens de treinamento. Tal escolha se deve ao fato de ser o dobro da resolução de 640 x 640, considerada padrão e amplamente utilizada em treinamentos de modelos de detecção. A adoção de dimensões maiores buscou avaliar o impacto da resolução na precisão dos modelos, mantendo proporcionalidade com o padrão mais comum.

Figura 9 – Fluxograma de conversão das dimensões das imagens

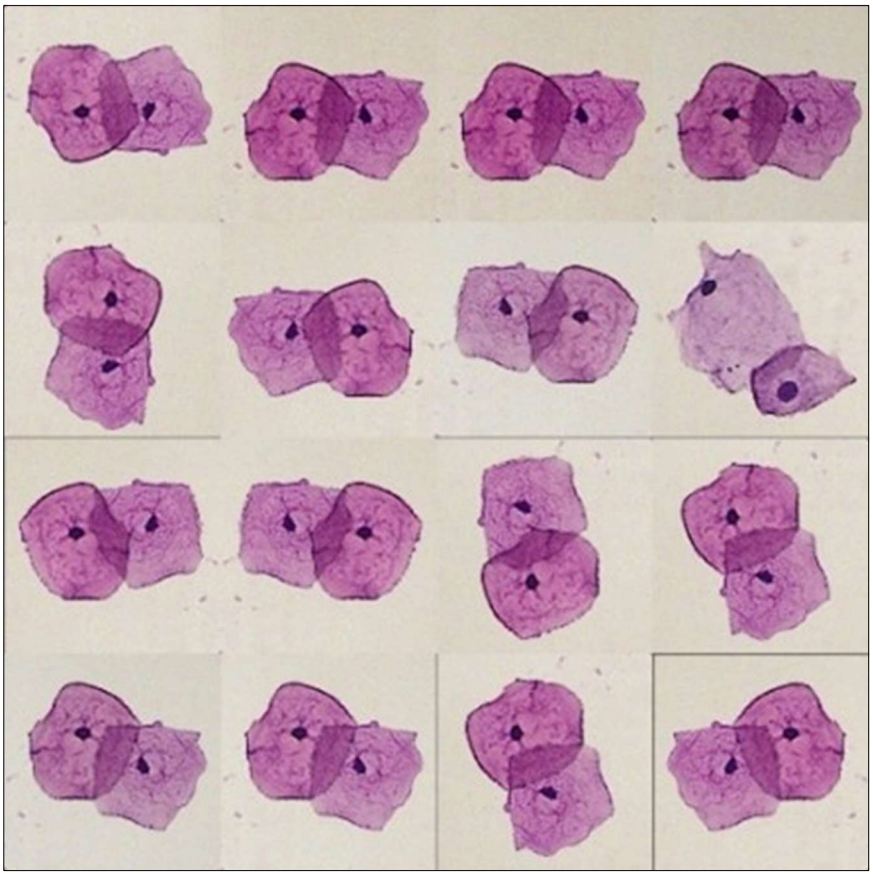


Fonte: Elaborado pelo Autor (2025).

Já para o segundo *dataset*, inicialmente foi considerada a resolução 2000 x 2000 *pixels*. Contudo, durante o processo de treinamento, o YOLO rejeitou essas dimensões, uma vez que sua arquitetura exige que as imagens sejam múltiplas de 32. Por essa razão, as imagens foram redefinidas para 2016 x 2016 *pixels*, valor que atende ao requisito estrutural do modelo e, ao mesmo tempo, mantém a proposta de explorar resoluções superiores ao padrão tradicional.

O processo de conversão dos arquivos, apresentado na Figura 9, foi implementado na linguagem *Python* com o auxílio da biblioteca OpenCV. De forma geral, o procedimento inicia-se com a abertura da pasta que contém as imagens, as quais são lidas por meio de um laço de repetição. Em seguida, são calculadas as dimensões da nova imagem preservando suas proporções originais. Por fim, realiza-se o redimensionamento, que pode ser tanto 1280 x 1280 quanto 2016 x 2016 *pixels*, conforme os parâmetros definidos.

Figura 10 –RepoMedUNM, imagem consolidada a partir de outros 16 arquivos



Fonte: elaborado pelo Autor (2025).

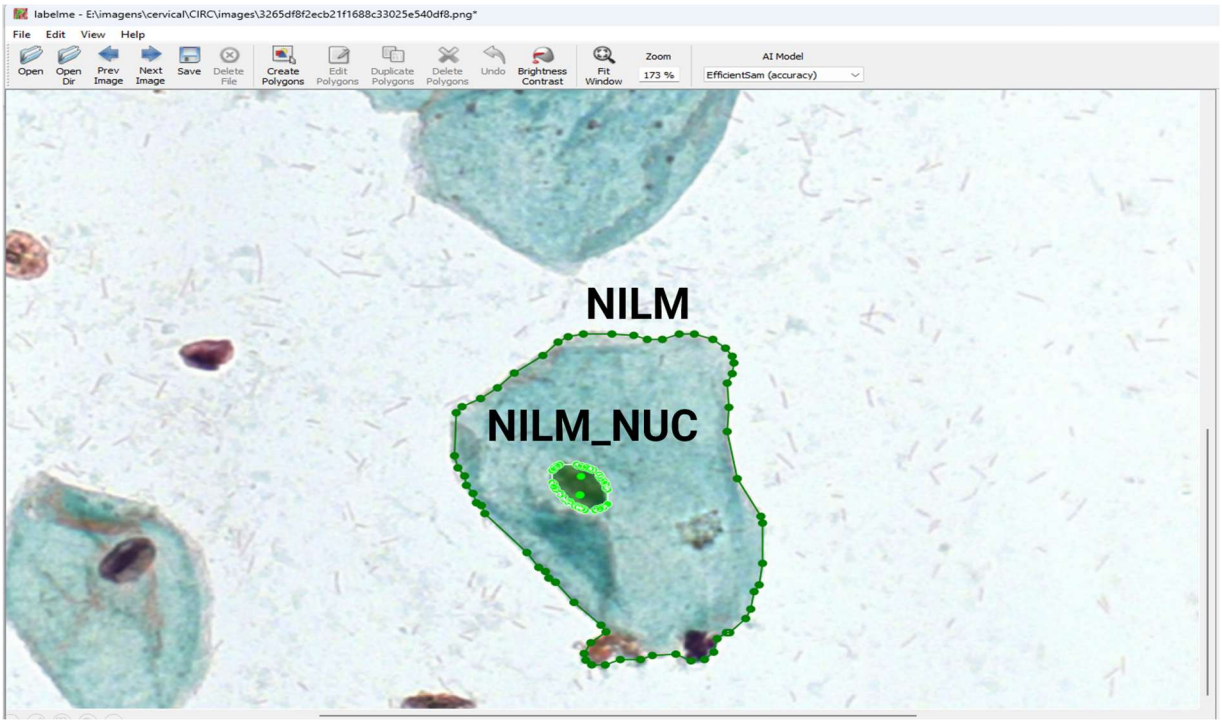
Para amostras de imagens pequenas e isoladas, como as que estão em presente no conjunto de dados RepoMedUNM, foi aplicada uma etapa específica de pré-processamento (Figura 10). Nessa etapa, utilizou-se o *Microsoft Paint* para consolidar grupos de dezesseis imagens individuais em um único arquivo composto. Essa estratégia mostrou-se particularmente vantajosa para conjuntos de dados de treinamento maiores (superiores a 640 x 640 *pixels*), pois otimizou significativamente o processo de treinamento.

3.1.3.2 Anotação das Imagens

No contexto do pré-processamento, a rotulação das imagens é uma tarefa trabalhosa, principalmente porque segue o padrão PASCAL VOC (*Visual Object Classes*), que se consolidou por ser amplamente utilizado na detecção e segmentação de objetos (EVERINGHAM, GOOL, *et al.*, 2010). O principal objetivo das anotações (rotulações) é criar um conjunto de dados para treinamento; além disso, elas também servem como um recurso valioso para avaliação do modelo. As anotações podem ser realizadas de forma programática ou manual, por meio do desenho de caixas delimitadoras ou polígonos para segmentação de

instâncias (ao redor do objeto) (RAGAB, ABDULKADIR, *et al.*, 2024, p. 57819).

Figura 11 – Anotação das imagens de células cervicais em duas classes, uma para a célula toda compreendo citoplasma e núcleo e outra somente para núcleo



Fonte: Elaborado pelo Autor (2025).

Em arquiteturas como YOLO, as anotações são representadas em arquivos de configuração contendo as coordenadas normalizadas x e y do ponto central da caixa delimitadora, juntamente com os valores relativos da largura e altura da imagem, variando de 0 a 1. Cada linha de anotações em um arquivo corresponde a um objeto anotado na imagem, associado a uma classe numérica (RAGAB, ABDULKADIR, *et al.*, 2024, p. 57820).

Para as anotações de todas as imagens, o processo foi todo realizado manualmente utilizando o *software* gratuito *Labelme*, desenvolvido pelo *MIT Computer Science and Artificial Intelligence Laboratory* (CSAIL). Além de permitir anotações por meio de caixas delimitadoras, o *software* possibilita segmentações utilizando polígonos delimitados por pontos definidos com cliques ao redor do objeto (RUSSELL, TORRALBA, *et al.*, 2008, p. 158).

As imagens utilizadas neste trabalho seguem o formato de segmentação de objetos, cuja ideia principal é segmentar a célula inteira, compreendendo citoplasma e núcleo, com classes correspondentes às classificações em TBS, tais como: NILM, ASCUS, LSILS, HSIL e SCC. Em paralelo, os núcleos das células também foram segmentados como classes independentes rotuladas em TBS, tais como: NILM_NUC, ASCUS_NUC, LSILS_NUC,

HSIL_NUC e SCC_NUC, conforme pode-se observar na Figura 11, a qual ilustra a tala do *software* LABELME no processo de anotação da célula em forma de polígono.

Quadro 5 - SIPaKMeD, equivalência das classes Tipo celulares para TBS

Tipos de células	**TBS**
Superficial/Intermediária	NILM
Parabasal	NILM
Coilocitótico	LSIL
Metaplástica	HSIL
Disqueratótico	HSIL

Fonte: Elaborado pela Autor de (WANA, SUNA, *et al.*, 2023).

O conjunto de dados SIPaKMeD foi anotado, e as classes originais foram convertidas de tipos celulares para classificações TBS. Os tipos celulares desse *dataset* foram categorizados em três classes normais, uma lesão de baixo grau e outra grave. A conversão destas classes e a correlação entre elas pode ser observada no Quadro 5 (WANA, SUNA, *et al.*, 2023, p. 6-7).

3.1.3.3 Aumento de Dados

Uma estratégia fundamental no treinamento de modelos em visão computacional em conjunto de dados é a técnica de *data augmentation* (aumento de dados), segundo a pesquisa de *Bochkovskiy, Wang e Liao* (2020, p. 2), ele considera o aumento de dado uma prática frequentemente adotada que beneficia a tarefa de detecção de objetos.

O objetivo da aplicação dessa técnica é aumentar artificialmente o número de amostras do *dataset* por meio de transformações de *pixels* aplicadas às imagens originais (PEREZ e WANG, 2017, p. 1530). As principais operações aplicadas incluem rotação, translação, espelhamento, variação de brilho, contraste e distorções (SHORTEN e KHOSHGOFTAAR, 2019, p. 3).

Modelos de detecção de objetos, especialmente as versões e variantes do modelo YOLO, exigem um conjunto de dados extenso e diversificado durante o treinamento para alcançar alta precisão e robustez. Nesse contexto, a técnica de aumento de dados é amplamente empregada para expandir artificialmente o conjunto de treino (ULTRALYTICS, 2025). O aumento de dados é uma prática comum nos *frameworks* da *Ultralytics* YOLO, que já implementam nativamente um conjunto abrangente de técnicas de *augmentation*, como ajustes de cor (HSV), rotação, translação e mosaico, integradas ao processo padrão de treinamento (RUMAN, 2023).

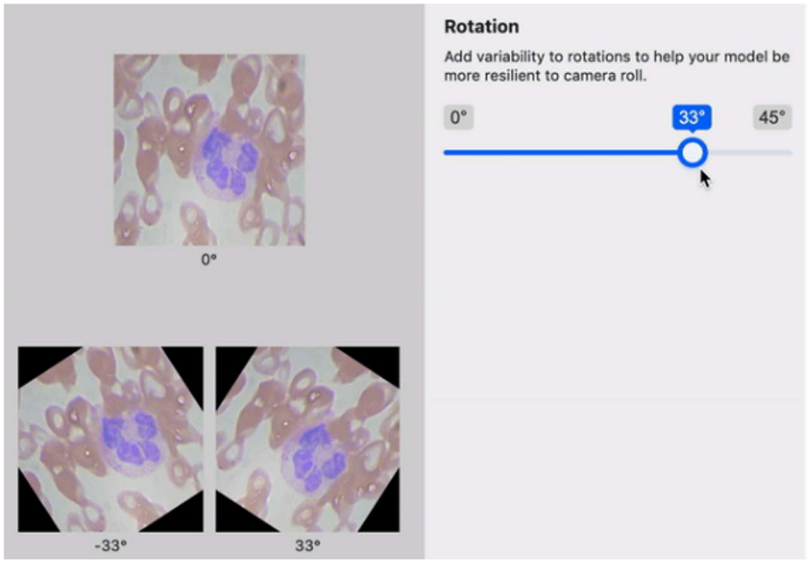
Tabela 4 – Distribuição de células do *dataset*

TBS	Células	% do total
NILM	8,579	53.07
ASCUS	606	3.75
ASCH	925	5.72
LSIL	2,385	15.52
HSIL	3,509	21.70
SCC	161	0.99
Total		**16,165**

Fonte: Elaborado pela Autor (2025).

A abordagem inicial deste trabalho envolveu o uso exclusivo de dois conjuntos de dados de imagens; no entanto, foi identificado um desequilíbrio significativo entre as classes. As células normais, em particular, predominaram de forma expressiva sobre os demais tipos de células atípicas, constituindo mais da metade da contagem total de células. Com o intuito de corrigir esse problema, o conjunto de dados RepoMedUNM foi posteriormente integrado aos dois iniciais. Contudo, mesmo após essa incorporação, o conjunto de dados final manteve uma distribuição de amostras desequilibrada (Tabela 4). Para mitigar essa disparidade, as classes minoritárias serão sobreamostradas por meio de técnicas de aumento de dados, expandindo efetivamente sua representação dentro do conjunto de dados.

Figura 12 – Primeiro estágio do aumento de dados no qual as imagens foram rotacionadas utilizando o *software Roboflow*



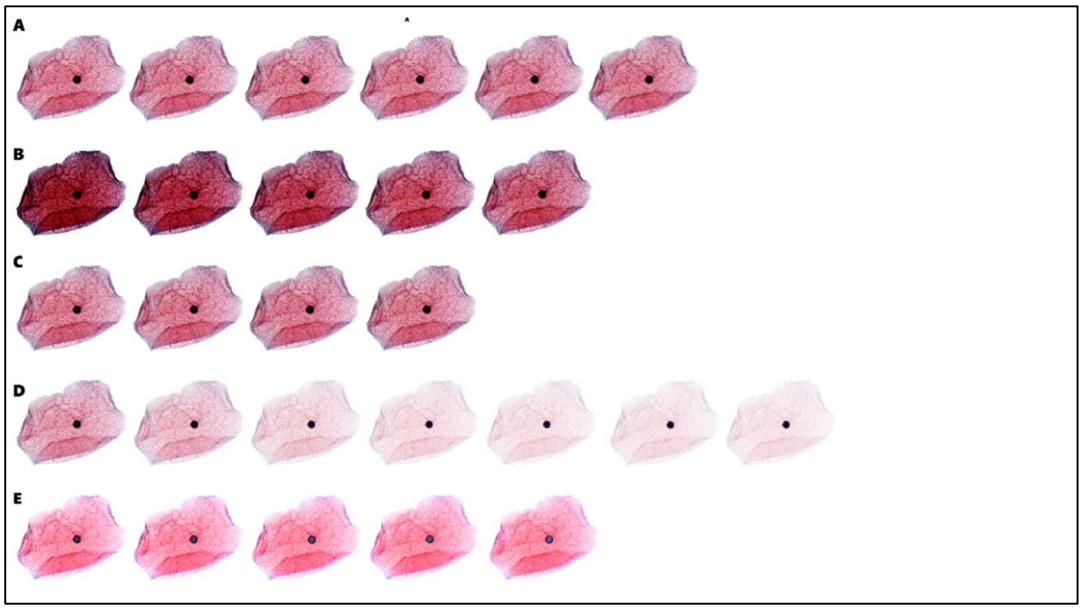
Fonte: Elaborado pelo Autor (2025).

Inicialmente, o conjunto de dados combinado SIPaKMeD e CRIC compreendia 820 arquivos de imagem. As imagens do RepoMedUNM foram consolidadas em 17 arquivos, cada cada um contendo 16 imagens individuais, elevando todo o conjunto preliminar para um total de 837 arquivos de imagem.

O primeiro estágio de aumento consistiu na aplicação de transformações rotacionais. Novos arquivos de imagem foram gerados a partir da rotação de cada imagem original em incrementos de 5 graus, tanto positivos quanto negativos, abrangendo o intervalo de 5 a 355 graus. Esse processo foi executado usando o aplicativo *Roboflow*, que cria automaticamente arquivos anotados após as rotações, conforme é demonstrado na Figura 12.

O segundo estágio concentrou-se exclusivamente na manipulação das propriedades da matriz de cores, nesta etapa também foi empregado o *software Roboflow*. As imagens foram alteradas apenas editando as matrizes dos canais de cores. Desta forma, foram selecionadas algumas propriedades que permitem variações da imagem, tais como: saturação, desfoque gaussiano, gaussiano médio, luz, nitidez e conversão para escala de cinza.

Figura 13 – Exemplo das alterações das imagens na segunda etapa do aumento de dados



Fonte: Elaborado pelo Autor (2025).

Para demonstrar o processo de aumento de dados nas propriedades dos canais de cores das imagens, a Figura 16 apresenta exemplos representativos da segunda etapa desse procedimento. Na Figura 1-A, observam-se algumas amostras resultantes da aplicação da propriedade *Average Blur*, que introduz variações de desfoque controlado. Já na Figura 13-B, são exibidas alterações relacionadas ao escurecimento da imagem, recurso particularmente útil por reproduzir condições comuns nas lâminas analisadas em microscópio, onde a iluminação nem sempre é uniforme.

Na sequência, a Figura 13-C ilustra exemplos de desfoque gaussiano, enquanto a Figura 13-D mostra ajustes de intensidade luminosa. Essas transformações não apenas

ampliaram o número de amostras disponíveis no *dataset* final, como também contribuíram para simular e mitigar problemas recorrentes de desfoque das objetivas e variações de iluminação do microscópio e das lâminas

Tabela 5 – Distribuição de células após aumento de dados

TBS	Células	% do total
NILM	13.591	8,75
ASCUS	13.591	8,75
ASCH	**12.425**	**7,99**
LSIL	13.591	8,75
HSIL	13.591	8,75
SCC	**8.870**	**5,71**
NILM_NUC	13.591	8,75
ASCUS_NUC	13.591	8,75
ASCH_NUC	**12.425**	**7,99**
LSIL_NUC	13.591	8,75
HSIL_NUC	13.591	8,75
SCC_NUC	**8.870**	**5,71**
Total	**155.318**	

Fonte: Elaborado pela Autor (2025).

Por fim, a Figura 13-E exibe alguns exemplos de modificações na saturação das imagens, recurso que adiciona diversidade cromática às amostras e fortalece a robustez do treinamento dos modelos.

O processo de aumento de dados foi aplicado a duas bases de dados com imagens de diferentes dimensões: 1280 x 1280 e 2016 x 2016 *pixels,* respectivamente. É importante destacar que ambas as bases de dados contêm as mesmas imagens em dimensões distintas, com o intuito de proporcionar maior diversidade nos experimentos de inferência e melhor desempenho no treinamento dos modelos de visão computacional.

Para expandir significativamente o conjunto de dados destinado ao treinamento, os exemplares de imagens foram ampliados de 16.165 para 155.318 amostras de células cervicais por meio do aumento de dados resultante dos dois estágios. Uma estimativa das proporções das classes é listada na Figura 5, na qual é possível perceber uma distribuição de um pouco mais de dez porcento na maioria das classes. Quanto as classes ASCH e ASCH_NUC obtiveram 7,99% de participação no total, já as classes SCC e SCC_NUC que antes tinham menos de 1% de participação, atingiram 5,71% de representação, mesmo sendo pouco foi um aumento substancial.

3.1.4 Quarta Fase de Modelagem

Nesta fase “Modelagem”, discutem-se os procedimentos e experimentos empregados para a seleção e aplicação do algoritmo de detecção YOLLOv11 nas variantes *nano* (n) e *small* (s), com foco na tarefa de segmentação de instâncias. Essa etapa envolve a partição dos dados, a configuração de hiperparâmetros consistentes e o treinamento dos modelos com os *datasets* de diferentes resoluções, a fim de comparar o desempenho e determinar a combinação ideal de arquitetura e dimensão de imagem que otimize a precisão e a velocidade do modelo.

Para atender aos objetivos deste presente trabalho, serão conduzidos experimentos tanto na fase de modelagem quanto na realização dos testes de inferência. No que diz respeito à escolha da CNN, adotou-se o YOLOv11. Apesar da versão atual ser o do YOLOv12, durante o período do desenvolvimento deste presente trabalho, preferiu-se utilizar a versão 11, com base no estudo de Nidhal (2024), que realizou uma avaliação comparando os modelos YOLOv12 e YOLOv11.

Os resultados demonstraram maior precisão do YOLOv12 em relação a outras versões. No entanto, essa maior precisão implicou maior complexidade computacional, exigindo mais hardware. Essa compensação levou à seleção do YOLOv11 como opção preferencial para esta pesquisa, devido ao seu excelente equilíbrio entre velocidade e tamanho do modelo, sem custo computacional significativo.

Os *datasets* finais foram treinados com o modelo YOLOv11 utilizando suas duas variantes menores, *Small* e *Nano*, em tarefas de detecção por caixas delimitadoras e segmentação de instâncias, conforme listado no Quadro 6.

As duas variantes do YOLOv11 foram treinadas na tarefa de detecção de segmentação de instâncias, pois uma vez o modelo é treinado nesta modalidade, beneficia-se de obter tanto detecções em caixa delimitadoras quanto segmentações de objetos, em um único modelo. O YOLOv11, assim como outros modelos de segmentação de instâncias de última geração, opera em um paradigma de "*detectar e então segmentar*" (KHANAM e HUSSAIN, 2024; REDMON, DIVVALA, *et al.*, 2016).

Isso significa que o modelo primeiro realiza a detecção por caixa delimitadora (*Bounding Boxes*, BB) para localizar objetos e, posteriormente, gera máscaras em nível de pixel

para essas instâncias detectadas. Consequentemente, uma avaliação abrangente de seu desempenho exige a análise de ambas as tarefas de detecção.

Quadro 6 – Listagem dos experimentos de treinamento

MODELO	TAREFA DE DETECÇÃO	*DATASET* 1	*DATASET* 2
YOLOv11n	Caixa delimitadora	2016 x 2016	1280 x 1280
YOLOv11s			
YOLOv11n	Segmentação de instâncias	2016 x 2016	1280 x 1280
YOLOv11s			

Fonte: Elaborado pela Autor (2025).

Para atender aos objetivos definidos da pesquisa, a fase experimental inicial concentrou-se no treinamento do modelo, empregando dois conjuntos de dados distintos. Esses conjuntos, derivados de uma mesma fonte, mas processados em resoluções diferentes, foram utilizados para treinar o modelo YOLOv11 em suas duas variantes. Além disso, cada conjunto de dados foi rigorosamente particionado em subconjuntos de treinamento, validação e teste com uma distribuição de 80%, 10% e 10%, respectivamente.

Hiperparâmetros consistentes foram aplicados em todas as execuções de treinamento: um número elevado e uniforme de épocas foi definido em ambas as variantes, porém interrompidas por meio da implementação de parada antecipada (*Early Stoppin*), visando evitar sobreajuste (*overfitting*). O treinamento foi configurado especificamente para uma tarefa de segmentação, utilizando um tamanho de lote de 16, um decaimento de peso de 0,0005 e uma taxa de aprendizado inicial de 0,001.

3.1.4.1 Treinamento dos datasets 1280 x 1280 e 2016 x 2016 pixels

Entre os objetivos desta pesquisa, propõe-se comparar o treinamento de dois *datasets* com imagens com resoluções 1280 x 1280 e 2016 x 2016 *pixels*, respectivamente, combinando-os com as duas variantes do YOLOv11 *nano* e *small*. Assim, iniciou-se o treinamento com o modelo YOLOv11n preferencialmente utilizando o *dataset* de maior resolução, com imagens de 2016 x 2016 *pixels*.

A razão pela qual dois *datasets,* com resoluções diferentes foram aplicados no treinamento das duas variantes do YOLOv11, justifica-se pelo fato de esses modelos reduzidos não atingirem uma precisão superior a 60%, mesmo com o aumente do número de amostras no treinamento (PIMENTEL, REIS, *et al.*, 2023). Além disso, segundo Liang (2023), que observaram a relação entre a precisão dos modelos YOLO e a resolução das imagens de

treinamento, reforça-se a necessidade de utilizar resoluções mais altas. Com base nessa premissa, justificam-se experimentos conduzidos em conjunto de dados que contemplam imagens com resoluções superiores a 640 x 640 *pixels*.

No decorrer do treinamento do modelo YOLOv11n com o *dataset* com resolução 2016 x 2016 *pixels*, exatamente na primeira época, houve um estouro de memória e, apesar da aplicação de medidas para mitigar o problema, tais como um amplo ajuste de hiperparâmetros, incluindo a redução sequencial do *batch size* de 64 até 1, o problema persistiu e não avançou além da primeira época. Em consequência disso, esse *dataset* foi descartado dos experimentos propostos devido às limitações das configurações do hardware descritas neste trabalho.

Já no treinamento do modelo YOLOv11n, com o *dataset* de resolução 1280 x 1280 *pixels*, o mesmo problema ocorreu no início da segunda época. Porém, na primeira bateria de ajustes, o problema foi resolvido ao desabilitar o hiperparâmetro *post-epoch validations*, definindo-o com *False* (falso). Contudo, o processo de treinamento nesta base de dados consumiu 50% de memória da GPU.

No entretanto, o treinamento da variante YOLOv11s, que também foi realizado no mesmo *dataset*, não apresentou nenhum problema com consumo de memória e não precisou ajustes de hiperâmetros. Além disso, o modelo *small* atingiu melhor precisão após somente 6 épocas, levando 78 horas em cada uma, em comparado ao modelo YOLOv11n, que precisou de 11 para atingir melhor precisão, com um tempo de 36 horas por época.

3.1.5 Quinta Fase de Avaliação

Nesta fase de "Avaliação", os modelos treinados foram avaliados com base nos objetivos de negócio e nas métricas de desempenho em tarefas de detecção de imagens nas variantes *nano* e *small* do YOLLO11. Diferentemente de tarefas de classificação pura, a detecção de objetos exige a medição tanto da correção da identificação da classe quanto da precisão (ALCARAZ-CHAVEZ, TÉLLEZ-ANGUIANO, *et al.*, 2024).

Desta forma, foi adotada a *Média das Precisões Médias* (do inglês *Mean Average Precisionm*, mAP), métrica utilizada e considerada mais robusta e amplamente aplicada para a avaliação de modelos, especialmente para os da família YOLO (KHANAM e HUSSAIN, 2024; REDMON, DIVVALA, *et al.*, 2016).

Além das métricas de acurácia, o modelo foi avaliado quanto à sua latência, ou seja,

quadro por segundo (*Frames Per Second,* FPS), confrontando os resultados com os critérios de sucesso do negócio e realizando testes de inferência em vídeo em tempo real com câmeras CMOS.

E por fim, para garantir que seja avaliado o melhor desempenho entre as duas variantes *nano* e *small*, foi conduzido uma avaliação alternando essas variantes, por meio de 5-*Fold Cross Validation* em uma nova base de dados derivada da original. Esse experimento visa determinar qual das variantes obteve, de fato, o melhor desempenho e, consequentemente, servirá de base para adoção do modelo mais adequado. Mais detalhes da avaliação cruzada foram discutidos na seção 4.2.

3.1.6 Sexta Fase de Implantação

A fase de implantação visa a proposição de um sistema final de detecção automatizado que seja totalmente replicável e de baixo custo, utilizando a variante do YOLLOv11 mais eficiente e precisa, conforme os resultados da avaliação.

O Plano de Preparação da Implantação será resumido em três etapas:

1. O modelo final será treinado em um banco de dados final, treinado na melhor arquitetura escolhida após o processo de avaliação;
2. A disponibilização do modelo será feita por meio de um executável portátil confeccionados em *Python*;
3. Pretende-se aplicar o modelo em parceria com instituições médicas para aprimorar o protótipo inicial em condições de uso real.

Para a tarefa de "Plano de Monitoramento e Manutenção", será estabelecido um ciclo de refinamento sob supervisão (*feedback loop*), com o objetivo de comparar as predições do modelo treinado em condições do mundo real. Os procedimentos serão:

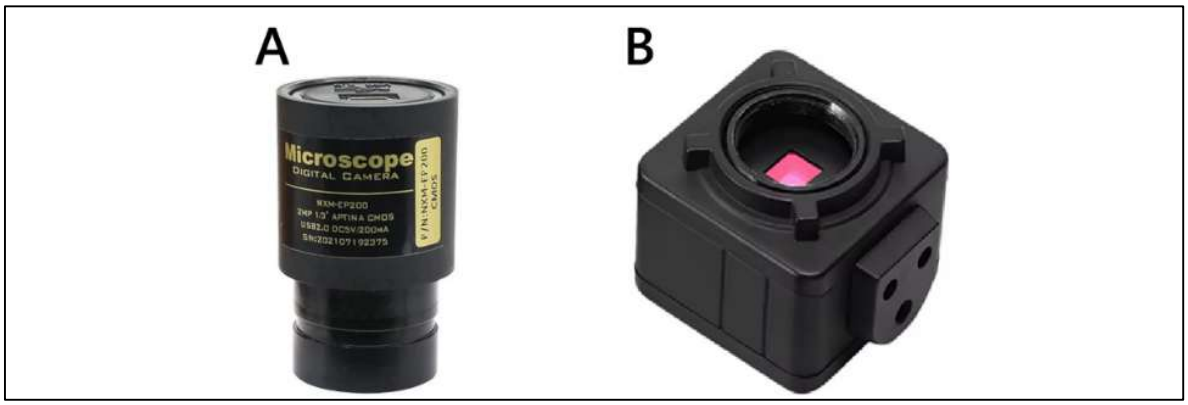
1. Os resultados das previsões erradas (falsos positivos e falos negativos) serão contabilizados e avaliados;
2. Todas as imagens que não foram corretamente detectadas pelo modelo serão coletadas e reinseridas no banco de dados definitivo, que será novamente utilizado para treinamento;
3. A quantidade de imagens coletadas também servirá como métrica para avaliar a redução de exames falsos negativos.

Devido a Lei Geral de Proteção de Dados (LGPD), o plano de implementação de dados não foi posto em ação e deixados para trabalhos futuros, no qual se pretende estabelecer parceria com uma instituição médica para não só implantar, mas avaliar a diminuição dos falsos negativos.

3.2 EXPERIMENTOS DE INFERÊNCIA DE VÍDEO EM TEMPO REAL

Uma vez concluída as fases de preparação de dados e modelagem, é possível realizar os experimentos de inferência de vídeo em tempo real. Porém, antes da realização destes experimentos, houve a aquisição de equipamentos financiados pela Universidade Estadual do Maranhão.

Figura 14 – (A) Câmera USB CMOS EYE PIECE de 2 MP. (B) Câmera USB CMOS 5 MP



Fonte: www.aliexpress.com.

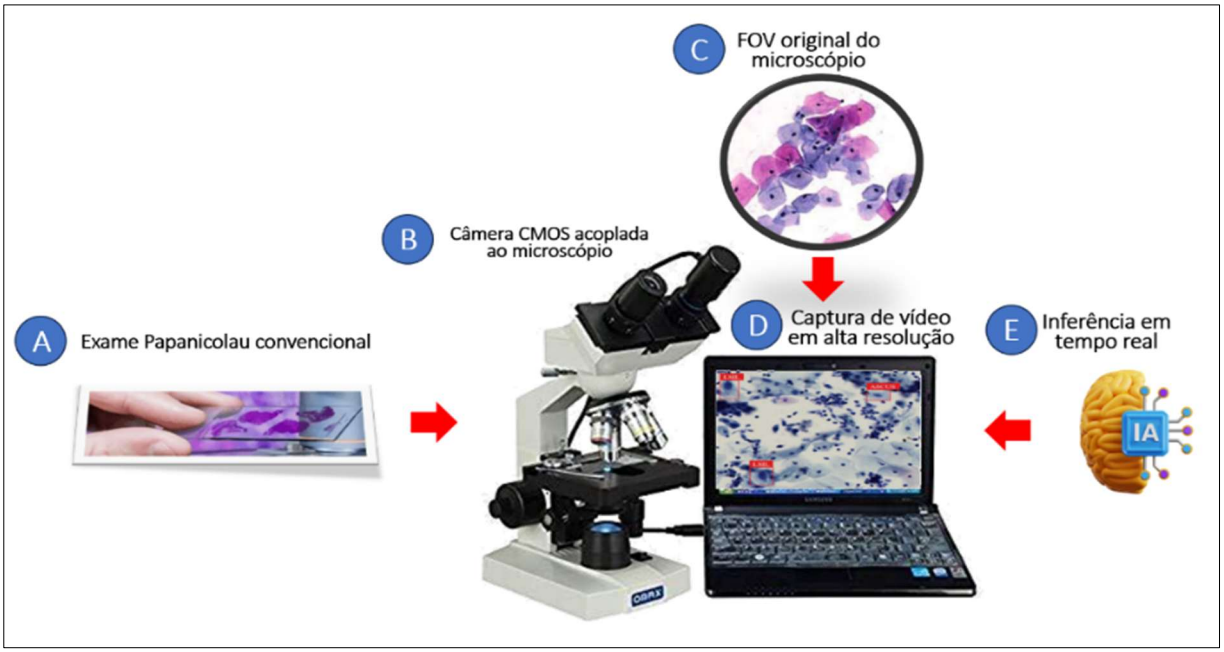
Os experimentos iniciais de inferência em vídeo em tempo real se iniciaram por meio da aquisição de câmeras para microscópio. A princípio, o objetivo era a adquirir uma câmera com o menor custo possível; assim, a melhor escolha foi uma câmera USB CMOS *Eye Piece* para microscópio de 2 *Mega Pixels* (MP) (Figura 14-A). Após a aquisição, foi necessário a compra de outra câmera compatível com a primeira, mas com resolução de 5 MP (Figura 14-B).

Para que os objetivos propostos neste trabalho fossem atingidos, foi elaborado um plano de automação do exame de Papanicolau, conforme pode ser observado na Figura 15. Este esquema serviu de base para realização de todos os experimentos de inferência em vídeo em tempo real.

Neste contexto, para que os experimentos pudessem ocorrer dentro deste plano de automação, não foi necessário adquirir nenhum equipamento adicional além das câmeras, o que

possibilita a utilização de toda a estrutura já existente em laboratórios tradicionais, nos quais o exame é realizado (Figura 15-A).

Figura 15 – Planejamento de automatização do exame de Papanicolau para realização dos experimentos



Fonte: Elaborado pelo Autor (2025).

Portanto, no planejamento dos experimentos, a imagem do FOV do microscópio será capturada pela câmera CMOS, a qual é montada em uma das objetivas do microscópio binocular (Figura 15-B). Na ocasião as câmeras de 2 e 5 MP foram testadas e avaliadas quanto a qualidade da imagem.

Quanto à participação de um citologista, este poderá examinar a lâmina por meio da objetiva do microscópio (Figura 15-C) ou acompanhar pela tela do computador, que ampliará todo o campo de visão em alta resolução (Figura 15-D). Por fim, paralelamente, foram realizados os experimentos de inferência de vídeo, nos quais foi possível determinar qual das variantes YOLOv11 atingiu melhor precisão e velocidade em tempo real (Figura 15-D).

No que diz respeito à aquisição de vídeo a partir do microscópio, sugerida no planejamento de experimentação, optou-se por câmeras USB CMOS, pelo fato de esses equipamentos capturarem vídeo em tempo real sem necessidade de configuração adicional, já que funcionam como uma webcam USB em ambiente *Windows*. Por esta razão, para alcançar melhor desempenho e inferência, propõe-se o mesmo procedimento utilizado em captura de

vídeo por câmeras encontradas em computadores e *notebooks*.

Figura 16 – Algoritmo proposto para inferência de vídeo

Algorithm 1 Inferência de vídeo em tempo real com YOLO

```
Require: fonte_vídeo, modelo_yolo
Ensure: exibição_inferência_tempo_real
 1: dispositivo_captura ← inicializar_captura_vídeo(fonte_vídeo)
 2: definir_resolução(dispositivo_captura, largura=1280, altura=720)
 3: modelo_yolo ← carregar_modelo(modelo_yolo)
 4: while dispositivo_captura_aberto do
 5:    ret, quadro_entrada ← ler_quadro(dispositivo_captura)
 6:    if não ret then
 7:       interromper
 8:    end if
 9:    resultados_inferência ← modelo_yolo(quadro_entrada)[0]
10:    quadro_anotado ← plotar_resultados(resultados_inferência,
       máscaras=False, caixas=False)
11:    exibir_quadro(quadro_anotado)
12:    if verificar_tecla_pressionada('q') then
13:       interromper
14:    end if
15: end while
16: liberar_dispositivo_captura(dispositivo_captura)
17: destruir_todas_janelas()
```

Fonte: Adaptado pelo Autor de (ULTRALYTICS, 2023).

Como parte do plano de automação, desenvolveu-se um código simples de inferência em vídeo, conforme pode ser observado no algoritmo exibido na Figura 16 (ULTRALYTICS, 2023). O destaque principal neste algoritmo se concentra nas seguintes linhas: 1, 3, 5, 9, 10 e 11. A linha 1 obtém o vídeo direto da câmera, exatamente como se procede com uma *webcam*. Na linha 3, o modelo YOLOv11, já com os pesos treinados, é carregado, e este processo se repetirá para cada variante.

Mais abaixo, a linha 5, dentro de um *loop*, captura uma imagem extraída de um *frame* do vídeo. As linhas 9 e 10 encarregam-se da inferência do modelo. Neste passo, o YOLO fornece uma matriz repleta de informações, a partir da qual é possível obter coordenadas para a plotagem dos polígonos que contornam as células detectadas. Além disso, a matriz também fornece as respectivas classes com suas pontuações de detecção.

Além disso, é importante destacar que a câmera USB CMOS oferece uma imagem com uma resolução 640 x 480 *pixels* por padrão. Desta forma, a imagem capturada na linha 5

tem exatamente essa resolução. Apesar do treinamento dos modelos em alta resolução, as imagens inferidas na linha 9 utilizam a mesma configuração. Contudo, pode-se recorrer a dois recursos para aumentar a resolução da câmera:

- O primeiro, configurando a resolução própria, que é a opção escolhida no algoritmo, como exibido na linha 2, que definiu a imagem para 1280 x 720 *pixels*, essa foi exatamente a resolução utilizada na inferência dos modelos.

- A segunda opção, seria utilizar a resolução padrão da câmera e redefinir a dimensão da imagem por meio da biblioteca OPENCV com a inclusão da instrução cv2.*resize* antes da linha 11, que atualiza a imagem na tela.

4 RESULTADOS EXPERIMENTAIS

Nesta seção, apresentam-se de forma geral os resultados produzidos pelos experimentos correlacionados as questões desta pesquisa. Todos os experimentos foram executados de forma sequencial e visam atender os pré-requisitos estabelecidos para conclusão deste trabalho. Desta forma, os resultados refletem exatamente os dados extraídos do treinamento das duas variantes do YOLOv11 e suas respectivas avaliações, bem como as análises e observações sobre as inferências de vídeo em tempo real.

4.1 PRECISÃO DOS MODELOS

Os experimentos dedicados ao treinamento das variantes YOLOv11n e YOLOv11s produziram resultados apenas quando aplicados ao conjunto de dados com imagens de resolução 1280 x 1280 *pixels*. Nesse cenário, os modelos alcançaram bom desempenho com imagens nesta resolução, conforme pode ser observado no Quadro 7.

Quadro 7 – Precisão e *Recall* do treinamento das variantes do YOLOv11

MODELO	Tarefas	mAP	Recall
YOLOv11n	BB	98,00	97,54
	SEG.	95,40	94,60
YOLOv11s	BB	99,31	99,20
	SEG.	96,20	96,60

Fonte: Elaborado pela Autor (2025).

Quanto ao processo de avaliação de desempenho dos modelos com a variante YOLOv11n na tarefa BB, após 11 épocas, o modelo atingiu uma mAP de 98% e um *Recall* de 97,54%. Já na tarefa de Segmentação de instâncias, rendeu um mAP de 95,4% e uma *Recall* de 94,6% (Quadro 7).

Quadro 8 – Precisão e *Recall* das classes do treinamento da variante YOLOv11n

CLASSE	**BB**		**SEGMENTAÇÃO**	
	mAP	***Recall***	**mAP**	***Recall***
NILM	0.983	0.992	0.99	0.989
ASCUS	0.995	0.990	0.995	0.990
ASCH	0.982	0.959	0.967	0.945
LSIL	**0.994**	**0.994**	**0.991**	**0.991**
HSIL	0.998	0.996	0.949	0.948
SCC	**0.99**	**0.975**	**0.719**	**0.708**
NILM_NUC	0.982	0.987	0.976	0.982
ASCUS_NUC	0.985	0.990	0.984	0.989
ASCH_NUC	0.955	0.95	0.953	0.949
LSIL_NUC	0.986	0.981	0.98	0.975
HSIL_NUC	0.955	0.93	0.953	0.929
SCC_NUC	0.981	0.962	0.978	0.960

Fonte: Elaborado pela Autor (2025).

As variantes *small* do YOLOv11 obtiveram um desempenho equivalente em comparação as variantes *nano*. Porém, como esperado, atingiram métricas um pouco maior. Na tarefa de detecção BB a precisão alcançou um mAp de 99,31% e um *Recall* de 99,20. E na tarefa de segmentação de instâncias, a precisão obteve um mAP de 96,20% e um *Recall* de 96,20% (Quadro 7). Pode-se observar que ambas as variantes atingiram resultados semelhantes, sendo os modelos *small* ligeiramente superiores.

Para avaliar o desempenho individuais de cada classe da variante *nano*, cujas métricas resultantes podem ser observadas no Quadro 8. Apesar das classes NILM serem predominantes, o trabalho de aumento de dados contribuiu para balancear e equilibrar a precisão entre todas as classes. Em uma análise simples, observa-se que a classe LSIL se destacou em todas as métricas das tarefas de detecção. Entretanto, a classe SCC, que tinha a menor representatividade, surpreendeu com um mAP de 99%, porém, todas as demais métricas foram mais baixas em relação às outras classes.

Quadro 9 – Precisão e *Recall* das classes do treinamento da variante YOLOv11s

CLASSE	**BB**		**SEGMENTAÇÃO**	
	mAP	***Recall***	**mAP**	***Recall***
NILM	0.996	0.993	0.994	0.991
ASCUS	0.998	0.997	0.998	0.997
ASCH	0.991	0.988	0.984	0.981
LSIL	0.996	0.989	0.995	0.988
HSIL	0.997	**1.00**	0.965	0.967
SCC	0.997	0.987	**0.764**	**0.755**
NILM_NUC	0.994	0.996	0.988	0.991
ASCUS_NUC	0.996	0.997	0.995	0.995
ASCH_NUC	0.986	0.99	0.987	0.987
LSIL_NUC	0.994	0.994	0.990	0.989
HSIL_NUC	0.98	0.981	0.980	0.979
SCC_NUC	0.992	0.983	0.992	0.982

Fonte: Elaborado pela Autor (2025).

A mesma análise de desempenho individual das classes na variante YOLOv11s, apresentou um resultado mais homogêneo (Quadro 9), com apenas a classe HSIL, se destacando sobre as demais, alcançando 100% de precisão. A classe SCC também apresentou um mAP inferior apenas na tarefa de segmentação. Em termos de precisão média, a variante *small* do YOLOv11 mostrou-se superior ao *nano*. Porém, o nano não apresentou uma diferença percentual tão significativa em relação a variante *small*.

4.2 AVALIAÇÃO 5-*FOLD CROSS VALIDATION*

A seleção de modelos e a estimativa de riscos são etapas críticas no aprendizado de máquina, cuja avaliação cruzada surge como uma ferramenta essencial. Segundo Arlot e Celisse (2010, p. 51), a avaliação cruzada baseia-se na ideia de reutilização dos dados, onde um conjunto de dados original é dividido para que o modelo seja treinado em subconjuntos distintos. Essa abordagem permite estimar o erro de generalização de um algoritmo sem a necessidade de um conjunto de testes independente e massivo.

A respeito da validação cruzada de 5-*folds* (ou V-*Fold*, cujo V=5), o procedimento consiste em particionar os dados em 5 blocos (dobras) de tamanhos aproximadamente iguais. Conforme detalhado por Alort e Celisse (2010, p. 50), assim, cada variante do YOLOv11 foi treinada cinco vezes; em cada iteração. Desta forma, quatro blocos foram utilizados para o ajuste do modelo (treino), enquanto o bloco remanescente é reservado para a avaliação (validação). Esse ciclo garante que cada observação do conjunto de dados seja utilizada para testes exatamente uma vez.

Neste trabalho utilizou-se o conjunto de dados com resolução 1280 x 280 *pixels*. todo o conjunto de dados foi dividido em 5 partes iguais (ou o mais próximo possível disso), cada uma destas partes corresponde a um *fold* ou dobras. O processo ocorreu em cinco rodadas independentes. Em cada rodada, a configuração muda:

- **Iteração 1:** As dobras 1, 2, 3 e 4 são usadas para o treinamento. A dobra 5 é usada para a validação.
- **Iteração 2:** As dobras 1, 2, 3 e 5 são usadas para o treinamento. A dobra 4 é usada para a validação.
- **Iteração 3:** As dobras 1, 2, 4 e 5 são usadas para o treinamento. A dobra 3 é usada para a validação.
- **Iteração 4:** As dobras 1, 3, 4 e 5 são usadas para o treinamento. A dobra 2 é usada para a validação.
- **Iteração 5:** As dobras 2, 3, 4 e 5 são usadas para o treinamento. A dobra 1 é usada para a validação.

Para uma avaliação comparativa mais ampla das variantes do YOLOv11, aplicada às tarefas de detecção de objetos e segmentação de instâncias, foi empregada uma validação cruzada de *5-fold* sob as duas configurações de treinamento: YOLOv11n treinado por 11 épocas e YOLOv11s treinado por 6 épocas. O número diferente de épocas deve-se ao fato de ambos terem incorporado uma parada antecipada. O objetivo foi analisar como a duração do treinamento impactou o desempenho do modelo em termos de precisão, *recall* e *F1-score*. Assim, o Quadro 12 resume as métricas médias obtidas nos cinco *folds* para ambas as configurações.

Quadro 10 – Avaliação geral obtida por meio de 5-*Fold Cross Validation*

Modelo	**Tarefa**	**mAP**	***Recall***	**F1-*score***
YOLOv11n	BB	0.841	0.830	0.835
YOLOv11n	SEG.	0.861	0.790	0.795
YOLOv11s	BB	0.902	0.913	0.913
YOLOv11s	SEG.	0.878	0.877	0.878

Fonte: Elaborado pela Autor (2025).

O modelo YOLOv11s, treinado por menos épocas, superou consistentemente o YOLOv11n em todas as três métricas, tanto para predição de caixa delimitadora quanto para segmentação de instâncias. O modelo YOLOv11s alcançou uma pontuação F1 média de 91,30% para detecção por caixas delimitadoras e 87,80% para segmentação de instâncias, demonstrando generalização superior mesmo com um período de treinamento menor.

Quadro 11 – Avaliação detalhada 5-*Fold Cross Validation*

Fold	**Modelo**	**BB**			**SEGMENTAÇÃO**		
		mAP	***Recall***	**F1**	**mAP**	***Recall***	**F1**
01	YOLOv11n	0.869	0.878	0.873	0.826	0.838	0.832
	YOLOv11s	0.898	0.910	0.904	0.874	0.870	0.872
02	YOLOv11n	0.829	0.824	0.827	0.789	0.774	0.782
	YOLOv11s	0.920	0.925	0.922	0.880	0.886	0.883
03	YOLOv11n	0.837	0.816	0.826	0.797	0.778	0.787
	YOLOv11s	0.922	0.922	0.922	0.884	0.890	0.887
04	YOLOv11n	0.855	0.810	0.832	0.815	0.773	0.793
	YOLOv11s	0.910	0.897	0.903	0.884	0.864	0.874
05	YOLOv11n	0.814	0.822	0.818	0.776	0.785	0.780
	YOLOv11s	0.911	0.913	0.912	0.869	0.875	0.872
06	YOLOv11n	0.832	0.828	0.830	0.800	0.788	0.794
	YOLOv11s	0.911	0.913	0.912	0.869	0.875	0.872
07	YOLOv11n	0.840	0.831	0.835	0.802	0.792	0.797
08	YOLOv11n	0.836	0.825	0.831	0.799	0.789	0.794
09	YOLOv11n	0.838	0.834	0.836	0.801	0.791	0.796
10	YOLOv11n	0.845	0.829	0.837	0.803	0.790	0.796
11	YOLOv11n	0.856	0.833	0.844	0.803	0.792	0.797

Fonte: Elaborado pela Autor (2025).

Para melhor compreender a variabilidade entre todos os cinco *folds*, a Quadro 11 apresenta uma análise detalhada da precisão, *recall* e pontuação F1 para cada *fold*. Esse nível de granularidade ajuda a avaliar a consistência e a robustez de cada modelo. Como se pode observar, o YOLOv11s supera o YOLOv11n de forma consistente em todas as *folds* e em todas as três métricas. Notavelmente, os ganhos na pontuação F1 são particularmente significativos na tarefa de segmentação de instâncias, na qual o YOLOv11s melhora a média percentualmente.

O desempenho do YOLOv11n desafia a suposição de que um treinamento mais longo necessariamente produz uma melhor generalização. Os resultados da pontuação F1, em especial, indicam uma relação mais equilibrada entre precisão e *recall*.

Algumas observações podem ser extraídas dessa avaliação cruzada, tais como:

- O YOLOv11s demonstrou maior precisão e consistência, especialmente na segmentação, uma tarefa tipicamente mais sensível em detalhes e precisão espacial.
- O uso da pontuação F1 forneceu uma visão de desempenho mais realista quando precisão e o *Recall* divergem ligeiramente, como observado em alguns dos cinco *folds* do YOLOv11n.

Esta avaliação comparativa demonstra que o YOLOv11s, apesar de uma duração de treinamento significativamente menor, alcança resultados superiores tanto em tarefas de detecção quanto de segmentação. No entretanto é possível notar que a vantagem da variante YOLOv11 *small* em relação a *nano* não é tão significativa.

4.3 INFERÊNCIA EM BANCO DE DADOS EXTERNO

Para a avaliação e teste do desempenho de detecção inicial, os modelos YOLOv11n e YOLOv11s foram submetidos a testes de inferência em um conjunto de dados não visto anteriormente pelos modelos, especificamente o conjunto de dados APACC. Este conjunto de dados foi selecionado devido às suas características particularmente desafiadoras. Os principais desafios apresentados incluem um campo de visão significativamente mais estreito em comparação às imagens do treinamento.

Além disso, muitas imagens apresentam desfoque substancial, resultando em imagens frequentemente indistintas. Adicionalmente, há um número considerável de imagens subexpostas. Uma vantagem crucial do conjunto de dados APACC, no entanto, é a

disponibilidade de anotações de verdade fundamental (imagens já anotadas no APACC por caixas delimitadoras), o que permitem uma comparação quantitativa direta dos resultados de inferência.

O experimento de inferência avaliou as detecções de células cervicais geradas pelos modelos YOLOv11n e YOLOv11s em comparação com suas respectivas imagens de verdade fundamental já anotadas. A variante nano apresentou consistentemente maior sensibilidade de detecção para instâncias celulares, identificando um número maior de classes de células inteiras e somente núcleos em comparação com a variante *small*. Apesar dessa maior sensibilidade, sua precisão na detecção de classes individuais foi notavelmente menor, em aproximadamente 3 a 10 pontos percentuais, quando comparada à outra variante.

Inesperadamente, o modelo YOLOv11n apresentou um tempo médio de processamento por maior. Embora sua velocidade de inferência bruta pudesse ser inerentemente mais rápida devido à sua arquitetura menor, o número significativamente maior de detecções geradas por sua maior sensibilidade aumentou substancialmente a carga de pós-processamento, estendendo assim o tempo total de detecção.

O modelo YOLOv11s, por outro lado, demonstrou eficiência superior, exibindo consistentemente percentuais de acerto por classe individuais mais altos em comparação com o modelo nano. Seu tempo de inferência apresentou variância significativamente menor do que o do YOLOv11n, principalmente devido à sua menor sensibilidade de detecção. Essa sensibilidade reduzida levou a certas observações quanto ao modelo YOLOv11s, que se concentrou predominantemente em objetos com maiores pontuações de confiança, filtrando, assim, detecções potencialmente imprecisas que possivelmente sobrecarregariam o tempo de detecção.

No geral, o modelo YOLOv11s teve um desempenho substancialmente melhor do que a variante nano, particularmente em termos de precisão. Apesar dessas diferenças de desempenho, ambos os modelos alcançaram velocidades de inferência muito rápidas, tornando-os adequados para aplicações de detecção em tempo real.

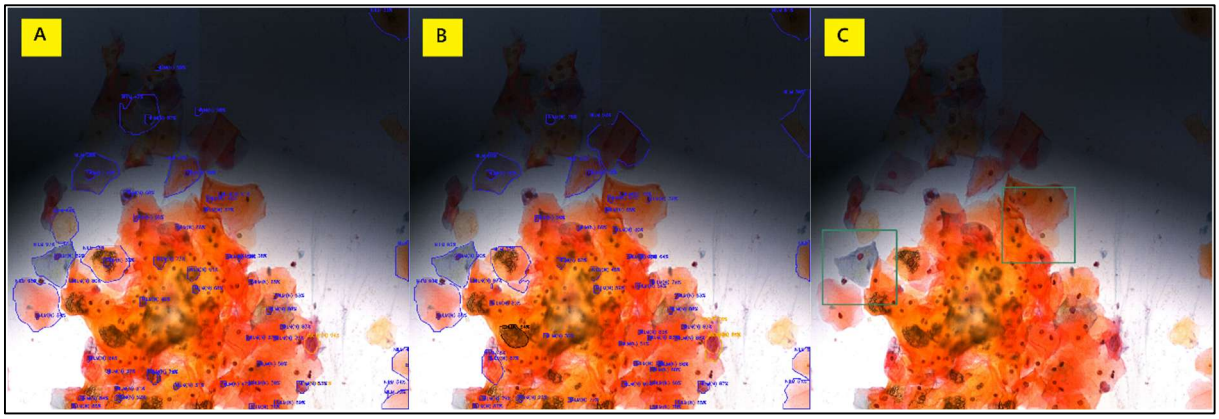
A comparação do desempenho coletivo dos dois modelos revelou sua alta assertividade na localização de objetos, identificando consistentemente um número maior de objetos em comparação as imagens de verdade absoluta do banco de dados APACC. Especificamente, a detecção de células normais (NILM) demonstrou a maior precisão em

ambos os modelos.

Além disso, ambos exibiram grande capacidade de detectar com precisão pequenos objetos e apresentaram desempenho excepcional na identificação de células com imagens borradas. Notavelmente, a presença de imagens subexpostas não prejudicou o desempenho da detecção, em ambos os modelos localizando com sucesso todas as células relevantes nessas condições adversas.

O principal objetivo da avaliação dos dois modelos YOLOv11 foi comparar sua capacidade de detectar células cervicais, especialmente em agrupamentos celulares sobrepostos. Notavelmente, ambos os modelos alcançaram detecção precisa de núcleos celulares, mesmo em condições de imagem subótimas.

Figura 17 – Resultados de inferência comparativa dos modelos YOLOv11n e YOLOv11s no conjunto de dados APACC. (A) O YOLOv11n exibe maior sensibilidade de detecção. (B) YOLOv11s demonstra precisão superior. (C) Imagem de referência do *dataset* APACC com BB de verdades fundamentais já inferidas



Fonte: Adaptado pelo Autor 2025.

A Figura 17 , demonstra os resultados de inferência, comparando os modelos YOLOv11n e YOLOv11s, os quais foram avaliados e testados no conjunto de dados APACC. Conforme pode ser observado, as imagens ilustram claramente que ambas as variantes do YOLOv11 capturaram um número significativamente maior de objetos em comparação as imagens de verdade fundamental no conjunto de dados APACC. Especificamente, o YOLOv11n demonstrou melhor sensibilidade ao detectar uma quantidade maior de objetos do que o YOLOv11s.

No entanto, essa maior sensibilidade geralmente está correlacionada com uma precisão de detecção ligeiramente menor (Figura 17-A). Por outro lado, o modelo YOLOv11s (Figura 17-B) demonstrou assertividade e precisão geral superior. A diferença substancial na

capacidade de detecção entre as duas variantes do YOLOv11 é evidente, particularmente na sua capacidade de identificar células em condições adversas, como desfocada e dentro de aglomerados de células densamente sobrepostas, as quais foram inferidas no *dataset* APACC (Figura 17-C).

4.4 INFERÊNCIA DE VÍDEO EM TEMPO REAL

Para avaliar a velocidade de inferência por meio da taxa de FPS, foi necessário utilizar vídeos públicos de exames de Papanicolau. Porém, inicialmente, conforme o planejamento dos experimentos com vídeo em tempo real, foram empregadas duas câmeras USB CMOS, uma de 2 megapixels (MP) e outra de 5 MP. A câmera de 2 MP apresentou qualidade de imagem inconsistente, caracterizada por faixas verticais quase imperceptíveis na tela (

Figura 18). Esse fato interferiu significativamente na detecção celular, resultando em variações perceptíveis no desempenho de ambas as variantes do modelo YOLOv11. Em contraste, a câmera de 5 MP forneceu um fluxo de imagens mais limpo e estável.

Figura 18 – A câmera CMOS de 2MP demonstrou qualidade de imagem inferior com barras verticais deslizantes, o que na prática afetou a detecção de células cervicais

Fonte: Adaptado pelo Autor 2025.

No decorrer dos experimentos, a latência para ambos os modelos foi quase imperceptível quando a câmera, montada no microscópio, serviu como fonte de imagens em tempo real. Essa latência mínima pode ser atribuída aos atrasos associados a ajustes manuais, como foco e posicionamento da lâmina, juntamente com a confirmação visual pelo microscópio.

Esse processo ocorre antes mesmo que o resultado da inferência do modelo seja

exibido na tela do computador. Essas operações manuais introduzem atrasos na percepção visual que superam a latência computacional do próprio processo de inferência em tempo real. Devido a este fator, para avaliar de forma mais precisa a velocidade de inferência e a taxa de FPS, foram utilizados vídeos públicos de exames de Papanicolau.

Um citologista experiente avaliou todos os modelos usando um conjunto de 100 lâminas. No entanto, devido às restrições impostas pela Lei Geral de Proteção de Dados, que proíbe a divulgação de dados sensíveis de pacientes, os resultados detalhados desta avaliação interna não puderam ser acessados. Consequentemente, duas descobertas particularmente relevantes desta avaliação foram selecionadas para replicação e validação, por meio de vídeos de domínio público. Assim, conduziu-se um experimento para avaliar os modelos utilizando dois vídeos distintos (Quadro 12).

Quadro 12 - Vídeos públicos de exames de Papanicolau para estudo de caso

Vídeo	Duração	Número médio de objetos/células por quadro
1	2:33	10 a 15 objetos
2	2:51	50 a 70 objetos

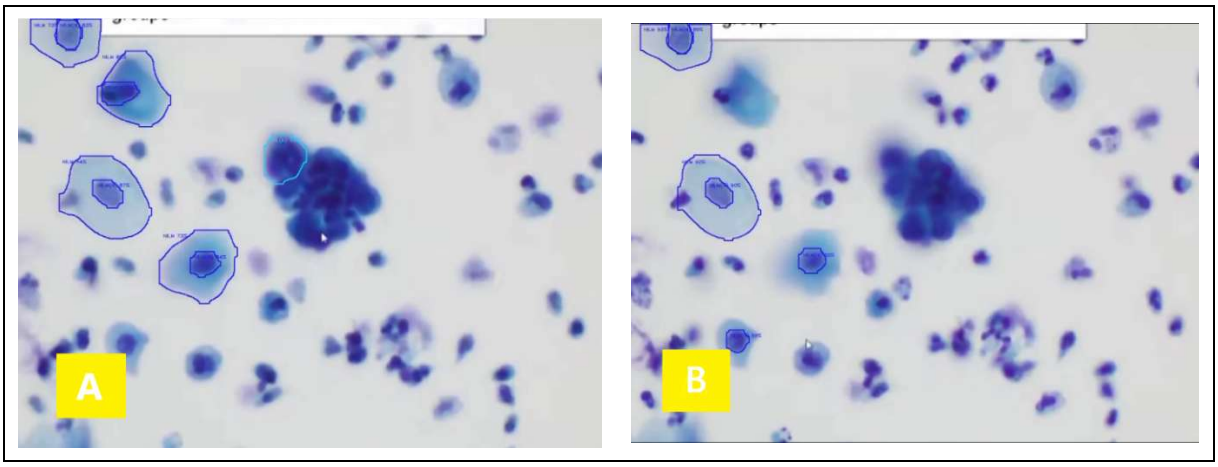
Fonte: Elaborado pela Autor (2025).

O primeiro vídeo[1], com duração de 2 minutos e 33 segundos, apresentou uma média de 10 a 15 objetos detectáveis por quadro. O segundo vídeo[2], com duração de 2 minutos e 51 segundos, apresentou uma maior densidade, variando de 50 a 71 objetos por quadro. Os modelos YOLOv11n e YOLOv11s foram aplicados a cada vídeo. Inicialmente, a inferência foi realizada para as tarefas de detecção de caixas delimitadoras, seguida pelo mesmo processo de segmentação.

Além disso, é possível observar que o vídeo 1 apresenta uma densidade menor de objetos por quadros (Figura 19-A) e em termos sensibilidade, a variante YOLOv11n manteve uma sensibilidade maior, detectando mais objetos, porém, a precisão individual de cada objeto foi menor. No entretanto, a variante YOLOv11s detectou um número menor de objetos por quadro, porém, a precisão é maior em comparação a variante YOLOv11n (Figura 19-B).

[1] Video 1: https://youtu.be/nTnt1Orz2jQ
[2] Video 2: https://youtu.be/G7yncyKUA08?t=15

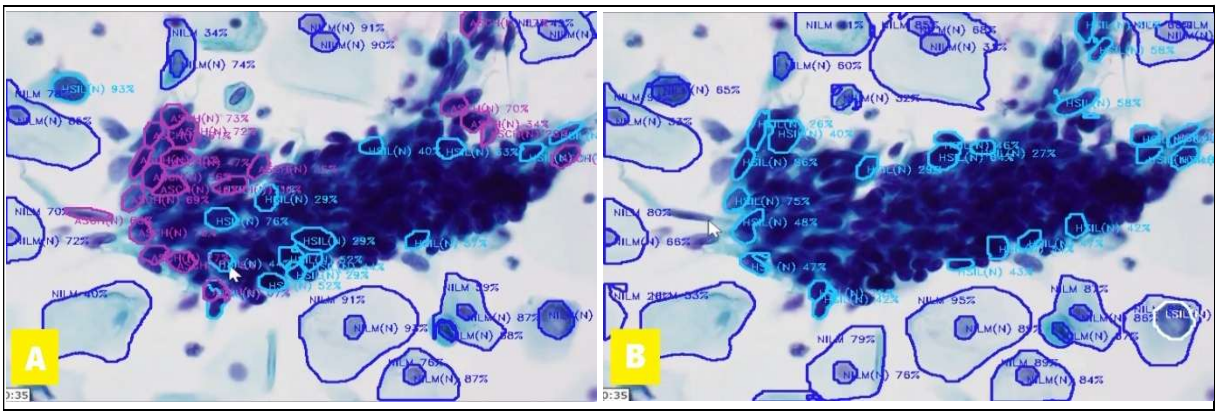
Figura 19 – Imagens de inferência do Vídeo 1 mostrando a detecção de 5 a 10 objetos por quadro. (A) A variante nano detecta um número maior de objetos. (B) Em contraste, a variante *small* produz menos detecções



Fonte: Elaborado pela Autor (2025).

Em contrate, o vídeo 2 tem uma densidade maior de objetos e a variante YOLOv11n exibiu o mesmo comportamento detectando muito mais objetos por quadro com uma precisão menor. Conforme é observado na Figura 20-A, algumas células HSIL tem seus núcleos detectados como classes ASCH. O mesmo não ocorre na variante YOLOv11s, que apesar de detectar menos objetos, tem uma precisão melhor identificando corretamente as células com lesões HSIL, além disso, detectou com exatidão a única célula com lesão LSIL, que se encontra no canta inferior direito da Figura 20-B.

Figura 20 – O Vídeo 2 contém aproximadamente 50 a 75 objetos por quadro. (A) O YOLOv11n apresenta menor precisão, mas detecta um número maior de objetos. (B) O YOLOv11s realiza menos detecções, porém alcança maior precisão



Fonte: Elaborado pela Autor (2025).

As métricas de desempenho resultantes, incluindo taxa de FPS, velocidade de inferência principal, tempo de pré-processamento, tempo de pós-processamento e tempo total

médio por imagem, estão detalhados no Quadro 13, que ilustra as taxas de FPS comparativas entre as duas variantes do YOLOv11, destacando várias observações importantes:

- A detecção de caixas delimitadoras pelo YOLOv11n apresenta uma redução significativa de FPS no Vídeo 2. Essa queda acentuada é diretamente atribuída a um aumento drástico no tempo de pós-processamento, passando de 3,7 para 75 ms, evidenciando sua limitada escalabilidade diante de uma alta densidade de objetos.
- O YOLOv11s para detecção de caixas delimitadoras, por outro lado, manteve relativa estabilidade, apresentando apenas uma pequena diminuição na taxa de FPS mesmo em cenários com alta densidade de objetos.
- Para tarefas de segmentação, ambos os modelos demonstram alta estabilidade, com variações mínimas de FPS entre os dois vídeos, independentemente da densidade de objetos.
- De modo geral, o YOLOv11n apresenta desempenho superior em FPS tanto na detecção de caixas delimitadoras quanto na segmentação. No entanto, sua robustez diminui consideravelmente com o aumento da complexidade visual, principalmente na detecção de caixas delimitadoras.

Quadro 13 - YOLOv11 *nano* e *small*: resultados da inferência (tempo em ms.)

YOLOv11	Vídeo	FPS	T. Pré-processamento	T. Pós-processamento	Tempo de Inferência
(n) BB	1	**21,36**	24,9	3,7	35,9
	2	**7,49**	28,4	75,0	110,1
(s) BB	1	12,03	60,5	4,0	71,7
	2	10,46	59,2	9,8	75,7
(n) SEG.	1	**15,76**	24,9	3,8	35,1
	2	**15,56**	25,0	3,8	36,1
(s) SEG.	1	9,64	60,7	4,0	72,0
	2	9,65	60,7	4,0	72,1

Fonte: Elaborado pela Autor (2025).

Os experimentos sobre a velocidade de inferência usando arquivos de vídeo revelaram padrões distintos e compensações que exigem consideração cuidadosa em estudos futuros. Na avaliação das variantes do YOLOv11 para métricas de detecção baseadas em vídeo, parâmetros como velocidade de inferência, tempos de pré-processamento, pós-processamento e tempo total de inferência foram meticulosamente avaliados.

Os dados desses experimentos indicam que o YOLOv11n supera significativamente o YOLOv11s em cenários com baixa densidade de objetos, alcançando um aumento de

aproximadamente 73% na taxa de FPS no Vídeo 1. No entanto, quando submetido a altas cargas computacionais (Vídeo 2), o YOLOv11n apresentou uma drástica redução de aproximadamente 35% na taxa de quadros, de 21,36 para 7,49 FPS.

Essa queda atribui-se principalmente a um aumento abrupto no tempo de pós-processamento de 3,7 para 75,0 ms., evidenciando a desempenho limitada do modelo nano em ambientes com alta densidade de objetos. Em contraste, o YOLOv11s demonstrou maior estabilidade sob variação de carga, apresentando uma redução de apenas 13% na taxa de quadros entre os dois vídeos. Esse comportamento sugere que, embora mais lento em contextos menos exigentes, o modelo *small* oferece robustez e consistência superiores em cenários mais desafiadores.

O desempenho da segmentação apresentou um perfil notavelmente diferente. O YOLOv11n manteve uma taxa de FPS praticamente estável (15,76 vs. 15,56) entre os dois vídeos, demonstrando uma notável consistência apesar do aumento no número de objetos. A estabilidade do seu tempo total por imagem (~ 36 ms.) reforça ainda mais essa eficiência.

Esse resultado sugere que, diferentemente da detecção de caixas delimitadoras, o custo computacional da segmentação é menos sensível à densidade de objetos na implementação analisada. O YOLOv11s, embora também tenha mantido a estabilidade, operou com um desempenho absoluto inferior, com FPS em torno de 9,6 em ambos os vídeos e um tempo total por imagem, quase o dobro do YOLOv11n (~72 vs. ~36 ms.). Isso indica um custo computacional inerentemente maior por quadro para a segmentação do YOLOv11s, sem um ganho perceptível em estabilidade adicional em comparação com o YOLOv11n para essa tarefa.

5 CONCLUSÃO

Em um contexto geral, os resultados produzidos nesta pesquisa foram satisfatórios; porém, há certas considerações que é necessário ressaltar. Nesta seção, discorre-se sobre os principais destaques que tornaram possível concretizar os objetivos deste trabalho.

No que diz respeito aos trabalhos correlatos descritos na seção 1.7, houve resultados significativos em comparação a esta pesquisa. As propostas similares de microscopia assistida por IA atingiram seus objetivos com êxito, mas, quando comparadas a esta pesquisa, os resultados aqui obtidos se destacaram. Quanto ao tempo de inferência dos modelos propostos, a maioria foi apresentada em segundos, enquanto esta pesquisa obteve sucesso com o tempo de detecção por imagem em milissegundos, superando todos as soluções propostas. No quesito do desempenho dos modelos, obteve-se uma precisão de 99,30% no melhor modelo treinado do YOLOv11.

Quanto às três perguntas deste trabalho, em resposta a questão de pesquisa “Q01”, que levanta a possibilidade de utilizar apenas bases de dados públicas para produzir um modelo treinado de IA em visão computacional para a detecção de células cervicais, não foi possível obter uma resposta precisa. As bases públicas revelaram uma proporção desbalanceada na distribuição nas amostras das classes de imagens cervicais, pois, mesmo unido três diferentes *datasets*, as classes referentes a imagens normais apresentaram mais de 50% de representatividade sobre as demais.

Tal fato exigiu um trabalho maior na aplicação de técnicas de aumento de dados para mitigar o desbalanceamento das amostras. Apesar do modelo ter alcançado uma ótima precisão final, sugere-se uma próxima interação com instituições médicas para criação de um banco dedados com distribuição das classes de forma mais balanceada.

Apesar da obtenção de uma resposta não tão positiva sobre a questão “Q01”, isso não impediu uma resposta totalmente favorável para à questão de pesquisa “Q02”, que levanta a possibilidade de, a partir apenas de recursos de domínios públicos, propor uma solução de automatização de citologia cervical alternativa que possa ser totalmente reproduzida por outros pesquisadores.

Considerando que as bases de dados públicas não são totalmente indicadas para desenvolvimento de um sistema automatizado para citologia cervical convencional, ainda

assim, foi possível obter um protótipo totalmente funcional, indicando a viabilidade de reproduzir todos esses experimentos por outros pesquisadores ou, simplesmente, aprimorar este presente trabalho em novas interações.

Quanto a última questão de pesquisa "Q03", que indaga se arquiteturas menores podem melhorar o desempenho de treinamento ao aumentar a precisão mediante a implementação de imagens com dimensões superiores a 640 x 640 *pixels*, a resposta foi obtida com exatidão. Tanto o modelo YOLOv11n quanto o YOLOv11s alcançaram precisão superior a 90%.

Em relação às cinco hipóteses levantadas neste estudo, correlacionadas às três questões da pesquisa, como pode-se observar na seção 1.4, analisou-se:

- **H1:** o uso de câmeras CMOS acopladas a microscópios ópticos permite a captura de imagens com qualidade suficiente para inferência em tempo real por modelos de visão computacional. Esta hipótese se mostrou totalmente viável, pois as câmeras foram suficientes como fonte de imagens de video em tempo real.

- **H2:** a aplicação de modelos de arquitetura reduzida, como YOLOv11 nas variantes *nano* e *small*, proporciona desempenho satisfatório (mAP >= 60%) na detecção de células cervicais em video em tempo real. Esta hipótese também se mostrou verdadeira, pois a limitação de precisão dos modelos foi superada e as duas variantes atingiram desempenho superior 90%. Parte deste resultado está diretamente ligada à hipótese **H3:** o uso de imagens com resoluções superiores a 640 x 640 *pixel* melhora a precisão e o *recall* dos modelos de detecção, sem comprometer significativamente a velocidade de inferência. A utilização de imagens com resoluções superiores no treinamento se mostrou eficaz e contribuiu para o bom desempenho dos modelos.

- **H4*:*** a automação do exame de Papanicolau convencional com inteligência artificial reduz a incidência de falso negativos em comparação ao método CML, contribuindo para maior confiabilidade diagnóstica. Esta hipótese ficou associada a trabalhos futuros, pois, devido às limitações encontradas durante o desenvolvimento desta pesquisa, não foi possível implantar o protótipo inicial em uma instituição médica com estrutura laboratorial.

- **H5:** as bases de dados públicas, juntamente com a aplicação de técnicas de aumento de dados (*data augmentation*), fornecem amostras suficiente de lesões celulares capazes de gerar bom desempenho no treinamento de modelos CNN do tipo YOLO. Apesar das implicações das bases de dados públicas, em relação ao numero de exemplares de células normais representarem 50% das amostras totais, a aplicão da técnica de aumento de dados mostrou-se eficiente, mitigou o problema e proporcionou bom resultado geral no desempenho dos modelos em ambas as variantes do YOLOv11.

Por fim, em relação aos objetivos desta pesquisa, pode-se afirmar de forma mais direta que todos foram atingidos. No que diz respeito ao que foi proposto incialmente, os experimentos, em sua totalidade, foram bem-sucedidos e revelaram achados importantes para a conclusão deste estudo.

Em resumo, foi possível desenvolver uma solução automatizada assistida por IA que responde à questão norteadora deste presente trabalho: "***uma solução baseada em IA utilizando uma CNN YOLOv11, pode auxiliar na detecção automatizada de lesões percursoras de câncer em tempo real, visando mitigar a sobrecarga dos citologistas e aumentar a precisão e velocidade diagnóstica dos exames de Papanicolau?***".

5.1 ACHADOS RELEVANTES

Durante a condução dos experimentos deste presente trabalho, alguns resultados foram considerados relevantes do ponto de vista do que se esperava. Nesta seção descreve-se pontos importantes sobre os resultados obtidos durante toda a fase de experimentação.

Primeiro, destaca-se a os experimentos com de inferência de vídeo com câmeras digitais acopladas ao microscópio, revelando que, mesmo com um sistema com baixa latência, este ainda se mostra útil para como ferramenta de análise assistida por IA. Ou seja, mesmo que ainda lento, a latência não é percebida durante o uso do microscópio. Tal fato levou à necessidade de experimentação para avaliação de desempenho da taxa de FPS em vídeos públicos.

Os experimentos em vídeos públicos, também revelaram certos resultados inesperados. Os dois vídeos continham diferentes densidades de objetos por cenas. O vídeo que continha menos números de objetos obteve uma boa taxa de FPS. Porém, quando o número de

objetos aumentou, a taxa de FPS caiu. Este resultado revelou dois achados relevantes:

- Quanto maior o número de objetos na cena, mais tempo de pós-processamento para cada objeto será demandado e menor será a taxa de FPS;
- Outro fator relacionando a densidade de objetos foi o número de classes dobrado, ou seja, uma classe para a célula inteira e outra para apenas para o núcleo. Apesar de ser uma proposta ambiciosa, percebeu-se que, quanto maior o número de classes, maior é o custo computacional exigido no tempo de inferência em vídeo em tempo real.

Em relação à precisão das variantes menores do YOLOv11, as imagens com resoluções superiores a 640 x 640 *pixels*, que, neste trabalho, operam com o dobro :1280 x 1280 *pixels*, produziram resultados além do esperado, como:

- O tempo de treinamento foi elevado: para a variante *nano* foram 11 épocas com tempo 36 horas cada. Já a variante *small* demandou 6 épocas de 78 horas cada, um tempo mais longo se comparado aos modelos que treinados com imagens com tamanho padrão de 640 x 640 *pixels*;
- As imagens com resolução de 1280 x 1280 *pixels* possibilitaram a superação do limite de precisão, especialmente na variante nano do YOLOv11, que costumava não passar de 60%, acabando por atingir uma precisão de 99%;
- Apesar da grande diferença no número de épocas e no tempo de treinamento, tanto o modelo *nano* quanto o *small* obtiveram praticamente o mesmo desempenho. O *nano*, na tarefa de detecção por caixas delimitadoras, atingiu um mAP de 98%, enquanto o *small* alcançou 99,30%. O mesmo ocorreu na tarefa de segmentação de instâncias, em que o *nano* obteve um mAP de 95% e o *small* 96%. Em ambos os casos, a diferença foi de aproximadamente 1%.

Outro fator percebido durante os experimentos de inferência de vídeo foi o fato de que a variante *nano* demonstrou maior sensibilidade de detecção de objetos, capturando muito mais células cervicais, embora os percentuais de confiança tenham sido relativamente mais baixos. Ao comparar com a variante *small*, observou-se menor sensibilidade, mas a detecção dos objetos concentrou-se apenas nas células com percentuais de confiança mais elevados. Esse resultado comprovou o melhor desempenho na variante *small* em termos de precisão. Além disso, o modelo *nano* se mostrou-se menos preciso na detecção da classe HSIL em comparação

a variante *small*.

Vale registrar que, apesar da quantidade de classes e das condições extremas nos testes de inferência com alta densidade de objetos por cena, o modelo *nano*, ainda é o mais rápido em inferência em tempo real, apresentando o menor tempo de detecção por imagem (35,9 milissegundos) a uma taxa de 21,9 FPS, comprovando sua eficiência e rapidez.

5.2 LIMITAÇÕES DA PESQUISA

Devido a Lei Geral de Proteção de Dados, não foi possível utilizar nesta pesquisa nenhum tipo de dados ou material genético de pacientes, bem como trabalhar com alguma instituição médica com estrutura laboratorial. Tal restrição, limitou a coleta de amostras diretamente de lâminas em condições reais.

Em consequência desta limitação, não houve um envolvimento de uma equipe especializada para avaliar de forma extensiva o protótipo inicial. Entretando, é importante salientar, houve acompanhamento por parte de um único citologista especializado que testou, avaliou e aprovou o protótipo inicial de forma independente.

E por fim, como não houve uma interação com uma entidade médica ou laboratorial, não foi possível a implantação, bem como a possibilidade de coleta de imagens oriundas de falhas de deteção do modelo, que potencialmente seriam uteis como métrica quantitativa para avaliar a diminuição dos exames falsos negativos.

5.3 TRABALHOS FUTUROS

Ao final desta primeira interação, com uma avaliação dos resultados, percebeu-se que ainda há oportunidades de aprimoramento deste protótipo inicial, que, apesar de funcional, pode ser melhorado e continuado em pesquisas baseadas no mesmo princípio deste trabalho.

Tendo em vista a necessidade de parceria com entidades médicas, pretende-se trabalhar com laboratórios que possam interagir com o protótipo e fornecer feedback, de modo a mensurar o progresso na redução dos exames falsos negativos.

Ainda neste sentido, em uma nova interação pretende-se adquirir uma permissão concedidas por meio comitês de éticas especializados neste tipo de pesquisa, para que se possa utilizar dados de pacientes, para a digitalização de lâminas em condições reais de uso.

5.4 CONSIDERAÇÕES FINAIS

Considerando o pouco tempo de desenvolvimento desta pesquisa, um ano e oito meses, dedicados ao estudo, seleção de publicações, levantamento teórico, montagem das bases de dados, ajustes de hiperparâmetros, treinamento e condução de experimentos, considera-se que o objetivo principal deste trabalho foi atingido com sucesso.

Portanto, tendo em vista toda a classe feminina que é atingida por consequência dos exames falsos negativos, dedica-se esta pesquisa como uma contribuição à sociedade.

Como contribuição para área de TI, esta pesquisa abre portas para mais estudos no âmbito de detecção de vídeo em tempo real, tirando proveito da velocidade das variantes menores do modelo YOLOv11 e vencendo a barreira do limite de precisão.

E por fim, como contribuição a classe acadêmica, em especial às pesquisas desenvolvidas no Brasil, dedica-se esta pesquisa como legado, para que se possa ser continuada e cada vez mais beneficiar os exames assistidos por IA por meio vídeo em tempo real.

REFERÊNCIAS

ACKERMANN, A. E. F.; SELLITTO, M. A. Métodos de previsão de demanda: uma revisão de literatura. **Innovar**, 32, Julo-Setembro 2022. 83-99.

ALCARAZ-CHAVEZ, J. E. et al. Real-Time Tracking and Detection of Cervical Cancer Precursor Cells: Leveraging SIFT Descriptors in Mobile Video Sequences for Enhanced Early Diagnosis. **Algorithms**, Michoacán, Mexico, 17, n. 7, 12 jul. 2024. Disponivel em: <https://d1wqtxts1xzle7.cloudfront.net/116743261/algorithms_17_00309-libre.pdf?1720831285=&response-content-disposition=inline%3B+filename%3DReal_Time_Tracking_and_Detection_of_Cerv.pdf&Expires=1749702097&Signature=JwcNwPViybAEUXUVoMgn9AlLrBsz4AtP8C0SyMdtm>. Acesso em: 5 jan. 2025.

ARLOT, S.; CELISSE, A. A survey of cross-validation procedures for model selection (Uma pesquisa sobre procedimentos de validação cruzada para seleção de modelos). **Statistics Surveys**, 4, 2010. 40–79. Acesso em: 20 Jan 2026.

BADI, M. et al. TensorFlow: Large-scale machine learning on heterogeneous systems. **Machine Learning**, 6, 1991. 37–66. Acesso em: 12 Jan 2025.

BARROS, A. L. D. S. et al. **Técnico em citopatologia:** caderno de referência 1: citopatologia ginecológica. [S.l.]: Ministério da Saúde (Brasil), 2012. 194 p. Acesso em: 1 Dez 2024.

BASAK, et al. Cervical cytology classification using PCA and GWO enhanced deep features selection. **SN Computer Science**, 2, n. 369, 7 jun. 2021.

BOCHKOVSKIY, A.; WANG, C.-Y.; LIAO, H.-Y. M. YOLOv4: Optimal Speed and Accuracy of Object Detection, 23 Abril 2020. Acesso em: 10 Fev 2026.

BORDWELL, D.; THOMPSON, K. **A Arte do Cinema:** Uma Introdução. 5. ed. São Paulo: Unifesp, 2018.

BOUREAU, Y.; PONCE, J.; LECUN, Y. A theoretical analysis of feature pooling in visual recognition. **Proceedings of the 27th International Conference on Machine Learning (ICML).**, 2010. 111-118. Disponivel em: <https://www.rogerioferis.com/VisualRecognitionAndSearch2013/material/Class3Sparse2.pdf>. Acesso em: 18 Jan 2025.

BRASIL, MINISTÉRIO DA FAZENDA. **Princípios Básicos:** o que você precisa saber sobre as transferências fiscais da União. Brasília, DF: Presidência da República, mar. 2016. 22 p. Disponivel em: <https://cdn.tesouro.gov.br/sistemas-internos///apex//producao//sistemas//thot//arquivos//publicacoes/28549_909191/anexos/4540_910628///pge_cartilha_principios_basicos.pdf?v=1281>. Acesso em: 2022 Janeiro 27.

CAETANO, R. et al. Cost-effectiveness of early screening of cervix neoplasms in Brazil. **Revista de Saúde Coletiva**, 16, n. 1, 2006. 99-118. Disponivel em: <https://www.scielosp.org/pdf/physis/2006.v16n1/99-118/pt>. Acesso em: 29 Dezembro 2023.

CHAPMAN, P. et al. **CRISP-DM 1.0:** Step-by-step data mining guide. [S.l.]: [s.n.], 2000. Disponivel em: <https://www.semanticscholar.org/paper/CRISP-DM-1.0%3A-Step-by-

step-data-mining-guide-Chapman-Clinton/54bad20bbc7938991bf34f86dde0babfbd2d5a72>. Acesso em: 10 Julho 2022.

COLONELLI,. Avaliação do desempenho da citologia em meio líquido versus citologia convencional no Sistema Único de Saúde. **Universidade Federal de São Paulo (UNIFESP)**, São Paulo, 2014. Dissertação de Mestrado, com número de páginas (102 f.). Disponivel em: <https://www.academia.edu/download/38237195/Avaliacao_do_desempenho_da_citologia_em_meio_liquido_versus_citologia_convencional_no_SUS.pdf>. Acesso em: 14 Nov 2024.

CONSOLARO , E. L.; MARIA-ENGLER,. **Citologia Clínica Cérvico-Vaginal:** Texto e Atlas. São Paulo: Roca, 2012. Acesso em: 18 Nov 2024.

COX, T. J.; BARBARA,. Liquid-based cytology: evaluation of effectiveness, cost-effectiveness, and application to present practice. **Journal of the National Comprehensive Cancer Network**, 6, 2004. 597-611. Disponivel em: <https://jnccn.org/view/journals/jnccn/2/6/article-p597.xml?content=contributorNotes-7604>. Acesso em: 16 Nov 2024.

DUARTE, G. J. et al. **Guia brasileiro de análise de dados:** armadilhas & soluções. Brasília/DF: Enap, 2021. 251 p. ISBN ISBN: 978-65-87791-25-8.

EVERINGHAM, M. et al. The PASCAL Visual Object Classes (VOC) Challenge. **International Journal of Computer Vision**, 88, n. 2, 2010. 303–338. Disponivel em: <https://www.pure.ed.ac.uk/ws/portalfiles/portal/7879113/ijcv_voc09.pdf>. Acesso em: 23 Jul 2025.

FELZENSZWALB, P. F.; HUTTENLOCHER, D. P. Efficient Graph-Based Image Segmentation. **International Journal of Computer Vision**, 59, n. 2, 2004. 167–181. Disponivel em: <https://link.springer.com/article/10.1023/B:VISI.0000022288.19776.77>. Acesso em: 29 Mar 2025.

FERREIRA, D. S. et al. Saliency-driven system models for cell analysis with deep learning, 182, 2019. 105053. Disponivel em: <https://www.sciencedirect.com/science/article/pii/S0169260718316870>. Acesso em: 11 Nov 2024.

FILHO, A. L. et al. DCS liquid-based system is more effective than conventional smears to diagnosis of cervical lesions: study in high-risk population with biopsy-based confirmation. **Gynecologic Oncology**, 97, n. 3, 2005. 497-500. Disponivel em: <linkinghub.elsevier.com/retrieve/pii/S009082580500082X>. Acesso em: 10 Janeiro 2025.

FUKUSHIMA, K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. **Biological Cybernetics**, 36, n. 4, 1980. 193-202. Disponivel em: <https://www.cs.princeton.edu/courses/archive/spr08/cos598B/Readings/Fukushima1980.pdf>. Acesso em: 15 Dez 2024.

GAVRILESCU, R. et al. Faster R-CNN: An Approach to Real-Time Object Detection. **International Conference on Electrical and Power Engineering (EPE)**, 18-19 Out 2018.

Disponivel em: <https://ieeexplore.ieee.org/abstract/document/8559776/authors#authors>. Acesso em: 26 Nov 2024.

GÉRON, A. **Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow:** Concepts, Tools, and Techniques to Build Intelligent Systems. 2a. ed. [S.l.]: O'Reilly Media, 2019. Acesso em: 120 Jan 2025.

GIL, A. C. **Como elaborar projetos de pesquisa**. 6. ed. São Paulo-SP: Atlas, 2017.

GIRIANELLI, V. R. et al. Comparação do desempenho do teste de captura híbrida II para HPV, citologia em meio líquido e citologia convencional na. **Revista Brasileira de Cancerologia**, 50, n. 3, 2004. 225-226. Disponivel em: <file:///C:/Users/fcalp/Downloads/2027-Texto%20do%20artigo-14737-1-10-20210514%20(1).pdf>. Acesso em: 10 Mar 2025.

GIRSHICK, R. Fast R-CNN. **In Proceedings of the IEEE International Conference on Computer Vision (ICCV)**, 2015. 1440-1448. Disponivel em: <https://ieeexplore.ieee.org/document/7410526>. Acesso em: 18 Jan 2025.

GIRSHICK, R. et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**, 2014. 580-587. Disponivel em: <https://openaccess.thecvf.com/content_cvpr_2014/papers/Girshick_Rich_Feature_Hierarchies_2014_CVPR_paper.pdf>. Acesso em: 13 Dez 2024.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. [S.l.]: MIT Press, 2016. Acesso em: 29 Mar 2025.

GUPTA, R. et al. Artificial intelligence-driven digital cytology-based cervical cancer screening: is the time ripe to adopt this disruptive technology in resource-constrained settings? A literature review. **Journal of Digital Imaging**, 36, n. 4, 2023. 1643-1652. Disponivel em: <0.1007/s10278-023-00821-0>. Acesso em: 23 Nov 2024.

HASHMI, et al. Comparison of liquid-based cytology and conventional Papanicolaou smear for cervical cancer screening: an experience from Pakistan. **Cureus**, 12, n. 12, 2020. 12018. Disponivel em: <https://assets.cureus.com/uploads/original_article/pdf/48210/1612431941-1612431937-20210204-18204-jpkv6a.pdf>. Acesso em: 22 Jan 2024.

HE, et al. Mask R-CNN. **Proceedings of the IEEE International Conference on Computer Vision (ICCV)**, 2017. 2961–2969. Disponivel em: <https://openaccess.thecvf.com/content_ICCV_2017/papers/He_Mask_R-CNN_ICCV_2017_paper.pdf>. Acesso em: 27 Jan 2025.

JANTZEN, ; NORUP, ; GEORGIOS , D. Pap-smear benchmark data for pattern classification. **Nature inspired smart information systems (NiSIS 2005)**, 2025. 1-9. Disponivel em: <https://orbit.dtu.dk/en/publications/pap-smear-benchmark-data-for-pattern-classification>. Acesso em: 20 Mai 2025.

JIANG, et al. A systematic review of deep learning-based cervical cytology screening: from cell identification to whole slide image analysis. **Artificial Intelligence Review**, 56, n. Supplement 2, 2023. 2687–2758. Disponivel em:

<https://link.springer.com/content/pdf/10.1007/s10462-023-10588-z.pdf>. Acesso em: 10 Novembro 2023.

JOCHER, et al. YOLOv5: Implementation Guide and Loss Analysis, 2020. Acesso em: 11 Nov 2024.

KHANAM, ; HUSSAIN,. YOLOv11: An Overview of the Key Architectural Enhancements. **ArXiv**, arXiv:2410.17725, 23 Out 2024. Disponivel em: <https://arxiv.org/pdf/2410.17725>. Acesso em: 18 Jan 2025.

KHANAM, R.; HUSSAIN, M. YOLOv11: An Overview of the Key Architectural Enhancements, 2024. Disponivel em: <https://arxiv.org/abs/2410.17725>.

KIM,. Tutorial: Convolutional Neural Networks (CNN) with TensorFlow 2.0. **I-Systems Research Group**, 2021. Disponivel em: <https://i-systems.github.io/tutorial/KIM/210818/01_CNN_tf2.html>. Acesso em: 22 Set 2025.

KOSS, G.; GOMPEL ,. **Koss' Gynecologic Cytopathology with Histologic and Clinical Correlations**. 4ª. ed. [S.l.]: Roca, 2006. Acesso em: 12 Jan 2025.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. ImageNet Classification with Deep Convolutional Neural Networks. **Advances in Neural Information Processing Systems**, 25, 2012. Acesso em: 17 Nov 2024.

KUNDU,. YOLO: Algorithm for Object Detection Explained [+Examples]. **V7 Labs**, 2023. Disponivel em: <https://www.v7labs.com/blog/yolo-object-detection:cite[1]>. Acesso em: 9 Set 2025.

KUPAS, D. et al. Model of Watershed Segmentation in Deep Learning Method to Improve Identification of Cervical Cancer at Overlay Cells. **In International Conference on Neural Information Processing**, 12, n. 2, 2023. 813-819. Disponivel em: <https://www.temjournal.com/content/122/TEMJournalMay2023_813_819.pdf>. Acesso em: 12 Dez 2024.

LECUN, et al. Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, 86, n. 11, 1998. 2278-2324. Acesso em: 18 Fev 2025.

LECUN, Y. et al. Handwritten Digit Recognition: Applications of Neural Net Chips and Automatic Learning. **Neuro-computing: Algorithms, Architectures and Applications**, 68, n. 303, 2012. Acesso em: 11 Nov 2025.

LEE, Y.-M. et al. Beyond the Microscope: A Technological Overture for Cervical Cancer Detection. **Diagnostics**, 13, n. 19, 2023. 3079. Disponivel em: <https://www.mdpi.com/2075-4418/13/19/3079>. Acesso em: 17 Dez 2024.

LIANG, et al. Exploring Contextual Relationships for Cervical Abnormal Cell Detection. **IEEE Journal of Biomedical and Health Informatics**, 27, n. 8, 2023. 4086-4097. Disponivel em: <https://ieeexplore.ieee.org/document/10124979>. Acesso em: 26 Fev 2025.

LIANG, Y. et al. Exploring contextual relationships for cervical abnormal cell detection. **IEEE Journal of Biomedical and Health Informatics**, 27, n. 8, 2023. 4086-4097. Disponivel em: <https://ieeexplore.ieee.org/abstract/document/10124979>. Acesso em: 20 dez. 2024.

LIN, M.; CHEN, Q.; YAN, S. Network in Network. **arXiv:1312.4400**, 2013. Disponivel em: <https://arxiv.org/abs/1312.4400>.

LIU, W. et al. SSD: Single Shot MultiBox Detector. **Computer Vision – ECCV 2016: 14th European Conference**, Amsterdam, The Netherlands, 2016. Disponivel em: <https://arxiv.org/pdf/1512.02325>. Acesso em: 22 Mar 2025.

MAAS, L.; HANNUN, Y.; NG, Y. Rectifier Nonlinearities Improve Neural Network Acoustic Models. **Proc. ICML (Proceedings of the International Conference on Machine Learning)**, 30, n. 1, Jun 2013. 3. Disponivel em: <https://www.awnihannun.com/papers/relu_hybrid_icml2013_final.pdf>. Acesso em: 17 Jan 2025.

MAKDE, M. M.; SATHAWANE, P. Liquid-based cytology: Technical aspects. **Technical aspects. Cytojournal**, 19, 2022. 1-10. Disponivel em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9345114/>. Acesso em: 19 Nov 2024.

MARTÍNEZ-PLUMED, F. et al. CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. **IEEE Transactions on Knowledge and Data Engineering**, v. 33, p. 3048-3061, 01 Agosto 2021. ISSN 8. Disponivel em: <https://ieeexplore.ieee.org/document/8943998>. Acesso em: 08 Junho 2022.

MESA ,. Understanding YOLOv5 Loss: A Comprehensive Analysis. **Medium**, 2024. Disponivel em: <https://pgmesa.medium.com/understanding-yolov5-loss-a-comprehensive-analysis-ffe35e6203e0>. Acesso em: 10 Jan 2025.

MOHAMMED, S. Y. Architecture review: Two-stage and one-stage object detection. **Franklin Open**, 2025. 100322. Disponivel em: <https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/Redmon_You_Only_Look_CVPR_2016_paper.pdf>. Acesso em: 123 Mar 2025.

MURCH, W. **In the Blink of an Eye:** A Perspective on Film Editing. 2a. ed. Los Angeles: Silman-James Press, 2001.

NAIR, ; HINTON, E. Rectified linear units improve restricted boltzmann machines. **Proceedings of the 27th international conference on machine learning (ICML-10)**, 2010. 807-814. Disponivel em: <https://www.cs.toronto.edu/~hinton/absps/reluICML.pdf>. Acesso em: 1 Dez 2024.

NAKISIGE, C.; SCHWARTZ , M.; NDIRA, A. O. Cervical cancer screening and treatment in Uganda. **Gynecologic Oncology Reports**, 20, 2017. 37-40. Disponivel em: <https://www.sciencedirect.com/science/article/pii/S2352578917300115>. Acesso em: 8 Dezembro 2024.

NEUBECK , ; GOOL,. Efficient Non-Maximum Suppression. **18th International Conference on Pattern Recognition (ICPR'06)**, Hong Kong, 3, 20-24 Ago 2006. 850-855. Disponivel em: <https://www.semanticscholar.org/paper/Efficient-Non-Maximum-Suppression-Neubeck-Gool/52ca4ed04d1d9dba3e6ae30717898276735e0b79>. Acesso em: 11 Mar 2025.

NIDHAL , J. et al. YOLO Evolution: A Comprehensive Benchmark and Architectural Review of YOLOv12, YOLO11, and Their Previous Versions. **arXiv**, 2024. Disponivel em: <https://arxiv.org/pdf/2411.00201>. Acesso em: 30 Mar 2025.

NIELSEN, M. A. **Neural Networks and Deep Learning**. [S.l.]: Determination Press, 2015. Disponivel em: <http://neuralnetworksanddeeplearning.com/>. Acesso em: 23 Mar 2025.

O guia "Fine-Tuning in Machine Vision: A Beginner's Guide". **UNITX LABS**, 2025. Disponivel em: <https://www.unitxlabs.com/resources/fine-tuning-machine-vision-guide/>. Acesso em: 29 Mar 2025.

OLIVER, M. The Comprehensive Guide to Training and Running YOLOv8 Models on Custom *Datasets*. **Medium**, 2024. Disponivel em: <https://medium.com/data-science/the-comprehensive-guide-to-training-and-running-yolov8-models-on-custom-datasets-22946da259c3>. Acesso em: 15 Jan 2025.

OPENCV Selective Search for Object Detection. **PyImageSearch**, 2020. Disponivel em: <https://pyimagesearch.com/2020/06/29/opencv-selective-search-for-object-detection/>. Acesso em: 15 Mar 2025.

OROZCO-MONTEAGUDO, M. et al. Combined Hierarchical Watershed Segmentation and SVM Classification for Pap-Smear Cell Nucleus Extraction. **Computación y Sistemas**, 16, n. 2, 2012. 133-145. Disponivel em: <https://www.polibits.cidetec.ipn.mx/ojs/index.php/CyS/article/viewFile/1379/1428>. Acesso em: 29 nov. 2024.

PADILLA, R.; NETTO, S. L.; DA SILVA, E. A. A Survey on Performance Metrics for Object Detection Algorithms. **In 2020 7th International Conference on Computer and Communications Management (ICCCM)** , 2020. 248-255. Disponivel em: <https://www.researchgate.net/profile/Rafael-Padilla/publication/343194514_A_Survey_on_Performance_Metrics_for_Object-Detection_Algorithms/links/5f1b5a5e45851515ef478268/A-Survey-on-Performance-Metrics-for-Object-Detection-Algorithms.pdf>. Acesso em: 12 out. 2024.

PADILLA, R.; NETTO, S.; SILVA, E. D. A Survey on Performance Metrics for Object-Detection Algorithms. **2020 International Conference on Systems, Signals and Image Processing**, 2020. 248-255. Disponivel em: <Padilla/publication/343194514_A_Survey_on_Performance_Metrics_for_Object-Detection_Algorithms/links/5f1b5a5e45851515ef478268/A-Survey-on-Performance-Metrics-for-Object-Detection-Algorithms.pdf>. Acesso em: 15 Nov 2024.

PEREZ, L.; WANG,. The effectiveness of data augmentation in image classification using deep learning. **arXiv preprint**, 2017. arXiv:1712.04621. Acesso em: 29 Jun 2025.

PIMENTEL, C. S. et al. Convolutional Neural Networks for Detecting Ship Hatch Closing Moment: A Case Study Using YOLO Family. **Workshop de Visão Computacional (WVC)**, 12 Nov 2023. 54-59. Disponivel em: <10.5753/wvc.2023.27532:cite[5].>. Acesso em: 20 Nov 2023.

PIMENTEL, F. C. S. et al. Convolutional Neural Networks for Detecting Ship Hatch Closing Moment: A Case Study Using YOLO Family. **In Workshop de Visão Computacional (WVC)**, nov. 2023. 54-59. Disponivel em:

<https://sol.sbc.org.br/index.php/wvc/article/view/27532/27344>. Acesso em: 20 nov. 2023.

PLISSITI, M. E. et al. Sipakmed: A new dataset for feature and image based classification of normal and pathological cervical cells in pap smear images. **25th IEEE International Conference on Image Processing (ICIP)**, 2018. 3144-3148. Disponivel em: <https://www.academia.edu/download/77749683/C072_Plissiti_icip_2018_Athens.pdf>. Acesso em: 25 Out 2025.

PLOTNIKOVA, V.; DUMAS, M.; MILANI, F. P. Adapting the CRISP-DM Data Mining Process: A Case Study in the Financial Services Domain. **Research Challenges in Information Science. RCIS 2021. Lecture Notes in Business Information Processing**, Limassol, Cyprus, 415, 08 Maio 2021. 55-71. Disponivel em: <https://www.semanticscholar.org/paper/Adapting-the-CRISP-DM-Data-Mining-Process%3A-A-Case-Plotnikova-Dumas/a4e44d5ce058e6b815631947db23044a22f8b114>. Acesso em: 10 Julho 2022.

RAGAB, M. G. et al. A Comprehensive Systematic Review of YOLO for Medical Object Detection (2018 to 2023). **IEEE Access**, 12, 2024. 57815-57836. Disponivel em: <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=10494845>. Acesso em: 23 Mar 2025.

RAGAB, M. G. et al. A Comprehensive Systematic Review of YOLO for Medical Object Detection (2018 to 2023). **IEEE Access**, 12, 2024. 57815-57836. Disponivel em: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10494845>. Acesso em: 17 Dez 2024.

RAHAMAN, et al. DeepCervix: A deep learning-based framework for the classification of cervical cells using hybrid deep feature fusion techniques. **Computers in Biology and Medicine**, 136, 2021. 104649. Disponivel em: <https://pubmed.ncbi.nlm.nih.gov/34332347/>. Acesso em: 25 Out 2024.

RAHAMAN, M. et al. A Survey for Cervical Cytopathology Image Analysis Using Deep Learning. **IEEE Access**, 8, 2020. 61687–61710. Disponivel em: <https://ieeexplore.ieee.org/document/9046839>. Acesso em: 12 out. 2024.

REDMON, J. et al. You Only Look Once: Unified, Real-Time Object Detection. **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**, 2016. Disponivel em: <https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/Redmon_You_Only_Look_CVPR_2016_paper.pdf>.

REN, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. **Advances in Neural Information Processing Systems (NIPS)**, 2015. 91-99. Disponivel em: <https://proceedings.neurips.cc/paper/2015/hash/14bfa6bb14875e45bba028a21ed38046-Abstract.html>. Acesso em: 28 Mar 2025.

REN, S. et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, 39, n. 6, 2017. Disponivel em:

<https://proceedings.neurips.cc/paper_files/paper/2015/file/14bfa6bb14875e45bba028a21ed38046-Paper.pdf>. Acesso em: 21 Dez 2024.

REZENDE, et al. Cric searchable image database as a public platform for conventional pap smear cytology data, 151, 2021. Disponivel em: <https://www.nature.com/articles/s41597-021-00933-8>. Acesso em: 15 Nov 2024.

RIANA, D. et al. Model of Watershed Segmentation in Deep Learning Method to Improve Identification of Cervical Cancer at Overlay Cells. **In International Conference on Neural Information Processing**, 2023, 12, n. 2. 813-819. Disponivel em: <https://www.temjournal.com/content/122/TEMJournalMay2023_813_819.pdf>. Acesso em: 26 Nov 2024.

RUMAN. Data Augmentation using Ultralytics YOLO. **YOLO Docs Ultralytics YOLO**, 2023. Disponivel em: <https://docs.ultralytics.com/guides/yolo-data-augmentation/>. Acesso em: 11 Nov 2025.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. **Nature**, 323, n. 6088, 1986. 533-536. Disponivel em: <https://blog.waqasrana.me/assets/papers/rumelhart1986.pdf>. Acesso em: 23 Nov 2024.

RUSSELL, ; NORVIG,. **Artificial Intelligence:** A Modern Approach. 4a. ed. [S.l.]: Pearson, 2020. Acesso em: 19 Mar 2025.

RUSSELL, B. C. et al. A database and web-based tool for image annotation. **International Journal of Computer Vision**, 77, n. 1-2, 2008. 157-173. Disponivel em: <https://link.springer.com/article/10.1007/s11263-007-0090-8>. Acesso em: 13 Abr 2025.

SALT, B. **Film Style and Technology:** History and Analysis. 3a. ed. London: Starword, 2009.

SAMPAIO, A. F.; ROSADO, L.; VASCONCELOS, M. J. M. Towards the mobile detection of cervical lesions: a region-based approach for the analysis of microscopic images. **IEEE Access**, 9, 8 nov. 2021. 152188-152205. Disponivel em: <https://ieeexplore.ieee.org/abstract/document/9606873>. Acesso em: 16 nov. 2024.

SANTOS, M. D. O. et al. Estimativa 2023: Incidência de Câncer no Brasil. **Instituto Nacional de Câncer (INCA)**, 2023. Disponivel em: <https://www.gov.br/inca/pt-br/assuntos/cancer/numeros/estimativa>. Acesso em: 08 Maio 2024.

SANTOS, S. D. Útero: anatomia, função, câncer de colo de útero e mais. **Biologia Net**, 2025. Disponivel em: <https://www.biologianet.com/anatomia-fisiologia-animal/utero.htm>. Acesso em: 21 Set 2025.

SCHERER, D.; MÜLLER, A.; BEHNKE, S. Evaluation of pooling operations in convolutional architectures for object recognition. **International Conference on Artificial Neural Networks (ICANN)**, Berlin, 2010. 92-101. Disponivel em: <https://ais.uni-bonn.de/papers/icann2010_maxpool.pdf>. Acesso em: 22 Jan 2025.

SCHILITHZ, A. O. C. et al. Estimativa 2016: Incidência de câncer no Brasil. **INCA - Instituto Nacional de Câncer ou Ministério da Saúde do Brasil**, 2015. Disponivel em: <https://ninho.inca.gov.br/jspui/bitstream/123456789/11691/1/Estimativa%202016%20In

cid%C3%AAncia%20de%20C%C3%A2ncer%20no%20Brasil.%202015.pdf>. Acesso em: 25 nov. 2024.

SCHRÖER, C.; KRUSE, F.; GÓMEZ, J. M. A Systematic Literature Review on Applying CRISP-DM Process Model. **Procedia Computer Science**, 181, 2021. 526-534. Disponivel em: <https://www.sciencedirect.com/science/article/pii/S1877050921002416>. Acesso em: 8 Junho 2022.

SEKAR, M. **Machine Learning for Auditors:** Automating Fraud Investigations Through Artificial Intelligence. Calgary, AB, Canada: Apress, 2022. 241 p.

SELECTIVE Search for Object Detection | R-CNN. **GeeksforGeeks**, 2021. Disponivel em: <https://www.geeksforgeeks.org/machine-learning/selective-search-for-object-detection-r-cnn/>. Acesso em: 15 Mar 2025.

SHORTEN, C.; KHOSHGOFTAAR, T. A survey on Image Data Augmentation for Deep Learning. **Journal of Big Data**, 6, n. 1, 2019. 1529-1556. Disponivel em: <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0197-0>. Acesso em: 13 Jan 2025.

SIMONYAN, K.; ZISSERMAN, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. **arXiv**, 1409.1556, 2014. Disponivel em: <https://arxiv.org/pdf/1409.1556>. Acesso em: 29 Dez 2024.

SINGH, D. et al. Global Estimates of Incidence and Mortality of Cervical Cancer in 2020: A Baseline Analysis of the WHO Global Cervical Cancer Elimination Initiative. **The Lancet Global Health**, 2, 2023. 197–206. Disponivel em: <https://www.thelancet.com/journals/langlo/article/PIIS2214-109X(22)00501-0/fulltext>. Acesso em: 10 Novembro 2024.

SOLOMON , D.; NAYAR,. **The Bethesda System for Reporting Cervical Cytology**. 2ª. ed. Rio de Janeiro: Revinter Ltda, 2024.

STABILE, S. A. B. et al. Estudo comparativo dos resultados obtidos pela citologia oncótica cérvico-vaginal convencional e pela citologia em meio líquido. **einstein (São Paulo)**, 10, n. 4, 2012. 466-472. Disponivel em: <https://www.scielo.br/j/eins/a/TP8s8FFQVc4GvMWvFk6pBGw/?lang=pt>. Acesso em: 5 Dezembro 2024.

SULAIMAN, S. N. et al. Overlapping cells separation method for cervical cell images. **INTERNATIONAL CONFERENCE ON INTELLIGENT SYSTEMS DESIGN AND APPLICATIONS**, Cairo, 10, 2010. 1218–1222. Disponivel em: <https://www.academia.edu/download/8660361/overlapping%20cells%20separation.pdf>. Acesso em: 2 Jan 2025.

SUPPORT Vector Machine (SVM). **Analytics Vidhya**, 2021. Disponivel em: <https://www.analyticsvidhya.com/blog/2021/10/support-vector-machinessvm-a-complete-guide-for-beginners/:cite[6]>. Acesso em: 20 Mar 2025.

SZEGEDY, C. et al. Going Deeper with Convolutions. **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**, 2015. 1-9. Disponivel em: <https://www.cv-

foundation.org/openaccess/content_cvpr_2015/papers/Szegedy_Going_Deeper_With_2015_CVPR_paper.pdf>. Acesso em: 26 Nov 2024.

TANAKA , S. et al. From Microscope to AI: Developing an Integrated Diagnostic System for Endometrial Cytology. **Research Square, Nature Spring Preprint Server**, 2024. Disponivel em: <https://www.researchsquare.com/article/rs-4205271/v3.pdf>. Acesso em: 19 Mar 2025.

TANG, H.-P. et al. Cervical cytology screening facilitated by an artificial intelligence microscope: a preliminary study. **Cancer cytopathology**, 129, n. 6, set. 2021. 677-678. Disponivel em: <https://acsjournals.onlinelibrary.wiley.com/doi/pdfdirect/10.1002/cncy.22425>. Acesso em: 24 nov. 2024.

TERASAKI, M. et al. From Microscope to AI: Developing an Integrated Diagnostic System for Endometrial Cytology, n. 3, jul. 2024. 2. Disponivel em: <https://www.researchsquare.com/article/rs-4205271/v3.pdf>. Acesso em: 15 jan. 2025.

TIAN, ; YE, ; DOERMANN,. Breaking the CNN Mold: YOLOv12 Brings Attention to Real-Time Object Detection. **PyImageSearch**, 2025. Disponivel em: <https://pyimagesearch.com/2025/07/07/breaking-the-cnn-mold-yolov12-brings-attention-to-real-time-object-detection/:cite[3]>. Acesso em: 25 Set 2025.

UIJLINGS, J. et al. Selective Search for Object Recognition. **International Journal of Computer Vision**, 104, n. 2, 2013. 154-171. Disponivel em: <https://staff.fnwi.uva.nl/th.gevers/pub/GeversIJCV2013.pdf>. Acesso em: 12 Out 2024.

ULTRALYTICS. YOLOv8 (Version 8.0), 2023. Disponivel em: <https://github.com/ultralytics/ultralytics>. Acesso em: 28 Fev 2025.

ULTRALYTICS. YOLOv11 architecture diagram. Ultralytics Documentation, 2024. Disponivel em: <from https://docs.ultralytics.com/models/yolo11>. Acesso em: 23 Jul 2025.

ULTRALYTICS. The Ultimate guide to data augmentation in 2025. **Ultralytics YOLO Vision**, 2025. Disponivel em: <https://www.ultralytics.com/blog/the-ultimate-guide-to-data-augmentation-in-2025#:~:text=Image%20data%20augmentation%20is%20a,have%20on%20computer%20vision%20applications.>. Acesso em: 09 Nov 2025.

VARGAS-CARDONA, H. D. et al. Artificial intelligence for cervical cancer screening: Scoping review, 2009–2022. **International Journal of Gynecology and Obstetrics**, 165, n. 2, 2024. 66-578. Disponivel em: <https://obgyn.onlinelibrary.wiley.com/doi/pdf/10.1002%2Fijgo.15179>. Acesso em: 22 Nov 2024.

WANA, T. et al. A cascaded learning approach for TBS classification of cervical cells by fusing light-weight LBCNet and handcrafted features, 5 dez. 2023. Disponivel em: <file:///C:/Users/fcalp/Downloads/ssrn-4643012%20(1).pdf>. Acesso em: 18 dez. 2024.

WAZLAWICK, R. S. **Metodologia de pesquisa para ciência da computação**. 3. ed. Rio de Jeneiro-RJ: LTC, 2021.

WEIDMAN, S. **Deep Learning from Scratch:** Building with Python from First Principles. Sebastopol: O'Reilly Media, 2019. Acesso em: 12 Mar 2025.

WHAT is YOLO? The Ultimate Guide. **Roboflow Blog**, 2025. Disponivel em: <https://blog.roboflow.com/guide-to-yolo-models/>. Acesso em: 26 Set 2025.

WILLIAM, W. et al. A pap-smear analysis tool (PAT) for detection of cervical cancer from pap-smear images. **Biomedical engineering online**, 18, n. 16, 12 fev. 2019. Disponivel em: <https://link.springer.com/content/pdf/10.1186/s12938-019-0634-5.pdf>. Acesso em: 29 nov. 2024.

ZETTL, H. **Sight, Sound, Motion:** Applied Media Aesthetics. 8a. ed. Boston: Cengage Learning, 2016.