

Augusto Almondes Teixeira
Orientador: Antonio Oseas de Carvalho Filho

**Método Automático para Detecção e
Segmentação de Bacilos de Koch Utilizando
YOLOv11 e Esquemas de Cores**

Picos - PI
17 de março de 2025

Augusto Almondes Teixeira
Orientador: Antonio Oseas de Carvalho Filho

Método Automático para Detecção e Segmentação de Bacilos de Koch Utilizando YOLOv11 e Esquemas de Cores

Modelo de Trabalho de Conclusão de Curso em Bacharelado em Sistemas de Informação na Universidade Federal do Piauí. Este modelo está em conformidade com as normas ABNT.

Universidade Federal do Piauí
Campus Senador Heuvídio Nunes de Barros
Bacharelado em Sistemas de Informação

Picos - PI
17 de março de 2025

FICHA CATALOGRÁFICA
Serviço de Processamento Técnico da Universidade Federal do Piauí
Biblioteca José Albano de Macêdo

T266m Teixeira, Augusto Almondes.

Método automático para detecção e segmentação de bacilos de koch utilizando YOLOv11 e esquemas de cores / Augusto Almondes Teixeira – 2025.

42 f.

1 Arquivo em PDF.

Indexado no catálogo *online* da biblioteca José Albano de Macêdo-CSHNB Aberto a pesquisadores, com restrições da Biblioteca.

Trabalho de Conclusão de Curso (Graduação) – Universidade Federal do Piauí, Curso de Bacharelado em Sistemas de Informação, Picos, 2025.

“Orientador: Antonio Oseas de Carvalho Filho”.

1. Sistemas informacionais – monitoramento. 2. Visão computacional. 3. Tecnologia da informação – YOLOv11. I. Teixeira, Augusto Almondes. II. Carvalho Filho, Antonio Oseas de. III. Título.

CDD 005.7

Elaborada por Maria Letícia Cristina Alcântara Gomes
Bibliotecária CRB nº 03/1835

MÉTODO AUTOMÁTICO PARA DETECÇÃO E SEGMENTAÇÃO DE BACILOS DE KOCH UTILIZANDO YOLOV11 E ESQUEMA DE CORES

AUGUSTO ALMONDES TEIXEIRA

Monografia aprovada como exigência parcial para obtenção do grau de Bacharel em Sistemas de Informação.

Data de Aprovação

Picos – PI, 25 de junho de 2025

Documento assinado digitalmente
 **ANTONIO OSEAS DE CARVALHO FILHO**
Data: 05/07/2025 09:31:30-0300
Verifique em <https://validar.it.gov.br>

Prof. Dr. Antonio Oseas de Carvalho Filho

Documento assinado digitalmente
 **PATRICK RYAN SALES DOS SANTOS**
Data: 04/07/2025 18:37:42-0300
Verifique em <https://validar.it.gov.br>

Patrick Ryan Sales dos Santo

Documento assinado digitalmente
 **VITORIA DE CARVALHO BRITO RODRIGUES**
Data: 04/07/2025 18:23:28-0300
Verifique em <https://validar.it.gov.br>

Vitória de Carvalho Brito Rodrigues

Agradecimentos

Agradeço primeiramente a Deus por ter me dado saúde, paciência e força para enfrentar os desafios ao longo dessa caminhada. Foi em meio às incertezas e ao cansaço que percebi como a fé pode nos manter firmes. Em cada noite mal dormida, em cada dúvida persistente, em cada pequeno avanço, senti Sua presença guiando meus passos. Agradecer aos meus pais, meu amor eterno e minha maior inspiração, sou grato por todo o suporte e carinho. Vocês foram meu porto seguro quando o cansaço e o desânimo ameaçaram me parar. Obrigado por me ensinarem o valor do esforço, da honestidade e da resiliência. Cada conquista minha é, também, de vocês.

Sou grato aos meus professores, em especial ao professor Antonio Oseas de Carvalho Filho, por sua orientação dedicada, paciência e incentivo. Obrigado por acreditar no potencial desse trabalho e me impulsionar a dar sempre o meu melhor. Seus ensinamentos vão além do conteúdo acadêmico, são valores e exemplos que levarei para a vida.

Aos amigos que conheci ao longo dessa jornada e que dividiram comigo não apenas salas de aula, mas conhecimento. Obrigado pela parceria, pelas conversas, pelas trocas de conhecimentos, motivações e por tornarem essa jornada mais leve e significativa. Agradeço a todos que participaram, de forma direta ou indireta, da minha conquista.

Por fim, agradeço a mim mesmo. Por não desistir quando teria sido mais fácil, por acreditar mesmo quando parecia difícil, por seguir em frente mesmo sem certezas. Chegar até aqui é, antes de tudo, um ato de resistência, coragem e crescimento.

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

A Tuberculose persiste como um grave problema de saúde pública em todo o mundo, cujos métodos mais comuns de diagnósticos apresenta limitações de tempo e precisão. Neste contexto, abordagens de Visão Computacional e Aprendizado Profundo surgem como alternativas promissoras para automatizar e otimizar a detecção do *Mycobacterium tuberculosis*, bactéria causadora da doença. O objetivo geral deste trabalho foi propor e avaliar um método computacional para a detecção e segmentação automatizada de bacilos da Tuberculose em imagens de baciloscopia, utilizando a arquitetura YOLOv11 e investigando o impacto de diferentes estratégias de pré-processamento de cores. Para isso, foi utilizado um banco de dados público composto por imagens de baciloscopia, organizado em quatro conjuntos distintos: (a) RGB original, (b) convertido para o espaço de cores LUV, (c) RGB com conversão em escala de cinza seguida por técnica de contraste limitada à equalização do histograma adaptável (CLAHE), e (d) combinação do espaço de cores LUV e escala de cinza com CLAHE. O desempenho foi medido utilizando as métricas de precisão, recall e mAP50. Os resultados indicam que o modelo treinado com o conjunto de dados em LUV obteve o melhor desempenho, alcançando, na tarefa de detecção, 97,60% de precisão, 95,60% de recall e 98,30% de mAP. Na segmentação, a mesma abordagem atingiu 84,60% de precisão, 82,80% de recall e 84,80% de mAP, superando os demais modelos. Conclui-se que a arquitetura YOLOv11 é eficaz para a tarefa e que a escolha do pré-processamento é um fator decisivo, com o espaço de cores LUV se mostrando superior ao dissociar a luminância da crominância e minimizar interferências de iluminação.

Palavras-chave: Tuberculose; Baciloscopia de escarro; YOLOv11; Visão Computacional.

Abstract

Tuberculosis remains a serious global public health problem, and its most common diagnostic methods have limitations in speed and accuracy. In this context, Computer Vision and Deep Learning approaches emerge as promising alternatives to automate and optimize the detection of *Mycobacterium tuberculosis*, the bacterium that causes the disease. The main objective of this work was to propose and evaluate a computational method for the automated detection and segmentation of Tuberculosis bacilli in smear microscopy images, using the YOLOv11 architecture and investigating the impact of different color preprocessing strategies. For this, a public dataset of microscopy images was used, organized into four distinct sets: (a) original RGB, (b) converted to the LUV color space, (c) RGB converted to grayscale followed by Contrast Limited Adaptive Histogram Equalization (CLAHE), and (d) a combination of the LUV color space and grayscale with CLAHE. Performance was measured using precision, recall, and mAP50 metrics. The results indicate that the model trained with the LUV dataset achieved the best performance, reaching, in the detection task, 97.60% precision, 95.60% recall, and 98.30% mAP. In the segmentation task, the same approach achieved 84.60% precision, 82.80% recall, and 84.80% mAP, outperforming the other models. It is concluded that the YOLOv11 architecture is effective for the task and that the choice of preprocessing is a decisive factor, with the LUV color space proving superior by decoupling luminance from chrominance and minimizing illumination interferences.

Lista de ilustrações

Figura 1 – Fotomicrografia de esfregaço de escarro corado pelo método de Ziehl-Neelsen, evidenciando BAAR (regiões vermelhas na imagem).	16
Figura 2 – Arquitetura do modelo YOLO composta por 24 camadas convolucionais, que atuam como extratoras de características, seguidas por 2 camadas totalmente conectadas, que realizam as predições finais. Fonte: (REDMON et al., 2016).	20
Figura 3 – Representação esquemática do fluxo de trabalho para desenvolvimento e avaliação do modelo de detecção de MT. As etapas incluem a preparação dos dados, o pré-processamento por canais de cor, o treinamento e a avaliação de resultados. Fonte: (Autor).	25
Figura 4 – Exemplo visual de uma imagem da base de dados com sua respectiva máscara. O modelo da esquerda (a) é uma imagem em mosaico com a presença de MT em algumas sub-imagens. No segundo modelo (b), é representado a máscara, com regiões em branco representando a localização dos MT na primeira imagem. Fonte: (Autor).	26
Figura 5 – Efeito de diferentes técnicas de pré-processamento sobre uma imagem amostra do conjunto de dados: (a) Imagem original no espaço de cor RGB; (b) Imagem convertida para o espaço de cor LUV; (c) Imagem em escala de cinza obtida após aplicação da técnica CLAHE à imagem RGB; (d) Imagem em escala de cinza obtida a partir da conversão da imagem LUV e posterior aplicação do CLAHE. Fonte: (Autor).	28
Figura 6 – Comparação visual das detecções e segmentações nas diferentes bases de dados	32

Lista de tabelas

Tabela 1 – Resumo dos principais trabalhos relacionados.	23
Tabela 2 – Hiperparâmetros utilizados no treinamento do modelo.	29
Tabela 3 – Métricas das predições de caixas delimitadoras (<i>bouding boxes</i>)	30
Tabela 4 – Métricas das segmentações realizadas pelos modelos	31

Lista de abreviaturas e siglas

TB	Tuberculose
MT	<i>Mycobacterium tuberculosis</i>
OMS	Organização Mundial da Saúde
MS	Ministério da Saúde
BE	Baciloscopia de escarro
VC	Visão Computacional
AP	Aprendizado Profundo
BAAR	Bacilos Álcool-Ácido-Resistentes
AM	Aprendizado de Máquina
IA	Inteligência Artificial
RNC	Redes Neurais Convolucionais
YOLO	<i>You Only Look Once</i>

Sumário

1	Introdução	12
1.1	Objetivos	13
1.1.1	Objetivo Geral	13
1.1.2	Objetivos Específicos	13
1.2	Estrutura do Trabalho	14
2	Referencial Teórico	15
2.1	Tuberculose	15
2.2	Baciloscopia do escarro	16
2.3	Visão Computacional	17
2.4	Aprendizado de Máquina	17
2.4.1	Aprendizado Profundo	18
2.4.1.1	Rede Neural Convolutacional	18
2.5	Arquitetura YOLO	19
2.5.1	YOLOv11	19
2.6	Métricas de Avaliação	20
3	Trabalhos Relacionados	22
4	Matérias e Métodos	25
4.1	Base de Dados	26
4.2	Pré-processamento	26
4.3	Treinamento com modelo YOLOv11	27
5	Resultados e Discussão	30
5.1	Avaliação de Desempenho da Detecção e Segmentação	30
5.2	Estudos de Caso	31
6	Produção Científica e Técnica	34
6.1	Artigos Científicos Submetidos	34
6.2	Registros de Software Produzidos	34
7	Conclusão	36
	Referências	38

1 Introdução

A Tuberculose (TB) configura-se como uma doença de caráter infeccioso, cujo agente etiológico é o *Mycobacterium tuberculosis* (MT). Esta condição acomete, de forma preponderante, os pulmões, embora possa também atingir outras estruturas orgânicas, tais como os rins, o sistema esquelético, o sistema nervoso e os linfonodos. A propagação da doença dá-se, majoritariamente, pela via aérea, quando indivíduos portadores da infecção expõem, ao tossir ou espirrar, minúsculas gotículas contendo o microrganismo patogênico. A antiguidade da Tuberculose é atestada por registros históricos, que revelam indícios da infecção em múmias do Egito Antigo, bem como documentação de seu estudo na Grécia Clássica (MARTINS; MIRANDA, 2020). A TB voltou a figurar como a enfermidade infecciosa de maior mortalidade mundial, ultrapassando a COVID-19, que ocupou tal posto por um triênio (2020, 2021, 2022). Destarte, a Tuberculose persiste como um relevante problema de saúde pública em escala global (WHO, 2024).

Conforme o relatório global emitido pela Organização Mundial da Saúde (OMS) referente ao ano de 2023, a Tuberculose foi responsável por 1,25 milhão de fatalidades, sendo que 161 mil desses óbitos ocorreram em pessoas que conviviam com o HIV (WHO, 2024). Em território brasileiro, os indicadores da doença permanecem em patamares alarmantes. Levantamentos do Ministério da Saúde (MS) revelam a notificação anual de aproximadamente 80 mil novos diagnósticos da infecção e um contingente de cerca de 5,5 mil mortes atribuídas à TB (BRASIL, 2023). A cada ano, observa-se um agravamento no impacto dessa enfermidade, impulsionado pelo aumento no número de mortes. Ressalta-se ainda que a maior concentração de casos é verificada em nações com renda nacional bruta considerada média ou baixa (WHO, 2024).

Não obstante os progressos na tecnologia terem fomentado o desenvolvimento de novas abordagens para a detecção e o tratamento da Tuberculose, a aplicação generalizada de muitos desses métodos permanece restrita devido aos elevados custos de produção e implementação (GIOSEFFI; BATISTA; BRIGNOL, 2022). Entre as ferramentas diagnósticas disponíveis, encontram-se o exame microscópico, a radiografia do tórax, o teste tuberculínico, a cultura do MT e o ensaio de liberação de interferon-gama. Entretanto, a baciloscopia de escarro (BE), que se fundamenta na análise microscópica de amostras de expectoração dos pacientes, continua sendo a técnica mais difundida.

A identificação e a quantificação do MT são conduzidas por técnicos qualificados, responsáveis pelo exame de lâminas em microscópio (GOMIDE et al., 2023). Embora este constitua o procedimento padrão, sua efetividade pode ser otimizada, uma vez que depende de uma avaliação manual. Adicionalmente, configura-se como um exame que requer tempo, meticulosidade e um profissional com vasta experiência, sobretudo nas etapas iniciais da patologia, quando o diagnóstico se torna mais intrincado (RODRIGUES

et al., 2022).

Em vista do cenário apresentado e alavancados pelo progresso tecnológico, têm sido propostas diversas metodologias inovadoras para colaborar no enfrentamento da TB e na otimização de sua terapêutica, como é o caso da implementação da Visão Computacional (VC) baseada em Aprendizado Profundo (AP), em inglês *Deep Learning*. A utilização dessa abordagem tecnológica no âmbito médico tem se intensificado progressivamente ao longo dos últimos anos. No que tange ao diagnóstico da TB, tais métodos podem ser aplicados para automatizar e refinar a detecção do MT em amostras de expectoração, conferindo ao processo maior acurácia e agilidade. Adicionalmente, estratégias que envolvem a manipulação do espaço de cores em imagens clínicas têm demonstrado impacto favorável no desempenho de redes neurais. Essas abordagens concorrem de maneira expressiva para a identificação antecipada da patologia, viabilizando, assim, um manejo terapêutico mais eficiente.

O presente trabalho emerge da necessidade de aprimorar os métodos de diagnóstico da TB, superando as limitações das técnicas tradicionais, como a BE manual. Neste contexto, destaca-se como principal diferencial desse trabalho a aplicação da mais recente arquitetura YOLOv11 para a detecção e segmentação do MT, conjugada a uma investigação minuciosa dos efeitos de distintas técnicas de pré-processamento de imagem. Tal enfoque metodológico viabiliza uma avaliação mais acurada da efetividade desses procedimentos na automatização do diagnóstico da Tuberculose.

1.1 Objetivos

Alinhado à problemática da otimização do diagnóstico da TB, este trabalho tem como propósito principal investigar e desenvolver uma solução computacional para a detecção automatizada do MT. Para tal, serão empregadas tecnologias de VC e AP aplicadas a imagens de BE.

1.1.1 Objetivo Geral

O objetivo geral deste estudo é propor e avaliar um método computacional inovador, fundamentado na arquitetura de AP YOLOv11 e em técnicas de pré-processamento de imagem, para a detecção e segmentação automatizada do MT em imagens de BE, visando contribuir para um diagnóstico mais preciso, eficiente e ágil da TB.

1.1.2 Objetivos Específicos

Para alcançar o objetivo geral, foram estabelecidos os seguintes objetivos específicos:

- Avaliar a eficácia da arquitetura YOLOv11 para a detecção e segmentação de MT em imagens de BE.

- Investigar e analisar comparativamente o impacto de diferentes alterações no espaço de cores na *performance* do modelo de detecção e segmentação.
- Validar a eficácia da melhor abordagem de pré-processamento encontrada para a tarefa de segmentação precisa dos contornos dos bacilos.

1.2 Estrutura do Trabalho

Este trabalho está organizado em 7 capítulos. Além desta Introdução, o Capítulo 2 apresenta os fundamentos teóricos que embasam a pesquisa; o Capítulo 3 discute os principais trabalhos relacionados; o Capítulo 4 descreve a metodologia adotada; o Capítulo 5 traz a análise dos resultados obtidos; o Capítulo 6 detalha as produções científicas e técnicas derivadas deste estudo; e, por fim, o Capítulo 7 apresenta as conclusões e sugestões para pesquisas futuras.

2 Referencial Teórico

Este capítulo visa estabelecer uma base conceitual fundamental para a compreensão deste projeto. Serão abordados conceitos-chave que sustentam a estrutura da pesquisa, como a TB, o Aprendizado de Máquina e a VC.

2.1 Tuberculose

A palavra “tuberculose” tem origem no latim *tuberculum*, derivado de *tuber* (tumor, em português), em referência às lesões características que a doença provoca nos pulmões (GIOSEFFI; BATISTA; BRIGNOL, 2022). Trata-se de uma enfermidade milenar, com registros históricos desde a Grécia e Roma antigas, além de evidências em múmias do Egito antigo, o que comprova sua longa existência ao longo da história da humanidade (CUNHA et al., 2023).

A TB é uma doença infectocontagiosa e endêmica que acomete, predominantemente, os pulmões. Seus sintomas mais comuns incluem tosse seca persistente e dores na região torácica, podendo levar ao óbito, caso não seja tratada adequadamente. É causada pelo MT, também conhecido como Bacilo de Koch (GIOSEFFI; BATISTA; BRIGNOL, 2022). A transmissão ocorre principalmente por via aérea, através da inalação de aerossóis contaminados expelidos por pessoas infectadas. Essas bactérias apresentam crescimento lento e possuem afinidade especial pelo tecido pulmonar (MARTINS; MIRANDA, 2020).

A propagação da TB está fortemente relacionada às condições socioeconômicas da população. Como outras doenças infecciosas, tende a se disseminar em regiões com alta densidade populacional e infraestrutura urbana precária (SOEIRO; CALDAS; FERREIRA, 2022). Por essa razão, sua incidência é geralmente maior nas periferias dos grandes centros urbanos, embora possa acometer indivíduos em qualquer contexto geográfico, inclusive em áreas rurais. A partir de 2015, observou-se um aumento contínuo no número de casos, possivelmente associado à desaceleração econômica e à redução de investimentos em políticas sociais, o que impactou negativamente o bem-estar da população (SILVA; GALVÃO, 2024).

Conforme a OMS, mais de 10 milhões de pessoas adoecem de TB a cada ano. Em 2023, foram estimados aproximadamente 10,8 milhões de casos incidentes, com uma taxa de 134 casos por 100.000 habitantes. No mesmo período, a doença foi responsável por cerca de 1,25 milhão de mortes em todo o mundo (WHO, 2024). Caso as tendências atuais se mantenham sem a implementação de medidas eficazes, as projeções indicam que, até 2030, o controle da tuberculose poderá regredir aos patamares críticos observados no início dos anos 2000 (SILVA; GALVÃO, 2024).

2.2 Baciloscopia do escarro

A BE é um exame realizado em laboratório capaz de detectar e quantificar MT em amostras de escarro. Esse é o exame mais utilizado para detectar as bactérias causadoras da TB, além de ser comumente realizado para diagnósticos e no tratamento da doença. Quando executada de forma correta em todas as etapas, essa técnica permite detectar grande parte dos casos.

A metodologia mais utilizada na BE é a de Ziehl-Neelsen, uma técnica de coloração usada em parasitologia e microbiologia para a detecção e diferenciação de vários patógenos. No processo de coloração, o esfregaço de escarro é fixado em uma lâmina e adicionado em seguida um corante vermelho (fucsina), absorvido pelo bacilo. Ao aplicar o descolorante na lâmina, somente os bacilos álcool-ácido-resistentes (BAAR) continuaram com o tom avermelhado, o que facilita a identificação e contagem dos mesmos (BRASIL, 2020).

A Figura 1 apresenta uma fotomicrografia de um esfregaço de escarro (obtida de uma fonte privada) capturada em lâmina corada pelo método de Ziehl-Neelsen, evidenciando a presença de BAAR. A análise detalhada da imagem permite inferir a complexidade inerente ao processo de identificação e quantificação manual do MT.

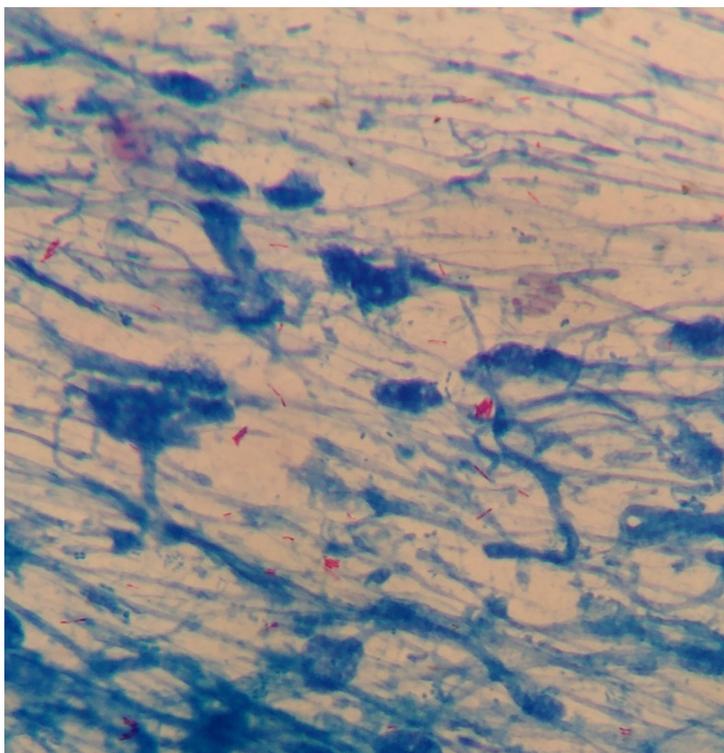


Figura 1 – Fotomicrografia de esfregaço de escarro corado pelo método de Ziehl-Neelsen, evidenciando BAAR (regiões vermelhas na imagem).

2.3 Visão Computacional

A VC é uma área da IA que capacita sistemas computacionais a extrair, processar e interpretar informações significativas a partir de imagens e vídeos (SANTOS et al., 2020). O crescimento desse método deve-se primariamente à superação de barreiras históricas: a dependência da intervenção humana, que tornava os processos lentos e suscetíveis a erros, e as limitações de hardware, que restringiam a capacidade e a velocidade do treinamento de modelos complexos (CONSTÂNCIO; CARVALHO; TSUNODA, 2022).

As aplicações da tecnologia são vastas e impactam múltiplos setores. Em tarefas de processamento gráfico, por exemplo, é utilizada para reconhecimento de objetos, segmentação de imagens e reconstrução de ambientes tridimensionais a partir de dados visuais do mundo real (ZHOU, 2023). No setor da saúde, a VC desempenha um papel fundamental na análise de exames de imagem, auxiliando na detecção de anomalias e contribuindo para diagnósticos médicos mais rápidos e precisos (LINGINANI; MUDDANA, 2023).

O processo de VC é organizado em uma sequência de etapas. Inicia-se com a aquisição da imagem, na qual uma imagem bidimensional, tridimensional ou uma sequência de imagens é obtida por sensores. Em seguida, a imagem passa pelo pré-processamento, uma fase de tratamento para facilitar a identificação de objetos, como o destaque de contornos e formas. Depois, ocorre a extração de características matemáticas da imagem, como textura e formato, e a detecção e segmentação para destacar e isolar as regiões de interesse. Por fim, o processamento de alto nível classifica os objetos em categorias e valida os dados, convertendo a informação visual em conhecimento prático (LEOCÁDIO et al., 2021).

2.4 Aprendizado de Máquina

O AM, uma subcategoria essencial da inteligência artificial (IA), permite que sistemas computacionais aprendam a partir de dados sem a necessidade de programação explícita (LUDERMIR, 2021). Seus algoritmos aprimoram seu desempenho progressivamente à medida que são expostos a mais exemplos de treinamento, adaptando-se e extraíndo conhecimento diretamente dos dados (IZBICKI; SANTOS, 2020). Essa área compreende diversas abordagens, como o aprendizado supervisionado, não supervisionado e por reforço.

No âmbito do AM, o aprendizado não supervisionado caracteriza-se pela utilização de algoritmos treinados com dados que não possuem rótulos predefinidos. Nesses cenários, os algoritmos operam identificando e agrupando dados a partir das similaridades detectadas em seus atributos, estabelecendo conexões que o próprio algoritmo determina serem significativas. Uma vez formados esses agrupamentos, torna-se essencial uma análise subsequente para desvendar o sentido e a relevância de cada conjunto dentro do domínio do

problema em estudo (LUDERMIR, 2021).

Para o aprendizado supervisionado, o processo de treinamento depende da disponibilização de uma resposta esperada para cada exemplo fornecido ao algoritmo. Cada exemplo é estruturado com base em um conjunto de atributos e sua classe correspondente (ALVES, 2024). A meta do algoritmo é aprender essa relação para, então, ser capaz de identificar com precisão a classe de exemplos futuros que ainda não possuem rótulo. É importante distinguir que, para rótulos de classe discretos, o desafio é conhecido como classificação, enquanto para prever valores contínuos, o problema é denominado regressão (LUDERMIR, 2021).

Já as técnicas de aprendizado por reforço são primariamente voltadas para a solução de problemas de tomada de decisão sequencial. Diferentemente de outras abordagens, elas não se concentram no aprendizado de representações complexas dos dados ou na resolução de problemas com alta dimensionalidade (KINOSHITA; MENEGHETTI; BIANCHI, 2022). Essa é uma técnica de AM que capacita um agente a aprender a tomar decisões em um ambiente incerto. Esse aprendizado ocorre por meio do recebimento de reforços, positivos ou negativos, em resposta a cada ação executada ao longo do tempo (SALES et al., 2023).

2.4.1 Aprendizado Profundo

O AP é empregado para discernir e modelar relações complexas entre dados de entrada e as previsões de saída resultantes. Inspirando-se na organização neural do cérebro humano, esta metodologia consiste em redes de nós interconectados (neurônios artificiais) que processam informações em camadas sucessivas (MARQUES et al., 2022). Diferentes arquiteturas de AP foram desenvolvidas para otimizar o tratamento de tipos de dados específicos. Apesar dos desafios inerentes, como a necessidade de grandes conjuntos de dados e consideráveis recursos computacionais, o potencial transformador do AP na tecnologia e na resolução de problemas complexos continua expressivo (HODONOU, 2024).

2.4.1.1 Rede Neural Convolutacional

As Redes Neurais Convolutacionais (RNC), uma classe de algoritmos fundamentada em AP, são predominantemente empregadas no processamento e análise de imagens. Diferentemente de abordagens convencionais onde a definição de filtros para extração de características era um processo manual, as RNCs aprendem esses filtros automaticamente durante o treinamento (PEREIRA et al., 2025). Sua arquitetura é caracterizada por múltiplas camadas treináveis, organizadas hierarquicamente. Cada camada é projetada para extrair características progressivamente mais complexas dos dados de entrada, aproveitando as informações processadas pelas camadas anteriores. Essa estrutura permite que as RNCs preservem eficazmente as relações espaciais locais dos pixels, aspecto crucial para a interpretação de dados visuais (OLIVEIRA, 2020).

A arquitetura de uma RNC compreende, essencialmente, três tipos de camadas que desempenham papéis distintos: a camada convolucional, a camada de agrupamento (comumente referida como *pooling*), e a camada totalmente conectada (*fully connected layer*) (PEREIRA et al., 2025).

O processamento em uma RNC geralmente se inicia com blocos de camadas convolucionais, que aplicam um conjunto de filtros aprendíveis à imagem de entrada para gerar mapas de características, detectando padrões locais. Entre as camadas convolucionais, ou após um bloco delas, são inseridas camadas de agrupamento. Estas têm a função de reduzir a resolução espacial dos mapas de características, o que contribui para a diminuição do número de parâmetros e para a criação de representações mais robustas a pequenas variações na entrada. Por fim, as características de alto nível extraídas são geralmente achatadas (*flattened*) e processadas por uma ou mais camadas totalmente conectadas, que integram as informações para realizar a tarefa de inferência final, como a classificação. (BARROS et al., 2021).

2.5 Arquitetura YOLO

A arquitetura YOLO (*You Only Look Once*) representa uma nova abordagem para a detecção de objetos, ao enquadrar a tarefa como um problema de regressão unificado. Diferentemente de métodos tradicionais baseados em classificadores, que exigem pipelines complexos e múltiplos estágios, o YOLO processa imagens inteiras em uma única avaliação por meio de uma única rede neural Redmon et al. (2016). O sistema divide a imagem de entrada em uma grade de $S \times S$ células. Caso o centro de um objeto recaia sobre uma célula da grade, esta se torna responsável por sua detecção. Cada célula prevê B caixas delimitadoras (*bounding boxes*) e suas respectivas pontuação de confiança. Essas pontuações refletem tanto a probabilidade de um objeto estar contido na caixa quanto a precisão da predição da própria caixa (GOMES, 2022).

A Figura 2 ilustra a arquitetura da RNC do YOLO, composta por 24 camadas convolucionais e 2 camadas totalmente conectadas que processam uma imagem de entrada. As camadas convolucionais extraem características da imagem de forma progressiva. Após a extração de características, as camadas totalmente conectadas geram o tensor de predição final, que unifica as informações de caixas delimitadoras, pontuação de confiança e probabilidades de classe para a imagem inteira Redmon et al. (2016).

2.5.1 YOLOv11

Desde sua concepção original, a arquitetura YOLO passou por uma notável evolução. Com o YOLOv3, Redmon e Farhadi (2018) aprimoraram a detecção em múltiplas escalas, estabelecendo um padrão na indústria por anos. Posteriormente, os trabalhos de Bochkovskiy, Wang e Liao (2020) e Khanam e Hussain (2024) focaram na otimização do

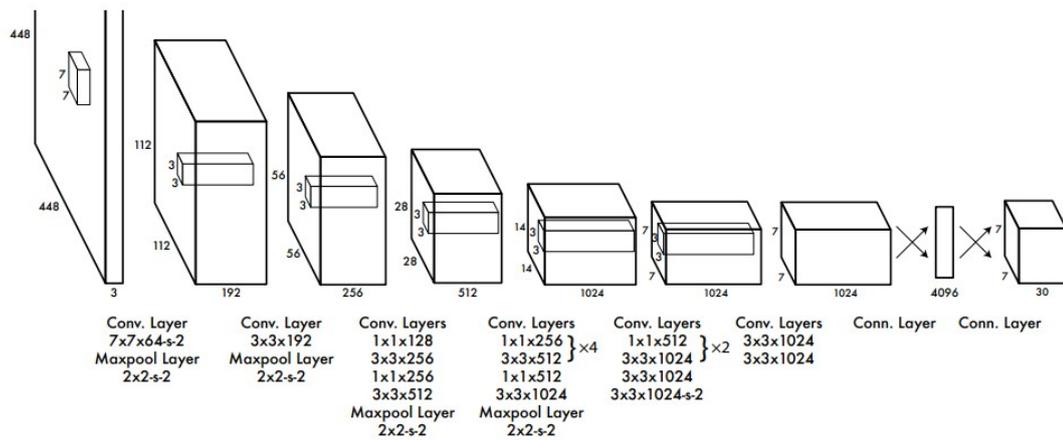


Figura 2 – Arquitetura do modelo YOLO composta por 24 camadas convolucionais, que atuam como extratoras de características, seguidas por 2 camadas totalmente conectadas, que realizam as previsões finais. Fonte: (REDMON et al., 2016).

treinamento e da inferência, introduzindo um conjunto de heurísticas e melhorias com o YOLOv4 e o YOLOv5, conhecidas como '*Bag of Freebies*' e '*Bag of Specials*'. Mais recentemente, uma mudança significativa ocorreu com a transição para abordagens *anchor-free*, que simplificam a cabeça de detecção e melhoram o desempenho, uma tendência consolidada em versões como o YOLOv8 (YASEEN, 2024).

Conforme descrito por Huang et al. (2024), a arquitetura do YOLOv11 introduz inovações em seus três componentes essenciais: o *backbone*, o *neck* e a *head*. O *backbone* foi otimizado com a substituição do módulo C2F pelo C3K2, que aprimora a extração de características da imagem sem elevar a carga computacional. No *neck*, a inclusão do módulo C2PSA atua como um mecanismo de atenção, permitindo que o modelo foque nas características mais pertinentes da imagem. Por fim, a *head* (cabeça de detecção) foi redesenhada com uma abordagem desacoplada e *anchor-free*, onde a regressão de caixas delimitadoras é feita com convoluções padrão e a classificação de objetos emprega convoluções de maior eficiência, o que resulta em uma redução de custos computacionais e um ganho de precisão.

2.6 Métricas de Avaliação

A avaliação do desempenho é um pilar central em projetos de VC, onde métricas objetivas traduzem o desempenho em indicadores numéricos, permitindo uma análise criteriosa da eficácia do modelo. Para este trabalho, foram selecionadas três métricas essenciais, calculadas automaticamente pelo modelo YOLO durante o treinamento: precisão, recall e *Mean Average Precision* (mAP, especificamente o mAP50). Adicionalmente, essas métricas podem ser obtidas utilizando outras ferramentas do próprio framework YOLO.

As métricas escolhidas avaliam o desempenho do modelo sob diferentes óticas. A Precisão foca nos acertos do modelo, medindo a proporção de detecções corretas dentre todas

as detecções realizadas. O Recall, em contrapartida, avalia a capacidade do modelo de não deixar passar objetos relevantes, medindo quantos dos objetos que realmente estavam presentes foram encontrados (SOUZA, 2020). Como um modelo pode ter alta precisão, mas baixo recall (ou vice-versa), a métrica mAP50 é utilizada como o principal indicador de desempenho. Ela consolida a relação entre precisão e recall para todas as classes de objetos em um único número, fornecendo uma avaliação global e mais completa da capacidade do modelo de detectar e classificar corretamente.

3 Trabalhos Relacionados

A busca por automatizar a análise de MT tem impulsionado o uso do AM e do processamento de imagens na área da saúde. Neste capítulo, serão discutidos os estudos mais relevantes com abordagens similares a desse trabalho, explorando suas contribuições e limitações.

Uma contribuição notável no campo foi a de [Reis \(2022\)](#), que explora uma arquitetura híbrida para a análise de imagens de BE. A solução emprega o MLP-Mixer como classificador em conjunto com o detector YOLOv7 ([WANG; BOCHKOVSKIY; LIAO, 2022](#)), uma combinação que busca otimizar tanto a identificação quanto a localização dos MT. O desempenho do método foi validado quantitativamente, alcançando um F1-score de 76,46%, que reflete o balanço entre uma precisão de 81,36% e um recall de 71,36%. Esses números demonstram que o sistema, quando identifica um MT, tende a estar correto, mas ainda existindo espaço para melhorias em algumas métricas, como o recall, que se refere a capacidade de encontrar todos os MT presentes na amostra. Este ajuste fino entre precisão e recall é um desafio comum na área, e os resultados do estudo servem como um importante ponto de partida, mostrando a viabilidade da abordagem e, ao mesmo tempo, sinalizando caminhos para trabalhos futuros.

[Santos, Brito e Filho \(2023\)](#), em seu estudo, propuseram uma análise sistemática para a detecção do MT. Um diferencial do trabalho foi a investigação aprofundada de como diferentes componentes afetam o resultado, testando não apenas os diversos *backbones* que compõem a arquitetura, mas também o impacto de quatro espaços de cores (RGB, HSV, LUV e XYZ). Utilizando um conjunto de dados público e uma metodologia de validação cruzada com $k\text{-fold}=5$, o estudo reportou um F1-score de 0,774. Um aspecto particularmente notável foi o alto recall de 0,925, sugerindo que o modelo é bastante sensível e eficaz em localizar a maioria dos MT presentes. Em contrapartida, a precisão de 0,694 indica um desafio apontado pelos autores, onde a semelhança entre MT e outros artefatos nas lâminas pode levar a detecções falso-positivas.

O trabalho de [López \(2018\)](#) apresentou uma metodologia para a detecção do MT que se divide em duas etapas principais, sendo a primeira a classificação de fragmentos de imagem e, posteriormente, a detecção em imagens completas. Uma contribuição particular do estudo foi o uso de imagens de foco estendido, a partir das quais o autor construiu uma base de dados própria para treinar seus modelos de RNC. Na primeira etapa, de classificação de patches, a abordagem demonstrou grande sucesso, com os modelos de duas e três camadas convolutivas atingindo uma acurácia de 99% nos formatos de imagem R-G e RGB. Na etapa seguinte, que aplica o classificador treinado sobre a imagem inteira por uma janela deslizante e do algoritmo *Non-maximum Suppression*, os resultados refletem os desafios de uma aplicação em um cenário real. Foi obtido um F1-score de 68,47%, um

recall alto de 86,15% e uma precisão mais contida de 56,82%.

Um estudo notável foi o de [Carvalho \(2023\)](#), que realizou uma análise abrangente de técnicas de AP aplicadas à detecção de MT em imagens de BE. Os autores avaliaram diversas arquiteturas de RNC e métodos de detecção de objetos, testando-os em múltiplos conjuntos de dados, incluindo uma nova base de dados pública criada durante a pesquisa. Dentre as abordagens testadas, a arquitetura Faster R-CNN alcançou o melhor desempenho geral, obtendo um F1-Score de até 91,2%, com uma sensibilidade de 98,4% e uma precisão de 85,1%. Este resultado foi particularmente proeminente quando a arquitetura foi combinada com uma RNC adicional para a redução de falsos positivos. Adicionalmente, estratégias como aumento de dados (*data augmentation*), particionamento de imagens e um sistema de votação para anotação foram implementadas para otimizar o desempenho do modelo.

O estudo de [Xiong et al. \(2018\)](#) propôs um sistema de IA (TB-AI) baseado em uma RNC para a detecção de MT. Após um processo iterativo de refinamento do método, o sistema final alcançou uma alta sensibilidade de 97,94% e uma especificidade de 83,65%. Este balanço nos resultados evidencia o potencial da ferramenta para reduzir a omissão de diagnósticos (falsos negativos), mas, ao mesmo tempo, revela sua dificuldade em diferenciar bacilos patogênicos de artefatos ou contaminantes, reforçando a necessidade da validação final por um patologista.

A Tabela 1 apresenta um resumo das métricas dos trabalhos expostos anteriormente.

Tabela 1 – Resumo dos principais trabalhos relacionados.

Referência	Método	Dados Utilizados	Métricas Principais
(REIS, 2022)	MLP-Mixer para classificação e YOLOv7 para detecção	Imagens de BE	mAP: 74,08%, Precisão: 81,36%, Recall: 71,36%, F1-Score: 76,46%
(SANTOS; BRITO; FILHO, 2023)	EfficientDet com múltiplos <i>backbones</i>	Imagens de BE (espaços de cor diversos)	IoU: 0,523, Recall: 0,925, Precisão: 0,694, F1-Score: 0,774
(LÓPEZ, 2018)	CNNs para detecção automática	Imagens de BE (foco estendido)	Precisão: 56,82%, Recall: 86,15%, F1-Score: 68,47% (imagens completas)
(CARVALHO, 2023)	Faster R-CNN com RNC para redução de falsos positivos	Imagens de BE (Dataset público próprio)	Precisão: 85,1%, Sensibilidade: 98,4%, F1-Score: 91,2%
(XIONG et al., 2018)	CNN para detecção de MT (TB-AI)	Lâminas com coloração Ziehl-Neelsen	Sensibilidade: 97,94%, Especificidade: 83,65%

A análise dos trabalhos apresentados revela um cenário de contínua evolução, com um foco predominante na detecção do MT. Abordagens como Faster R-CNN e as diversas versões da arquitetura YOLO mostram-se promissoras para a localização de MT por meio de *bounding boxes*. Contudo, um exame mais detalhado das métricas reportadas indica desafios persistentes. Estudos que alcançam alta sensibilidade frequentemente o fazem em

detrimento de uma precisão mais baixa, sugerindo que a diferenciação fina entre os MT e os artefatos da lâmina ainda é complexa.

Neste contexto, a segmentação de instância, que consiste na delimitação precisa do contorno de cada MT, emerge não como uma lacuna não explorada, mas como um passo adiante para refinar a análise. Ao gerar uma máscara exata para cada MT, é possível não apenas localizar, mas também extrair características morfológicas detalhadas, o que pode levar a uma identificação mais robusta e à redução de falsos positivos.

Diante disso, o presente trabalho busca avançar em relação às metodologias anteriores, propondo um pipeline que visa não apenas detectar, mas segmentar os MT com maior precisão. Para tal, esta pesquisa se diferencia em dois eixos principais, sendo a arquitetura de detecção e o pré-processamento das imagens.

Em primeira instância, adota-se uma versão mais recente da arquitetura YOLO, explorando seus avanços para obter uma delimitação mais acurada dos patógenos. Em segundo lugar, o tratamento das imagens recebe atenção especial, com uma investigação sistemática sobre o pré-processamento. Esta pesquisa avalia tanto o impacto da conversão para o espaço de cores LUV, em contraposição ao tradicional RGB, quanto o efeito de um pipeline de melhoria de contraste. Este pipeline consiste na conversão da imagem para escala de cinza e na subsequente aplicação da equalização de histograma adaptativa por contraste limitado (CLAHE). Conforme discutido por diversos autores, a variabilidade na coloração e iluminação das lâminas é um desafio recorrente (XIONG et al., 2018; LÓPEZ, 2018), e as estratégias aqui propostas visam justamente aprimorar o contraste e reduzir tais interferências, favorecendo um desempenho superior do modelo de segmentação.

4 Matérias e Métodos

Neste capítulo, são apresentadas todas as etapas realizadas para o desenvolvimento de um modelo de detecção e segmentação de MT, além da análise do canal de cor mais adequado para o treinamento. As etapas metodológicas incluem: (i) aquisição e organização da base de dados; (ii) pré-processamento das imagens em diferentes espaços de cor, gerando novas bases de dados; (iii) treinamento do modelo com os novos *datasets*; e (iv) avaliação dos resultados obtidos. A Figura 3 apresenta um resumo visual dessas etapas.

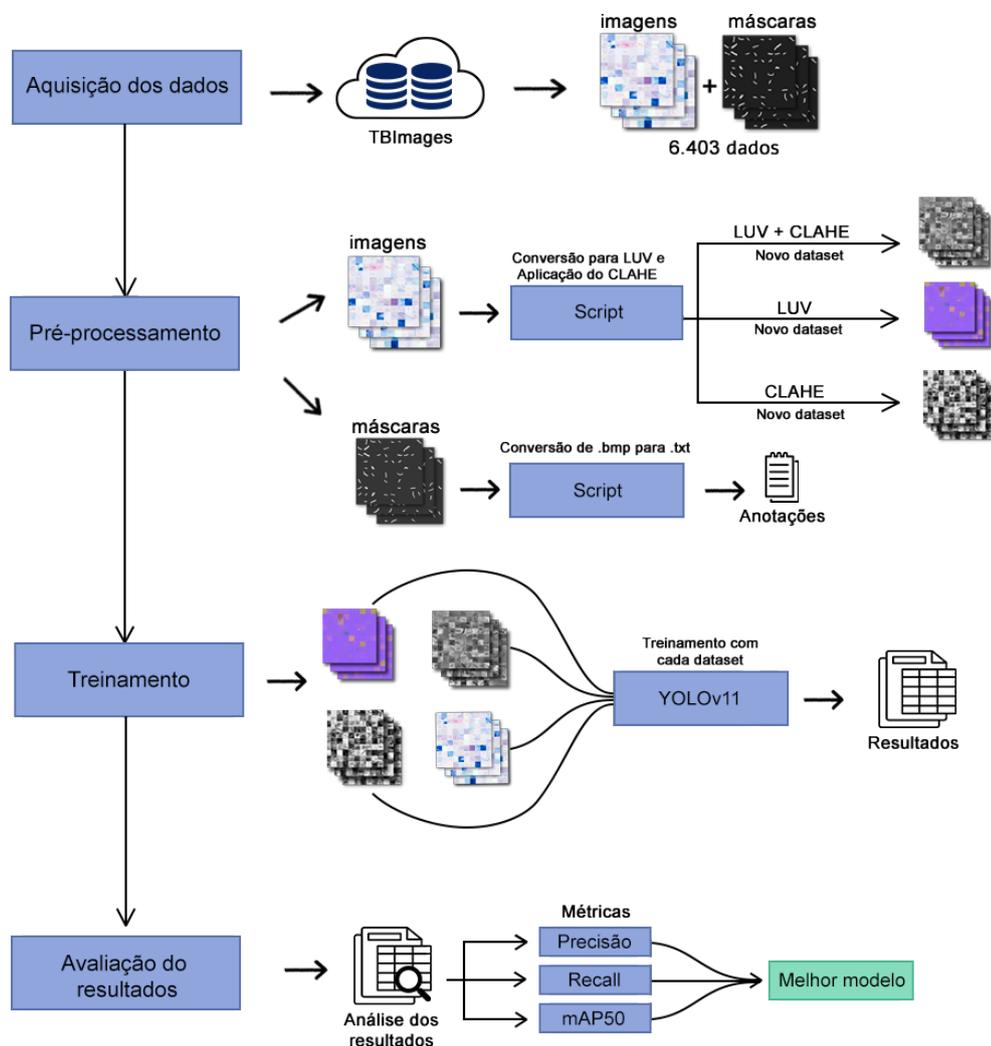


Figura 3 – Representação esquemática do fluxo de trabalho para desenvolvimento e avaliação do modelo de detecção de MT. As etapas incluem a preparação dos dados, o pré-processamento por canais de cor, o treinamento e a avaliação de resultados. Fonte: (Autor).

4.1 Base de Dados

Foi utilizado nesse trabalho imagens de MT. Essa base foi disponibilizada pelo Grupo de Pesquisa em Reconhecimento e Otimização de Padrões da Universidade Federal do Amazonas - UFAM (SERRÃO et al., 2024). O conjunto completo utilizado na pesquisa é composta por um total de 6.403 arquivos, que se dividem em duas categorias. A primeira é composta por 2.852 imagens, extraídas dos campos digitais das lâminas S01 a S15, todas no canal de cor RGB. A segunda possui 3.551 imagens, compostas por máscaras binárias. Essas máscaras funcionam como uma espécie de gabarito, indicando em branco os pixels das regiões que possuem MT e em preto o *background* da imagem.

Todas as imagens possuem resolução de 400×400 pixels, tanto as lâminas originais quanto suas respectivas máscaras. Os arquivos estão no formato BMP (*Bitmap*), um tipo de imagem rasterizada amplamente utilizado para armazenar dados visuais digitais sem compressão. Em termos de organização, as imagens são apresentadas como mosaicos em grade, contendo 100 sub-imagens com resolução de 40×40 cada, com uma mistura de áreas com manchas positivas e negativas (com ou sem presença de MT, respectivamente). A Figura 4 fornece um exemplo visual de uma imagem da base de dados e juntamente com sua máscara correspondente.

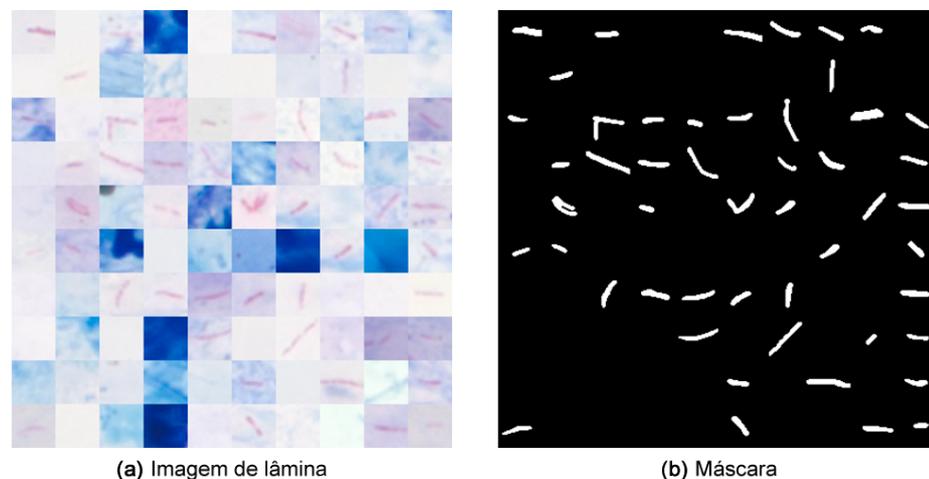


Figura 4 – Exemplo visual de uma imagem da base de dados com sua respectiva máscara. O modelo da esquerda (a) é uma imagem em mosaico com a presença de MT em algumas sub-imagens. No segundo modelo (b), é representado a máscara, com regiões em branco representando a localização dos MT na primeira imagem. Fonte: (Autor).

4.2 Pré-processamento

A primeira etapa do pré-processamento consistiu na verificação e organização dos dados. Neste ponto, identificou-se uma inconsistência no pareamento entre as imagens e suas respectivas máscaras em todos os conjuntos (treino, validação e teste). Para solu-

cionar essa disparidade, foi aplicado um filtro para remover todos os arquivos que não possuíam um par correspondente, garantindo uma relação unívoca entre cada imagem e sua máscara.

Após essa limpeza, a contagem de amostras foi alterada, e o conjunto de teste ficou com um número inferior aos 250 arquivos desejados. A fim de compor um conjunto de teste robusto com 250 pares, os arquivos faltantes foram remanejados do conjunto de treino, que possuía o maior volume de dados validados. Ao final desse processo de organização, a base de dados foi consolidada com um conjunto de treino de 2.032 pares imagem-máscara e um de validação com 62 pares, além do conjunto de teste finalizado com 250 pares. Essa organização resultou na seguinte distribuição de dados: 87% para treino, 10% para teste e 3% para validação.

Para aprimorar a qualidade das imagens e, conseqüentemente, o desempenho dos modelos, foram investigadas quatro técnicas de pré-processamento. Esse processo resultou na criação de quatro conjuntos de dados distintos para análise comparativa:

- RGB (Base): Imagens originais, não modificadas, no espaço de cor RGB.
- LUV: Imagens convertidas para o espaço de cor LUV. Esse espaço apresenta a vantagem de desacoplar a luminância (L) da crominância (UV), característica que pode auxiliar na separação de características texturais de estruturas microscópicas das variações de cor.
- RGB + escala de cinza + CLAHE: As imagens originais em RGB foram inicialmente convertidas para escala de cinza. Esta etapa é fundamental, pois descarta os dados de cor e permite que o CLAHE atue exclusivamente sobre as variações de intensidade, o que é essencial para acentuar os sutis contornos morfológicos dos MT em relação ao fundo. Em seguida, foi aplicado o método *Contrast Limited Adaptive Histogram Equalization* (CLAHE) para melhorar o contraste local e mitigar os efeitos de iluminação não uniforme.
- LUV + escala de cinza + CLAHE: Abordagem em duas etapas, na qual as imagens foram primeiramente convertidas para o espaço de cor LUV. Posteriormente, essa representação em LUV foi transformada em escala de cinza, criando um mapeamento de luminância único, influenciado pelas propriedades perceptuais do espaço LUV. Por fim, aplicou-se o CLAHE sobre essa imagem em escala de cinza para realçar o contraste local.

4.3 Treinamento com modelo YOLOv11

Para a tarefa de detecção e segmentação de MT, a arquitetura de RNC selecionada foi a YOLO. A escolha se baseia no notável equilíbrio entre velocidade de processamento e

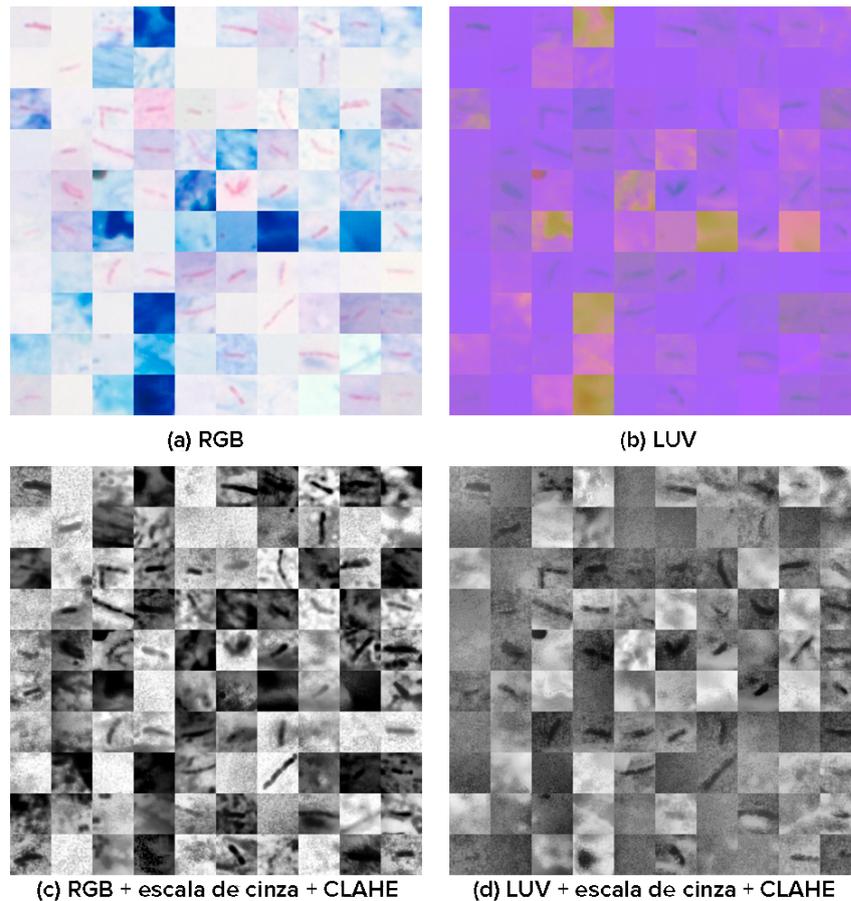


Figura 5 – Efeito de diferentes técnicas de pré-processamento sobre uma imagem amostra do conjunto de dados: (a) Imagem original no espaço de cor RGB; (b) Imagem convertida para o espaço de cor LUV; (c) Imagem em escala de cinza obtida após aplicação da técnica CLAHE à imagem RGB; (d) Imagem em escala de cinza obtida a partir da conversão da imagem LUV e posterior aplicação do CLAHE. Fonte: (Autor).

precisão que a YOLO oferece, tornando-a particularmente eficaz para tarefas de detecção de objetos, conforme detalhado na Seção 2.5. Para esse problema, foi empregada a versão YOLOv11 (Seção 2.5.1), que se destaca das anteriores por ganhos em acurácia, rapidez de inferência e otimização do uso computacional.

Os hiperparâmetros do treinamento foram ajustados empiricamente com o objetivo de otimizar o desempenho do modelo. Os valores estabelecidos, detalhados na Tabela 2, buscaram um equilíbrio entre a estabilidade da convergência e a eficiência computacional. A taxa de aprendizado, por exemplo, foi complementada por uma etapa de aquecimento *warm-up* para estabilizar o início do processo. Adicionalmente, para prevenir o *overfitting* e otimizar o processo, foi empregada a técnica de parada antecipada (*early stopping*). Foi definido um parâmetro de *patience* de 20 épocas. Para monitorar a convergência e a estabilidade do processo, foram analisados os gráficos da função de perda gerados pelo framework de treinamento, permitindo verificar o progresso do aprendizado e identificar potenciais anomalias.

O trabalho computacional foi dividido em duas máquinas distintas. Para a realiza-

Tabela 2 – Hiperparâmetros utilizados no treinamento do modelo.

Hiperparâmetro	Valor
Taxa de Aprendizado (<i>Learning Rate</i>)	$3 \cdot 10^{-4}$
Épocas de Aquecimento (<i>Warm-up</i>)	3
Tamanho do Lote (<i>Batch Size</i>)	16
Número Total de Épocas	200
<i>early stopping</i>	20

ção do pré-processamento das imagens, foi utilizado um hardware construído com um processador Intel i5 de 11^a geração, 16 GB de memória RAM e um SSD SATA para armazenamento dos dados. Para a fase de treinamento foi necessário um hardware com maior poder computacional, sendo conduzida em uma máquina equipada com uma GPU NVIDIA GeForce RTX 2080 Ti, mantendo também um SSD SATA para armazenar os dados de treinamento.

5 Resultados e Discussão

Neste capítulo, serão apresentados os resultados após o treinamento de cada modelo utilizando os quatros diferentes conjuntos de dados. O foco principal será a análise das métricas de avaliação, conforme descritas na Seção 2.6. A análise considerou quatro variações de modelos, treinados com as seguintes bases de dados: (a) RGB, (b) LUV, (c) RGB em escala de cinza + CLAHE e (d) LUV em escala de cinza + CLAHE. Essa abordagem permite avaliar de forma clara a influência de cada técnica de pré-processamento no desempenho do modelo.

5.1 Avaliação de Desempenho da Detecção e Segmentação

A Tabela 3 apresenta os resultados da detecção de MT em imagens de BE. Durante o treinamento, as técnicas descritas na Seção 4.2 foram aplicadas para determinar qual representação de cor melhor se adequa à metodologia do estudo. Os resultados indicam que o espaço de cores LUV se destacou por atingir as métricas mais altas, com uma precisão de 97,60%, recall de 95,60% e um mAP de 98,30%.

Tabela 3 – Métricas das predições de caixas delimitadoras (*bouding boxes*)

Base de dados	Precisão	Recall	mAP
RGB	97,30%	95,30%	97,90%
LUV	97,60%	95,60%	98,30%
RGB + escala de cinza + CLAHE	84,00%	85,40%	89,00%
LUV + escala de cinza + CLAHE	89,30%	81,10%	88,20%

A análise das métricas mostrou um desempenho superior para o modelo treinado com o dataset em LUV, que superou as outras abordagens em precisão, recall e mAP. Isso pode ser atribuído à forma como o LUV processa a imagem, separando a luminância da cromaticidade, reduzindo assim a interferência da iluminação e aprimorando a diferenciação das estruturas morfológicas dos MT. Embora o conjunto de dados no espaço de cores RGB tenha produzido métricas ligeiramente inferiores, a diferença mínima sugere que a conversão para LUV serviu para refinar a detecção, aumentando a precisão sem comprometer a sensibilidade geral do modelo. Por outro lado, as abordagens que aplicaram a conversão para escala de cinza seguida pela técnica CLAHE resultaram em um desempenho inferior. Isso ocorreu tanto com a base de dados derivada do RGB quanto com a derivada do LUV.

A Tabela 4 apresenta os resultados da segmentação MT. Observa-se que o conjunto de dados utilizando o espaço de cores LUV obteve o melhor desempenho geral, atingindo

uma precisão de 84,60%, um recall de 82,80% e um mAP de 84,80%, superando os valores obtidos com o conjunto de dados baseado em RGB.

Tabela 4 – Métricas das segmentações realizadas pelos modelos

Base de dados	Precisão	Recall	mAP
RGB	83.50%	81.50%	83.70%
LUV	84.60%	82.80%	84.80%
RGB + escala de cinza + CLAHE	71.50%	72.70%	74.50%
LUV + escala de cinza + CLAHE	70.80%	64.30%	65.70%

Na segmentação, o modelo treinado com o espaço de cores LUV manteve as métricas elevadas, reforçando a hipótese de que a separação da luminância e da cromaticidade melhora a definição dos contornos dos MT, permitindo uma segmentação mais precisa. No entanto, a aplicação da conversão para escala de cinza seguida pela técnica CLAHE teve um impacto consistentemente negativo, levando a uma redução significativa em todas as métricas de segmentação. Esse comportamento, observado tanto no fluxo que partiu do RGB quanto no que partiu do LUV.

A principal causa para essa queda de performance nos modelos que utilizam escala de cinza unido com o CLAHE é, muito provavelmente, a remoção da informação de cor durante a conversão para escala de cinza. O método de coloração Ziehl-Neelsen, base do diagnóstico visual, utiliza a cor vermelha como o principal diferenciador dos MT. Ao eliminar essa característica, o modelo foi forçado a basear-se apenas em atributos morfológicos e de textura, que são menos distintivos.

5.2 Estudos de Caso

Neste estudo de caso, uma avaliação visual das detecções e segmentações de MT foi conduzida para validar a precisão dos modelos na identificação de MT pequenos e de baixa opacidade. A análise foi realizada utilizando uma imagem do conjunto de dados caracterizada por MT de tamanho pequeno e baixa visibilidade, possibilitando levar a imprecisões nas previsões. Os resultados foram comparados visualmente entre os quatro conjuntos de dados utilizados no treinamento: (a) RGB, (b) LUV, (c) RGB + escala de cinza + CLAHE e (d) LUV + escala de cinza + CLAHE.

A Figura 6 apresenta os resultados de detecção e segmentação de MT obtidos a partir dos diferentes conjuntos de dados. O modelo treinado com o conjunto de dados no espaço de cores LUV obteve o melhor desempenho, detectando 51 MT de um total de 52. Em comparação, o modelo treinado com o conjunto de dados RGB também detectou 51 MT, demonstrando desempenho semelhante. O modelo treinado com o conjunto de dados LUV + escala de cinza + CLAHE detectou o mesmo número dos modelos anteriores, enquanto

o modelo utilizando RGB + escala de cinza + CLAHE apresentou o pior desempenho, detectando 46 MT.

Além de analisar o número de detecções feitas por cada modelo, as pontuações de confiança de predição também foram examinadas para avaliar a confiabilidade de cada abordagem na detecção de MT pequeno ou de baixa visibilidade. O modelo treinado com o canal de cores LUV mostrou uma ligeira redução na confiança, com valores não excedendo 90%. Da mesma forma, na tarefa de segmentação, alguns MT não foram totalmente identificados. O modelo baseado no canal de cores RGB produziu resultados comparáveis, embora algumas predições tenham pontuações de confiança muito baixas e dificuldades na segmentação dos MT. O modelo treinado com o conjunto de dados LUV + escala de cinza + CLAHE exibiu uma ligeira diminuição nos valores de confiança em comparação com os modelos anteriores, além de perder alguns MT e produzir detecções duplicadas do mesmo MT. O modelo baseado em RGB + escala de cinza + CLAHE enfrentou desafios significativos tanto na detecção quanto na segmentação, com pontuações de confiança extremamente baixas.

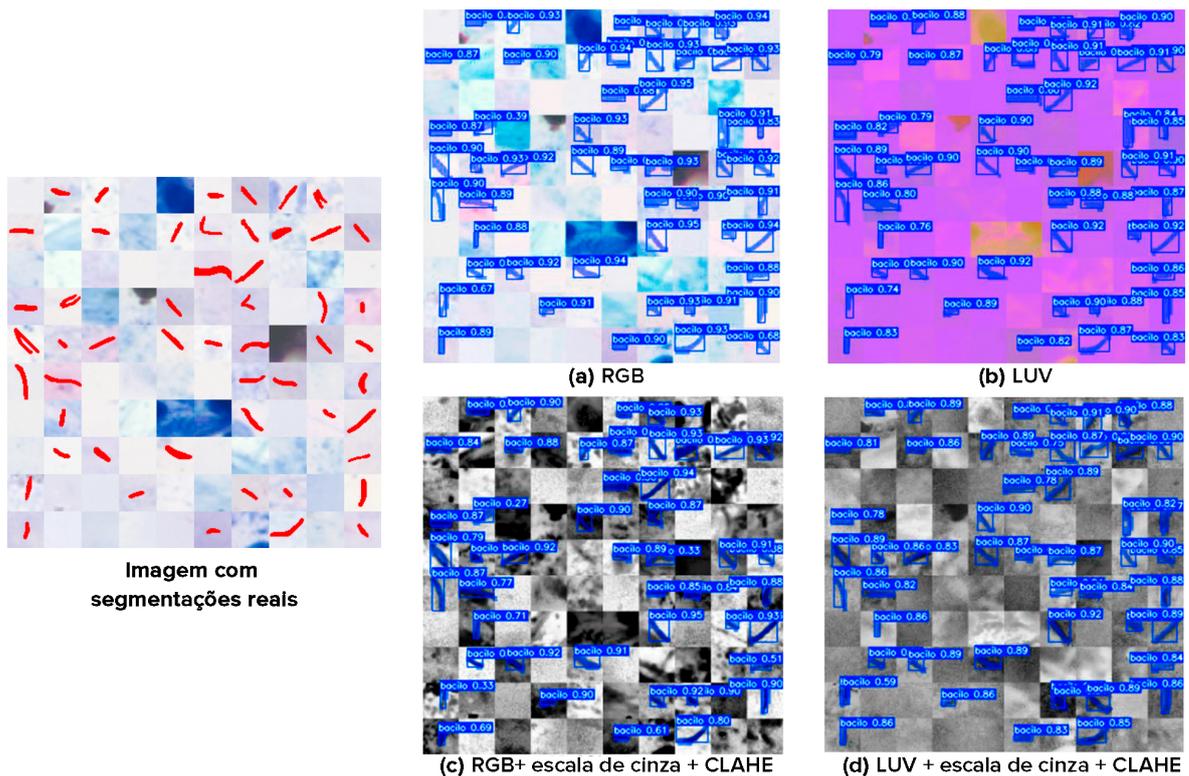


Figura 6 – Comparação visual das detecções e segmentações nas diferentes bases de dados

Esses resultados demonstram que, enquanto as métricas quantitativas sugerem um ótimo desempenho geral, a análise qualitativa expõe que o modelo tem limitações em identificar MT com características mais desafiadoras. Essa questão pode afetar diretamente a aplicabilidade da ferramenta na prática, como em cenários clínicos, onde a detecção precisa de MT pequenas ou de baixa visibilidade é essencial para um diagnóstico

confiável. Portanto, os resultados indicam que, para avançar, existe a necessidade de melhorias. Futuros trabalhos podem explorar otimizações no treinamento do modelo, como o uso de técnicas de aumento de dados mais direcionadas para simular casos complexos, além de ajustes na arquitetura da rede e a incorporação de métodos de segmentação mais avançados.

6 Produção Científica e Técnica

O desenvolvimento deste trabalho transcendeu a análise teórica e a experimentação, resultando em produções científicas e tecnológicas tangíveis que fortalecem a relevância e o impacto da pesquisa. Como fruto deste esforço, foram submetidos artigos a veículos acadêmicos e realizados registros de propriedade intelectual dos sistemas computacionais desenvolvidos.

6.1 Artigos Científicos Submetidos

Com o objetivo de disseminar os resultados obtidos e contribuir com o avanço da comunidade científica, foram elaborados e submetidos dois artigos a veículos de prestígio, atualmente em processo de avaliação por pares:

- **Metodologia Automática para Classificação de Bacilos Utilizando a Arquitetura Faster R-CNN:** Submetido à Revista de Sistemas de Computação (RSC). Este artigo aprofunda a investigação sobre a classificação de MT, validando a eficácia de arquiteturas de RNC no contexto do diagnóstico automatizado da TB.
- **Multitask Learning Based on YOLOv11 for Automatic Detection and Segmentation of Tuberculosis Bacilli in Sputum Smear Microscopy:** Submetido à Conferência Brasileira sobre Sistemas Inteligentes (BRACIS). Este trabalho apresenta a principal contribuição desta monografia, detalhando a aplicação inovadora da arquitetura YOLOv11 em uma abordagem de aprendizagem multitarefa, com foco na detecção e segmentação simultânea de MT.

6.2 Registros de Software Produzidos

Visando à proteção da propriedade intelectual e ao reconhecimento da inovação tecnológica gerada, foram realizados os registros dos softwares desenvolvidos junto ao Instituto Nacional da Propriedade Industrial (INPI). Tais registros não apenas asseguram os direitos autorais sobre os programas, como também atestam o potencial de aplicação prática e o caráter inovador das soluções propostas.

Até o momento, foram obtidos os seguintes registros oficiais:

- **TbKoch:** Plataforma voltada ao diagnóstico automatizado da tuberculose e à gestão de pacientes.
- **TBKoch API:** Interface programável para predição de tuberculose com uso de técnicas de AP.

A formalização dessas produções evidencia que este trabalho não apenas gerou conhecimento científico relevante, como também resultou em soluções tecnológicas com potencial de impacto direto na área da saúde, especialmente na otimização do diagnóstico da tuberculose.

7 Conclusão

Este trabalho partiu da necessidade de aprimorar o diagnóstico da TB, propondo uma abordagem computacional baseada em VC e AP para superar as limitações das técnicas manuais. Para isso, foi desenvolvido e avaliado um método de detecção e segmentação automatizada do MT, utilizando a moderna arquitetura YOLOv11 em conjunto com uma investigação sistemática sobre o impacto de diferentes esquemas de cores no desempenho do modelo.

Os objetivos propostos para esta pesquisa foram plenamente alcançados. Primeiramente, avaliou-se a eficácia da arquitetura YOLOv11, que demonstrou alto desempenho, com o modelo treinado em LUV atingindo 97,60% de precisão na detecção e 84,60% na segmentação. Em segundo lugar, investigou-se o impacto dos esquemas de cores, confirmando que a representação LUV foi superior às demais, inclusive ao tradicional RGB e às abordagens com CLAHE. Por fim, validou-se a abordagem LUV como a mais eficaz para a segmentação precisa dos MT, reforçando as hipóteses iniciais do estudo.

A principal contribuição deste trabalho para a academia e para a tecnologia reside em três pilares. Primeiro, a validação da arquitetura de ponta YOLOv11 para a tarefa específica de análise de imagens de BE, fornecendo um novo *benchmark* de desempenho para a comunidade. Segundo, a demonstração de uma abordagem eficiente de aprendizagem multitarefa, na qual um único modelo realiza simultaneamente a detecção e a segmentação, otimizando o processo computacional. Por fim, este estudo transcendeu a teoria ao gerar produtos concretos, como a submissão de artigos científicos a veículos qualificados e o registro de dois softwares (“TbKoch” e “TBKoch API”) junto ao INPI, o que atesta o caráter inovador e o potencial de aplicação prática da solução desenvolvida.

Apesar dos resultados promissores, reconhece-se que o modelo apresentou limitações na identificação de MT de baixa visibilidade, conforme observado na análise qualitativa. Diante disso e dos avanços alcançados, abrem-se caminhos para futuras investigações. Sugere-se:

- Otimização do treinamento com o uso de técnicas de aumento de dados mais direcionadas para simular casos desafiadores, como MT pequenos ou com pouca coloração.
- Exploração de outras técnicas de pré-processamento que possam realçar o contraste sem descartar a crucial informação de cor.
- Validação do modelo em conjuntos de dados de diferentes fontes e laboratórios para testar sua capacidade de generalização.
- Realização de um estudo de validação clínica, comparando o desempenho da ferramenta com o de técnicos de laboratório em um fluxo de trabalho real.

Conclui-se, portanto, que este trabalho cumpre seu papel ao oferecer uma solução tecnológica robusta, validada e com potencial de impacto real, contribuindo com um passo significativo para a automação e democratização do diagnóstico da TB.

Referências

- ALVES, L. F. V. *Avaliação Antecipada do Risco de Desistência em Disciplinas de Programação da ECT/UFRN: uma abordagem de aprendizado de máquina supervisionado e não supervisionado*. Dissertação (B.S. thesis) — Universidade Federal do Rio Grande do Norte, 2024. Citado na página 18.
- BARROS, C. M. et al. Aplicando redes neurais convolucionais em imagens para reconhecimento automatizado de abelhas melíferas (*apis mellifera* l.). In: SBC. *Workshop de Computação Aplicada à Gestão do Meio Ambiente e Recursos Naturais (WCAMA)*. [S.l.], 2021. p. 19–28. Citado na página 19.
- BOCHKOVSKIY, A.; WANG, C.-Y.; LIAO, H.-Y. M. *YOLOv4: Optimal Speed and Accuracy of Object Detection*. 2020. Disponível em: <<https://arxiv.org/abs/2004.10934>>. Citado na página 19.
- BRASIL, E. A. *Exame de Baciloscopia para Tuberculose*. 2020. Disponível em: <<https://www.gov.br/pt-br/servicos-estaduais/exame-de-baciloscopia-para-tuberculose-1>>. Citado na página 16.
- BRASIL, M. da S. *Tuberculose*. 2023. Disponível em: <<https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/t/tuberculose>>. Citado na página 12.
- CARVALHO, T. F. M. Técnicas de inteligência computacional aplicadas ao diagnóstico de tuberculose. 2023. Citado na página 23.
- CONSTÂNCIO, A. S.; CARVALHO, D. R.; TSUNODA, D. F. Aplicações de visão computacional na saúde: revisão de literatura incrementada com técnicas de processamento de linguagem natural. *Research, Society and Development*, v. 11, n. 10, p. e218111032942–e218111032942, 2022. Citado na página 17.
- CUNHA, C. S. et al. O papel da imprensa na disseminação dos tratamentos para a tuberculose no rio de janeiro no ano de 1900. *Saberes Interdisciplinares*, v. 16, n. 29, p. 55–67, 2023. Citado na página 15.
- GIOSEFFI, J. R.; BATISTA, R.; BRIGNOL, S. M. Tuberculose, vulnerabilidades e hiv em pessoas em situação de rua: revisão sistemática. *Revista de Saúde Pública*, SciELO Brasil, v. 56, p. 43, 2022. Citado 2 vezes nas páginas 12 e 15.
- GOMES, J. V. E. Detecção de objetos com a arquitetura yolo. 2022. Citado na página 19.
- GOMIDE, J. V. B. et al. Detecção e contagem automáticas de bacilos álcool ácido resistentes para o diagnóstico da tuberculose. *Código 31: revista de informação, comunicação e interfaces*, v. 1, n. 1, 2023. Citado na página 12.
- HODONOU, L. G. *Geração e Detecção de Objetos em Documentos por Modelos de Redes Neurais de Aprendizagem Profunda (DeepDocGen)*. Tese (Doutorado) — PUC-Rio, 2024. Citado na página 18.

- HUANG, J. et al. Lw-yolo11: A lightweight arbitrary-oriented ship detection method based on improved yolo11. *Sensors*, MDPI, v. 25, n. 1, p. 65, 2024. Citado na página 20.
- IZBICKI, R.; SANTOS, T. M. dos. *Aprendizado de máquina: uma abordagem estatística*. [S.l.]: Rafael Izbicki, 2020. Citado na página 17.
- KHANAM, R.; HUSSAIN, M. *What is YOLOv5: A deep look into the internal features of the popular object detector*. 2024. Disponível em: <<https://arxiv.org/abs/2407.20892>>. Citado na página 19.
- KINOSHITA, J. K.; MENEGHETTI, D. D. R.; BIANCHI, R. A. da C. Aprendizado por reforço profundo com redes recorrentes aplicado à negociação do minicontrato futuro de dólar. In: SBC. *Brazilian Workshop on Artificial Intelligence in Finance (BWAIF)*. [S.l.], 2022. p. 13–24. Citado na página 18.
- LEOCÁDIO, R. R. et al. Detecção de abelhas nativas em colmeias em campo utilizando visão computacional. In: SBC. *Workshop de Computação Aplicada à Gestão do Meio Ambiente e Recursos Naturais (WCAMA)*. [S.l.], 2021. p. 59–68. Citado na página 17.
- LINGINANI, I.; MUDDANA, A. 4 computer vision for healthcare applications. *Computer Vision: Applications of Visual AI and Image Processing*, Walter de Gruyter GmbH & Co KG, v. 15, p. 73, 2023. Citado na página 17.
- LÓPEZ, Y. P. Detecção do mycobacterium tuberculosis em imagens de baciloscopia de campo claro utilizando redes neurais convolutivas. Universidade Federal do Amazonas, 2018. Citado 3 vezes nas páginas 22, 23 e 24.
- LUDERMIR, T. B. Inteligência artificial e aprendizado de máquina: estado atual e tendências. *Estudos Avançados*, SciELO Brasil, v. 35, p. 85–94, 2021. Citado 2 vezes nas páginas 17 e 18.
- MARQUES, C. R. et al. Um estudo comparativo entre diferentes modelos de redes neurais de aprendizado profundo aplicados ao problema da inversão sísmica. 2022. Citado na página 18.
- MARTINS, V. D. O.; MIRANDA, C. V. D. Diagnóstico e tratamento medicamentoso em casos de tuberculose pulmonar: revisão de literatura. *Revista saúde multidisciplinar*, v. 7, n. 1, 2020. Citado 2 vezes nas páginas 12 e 15.
- OLIVEIRA, P. D. S. d. Uso de aprendizagem de máquina e redes neurais convolucionais profundas para a classificação de áreas queimadas em imagens de alta resolução espacial. 2020. Citado na página 18.
- PEREIRA, A. D. et al. Redes neurais convolucionais para análise de mamografias. Universidade Tecnológica Federal do Paraná, 2025. Citado 2 vezes nas páginas 18 e 19.
- REDMON, J. et al. *You Only Look Once: Unified, Real-Time Object Detection*. 2016. Disponível em: <<https://arxiv.org/abs/1506.02640>>. Citado 3 vezes nas páginas 8, 19 e 20.
- REDMON, J.; FARHADI, A. *YOLOv3: An Incremental Improvement*. 2018. Disponível em: <<https://arxiv.org/abs/1804.02767>>. Citado na página 19.

- REIS, F. J. D. S. Uma abordagem automatizada para detecção e classificação do bacilo da tuberculose. 2022. Citado 2 vezes nas páginas 22 e 23.
- RODRIGUES, F. M. et al. Metodologia automática para detecção de bacilos de tuberculose utilizando retinanet e modelos de cores. In: SBC. *Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)*. [S.l.], 2022. p. 334–345. Citado na página 13.
- SALES, K. et al. Aceleração de preenchimento da tabela q em ferramenta de automação de teste de dispositivos android baseada em aprendizado por reforço. In: SBC. *Simpósio Brasileiro de Engenharia de Sistemas Computacionais (SBESC)*. [S.l.], 2023. p. 43–48. Citado na página 18.
- SANTOS, P. R. S dos; BRITO, V. de C.; FILHO, A. O. de C. Kochdet: Arquitetura profunda baseada em bfpn para detecção de bacilos de kock. *Revista de Sistemas e Computação-RSC*, v. 13, n. 1, 2023. Citado 2 vezes nas páginas 22 e 23.
- SANTOS, T. T. et al. Visão computacional aplicada na agricultura. In: MASSRUHÁ, SMFS; LEITE, MA de A.; OLIVEIRA, SR de M.; MEIRA, CAA ... , 2020. Citado na página 17.
- SERRÃO, M. K. M. et al. Automatic bright-field smear microscopy for diagnosis of pulmonary tuberculosis. *Computers in Biology and Medicine*, Elsevier, v. 172, p. 108167, 2024. Citado na página 26.
- SILVA, M. T.; GALVÃO, T. F. Incidência de tuberculose no brasil: análise de série temporal entre 2001 e 2021 e projeção até 2030. *Revista Brasileira de Epidemiologia*, SciELO Brasil, v. 27, p. e240027, 2024. Citado na página 15.
- SOEIRO, V. M. d. S.; CALDAS, A. d. J. M.; FERREIRA, T. F. Abandono do tratamento da tuberculose no brasil, 2012-2018: tendência e distribuição espaço-temporal. *Ciência & saúde coletiva*, SciELO Brasil, v. 27, n. 03, p. 825–836, 2022. Citado na página 15.
- SOUZA, M. P. d. Explicabilidade de modelos de visão computacional aplicados à detecção de covid-19 em radiografias. 2020. Citado na página 21.
- WANG, C.-Y.; BOCHKOVSKIY, A.; LIAO, H.-Y. M. *YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors*. 2022. Disponível em: <<https://arxiv.org/abs/2207.02696>>. Citado na página 22.
- WHO, W. H. O. *Tuberculosis*. 2024. Disponível em: <<https://www.who.int/news-room/fact-sheets/detail/tuberculosis>>. Citado 2 vezes nas páginas 12 e 15.
- XIONG, Y. et al. Automatic detection of mycobacterium tuberculosis using artificial intelligence. *Journal of thoracic disease*, v. 10, n. 3, p. 1936, 2018. Citado 2 vezes nas páginas 23 e 24.
- YASEEN, M. *What is YOLOv8: An In-Depth Exploration of the Internal Features of the Next-Generation Object Detector*. 2024. Disponível em: <<https://arxiv.org/abs/2408.15857>>. Citado na página 20.
- ZHOU, X. Application and analysis of computer vision algorithms in graphics and image processing. *International Journal of Informatics and Information Systems*, v. 6, n. 1, p. 8–15, 2023. Citado na página 17.



TERMO DE AUTORIZAÇÃO PARA PUBLICAÇÃO DIGITAL NA BIBLIOTECA “JOSÉ ALBANO DE MACEDO”

Identificação do Tipo de Documento

- Tese
- Dissertação
- Monografia
- Artigo

Eu, **Augusto Almondes Teixeira**, autorizo com base na Lei Federal nº 9.610 de 19 de Fevereiro de 1998 e na Lei nº 10.973 de 02 de dezembro de 2004, a biblioteca da Universidade Federal do Piauí a divulgar, gratuitamente, sem ressarcimento de direitos autorais, o texto integral da publicação Método Automático para Detecção e Segmentação de Bacilos de Koch Utilizando YOLOv11 e Esquemas de Cores de minha autoria, em formato PDF, para fins de leitura e/ou impressão, pela internet a título de divulgação da produção científica gerada pela Universidade.

Picos-PI 25 de Julho de 2025.

Documento assinado digitalmente
gov.br **AUGUSTO ALMONDES TEIXEIRA**
Data: 04/07/2025 13:42:32-0300
Verifique em <https://validar.iti.gov.br>

Assinatura