



**UFAM**

UNIVERSIDADE FEDERAL DO AMAZONAS

FACULDADE DE TECNOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

MIKAELA KALLINE MACIEL SERRÃO

DIAGNÓSTICO DA TUBERCULOSE EM MICROSCOPIA DE CAMPO CLARO  
USANDO REDES PROFUNDAS

MANAUS

2024

UNIVERSIDADE FEDERAL DO AMAZONAS  
FACULDADE DE TECNOLOGIA  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

MIKAELA KALLINE MACIEL SERRÃO

DIAGNÓSTICO DA TUBERCULOSE EM MICROSCOPIA DE CAMPO CLARO  
USANDO REDES PROFUNDAS

Tese apresentada ao Programa de Pós-Graduação em Engenharia Elétrica, área de concentração Controle e Automação de Sistemas, como requisito parcial à obtenção do título de Doutora em Engenharia Elétrica.

Orientador: Prof. Dr. Cícero Ferreira Fernandes Costa Filho  
Co-Orientadora: Profa. Dra. Marly Guimarães Fernandes Costa

MANAUS

2024

## Ficha Catalográfica

Ficha catalográfica elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

S487d Serrão, Mikaela Kalline Maciel  
Diagnóstico da tuberculose em microscopia de campo claro  
usando redes profundas / Mikaela Kalline Maciel Serrão . 2024  
118 f.: il. color; 31 cm.

Orientador: Cícero Ferreira Fernandes Costa Filho  
Coorientadora: Marly Guimarães Fernandes Costa  
Tese (Doutorado em Engenharia Elétrica) - Universidade Federal  
do Amazonas.

1. diagnóstico da tuberculose pulmonar. 2. segmentação  
semântica. 3. microscopia de campo claro. 4. convolução separada  
por canal. 5. mecanismo de atenção de canal. I. Costa Filho, Cícero  
Ferreira Fernandes. II. Universidade Federal do Amazonas III.  
Título



**Poder Executivo  
Ministério da Educação  
Universidade Federal do Amazonas  
Faculdade de Tecnologia  
Programa de Pós-graduação em Engenharia Elétrica**

MIKAELA KALLINE MACIEL SERRÃO

**DIAGNÓSTICO DA TUBERCULOSE EM MICROSCOPIA DE CAMPO CLARO  
USANDO REDES PROFUNDAS.**

Tese apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal do Amazonas, como requisito parcial para obtenção do título de Doutor em Engenharia Elétrica na área de concentração Controle e Automação de Sistemas.

Aprovado em 01 de abril de 2024.

**BANCA EXAMINADORA**

DocuSigned by:

*Cícero Ferreira Fernandes Costa Filho*

13A5DC8FCCFE4CC

Prof. Dr. Cícero Ferreira Fernandes Costa Filho, Presidente  
Universidade Federal do Amazonas

DocuSigned by:

*Jose Raimundo Gomes Pereira*

F59DD0C4306D49D

Prof. Dr. José Raimundo Gomes Pereira, Membro  
Universidade Federal do Amazonas

DocuSigned by:

*Maurício Morishi Ogusku*

DE75D2FB7A874D5

Prof. Dr. Maurício Morishi Ogusku  
Instituto Nacional de Pesquisas da Amazônia

DocuSigned by:

*Luciana Botinelly Mendonça Fujimoto*

7A148E038F9A41D

Prof<sup>a</sup>. Dra. Luciana Botinelly Mendonça Fujimoto  
Universidade Federal do Amazonas

DocuSigned by:

*Fernando José Ribeiro Sales*

4982E45D2C4043B

Prof. Dr. Fernando José Ribeiro Sales  
Universidade Federal de Pernambuco

## AGRADECIMENTOS

Agradeço primeiramente à Deus pela minha vida, pela vida da minha família, por ter me dado saúde e forças para superar as dificuldades durante o desenvolvimento deste trabalho e por cuidar de mim em todos os momentos.

Agradeço aos meus orientadores, Prof. Dr. Cícero Ferreira Fernandes Costa Filho e Prof.<sup>a</sup> Dra. Marly Guimarães Fernandes Costa pelas oportunidades que me deram durante esses anos que me fizeram crescer no âmbito profissional e pessoal. E, principalmente, pela orientação e suporte dado no desenvolvimento deste trabalho.

Nada disso seria possível sem o apoio e suporte constante da minha mãe Valdenoura, meu pai Josivan, do meu irmão Levi que tem tornado minha vida mais leve e alegre desde que nasceu, e do meu esposo Melquisedeque que esteve ao meu lado me incentivando, torcendo por mim e me dando toda a estrutura necessária durante o desenvolvimento deste trabalho.

À Fundação CAPES pelo fornecimento da bolsa que permitiu dedicação exclusiva ao desenvolvimento da tese.

À Universidade Federal do Amazonas e, em especial, ao Centro de Tecnologia Eletrônica e da Informação – CETELI – pela concessão de toda infraestrutura para realização deste trabalho.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES).

Esta pesquisa, realizada no âmbito do Projeto Samsung-UFAM de Ensino e Pesquisa (SUPER), de acordo com o Artigo 39 do Decreto nº10.521/2020, foi financiada pela Samsung Eletrônica da Amazônia Ltda, nos termos da Lei Federal nº8.387/1991, através do convênio 001/2020 firmado com a UFAM e FAEPI, Brasil.

## RESUMO

A baciloscopia automatizada para o diagnóstico da tuberculose pulmonar tem sido tema de muitos estudos publicados nas últimas décadas. A maioria deles trata de uma etapa preliminar do diagnóstico, a detecção de bacilos, enquanto, conforme determinado pela Organização Mundial da Saúde (OMS), a baciloscopia compreende a detecção e relato do número de bacilos encontrados em até 100 campos microscópicos para a identificação do diagnóstico que pode ser Negativo, Esparso, 1+, 2+ ou 3+. O diagnóstico da tuberculose pulmonar na baciloscopia de campo claro depende da atenção de um técnico treinado que, com uma alta demanda laboratorial, pode ser sobrecarregado, enquanto o diagnóstico automatizado requer pouca ou, em alguns casos, nenhuma interpretação por parte de um técnico. Até onde sabemos, este trabalho propõe o primeiro método automático para diagnóstico da tuberculose pulmonar em baciloscopia de campo claro, de acordo com as determinações da OMS. O método proposto compreende uma etapa de segmentação semântica, utilizando uma rede neural profunda, seguida de uma etapa de filtragem visando reduzir o número de objetos detectados que não são bacilos (falso positivos): filtragem baseada nas características de cor e forma do objeto. Na etapa de segmentação semântica, diferentes configurações de codificadores foram avaliadas, usando camadas de convolução separada por canal e mecanismo de atenção de canal. Na etapa de filtragem, dois modelos foram avaliados, o primeiro com um filtro de cor e um filtro de forma, e o segundo com dois filtros de cor e um filtro de forma. O segundo modelo de filtragem apresentou melhor desempenho. O método proposto foi avaliado com um conjunto de dados grande, robusto e anotado projetado para esse fim, composto por 250 conjuntos de teste, em que cada classe diagnóstica tem 50 conjuntos. As seguintes métricas de desempenho foram obtidas com o método proposto para a classificação das 5 classes diagnósticas: precisão média de 0,894, sensibilidade média de 0,896 e *f1-score* médio de 0,895. Além disso, o método apresentou um tempo de diagnóstico de aproximadamente 7 minutos no caso em que são necessários mais campos digitais para a análise. Portanto, os resultados alcançados mostraram a possibilidade de realizar o diagnóstico automático da tuberculose por meio da baciloscopia de campo claro.

**Palavras-chave:** diagnóstico da tuberculose pulmonar, segmentação semântica, microscopia de campo claro, convolução separada por canal, mecanismo de atenção de canal, detecção de bacilos.

## ABSTRACT

The use of automatic smear microscopy for the diagnosis of pulmonary tuberculosis has been the subject of many studies published in recent decades. Most of them deal with a preliminary stage of diagnosis, the bacilli detection, while, as determined by World Health Organization (WHO), smear microscopy comprises the detection and reporting of the number of bacilli found in up to 100 microscopic fields for the prescription of the diagnosis that can be Negative, Scanty, 1+, 2+ or 3+. The diagnosis of pulmonary tuberculosis using bright-field smear microscopy depends on the attention of a trained technician who, with high laboratory demands, can be overloaded. While automated diagnosis requires little or no interpretation by a technician. As far as we know, this work proposes the first automatic method for diagnosing pulmonary tuberculosis using bright-field smear microscopy, according to the WHO guidelines. The proposed method comprises a semantic segmentation step, using a deep neural network, followed by a filtering step aiming to reduce the number of objects detected that are not bacilli (false positives): filtering based on color and shape features of the object. In semantic segmentation step, different encoder configurations were evaluated, using depth-wise separable convolution layers and channel attention mechanism. In the filtering step, two models were evaluated, the first with a color filter and a shape filter, and the second with two color filters and a shape filter. The second filtering model showed better performance. The proposed method was evaluated with a large, robust, and annotated dataset designed for this purpose, consisting of 250 test sets, where each diagnostic class has 50 sets. The following performance metrics were obtained with the proposed method for classifying the 5 diagnostic classes: average precision of 0.894, average recall of 0.896 and average f1-score of 0.895. Furthermore, the method presented a diagnostic time of approximately 7 minutes in the case where more digital fields are needed for the analysis. Therefore, the results achieved showed the possibility of automatically diagnosing tuberculosis using bright-field smear microscopy.

**Keywords:** pulmonary tuberculosis diagnosis, semantic segmentation, smear microscopy, depth-wise separable convolution, channel attention mechanism, bacilli detection.

## LISTA DE FIGURAS

Figura 1. Métodos de detecção de bacilos usados nos artigos revisados. ....	22
Figura 2. Preparação do esfregaço.....	33
Figura 3. Coloração do esfregaço.....	34
Figura 4. Exemplos de lâminas coradas. ....	35
Figura 5. Divisão do campo microscópico. ....	36
Figura 6. Esquema do cubo do espaço de cor RGB. ....	38
Figura 7. Esquema do hexacone do espaço de cor HSV. ....	39
Figura 8. Ilustração de uma estrutura de Árvore de Decisão.....	40
Figura 9. Ilustração de um modelo Random Forest. ....	42
Figura 10. Estrutura básica de um neurônio artificial. ....	43
Figura 11. Uma arquitetura de rede neural. ....	44
Figura 12. Arquitetura LeNet-5.....	47
Figura 13. Representação gráfica da convolução. ....	48
Figura 14. Ilustração da região da camada de entrada conectada à uma camada de saída.....	49
Figura 15. Ilustração de uma camada convolutiva em uma entrada $6 \times 6 \times 3$ , tamanho do <i>kernel</i> $3 \times 3$ , profundidade igual a 2, passo igual a 1 e sem preenchimento de zeros. ....	50
Figura 16. Operação de <i>pooling</i> . ....	51
Figura 17. Convolução transposta de passo 1.....	52
Figura 18. Exemplo do resultado da camada de <i>dropout</i> . ....	53
Figura 19. Processo de segmentação semântica. ....	55
Figura 20. Bloco de atenção de canal SENet.....	57
Figura 21. Camada <i>depthwise separable convolution</i> . ....	57
Figura 22. Exemplos de campos digitais da lâmina S05 e região TR1. ....	59
Figura 23. Exemplo de retalhos.....	61
Figura 24. Exemplo de imagem mosaico do banco de dados BD1.....	62
Figura 25. Número de campos digitais versus número de bacilos por campo digital usados na construção de 250 conjuntos para os diagnósticos. ....	63
Figura 26. Exemplos do banco de dados BD3.....	64
Figura 27. Metodologia da detecção dos bacilos e diagnóstico da Tb. ....	64
Figura 28. Divisão de um campo digital de $1040 \times 1388$ pixels em uma grade de $3 \times 4$ imagens de $400 \times 400$ pixels.....	66

Figura 29. Arquitetura proposta em SERRAO et al. (2020). .....	67
Figura 30. Codificador usando mecanismo de atenção de canal SENet. ....	67
Figura 31. Codificador usando <i>depthwise separable convolution</i> .....	68
Figura 32. Exemplos de objetos segmentados em regiões com problemas de coloração. ....	69
Figura 33. Ilustra a etapa de segmentação semântica.....	70
Figura 34. Diagrama do modelo de filtragem 1.....	71
Figura 35. Diagrama do modelo de filtragem 2.....	71
Figura 36. Arquitetura utilizada na implementação dos filtros de cor. ....	72
Figura 37. Ilustração de todos os planos utilizados na entrada das redes CNN2, CNN3 e CNN4 de um retalho. ....	73
Figura 38. Aplicação do filtro de cor do modelo de filtragem 1. ....	73
Figura 39. Aplicação dos filtros de cor do modelo de filtragem 2. ....	74
Figura 40. Imagem binarizada de um bacilo com o formato alongado, mostrando o maior eixo <i>Rmax</i> e menor eixo <i>Rmin</i> . ....	76
Figura 41. Importância das características geométricas utilizadas na entrada do classificador RF1. Excentricidade é a característica com maior influência na decisão do classificador, seguida pela característica de maior eixo. ....	76
Figura 42. Aplicação do filtro de forma. ....	77
Figura 43. Matriz de confusão mostrando o resultado do diagnóstico da Tb usando segmentação semântica + filtro de área, aplicado o banco de dados BD2.....	80
Figura 44. Matrizes de confusão do diagnóstico da Tb usando os dois modelos de filtragem. ....	81

## LISTA DE TABELAS

Tabela 1. Resultados dos filtros de cor CNN2, CNN3 e CNN4.....	74
Tabela 2. Resultados das arquiteturas avaliadas para a tarefa de segmentação semântica. ....	78
Tabela 3. Redução no número de FPs resultante do uso dos 2 modelos de filtragem, em 10 conjuntos de 164 imagens do conjunto de dados BD2.....	79
Tabela 4. Desempenho do método proposto para diagnóstico de Tb utilizando segmentação semântica + filtro de área + modelo de filtragem 1.....	81
Tabela 5. Desempenho do método proposto para diagnóstico de Tb utilizando segmentação semântica + filtro de área + modelo de filtragem 2.....	81
Tabela 6. Desempenho temporal de cada etapa do método proposto para 11 imagens com diferentes números de bacilos.....	82
Tabela 7. Desempenho temporal de cada classe diagnóstica de Tb pulmonar por baciloscopia de campo claro.....	83

## LISTA DE QUADROS

Quadro 1. Descrição das classes diagnósticas da Tb em termos de contagem de bacilos encontrados em campos microscópicos.....	18
Quadro 2. Resumo da análise realizada nas Revisão Bibliográfica. ....	30
Quadro 3. Descrição dos bancos de dados utilizados na implementação.....	60
Quadro 4. Distribuição das imagens de treinamento, validação e teste do banco de dados BD1. ....	61
Quadro 5. Correspondência entre o diagnóstico da Tb em campos microscópicos e campos digitais. ....	62

## LISTA DE SIGLAS

ADAM	<i>Adaptive Moment Estimation</i>
ANN	<i>Artificial Neural Network</i>
AUC	Área sob a Curva ROC
BAAR	Bacilos Álcool-Ácido Resistentes
BD1	Banco de Dados 1
BD2	Banco de Dados 2
BD3	Banco de Dados 3
CAT	<i>Channel Area Thresholding</i>
CCD	<i>Coupled Charge Device</i>
CNN	<i>Convolutional Neural Network</i>
CNN1	Modelo de Segmentação
CNN2	Filtro de cor
DBN	<i>Deep Belief Network</i>
DP	Desvio Padrão
EDF	Imagens de Foco Estendido
<i>Faster-RCNN</i>	<i>Faster Region-based Convolution Neural Network</i>
FN	Falso Negativo
FP	Falso Positivo
HIV	Vírus da Imunodeficiência Humana
HOG	<i>Histogram of Oriented Gradients</i>
HSV	<i>Hue, Saturation, Value</i>
INPA	Instituto Nacional de Pesquisa da Amazônia

kNN	<i>k-Nearest-Neighbor</i>
LoH	<i>Location-oriented Histogram</i>
MLP	<i>Multilayer Perceptron</i>
NAAT	<i>Nucleic Acid Amplification Tests</i>
OMS	Organização Mundial da Saúde
PCA	<i>Principal Component Analysis</i>
ReLU	<i>Rectifier Linear Unity</i>
RF	<i>Random Forest</i>
RF1	Filtro de forma
RGB	<i>Red, Green, Blue</i>
RMSProp	<i>Root Mean Square Propagation</i>
SENet	<i>Squeeze Excitation Net</i>
SGDM	<i>Stochastic Gradient Descent with Momentum</i>
SURF	<i>Speeded-Up Robust Features</i>
SUS	Sistema Único de Saúde
SVM	<i>Support Vector Machine</i>
Tb	Tuberculose
TDIDT	<i>Top-Down Induction of Decision Trees</i>
TR	<i>Tile Region</i>
TRM-TB	Teste Rápido Molecular para Tuberculose
VN	Verdadeiros Negativo
VP	Verdadeiros Positivo
ZN	<i>Ziehl-Neelsen</i>

## SUMÁRIO

1	INTRODUÇÃO.....	15
1.1	Objetivo Geral.....	19
1.2	Objetivos Específicos .....	19
1.3	Organização do Trabalho.....	19
2	REVISÃO BIBLIOGRÁFICA.....	21
2.1	Detecção de Bacilos em Microscopia de Campo Claro.....	21
2.1.1	Identificação do Bacilo com uma Etapa.....	23
2.1.2	Identificação do Bacilo com duas Etapas.....	24
2.1.3	Identificação do Bacilo com Detecção de Objetos.....	26
2.2	Conclusão.....	27
3	REFERENCIAL TEÓRICO.....	32
3.1	Baciloscopia de Campo Claro.....	32
3.1.1	Método de Preparação do Esfregaço .....	33
3.1.2	Método de Coloração do Esfregaço.....	34
3.1.3	Método de Leitura e Interpretação dos Resultados .....	35
3.2	Espaços de Cor.....	37
3.2.1	RGB.....	37
3.2.2	HSV .....	38
3.3	Aprendizado de Máquina.....	39
3.4	Árvores de Decisão .....	40
3.4.1	Random Forest.....	41
3.5	Redes Neurais Artificiais .....	42
3.5.1	Processo de Treinamento da Rede Neural.....	45
3.5.2	Algoritmo <i>Backpropagation</i> .....	45
3.6	Redes Neurais Convolucionais .....	46
3.6.1	Camada Convolutiva .....	47

3.6.2	Camada ReLU .....	50
3.6.3	Camada de subamostragem ( <i>pooling</i> ) .....	51
3.6.4	Camada de Convolução Transposta .....	51
3.6.5	Camada de <i>dropout</i> .....	52
3.6.6	Camada de normalização em lote.....	53
3.6.7	Regularização L2.....	54
3.6.8	Segmentação Semântica .....	54
3.7	Método de Otimização ADAM.....	55
3.8	Mecanismo de Atenção de Canal ( <i>Channel Attention Mechanism</i> ) .....	56
3.9	Convolução Separada por Canal ( <i>Depthwise Separable Convolution</i> ) .....	57
4	MATERIAIS E MÉTODOS.....	58
4.1	Materiais .....	58
4.1.1	Banco de Dados Original.....	58
4.1.2	Banco de Dados para Segmentação Semântica – BD1 .....	60
4.1.3	Banco de Dados para Diagnóstico – BD2 .....	62
4.1.4	Banco de dados Etapa de Filtragem – BD3.....	63
4.2	Métodos .....	64
4.2.1	Métricas de Avaliação .....	65
4.2.2	Pré-processamento.....	65
4.2.3	Segmentação Semântica .....	66
4.2.4	Redução de Falsos Positivos .....	70
4.2.5	Ambiente de Desenvolvimento .....	77
5	RESULTADOS E DISCUSSÃO .....	78
5.1	Resultados.....	78
5.1.1	Resultados do Diagnóstico .....	79
5.1.2	Desempenho Temporal.....	82
5.2	Discussão .....	83

5.3	Conclusão.....	86
6	REFERÊNCIAS .....	88
	APÊNDICE A – ARTIGO PUBLICADO NO IX CONGRESSO LATINO-AMERICANO DE ENGENHARIA BIOMÉDICA EM 2022.....	94
	APÊNDICE B – ARTIGO PUBLICADO NA REVISTA ACADÊMICA <i>COMPUTERS IN BIOLOGY AND MEDICINE</i> EM 2024.....	104

## 1 INTRODUÇÃO

De acordo com a Organização Mundial da Saúde (OMS), em 2022, estima-se que 10,6 milhões de pessoas desenvolveram a tuberculose (Tb) no mundo todo. A taxa de incidência da Tb foi de 133 novos casos por 100.000 habitantes por ano. O número de mortes causadas pela Tb em pessoas com sorologia para o Vírus da Imunodeficiência Humana (HIV) negativa foi cerca de 1,13 milhão e um adicional de 167.000 mortes entre pessoas com sorologia positiva.

Após dois anos do impacto prejudicial da pandemia de COVID-19 no número de pessoas que morrem ou adoecem por Tb, houve uma recuperação global em 2022. No entanto, a Tb ainda permanece como a segunda principal causa de morte, após o COVID-19, e causou quase o dobro de mortes que o HIV/AIDS. As metas globais de melhorias para redução da Tb continuam difíceis de serem alcançadas (WHO, 2023).

No Brasil, conforme descrito no Boletim Epidemiológico do Ministério da Saúde (MINISTÉRIO DA SAÚDE, 2023), foram registrados 36,3 casos de Tb por 100.000 habitantes em 2022, o que corresponde a uma recuperação parcial no número de pessoas diagnosticadas com a doença antes da pandemia de COVID-19 em 2020. No entanto, tanto o número de casos quanto a incidência da Tb não atingiram os níveis pré-pandêmicos observados. Um total de 13 estados registraram coeficiente de incidência superior ao nacional. Dentre estes, destacam-se o Amazonas, com 84,1 casos por 100.000 habitantes, Roraima, com 75,9 casos por 100.000 habitantes, e o Rio de Janeiro, com 68,6 casos por 100.000 habitantes, apresentando os coeficientes mais elevados. E Manaus é a capital com maior coeficiente de incidência da Tb, com 115,8 casos a cada 100.000 habitantes. Esse cenário está distante das metas e compromissos globais e nacionais pelo fim da Tb até 2035. Portanto, abordagens relacionadas ao fortalecimento dos sistemas de saúde, inovação tecnológica e combate à pobreza são urgentes para a eliminação da Tb (MINISTÉRIO DA SAÚDE, 2023).

Tuberculose é uma doença infecciosa causada pelo agente *Mycobacterium tuberculosis*, conhecido como bacilo de Koch. Ela pode atingir uma série de órgãos ou sistemas, mas a forma pulmonar é mais frequente. A apresentação da Tb na forma pulmonar é a responsável pela transmissão da doença. Quando uma pessoa acometida pela doença tosse, gotículas com bacilos são liberadas no ambiente e podem ser inaladas por outras pessoas que podem contrair a doença. Na última década, foram desenvolvidos novos testes moleculares para o diagnóstico da Tb que são mais precisos, conhecidos como testes de amplificação de ácidos nucleicos (NAAT, do inglês *Nucleic Acid Amplification Tests*), no entanto, eles são mais caros (KHUTLANG et al., 2010). O Teste Rápido Molecular para Tuberculose (TRM-TB) é um teste de amplificação de

ácidos nucleicos capaz de detectar simultaneamente o *Mycobacterium tuberculosis* e a resistência a rifampicina, fármaco usado no tratamento da Tb. Ele foi incorporado ao Sistema Único de Saúde (SUS) em 2014 e é considerado rápido pois fornece o resultado em menos de uma hora e meia. O TRM-TB é indicado para o diagnóstico de novos casos de Tb pulmonar e extrapulmonar e triagem de resistência à rifampicina, porém, para o acompanhamento do tratamento da Tb, deve-se utilizar a baciloscopia (MINISTÉRIO DA SAÚDE, 2022a). Além disso, por ser um método com baixos custos de infraestrutura laboratorial, a baciloscopia de campo claro é o método diagnóstico mais utilizado em países de baixa e média renda (COSTA FILHO et al., 2015).

Independentemente do resultado da baciloscopia, para todos os casos de suspeita de Tb, sendo novos casos ou retratamentos, deve ser realizada a cultura de micobactérias. Esse é um método de elevada sensibilidade e especificidade no diagnóstico da Tb. Quando realizada com alta qualidade técnica, é capaz de detectar de 10 a 100 bacilos cultiváveis por mililitro de escarro, e é considerada o padrão-ouro para o diagnóstico da Tb (MINISTÉRIO DA SAÚDE, 2022b). A desvantagem desse método é a demora no resultado. No Brasil, os métodos automáticos disponíveis realizam a semeadura da amostra em meio líquido, no qual o tempo de resultado varia de 5 a 12 dias, quando positivo, e 42 dias, quando negativo (MINISTÉRIO DA SAÚDE, 2018).

A baciloscopia de campo claro consiste na preparação de esfregaços de amostras de escarro em lâminas que são coradas com os métodos de *Ziehl-Neelsen* (ZN) ou *Kinyoun*. Esse processo confere aos bacilos da tuberculose uma coloração vermelha e aos demais elementos da amostra uma coloração azul. Com um microscópio de campo claro, um técnico especialista visualiza as lâminas e contabiliza o número de bacilos em até 100 campos microscópicos. O diagnóstico da doença é feito de acordo com os cinco diagnósticos propostos pela Organização Mundial da Saúde (WHO, 1998). A quantidade de bacilos contabilizados é associada a um diagnóstico da doença. Geralmente, a baciloscopia é considerada um método de diagnóstico de baixa sensibilidade, quando comparada com a cultura, pois a sensibilidade depende de vários fatores como tipo de lesão, tipo e número de amostras coletadas, técnica de preparo do esfregaço, qualidade dos corantes, atenção e persistência do microscopista e qualidade do microscópio. O número mínimo de bacilos da Tb necessários para produzir um resultado positivo tem sido estimado entre 5000 e 10000 por mililitro (MINISTÉRIO DA SAÚDE, 2022b).

A contagem manual dos bacilos é uma tarefa trabalhosa que depende da experiência do microscopista (COSTA FILHO et al., 2015). Em razão disso, a automatização desta análise traz várias vantagens, como a redução substancial na carga de trabalhos dos médicos, melhorando

a sensibilidade do diagnóstico e aumentando o número de imagens que pode ser analisada pelo computador (AYAS; EKINCI, 2014).

Muitos estudos de visão computacional abordam o tema do diagnóstico microscópico da Tb. No entanto, limitam-se ao desenvolvimento de métodos de detecção de bacilos em imagens de campo microscópico e não à automação da baciloscopia. Os métodos desenvolvidos processam imagens de microscopia e utilizam diferentes técnicas computacionais para detectar os bacilos, como os classificadores de aprendizado de máquina tradicionais e os classificadores de aprendizado de máquina profundo, que são técnicas mais recentes. Os classificadores de aprendizado de máquina tradicionais são: classificador Bayes (KHUTLANG et al., 2010), regressão linear (KHUTLANG et al., 2010), análise discriminante quadrática (KHUTLANG et al., 2010), máquina de vetor de suporte (SVM, do inglês *Support Vector Machine*) (COSTA FILHO et al., 2015), floresta aleatória (RF, do inglês *Random Forest*) (AYAS; EKINCI, 2014), agrupamento *k-means* (RICO-GARCIA et al., 2015), *fuzzy-c-means-clustering* (MITHRA; EMMANUEL, 2017), árvores de decisão (SOANS et al., 2016), *k-nearest-neighbor* (kNN) (YOUSEFI et al., 2020), otimização por exame de partículas (SUGIRTHA; MURUGES, 2017), técnicas de processamento digital de imagens (ZHAI et al., 2010), perceptron multicamadas (MLP, do inglês *Multilayer Perceptron*) (RULANINGTYAS; SUKSMONO; MENGKO, 2011), segmentação *watershed* (SHAH et al., 2016), rede *Gaussian-fuzzy-neural* (MITHRA; SAM EMMANUEL, 2021). Os classificadores de aprendizado profundo usados para detecção de bacilos são: redes neurais convolucionais (CNN, do inglês *Convolutional Neural Network*) (LÓPEZ et al., 2017), (XIONG et al., 2018), (QUINN et al., 2016), (PANICKER et al., 2018), (DINESH JACKSON SAMUEL; RAJESH KANNA, 2019), (SERRAO et al., 2020), (SWETHA; SANKARAGOMATHI; THANGAMALAR, 2020) e *deep belief neural networks* (MITHRA; SAM EMMANUEL, 2019).

Técnicas tradicionais de aprendizado de máquina permitem que o computador discrimine classes de diferentes padrões por meio do treinamento de uma máquina, sem que esta seja programada explicitamente para esse fim. O treinamento possibilita a classificação de novos dados ainda não visualizados, habilidade conhecida como generalização. As técnicas de aprendizado de máquina dividem-se em clássicas, ou rasas, e profundas. A principal característica das técnicas de aprendizado clássica é utilizarem estruturas com poucas camadas para realizar as tarefas de classificação ou regressão. Já as técnicas de aprendizado profundo utilizam muitas camadas para realizarem as mesmas tarefas. As técnicas de aprendizado de máquina profundo são biologicamente inspiradas no córtex visual dos mamíferos e são mais usadas para classificação de imagens e detecção de objetos. Nas técnicas clássicas de

aprendizado de máquina, a entrada é um conjunto de características extraídas das imagens pelo operador, enquanto nas técnicas de aprendizado profundo a entrada é a própria imagem. Nas redes profundas, as camadas iniciais da rede são as responsáveis pela extração de características, que são aprendidas através do método de propagação reversa.

Na literatura, é observado que muitos métodos têm sido desenvolvidos para identificar e contar bacilos da Tb, no entanto, nenhum trabalho relatou o diagnóstico da Tb conforme as diretrizes da baciloscopia de escarro mostradas no Quadro 1. A principal limitação dos trabalhos da literatura é a falta de banco de dados que permita o diagnóstico da Tb com baciloscopia de escarro.

Quadro 1. Descrição das classes diagnósticas da Tb em termos de contagem de bacilos encontrados em campos microscópicos.

<b>Classe Diagnóstica</b>	<b>Contagem de bacilos em campos microscópicos</b>
Negativo	Nenhum bacilo na leitura de 100 campos microscópicos
Esparso	Contagem entre 1 e 9 bacilos na leitura de 100 campos microscópicos
1 +	Contagem entre 10 e 99 bacilos na leitura de 100 campos microscópicos
2 +	Contagem média entre 1 e 10 bacilos por campo na leitura de 50 campos microscópicos
3 +	Contagem média de maior que 10 bacilos por campo na leitura de 20 campos microscópicos

Fonte: Adaptado de (RIEDER et al., 2007).

Métodos automáticos de detecção de patógenos por microscopia geralmente resultam em altas taxas de sensibilidade, indicando que o método classifica corretamente objetos de interesse. Entretanto, a desvantagem desses métodos é o alto número de falsos positivos detectados (AZIF; NUGROHO; WIBIRAMA, 2018). Para o diagnóstico automático da Tb, é necessário contar o número de bacilos em um conjunto de campos digitais que correspondem a 100 campos microscópicos. Conforme descrito nas seções seguintes, detectar um único falso positivo é suficiente para comprometer o diagnóstico negativo da Tb, que só ocorre quando nenhum bacilo é detectado em 100 campos microscópicos.

Anteriormente, desenvolvemos um método de detecção de bacilo em imagens-mosaico de esfregaço usando uma abordagem de aprendizado profundo (SERRAO et al., 2020). Este estudo pretende dar continuidade ao trabalho anterior, visando principalmente minimizar os falsos positivos na contagem automática de bacilos e, por conseguinte, aumentar a especificidade, possibilitando o diagnóstico automático da Tb. As duas principais etapas do método proposto

para o diagnóstico automático são: 1) Utilizar a segmentação semântica para a segmentação de objetos candidatos a bacilo; 2) Usar filtros para reduzir os falsos positivos. Na etapa 1, propomos uma melhoria da segmentação semântica proposta em SERRAO et al. (2020), através da utilização de mecanismo de atenção e convolução separada por canal. Na etapa 2, exploramos duas propriedades fundamentais do bacilo da Tb que ajudam a diferenciá-lo do background e de artefatos, a cor e forma. Baseado nessas duas propriedades, propomos a utilização de filtros de cor e filtro de forma. Com o método de detecção de bacilos estabelecido através das etapas anteriormente listadas, prosseguimos para a construção de uma base de dados que permitisse a avaliação do referido método para o diagnóstico da Tb, conforme as diretrizes estabelecidas pela Organização Mundial da Saúde.

### **1.1 Objetivo Geral**

Otimizar um método para o diagnóstico automático da tuberculose em imagens de baciloscopia de campo claro utilizando segmentação semântica e técnicas para minimização de falsos positivos.

### **1.2 Objetivos Específicos**

1. Aprimorar um método para segmentação semântica de Bacilos Álcool-Ácido Resistentes (BAAR) em imagens de baciloscopia de campo claro, utilizando *autoencoders*, através da utilização de mecanismos de atenção e convolução separada por canal.
2. Avaliar o desempenho de filtros baseados nas propriedades de cor e forma do bacilo para redução dos falsos positivos, resultantes da segmentação semântica de bacilos da tuberculose através de *autoencoders*, em imagens de baciloscopia de campo claro.
3. Construir uma base de dados para o diagnóstico automático da tuberculose, de acordo com os critérios estabelecidos em (BRASIL, 2008).
4. Avaliar um método automático de diagnóstico da tuberculose por meio de imagens de baciloscopia de campo claro, de acordo com os critérios estabelecidos em (BRASIL, 2008).

### **1.3 Organização do Trabalho**

Este trabalho está organizado conforme a divisão descrita a seguir:

- Capítulo 1: Introdução;

- Capítulo 2: Revisão Bibliográfica;
- Capítulo 3: Referencial Teórico;
- Capítulo 4: Materiais e Métodos;
- Capítulo 5: Resultados e discussão;

O Capítulo 1 apresentou um contexto do trabalho, citou a problemática da doença da Tb, a importância do controle dessa doença e qual o objetivo geral e objetivos específicos deste trabalho.

O Capítulo 2 apresenta trabalhos que foram publicados na literatura na área de reconhecimento de bacilos em imagens de baciloscopia de campo claro por meio de técnicas de aprendizado de máquina tradicional e aprendizado de máquina profundo que utilizam apenas uma etapa ou mais para a detecção do bacilo.

O Capítulo 3 apresenta os principais fundamentos teóricos utilizados no desenvolvimento deste trabalho: baciloscopia de campo claro, planos de cores de imagens, árvores de decisão, redes neurais artificiais, redes neurais convolucionais e seu desenvolvimento na segmentação de imagens.

No Capítulo 4 são apresentados os materiais utilizados no desenvolvimento da tese, as características do banco de dados e a metodologia proposta para o desenvolvimento das etapas de segmentação semântica e filtragem de falsos positivos.

Por fim, no Capítulo 5 são apresentados os resultados alcançados, discussão e conclusão do trabalho.

## 2 REVISÃO BIBLIOGRÁFICA

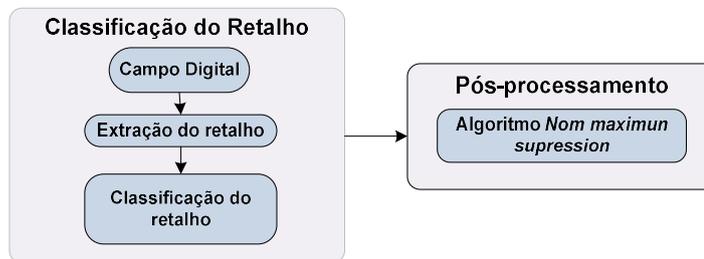
Neste capítulo será apresentada a revisão bibliográfica realizada destacando as principais técnicas descritas na literatura no que diz respeito à detecção de *Mycobacterium tuberculosis* em imagens de baciloscopia de campo claro, coradas com as técnicas de *Ziehl-Neelsen* ou *Kinyoun* (BRASIL, 2008), que são técnicas similares, pois usam as mesmas soluções de coloração, diferenciando-se esta última apenas por ser uma técnica a frio e usa diferentes concentrações de soluções de coloração. Buscou-se citar os diferentes métodos de identificação do bacilo da Tb, que pode ser em uma única etapa, em duas etapas ou utilizando técnicas de detecção mais avançadas. Além disso, foram identificadas as técnicas de aprendizado de máquina empregadas, características do banco de dados utilizados e os desempenhos alcançados pelos autores.

### 2.1 Detecção de Bacilos em Microscopia de Campo Claro

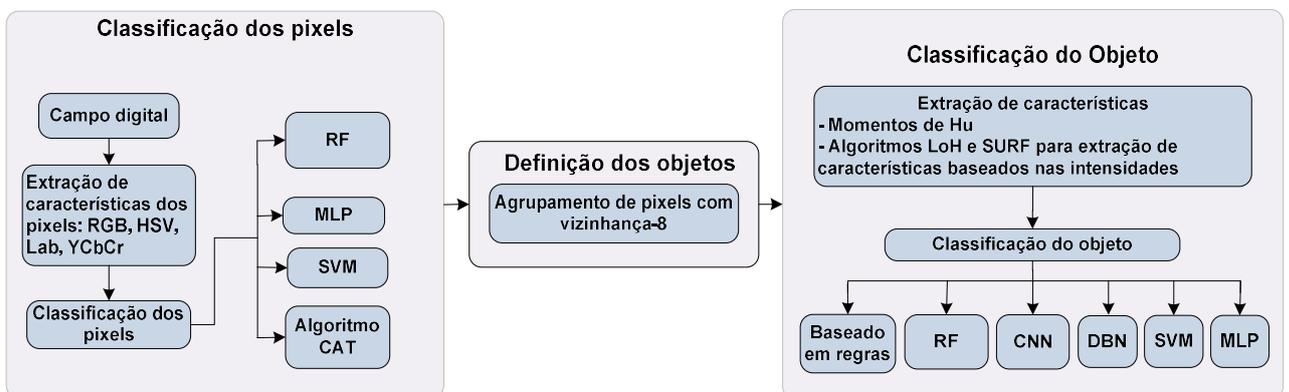
Esta seção apresenta as principais técnicas descritas na literatura para detecção do *Mycobacterium tuberculosis* em imagens de microscopia de campo claro de amostras coradas com as técnicas *Ziehl-Neelsen* ou *Kinyoun*. Os trabalhos publicados sobre a identificação dos bacilos da Tb em imagens de baciloscopia de campo claro podem ser divididos em três grupos. O primeiro grupo caracteriza-se por apresentar apenas uma etapa para a identificação dos bacilos: uma etapa de classificação do retalho, que classifica o retalho como positivo (retalho com bacilo) ou negativo (retalho sem bacilo). A metodologia consiste na extração de um retalho do campo digital seguida pela classificação desse retalho utilizando métodos de aprendizado de máquina profundo. É necessário um pós-processamento para eliminar as classificações sobrepostas. O segundo grupo caracteriza-se por apresentar duas etapas para a identificação do bacilo: classificação dos pixels e classificação do objeto. A metodologia consiste na primeira etapa responsável pela extração de características dos pixels do campo digital em diferentes espaços de cor e classificá-los com métodos de aprendizado de máquina. A segunda etapa é a classificação dos objetos formados na primeira etapa, a partir da extração de características deles. O terceiro grupo é caracterizado por utilizar redes neurais especializadas na detecção de objetos, como as redes *Faster Region-based Convolution Neural Network (Faster-RCNN)*. A metodologia consiste em duas etapas: uma para detecção de bacilos com redes especializadas em detecção de objetos e outra para filtragem de falsos positivos por meio de um modelo de aprendizado de máquina profundo. O quarto grupo é caracterizado por utilizar redes desenvolvidas para realizar a segmentação dos bacilos em imagens-mosaico. A Figura 1 ilustra

a metodologia utilizada para detecção de bacilos nos artigos revisados: em (a), (b), (c) e (d) mostra a metodologia dos grupos 1, 2, 3, e 4, respectivamente.

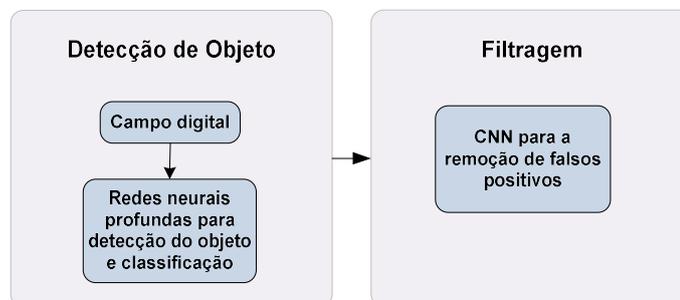
Figura 1. Métodos de detecção de bacilos usados nos artigos revisados.



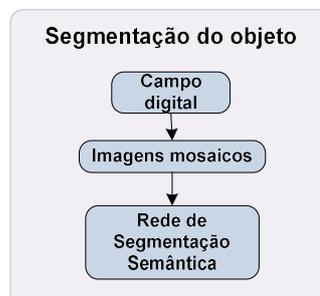
(a) Metodologia utilizada no grupo 1.



(b) Metodologia utilizada no grupo 2.



(c) Metodologia utilizada no grupo 3.



(d) Metodologia utilizada no grupo 4.

### 2.1.1 Identificação do Bacilo com uma Etapa

Em relação aos artigos do primeiro grupo, em QUINN et al. (2016), a classificação do retalho foi realizada com uma CNN. Os autores implementaram um método para três diferentes patógenos: malária em amostras de esfregaço de sangue, bacilos da tuberculose em amostras de escarro e parasitas intestinais em amostras de fezes. Cada imagem coletada foi reduzida e depois dividida em retalhos sobrepostos, com o fator de redução e o tamanho do retalho determinados pelo tipo de patógeno a ser reconhecido em cada caso. De acordo com os autores, como a maior parte de cada imagem não contém objetos de interesse, os patógenos, o número de retalhos negativos foi desproporcionalmente grande em comparação com o número de retalhos positivos. Duas ações foram tomadas para obter um conjunto de dados balanceado. Primeiro, retalhos negativos foram descartados aleatoriamente para que houvesse no máximo 100 vezes o número de retalhos positivos. A segunda ação foi o aumento de dados aplicado em retalhos positivos, girando e invertendo os retalhos, resultando em 7 retalhos positivos extras para cada retalho original. Uma rede com quatro camadas ocultas (camada de convolução, camada *pooling*, camada de convolução e camada totalmente conectada) foi usada para detectar os três tipos de patógenos. Uma divisão 50/50 para o treinamento/teste foi usada e o treinamento foi executado em 500 épocas. Em relação à tuberculose, o banco de dados foi composto por 630.284 retalhos (9% positivos e 91% negativos). Os autores usaram o algoritmo *non-maximum suppression* para remover os muitos bacilos detectados em duplicidade que ocorreram nas bordas dos retalhos. Para a identificação dos bacilos, a rede alcançou uma área sob a curva ROC (AUC) de 99% e uma acurácia média de 93%.

Em LÓPEZ et al. (2017), os autores também usaram uma CNN para a classificação de retalhos. O banco de dados consistia em 29.310 retalhos de 40×40 pixels, foi completamente balanceado e dividido em 60% para treinamento, 20% para validação e 20% para teste. Para obter melhores resultados, os autores usaram a normalização de intensidade dos retalhos. Três CNNs foram avaliadas: a primeira com uma camada de convolução, a segunda com duas camadas de convolução e a terceira com três camadas de convolução. Além disso, três entradas diferentes foram avaliadas: retalhos RGB, retalhos em tons de cinza e retalhos de intensidade obtidos a partir da subtração dos canais R e G (R-G). Os autores utilizaram o método de regularização *dropout* no treinamento da CNNs. O melhor modelo, aquele com três camadas de convolução treinado com retalhos RGB, foi avaliado em imagens de baciloscopia completa. A busca em toda a imagem utiliza uma janela deslizante de 40×40 pixels. Cada região de 40×40 pixels é classificada como positiva ou negativa. Os autores usaram o algoritmo *non-maximum suppression* para excluir os muitos bacilos sobrepostos que ocorrem nas bordas do retalho. Para

a classificação do retalho como positivo ou negativo, obteve-se um valor de AUC de 99% e uma precisão superior a 98%. Na detecção de bacilos em toda a imagem, foram obtidos os seguintes resultados: uma precisão de 73,68%, um *recall* de 92,63% e um *F1-score* de 80,07%.

Em XIONG et al. (2018), os autores usaram a abordagem de transferência de aprendizado baseada em CNNs, usando o banco de dados CIFAR-10, que foi desenvolvido para imagens pequenas, para classificar um retalho como positivo ou negativo. A entrada da rede foi de 32×32 pixels de regiões positivas e negativas. O banco de dados consistia em 578.191 retalhos positivos e 2.510.307 retalhos negativos. Os autores afirmam que o principal desafio era que os objetos de interesse eram extremamente pequenos em comparação com a imagem da lâmina: com ampliação de 40×, o tamanho de um bacilo da Tb era de apenas cerca de 20×4 pixels. O modelo de classificação foi treinado para discriminar entre os retalhos negativos e positivos. O algoritmo do gradiente descendente estocástico foi usado como método de otimização com tamanho do lote de 512. A taxa de aprendizado inicial foi definida como 0,05 e foi multiplicada por 0,1 a cada 5 épocas. O modelo foi treinado por 25 épocas no total. Os autores não usaram o algoritmo *non-maximum suppression* para excluir os bacilos sobrepostos que ocorrem na borda. Os autores obtiveram os seguintes resultados: sensibilidade de 97,94% e especificidade de 83,65%.

Em IBRAHIM et al. (2021), os autores, em vez de classificar retalhos como positivo ou negativo, classificam imagens de lâminas em infectada ou saudável usando transferência de aprendizado do modelo AlexNet. O banco de dados consiste em 178 imagens de 4416×3312 pixels utilizados no experimento A, 524 imagens de 3456×4608 pixels utilizados no experimento B e 530 imagens de 3456×4608 utilizadas no experimento C. Em todos os experimentos, foi utilizada a proporção 70/30 para treinamento e teste, e as imagens foram redimensionadas para o tamanho 227×227×3 pixels. Após o treinamento, o modelo foi testado em imagens de lâminas classificadas por um patologista. O melhor resultado encontrado foi de acurácia de 98,73%, sensibilidade de 98,59% e especificidade de 98,84% para o experimento C.

### **2.1.2 Identificação do Bacilo com duas Etapas**

Em relação aos artigos do segundo grupo, em AYAS; EKINCI (2014), os autores utilizaram um algoritmo supervisionado para classificação de pixels. Usando um classificador *Random Forest* (RF), os pixels primeiramente são classificados como pertencentes a bacilo ou não. Componentes de cores RGB foram usados como entrada do classificador RF. Os pixels

próximos rotulados como pertencentes a um bacilo foram agrupados em objetos que são rotacionados, redimensionados e posicionados no centro de caixas delimitadoras de 30×30 pixels, que são usadas para a classificação do objeto como bacilo ou não. Para a classificação de objetos, os autores usaram outro classificador RF e momentos de Hu como variáveis de entrada. O banco de dados utilizados pelos autores incluiu 116 imagens com tamanho de 640×480 pixels, divididas em 40 imagens para treinamento e 76 imagens para teste. Os resultados obtidos para a identificação do bacilo foram: acurácia de 67,98%, sensibilidade de 89,34% e especificidade de 62,89%.

Em COSTA FILHO et al. (2015), para a classificação dos pixels, foram extraídas 30 características dos espaços de cor RGB, HSI, YCbCr e Lab. Dois classificadores foram comparados: MLP e SVM. Os pixels classificados foram então agrupados em objetos. Na sequência, foram utilizados 3 tipos de filtros para separar os objetos correspondentes a bacilos de ruídos: um filtro de tamanho, para remover objetos com áreas muito grandes ou muito pequenas, um filtro geométrico que elimina objetos com base em sua excentricidade e um filtro baseado em regras, que usa um parâmetro de proporção de cores no espaço de cor RGB. Os autores utilizaram um banco de dados com 120 imagens de 2816×2112 pixels. Os melhores resultados foram obtidos com o classificador SVM: sensibilidade de 96,80% e acurácia de 96,62%.

Em MITHRA; SAM EMMANUEL (2019), os autores usaram o algoritmo *channel area thresholding* (CAT) na etapa de classificação de pixels, que é um método de segmentação multinível baseada em canal. Um limiar para o bacilo com base na intensidade do canal verde da imagem RGB e limiares para a execução sucessiva do algoritmo são definidos e utilizados para realizar a segmentação dos objetos de interesse. Em seguida, eles extraíram características dos objetos segmentados usando o algoritmo histograma orientado a localização (LoH, do inglês *Location-oriented Histogram*) e características robustas aceleradas (SURF, do inglês *Speeded-Up Robust Features*). Um modelo *Deep Belief Network* (DBN) é treinado para classificar objetos como bacilos ou não. A partir da contagem dos bacilos, se o número de bacilos em uma imagem for menor ou igual a 5, ela é classificada como baixo nível de infecção. Se o número de bacilos em uma imagem for maior que 5, ela é classificada como alto nível de infecção. O sistema foi treinado com três categorias diferentes de imagens tiradas do banco de dados ZNSM-iDB, com nenhum bacilo, poucos bacilos e bacilos sobrepostos em cerca de 275 imagens. O sistema obteve os seguintes resultados: acurácia de 97,55%, sensibilidade de 97,86% e especificidade de 98,23%.

Em SWETHA; SANKARAGOMATHI; THANGAMALAR (2020), os autores também utilizaram o algoritmo CAT na etapa de classificação de pixels. Na etapa de classificação de objetos, eles extraíram características com o algoritmo SURF combinado com um método descritor de características de objetos, o histograma de gradientes orientados (HOG, do inglês *Histogram of Oriented Gradients*). Para classificar os objetos, os autores utilizaram uma rede CNN que combina em sua entrada a imagem original da baciloscopia de escarro com as características extraídas pelos algoritmos SURF e HOG. A gravidade da Tb é classificada em leve, moderada e grave. O método obteve os seguintes resultados: acurácia de 99,40%, sensibilidade de 94,74% e especificidade de 99,41%.

### 2.1.3 Identificação do Bacilo com Detecção de Objetos

Em relação aos artigos do terceiro grupo, em EL-MELEGY et al. (2019), os autores usaram um método de aprendizado profundo de última geração, o modelo *Faster-RCNN* que identifica os bacilos da tuberculose com caixas delimitadoras. Os autores adicionaram uma segunda etapa na metodologia proposta, que consiste em uma CNN que classifica as caixas delimitadoras em bacilos da Tb verdadeiros positivos ou falsos positivos, com o objetivo de realizar uma filtragem nos objetos detectados. A segunda etapa foi necessária, pois os autores identificaram uma alta taxa de falsos positivos nos experimentos com a *Faster-RCNN*. A CNN reduziu a taxa de falsos positivos em 20%. O banco de dados utilizado consistia em 500 imagens de tamanho 400×400 pixels contendo um total de 2500 bacilos. Os dados foram separados em 80% para treinamento e 20% para teste. O modelo CNN foi treinado com entradas de 20×20 pixels e com o mesmo conjunto de dados, no entanto, caixas delimitadoras de falsos positivos foram consideradas para as amostras de treinamento da classe não-Tb e caixas delimitadoras de verdadeiro positivos foram consideradas para as amostras de treinamento da classe Tb. O método obteve o seguinte resultado: *recall* de 98,4%, precisão de 85,1% e *F-score* de 91,2%. O baixo valor obtido para a precisão indica que os autores não conseguiram reduzir significativamente a quantidade de falsos positivos.

Em KUOK et al. (2019), os autores propuseram também um sistema de detecção de bacilos em dois estágios: identificação dos bacilos utilizando uma *Faster-RCNN* e classificação do objeto detectado como bacilo com uma CNN. Um mecanismo de refinamento que realimenta a caixa delimitadora de saída na entrada da *Faster-RCNN* foi usado com a finalidade de obter caixas delimitadoras mais precisas. O banco de dados consiste em 19.234 imagens de 1280×960 pixels. A entrada da CNN no segundo estágio é um retalho de 48×48 pixels O

sistema alcançou uma precisão de cerca de 86%. Os autores citam que muitos bacilos candidatos tornaram a tarefa difícil, e uma CNN mais poderosa deve ser implementada para melhorar as falsas detecções.

#### **2.1.4 Identificação do Bacilo com Segmentação Semântica**

Em relação ao artigo do quarto grupo, o trabalho desenvolvido anteriormente (SERRAO et al., 2020) utilizou rede de segmentação semântica para segmentar os bacilos em imagens-mosaico. As imagens-mosaico são compostas por retalhos positivos (contendo bacilos) e retalhos negativos (sem bacilos). Em média, uma imagem-mosaico contém 50% retalhos positivos e 50% retalhos negativos. Três arquiteturas de CNN foram avaliadas para realizar a segmentação de objetos de interesse, os bacilos. Após a segmentação, uma etapa com filtro de área foi aplicada para remover ruídos. O objetivo desse filtro foi descartar objetos pequenos e classificar como bacilo apenas objetos com área maior que 25 pixels. Os melhores resultados obtidos para a detecção de bacilos foi uma acurácia de 99,66%, precisão de 99,35%, sensibilidade de 99,99%, especificidade de 99,34% e *F1-score* de 99,67%.

Considerando a Taxa de Falsos Positivos (1-especificidade) de 0,66% e o fato de cada imagem-mosaico ter uma média de 50 retalhos negativos, a quantidade de Falsos Positivos (FP) é de 0,33 por imagem-mosaico, que resulta em uma média de 4,86 FP por campo microscópico. Quando aplicado em um conjunto de 100 campos microscópicos de um paciente com diagnóstico negativo, resulta em 486 FP. De acordo com o Quadro 1, na Seção 1, esse número de FP poderia ser classificado como diagnóstico 2+, o segundo mais severo diagnóstico da Tb. Portanto, é necessária uma especificidade bem superior àquela relatada acima para obter um diagnóstico automático confiável da Tb com a baciloscopia de campo claro.

## **2.2 Conclusão**

Resumindo os artigos revisados, os métodos para identificação de bacilos do grupo 1, precisam de alta carga computacional, devido o algoritmo *non-maximum suppression*. Os métodos do grupo 2 envolvem agrupamento de pixels vizinhos para definir um bacilo e podem gerar falsos negativos, pois as regiões de um bacilo podem ser muito heterogêneas, gerando descontinuidade quando detectadas pelo método de detecção de pixels. Assim, ao invés de detectar um bacilo, são detectados vários fragmentos, que podem ser confundidos com ruídos. Além disso, a análise pixel a pixel em larga escala pode ser demorada.

Os métodos do grupo 3 envolvem o uso de redes neurais desenvolvidas para a detecção de objetos grandes e, portanto, não são otimizadas para detecção de bacilos. Dois trabalhos foram encontrados na literatura que utilizam a rede *Faster R-CNN* desenvolvida para detecção de objetos, estes apresentaram uma alta ocorrência de falsos positivos, e conseqüentemente um valor de precisão muito baixo, o que atesta essa inadequação.

O único artigo do grupo 4, que utiliza segmentação semântica com imagens-mosaico, relatou especificidade de 99,34%.

Em geral, os artigos revisados não propuseram o diagnóstico da Tb em imagens de baciloscopia de campo claro. Eles ainda focam apenas na detecção de bacilos. Em SWETHA; SANKARAGOMATHI; THANGAMALAR (2020) e MITHRA; SAM EMMANUEL (2019), os autores propuseram apenas um método para avaliar o nível de infecção por Tb em imagens de baciloscopia, classificando em categorias como baixo, médio e alto nível de infecção. E em IBRAHIM et al. (2021), os autores classificam imagens de lâminas em infectada ou saudável.

A principal deficiência identificada nos artigos revisados anteriormente, que impedem o diagnóstico da tuberculose pela baciloscopia de campo claro, é que os valores de especificidade para detecção de bacilos estão abaixo de 99,34%, o que leva a uma alta taxa de falsos positivos quando é feita uma análise na quantidade de campos digitais necessários para o diagnóstico da Tb.

Os artigos revisados não utilizaram convolução separada por canal ou mecanismos de atenção de canal para segmentação de bacilos. Visando melhorar a especificidade, neste trabalho avaliamos o uso destas técnicas para melhor explorar canais de cores na segmentação de bacilos.

O método de identificação de bacilos utilizado neste trabalho, difere dos métodos descritos anteriormente, no seguinte sentido:

- a) A segmentação dos bacilos, apresentada em SERRAO et al. (2020), é feita por segmentação semântica. Em que foi avaliada a convolução separada por canal e mecanismo de atenção por canal para a segmentação de bacilos. Isso só foi possível porque a área dos bacilos foi anotada manualmente nas imagens do banco de dados usadas para o treinamento. Dessa forma, os pixels do bacilo são marcados com nível lógico 1, enquanto o fundo é marcado com nível lógico 0.
- b) Após a segmentação aplicamos dois filtros especializados. Esses filtros focam em características específicas dos bacilos, que ajudam a diferenciá-los dos ruídos: cor e forma.

O processo de anotação do bacilo utilizado neste estudo permite um melhor aprendizado do classificador no sentido de diferenciar os pixels de um bacilo dos pixels do fundo. O processo de filtragem permite a eliminação progressiva de falsos positivos, avaliando as diferentes características. E a elaboração de um banco de dados de 50 conjuntos de cada um dos cinco diagnósticos mostrados no Quadro 1, na Seção 1, permitirá a avaliação do método no diagnóstico da tuberculose. O Quadro 2 mostra um resumo dos trabalhos apresentados nessa seção.

Quadro 2. Resumo da análise realizada nas Revisão Bibliográfica.

Referência	Materiais	Métodos				Resultados
		Primeiro Estágio		Segundo Estágio		
		Entrada	Identificação do Objeto	Entrada	Classificação do Objeto	
(QUINN et al., 2016)	928 imagens			Retalhos RGB do tamanho aproximado de um bacilo	CNN + <i>non-maximum supression</i>	AUC: 99% Acurácia média: 93%
(LÓPEZ et al., 2017)	492 imagens de 1388×1040 pixels			Retalhos RGB de 40×40 pixels	CNN + <i>non-maximum supression</i>	AUC: 99% Precisão: 98%
(XIONG et al., 2018)	246 imagens			Retalhos RGB de 32×32 pixels	CNN baseado no modelo CIFAR-10 + <i>non-maximum supression</i>	Sensibilidade: 97,94% Especificidade: 83,65%
(IBRAHIM et al., 2021)	178 imagens de 4416×3312 pixels; 1054 imagens de 3456×4608			Imagem RGB de 227×227 pixels	AlexNet	Acurácia: 98,73% Sensibilidade: 98,59% Especificidade: 98,84%
(AYAS; EKINCI, 2014)	116 imagens de 640×480 pixels	Características do plano RGB de cada pixel	<i>Random Forest</i>	Momentos de Hu dos objetos formados pelo agrupamento dos pixels	<i>Random Forest</i>	Acurácia: 67,98% Sensibilidade: 89,34% Especificidade: 62,89%
(COSTA FILHO et al., 2015)	120 imagens de 2816×2112 pixels	30 características dos planos RGB, HSI, YCbCr e Lab de cada pixel	SVM	Características de área, excentricidade e intensidades RGB	Filtro de área + Filtro geométrico + Filtro de regras	Acurácia: 96,62% Sensibilidade: 96,80%

(MITHRA; SAM EMMANUEL, 2019)	275 imagens	Intensidades no canal verde da imagem	CAT	Extração de características dos objetos segmentados com os algoritmos LoH e SURF	<i>Deep Belief Network</i>	Acurácia: 97,55% Sensibilidade: 97,86% Especificidade: 98,23%
(SWETHA; SANKARAGOMATHI ; THANGAMALAR, 2020)		Intensidades no canal verde da imagem	CAT	Imagem original + características extraídas com os algoritmos SURF e HOG	CNN	Acurácia: 99,40% Sensibilidade: 94,74% Especificidade: 99,41%.
(EL-MELEGY et al., 2019)	500 imagens de 400×400 pixels	Imagem RGB de 400×400 pixels	<i>Faster-RCNN</i>	Retalho de 20×20 pixels	CNN	<i>Recall</i> : 98,4% Precisão: 85,1% <i>F-score</i> : 91,2%
(KUOK et al., 2019)	19.234 imagens de 1280×960 pixels	Imagem RGB de 1280×960 pixels	<i>Faster-RCNN</i> refinada	Retalho de 48×48 pixels	CNN	Precisão: 86%
(SERRAO et al., 2020)	5000 imagens-mosaico de 400×400 pixels	Imagem RGB de 400×400 pixels	CNN			Acurácia: 99,66% Precisão: 99,35% Sensibilidade: 99,99% Especificidade: 99,34% <i>F1-score</i> : 99,67%

### 3 REFERENCIAL TEÓRICO

Neste capítulo serão abordados os principais conceitos necessários para o entendimento deste trabalho. A Seção 3.1 apresenta as etapas do exame de baciloscopia de campo claro. A Seção 3.2 aborda os conceitos básicos de planos de cores de imagens. A Seção 3.3, introduz sobre aprendizado de máquina, e nas seções seguintes são abordados os modelos utilizados na implementação do trabalho: árvores de decisão na Seção 3.4, redes neurais artificiais na Seção 3.5 e redes neurais convolucionais na Seção 3.6.

#### 3.1 Baciloscopia de Campo Claro

Nas últimas décadas houve um progresso notável na qualidade dos serviços de diagnóstico citobacteriológico e na velocidade desses diagnósticos nos países industrializados. Países industrializados onde a prevalência da doença é baixa ou países onde a doença é endêmica possuem necessidades de diagnóstico diferentes. Portanto, é importante considerar o contexto epidemiológico ao implementar estratégias de diagnóstico para a Tb. Na maior parte do mundo, onde a Tb é um grande problema de saúde pública, o progresso no diagnóstico ainda não foi alcançado e o exame de baciloscopia continua sendo necessário. Dessa forma, diferentes ferramentas são necessárias para diagnosticar a Tb (STEINGART et al., 2006)

De acordo com RIEDER et al. (2007), a sensibilidade (proporção de casos identificados) da baciloscopia de campo claro no diagnóstico da Tb está longe de ser perfeita. Entretanto, a sensibilidade na identificação de transmissores do bacilo da Tb é superior a 80%. Considerando que pacientes com baciloscopia positiva têm muito mais chances de transmitir o agente do que pacientes negativos, a baciloscopia de escarro é uma das ferramentas mais eficazes e eficientes para a detecção de casos em um programa nacional de controle da tuberculose, identificando os casos que devem ter prioridade máxima no controle da Tb.

A baciloscopia inclui a coleta do escarro, preparo e coloração do esfregaço, leitura do esfregaço e registro do número de bacilos encontrados. A lâmina de esfregaço é feita a partir do escarro do paciente, que é seco, fixado e corado usando técnicas de coloração padrão, como as técnicas *Ziehl-Neelsen* ou *Kinyoun* (LUMB et al., 2013; RIEDER et al., 2007). Ambas as técnicas usam as mesmas soluções de coloração, diferenciando-se que a última é a frio e usa diferentes concentrações de soluções de coloração. A coloração do esfregaço permite a identificação dos bacilos da Tb que são anotados e contabilizados para o diagnóstico. As seções

seguintes detalham o método de preparação do esfregaço, método de coloração do esfregaço e método de leitura e interpretação dos resultados.

### 3.1.1 Método de Preparação do Esfregaço

De acordo com o Manual de Recomendações para o Diagnóstico Laboratorial de Tuberculose e Micobactérias não Tuberculosas de Interesse em Saúde Pública no Brasil (MINISTÉRIO DA SAÚDE, 2022b), deve ser coletada uma amostra de escarro espontâneo que é distendida diretamente sobre a lâmina. É indicado o uso da porção mais purulenta do escarro para evitar resultados falso negativos e utilizar lâminas sem prévio uso, para evitar resultados falsos positivos. Cuidados como utilizar a porção mais purulenta do escarro para evitar falsos negativos, utilizar lâminas sem prévio uso para evitar falsos positivos e manter a área de trabalho com iluminação adequada para evitar troca de amostras são precauções importantes que o técnico de laboratório deve ter para uma correta preparação das lâminas.

O processo de preparação é descrito a seguir. Primeiramente, o técnico deve retirar a partícula maior e mais purulenta da amostra com o auxílio de palitos, e depositar em uma lâmina. Em seguida, deve distender a amostra até o extremo oposto da lâmina com o palito na horizontal e espalhar a amostra até que cubra 2/3 da lâmina, sem deixar espaços vazios, até obter um esfregaço homogêneo. As lâminas são colocadas para secar em temperatura ambiente até ficarem completamente secas. Por fim, a lâmina com o esfregaço voltado para cima é passada rapidamente por três vezes sobre a chama do bico de Bunsen, para a fixação do esfregaço. Na Figura 2 mostra-se o fluxo da preparação do esfregaço: em (a) é apresentada uma amostra de escarro sendo colocada sob a lâmina; em (b) é apresentada a amostra sendo distendida na lâmina; e (c) apresenta uma amostra no processo de fixação na lâmina, depois de completamente seca.

Figura 2. Preparação do esfregaço.



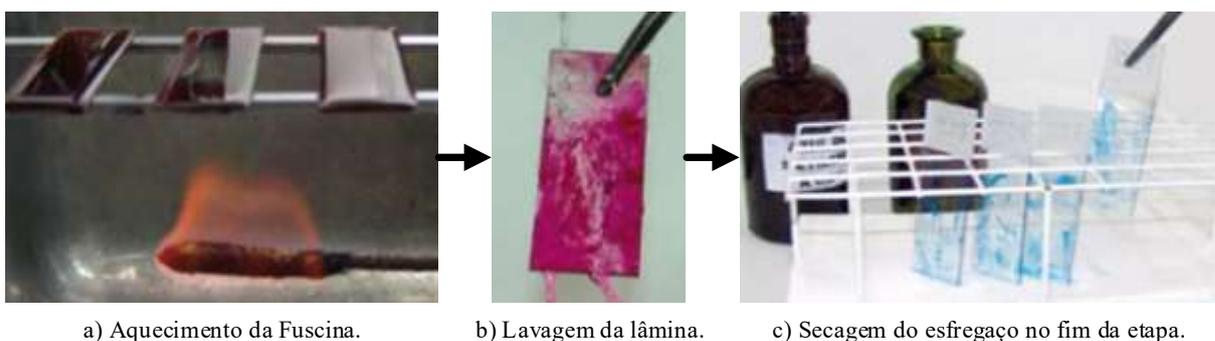
Fonte: Adaptado de (BRASIL, 2008).

### 3.1.2 Método de Coloração do Esfregaço

Os bacilos são denominados Bacilos Álcool-Ácido Resistentes (BAAR) porque possuem em sua parede celular o ácido micólico. Eles são corados de vermelho ao entrar em contato com o corante fucsina fenicada e persistem na coloração ao entrarem em contato com soluções álcool-ácidas utilizadas para descoloração. Assim, o método de coloração do esfregaço é realizado para que somente os bacilos da Tb assumam uma coloração vermelha, em contraste com o restante da amostra que assume uma coloração azul, diferenciando-os na amostra e a sua visualização é facilitada. Vários métodos de coloração têm sido desenvolvidos, mas a padronização utilizando um único método em todo país é crucial. O método de coloração *Ziehl-Neelsen* é recomendado pelo Ministério da Saúde, uma vez que possibilita a identificação de BAAR e utiliza o microscópio ótico comum para a leitura do esfregaço.

No método *Ziehl-Neelsen*, o corante fucsina fenicada é aplicado no esfregaço e aquecido até que ocorra a emissão de vapores visíveis, sem ferver o corante. Esse processo deve ser realizado 3 vezes. Em seguida, a lâmina é lavada para remover o excesso de corante, e posteriormente é aplicada a solução descorante álcool-ácida, que remove o corante das bactérias não ácido resistentes. Após um minuto, a lâmina é lavada novamente para remover toda a solução descorante, e é coberta com a solução azul de metileno. Essa solução cora as outras bactérias e elementos da amostra que foram descorados pela solução álcool-ácido. Proporcionando o contraste necessário para visualizar o BAAR. Por fim, deve-se lavar a lâmina para eliminar o excesso da solução azul e colocá-la para secar. A Figura 3 representa alguns dos processos citados, como: a) o aquecimento da fucsina após sua aplicação; b) a lavagem da lâmina após o aquecimento da fucsina; c) a secagem das lâminas após ser aplicado o álcool-ácido e o azul de metileno. O método *Ziehl-Neelsen* difere do método *Kinyoun*, pois não é feito o aquecimento do corante fucsina. Nesse último, após a aplicação da fucsina nas lâminas já é feita a lavagem da lâmina com água corrente.

Figura 3. Coloração do esfregaço.



a) Aquecimento da Fucsina.

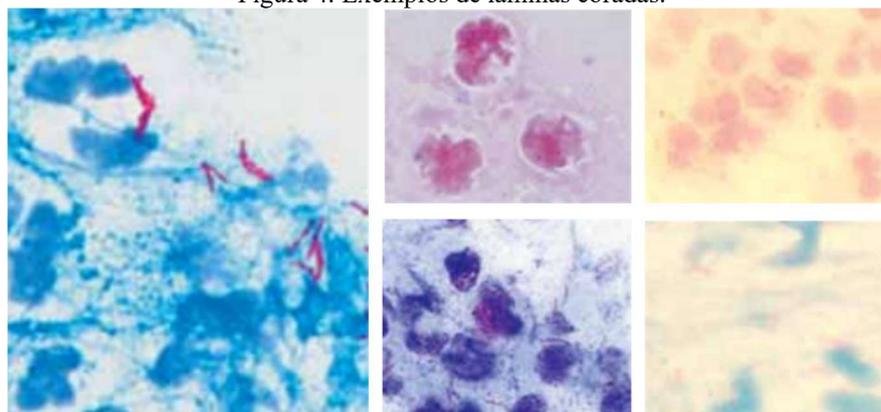
b) Lavagem da lâmina.

c) Secagem do esfregaço no fim da etapa.

Fonte: Adaptado de (BRASIL, 2008).

Durante o processo de coloração alguns problemas podem resultar em um esfregaço com coloração muito rosa e dificultar a identificação dos bacilos. Descolorização insuficiente, concentração do ácido muito baixa ou aplicado por curto tempo, deixar o corante secar no esfregaço e esfregaço muito espesso podem comprometer a coloração. A Figura 4 mostra em (a) um exemplo de lâmina que foi corretamente preparada e em (b) exemplos de lâminas com problemas de coloração.

Figura 4. Exemplos de lâminas coradas.



(a) Lâmina corada corretamente

(b) Lâminas com problemas de coloração

Fonte: Adaptado de (LUMB et al., 2013).

### 3.1.3 Método de Leitura e Interpretação dos Resultados

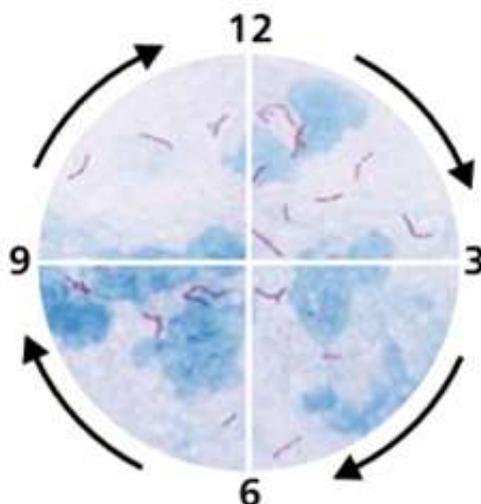
Considerando que o número de bacilos álcool-ácido resistentes indica o quão infectado o paciente está, várias escalas de classificação para o número de BAAR encontrados em um esfregaço foram propostas ao longo dos anos (LUMB et al., 2013; RIEDER et al., 2007; WHO, 1998). A escala da OMS propõe cinco diagnósticos para relatar os resultados da leitura de um esfregaço: Negativo, Esparso, 1+, 2+ e 3+. Conforme mostrado no Quadro 1, na Seção 1, cada escala de classificação corresponde à uma contagem de bacilos visualizados em campos microscópicos.

A leitura da baciloscopia é realizada utilizando um microscópio de campo claro, e é suficiente examinar uma extensão de um esfregaço de 20 mm, que corresponde a 100 campos microscópicos, para relatar um resultado negativo. O esfregaço é analisado da esquerda para a direita. Os bacilos da Tb aparecem em vermelho brilhante contra o material de fundo contrastado em azul. Eles são bastante variáveis em forma, de fragmentos curtos a tipos alongados. Eles podem ser uniformemente corados, podem ter lacunas ou podem até parecer granulares. Eles ocorrem isoladamente ou em pequenos grupos e raramente em grandes

aglomerados. A aparência típica dos bacilos é de bacilos finos e ligeiramente curvos. Se forem observadas estruturas com coloração correta e morfologia compatível, devem ser consideradas bacilos álcool-ácido resistentes e relatadas como tal na análise de baciloscopia (RIEDER et al., 2007).

Visando facilitar o processo de leitura, o Manual de Recomendações para o Diagnóstico Laboratorial de Tuberculose e Micobactérias não Tuberculosas de Interesse em Saúde Pública no Brasil (MINISTÉRIO DA SAÚDE, 2022b) sugere que cada campo microscópico seja dividido mentalmente em quatro quadrantes, conforme a Figura 5, utilizando os números 12, 3, 6 e 9 para as divisões dos quadrantes. O técnico inicia a leitura no quadrante superior direito, entre os números 12 e 3, e contabiliza os BAAR presentes em cada quadrante. Ao completar 20 campos, deve ser calculada a média de bacilos por campo. Caso a média seja maior que 10 BAAR por campo, então a leitura é encerrada e dado o diagnóstico “3+” para a amostra. Caso contrário, deve-se continuar a leitura até completar 50 campos. Após completar a contagem nos 50 campos, deve ser calculada a média de bacilos por campo. Caso a média obtida seja de 1 a 10 BAAR por campo, então a leitura é encerrada e é dado o diagnóstico “2+” para a amostra. Caso contrário, deve-se continuar a leitura até completar 100 campos. Após completar a contagem dos 100 campos, os resultados são somados. Se for encontrado um total de 10 a 99 BAAR nos 100 campos é dado o diagnóstico “1 +” para amostra. Se for encontrado de 1 a 9 BAAR nos 100 campos observados, deve ser relatado o número exato de BAAR encontrados. Caso não sejam encontrados BAAR nos 100 campos observados é dado o diagnóstico “negativo”.

Figura 5. Divisão do campo microscópico.



Fonte: (BRASIL, 2008).

### 3.2 Espaços de Cor

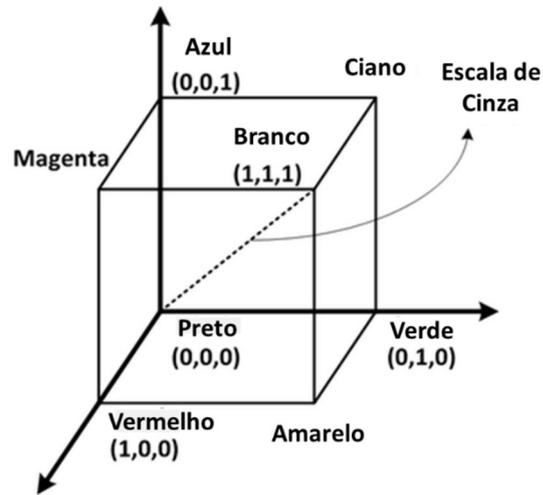
Este trabalho utiliza imagens de microscopia para a identificação e classificação do bacilo da Tb. Essas imagens correspondem aos campos microscópicos visualizados pelo microscopista e podem ser trabalhadas em diferentes espaços de cor. O propósito de um espaço de cor é facilitar a especificação de cores de alguma forma padrão, com o objetivo de tornar o conceito abstrato do espectro de cores em algo compreensível. Em essência, um espaço de cor é uma especificação de um sistema de coordenadas e um subespaço dentro desse sistema, de modo que cada cor seja representada por um único ponto contido nesse subespaço (GONZALEZ; WOODS, 2018). Existem vários espaços de cor em uso hoje, porém, neste trabalho dois espaços serão utilizados e combinados nas implementações da metodologia: RGB (do inglês *Red, Green, Blue*) e HSV (do inglês *Hue, Saturation, Value*). Enquanto o espaço de cor RGB se refere ao processamento biológico das cores no sistema visual humano, e define uma cor com porcentagens de tons de vermelho, verde e azul, o espaço de cor HSV corresponde à percepção humana da similaridade de cores, definindo uma cor pelo matiz (tom da cor), saturação (profundidade ou pureza da cor) e valor (brilho da cor).

#### 3.2.1 RGB

O espaço de cor RGB é baseado no sistema de coordenadas cartesianas. Esse espaço de cor é representado como um cubo, no qual mapeia os valores vermelho, verde e azul nos eixos x, y e z. A Figura 6 mostra o cubo que corresponde ao espaço de cor RGB. Os valores primários RGB estão em três cantos; as cores secundárias ciano, magenta e amarelo estão em outros três cantos; o preto está na origem; e o branco está no canto mais distante da origem. A escala de cinza se estende do preto ao branco ao longo da linha que une esses dois pontos. As cores são pontos no cubo ou dentro dele e são definidas por vetores que se estendem desde a origem (GONZALEZ; WOODS, 2018).

As imagens representadas no espaço de cor RGB consistem em três imagens componentes, uma para cada cor primária. E essas componentes combinadas produzem uma imagem colorida. No entanto, o espaço de cor RGB depende do dispositivo, uma vez que diferentes monitores podem produzir cores visualmente diferentes, ainda que tenham os mesmos valores RGB especificados. E embora seja muito utilizado e corresponda ao processamento biológico das cores no sistema visual humano, o espaço de cor RGB não é intuitivo para os humanos, pois não se correlaciona com a distinção perceptiva humana entre as cores (LOESDAU; CHABRIER; GABILLON, 2014).

Figura 6. Esquema do cubo do espaço de cor RGB.



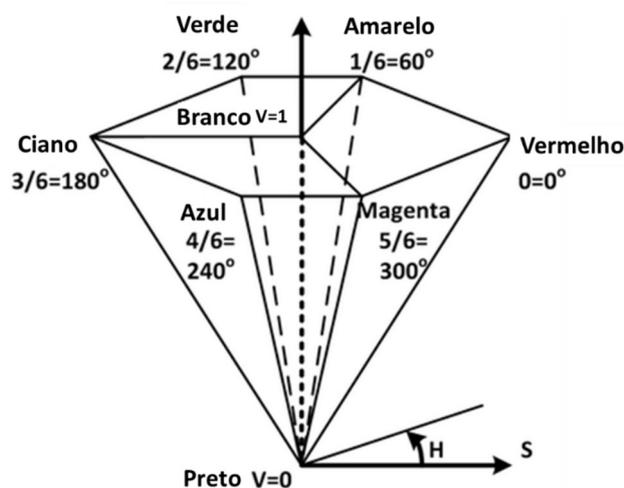
Fonte: (LEE; WANG; PLATANIOTIS, 2014)

### 3.2.2 HSV

O espaço de cor HSV mapeia as cores primárias RGB em dimensões que obedecem a diferenciação humana entre as cores: as dimensões matiz, saturação e valor. E tem a vantagem de separar as informações cromáticas (matiz e saturação) das acromáticas (valor) (LOESDAU; CHABRIER; GABILLON, 2014).

O espaço de cor HSV pode ser representado em um hexacone em três dimensões, mostrado na Figura 7. Conforme BORA; KUMAR GUPTA; KHAN (2015), na dimensão H, o matiz é um ângulo no intervalo de  $[0, 360^\circ]$  relativo ao eixo vermelho, onde vermelho é o ângulo  $0^\circ$ , amarelo é o ângulo  $60^\circ$ , verde o ângulo  $120^\circ$ , ciano o ângulo  $180^\circ$ , azul o ângulo  $240^\circ$ , magenta  $300^\circ$  e vermelho novamente o ângulo  $360^\circ$ . Na dimensão S, a saturação descreve o quão puro é o matiz em relação a uma referência branca, entendido como a pureza ou profundidade da cor. É medido como a distância radial do eixo central, com valor 0, e superfície externa, com valor 1. Uma cor com valor de saturação igual a 1 será a cor mais pura possível, enquanto 0 de saturação produz tons de cinza. A dimensão V indica a quantidade de luz que ilumina uma cor, ou seja, controla o brilho de uma cor. Uma cor com brilho igual a 0 é preto puro, enquanto uma cor de brilho igual a 1 não contém preto misturado à cor.

Figura 7. Esquema do hexacone do espaço de cor HSV.



Fonte: (LEE; WANG; PLATANIOTIS, 2014)

### 3.3 Aprendizado de Máquina

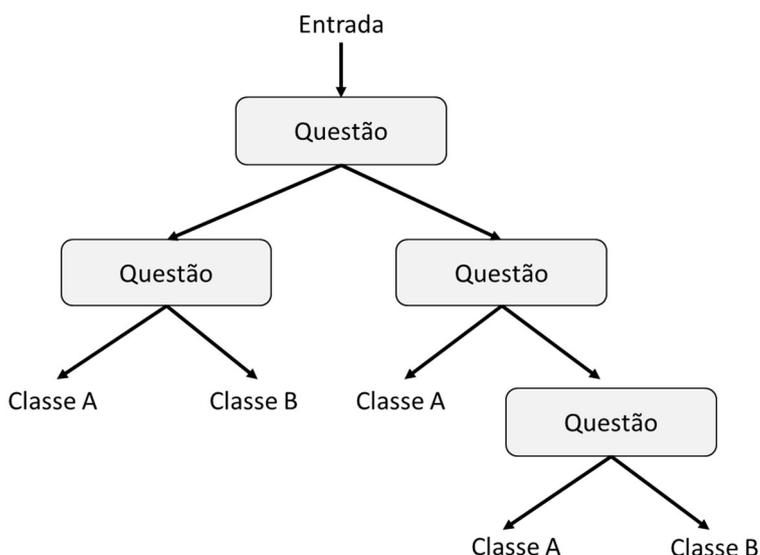
Aprendizado de máquina pode ser definido como um conjunto de algoritmos capazes de aprender automaticamente por meio da experiência, através do processamento de dados (SEWELL, 2009). Ou seja, são algoritmos que analisam amostras de um banco de dados buscando melhorar um critério de desempenho em uma determinada tarefa. Os algoritmos de aprendizado de máquina podem ser classificados como algoritmos de aprendizado supervisionado e não supervisionado. Algoritmos de aprendizado supervisionado analisam dados de uma amostra associada a uma saída, como por exemplo os algoritmos: árvores de decisão, SVM, kNN, rede neural artificial (ANN, do inglês *Artificial Neural Network*) e CNN. Enquanto algoritmos de aprendizado não supervisionado analisam os dados e aprendem propriedades úteis da estrutura desses dados, sem ter uma saída associada, como por exemplo o *clustering* que divide o banco de dados em grupos semelhantes, e a Análise de Componentes Principais (PCA, do inglês *Principal Component Analysis*) que faz redução de dimensionalidade dos dados (GOODFELLOW; BENGIO; COURVILLE, 2016).

Neste trabalho, foram utilizadas as técnicas de aprendizado supervisionado, visto que as amostras têm uma saída associada, que corresponde à informação se é um bacilo ou não. Dois algoritmos foram utilizados na metodologia: redes neurais convolucionais e *Random Forest*. Nas seções seguintes são abordados conceitos importantes para a compreensão dos algoritmos de RF e CNN expostos nas seções 3.4.1 e 3.6, respectivamente.

### 3.4 Árvores de Decisão

Árvores de Decisão são modelos de aprendizado de máquina supervisionado que realizam tarefas de classificação (saída discreta) ou regressão (saída contínua), a partir de dados de entrada estruturados em árvores. Essas estruturas são semelhantes a fluxogramas, e dispõem de facilidade de interpretação e visualização. Na Figura 8 é ilustrada uma estrutura de árvore de decisão, na qual os parâmetros aprendidos são questões sobre os dados. Cada “Questão” ilustrada na Figura 8 é chamada de nó e analisa o valor de uma característica dos dados. A questão superior representa a raiz da árvore, e por isso é denominada por nó raiz. As questões seguintes representam os ramos da árvore, e são denominadas por nó de decisão. E as classes representam as folhas, também denominadas por nó terminal. Cada nó da árvore está associado a uma região dos dados e os nós de decisão dividem essa região em sub-regiões (GOODFELLOW; BENGIO; COURVILLE, 2016). Portanto, os dados de entrada são divididos em regiões não sobrepostas.

Figura 8. Ilustração de uma estrutura de Árvore de Decisão.



A classificação de um novo exemplo inicia no nó raiz, e o valor da característica que corresponde a este nó é considerado. Então o exemplo é movido para o ramo que satisfaz a questão sobre essa característica, chegando ao nó de decisão correspondente, que vai analisar uma nova característica dos dados. Esse processo é repetido até chegar ao nó terminal, a folha, que corresponde a um valor alvo, uma classe para tarefas de classificação. Para todos os exemplos que chegam à mesma folha, o mesmo valor alvo será previsto. Apenas uma folha e os nós no caminho do nó raiz à folha são ativados quando uma entrada é fornecida.

As árvores de decisão são aprendidas com um algoritmo *Top-Down Induction of Decision Trees* (TDIDT), particionamento recursivo ou aprendizado de divisão e conquista. O

algoritmo seleciona a melhor característica para a raiz da árvore, divide o conjunto de dados em conjuntos separados e adiciona nós e ramos correspondentes à árvore. O critério de divisão para características discretas tem a forma  $t \leftarrow (A = v)$ , onde  $v$  é um valor possível da característica escolhida  $A$ , e para características numéricas tem a forma  $t \leftarrow (A < v_t)$ , que indica se o valor da característica está acima ou abaixo de um certo limiar  $v_t$  (SAMMUT; WEBB, 2010). A impureza é uma medida utilizada para avaliar a homogeneidade das classes nos nós filhos de um nó. A diminuição da impureza é utilizada pelos critérios de divisão visando selecionar a melhor divisão em cada nó. Sendo assim, quanto menor for a impureza, mais similares são as amostras dentro de uma classe e mais clara é a separação entre as classes.

Critério de divisão, geração do modelo e *overfitting* são pontos fundamentais que devem ser analisados na construção de árvores de decisão (GULATI; SHARMA; GUPTA, 2016):

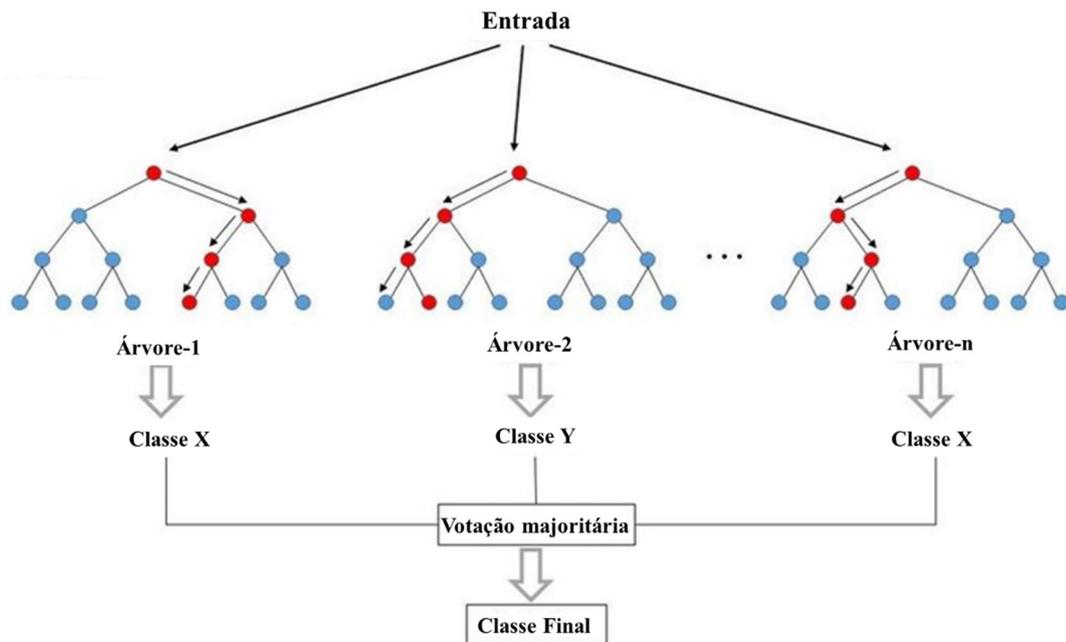
- O critério de divisão busca identificar qual a melhor característica para dividir os dados. Diferentes técnicas podem ser utilizadas como Entropia, Ganho de informação, Índice de Gini etc.
- Geração de modelo está relacionado ao tipo de dados utilizados, se eles constroem um modelo de regressão ou classificação.
- *Overfitting* usa a aplicação de métodos de poda para melhorar a generalização do modelo em dados não vistos. Os métodos de poda controlam o tamanho da árvore com parâmetros como: número mínimo de amostras por folha e máxima profundidade da árvore.

### 3.4.1 Random Forest

O modelo *Random Forest* (RF) é uma técnica de *ensemble learning* que permite maior confiabilidade nas predições, pois combina as predições de várias árvores de decisão especializadas. Ao final, as saídas das árvores são agrupadas para realizar uma predição. Para tarefas de classificação, a classe mais votada é considerada. E para tarefas de regressão, a média simples de todas as saídas das árvores é considerada. A Figura 9 ilustra um modelo RF. Para a construção de um modelo RF, o critério de decisão, tipo do modelo e métodos de poda são fundamentais. Para controlar o tamanho da árvore e do modelo, podem ser ajustados parâmetros como a quantidade total de árvores, o número mínimo de amostras por folha, o número mínimo de amostras necessárias para uma divisão, a profundidade máxima da árvore, o número máximo de nós por árvore etc. No entanto, é importante observar que esses parâmetros podem variar entre os diferentes simuladores.

O modelo Random Forest é uma combinação de árvores, onde cada árvore depende de um vetor do conjunto de dados aleatório amostrado de forma independente e com a mesma distribuição para todas as árvores (BREIMAN, 2001). Além da amostragem das observações, as árvores de decisão do modelo RF operam em uma amostragem aleatória das características. Cada classificador usa uma seleção aleatória de características para dividir cada nó. O conjunto de características é escolhido uniformemente ao acaso a partir de uma amostra *bootstrap* do conjunto de dados original (SAMMUT; WEBB, 2010). Amostragem *bootstrap* é uma técnica para obter aleatoriamente amostras de conjunto de dados, podendo ter dados omitidos ou repetidos. Essa aleatoriedade ajuda a reduzir a correlação entre as árvores individuais e aumenta a diversidade do modelo.

Figura 9. Ilustração de um modelo Random Forest.



Fonte: Adaptado de (DIMITRIADIS; LIPARAS, 2018)

### 3.5 Redes Neurais Artificiais

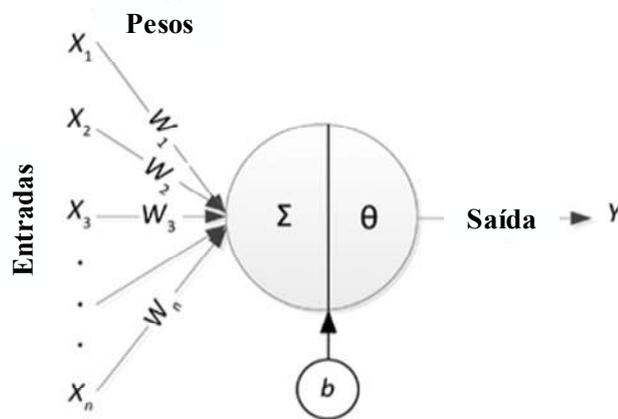
O cérebro biológico toma decisões a partir de uma comunicação complexa entre seus neurônios. Inspirado nas redes neurais do cérebro, o modelo de redes neurais artificiais é capaz de tomar decisões e realizar cálculos complexos por meio da comunicação entre dispositivos computacionais básicos, os neurônios artificiais (SHALEV-SHWARTZ; BEN-DAVID, 2014). A estrutura de um neurônio artificial é semelhante à estrutura de um neurônio do sistema nervoso, e está ilustrada na Figura 10. As entradas ( $X$ ) conectadas ao neurônio através dos pesos ( $W$ ) simulam a estrutura do dendrito, que atuam na recepção de estímulos de outros neurônios.

O somatório, a polarização ( $b$ ) e a função de ativação ( $\theta$ ) funcionam como o corpo da célula. A propagação da saída ( $Y$ ) é análoga ao axônio em um neurônio biológico, que transmite sinais elétricos para outros neurônios. A representação matemática de um neurônio artificial é dada na equação (1) e pode ser representada pela forma matricial, conforme a equação (2).

$$Y = \theta(\sum_{i=1}^n W_i X_i + b) \quad (1)$$

$$Y = \theta(W \cdot X + b), \text{ sendo } W = [W_1 \quad W_2 \quad \dots \quad W_n] \text{ e } X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} \quad (2)$$

Figura 10. Estrutura básica de um neurônio artificial.



Fonte: (AWAD; KHANNA, 2015)

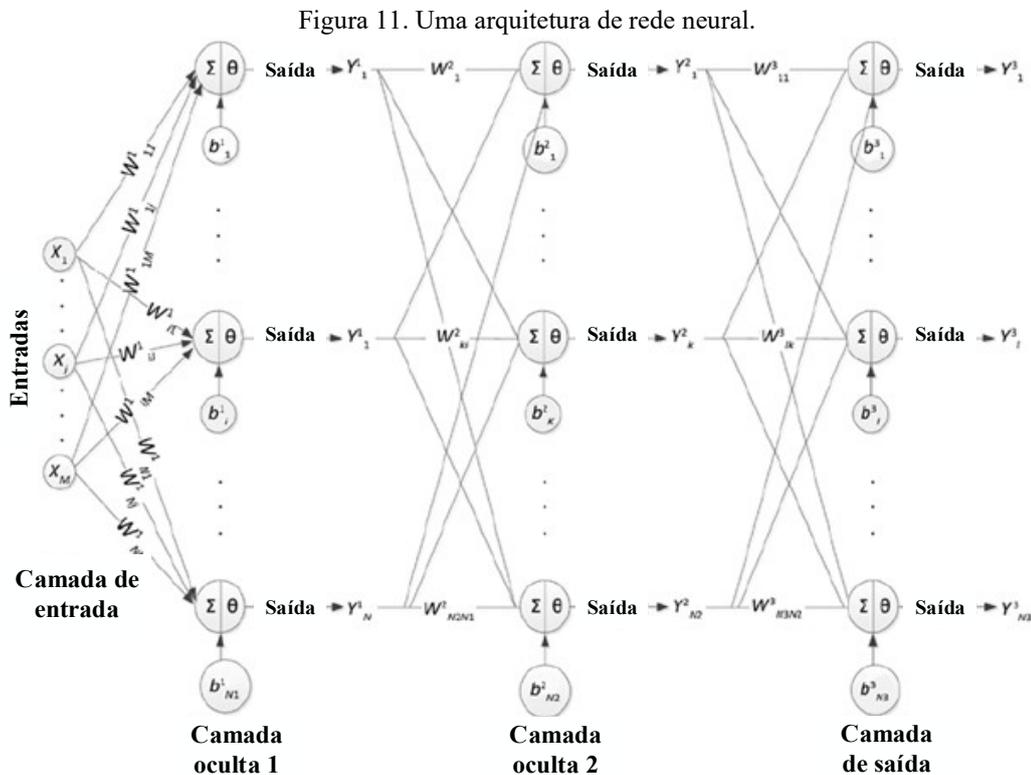
A entrada  $X$  representa um conjunto de características aplicado na entrada do neurônio artificial para determinar uma saída  $Y$ . A matriz de pesos  $W$  corresponde a importância de cada característica de entrada. Esses pesos são aprendidos durante o treinamento da rede neural através da análise de um conjunto de dados. A saída é predita pela função de ativação  $\theta$  que é ajustada pela polarização  $b$ , em conformidade com a tarefa a ser desempenhada pela rede neural. O desempenho do modelo de rede neural depende da escolha da função de ativação. As funções de ativação mais utilizadas são:

- Sigmoid:  $\theta(a) = \frac{1}{1+e^a}$
- Tangente hiperbólica:  $\theta(a) = \frac{e^a - e^{-a}}{e^a + e^{-a}}$
- Degrau:  $\theta(a) = \begin{cases} 0 & \text{se } a < 0 \\ 1 & \text{se } a > 0 \end{cases}$
- Linear:  $\theta(a) = \begin{cases} 0 & \text{se } a < 0 \\ a & \text{se } 0 \leq a \leq 1 \\ 1 & \text{se } a > 1 \end{cases}$

- Softmax:  $\theta(x_i) = \frac{e^{x_i}}{\sum_j e^{x_j}}$

Uma camada de neurônios é formada por vários neurônios artificiais conectados a um mesmo conjunto de entrada, operando em paralelo. A saída de uma camada é um vetor, em que cada componente corresponde a saída de um neurônio artificial dessa camada. A rede neural artificial, também chamada de perceptron multicamadas, é composta por várias camadas de neurônios, cada uma com sua própria matriz de peso, vetor de polarização e vetor de saída (AWAD; KHANNA, 2015). Na rede neural artificial, as camadas são completamente conectadas, ou seja, todos os neurônios de uma camada da rede estão conectados a todos os neurônios da camada seguinte.

A arquitetura de uma rede neural artificial é composta por três tipos principais de camadas. A primeira camada é a camada de entrada, que corresponde ao conjunto de características. Em seguida, temos as camadas de neurônios ocultas, cuja quantidade e número de neurônios são definidos com base na complexidade da tarefa que a rede deve realizar. Por fim, a camada de saída será projetada de acordo com a tarefa específica da rede, por exemplo, para tarefas de classificação multiclasse, a camada de saída terá a quantidade de neurônios igual a quantidade de classes que a rede deve classificar. A Figura 11 ilustra uma arquitetura de rede neural artificial com uma camada de entrada, duas camadas ocultas e uma camada de saída.



Fonte: (AWAD; KHANNA, 2015)

### 3.5.1 Processo de Treinamento da Rede Neural

O processo de treinamento de uma rede neural realiza a atualização dos pesos e polarizações de cada camada com o objetivo de melhorar seu desempenho e torná-la eficiente em reconhecer padrões de diferentes classes. Durante o processo de treinamento é calculado erro entre a saída  $Y$  predita pela rede e a saída desejada  $D$ . Assim, o treinamento busca reduzir gradualmente esse erro, modificando os valores dos pesos e polarizações (JAIN; MAO; MOHIUDDIN, 1996).

O erro entre a saída predita e a saída desejada é calculada por uma função de custo. Para tarefas de regressão, a função de custo do Erro Médio Quadrático é frequentemente usada. Em redes neurais modernas é usada a Máxima Probabilidade, também chamada por função de custo *cross-entropy* (GOODFELLOW; BENGIO; COURVILLE, 2016). A equação (3) mostra essa função de custo.

$$L_i = -\log (P(c(i)|X_i)) \quad (3)$$

Onde:

$L_i$  - Valor da perda ao apresentar um padrão  $X_i$  na entrada

$i \in \{0,1\}$  é o índice de cada classe do padrão  $X_i$

Baseado na função de custo escolhida, os pesos e polarizações são modificados por meio de um algoritmo de treinamento, de tal forma que a saída predita se aproxime da desejada. O algoritmo de *backpropagation* é o mais utilizado para atualização dos pesos e polarizações. É um algoritmo iterativo, em que o critério de parada é normalmente fixado pelo número máximo de épocas. Também é possível utilizar funções de erros regularizadoras ou o método da parada antecipada no treinamento, a fim de melhorar a capacidade da rede de fazer previsões corretas para amostras que não foram observadas no treinamento.

### 3.5.2 Algoritmo *Backpropagation*

Nas redes multicamadas é usado o algoritmo *backpropagation* na atualização dos pesos. Esse algoritmo consiste nos seguintes passos (AWAD; KHANNA, 2015):

- Inicialização: os pesos são inicializados em valores randômicos próximos a 0.
- Passagem para frente (*forward*): as entradas são alimentadas na rede e propagadas para a camada de saída e calculado o erro resultante.
- Passagem para trás (*backward*): os pesos e polarizações são atualizados com o algoritmo gradiente descendente estocástico:

$$W_{ij}^k(t + 1) = W_{ij}^k(t) - \alpha \frac{\partial E}{\partial W_{ij}^k} \quad (4)$$

$$b_i^k(t + 1) = b_i^k(t) - \alpha \frac{\partial E}{\partial b_i^k} \quad (5)$$

Em que  $\alpha$  é uma taxa de aprendizado ajustável,  $E$  é a função de custo,  $W_{ij}^k$  é o peso da conexão entre o  $j$ -ésimo neurônio na camada  $k - 1$  e o  $i$ -ésimo neurônio na camada  $k$ , e  $b_i^k$  é a polarização do  $i$ -ésimo neurônio na camada  $k$ . A escolha de  $\alpha$  afeta a velocidade de convergência do algoritmo de *backpropagation*. Valores altos para a taxa de aprendizado podem causar oscilações no algoritmo, e valores muito baixos podem resultar em uma convergência lenta.

É essencial que a função de custo e as funções de ativação sejam deriváveis, pois o cálculo da atualização de pesos é realizado com o gradiente da função de custo em relação a um peso, o qual utiliza a regra da cadeia e envolve as funções de ativação. Como o objetivo é diminuir a função de custo, a cada iteração o peso atual é subtraído do valor do gradiente multiplicado pela taxa de aprendizado. Uma vez que o cálculo do gradiente determina o sentido e a taxa de variação da função de custo em relação ao peso, o erro da rede é reduzido por meio dessa abordagem do gradiente descendente (AWAD; KHANNA, 2015).

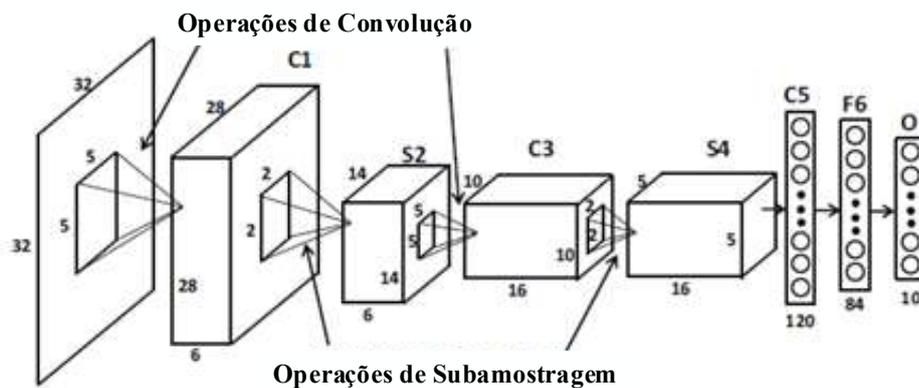
### 3.6 Redes Neurais Convolucionais

O processamento de imagens com redes ANN envolve arquiteturas muito complexas, pois elas trabalham apenas com entradas lineares e são inteiramente conectadas. Assim, uma imagem deve ser convertida em um vetor de grande dimensão, tendo cada neurônio da primeira camada conectado a esse vetor, o que implica em uma grande quantidade de parâmetros e um processamento muito lento. As redes neurais convolucionais foram criadas para permitir o processamento de imagens sem a sobrecarga de tantas conexões, utilizando camadas convolutivas que se conectam não com a imagem inteira, mas com uma região dela (AGGARWAL, 2018).

Redes neurais convolucionais são usadas na visão computacional para classificação de imagem e detecção de objeto. A arquitetura de uma CNN tem duas camadas principais: a camada de convolução e a de subamostragem. Todas as camadas da rede são tridimensionais. Na camada de entrada, a terceira dimensão refere-se ao número de planos da entrada. Por exemplo, uma imagem RGB tem três planos, enquanto uma imagem em nível de cinza tem apenas um plano. Nas camadas convolutivas, a terceira dimensão refere-se ao número de filtros convolutivos utilizados. Cada filtro convolutivo gera um mapa de características, que

representam o resultado de vários processamentos realizados por um filtro na imagem (AGGARWAL, 2018). Esse mapa de características é gerado por meio da operação de convolução do filtro com a entrada, esse filtro tridimensional tem a mesma profundidade que a entrada, mas com dimensão espacial menor. Assim, cada neurônio da camada de saída será conectado apenas a uma pequena região da camada de entrada, reduzindo consideravelmente a quantidade de parâmetros daquele neurônio. Na camada de subamostragem a terceira dimensão também corresponde a quantidade de filtros, uma vez que essa camada não modifica a profundidade, pois tem a finalidade de reduzir o tamanho do mapa de características. Para isso, a média ou o valor máximo de uma região de tamanho  $2 \times 2$ , ou  $3 \times 3$ , é calculado para compactar as dimensões espaciais das camadas por um fator de 2 ou 3, respectivamente. Outras camadas são utilizadas na construção de uma CNN e serão abordadas nas seções seguintes. Na Figura 12 é ilustrada a arquitetura LeNet-5, uma das primeiras CNN's desenvolvidas por (LECUN et al., 1998). Ela usa imagens de entrada em escala de cinza, possui uma camada convolutiva de *kernel*  $5 \times 5$  com profundidade igual 6, seguida de uma camada de subamostragem de tamanho  $2 \times 2$ . Em sequência, há outra camada convolutiva de *kernel*  $5 \times 5$  com profundidade igual 16, seguida por mais uma camada de subamostragem de tamanho  $2 \times 2$ . A arquitetura também inclui uma camada completamente conectada de 120 neurônios, outra de 84 neurônios e, por fim, a camada de saída com 10 neurônios.

Figura 12. Arquitetura LeNet-5.



Fonte: (AGGARWAL, 2018)

### 3.6.1 Camada Convolutiva

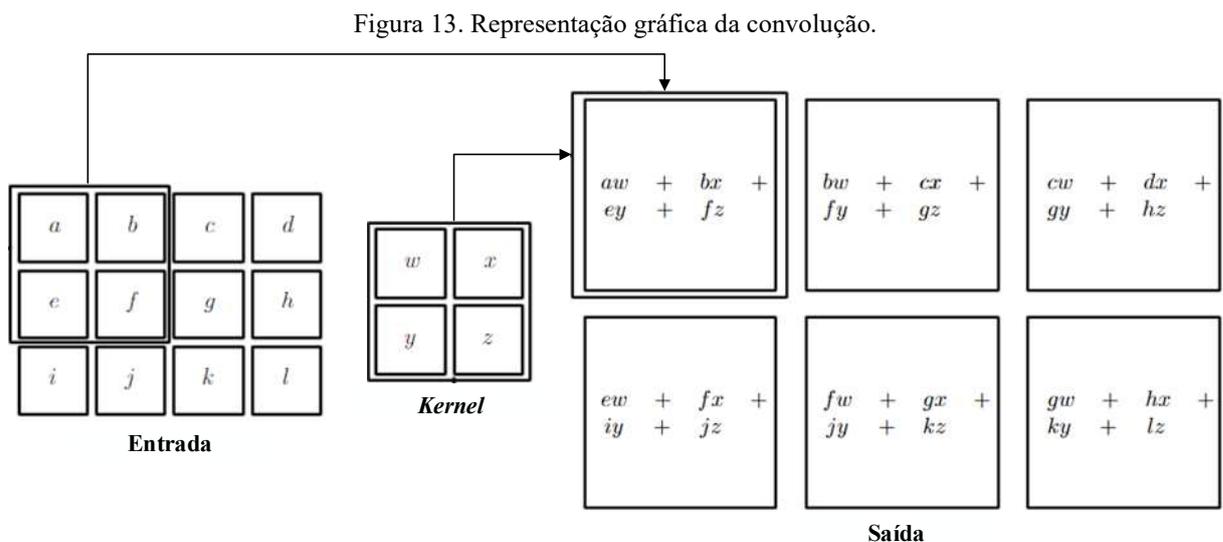
A camada convolutiva realiza uma operação chamada de convolução na entrada da camada que resulta em um mapa de características. Essa operação, mostrada na equação 6, é representada por um  $*$  e realiza o produto entre duas matrizes. Na equação (6),  $x$  é chamado de entrada,  $w$  é chamado de filtro ou *kernel*, e  $s(t)$  é a saída chamada de mapa de características.

$$s(t) = (x * w)(t) \quad (6)$$

Geralmente, as convoluções são realizadas em mais de um eixo por vez, então para uma entrada bidimensional, provavelmente a operação será com um *kernel* bidimensional (GOODFELLOW; BENGIO; COURVILLE, 2016). A operação matemática para entrada e *kernel* bidimensional é mostrada na equação (7). Onde  $S(i, j)$  se refere às posições do mapa de características que será gerado,  $K$  é o *kernel*,  $m$  e  $n$  se referem à dimensão de  $K$ .  $I$  é a imagem de entrada,  $i$  e  $j$  referem-se à dimensão de  $I$ . O *kernel* é uma matriz menor que a entrada e seus parâmetros são aprendidos pelo algoritmo de aprendizado.

$$S(i, j) = (K * I)(i, j) = \sum_{p=1}^m \sum_{q=1}^n I(i + p, j + q)K(p, q) \quad (7)$$

Na Figura 13 ilustra-se a operação de convolução graficamente. A entrada tem dimensão  $3 \times 4$ , o *kernel* tem dimensão  $2 \times 2$ , e a operação de convolução foi realizada com um passo de 1. Nota-se que, cada posição da saída, é a soma dos elementos da entrada multiplicados pelos elementos *kernel*, nas posições correspondentes.



Fonte: Adaptado de (GOODFELLOW; BENGIO; COURVILLE, 2016).

Na camada convolutiva é necessária a definição de alguns hiper parâmetros: tamanho do *kernel*, profundidade e passo. O tamanho do *kernel* está associado à região da entrada que o neurônio da camada de saída estará conectado. Em ANN cada neurônio da camada de saída se conecta a todos os neurônios da camada de entrada. Em CNN, cada neurônio na camada de saída se conecta a uma região de tamanho  $m \times n$  da camada de entrada, o qual corresponde ao tamanho do *kernel*, conforme ilustrado na Figura 14. Quando se tem uma entrada com múltiplos planos, a rede aprenderá os parâmetros de um *kernel* para cada plano da entrada. O neurônio da saída estará associado a uma região que abrange todos os canais da entrada. Por exemplo, se

utilizarmos um *kernel* de  $5 \times 5$  em uma entrada de 3 planos, cada neurônio da saída terá um total de  $5 \times 5 \times 3 = 75$  pesos. O hiper parâmetro profundidade se refere à quantidade de *kernels* na camada convolutiva e, conseqüentemente, ao número de neurônios relacionados a uma mesma região de entrada. Dessa forma, a rede pode identificar diferentes características em uma mesma região, como bordas verticais, bordas horizontais, entre outras. A saída terá a profundidade igual a quantidade de *kernels* definidos. O hiper parâmetro passo controla o deslocamento do *kernel* ao percorrer a entrada. Passo 1 indica um salto de uma coluna da matriz de entrada no deslocamento horizontal e uma linha no deslocamento vertical. Outro hiper parâmetro a ser definido é o preenchimento de zeros, que é usado para controlar a dimensão da saída. A operação de convolução reduz a entrada, e pode ser desejável manter a dimensão, para isso são acrescentados zeros nas bordas da entrada, mantendo algumas informações. A Figura 15 ilustra um exemplo de uma camada convolutiva com entrada de dimensão  $6 \times 6 \times 3$ , tamanho do *kernel*  $3 \times 3$ , profundidade igual a 2, passo igual a 1 e sem preenchimento de zeros. A saída é uma imagem de dimensão  $4 \times 4 \times 2$ . Nota-se que a profundidade de cada *kernel* corresponde a profundidade da entrada, e cada *kernel* extrai diferentes características de uma mesma região. A saída foi reduzida devido fato de não ter zeros nas bordas da entrada, tem profundidade igual a quantidade de *kernels*.

Figura 14. Ilustração da região da camada de entrada conectada à uma camada de saída.

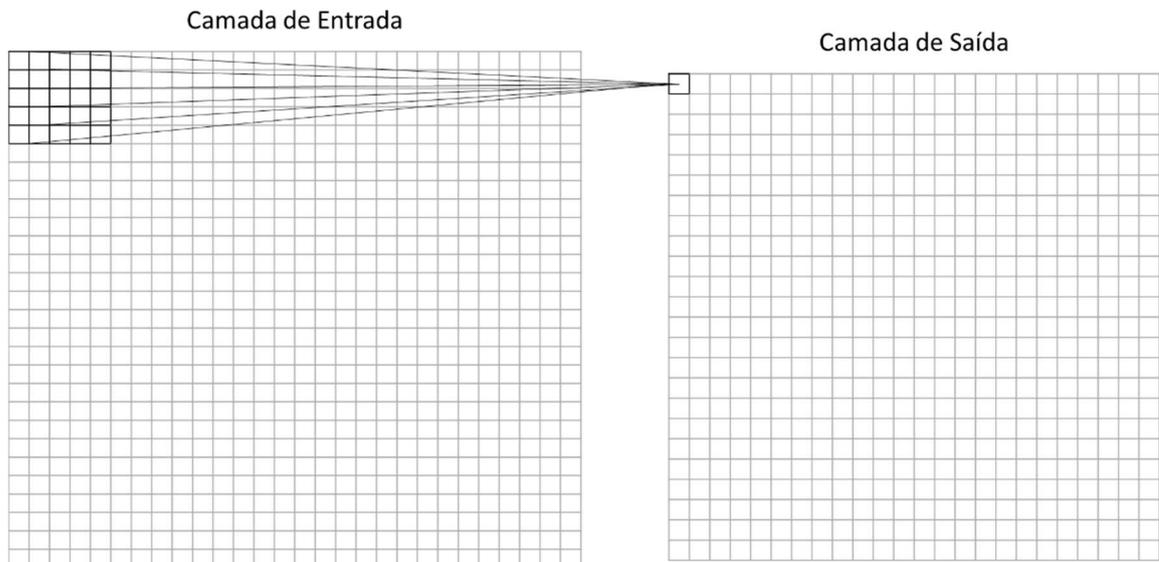
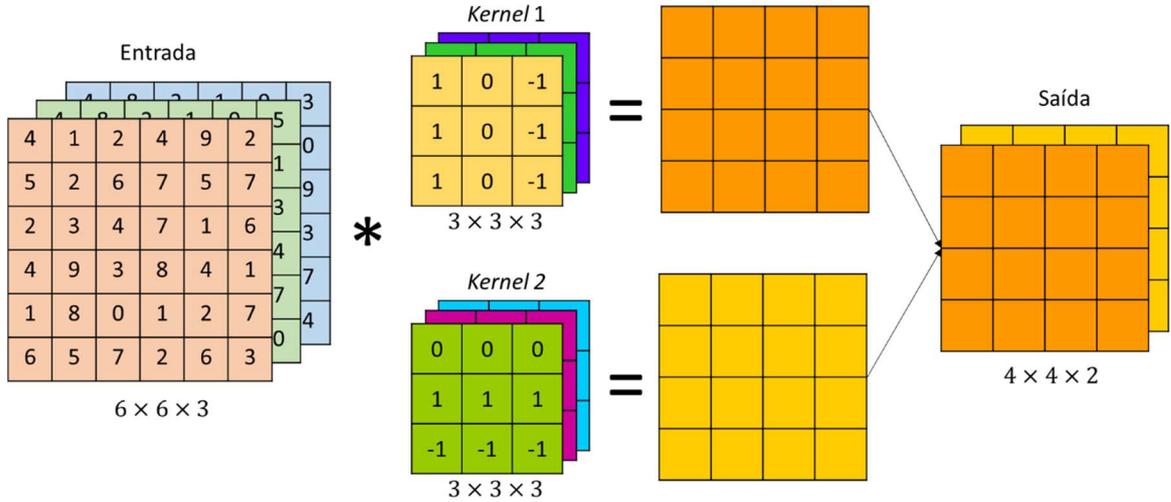


Figura 15. Ilustração de uma camada convolutiva em uma entrada  $6 \times 6 \times 3$ , tamanho do *kernel*  $3 \times 3$ , profundidade igual a 2, passo igual a 1 e sem preenchimento de zeros.



É possível calcular o tamanho da saída a partir dos parâmetros da camada convolutiva. Sendo  $W_1$  e  $H_1$  as dimensões da entrada,  $W_2$  e  $H_2$  as dimensões da saída,  $D_2$  a profundidade da saída,  $F$  é o tamanho do *kernel*,  $P$  é a quantidade de preenchimento de zeros usado nas bordas,  $S$  é o passo definido e  $K$  é a quantidade de *kernels*. As equações (8), (9) e (10) calculam o tamanho do volume de saída.

$$W_2 = \frac{W_1 - F + 2P}{S + 1} \quad (8)$$

$$H_2 = \frac{H_1 - F + 2P}{S + 1} \quad (9)$$

$$D_2 = K \quad (10)$$

As camadas convolutivas de uma CNN têm a função de gerar um mapa de características a partir de uma entrada. A entrada de uma camada convolutiva é o mapa de característica gerado pela camada anterior e tem as dimensões e profundidade definidas pelos hiper parâmetros da camada.

### 3.6.2 Camada ReLU

ReLU (*Rectifier Linear Unity*) é uma função de ativação não linear que substituiu as funções de ativação sigmoid e tangente hiperbólica nas camadas de modelos CNN, devido suas grandes vantagens em termos de velocidade e precisão (AGGARWAL, 2018). A camada ReLU realiza a ativação de cada unidade de neurônio da entrada da camada usando a equação (11). É aplicada após a camada de convolução para realizar a ativação dos neurônios.

$$f(x) = \max(0, x) \quad (11)$$

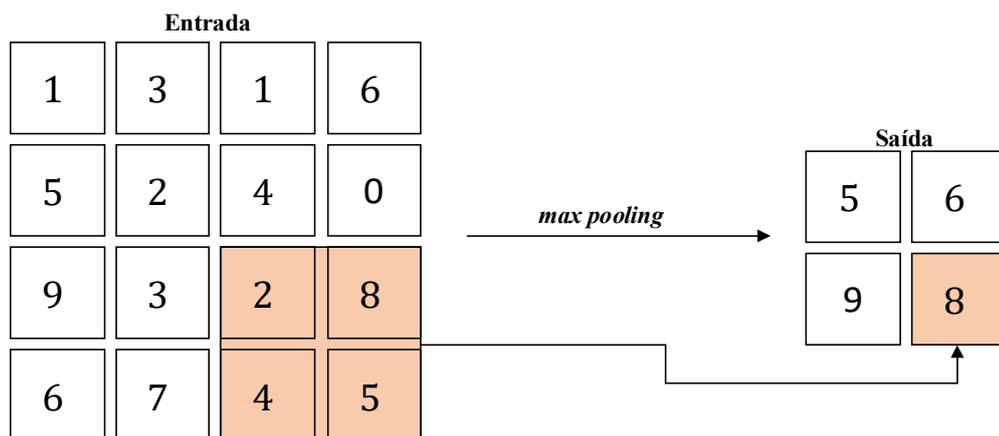
Em que  $x$  é qualquer valor real e representa o valor do neurônio em uma camada. A função é aplicada para cada um dos neurônios da camada. Esses valores são então passados para a próxima camada. Portanto, a aplicação da ReLU não altera as dimensões de uma camada, pois é um mapeamento simples de um para um (AGGARWAL, 2018).

### 3.6.3 Camada de subamostragem (*pooling*)

Uma arquitetura de CNN normalmente é composta por camadas de convolução, que produz um conjunto de ativações lineares e cada ativação é executada através de uma função não linear, como a ReLU, e camadas de *pooling* usadas para redução de dimensionalidade da saída (GOODFELLOW; BENGIO; COURVILLE, 2016).

A camada *pooling* substitui determinado local da entrada da camada por uma estatística resumida através de operações como média ponderada, média ou valor máximo de uma vizinhança retangular. Dessa forma, o mapa de características da entrada é reduzido na saída da camada, porém não tem a sua profundidade alterada. O tamanho típico da região sobre a qual se realiza o *pooling* é de  $2 \times 2$  (AGGARWAL, 2018). Na Figura 16 está ilustrado um exemplo de camada *pooling* que reduz a entrada extraindo o valor máximo de uma vizinhança de tamanho  $2 \times 2$ , operação denominada por *max pooling*. Com um passo de 2, a saída foi reduzida em um fator de 2.

Figura 16. Operação de *pooling*.

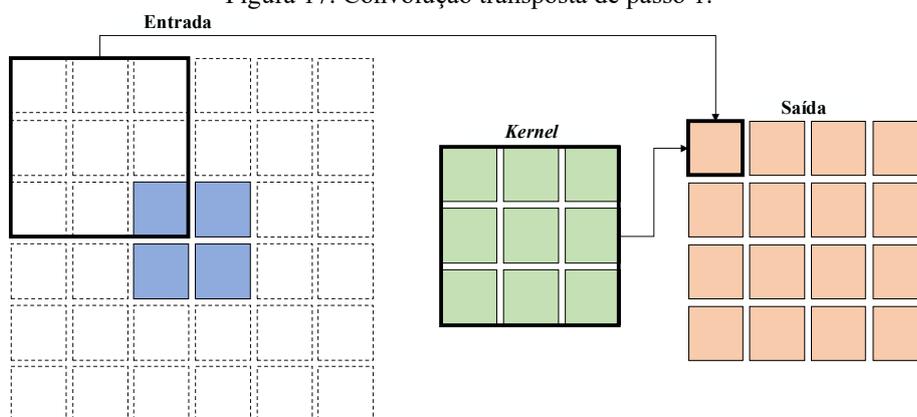


### 3.6.4 Camada de Convolução Transposta

A convolução transposta é uma convolução com um filtro que é derivado pela transposição e inversão dos filtros aprendidos na convolução (AGGARWAL, 2018). A camada de convolução transposta, também chamada de camada de deconvolução na literatura, funciona de forma oposta à camada de convolução. Enquanto as camadas de convolução são usadas para

compactar os dados de uma entrada em uma representação espacial, as camadas de convolução transposta são usadas para descompactar esses dados. Essas camadas associam uma única ativação dos mapas de características de entrada, gerados pelas camadas *pooling* e ReLU, com várias saídas por meio de filtros aprendidos que reconstróem a forma de um objeto de entrada (NOH; HONG; HAN, 2015). A Figura 17 ilustra uma operação de convolução transposta de um mapa de entrada de tamanho  $2 \times 2$ , um *kernel*  $3 \times 3$  que gera como resultado um mapa de saída de tamanho  $4 \times 4$ . É adicionada uma região preenchida com zeros ao mapa de entrada e o valor de saída é calculado da mesma forma que a operação de convolução.

Figura 17. Convolução transposta de passo 1.

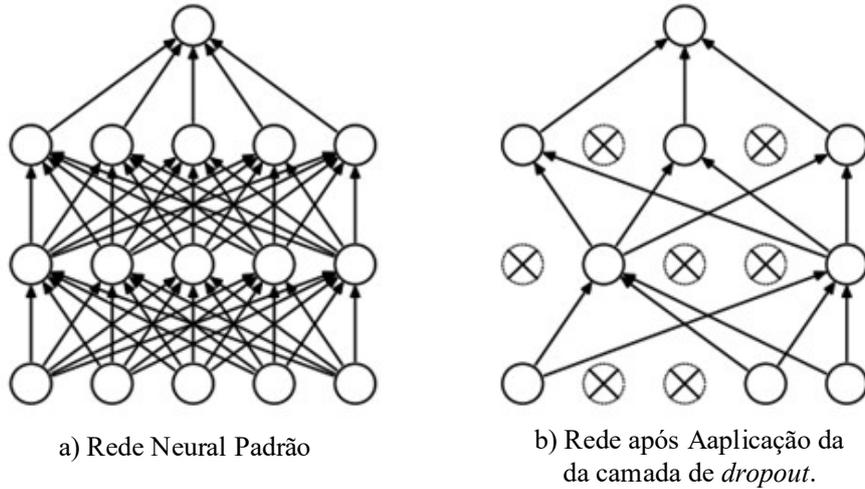


### 3.6.5 Camada de dropout

O *overfitting* refere-se ao fato da rede se especializar no conjunto de treinamento, alcançando um excelente desempenho, mas apresentar um desempenho ruim no conjunto de validação ou teste. Uma forma de reduzir esse problema é o uso de uma camada chamada *dropout*. *Dropout* é um método que cria um conjunto de redes neurais formadas pela remoção de neurônios. Se um neurônio for eliminado, então todas as conexões de entrada e saída desse neurônio também serão descartadas. Os neurônios são descartados apenas das camadas de entrada e ocultas da rede (AGGARWAL, 2018). Dessa forma, em cada etapa de treinamento, alguns neurônios são eliminados aleatoriamente e uma nova rede neural baseada na rede original é criada. A predição final da rede é obtida através de uma média da predição das diversas redes criadas. Cada neurônio é retido com probabilidade  $p$ , hiper parâmetro fixado antes do início de treinamento. Esse  $p$  pode ser escolhido usando um conjunto de validação ou pode ser definido um valor (SRIVASTAVA et al., 2014). Cada rede neural criada herda um subconjunto diferente de parâmetros da rede neural original, e esse compartilhamento de parâmetros torna possível representar muitos modelos com uma quantidade tratável de memória

(GOODFELLOW; BENGIO; COURVILLE, 2016). Na Figura 18 estão ilustradas duas redes: a) é uma rede neural padrão e b) é uma rede obtida pela aplicação do *dropout* na rede padrão.

Figura 18. Exemplo do resultado da camada de *dropout*.



Fonte: (SRIVASTAVA et al., 2014)

### 3.6.6 Camada de normalização em lote

A camada de normalização em lote, ou *batch normalization*, é uma técnica que busca melhorar a otimização do modelo reduzindo o efeito de uma convergência lenta de treinamento em modelos muito profundos. Modelos muito profundos envolvem a composição de várias funções ou camadas que são atualizadas simultaneamente, podendo gerar resultados inesperados, pois as funções são atualizadas sob a suposição de que as outras funções permanecem constante, uma vez que a atualização dos pesos assume que as outras camadas não mudam (GOODFELLOW; BENGIO; COURVILLE, 2016). Para reduzir esse efeito, camadas de normalização são adicionadas entre as camadas ocultas e cria ativações com uma mesma faixa de variação dos valores, evitando os resultados inesperados. Cada neurônio  $i$  nas camadas de normalização contém dois parâmetros adicionais, são eles  $\beta_i$  e  $\gamma_i$  que são aprendidos no treinamento. A ideia básica é que a saída do  $i$ -ésimo neurônio tenha uma média  $\beta_i$  e um desvio padrão  $\gamma_i$  sobre cada subconjunto do conjunto de treinamento (AGGARWAL, 2018).

Um efeito importante gerado pelo uso da normalização em lote é a capacidade da generalização da rede, pois, no treinamento com a camada de normalização em lote, um dado do conjunto de treinamento é visto em conjunto com outros dados no subconjunto, e a rede não produz valores específicos para aquele dado, contribuindo para a capacidade de generalização da rede (IOFFE; SZEGEDY, 2015).

### 3.6.7 Regularização L2

A regularização baseada em penalidade é a abordagem mais comum na redução de *overfitting*. A regularização L2 adiciona uma penalidade suave aos parâmetros. A penalidade adicional é definida pela soma dos quadrados de todos os pesos  $w$  da rede escalonado por um fator  $\lambda > 0$ , que é o parâmetro de regularização (AGGARWAL, 2018). A equação (12) define a função de custo regularizada  $E_R$ . Aumentar ou diminuir o valor desse parâmetro reduz a suavidade da penalidade. Para pequenos valores de  $\lambda$ , a rede prioriza minimizar a função de custo original, e para altos valores de  $\lambda$ , a rede prioriza minimizar a função de custo e a aprender pesos pequenos.

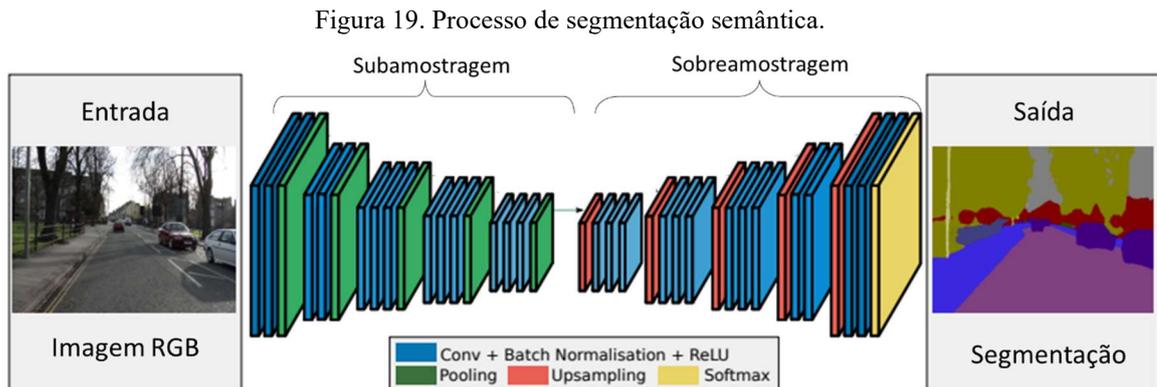
$$E_R = E + \lambda \sum_w w^2 \quad (12)$$

### 3.6.8 Segmentação Semântica

As redes neurais convolucionais podem ser usadas para tarefas que envolvem visão computacional em imagens e vídeos, entre elas estão tarefas de classificação de imagens, processamento de vídeo, detecção de objetos e segmentação semântica. Neste trabalho, as redes neurais convolucionais foram utilizadas nas implementações de dois modelos. Um de classificação de imagens e um de segmentação semântica. Para a tarefa de classificação de imagens, a imagem de entrada passa pelas camadas da rede e ao final a rede atribui uma classe para a imagem. Em segmentação semântica, a imagem de entrada passa pelas camadas da rede e ao final uma classe é atribuída para cada pixel da imagem. Na Figura 19 é ilustrado um exemplo do processo de segmentação semântica em imagens. Uma rede para a tarefa de segmentação é dividida em duas etapas: etapas de subamostragem (*pooling*) e sobreamostragem (*upsampling*). A etapa de subamostragem reduz a entrada até obter um espaço latente desta utilizando camadas como, por exemplo, a *maxpooling*. A etapa de sobreamostragem reconstrói o tamanho da entrada utilizando camadas de convolução transposta. O resultado é uma nova imagem, em que cada pixel tem uma classe atribuída. Camadas de convolução são combinadas às etapas de subamostragem e sobreamostragem para extrair as características e aprender as classes.

Por ser um modelo de aprendizado supervisionado, é necessária a classificação real de cada pixel. Dessa forma, os modelos de segmentação semântica têm, para cada amostra, uma máscara correspondente, onde cada pixel da máscara corresponde à classe do pixel na imagem. Nos trabalhos de QUINN et al. (2016), LÓPEZ et al. (2017), XIONG et al. (2018), EL-MELEGY et al. (2019) e KUOK et al. (2019) os autores utilizaram uma abordagem de classificação de retalhos, os quais continham apenas um bacilo. A rede recebia como entrada o

retalho e classificava como positivo ou negativo. Na abordagem deste trabalho, como será mostrado no capítulo Materiais e Métodos, também foi utilizada uma abordagem de classificação de retalhos, porém, antes dessa etapa, é utilizada uma rede de segmentação semântica para extrair os possíveis bacilos da imagem. A rede foi treinada com imagens formadas por um mosaico de retalhos para separar os bacilos do fundo e então identificá-los na imagem de entrada.



Fonte: Adaptado de (BADRINARAYANAN; KENDALL; CIPOLLA, 2017)

### 3.7 Método de Otimização ADAM

No treinamento da rede, um hiper parâmetro a ser definido é o método de otimização, que tem relação com a capacidade da rede de minimizar a função de custo e se tornar eficiente nas tarefas de classificação ou segmentação semântica do modelo. Muitos métodos de otimização podem ser usados. No trabalho SERRAO et al. (2020), desenvolvido anteriormente, ao qual o presente trabalho está dando continuidade, avaliou três métodos de otimização na tarefa de segmentação semântica: ADAM (do inglês *Adaptive Moment Estimation*), RMSProp (do inglês *Root Mean Square Propagation*) e SGDM (do inglês *Stochastic Gradient Descent with Momentum*). O método que apresentou melhor desempenho foi o ADAM, o qual foi utilizado nos treinamentos que serão apresentados no capítulo Materiais e Métodos.

O método ADAM é considerado bastante robusto para a escolha dos hiper parâmetros, embora às vezes precise ser alterada a taxa de aprendizado padrão sugerida (GOODFELLOW; BENGIO; COURVILLE, 2016). Ele é conhecido por ser eficiente em acelerar o treinamento de redes neurais artificiais e redes neurais convolucionais. É extremamente popular porque incorpora a maioria das vantagens de outros métodos de otimização (AGGARWAL, 2018).

O método ADAM ajusta a taxa de aprendizado para cada parâmetro da rede neural, ajustando duas médias ponderadas exponenciais para cada parâmetro, o  $A_i$  e  $F_i$ . As equações (13) a (16) correspondem às equações do método de otimização ADAM. Considerando  $A_i$  o

valor da média exponencial ponderada do primeiro momento,  $\rho$  o fator de decaimento, e  $E$  a função de custo,  $A_i$  é atualizado conforme a equação (13). Ao mesmo tempo, a média exponencialmente ponderada do segundo momento, denominada de  $F_i$ , é feita por um fator de decaimento  $\rho_f$  e é atualizada conforme a equação (14). A taxa de aprendizado  $\alpha_t$  depende do índice de iteração  $t$  e atualizada conforme a equação (15). Esse ajuste na taxa de aprendizado é um fator de correção na polarização que é importante nas iterações iniciais, onde ambos  $A_i$  e  $F_i$  são inicializados com 0 (AGGARWAL, 2018). E os parâmetros  $w_i$  do modelo são atualizados utilizando essa nova taxa de aprendizado, conforme a equação (16).

$$A_i \leftarrow \rho A_i + (1 - \rho) \left( \frac{\partial E}{\partial w_i} \right)^2; \forall i \quad (13)$$

$$F_i \leftarrow \rho_f F_i + (1 - \rho_f) \left( \frac{\partial E}{\partial w_i} \right); \forall i \quad (14)$$

$$\alpha_t = \alpha \left( \frac{\sqrt{1 - \rho^t}}{1 - \rho_f^t} \right) \quad (15)$$

$$w_i \leftarrow w_i - \frac{\alpha_t}{\sqrt{A_i}} F_i; \forall i \quad (16)$$

### 3.8 Mecanismo de Atenção de Canal (*Channel Attention Mechanism*)

Em visão computacional, mecanismos de atenção são métodos que focam a atenção nas partes mais importantes da imagem. Dependendo do método usado, há quatro categorias básicas de mecanismos de atenção: atenção de canal, atenção espacial, atenção temporal e atenção de ramo (GUO et al., 2022). Neste trabalho, dada a importância da informação de cor para a segmentação do bacilo, foi utilizado o mecanismo de atenção de canal, a *Squeeze Excitation Net* (SENet), proposto por HU et al. (2017). Atenção de canal gera um vetor de pontuações de atenção, que é multiplicado pelas informações presentes nos diversos canais de entrada. Equações (17) e (18) e a Figura 20 ilustram a operação do bloco SENet.

$$s = \sigma(W_2 \delta(W_1 \text{GAP}(X))) \quad (17)$$

$$Y = s \odot X \quad (18)$$

Onde:

$X$  – vetor de características

$s$  – vetor de pontuação

$\sigma$  – função sigmoid

$W_1$  – matriz da primeira camada completamente conectada

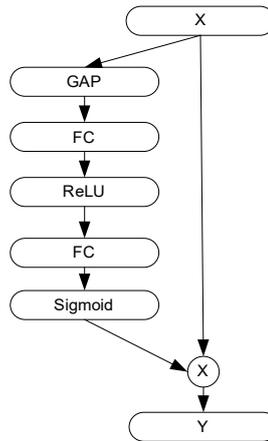
$\delta$  – função de ativação ReLU

$W_2$  – matriz da segunda camada completamente conectada

$\text{GAP}$  – operação *Global Average Pooling* de cada canal  $X$

$Y$  – saída

Figura 20. Bloco de atenção de canal SENet.

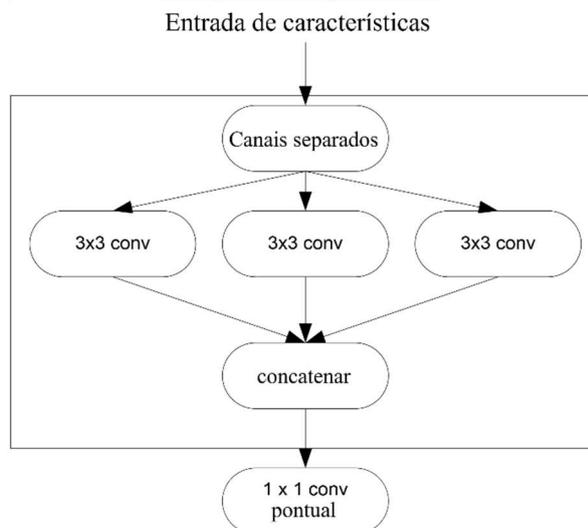


Fonte: Adaptado de GUO et al. (2022)

### 3.9 Convolução Separada por Canal (*Depthwise Separable Convolution*)

*Depthwise separable convolution*, proposto por CHOLLET (2017), é uma operação que realiza convolução espacial separada em cada canal de uma entrada de características, por meio da aplicação de um filtro por canal, combinando a saída por meio de uma convolução pontual (convolução  $1 \times 1$ ). Figura 21 ilustra esta operação. Da mesma forma que a convolução regular, *depthwise separable convolution* baseia-se na suposição de que as localizações espaciais em ativações intermediárias são altamente correlacionadas, mas os canais são altamente independentes. Em comparação com a convolução regular, *depthwise separable convolution* requer menos cálculo e menos parâmetros. Originalmente, foi usada na rede de classificação Xception.

Figura 21. Camada *depthwise separable convolution*.



## 4 MATERIAIS E MÉTODOS

A metodologia proposta para este trabalho consiste na aplicação de um modelo de segmentação semântica para segmentar os bacilos da tuberculose em imagens de microscopia, e na aplicação de filtros com o objetivo de remover os falsos positivos segmentados. Além disso, após a implementação do método de segmentação do bacilo e remoção dos falsos positivos, ele será avaliado para o diagnóstico da tuberculose. A ideia central é que cada objeto segmentado pelo modelo de segmentação semântica seja avaliado por um filtro de cor e um filtro de forma. O objeto é considerado falso positivo se não for classificado como bacilo nos filtros e é eliminado na contagem de bacilos para a determinação do diagnóstico. O diagnóstico da tuberculose será avaliado em um banco de dados construído para os 5 diagnósticos descritos no Quadro 1, na Seção 1, em que cada diagnóstico é composto por 50 conjuntos de campos digitais equivalentes a campos microscópicos.

No Capítulo 3 foram apresentados os principais conceitos necessários para iniciar a implementação deste trabalho. Neste capítulo serão expostos os materiais necessários para essa implementação e o detalhamento da metodologia.

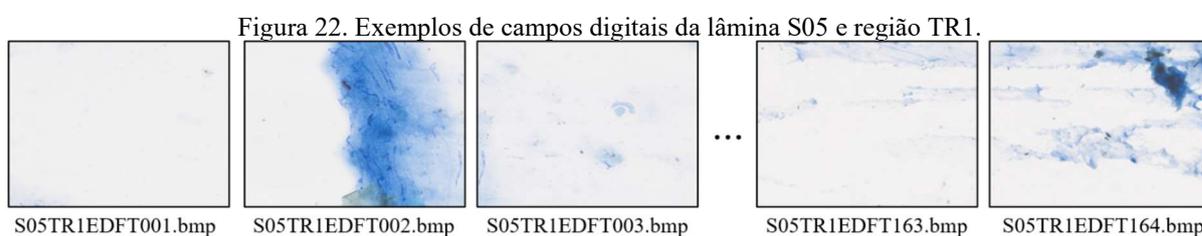
### 4.1 Materiais

#### 4.1.1 Banco de Dados Original

O banco de dados utilizado neste trabalho foi disponibilizado pelo Grupo de Pesquisa em Reconhecimento de Padrões e Otimização (<http://dgp.cnpq.br/dgp/espelhogrupo/4186>), da UFAM. Esse banco foi coletado no Laboratório de Micobacteriologia do Instituto Nacional de Pesquisa da Amazônia – INPA, protocolo 186/08. Para as aquisições das imagens, foi utilizada uma câmera digital Canon PowerShot A640 acoplada a um sensor de imagem CCD (*Coupled Charge Device*) de 10 megapixels com zoom óptico de 4×. O microscópio utilizado foi um Zeiss Axioskop 40, um aumento de 100× e abertura numérica de 1,25. O campo visual do microscópio estimado para a leitura pelo microscopista corresponde ao diâmetro de 0,228 mm, equivalente a uma área circular de 0,04 mm<sup>2</sup>. A leitura de 100 campos microscópicos equivale a uma área total de 4 mm<sup>2</sup>. O adaptador entre o microscópio e a câmera reduz a abertura do microscópio para 61% do campo microscópico. Assim, a área de um campo digital é aproximadamente 0,0243 mm<sup>2</sup>. Para avaliar uma área equivalente a 100 campos circulares, é necessária a aquisição de aproximadamente 164 campos digitais. A resolução de profundidade

das imagens de campos digitais é de 24 bits e a dimensão 1388×1040 pixels. A correspondência, no Quadro 1, na Seção 1, entre os campos microscópicos e campos digitais é: para diagnósticos negativo e esparso e 1+, a leitura de 164 campos digitais; para diagnósticos 2+ e 3+, a leitura de 82 e 32 campos digitais, respectivamente. Cada campo digital foi anotado por três especialistas em análise de esfregaço de escarro para diagnóstico da Tb.

O banco de dados original é composto por 25 lâminas, nomeadas por S01 a S25, onde cada lâmina corresponde a um paciente. De cada lâmina, 164 campos digitais foram extraídos de duas regiões denominadas TR (do inglês, *Tile Region*). A Figura 22 mostra exemplos de campos digitais da lâmina S05 e região TR1. Três bancos de dados foram elaborados com base no banco de dados original: Banco de Dados 1 (BD1), Banco de Dados 2 (BD2) e Banco de Dados 3 (BD3).



O banco de dados BD1 é composto por 8000 imagens de 400 × 400 pixels, que são formadas pela combinação de retalhos positivos (contendo bacilo) e negativos (sem bacilo) extraídos das lâminas S01 a S15. O BD1 foi usado para implementação do modelo de segmentação semântica. Lâminas de diferentes pacientes foram utilizadas nas etapas de treinamento e validação.

O banco de dados BD2 foi usado na avaliação do diagnóstico da Tb. Para as classes diagnósticas negativo e 1+, foram obtidos 150 conjuntos com 164 campos. Para a classe diagnóstica 2+ foram obtidos 50 conjuntos com 82 campos digitais. Para a classe diagnóstica 3+ foram obtidos 50 conjuntos com 32 campos digitais. Para a obtenção das imagens desses conjuntos, foram utilizados campos digitais extraídos das lâminas S16 a S25.

O banco de dados BD3 foi elaborado para treinar e testar os filtros de cor forma, e é composto por retalhos resultantes da segmentação semântica nos conjuntos de treinamento e validação do banco de dados BD1. Esse conjunto é formado por retalhos classificados como falsos positivos e verdadeiros positivos, visando treinar modelos com maior poder de diferenciação entre bacilos e artefatos classificados como bacilos. Ao total são 579 retalhos de 40×40 pixels, dos quais 219 retalhos são verdadeiros positivos e 360 retalhos são falsos positivos. A descrição dos três bancos de dados é mostrada no Quadro 3.

Quadro 3. Descrição dos bancos de dados utilizados na implementação.

Banco de Dados	Tamanho da Imagem	Lâminas	Número de imagens	Aplicação
BD1	Imagens Mosaicos de 400×400 pixels e suas respectivas máscaras binarizadas	S01-S15	8000 imagens mosaicos e suas máscaras binarizadas: 5000 imagens mosaicos com 50% de retalhos positivos e 50% negativos; 1500 imagens mosaicos com mais de 90% de retalhos negativos; 1500 com 100% de retalhos negativos	Treinamento, validação e teste do modelo de segmentação
BD2	Imagens completas de 1040×1388 pixels	S16-S25	250 conjuntos de campos digitais: 150 conjuntos de 164 campos digitais correspondem aos diagnósticos negativo, esparsos e 1+; 50 conjuntos de 82 campos digitais correspondem ao diagnóstico 2+; e 50 conjuntos de 32 campos digitais correspondem ao diagnóstico 3+	Avaliação do diagnóstico da Tuberculose
BD3	Retalhos de 40×40 pixels	S01-S15	579 retalhos: 219 retalhos verdadeiros positivos e 360 retalhos falsos positivos	Treinamento e teste dos filtros de cor e forma

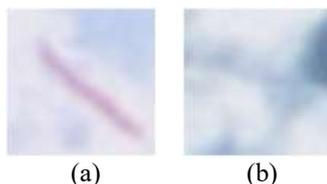
Todas as imagens usadas nesse estudo são imagens de foco estendido (EDF). As imagens de foco estendido foram obtidas pela aplicação de um método de fusão de imagens de foco múltiplo desenvolvido pelo Grupo de Pesquisa em Reconhecimento de Padrões e Otimização e publicado em COSTA et al. (2019). Nesse artigo é descrito todo o processo de fusão. Uma pilha de imagens é composta por 11 imagens do mesmo campo do microscópio. As imagens de uma pilha diferem das demais apenas porque foram adquiridas em diferentes profundidades focais. O processo de captura foi realizado da seguinte forma: a primeira imagem foi adquirida usando a função de autofoco do microscópio; então, usando um passo de distância focal de 0,52  $\mu\text{m}$ , foram adquiridas cinco imagens para cima e cinco imagens para baixo, em relação à primeira, respectivamente. Cada imagem foi anotada por três especialistas em baciloscopia de escarro para o diagnóstico da Tb.

#### 4.1.2 Banco de Dados para Segmentação Semântica – BD1

O banco de dados BD1, formado para o treinamento, validação e teste do modelo de segmentação semântica é composto por imagens mosaico, as quais são formadas pela combinação de retalhos dos campos digitais de tamanho 40×40 pixels. Os retalhos podem ser

positivos, aqueles que contém um bacilo centralizado, e negativos, aqueles que não contém bacilos. A Figura 23 mostra um exemplo de retalho positivo em (a) e um exemplo de retalho negativo em (b).

Figura 23. Exemplo de retalhos. Em (a) é mostrado um retalho positivo, e em (b) um retalho negativo.

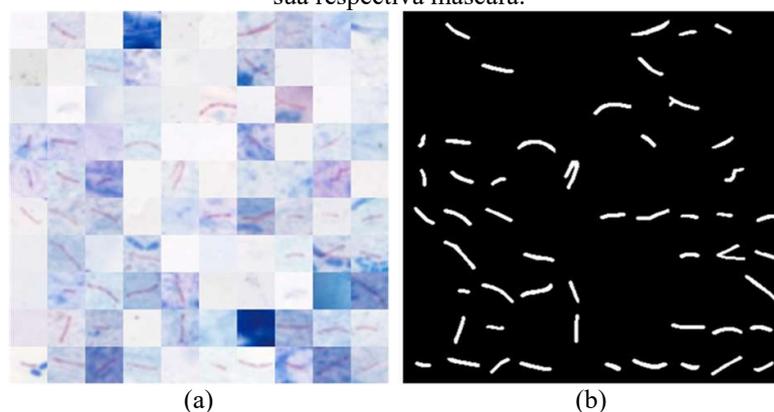


Cada retalho tem uma máscara binarizada correspondente de tamanho 40×40 pixels. Os retalhos positivos têm valor 0 para os pixels correspondente ao fundo e valor 255 para os pixels correspondentes ao bacilo. Os retalhos negativos têm o valor 0 em todos os pixels, pois todos correspondem ao fundo. As imagens dos retalhos positivos e negativo e suas respectivas máscaras compuseram as imagens mosaico do banco de dados BD1. Um total de 2636 retalhos positivos e 28848 retalhos negativos, todos das lâminas S01 a S15, foram utilizados na composição das imagens mosaico. Cada imagem é composta por 100 retalhos, e tem tamanho de 400×400 pixels. No total, foram geradas 8000 imagens mosaico e suas máscaras correspondentes. O conjunto de treinamento é formado por 6000 imagens, das quais 3000 são compostas por 50% de retalhos positivos e 50% de retalhos negativos aproximadamente, 1500 são compostas por mais de 90% de retalhos negativos e 1500 são compostas por 100% de retalhos negativos. Os conjuntos de validação e teste são formado por 1000 imagens mosaico com 50% de retalhos positivos e 50% de retalhos negativos. O Quadro 4 mostra a distribuição e composição das imagens dos conjuntos de treinamento, validação e teste. A Figura 24 mostra um exemplo de imagem mosaico e sua respectiva máscara.

Quadro 4. Distribuição das imagens de treinamento, validação e teste do banco de dados BD1.

<b>Conjunto</b>	<b>Quantidade de imagens</b>	<b>Composição das imagens</b>
Treinamento	3000	Aproximadamente 50% retalhos positivos e 50% retalhos negativos
	1500	Mais de 90% retalhos negativos
	1500	100% retalhos negativos
Validação	1000	Aproximadamente 50% retalhos positivos e 50% retalhos negativos
Teste	1000	Aproximadamente 50% retalhos positivos e 50% retalhos negativos

Figura 24. Exemplo de imagem mosaico do banco de dados BD1. Em (a) é mostrada a imagem-mosaico e em (b) sua respectiva máscara.



### 4.1.3 Banco de Dados para Diagnóstico – BD2

Em conformidade com o Quadro 1, na Seção 1, são estabelecidos cinco diagnósticos da Tb: negativo, esparso, 1+, 2+ e 3+. Cada diagnóstico é definido com a contagem de bacilos nos campos microscópicos correspondentes. O diagnóstico negativo é definido com a contagem de bacilos igual a 0, o esparso é determinado com a contagem de bacilos entre 1 e 9, o 1+ é determinado com a contagem de bacilos entre 10 e 90, o 2+ é determinado com a média de contagem entre 1 e 10 bacilos por campo microscópico, e o diagnóstico 3+ é determinado com a média de contagem maior que 10 bacilos por campo microscópico. Os primeiros três diagnósticos requerem a leitura em 100 campos microscópicos, o que corresponde a 164 campos digitais. Os diagnósticos 2+ e 3+ requerem a leitura em 50 e 20 campos microscópicos, o que corresponde a 82 e 32 campos digitais, respectivamente. O Quadro 5 resume essa correspondência entre diagnóstico da Tb em campos microscópicos e em campos digitais.

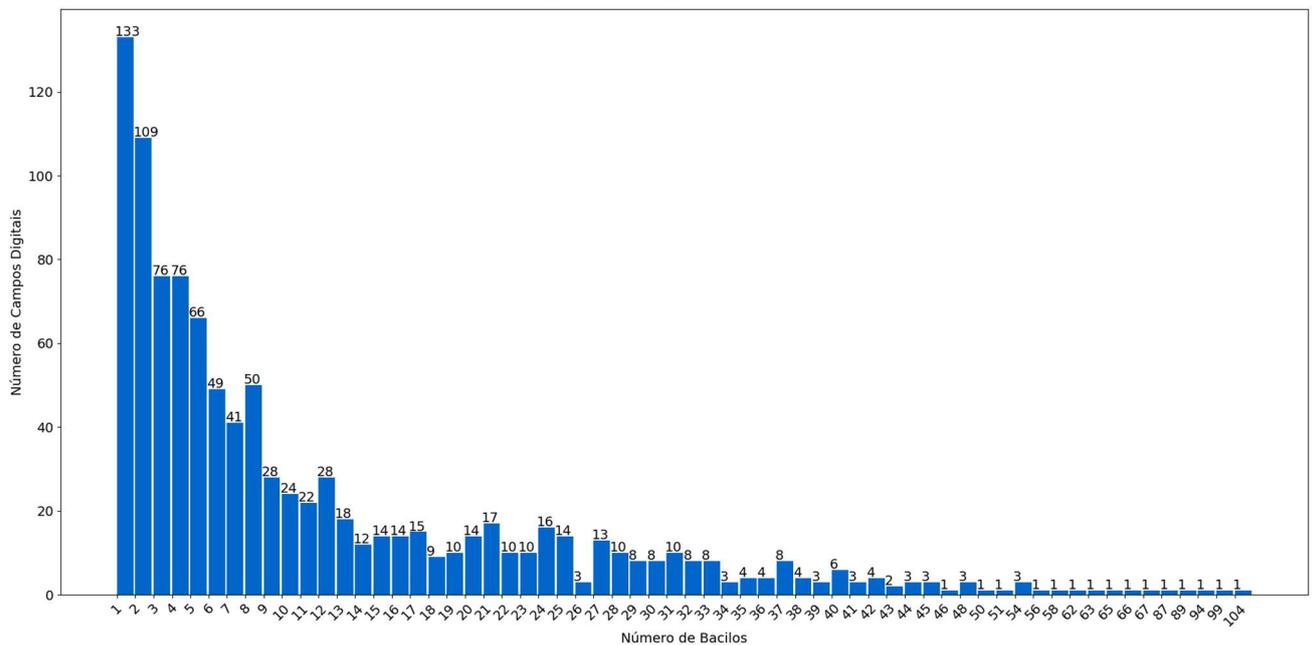
Quadro 5. Correspondência entre o diagnóstico da Tb em campos microscópicos e campos digitais.

<b>Diagnóstico</b>	<b>Campos Microscópicos</b>	<b>Campos Digitais</b>
Negativo	Nenhum bacilo em 100 campos microscópicos	Nenhum bacilo em 164 campos digitais
Esparso	Entre 1 e 9 em 100 campos microscópicos	Entre 1 e 9 bacilos em 164 campos digitais
1+	Entre 10 e 99 bacilos em 100 campos microscópicos	Entre 10 e 99 bacilos em 164 campos digitais
2+	Média entre 1 e 10 bacilos por campo em 50 campos microscópicos	Média entre 1 e 10 bacilos por campo em 82 campos digitais
3+	Média de maior que 10 bacilos por campo em 20 campos microscópicos	Média de maior que 10 bacilos por campo em 32 campos digitais

Fundamentado nessas informações, foi elaborado o banco de dados BD2 para avaliar a metodologia no diagnóstico da Tb. Para a constituição das imagens desse banco, foram utilizados campos digitais das lâminas S16 a S25. O número de bacilos por campo digital varia de 1 a 104, e um adicional de 2269 campos digitais negativos sem bacilos. A Figura 25 mostra a quantidade de campos digitais para cada variação de quantidade de bacilos. Essas quantidades de bacilos foram consideradas na construção do banco de dados BD2.

Foram gerados 50 conjuntos para cada diagnóstico. Cada conjunto representa um indivíduo a ser diagnosticado. Um total de 250 conjuntos foram gerados, sendo 150 conjuntos com 164 campos digitais, que correspondem aos diagnósticos Negativo, Esparso e 1+, 50 conjuntos com 82 campos digitais, que corresponde ao diagnóstico 2+, e 50 conjuntos com 32 campos digitais, que corresponde ao diagnóstico 3+. Cada conjunto tem a quantidade de bacilos total obedecendo a regra do diagnóstico associado mostrada no Quadro 5.

Figura 25. Número de campos digitais versus número de bacilos por campo digital usados na construção de 250 conjuntos para os diagnósticos.

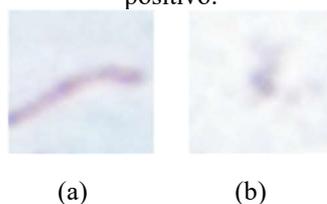


#### 4.1.4 Banco de dados Etapa de Filtragem – BD3

Para a implementação dos filtros de cor e forma, foi necessária a construção do banco de dados BD3. O banco de dados BD3 é composto por 579 retalhos, dos quais 219 são retalhos positivos classificados como bacilos pelo modelo de segmentação semântica, ou seja, considerados como verdadeiro positivo (VP), e 360 são retalhos negativos classificados como

bacilos pelo modelo de segmentação semântica, ou seja, considerados como falso positivo (FP). O objetivo de utilizar esses retalhos foi o de desenvolver os filtros com exemplos que são classificados erroneamente como bacilos pelo modelo de segmentação, e assim reduzir o número de falsos positivos. A Figura 26 mostra dois exemplos de amostras do banco de dados BD3: (a) mostra um exemplo de retalho verdadeiro positivo e (b) um exemplo de retalho falso positivo. O BD3 foi dividido em 80% para treinamento e 20% para teste.

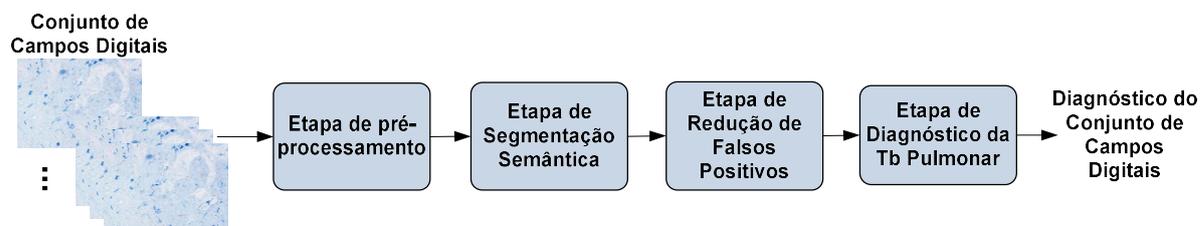
Figura 26. Exemplos do banco de dados BD3. Em (a) é um exemplo de verdadeiro positivo, e em (b) de falso positivo.



## 4.2 Métodos

Com a finalidade de identificar bacilos em campos digitais e prever o diagnóstico a partir da contagem dos bacilos detectados, este trabalho aborda uma metodologia dividida em quatro etapas: pré-processamento, segmentação semântica, filtragem e contagem dos bacilos. A Figura 27 mostra o fluxograma da metodologia deste trabalho.

Figura 27. Metodologia da detecção dos bacilos e diagnóstico da Tb.



Na etapa de pré-processamento a imagem do campo digital é dividida em imagens de  $400 \times 400$  pixels.

Na etapa de segmentação semântica, os bacilos são segmentados usando o modelo CNN, denominado CNN1. Em seguida, áreas com problemas de coloração e objetos muito pequenos ou muito grandes são removidos.

Na etapa de redução de falsos positivos, dois filtros são aplicados: um filtro de cor, implementado com CNN, denominado por CNN2; e um filtro de forma implementado com um classificador RF, denominado por RF1. Esses filtros são implementados com o objetivo de reduzir o número de falsos positivos obtidos na etapa de segmentação semântica, explorando um conjunto específico das propriedades dos objetos candidatos a bacilo e da sua vizinhança.

Conforme mostrado mais adiante, na seção 4.2.4, cada filtro classifica um objeto segmentado como bacilo ou não, e são aplicados em cascata.

#### 4.2.1 Métricas de Avaliação

Para avaliar o desempenho da CNN1, CNN2 e RF1, as métricas aplicadas foram: sensibilidade, especificidade, precisão e acurácia. A CNN1 foi treinada com imagens mosaico formadas por retalhos de 40×40 pixels. Esses retalhos podem ou não conter bacilos, portanto, na etapa de segmentação semântica é possível avaliar a taxa de verdadeiro negativo (VN). Da mesma forma, a CNN2 e o RF1 foram treinados com retalhos FP e VP, sendo possível avaliar a taxa de verdadeiro negativo (VN).

Para avaliar o desempenho do método no diagnóstico da Tb, utilizando as imagens originais de microscopia de campo claro, a sensibilidade, precisão e *f1-score* podem ser calculadas, porém a especificidade e acurácia não podem ser calculadas, uma vez que não é possível determinar a taxa de verdadeiro negativo (VN). As métricas são definidas pelas equações abaixo:

$$\textit{sensibilidade} = \frac{VP}{VP+} \quad (19)$$

$$\textit{especificidade} = \frac{VN}{VN+FP} \quad (20)$$

$$\textit{precisão} = \frac{VP}{VP+FP} \quad (21)$$

$$\textit{acurácia} = \frac{VP+VN}{VP+VN+FP+F} \quad (22)$$

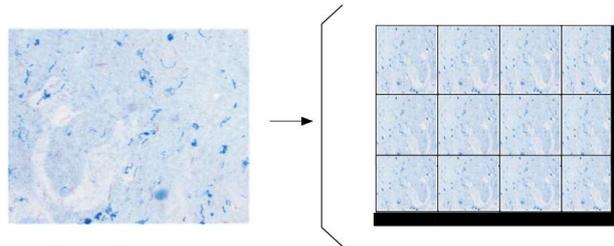
$$F1 - \textit{score} = 2 * \frac{\textit{Precisão} * \textit{Sensibilidade}}{\textit{Precisão} + \textit{Sensibilid}} \quad (23)$$

Onde: VP – Verdadeiro Positivo; FP – Falso Positivo; VN – Verdadeiro Negativo e FN – Falso Negativo.

#### 4.2.2 Pré-processamento

Na etapa de pré-processamento, cada campo digital de 1040×1388 pixels é dividido em uma grade de 3×4 imagens de tamanho 400×400 pixels, com as últimas linhas e últimas colunas preenchidas com zeros. A Figura 28 mostra essa divisão em um campo digital. Essas imagens, de 400×400 pixels, são usadas como entrada CNN1 que implementa a etapa de segmentação semântica.

Figura 28. Divisão de um campo digital de 1040×1388 pixels em uma grade de 3×4 imagens de 400×400 pixels.

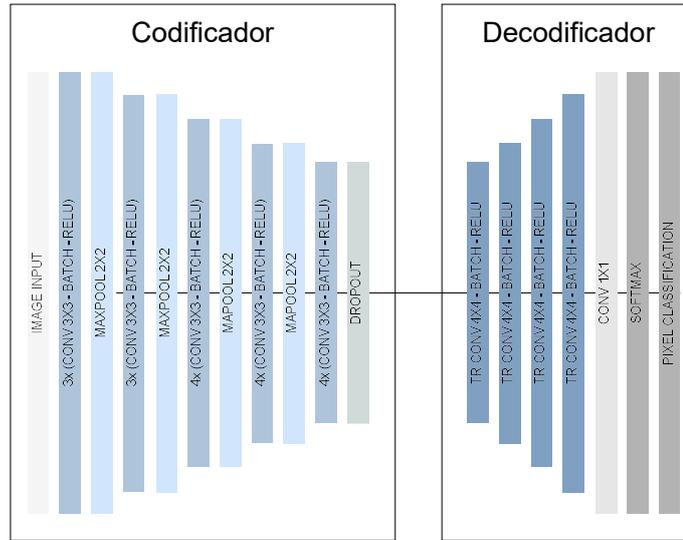


### 4.2.3 Segmentação Semântica

O modelo de segmentação semântica visa segmentar bacilos em imagens de microscopia. Em trabalho anteriormente desenvolvido (SERRAO et al., 2020), foi proposto utilizar uma arquitetura profunda denominada *autoencoder*, para segmentação de bacilo da Tb, utilizando imagens mosaico como entrada. A arquitetura do *autoencoder* é constituída por dois blocos: codificador (*encoder*) e decodificador (*decoder*). No trabalho citado, foram avaliadas 3 arquiteturas de *autoencoders*, três técnicas de otimização e 2 métodos para melhoria da generalização. A primeira arquitetura foi construída com 18 camadas de convolução, a segunda com 6 camadas de convolução e a terceira com 2 camadas de convolução. Cada arquitetura foi treinada usando as técnicas de otimização: SGDM, RMSProp e ADAM. E foram utilizados os métodos de generalização regularização  $L_2$  e a camada *dropout* individualmente e combinados. A arquitetura com melhor desempenho foi a com maior quantidade de camadas convolutivas. A Figura 29 mostra essa arquitetura. O codificador é composto por uma camada de entrada (que recebe uma imagem mosaico), 4 etapas de subamostragem, uma camada *dropout* com 0,3 de taxa. Cada etapa de subamostragem consiste em sequências de camadas: convolutiva com *kernel* 3×3, *batch normalization* e camada de ativação ReLU. Após cada sequência, há uma camada de *maxpooling* com tamanho 2×2. O decodificador é composto por 4 etapas de sobreamostragem, uma cada Softmax e uma camada de saída. Cada etapa de sobreamostragem consiste em uma camada de convolução transposta com *kernel* 4×4, camada de *batch normalization* e camada de ativação ReLU.

Os hiper parâmetros de treinamento foram: taxa de aprendizado inicial 0,001; fator de queda da taxa de aprendizado 0,5; máximo número de épocas 20; tamanho do lote 10; e critério de parada 6000 iterações. A função de perda utilizada foi a *cross-entropy*, que é o logaritmo negativo da probabilidade da classe verdadeira, conforme a equação (3) mostrada anteriormente.

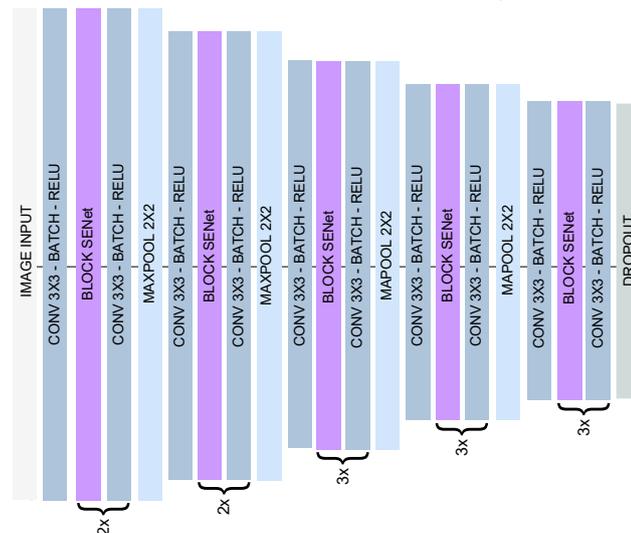
Figura 29. Arquitetura proposta em SERRAO et al. (2020).



Neste trabalho de tese, prospectou-se a utilização de duas novas arquiteturas para o codificador (*encoder*), utilizando mecanismo de atenção por canal SENet e convolução separada por canal (*depthwise separable convolution*).

O codificador com mecanismo de atenção de canal, ilustrado na Figura 30, consiste em uma camada de entrada, 4 etapas de subamostragem e uma camada *dropout* com 0,3 de taxa. Cada etapa de subamostragem consiste em uma camada de convolução com *kernel* 3×3 seguido por sequências de: bloco SENet e uma camada de convolução com *kernel* 3×3. Uma camada *maxpooling* com tamanho 2×2 é usada após essas sequências. Todas as camadas de convolução são seguidas por uma camada de *batch normalization* e uma camada de ativação ReLU.

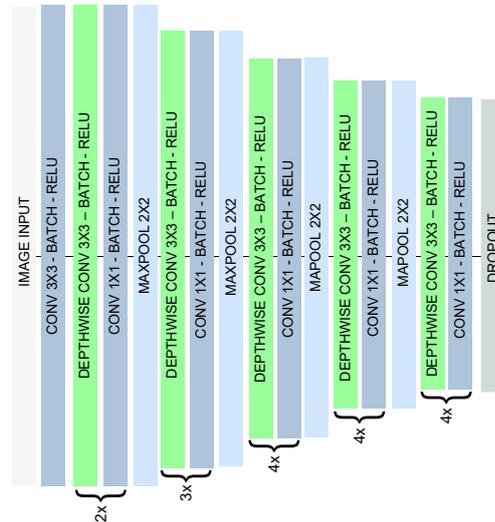
Figura 30. Codificador usando mecanismo de atenção de canal SENet.



O codificador com convolução separada por canal, ilustrado na Figura 31, consiste em uma camada de entrada seguida por uma camada de convolução com *kernel* 3×3, 4 etapas de subamostragem e uma camada *dropout* com 0,3 de taxa. Cada etapa de subamostragem consiste

em sequências de: camada *depthwise convolution* com *kernel*  $3 \times 3$  e uma camada de convolução com *kernel*  $1 \times 1$ . Após as sequências, há uma camada *maxpooling* com tamanho  $2 \times 2$ . Todas as camadas de convolução e *depthwise convolution* são seguidas por uma camada de *batch normalization* e uma camada de ativação ReLU.

Figura 31. Codificador usando *depthwise separable convolution*.



Nesse trabalho de tese, avaliou-se também a segmentação semântica dos bacilos através da rede U-Net. A arquitetura U-Net (RONNEBERGER; FISCHER; BROX, 2015) é amplamente utilizada em tarefas de segmentação semântica, particularmente no processamento de imagens médicas. Ela foi implementada para realizar segmentação precisa mesmo com poucas imagens de treinamento. A arquitetura consiste em um codificador com 4 etapas de subamostragem. Cada etapa de subamostragem é composta por duas camadas de convolução com *kernel*  $3 \times 3$  seguidas por uma camada *maxpooling* com tamanho  $2 \times 2$ . O decodificador consiste em 4 etapas de sobreamostragem. Cada etapa de sobreamostragem por uma camada de convolução transposta  $2 \times 2$ , uma concatenação com o mapa de características correspondente do codificador, e duas camadas de convolução com *kernel*  $3 \times 3$ . Esta arquitetura conecta camadas codificadoras e camadas decodificadoras. Os autores avaliaram a U-Net para segmentação de estruturas neuronais em registros microscópicos eletrônicos e segmentação celular em imagens microscópicas usando dois conjuntos de dados diferentes. Um bom desempenho foi alcançado em todas as tarefas avaliadas.

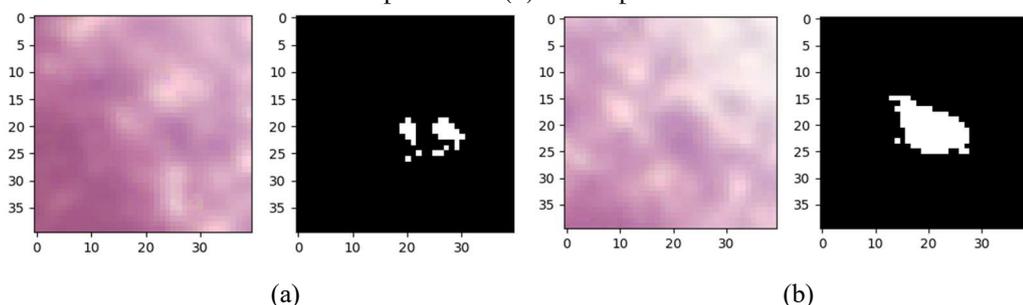
Os resultados de cada arquitetura avaliada são explorados na Seção 5.1.

Como mencionado anteriormente, ainda que o modelo de segmentação tenha especificidade maior que 99%, este valor ainda gera muitos falsos positivos, especialmente em imagens sem bacilos, que corresponde ao diagnóstico negativo da Tb. Como a rede de

segmentação semântica trabalha na classificação de pixel, é possível que um único pixel ou que poucos pixels, devido à presença de corpos estranhos no esfregaço ou problemas de coloração inferior a 50% da janela de 40×40 pixels, tenham características de cor semelhantes aos pixels dos bacilos e podem ser classificados como tal. Portanto, um filtro de área é aplicado para eliminar falsos positivos detectados pelo modelo de segmentação com área menor que 25 pixels e maior que 900 pixels, uma vez que correspondem a artefatos.

A preparação da lâmina é realizada para que os bacilos da Tb assumam uma coloração vermelha e os demais elementos presentes na lâmina uma coloração azul, de tal forma que os bacilos da Tb se destaquem e o microscopista realize a contagem deles. Porém, durante esse processo, podem ocorrer problemas de coloração que resultam em áreas dessa lâmina com a coloração dos bacilos, e não somente os bacilos da Tb. Uma vez que isso pode gerar erros pelo método desenvolvido, é necessário identificar regiões com problemas de coloração. Para isso, é investigado o valor do matiz (H) na vizinhança de todos os objetos segmentados. Uma janela de 40×40 pixels é centralizada em cada objeto segmentado e uma operação de limiar usando a componente H é aplicada nesta janela. Se mais de 50% dos pixels da janela tiver um valor H maior de 266°, uma vez que a solução utilizada na preparação das lâminas tem a coloração rosa escuro a vermelho brilhante, então a janela é considerada como uma região com problema de coloração e é substituída por uma região da mesma imagem que não apresenta problema de coloração. A Figura 32 mostra dois exemplos de objetos segmentados em área com problema de coloração.

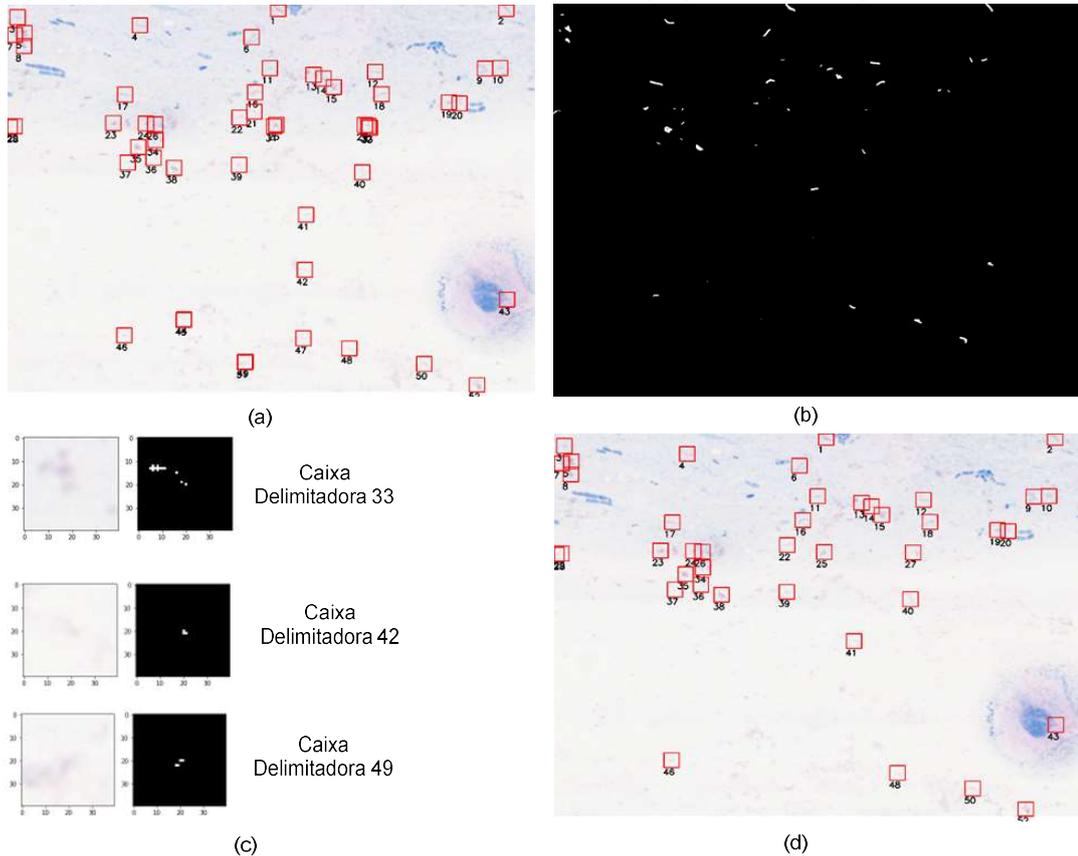
Figura 32. Exemplos de objetos segmentados em regiões com problemas de coloração. Em (a) é mostrado o exemplo 1 e em (b) o exemplo 2.



A Figura 33 (a) mostra uma imagem de entrada e o resultado da etapa de segmentação semântica, onde objetos candidatos a bacilo são evidenciados por uma caixa delimitadora. Algumas caixas delimitadoras estão sobrepostas porque os objetos detectados estão muito próximos. Por exemplo, as caixas que estão com sobreposição são: 44 e 45; 48 e 49; 27, 30, 32 e 33; 25 e 31; 24 e 26. Na Figura 33 (b) é mostrado resultado da segmentação semântica em

uma imagem binarizada, onde pixels brancos foram classificados como bacilo pela rede e pixels pretos foram classificados como não bacilos.

Figura 33. Ilustra a etapa de segmentação semântica. Em (a) é mostrada a imagem de saída da rede com cada objeto candidato a bacilo evidenciado em uma caixa delimitadora. Em (b) a imagem binarizada da saída da rede com os pixels de candidatos a bacilos com nível 1. Em (c) exemplos de candidatos a bacilos removidos pelo filtro de área. E em (d) imagem com candidatos a bacilos que permaneceram após o filtro de área.



A Figura 33 (c) mostra exemplos de três objetos segmentados mostrados na Figura 33 (a), caixas delimitadoras 33, 42 e 49, que correspondem a falsos positivos e foram removidos após a aplicação do filtro de área. A Figura 33 (d) mostra o resultado da etapa de segmentação semântica, após o filtro de área remover todos os objetos considerados artefatos. Cada bacilo candidato está destacado por uma caixa delimitadora vermelha.

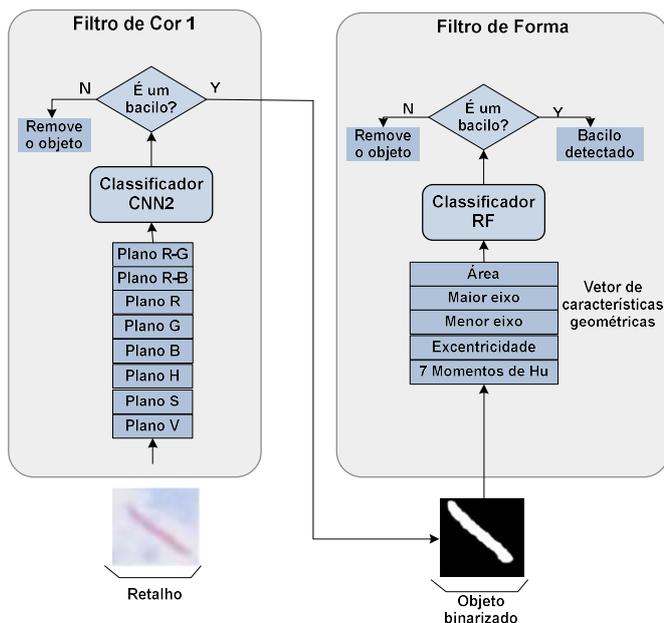
#### 4.2.4 Redução de Falsos Positivos

Com o objetivo de reduzir os falsos bacilos, cada objeto remanescente da etapa de segmentação semântica é submetido a um processo de filtragem, que utiliza filtros de cor e forma. O modelo de filtragem 1 é composto por um filtro de cor e um filtro de forma. O modelo de filtragem 2 é composto por dois filtros de cor e um filtro de forma.

Os filtros de cor foram implementados utilizando a mesma arquitetura de rede neural convolutiva, diferenciando-se pela entrada da rede.

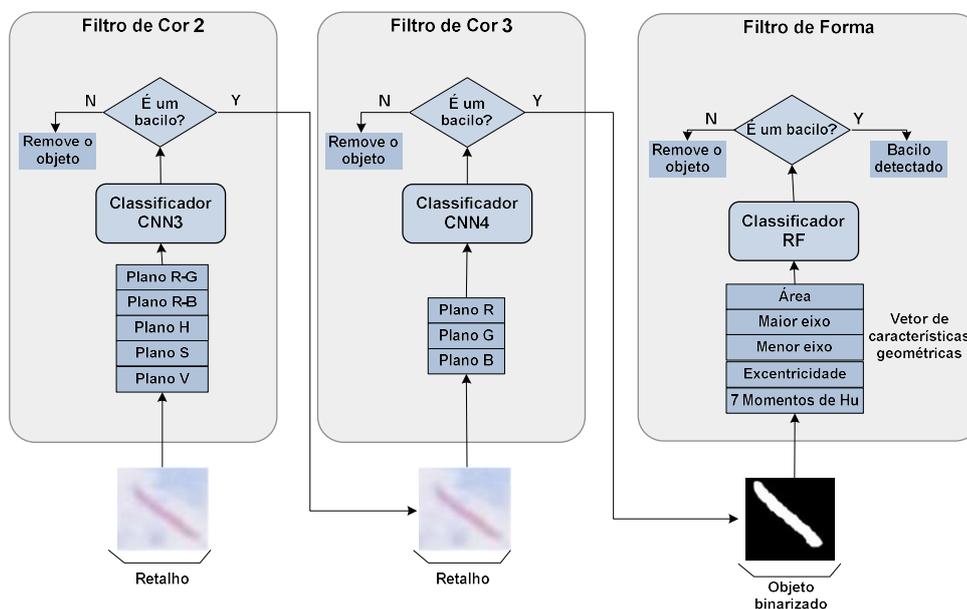
No modelo de filtragem 1, é usado um filtro de cor, denominado por CNN2, que foi implementado com uma entrada de 8 planos: R-G, R-B, R, G, B, H, S e V. Seguido por um filtro de forma, denominado por RF1, implementado com um classificador *Random Forest*, que tem como entrada um vetor de 11 características. A Figura 34 mostra o diagrama do modelo de filtragem 1.

Figura 34. Diagrama do modelo de filtragem 1.



No modelo de filtragem 2, o primeiro filtro de cor, denominado por CNN3, foi implementado com uma entrada de 5 planos: R-G, R-B, H, S e V, e o segundo filtro de cor, denominado por CNN4, foi implementado com uma entrada de 3 planos: R, G e B. Seguidos pelo mesmo filtro de forma RF1. A Figura 35 mostra o diagrama do modelo de filtragem 2.

Figura 35. Diagrama do modelo de filtragem 2.

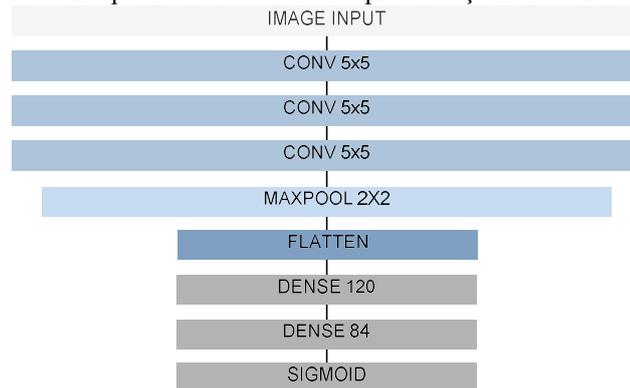


#### 4.2.4.1 Filtros de Cor

Essa etapa consiste no refinamento da segmentação semântica, que visa reduzir os falsos positivos por meio de filtros que focam nas características de cor dos objetos segmentados. A mesma arquitetura foi utilizada na implementação dos filtros de cor dos modelos de filtragem 1 e 2. Ela consiste em três camadas convolutivas com *kernel*  $5 \times 5$ , uma camada *pooling* de  $2 \times 2$ , uma camada *flatten*, uma camada densa de 120 neurônios, uma camada densa de 84 neurônios e uma camada de saída com a função sigmoid. A arquitetura está ilustrada na Figura 36.

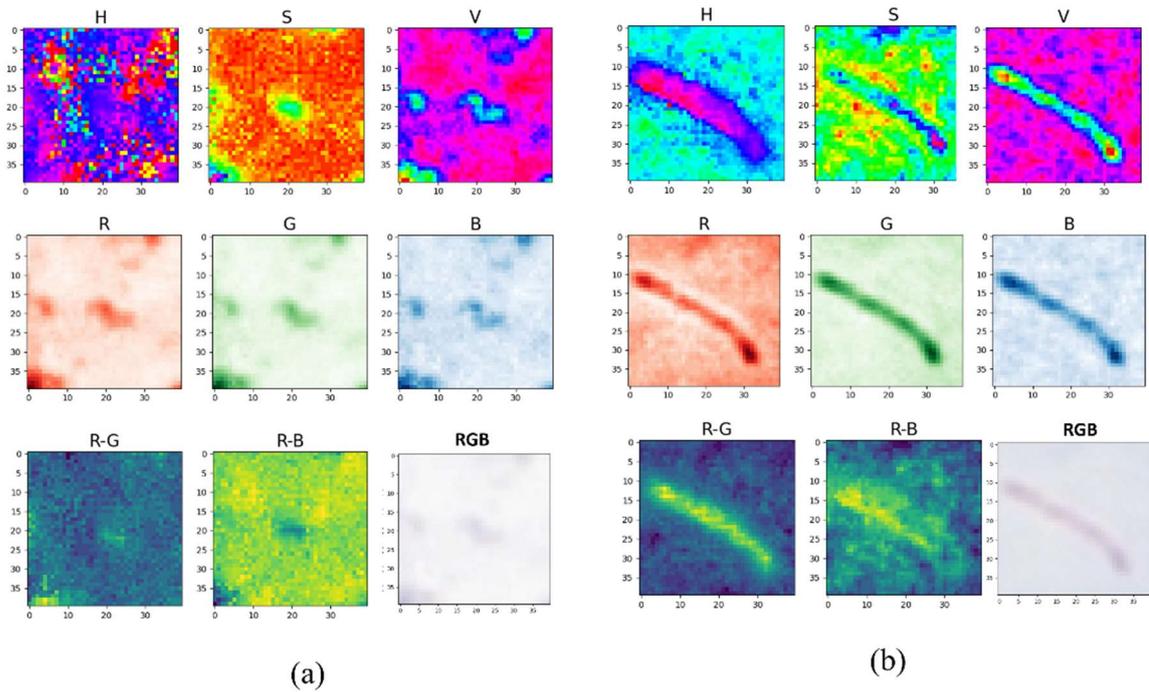
Os filtros de cor implementados diferenciam-se pela entrada, que combinam os espaços de cor RGB e HSV. O filtro de cor do modelo de filtragem 1 utiliza uma entrada de 8 planos: R-G, R-B, R, G, B, H, S, V, e é denominado de CNN2. Os filtros de cor do modelo de filtragem 2 utilizam uma entrada de 5 planos: R-G, R-B, H, S e V, denominado CNN3, e uma entrada de 3 planos: RGB, denominado CNN4.

Figura 36. Arquitetura utilizada na implementação dos filtros de cor.



O espaço de cor RGB foi utilizado neste trabalho porque é o espaço original no qual as imagens são representadas, e porque foi usado em outros estudos publicados anteriormente na literatura. A importância da característica R-G para segmentação de bacilo já foi verificada no trabalho anterior COSTA et al. (2008). O espaço de cor HSV foi escolhido por possuir características como o sistema visual humano, no qual a informação de cor (H e S) é desacoplada da informação de intensidade (V). A Figura 37 mostra dois exemplos de retalhos do banco de dados BD3, um falso positivo e um verdadeiro positivo, e ilustra todos os planos utilizados nas entradas dos filtros de cor de cada um desses retalhos: em (a) são mostrados os planos de um retalho FP mais o retalho em RGB e em (b) os planos de um retalho VP mais o retalho em RGB. Pode-se observar que objetos detectados sem a coloração de um bacilo se diferenciam principalmente nos planos R-G, R-B, H, S e V, além disso, a informação sobre o contexto do objeto detectado também é aprendida pelos modelos treinados com os planos R, G e B.

Figura 37. Ilustração de todos os planos utilizados na entrada das redes CNN2, CNN3 e CNN4 de um retalho. Em (a) é mostrado os planos de um retalho FP e em (b) um retalho VP.



A Figura 38 exemplifica a etapa do filtro de cor do modelo de filtragem 1. E a Figura 39 exemplifica a etapa do filtro de cor com o modelo de filtragem 2.

Figura 38. Aplicação do filtro de cor do modelo de filtragem 1. Em (a) exemplos de objetos candidatos a bacilo removidos pelo filtro de cor CNN2, que correspondem às caixas delimitadoras 3, 11, e 36; (b) imagem com os objetos remanescentes após o filtro de cor do modelo de filtragem 1.

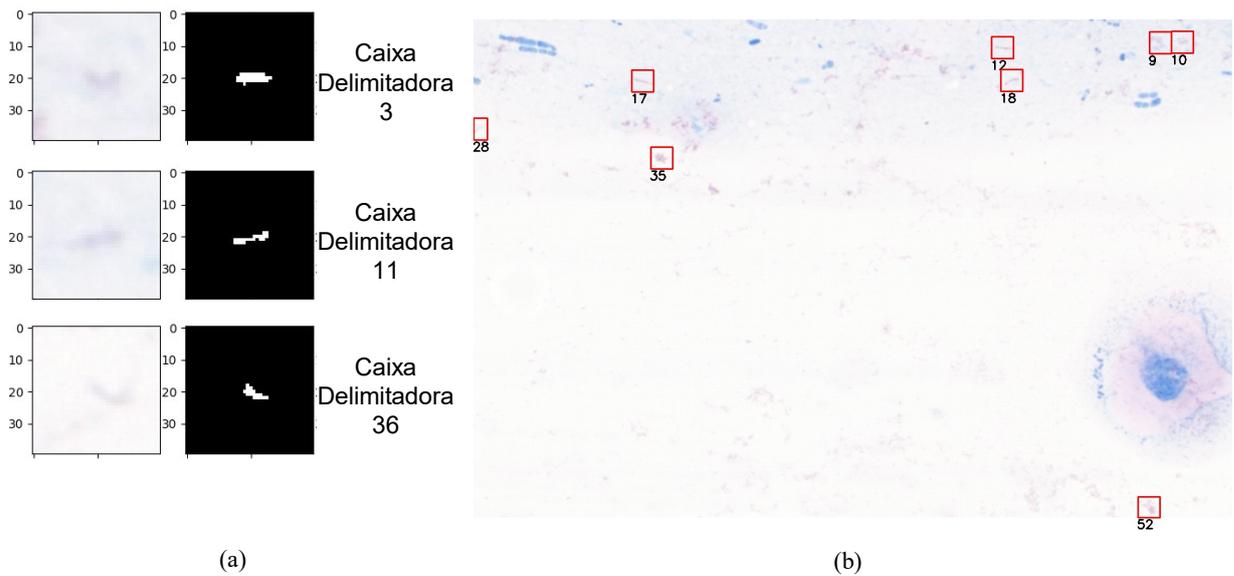
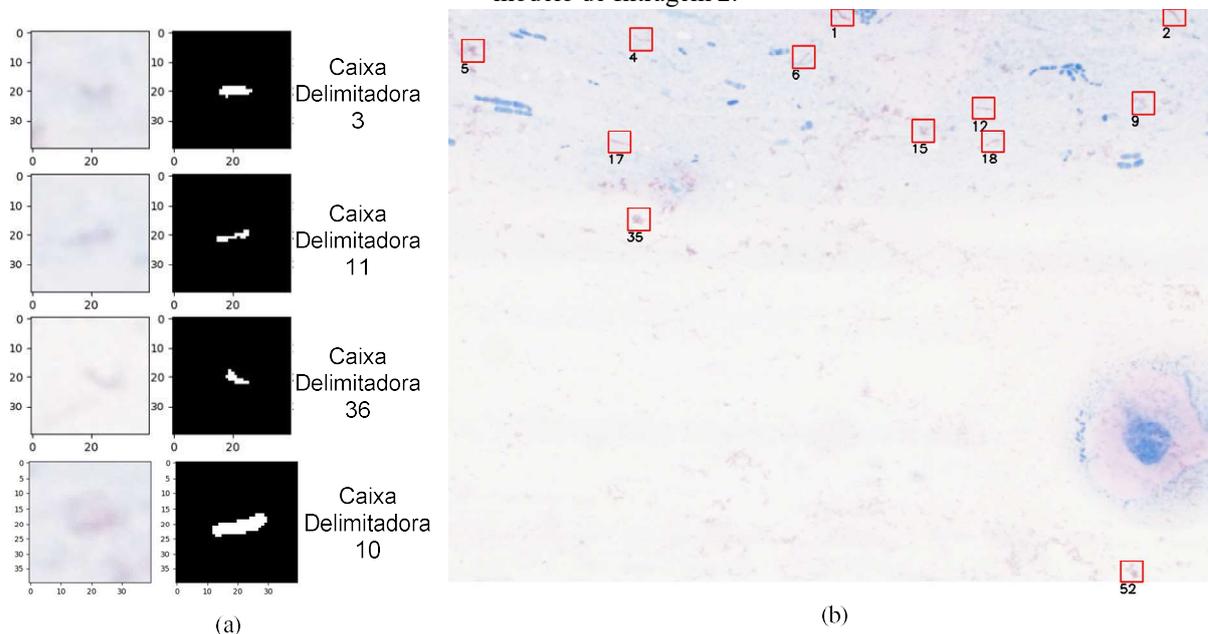


Figura 39. Aplicação dos filtros de cor do modelo de filtragem 2. Em (a) exemplos de objetos candidatos a bacilo removidos pelo filtro de cor CNN3, que correspondem às caixas delimitadoras 3, 11, e 36, e pelo filtro CNN4, que corresponde à caixa delimitadora 10; (b) imagem com os objetos remanescentes após os filtros de cor do modelo de filtragem 2.



A Figura 38 (a) mostra três objetos segmentados extraídos da Figura 33 (a), as caixas delimitadoras 3, 11 e 36, correspondendo a falsos positivos, que foram removidos pelo filtro de cor CNN2. A Figura 38 (b) mostra os objetos segmentados remanescentes do filtro de cor no modelo de filtragem 1.

A Figura 39 (a) mostra três objetos segmentados extraídos da Figura 33 (a), as caixas delimitadoras 3, 11 e 36, correspondendo a falsos positivos, que foram removidos pelo filtro de cor CNN3, e a caixa delimitadora 10, correspondendo a um falso positivo removido pelo filtro de cor CNN4. E a Figura 39 (b) mostra os objetos segmentados remanescentes dos filtros de cor no modelo de filtragem 2.

Embora esses objetos tenham a forma de bastonete como um bacilo, eles não têm a coloração avermelhada do bacilo e foram eliminados na etapa do filtro de cor.

A Tabela 1 mostra os resultados dos filtros de cor implementados e avaliados com o banco de dados BD3.

Tabela 1. Resultados dos filtros de cor CNN2, CNN3 e CNN4.

Modelo	Acurácia (%)	Precisão (%)	Sensibilidade (%)	Especificidade (%)
CNN2	92,45	90,62	85,29	95,83
CNN3	93,39	96,55	82,35	98,61
CNN4	91,50	85,71	88,23	93,05

#### 4.2.4.2 Filtro de Forma

A informação de cor foi considerada no processamento descrito anteriormente. Nessa etapa, o objetivo é focar a atenção na forma dos objetos segmentados. Para cada objeto segmentado, um conjunto de 11 características geométricas são extraídas: área, maior eixo, menor eixo, excentricidade e 7 momentos de Hu que são invariantes à translação, escala e rotação. Em (YOUSEFI et al., 2020) os autores também usam momentos de Hu para detecção de bacilos. As características citadas acima são descritas a seguir:

- Maior eixo e menor eixo: distância máxima e mínima do limite do bacilo ao centro de massa  $(\bar{x}, \bar{y})$ , correspondendo ao  $R_{max}$  e  $R_{min}$  na Figura 40.
- Excentricidade: a razão entre o eixo maior e eixo menor, conforme a equação (24).

$$excentricidade = \frac{R_{max}}{R_{min}} \quad (24)$$

- Momentos de Hu: são descritores utilizados para caracterizar a forma de um objeto binário em uma imagem. Existem 7 momentos Hu. Para calcular os momentos de Hu, inicialmente definimos os momentos relativos, que são centrados em torno do baricentro, conforme a equação (25).

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p \cdot (y - \bar{y})^q \cdot I(x, y) \quad (25)$$

Onde:

$(\bar{x}, \bar{y})$  – coordenadas do centróide

$p$  e  $q$ : ordem dos momentos nas direções  $x$  e  $y$

$I(x, y) \in \{0,1\}$

- Os 7 momentos de Hu são definidos pelas equações de (26) a (32).

$$M_1 = \mu_{20} + \mu_{02} \quad (26)$$

$$M_2 = (\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2 \quad (27)$$

$$M_3 = (\mu_{30} - 3\mu_{12})^2 + (3\mu_{21} - \mu_{30})^2 \quad (28)$$

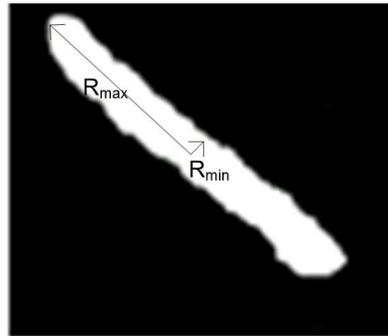
$$M_4 = (\mu_{30} + \mu_{12})^2 + (\mu_{21} + \mu_{03})^2 \quad (29)$$

$$M_5 = (\mu_{30} - 3\mu_{12})(\mu_{30} + \mu_{12})((\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2) + (3\mu_{21} - \mu_{03})(\mu_{21} + \mu_{03})(3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2) \quad (30)$$

$$M_6 = (\mu_{20} - \mu_{02})((\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2) + 4\mu_{11}(\mu_{30} + 3\mu_{12})(\mu_{21} + \mu_{03}) \quad (31)$$

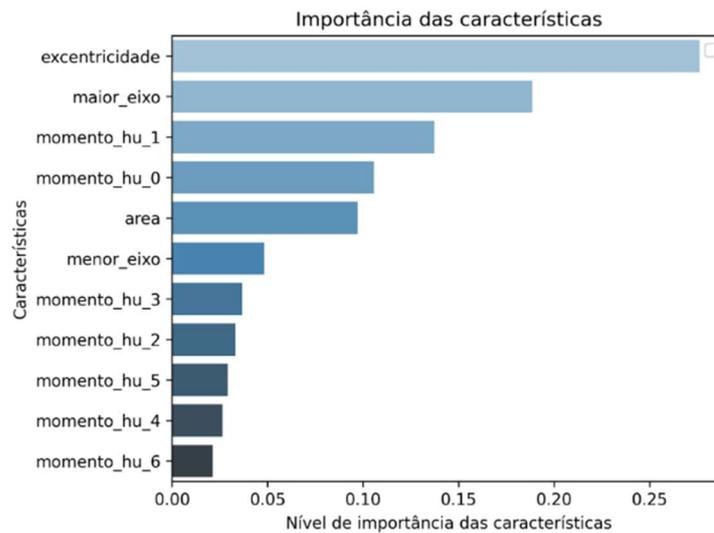
$$M_7 = (3\mu_{21} - \mu_{03})(\mu_{30} + \mu_{12})((\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2) - (\mu_{30} - 3\mu_{12})(\mu_{21} + \mu_{03})(3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2) \quad (32)$$

Figura 40. Imagem binarizada de um bacilo com o formato alongado, mostrando o maior eixo  $R_{max}$  e menor eixo  $R_{min}$ .



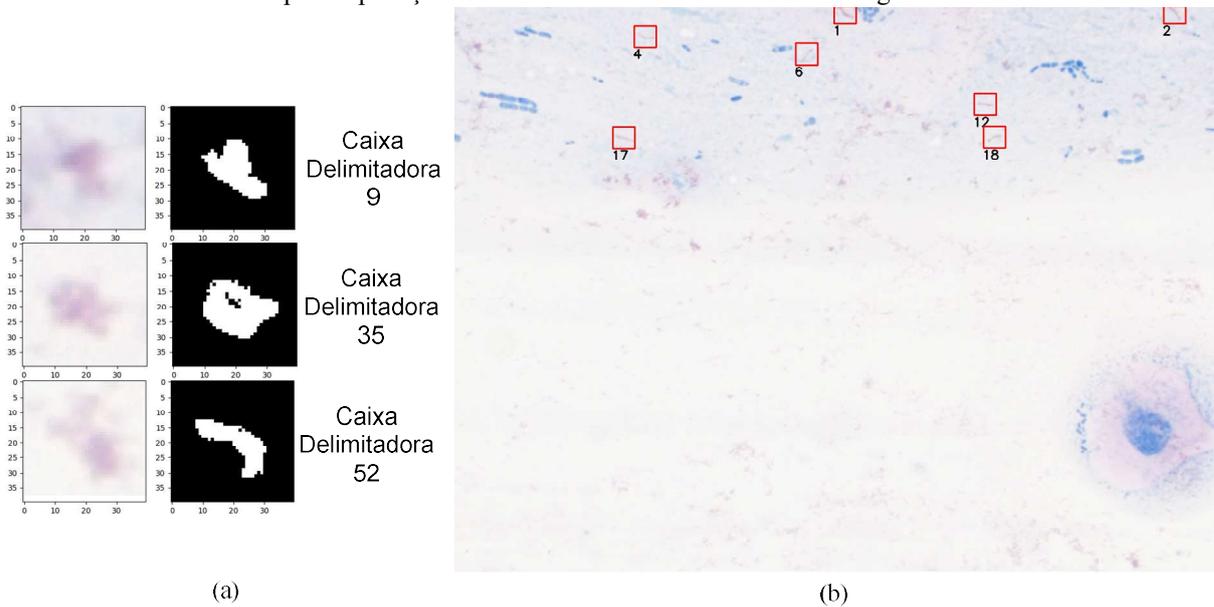
Um classificador *Random Forest*, denominado por RF1, foi treinado com o banco de dados BD3 para reconhecer bacilos e falsos positivos. Os parâmetros de treinamento do RF1 são: profundidade da árvore igual a 10 e número máximo de nós igual a 100. A Figura 41 mostra o nível de importância de todas as características utilizadas na entrada para a decisão do classificador, baseado na diminuição média das impurezas dos nós de cada árvore quando cada característica é usada para fazer a divisão. A excentricidade é a variável com maior influência na decisão do classificador. A segunda característica mais importante é o maior eixo do objeto, o que está de acordo com o fato de que o bacilo é uma forma alongada de bastonete.

Figura 41. Importância das características geométricas utilizadas na entrada do classificador RF1. Excentricidade é a característica com maior influência na decisão do classificador, seguida pela característica de maior eixo.



A Figura 42 (a) mostra exemplos de objetos segmentados extraídos da Figura 33 (a), caixas delimitadoras 9, 35 e 52, correspondendo a falsos positivos que foram removidos pela aplicação dos filtros de forma. A Figura 42 (b) mostra os objetos segmentados remanescentes após a aplicação do filtro de forma. Sendo essa a etapa final de filtragem.

Figura 42. Aplicação do filtro de forma. Em (a) exemplos de bacilos candidatos removido pelo filtro de forma, que correspondem às caixas delimitadoras 9, 35 e 52. Em (b) imagem com os bacilos candidatos remanescentes após a aplicação dos filtros de forma do modelo de filtragem 2.



A aplicação sequencial dos filtros resultou em uma redução progressiva de falsos positivos. Observa-se que, o filtro de cor foi responsável pela maior redução de falsos positivos, tanto no modelo de filtragem 1 quanto no modelo de filtragem 2. Isso pode ser observado comparando a Figura 33 (b), Figura 38 (b) e Figura 39 (b) e Figura 42 (b). O filtro de forma foi implementado e avaliado com o banco de dados BD3 e alcançou uma acurácia de 95,24%, precisão de 93,10%, sensibilidade de 96,13% e especificidade de 94,28%.

#### 4.2.5 Ambiente de Desenvolvimento

As simulações foram feitas em um computador com um processador Intel Core i5-9300H, CPU 2.40GHz, 8GB RAM, processador de unidade gráfica NVIDIA GeForce GTX 1650 (4 GB de RAM). O trabalho foi desenvolvido no ambiente Google Collaboratory, usando a versão do Python 3.7.13, Tensorflow 2.8.2. As bibliotecas skimage na versão 0.18.3, sklearn na versão 1.0.2 e Pandas na versão 1.3.5 foram usadas para desenvolver os filtros de cor e forma. Manipulações das imagens foram realizadas usando as bibliotecas NumPy na versão 1.21.6, OpenCV na versão 4.1.2 e Matplotlib na versão 3.2.2.

## 5 RESULTADOS E DISCUSSÃO

Neste capítulo serão apresentados os resultados obtidos. O modelo de segmentação semântica foi treinado e testado com o BD1. Assim como os modelos da etapa de filtragem foram implementados e testados com o conjunto de dados BD3. O diagnóstico da Tb, utilizando o banco de dados BD2, foi avaliado empregando os seguintes modelos: segmentação semântica e filtro de área, segmentação semântica + o modelo de filtragem 1 e segmentação semântica + modelo de filtragem 2.

### 5.1 Resultados

A Tabela 2 mostra o desempenho dos modelos de segmentação semântica dos bacilos e compara-os com resultados obtidos anteriormente. Conforme mostrado na Tabela 2, os melhores valores para acurácia, precisão, especificidade e *f1-score* foram obtidos com a rede usando camadas de convolução separada por canal no codificador do autoencoder. A melhor sensibilidade foi obtida com utilização do mecanismo de atenção de canal SENet no codificador do autoencoder. A rede U-Net apresentou os piores resultados, com exceção da sensibilidade, onde alcançou um valor melhor que o resultado apresentado em SERRAO et al. (2020).

Tabela 2. Resultados das arquiteturas avaliadas para a tarefa de segmentação semântica.

Rede de Segmentação Semântica	VP	FP	VN	FN	Acc (%)	Prec. (%)	Sens. (%)	Espec. (%)	<i>F1-score</i> (%)
Arquitetura proposta em (SERRAO et al., 2020) com convoluções regulares	49753	167	49657	536	99,30	99,67	98,93	99,66	99,30
Arquitetura com codificador usando convolução separada por canal	50064	140	49610	225	<b>99,64</b>	<b>99,72</b>	99,55	<b>99,72</b>	<b>99,64</b>
Arquitetura com codificador usando bloco de mecanismo de atenção SENet	50116	373	49514	173	99,45	99,26	<b>99,66</b>	99,25	99,46
Arquitetura U-Net	50016	2787	49162	273	98,97	98,56	99,40	98,55	98,98

O valor mínimo de falsos positivos, 140, obtido com a rede usando camadas de convolução separada por canal no codificador do autoencoder, ainda é muito alto. Para minimizar o número de falsos positivos, foram utilizados os filtros descritos na Seção 4.2.4.

Para avaliar os desempenhos desses filtros, foram selecionados 10 conjuntos de 164 imagens negativas de campo digital do banco de dados BD2. Nesses conjuntos foram então contabilizados o número de bacilos detectados. Em seguida, obteve-se a média e o desvio padrão (DP) dos valores obtidos para cada conjunto. A Tabela 3 mostra a redução do número de FPs com os modelos de filtragem 1 e 2. Conforme mostrado, o filtro de cor é responsável pela maior redução nos FPs. Ambos os métodos resultaram em uma redução superior a 96% no número de FPs. Nota-se também que a redução de FPs foi superior com o modelo de filtragem 2 com uma média de  $10 \pm 3$  FPs. Portanto, em um conjunto de 164 campos digitais, o método proposto apresentou um viés positivo médio de 10 FPs.

Tabela 3. Redução no número de FPs resultante do uso dos 2 modelos de filtragem, em 10 conjuntos de 164 imagens do conjunto de dados BD2.

<b>Etapa</b>	<b>Falsos Positivos (média <math>\pm</math> DP)</b>
Segmentação semântica	$546 \pm 125$
Segmentação semântica + modelo de filtragem 1	$20 \pm 7$
Segmentação semântica + modelo de filtragem 2	$10 \pm 3$

### 5.1.1 Resultados do Diagnóstico

Conforme dito anteriormente, para o diagnóstico da Tb, foi avaliado o resultado obtido com o banco de dados BD2 considerando a utilização dos seguintes modelos: segmentação semântica e filtro de área, segmentação semântica + o modelo de filtragem 1 e segmentação semântica + modelo de filtragem 2. As métricas utilizadas foram a matriz de confusão, a precisão, a sensibilidade e *f1-score*. O eixo horizontal da matriz de confusão mostra a classe do diagnóstico predita pelo método, enquanto o eixo vertical mostra a classe do diagnóstico real. A avaliação do diagnóstico foi realizada conforme descrito no Quadro 1, na Seção 1.

Os resultados apresentados na Figura 43 foram obtidos com a aplicação do modelo de segmentação semântica + filtro de área. Observa-se que, aplicando apenas o modelo de segmentação semântica + filtro de área, somente os conjuntos dos diagnósticos 2+ e 3+ foram

corretamente classificados. Os demais foram todos classificados como o segundo diagnóstico mais severo, o 2+.

Figura 43. Matriz de confusão mostrando o resultado do diagnóstico da Tb usando segmentação semântica + filtro de área, aplicado o banco de dados BD2.

Real	Negativo	0	0	0	50	0
	Esparso	0	0	0	50	0
	1+	0	0	0	50	0
	2+	0	0	0	50	0
	3+	0	0	0	0	50
		Negativo	Esparso	1+	2+	3+
		Predito				

Para a avaliação do diagnóstico utilizando a etapa de redução de FP, os modelos de filtragem 1 e 2 foram empregados. O modelo de filtragem 1 é aquele com um filtro de cor e um filtro de forma. E o modelo de filtragem 2 é aquele com dois filtros de cor e um filtro de forma. A Figura 44 apresenta as matrizes de confusão obtidas para o diagnóstico da Tb com a redução de falsos positivos. Em (a) é mostrado o resultado com a aplicação da segmentação semântica + filtro de área + modelo de filtragem 1. E em (b) é mostrado o resultado com a aplicação da segmentação semântica + filtro de área + modelo de filtragem 2. Observa-se que os filtros de cor e forma implementados reduziram a quantidade de FPs gerados na etapa de segmentação semântica. Com os modelos de filtragem 1 e 2, todos os conjuntos das classes diagnósticas 1+, 2+ e 3+ foram corretamente classificados, porém ainda ocorre confusão nas classes Negativo e Esparso. O modelo de filtragem 2 apresentou menor confusão entre essas classes.

O desempenho do método proposto para o diagnóstico da Tb utilizando segmentação semântica + filtro de área + modelo de filtragem 1 é mostrado na Tabela 4. O desempenho para o método utilizando segmentação semântica + filtro de área + modelo de filtragem 2 é mostrado na Tabela 5. O método que utiliza o modelo de filtragem 2 apresentou melhor desempenho com uma precisão média de 0,894, sensibilidade média de 0,896 e *f1-score* médio de 0,895.

Figura 44. Matrizes de confusão do diagnóstico da Tb usando os dois modelos de filtragem. Considerando a segmentação semântica + filtro de área e o modelo de filtragem 1 e em (a) e o modelo de filtragem 2 (b).

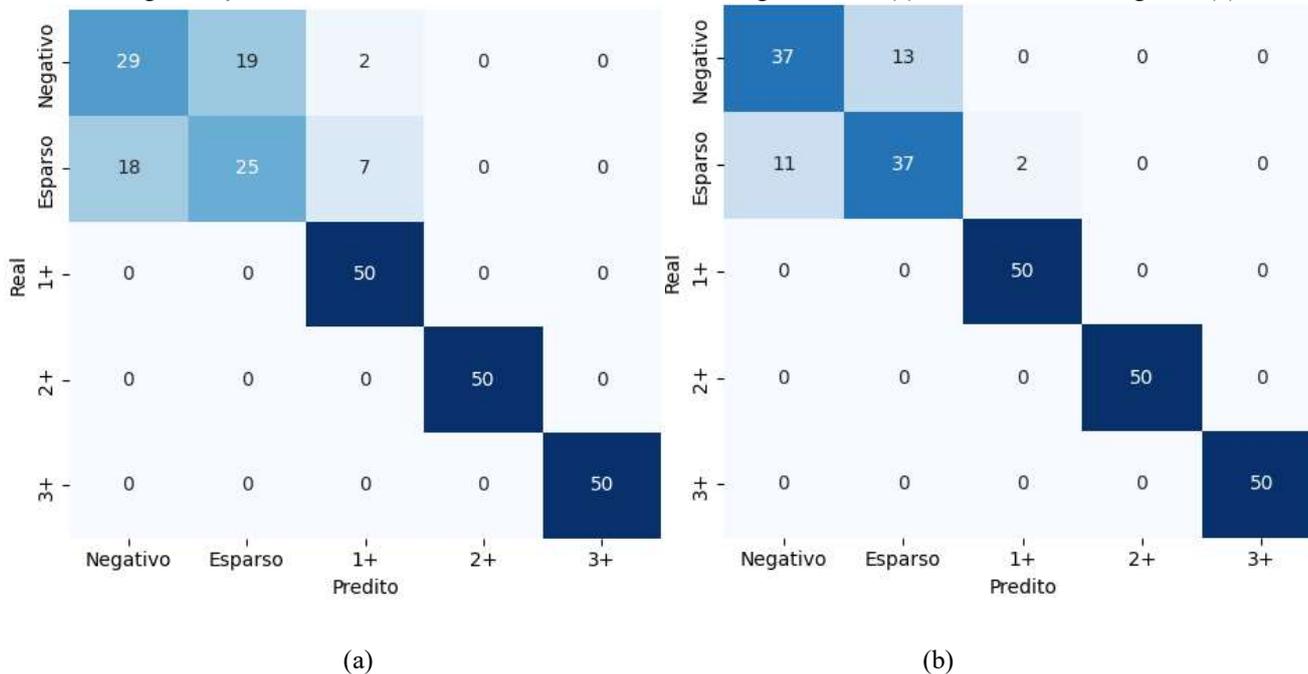


Tabela 4. Desempenho do método proposto para diagnóstico de Tb utilizando segmentação semântica + filtro de área + modelo de filtragem 1.

Classe	Precisão	Sensibilidade	<i>F1-score</i>
Negativo	0,617	0,580	0,598
Esparso	0,568	0,500	0,572
1+	0,847	1	0,917
2+	1	1	1
3+	1	1	1
<b>Média</b>	<b>0,806</b>	<b>0,816</b>	<b>0,817</b>

Tabela 5. Desempenho do método proposto para diagnóstico de Tb utilizando segmentação semântica + filtro de área + modelo de filtragem 2.

Classe	Precisão	Sensibilidade	<i>F1-score</i>
Negativo	0,771	0,740	0,755
Esparso	0,740	0,740	0,740
1+	0,961	1	0,980
2+	1	1	1
3+	1	1	1
<b>Média</b>	<b>0,894</b>	<b>0,896</b>	<b>0,895</b>

### 5.1.2 Desempenho Temporal

Para avaliar o desempenho temporal de cada etapa do método proposto e de cada classe diagnóstica, foi considerado o método de diagnóstico da Tb pulmonar que utiliza segmentação semântica + filtro de área + modelo de filtragem 2, pois esse foi o modelo que alcançou melhor desempenho no diagnóstico.

A Tabela 6 mostra o tempo em segundos para cada etapa do método proposto: pré-processamento, segmentação semântica e redução de FP. Para o levantamento dos dados da Tabela 6, foram analisadas 11 imagens de campo digital com diferentes números de bacilos, que variam de nenhum a 65 bacilos por imagem. Observa-se que o tempo das etapas de pré-processamento e segmentação semântica permanece similar para todos os campos digitais, independentemente da quantidade de bacilos na imagem. No entanto, a etapa de redução de FPs demora mais tempo, à medida que a quantidade de bacilos na imagem cresce. Isso ocorre porque todos os objetos segmentados passam pelos dois filtros de cor e pelo filtro de forma do método de filtragem 2.

A Tabela 7 mostra o tempo em segundos para cada classe diagnóstica da Tb pulmonar. O desempenho temporal para diagnóstico das classes Negativo, Esparso e 1+ é de aproximadamente 7 minutos. Porém, o desempenho temporal para o diagnóstico das classes 2+ e 3+ é de 3,8 minutos e 2,5 minutos, respectivamente. Isso porque, conforme mostra o Quadro 5, para as classes diagnósticas 2+ e 3+ é necessário menos campos digitais.

Tabela 6. Desempenho temporal de cada etapa do método proposto para 11 imagens com diferentes números de bacilos.

Campo Digital	Número de bacilos (padrão-ouro)	Tempo (s)			
		Pré-processamento	Segmentação Semântica	Redução de FP	Total
1	0	0.002	2.423	0.003	2.428
2	0	0.002	2.39	0.017	2.409
3	10	0.017	2.774	0.635	3.426
4	10	0.011	2.668	0.635	3.314
5	20	0.026	2.494	0.899	3.419
6	20	0.016	2.395	1.047	3.442
7	30	0.031	2.469	1.707	4.207
8	30	0.078	2.484	1.655	4.217
9	40	0.026	2.538	2.577	5.204
10	40	0.031	2.421	2.082	4.503
11	65	0.076	2.635	4.356	7.067

Tabela 7. Desempenho temporal de cada classe diagnóstica de Tb pulmonar por baciloscopia de campo claro.

<b>Classe</b>	<b>Tempo (s) (média ± SD)</b>
Negativo	410.349 ± 33.96
Esparso	422.745 ± 25.289
1+	426.059 ± 29.333
2+	230.244 ± 2.22
3+	150.597 ± 6.218

## 5.2 Discussão

A maior parte dos trabalhos publicados na literatura não propôs o diagnóstico da Tb pulmonar usando microscopia de campo claro segundo o método proposto pela OMS (RIEDER et al., 2007; WHO, 1998). Eles apenas focam na detecção automática de bacilos, ou na classificação da infecção por Tb em duas categorias, de baixo nível ou de alto nível (SWETHA; SANKARAGOMATHI; THANGAMALAR, 2020; MITHRA; SAM EMMANUEL, 2019). Contudo, a contagem de bacilos é apenas uma etapa intermediária na obtenção de um método automático de diagnóstico da Tb.

Uma das dificuldades para fazer o diagnóstico através de visão computacional decorre do fato de exigir a disponibilidade de um grande e robusto conjunto de dados de imagens, pois o diagnóstico em um único paciente requer a contagem do número de bacilos em até 164 imagens de campo digital por paciente.

Nesse trabalho, utilizamos um banco de dados que possibilitou não apenas desenvolver um método para detecção de bacilos, mas também avaliar o desempenho desse método de acordo com as diretrizes internacionais adotadas pelo Ministério da Saúde do Brasil (BRASIL, 2008).

Em estudos anteriores, que utilizaram a análise de retalhos ou imagens mosaicos (LÓPEZ et al., 2017), (SERRAO et al., 2020), as taxas de sensibilidade e especificidade para detecção de bacilos obtidas foram acima de 99%. No entanto, como demonstrado neste estudo, esses valores não foram suficientes para o diagnóstico da Tb, pois, a quantidade de falsos positivos detectados em até 164 imagens de campo digital é muito alta e impossibilita o diagnóstico.

Como observado na Figura 43, o diagnóstico é comprometido devido a ocorrência de muitos falsos positivos, ou seja, objetos classificados erroneamente como bacilos pelo modelo

de segmentação semântica, por serem semelhantes em cor ou em forma com bacilos verdadeiros.

Assim sendo, não é possível o diagnóstico confiável somente com a etapa de segmentação semântica, uma vez que os 50 conjuntos das classes diagnósticas Negativo, Esperso e 1+ foram todos classificados como sendo da classe 2+. Portanto, o principal objetivo do método proposto nesse trabalho foi a redução do número de falsos positivos após a etapa de segmentação semântica, usando uma etapa posterior de filtragem. Problemas com detecção de falsos positivos em exames de patógenos com testes microscópicos também foram relatados por outros autores (PRIYA; SRINIVASAN, 2016; SUZUKI et al., 2013; ELTER; HASSLMAYER; ZERFASS, 2011; EL-MELEGY et al., 2019).

Comparados com a literatura, os resultados obtidos para detecção de bacilos na segmentação semântica foram superiores aos resultados obtidos em trabalhos publicados anteriormente. Em RICO-GARCIA et al. (2015), os autores obtiveram uma acurácia de 98%, enquanto neste trabalho foi obtida uma acurácia de 99.66%, conforme a Tabela 2. Em COSTA FILHO et al. (2015), AYAS; EKINCI (2014) e MITHRA; EMMANUEL (2017), os autores obtiveram sensibilidade e especificidade de 96,80% e 96,62%, 89,35% e 62,89%, 96,4% e 97,7%, respectivamente. Neste trabalho, conforme mostrado na Tabela 2, obteve-se sensibilidade de 99,55% e especificidade de 99,72%. Este desempenho só foi possível graças ao uso de uma segmentação semântica utilizando camadas de convolução separadas por canal.

Ao comparar o método desenvolvido neste trabalho para detecção de bacilos com os métodos de classificação de retalhos encontrados nos trabalhos do grupo 2, apresentados na Revisão Bibliográfica, Seção 2, observamos que eles usam retalhos na sua entrada, enquanto o método proposto utiliza imagens mosaico. Uma vantagem do uso de imagens mosaico é que elas minimizam a contagem duplicada de bacilos, tornando desnecessária a aplicação do método *non-maximum supression*.

Dois modelos de filtragem foram avaliados para reduzir o número de FPs. O primeiro deles, o modelo de filtragem 1, que usa uma CNN com imagem de entrada de 8 planos de cor. O segundo deles, o modelo de filtragem 2, que usa duas CNNs para o filtro de cor, a primeira com imagem de entrada com 5 planos de cor, e a segunda com 3 planos de cor. Em ambos os modelos, aplica-se finalmente um filtro de forma, implementado com RF. Conforme mostrado na Figura 44 (b), usar o modelo de filtragem 2 após a segmentação semântica melhorou consideravelmente o desempenho do diagnóstico.

Observamos que não foi possível reduzir a zero o número de FPs. Após a aplicação do método de diagnóstico da Tb em 10 conjuntos de 164 imagens com diagnóstico Negativo,

observou-se, com o modelo de filtragem 2, um número de  $10 \pm 3$  FPs. Portanto, o método desenvolvido teve um viés positivo médio de 10 FPs por conjunto de 164 imagens.

A matriz de confusão na Figura 44 (b) mostra que o método proposto classificou corretamente todos os 50 conjuntos das classes diagnósticas 1+, 2+ e 3+. Para as classes Negativo e Esparso, conforme a Tabela 5, obteve-se uma precisão de 0,771 e 0,74, respectivamente. Para o diagnóstico da Tb, os valores médios da precisão, sensibilidade e *f1-score* obtidos com o modelo de filtragem 2 foram de 0,894, 0,896 e 0,895 respectivamente. A sensibilidade alcançada é superior à sensibilidade de 80% fixada pelo Ministério da Saúde nos Testes de Proficiência para os resultados positivos (MINISTÉRIO DA SAÚDE, 2022b).

Conforme mostrado na Tabela 7, o desempenho temporal para as classes diagnósticas Negativo, Esparso e 1+ é de aproximadamente 7 minutos. E o desempenho temporal para a classe diagnóstica 2+ é de 3,8 minutos e para a classe 3+ é de 2,5 minutos. Como as classes 2+ e 3+ tem menor quantidade de campos digitais, o tempo para o diagnóstico é menor. Em LUMB et al. (2013) os autores afirmam que para avaliar um diagnóstico de Tb, o técnico de microscopia deve levar aproximadamente 5 minutos. Contudo, o laboratório de microscopia de nível básico de um programa nacional de controle da tuberculose em países de baixa renda tem outras tarefas a realizar além da baciloscopia. Considerando que uma carga horária de 20 a 25 exames por dia é o máximo que se pode razoavelmente esperar de um único trabalhador e que a demanda diária pode ser maior, esses fatos podem resultar em sobrecarga do trabalhador e, conseqüentemente, baciloscopias de baixa qualidade (RIEDER et al., 2007). Portanto, a automatização do diagnóstico da Tb ajudará a aumentar a produtividade e a qualidade dos exames realizados.

Em resumo, as características do método proposto para diagnóstico da Tb nesse trabalho são as seguintes:

1. Implementa o diagnóstico automático da Tb usando 5 escalas de classificação (Negativo, Esparso, 1+, 2+ e 3+), de acordo com as recomendações da OMS.
2. Apresenta tempo de diagnóstico de 7 minutos no caso da necessidade de análise do maior número de campos.
3. Reduz a taxa de falsos positivos por meio de um método de filtragem eficiente.
4. Utiliza convolução separada por canal na segmentação semântica de bacilos.
5. Detecta e remove áreas da lâmina com problemas de coloração.
6. Apresenta precisão média de 0,894, sensibilidade média de 0,896 e *f1-score* médio de 0,895.

A microscopia de fluorescência tem, pelo menos, uma sensibilidade 10% maior que a baciloscopia de campo claro, quando as técnicas são analisadas no contexto do mundo real de pacientes com a doença da tuberculose ativa. Apesar disso, em muitos países, a baciloscopia de campo claro ainda seja o primeiro exame diagnóstico empregado para pacientes com suspeita de Tb, por ser um método de diagnóstico econômico e eficiente da detecção da maioria dos casos infecciosos, além de exigir um treinamento mais simples, pois a habilidade de identificar os bacilos por esse método é mais fácil de adquirir. Além disso, a solução Auramina-O, utilizada na microscopia de fluorescência, é cancerígena, trazendo riscos à saúde dos técnicos de laboratório (IMAZ, 2018).

### 5.3 Conclusão

A principal contribuição deste trabalho foi propor um diagnóstico automático e rápido de tuberculose pulmonar em baciloscopia de campo claro de acordo com as recomendações da Organização Mundial da Saúde. O método proposto consiste nas etapas: pré-processamento, segmentação semântica, redução de FPs e diagnóstico da Tb. A construção de um conjunto de dados grande, robusto e anotado possibilitou o desenvolvimento e avaliação do método proposto. Um conjunto de dados de 8000 imagens mosaicos de  $400 \times 400$  pixels e suas respectivas máscaras foi utilizado na etapa de segmentação semântica. Um conjunto de dados com retalhos VP e FP de  $40 \times 40$  pixels foi utilizado na etapa de redução de FPs. E um conjunto de dados de 250 conjuntos de imagens de campo digital de  $1388 \times 1040$  pixels foi utilizado na avaliação dos diagnósticos, sendo 50 conjuntos para cada classe diagnóstica. A etapa de segmentação semântica foi implementada com uma rede neural convolutiva elaborada com camadas de convolução separada por canal. Os objetos segmentados passam por um filtro de área para eliminar objetos muito pequenos ou muito grandes, e os objetos remanescentes seguem para a etapa de redução de FPs através de filtros baseados na cor e forma desses objetos. Os filtros de cor também foram implementados com redes neurais convolutivas, porém para realizar a tarefa de classificação. O primeiro filtro de cor aplicado tem como entrada um retalho de  $40 \times 40$  pixels, com 5 canais: R-G, R-B, H, S e V. O segundo filtro de cor aplicado tem como entrada um retalho de  $40 \times 40$  pixels de 3 canais: R, G e B. O terceiro filtro aplicado é o filtro de forma implementado com classificador *Random Forest* e tem como entrada um vetor de 11 características, que são área, maior eixo, menor eixo, excentricidade e 7 momentos de Hu. Os filtros foram aplicados em cascata para os objetos segmentados. As seguintes métricas de desempenho foram obtidas para a classificação das 5 classes diagnósticas com o método

proposto: precisão média de 89,4%, sensibilidade média de 89,6% e *f1-score* médio de 89,5%. O diagnóstico é obtido em aproximadamente 7 minutos, nos casos em que é necessário analisar mais imagens de campo digital.

## 6 REFERÊNCIAS

- AGGARWAL, C. C. **Neural Networks and Deep Learning**. New York: Springer, 2018.
- AWAD, M.; KHANNA, R. **Efficient learning machines: Theories, concepts, and applications for engineers and system designers**. New York: Apress Open, 2015.
- AYAS, S.; EKINCI, M. Random forest-based tuberculosis bacteria classification in images of ZN-stained sputum smear samples. **Signal, Image and Video Processing**, v. 8, n. 1, p. 49–61, 2014.
- AZIF, F. M.; NUGROHO, H. A.; WIBIRAMA, S. **Adaptive Threshold Determination Based on Entropy in Active Contour without Edge Method for Malaria Parasite Candidate Detection**. 2018 4th International Conference on Science and Technology (ICST). **Anais...**Yogyakarta: IEEE, 2018.
- BADRINARAYANAN, V.; KENDALL, A.; CIPOLLA, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 39, n. 12, p. 2481–2495, 1 dez. 2017.
- BORA, D. J.; KUMAR GUPTA, A.; KHAN, F. A. Comparing the Performance of L\*A\*B\* and HSV Color Spaces with Respect to Color Image Segmentation. **International Journal of Emerging Technology and Advanced Engineering** , v. 5, n. 2, p. 192–2013, 2015.
- BRASIL, M. DA S. **Manual Nacional de Vigilância Laboratorial da Tuberculose e outras Micobactérias**. Brasília, DFMinistério da Saúde, , 2008. Disponível em: <[http://bvsmms.saude.gov.br/bvs/publicacoes/manual\\_vigilancia\\_laboratorial\\_tuberculose.pdf](http://bvsmms.saude.gov.br/bvs/publicacoes/manual_vigilancia_laboratorial_tuberculose.pdf)>
- BREIMAN, L. Random Forests. **Machine Learning**, v. 45, n. 1, p. 5–32, 2001.
- CHOLLET, F. **Xception: Deep learning with depthwise separable convolutions**. Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017. **Anais...**Institute of Electrical and Electronics Engineers Inc., 6 nov. 2017.
- COSTA FILHO, C. F. F. et al. Automatic identification of tuberculosis mycobacterium. **Revista Brasileira de Engenharia Biomedica**, v. 31, n. 1, p. 33–43, 2015.
- COSTA, M. G. F. et al. Automatic identification of mycobacterium tuberculosis with conventional light microscopy. **30th Annual International Conf. of the IEEE Engineering in Medicine and Biology Society**, p. 382–385, 2008.
- COSTA, M. G. F. et al. Multi-focus image fusion for bacilli images in conventional sputum smear microscopy for tuberculosis. **Biomedical Signal Processing and Control**, v. 49, 2019.

DIMITRIADIS, S.; LIPARAS, D. How random is the random forest? RF algorithm on the service of structural imaging biomarkers for AD: from ADNI database. **Neural Regeneration Research**, v. 13, p. 962–970, 12 jun. 2018.

DINESH JACKSON SAMUEL, R.; RAJESH KANNA, B. Tuberculosis (TB) detection system using deep neural networks. **Neural Computing and Applications**, v. 31, n. 5, p. 1533–1545, 2019.

EL-MELEGY, M. et al. Identification of tuberculosis bacilli in ZN-stained sputum smear images: A deep learning approach. **IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops**, v. 2019- June, p. 1131–1137, 2019.

ELTER, M.; HASSLMEYER, E.; ZERFASS, T. **Detection of malaria parasites in thick blood film**. 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. **Anais...**Boston, MA, USA: 2011.

GONZALEZ, R. C.; WOODS, R. E. (RICHARD E. **Digital image processing**. Fourth Edition ed. New York: Pearson Education, 2018.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. [s.l.] The MIT Press, 2016.

GULATI, P.; SHARMA, A.; GUPTA, M. Theoretical Study of Decision Tree Algorithms to Identify Pivotal Factors for Performance Improvement: A Review. **International Journal of Computer Applications**, v. 141, n. 14, p. 19–25, 17 maio 2016.

GUO, M. H. et al. **Attention mechanisms in computer vision: A survey**. **Computational Visual Media** Tsinghua University, , 1 set. 2022.

HU, J. et al. Squeeze-and-Excitation Networks. 5 set. 2017.

IBRAHIM, A. U. et al. Automated detection of Mycobacterium tuberculosis using transfer learning. **Journal of Infection in Developing Countries**, v. 15, n. 5, p. 678–686, 1 maio 2021.

IMAZ, M. S. **Handbook for the bacteriologic diagnosis of Tuberculosis. Part 1: Smear Microscopy Update. Program “Strengthening of the Network of Tuberculosis Laboratories in the Region of the Americas”**Lima Organismo Andino de Salud – Convenio Hipólito Unanue, , 2018.

IOFFE, S.; SZEGEDY, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. 10 fev. 2015.

JAIN, A. K.; MAO, J.; MOHIUDDIN, K. M. Artificial Neural Network. **IEEE Computer**, v. 29, n. 3, p. 31–44, 1996.

KHUTLANG, R. et al. Classification of Mycobacterium tuberculosis in Images of ZN-Stained Sputum Smears. **IEEE Transactions on Information Technology in Biomedicine**, v. 14, n. 4, p. 949–957, 2010.

KUOK, C. P. et al. An effective and accurate identification system of Mycobacterium tuberculosis using convolution neural networks. **Microscopy Research and Technique**, v. 82, n. 6, p. 709–719, 1 jun. 2019.

LECUN, Y. et al. Gradient-Based Learning Applied to Document Recognition. **Proceedings of the IEEE**, v. 86, n. 11, p. 2278–2324, 1998.

LEE, D.; WANG, J.; PLATANIOTIS, K. Contribution of Skin Color Cue in Face Detection Applications. Em: **Lecture Notes in Computational Vision and Biomechanics**. [s.l: s.n.]. v. 11p. 367–407.

LOESDAU, M.; CHABRIER, S.; GABILLON, A. **Hue and Saturation in the RGB Color Space**. ICISP - International Conference on Image and Signal Processing. **Anais...2014**.

LÓPEZ, Y. P. et al. **Automatic classification of light field smear microscopy patches using Convolutional Neural Networks for identifying Mycobacterium Tuberculosis**. 2017 CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON). **Anais...2017**.

LUMB, R. et al. **Laboratory diagnosis of tuberculosis by sputum microscopy: the handbook**. [s.l: s.n.].

MINISTÉRIO DA SAÚDE. **Manual de Recomendações para o Controle da Tuberculose no Brasil**. Brasília, DF: [s.n.].

MINISTÉRIO DA SAÚDE. **Teste Rápido Molecular para Tuberculose (TRM-TB)**. Disponível em: <<https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/t/tuberculose/publicacoes/informativo-sobre-o-teste-rapido-molecular-para-tuberculose-trm-tb.pdf/view>>. Acesso em: 1 mar. 2024a.

MINISTÉRIO DA SAÚDE. **Manual de Recomendações para o Diagnóstico Laboratorial de Tuberculose e Micobactérias não Tuberculosas de Interesse em Saúde Pública no Brasil**. Brasília: [s.n.].

MINISTÉRIO DA SAÚDE. **Boletim Epidemiológico - Tuberculose 2023**. [s.l: s.n.].

MITHRA, K. S.; EMMANUEL, W. R. S. Segmentation and classification of mycobacterium from Ziehl Neelsen stained sputum images for tuberculosis diagnosis. **IEEE International Conference on Communication and Signal Processing**, p. 1672–1676, 2017.

MITHRA, K. S.; SAM EMMANUEL, W. R. Automated identification of mycobacterium bacillus from sputum images for tuberculosis diagnosis. **Signal, Image and Video Processing**, v. 13, n. 8, p. 1585–1592, 2019.

MITHRA, K. S.; SAM EMMANUEL, W. R. GFNN: Gaussian-Fuzzy-Neural network for diagnosis of tuberculosis using sputum smear microscopic images. **Journal of King Saud University - Computer and Information Sciences**, v. 33, n. 9, p. 1084–1095, 2021.

NOH, H.; HONG, S.; HAN, B. Learning Deconvolution Network for Semantic Segmentation. 17 maio 2015.

PANICKER, R. O. et al. Automatic detection of tuberculosis bacilli from microscopic sputum smear images using deep learning methods. **Biocybernetics and Biomedical Engineering**, v. 38, n. 3, p. 691–699, 2018.

PRIYA, E.; SRINIVASAN, S. Automated object and image level classification of TB images using support vector neural network classifier. **Biocybernetics and Biomedical Engineering**, v. 36, n. 4, p. 670–678, 2016.

QUINN, J. A. et al. **Deep Convolutional Neural Networks for Microscopy-Based Point of Care Diagnostics**. Proceedings of International Conference on Machine Learning for Health Care. **Anais...**2016. Disponível em: <<http://arxiv.org/abs/1608.02989>>

RICO-GARCIA, M. et al. **Detection of Mycobacterium tuberculosis in microscopic images of Ziehl-Neelsen-stained sputum smears**. 6th Latin American Conference on Networked and Electronic Media. **Anais...**2015.

RIEDER, H. L. et al. **Priorities for tuberculosis bacteriology services in low-income countries**. 2. ed. Paris: International Union Against Tuberculosis and Lung Disease, 2007.

RONNEBERGER, O.; FISCHER, P.; BROX, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. 18 maio 2015.

RULANINGTYAS, R.; SUKSMONO, A. B.; MENGKO, T. L. R. Automatic classification of tuberculosis bacteria using neural network. **Proceedings of the 2011 International Conference on Electrical Engineering and Informatics, ICEEI 2011**, n. July, p. 1–4, 2011.

SAMMUT, C.; WEBB, G. I. Decision Tree. Em: SAMMUT CLAUDE AND WEBB, G. I. (Ed.). **Encyclopedia of Machine Learning**. Boston, MA: Springer US, 2010. p. 263–267.

SERRAO, M. K. M. et al. Automatic Bacillus Detection in Light Field Microscopy Images Using Convolutional Neural Networks and Mosaic Imaging Approach. **Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS**, v. 2020- July, p. 1903–1906, 2020.

SEWELL, M. **Machine Learning**. Department of Computer Science University College London, 2009.

SHAH, M. I. et al. **Automatic detection and classification of tuberculosis bacilli from ZN-stained sputum smear images using watershed segmentation**. International Conference on Signal Processing. **Anais...**Vidisha, India: IET, 2016.

SHALEV-SHWARTZ, S.; BEN-DAVID, S. **Understanding machine learning: From theory to algorithms**. New York: Cambridge University Press, 2014. v. 9781107057

SOANS, R. S. et al. **Classification of Bacterial Morphotypes from Images of ZN-stained Sputum-smears Towards Diagnosing Drug-resistant TB**. 2016 International Conference on Signal Processing and Communications (SPCOM). **Anais...**: International Conference on Signal Processing and Communications SPCOM.345 E 47TH ST, NEW YORK, NY 10017 USA: IEEE, 2016.

SRIVASTAVA, N. et al. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. **Journal of Machine Learning Research**, v. 15, p. 1929–1958, 2014.

STEINGART, K. R. et al. Fluorescence versus conventional sputum smear microscopy for tuberculosis: a systematic review. **Lancet Infectious Diseases**, v. 6, n. 9, p. 570–581, 2006.

SUGIRTHA, E.; MURUGES, G. **Detection of Tuberculosis Bacilli from Microscopic Sputum Smear Images**. 2017 3rd International Conference on Biosignals, images and instrumentation (ICBSII). **Anais...**Chennai: 2017.

SUZUKI, C. T. N. et al. Automatic segmentation and classification of human intestinal parasites from microscopy images. **IEEE Transactions on Biomedical Engineering**, v. 60, n. 3, p. 803–812, 2013.

SWETHA, K.; SANKARAGOMATHI, B.; THANGAMALAR, J. B. **Convolutional Neural Network based Automated Detection of Mycobacterium Bacillus from Sputum Images**. Proceedings of the 5th International Conference on Inventive Computation Technologies, ICICT 2020. **Anais...**345 E 47TH ST, NEW YORK, NY 10017 USA: IEEE, 2020.

WHO, W. H. O. **Laboratory Services in Tuberculosis Control Organization and Management Part I**. Geneva, Switz : [s.n.].

WHO, W. H. O. **Global tuberculosis report 2023**. [s.l: s.n.]. Disponível em: <<https://iris.who.int/>>.

XIONG, Y. et al. Automatic detection of mycobacterium tuberculosis using artificial intelligence. **Journal of Thoracic Disease**, v. 10, n. 3, p. 1936–1940, 2018.

YOUSEFI, H. et al. Tuberculosis Bacilli Identification: A Novel Feature Extraction Approach via Statistical Shape and Color Models. **Proceedings - 19th IEEE International Conference on Machine Learning and Applications, ICMLA 2020**, p. 366–371, 2020.

ZHAI, Y. et al. Automatic identification of mycobacterium tuberculosis from ZN-stained sputum smear: Algorithm and system design. **2010 IEEE International Conference on Robotics and Biomimetics, ROBIO 2010**, p. 41–46, 2010.

## Evaluating Semantic Segmentation of Tuberculosis Bacilli in Bright Field Microscopy Using Different Color Spaces Components and Mosaic Images

M. K. Serrão<sup>1</sup>[0000-0002-8996-9285], I. M. Saldanha<sup>1</sup>[0009-0008-3122-6156], M. G. F. Costa<sup>1</sup>[0000-0002-6839-1402] and C. F. F Costa Filho<sup>1</sup>[0000-0003-3325-5715]

<sup>1</sup> Federal University of Amazonas/ Center for Research and Development in Electronic and Information Technology – CETELI, Manaus, Brazil  
Corresponding author: ccosta@ufam.edu.br

**Abstract.** Tuberculosis is one of the infectious diseases that causes the most victims in the world. The early diagnosis of the disease is fundamental for the treatment to be carried out quickly, saving lives and decreasing the number of infected people. The tuberculosis diagnosis is a time-consuming process that involves the analysis of up to 100 fields of conventional microscopy, which causes fatigue for the technician involved in the process. In order to assist the technician in counting bacilli in tuberculosis bright field microscopy examinations, automatic methods using deep neural networks are proposed. In this work, we evaluate the use of a semantic segmentation network for bacilli detection, using as input several color components from the RGB, HSV, YCbCr, and Lab spaces. The model that presented the best performance uses only the components of RGB color space as input to the network, with values of accuracy, precision, sensitivity, specificity, and f1-score above 99%.

**Keywords:** — Tuberculosis, Convolutional Neural Networks, Bacilli Detection, Color Spaces.

### 1 Introduction

Tuberculosis (TB) is caused by the bacteria *Mycobacterium tuberculosis* and most commonly affects the lungs. According to the World Health Organization (WHO), about 9.9 million people got sick with TB in 2020, and before the covid-19 pandemic, TB was the leading cause of death from a single infectious agent worldwide. [1].

For more than 100 years the bright field smear technique has been used in TB diagnosis and, because of its availability and simplicity, it remains the main technique in many settings with a high incidence of the disease [2]. In high-income countries, fluorescence microscopy is more widely used for diagnosing TB, but in middle- and low-income countries, where most TB cases occur, the standard diagnostic test is bright field smear microscopy, which uses a conventional microscope [3]. Bright field smear microscopy consists of preparing smears of sputum samples on slides stained with the Ziehl-Neelsen (ZN) or Kinyoun staining methods. This process gives the TB bacilli a pink coloration and the other elements of the sample a blue coloration. A technical

expert then uses a conventional microscope to view these slides and count the number of bacilli. The diagnosis is made according to the number of bacilli counted by the specialist.

Many works have been developed in the field of computer vision to identify bacilli in bright field sputum smear images. A pattern recognition technique that has been prominent in recent years for TB bacilli segmentation is the convolutional neural network (CNN) ([4], [5], [6], [7], [8] e [9]).

In [4], the authors used a CNN to identify malaria-causing pathogens in lung secretion smear samples, and intestinal parasites in stool samples. The input image of the network, in the RGB space, was a patch, i.e. a smaller image, with the size determined by the type of pathogen to be recognized in each case. In [5], the authors also use a CNN for bacilli segmentation. The input to the network was a 40x40 pixels patch and the network was evaluated with RGB images, gray-level images, and with the levels corresponding to R-G values, which corresponds to the difference between the R and G components of the RGB space, and intensity images, obtained from the RGB image. The networks with RGB and R-G input showed better results in comparison with the intensity image. Similarly, in [6] the authors used transfer learning, through a pre-trained model with the CIFAR-10 database, and trained the last dense layers for bacilli recognition. The network was trained with RGB patches of 32x32 pixels. In [9], the authors used a CNN to perform the bacilli segmentation task. The input RGB image has 400x400 pixels and is composed of 100 patches of 40 x 40 pixels, that can be positive or negative. Three CNN architectures were evaluated.

It can be observed that most of the cited works evaluate CNNs with the components of the RGB color space as inputs. This work proposes to evaluate a CNN with different inputs. The goal is to evaluate whether different combinations of the RGB, HSV, Lab, and YCbCr color spaces can improve the performance of the CNN in the detection of bacilli in bright field microscopy images.

The RGB color space was used in this work because it is the original space in which the images are represented, and was used in the other studies published in the literature. The HSV color space was chosen because it has characteristics similar to the human visual system, in which intensity information (H) is uncoupled from color information (S and V). To justify the choice of the YCbCr and Lab spaces, we will use the conclusion obtained in [5], about the importance of the R-G difference in the detection of bacilli. In the YCbCr space, the Cb component, where the R component is subtracted from the G component, would help the bacilli detection, while the Cr component would help the background detection. The inclusion of the Lab space was done to check the hypothesis established in [5], through the negation of the same, as in this space the R and G components are summed in the a component. Therefore, it is expected that this system does not present a good performance.

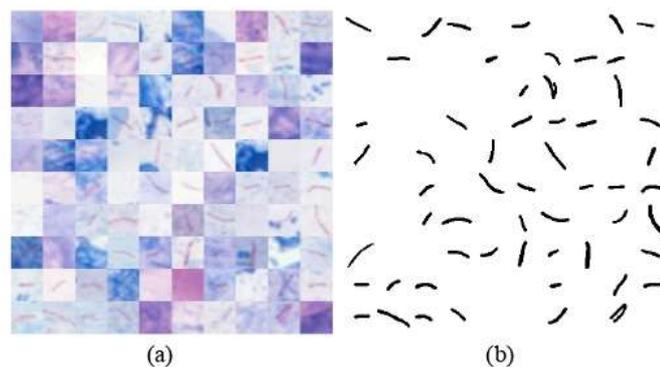
The following color space are combined in the CNN input: RGB; RGB and HSV; RGB, HSV, and Lab; RGB, HSV, and YCbCr; RGB, HSV, YCbCr, and Lab.

## 2 Materials and Methods

### 2.1 Materials

The database used in this work was provided by the Pattern Recognition and Optimization Research Group of UFAM [10]. This database consists of 8000 mosaic images

of 400x400 pixels and their corresponding gold standard images. The mosaic images are formed from the combination of 100 positive and negative 40x40 pixels patches. Positive patches contain a centralized bacillus, and negative patches contain no bacillus. All these patches were extracted from sputum smear images stained with the Kinyoun technique and had the classification performed by an expert. Fig. 1 shows an example of an RGB mosaic image and its gold standard image.



**Fig. 1.** (a) Mosaico image in RGB space and (b) gold standard image.

The database was divided into training, validation, and test sets. For the training set, 6000 mosaic images were used, of which 3000 are composed of approximately 50% positive and 50% negative patches, 1500 are composed of approximately 2% positive and 98% negative patches, and 1500 are composed of only negative patches. For the validation set, 1000 mosaic images consisting of approximately 50% positive and 50% negative patches were used. For the test set, 1000 mosaic images were also used, composed of approximately 50% positive and 50% negative patches.

## 2.2 Semantic Segmentation Network Architecture

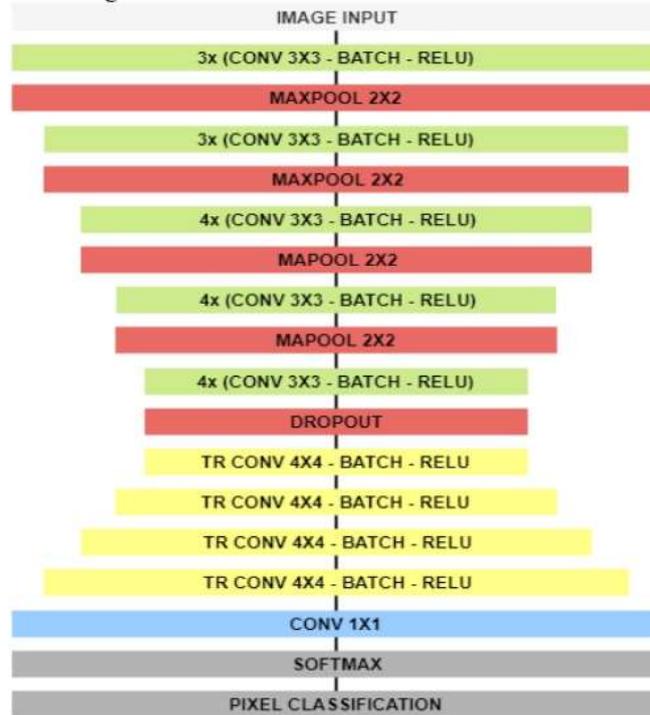
Convolutional neural networks are primarily used for image processing, performing both image classification and semantic segmentation tasks. The classification task aims to classify an image into categories, while semantic segmentation aims to extract regions of interest from the image. In [11], the authors investigated the ability of a convolutional neural network pre-trained with nonmedical images to classify pathologies in x-ray images. The network performs an important screening task by categorizing patients as healthy or having chest pathology that includes the classes: right pleural effusion, cardiomegaly, and abnormal mediastinum. In [12], the authors used a convolutional neural network, through the U-net architecture, to perform the task of semantic segmentation of brain tumors in magnetic resonance images.

The main layers of a convolutional network are the convolution layer, which has the function of extracting features, and the subsampling layer, which aims to reduce the dimensionality of the activation planes of the network [13]. The output of classification networks is a vector whose coordinate values indicate the probability that an image belongs to a particular class. In semantic image segmentation tasks, the network

classifies each pixel of the image as belonging or not belonging to a region of interest. For such networks, the output is also an image with the same dimensions as the input and transposed convolution layers are responsible for recovering an image in the output with the same dimensions as the input image.

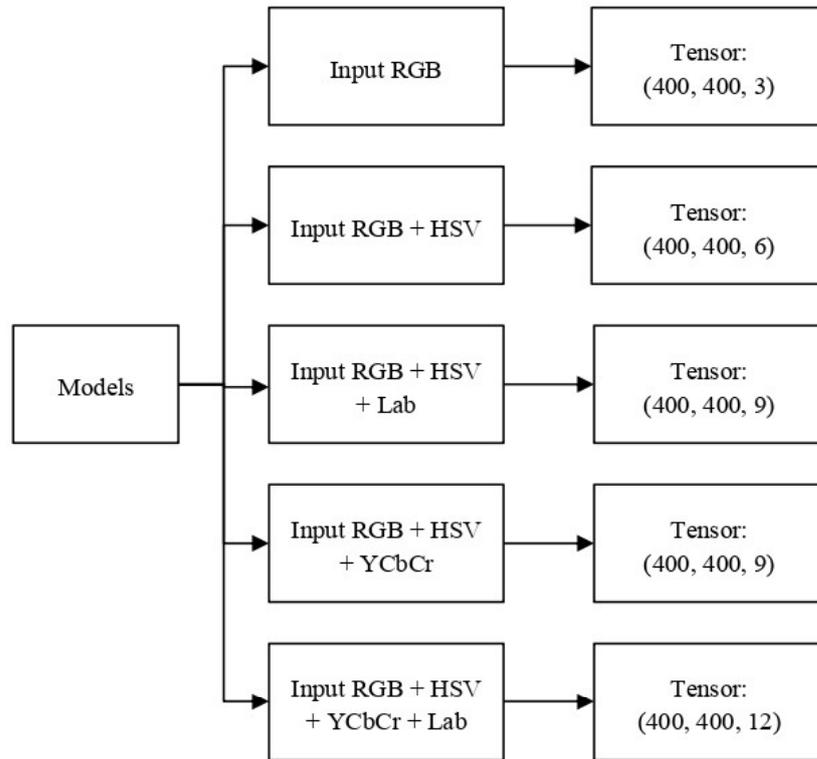
In [9] three different architectures were tested to perform semantic segmentation of tuberculosis bacilli in mosaic images. The architecture that achieved the best performance was the one that used the largest number of layers. The architecture used in this work was based on the best-performing architecture presented in [9].

The architecture used in this work, shown in Fig. 2, is composed of an input layer, 4 subsampling stages, a dropout layer, 4 oversampling stages, and the output layers. The subsampling stages are composed of sequences of convolution layers with a 3x3 filter, batch normalization layer, ReLU activation function, and max-pooling subsampling layer with a 2x2 filter. The over-sampling steps are composed of transposed convolution layer sequences with a 4x4 filter, batch normalization layer, and ReLU activation function. Finally, there is a convolution layer with a 1x1 size kernel, softmax activation function, and a pixel classification layer, which classifies each pixel as belonging to a bacillus or to the background.



**Fig. 2.** CNN architecture used for semantic segmentation of TB bacilli.

In order to evaluate the influence of different color space components on the performance of the semantic segmentation network of TB bacilli, input tensors with different color space components were used. Five models with different color spaces were evaluated to train the network. The list of these models are shown in Fig. 3.



**Fig. 3.** 5 Inputs with different color spaces were used in the 5 semantic bacilli segmentation models.

Each evaluated model has an input tensor in the for-mat: rows x columns x channels. The first model was trained with the components of the RGB color space and has the input tensor with dimensions 400x400x3. The second model was trained with a combination of the coordinates from the RGB and HSV color spaces and has the input tensor with dimensions 400x400x6. The third model was formed by a combination of the coordinates of the RGB, HSV, and Lab color spaces, and has the in-put tensor with dimensions 400x400x9. The fourth model was formed by a combination of the coordinates from the RGB, HSV, and YCbCr color spaces, and has the input tensor with dimensions 400x400x9. The last model was trained was formed by a combination of the coordinates of the RGB, HSV, YCbCr, and Lab color spaces, and has the input tensor with dimensions 400x400x12.

### 2.3 Training Parameters

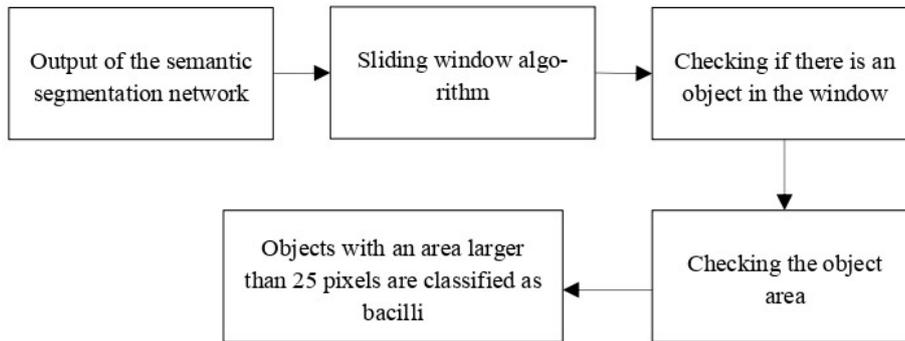
The ADAM optimization method was used to train the network. Table 1 shows the parameters used for training. The training was performed with MATLAB version 2020a, a 3.2 GHz Intel i7-8700 processor with 16GB of RAM, and a GeForce GTX 1070 GPU with 8GB of RAM.

**Table 1.** Parameters used in training the semantic segmentation network

Parameters	Values
Learning rate	0,001
Learning drop factor	0,05
Maximum number of epochs	20
Batch size	10
Stop condition	6000

## 2.4 Bacilli Classification

The CNN performs the classification of each pixel as bacillus or background. A set of 8-connected pixels, identified as bacilli, forms objects that can be bacilli or not. To classify these objects as belonging to bacillus or not, a post-processing step is performed on the binary images obtained from the output of the semantic segmentation network, as shown in the block diagram in Fig. 4. This figure shows a 40 x 40 pixels sliding window algorithm that runs through the entire image to check for 8-connected objects. Each iteration of the algorithm checks for 8-connected pixels in a sub-region of the image. Objects that have an area above the experimentally obtained threshold of 25 pixels are classified as bacilli. Otherwise, the object is disregarded. This process is carried out until the output image is fully evaluated.

**Fig. 4.** Post-processing applied in binary images obtained in semantic segmentation network output, for bacilli detection.

## 2.5 Evaluation Metrics

The following metrics were used to evaluate the performance of the models: accuracy, precision, sensitivity, specificity, and F1- score. These metrics are shown in equations (1) to (5). A true positive (TP) is when the semantic segmentation network detected a bacillus in a patch that contain a bacillus. A true negative (TN) is when the semantic segmentation network did not detect a bacillus in a patch that do not contain a bacillus. A false positive (FP) result is when the semantic segmentation network detected a

bacillus in a patch that do not contain a bacillus. A false negative (FN) result is when the semantic segmentation network did not detect a bacillus in a patch that contain a bacillus.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Sensitivity = \frac{TP}{TP+FN} \quad (3)$$

$$Specificity = \frac{TN}{TN+FP} \quad (4)$$

$$F1 - score = 2 \times \frac{Precision \times Sensitivity}{Precision+Sensitivity} \quad (5)$$

### 3 Results

Table 2 shows the results of the metrics for each of the models evaluated for the semantic segmentation of bacilli. From this table, one can conclude that, in general, the color plane that presented the best performance was the RGB, achieving an accuracy of 99.43%. This result is 2.53% better than the worst performing model, the one which used the color space components of RGB+HSV+YCbCr+Lab at inputs. The latter achieved an accuracy of 96.90%. The precision of the model that used the components of the color space RGB as inputs was 99.62%. This result is 5.11% better than the model that used the RGB+HSV+YCbCr+Lab color space components as inputs. Considering the sensitivity, the model with the components of the RGB+HSV+YCbCr+Lab color spaces as inputs showed a good performance, 99.54%. The specificity and F1-score values of the model with the components of RGB color space as inputs, were 99.63% and 99.43%, respectively.

Furthermore, it is possible to identify those combinations of the RGB, HSV, and YCbCr color spaces, in general, showed good results, with metric values above 99%.

Another observation is related to the Lab color space. In the models where this color space was used, the net-work presented a worse performance.

From these results, we conclude that the use of more color components at the network input does not contribute to the improvement of network performance. The model with the least amount of color components at the input, the one that used the RGB space, presented a much better performance than the models that used a larger amount of color components. Using the components of the RGB color space as inputs of a convolutional network is sufficient to perform the TB bacilli segmentation task with high performance.

Fig. 5 presents the confusion matrix of the test set for the RGB model. This model correctly predicted 49,631 bacilli, for a total of 50,015 bacilli.

The CNN network takes a few seconds to analyze a mosaic image, which is incomparably faster than the time taken by a technician to manually counting bacilli for a

diagnosis. Thus, using the method proposed in this study greatly speeds up the microscopic examination of slides.

In [8], using the faster region-based convolutional neural network (RCNN) and the ZNSM-iDB database for bacilli detection, the authors achieved a recall of 98.4%, a precision of 85.1% and an F1-score of 91.2%. In [6], using CNN for patch classification from a database collected from the Department of Pathology, Peking University First Hospital, the authors achieved a sensitivity of 97.94% and a specificity of 83.65%. In the present work, using CNN for bacilli detection, an accuracy of 99.43%, a precision of 99.62%, a sensitivity of 99.23%, a specificity of 99.63% specificity, and an F1-score of 99.43%.

**Table 2.** Performance of the 5 models were evaluated for semantic segmentation of bacilli

Models	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F1-score (%)
RGB	99.43	99.62	99.23	99.63	99.43
RGB+HSV	99.11	99.12	99.09	99.13	99.10
RGB+HSV+Lab	98.79	98.70	98.89	98.70	98.79
RGB+HSV+YCbCr	99.30	99.50	99.70	99.51	99.30
RGB+HSV+YCbCr+Lab	96.90	94.51	94.51	94.29	96.96



**Fig. 5.** Confusion matrix showing the performance on the test set of the model that uses the RGB color space components as input.

## 4 Conclusion

In this work, the performance of a convolutional neural network was evaluated to perform the semantic segmentation of TB bacilli, using different color spaces at the network input. The RGB model with input tensor 400x400x3, was the one that presented the best performance for the considered metrics, reaching performance higher than 99% for all of them, while the model with components of the RGB+HSV+YCbCr+Lab color spaces as inputs, presented a lower performance compared to the other evaluated models.

The main conclusion is that the RGB color space alone is sufficient to perform the TB bacilli segmentation task with high performance. With a larger number of color space components in the input, satisfactory results are achieved, but lower than the results obtained using only the RGB space color. The use of the Lab color space compromises network performance. We believe that the main reason for this is because in the Lab color space, the R and G components are grouped into a single component,  $a$ .

## 5 Acknowledgment

This research, carried out within the scope of the Samsung-UFAM Project for Education and Research (SUPER), according to Article 48 of Decree n°6.008/2006 (SUFRAMA), was funded by Samsung Electronics of Amazonia Ltda., under the terms of Federal Law n°8.387/1991 through agreement 001/2020, signed with UFAM and FAEPI, Brazil.

## 6 Conflict of Interest

The authors declare that they have no conflict of interest.

## References

1. World Health Organization (2021) Global Tuberculosis Report World Health Organization
2. Harries A, Kumar A (2018) Challenges and Progress with Diagnosing Pulmonary Tuberculosis in Low- and Middle-Income Countries. *Diagnostics (Basel)* vol 8, no. 4(78) DOI 10.3390/diagnostics8040078
3. Steingart K R, Henry M et al. (2006) Fluorescence versus conventional sputum smear microscopy for tuberculosis: a systematic review. *Lancet Infect. Dis.*, vol. 6, no. 9, pp 570–581 DOI 10.1016/S1473-3099(06)70578-3
4. Quinn J A, Nakasi R et al. (2016) Deep Convolutional Neural Networks for Microscopy-Based Point of Care Diagnostics. *International Conference on Machine Learning for Health Care Proc.* vol. 56, pp 1–12 at: <http://arxiv.org/abs/1608.02989>
5. López Y P, Costa Filho C. F. F et al. (2017) Automatic classification of light field smear microscopy patches using Convolutional Neural Networks for identifying Mycobacterium

- Tuber-culosis. Chil. Conf. Electr. Electron. Eng. Inf. Commun. Tech-nol. Proc. pp 1–5 DOI 10.1109/CHILECON.2017.8229512
6. Xiong X Y, Hou B a et al. (2018) Automatic detection of myco-bacterium tuberculosis using artificial intelligence. *J. Thorac. Dis.*, vol. 10, no. 3, pp. 1936–1940 DOI 10.21037/jtd.2018.01.91.
  7. Dinesh R, Saumel J et al. (2019) Tuberculosis (TB) detection system using deep neural networks, *Neural Comput. Appl.* vol. 31, no. 5, pp 1533–1545 DOI 10.1007/s00521-018-3564-4
  8. El-Melegy M, Mohamed D et al. (2019) Identification of tuberculosis bacilli in ZN-stained sputum smear images: A deep learning approach. *IEEE Comput. Soc. Conf. Comput. Vis. Pat-tern Recognit Proc.* vol. 2019-June, pp 1131–1137 DOI 10.1109/CVPRW.2019.00147
  9. Serrão M, Costa M et al. (2020) Automatic Bacillus Detection in Light Field Microscopy Images Using Convolutional Neural Networks and Mosaic Imaging Approach. *Int Conf IEEE Eng Med Biol Soc. Proc.* pp 1903-1906 DOI 10.1109/EMBC44109.2020.9176105.
  10. Pattern Recognition and Optimization Research Group (2008) at <http://dgp.cnpq.br/dgp/es-pelhogrupo/4186>
  11. Bar Y, Diamant I et al. (2015) Chest pathology detection using deep learning with non-medical training. *Int. Symp. Biomed. Imaging Proc.* vol. 2015-July, pp 294–297, DOI 10.1109/ISBI.2015.7163871.
  12. El Kaitouni S E I, Tairi H (2020) Segmentation of medical im-ages for the extraction of brain tumors: A comparative study between the Hidden Markov and Deep Learning approaches. *Int. Conf. Intell. Syst. Comput. Vision Proc.* pp 1-5 DOI 10.1109/ISCV49265.2020.9204319.
  13. Goodfellow I, Bengio Y, Courville A (2017) *Deep Learning*. MIT Press, London

# APÊNDICE B – ARTIGO PUBLICADO NA REVISTA ACADÊMICA *COMPUTERS IN BIOLOGY AND MEDICINE* EM 2024

## Automatic bright-field smear microscopy for diagnosis of pulmonary tuberculosis

Mikaela Kalline Maciel Serrão<sup>a</sup>, Marly Guimarães Fernandes Costa<sup>a</sup>,  
Luciana Botinelly Mendonça Fujimoto<sup>b</sup>, Maurício Morishi Ogusku<sup>c</sup>, Cicero Ferreira  
Fernandes Costa Filho<sup>a,\*</sup>

<sup>a</sup> R&D Center in Electronic and Information Technology, Federal University of Amazonas, Manaus, 69067-005, Brazil

<sup>b</sup> Faculty of Medicine, Federal University of Amazonas, Manaus, 69020-160, Brazil

<sup>c</sup> National Institute of Amazonian Research – INPA, 69080-971, Manaus, Brazil

### ARTICLE INFO

#### Keywords:

Pulmonary tuberculosis diagnosis  
Semantic segmentation  
Smear microscopy  
Attention mechanism  
Depth-wise separable convolution  
bacillus detection

### ABSTRACT

In recent decades, many studies have been published on the use of automatic smear microscopy for diagnosing pulmonary tuberculosis (TB). Most of them deal with a preliminary step of the diagnosis, the bacilli detection, whereas sputum smear microscopy for diagnosis of pulmonary TB comprises detecting and reporting the number of bacilli found in at least 100 microscopic fields, according to the 5 grading scales (negative, scanty, 1+, 2+ and 3+) endorsed by the World Health Organization (WHO). Pulmonary TB diagnosis in bright-field smear microscopy, however, depends upon the attention of a trained and motivated technician, while the automated TB diagnosis requires little or no interpretation by a technician. As far as we know, this work proposes the first automatic method for pulmonary TB diagnosis in bright-field smear microscopy, according to the WHO recommendations. The proposed method comprises a semantic segmentation step, using a deep neural network, followed by a filtering step aiming to reduce the number of false positives (false bacilli): color and shape filtering. In semantic segmentation, different configurations of encoders are evaluated, using depth-wise separable convolution layers and channel attention mechanism. The proposed method was evaluated with a large, robust, and annotated image dataset designed for this purpose, consisting of 250 testing sets, 50 sets for each of the 5 TB diagnostic classes. The following performance metrics were obtained for automatic pulmonary TB diagnosis by smear microscopy: mean precision of 0.894, mean recall of 0.896, and mean F1-score of 0.895.

### 1. Introduction

According to the World Health Organization (WHO), in 2020, it is estimated that 9.9 million people were affected by tuberculosis (TB) disease worldwide. In 2020, the number of deaths caused by TB in HIV-negative people was 1.3 million and an additional 214,000 in persons with HIV/AIDS [1]. About 85% of all tuberculosis deaths occurred in Africa and Southeast Asia. India was responsible for 34% of the total number of deaths from TB. The COVID-19 pandemic has severely increased the death rate from TB. In 2021, the WHO published a global estimate in which TB was the 13th leading cause of death and the leading cause of a single infectious agent. By 2020, because of the COVID-19 pandemic, TB became the second leading cause of death from a single infectious agent (WHO, 2021).

TB is caused by the agent *Mycobacterium tuberculosis*, referred to as Koch's bacillus, and its pulmonary form is the most common. When a person affected by the disease coughs, droplets with bacilli are released into the environment and can be inhaled by other people who may contract the disease. In recent decades, more precise molecular tests for TB diagnosis, known as nucleic acid amplification tests, have been developed. Nevertheless, they are more costly [2]. In low- and middle-income countries, bright-field smear microscopy is still the most commonly used method for diagnosing pulmonary TB because of the low direct costs of laboratory infrastructure to maintain this service [3].

Bright-field smear microscopy consists of preparing and staining smears of sputum samples with staining methods, counting and reporting the number of bacilli found in at least 100 smear microscopic fields, according to the 5 grading scales proposed by the WHO [4]. Bacilli

\* Corresponding author.

E-mail address: [ccosta@ufam.edu.br](mailto:ccosta@ufam.edu.br) (C.F.F. Costa Filho).

<https://doi.org/10.1016/j.combiomed.2024.108167>

Received 9 November 2023; Received in revised form 19 January 2024; Accepted 15 February 2024

Available online 29 February 2024

0010-4825/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).

visual counting is a labor-intensive task that depends on the expertise of the microscopist [3]. More details about smear microscopy are presented in the next section. On the other hand, with automation of this analysis, smear microscopy becomes simple enough to perform, requires little or no interpretation, substantially reduces technician workload and increases the number of exams performed per day [5,6]. Additionally, automatic methods of pathogen detection by microscopy usually result in high sensitivity rates, indicating that the method correctly classifies objects of interest. However, the disadvantage of these methods is the high number of false positives (FPs) detected [7].

In the literature, we identified many studies addressing the topic of the TB microscopic diagnosis. However, they are limited to the development of methods for detecting bacilli in microscopic field images rather than the automation of smear microscopy. The developed methods use classic machine learning classifiers and deep learning classifiers. The classic machine learning techniques used for bacilli detection are Bayes classifier [2], linear regression classifier [2], quadratic discriminant classifier [2], support vector machine (SVM) [3], random forest (RF) [5], k-means clustering [8], fuzzy c-means clustering [9], decision trees [10], k-nearest-neighbor [11], particle swarm optimization [12], digital image processing techniques [13], multilayer perceptron (MLP) [14], watershed segmentation [15] and Gaussian-fuzzy-neural network [16]. The deep learning classifiers used for bacilli detection are convolutional neural networks (CNN) [17–22], and deep belief neural networks [23].

This paper presents an automatic method for pulmonary TB diagnosis by sputum smear microscopy based on the WHO grading scale of 5 diagnostics classes (negative, scanty, 1+, 2+ and 3+), considering the number of bacilli found in the patient sputum smear (WHO, 1998), [24]. The main challenge solved by the proposed method is the reduction of the number of FPs in TB bacilli count in bright-field microscopy images. The main contributions of this paper are as follows.

- Propose a method for TB diagnosis combining semantic segmentation of bacilli with postprocessing steps to reduce FPs.
- In the semantic segmentation of bacilli, compare the performance of different encoder architectures: using regular convolution layers, using depth-wise separable convolution layers [25] and channel attention mechanisms [26].
- Compare the performance of bacilli semantic segmentation with the U-Net network [27], a network projected for biomedical image segmentation.
- Design and implement a database to evaluate the automatic diagnosis of TB.

The reason for exploring different encoder architectures in semantic segmentation is to identify which method is most suitable for processing color channel information, which is of fundamental importance in visually identifying bacillus in bright-field sputum smear microscopy.

## 2. Bright-field smear microscopy

In recent decades, remarkable progress has been made in the quality of cytobacteriological diagnostic services and the speed of such diagnoses in industrialized countries. However, differences in the epidemiological context in which pulmonary TB is diagnosed (whether in industrialized countries where the disease prevalence is low or in countries where the disease is endemic) have clear implications for the diagnostic needs of each setting. In most parts of the world, where TB is a major public health burden, this progress has not yet been achieved and microscopic examination of the stained sputum smear remains necessary. It is no longer in doubt, however, that different tools are needed to diagnose TB [6].

According to Rieder et al. [24], the sensitivity (proportion of actual cases identified) of bright-field smear microscopy in diagnosing pulmonary TB is far from perfect. Sensitivity in identifying transmitters of

**Table 1**

WHO scale, composed of 5 diagnostic classes, for reporting the results in terms of counting of bacilli in the microscopic field. (WHO, 1998) [24].

Result/Recording	Finding
Negative	No AFB found in at least 100 microscopic fields
Scanty/Record exact number of bacilli	1 to 9 AFB per 100 microscopic fields
1+	10 to 99 AFB per 100 microscopic fields
2+	1 to 10 AFB per field in at least 50 microscopic fields
3+	More than 10 AFB per microscopic field in at least 20 fields

tubercle bacilli is over 80%. This low sensitivity affects more patients of some groups, such as those living with HIV/AIDS, and in the laboratory diagnosis of extrapulmonary disease. Considering that sputum smear-positive patients are much more likely to transmit *M. tuberculosis* than negative patients, the sputum smear microscopy is one of the most effective and efficient tools for case-finding in a national tuberculosis control program, identifying the cases that should have the highest priority in tuberculosis control.

Smear microscopy includes sputum collection, preparation, staining smear, reading of smears and reporting the number of bacilli found. The smear slide is smeared with the patient's sputum, which is dried, fixed, and stained using standard staining techniques such as Ziehl-Neelsen or Kinyoun techniques (WHO, 1998) [24]. Both techniques use the same staining solutions. The difference is that the latter is a cold technique and uses different concentrations of staining solutions. In the stained smear, bacilli are called acid-fast bacilli (AFB) because the mycolic acid in the cell wall of bacillus render them resistant to decolorization with acid alcohol during the staining process. The stained smear slides are examined under ordinary light (bright-field) microscopy and the AFB appear bright red against the background material counterstained in blue. They are quite variable in shape, from short fragments to elongated types. They may be uniformly stained, may have gaps, or may even appear granular. They occur singly or in small groups, and rarely in large clumps. If structures are seen that have the correct color and a compatible morphology, they should be considered AFB and be reported as such in the smear microscopy analysis [24].

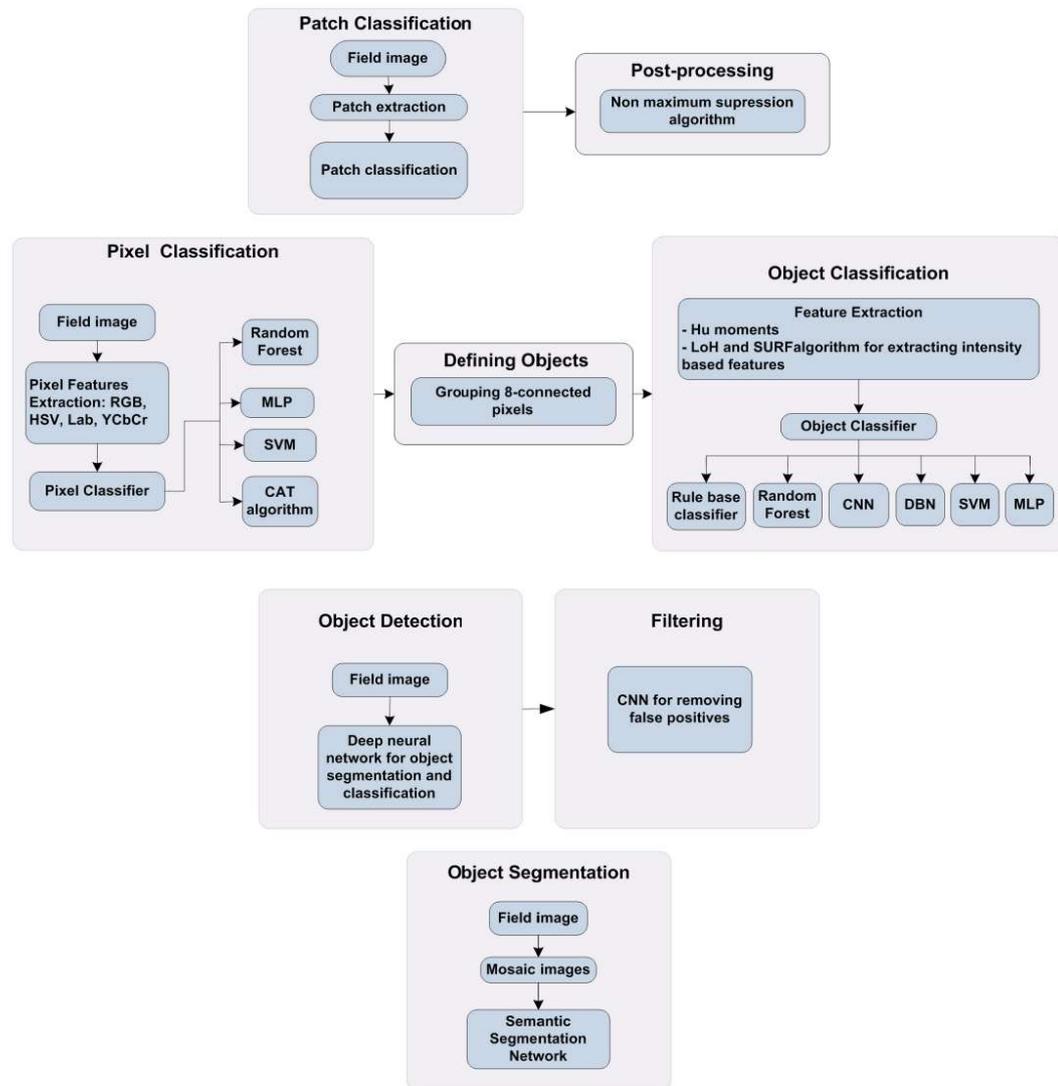
Considering that the number of acid-fast bacilli indicates how infectious the patient is, several grading scales for the number of AFB found in a smear have been proposed over the years (WHO, 1998), [24, 28]. The WHO scale comprises 5 groups for reporting the results from reading a smear: negative, scanty, 1+, 2+ and 3+. Table 1 shows the WHO grading scales for bright-field smear microscopy.

Results of smear microscopy analysis, however, depend upon the attention of a trained and motivated technician, microscope maintenance and laboratory quality control [6].

Although the structure of the health system varies from country to country, TB laboratory networks almost always provide services at peripheral, intermediate, and national levels. The main task of the basic level microscopy laboratory (the peripheral) of a national tuberculosis control program in low-income countries is the examination of sputum specimens for AFB. Usually, only one or two technicians work in a peripheral laboratory. Additionally, these workers have other tasks to perform besides sputum smear microscopy. Considering that a workload of 20–25 examinations per day is the maximum that can reasonably be expected from a single worker and that the daily demand may be higher, these facts may result in worker overload and, consequently, poor quality smear microscopies [24].

## 3. Literature review

This section presents the main techniques described in the literature for detecting *Mycobacterium tuberculosis* in bright-field smear microscopy images. The papers published can be classified into 4 groups. The



**Fig. 1.** Bacilli detection methods of reviewed papers: (a) group #1 of papers that use patch classification; (b) group #2 of papers that use pixel classification + object classification; (c) group #3 of papers that use deep neural networks for object detection and filtering; (d) group #4 of papers that use semantic segmentation.

first group [17–19] is characterized by presenting only one stage for bacilli identification: a patch classification step, which classifies the patch as positive or negative. The second group [3,5,23,29] is characterized by presenting 2 stages for bacillus identification: pixel classification and object classification. The third group [30] is characterized by using neural networks specialized in detecting objects, as Faster R-CNN networks. This group uses 2 stages, 1 for object detection and other for FP filtering. The fourth group [22] is characterized by using semantic segmentation. Fig. 1 (a) and (b), 1(c) and 1(d) show the methodologies used in groups 1, 2, 3 and 4, respectively.

Regarding the papers of the first group that use only one stage of patch classification, in Ref. [19], the patch classification was performed using CNN. The authors implemented a method for 3 different diagnoses: malaria in blood smear samples, tuberculosis in sputum smear samples, and intestinal parasites in stool samples. Each image collected was downsampled and then split up into overlapping patches, with the downsampling factor and patch size determined by the type of pathogen to be recognized in each case. According to the authors, since most images do not contain pathogen objects, the number of negative patches was disproportionately large compared to the number of positive ones.

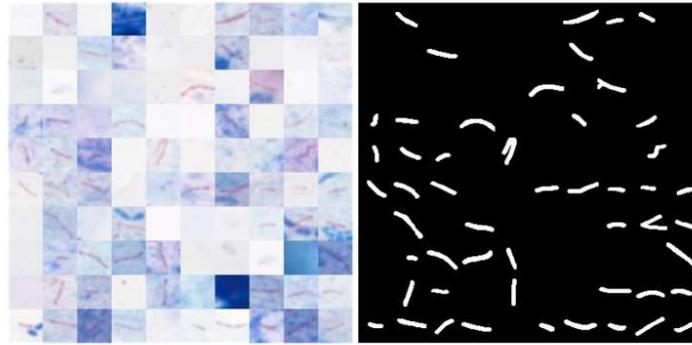


Fig. 2. (a) a  $400 \times 400$ -pixel mosaic image, composed of a  $10 \times 10$  array of  $40 \times 40$ -pixel patches; (b) corresponding mask of the mosaic image shown in (a).

Two actions were taken to obtain a balanced dataset. First, negative patches were randomly discarded so that there were at most 100 times the number of positive patches. Second, data augmentation was applied in positive patches by rotating and flipping the patches, giving 7 extra positive patches for each original one. The same network, 1 with 4 hidden layers (convolution, pooling, convolution, and fully connected layer), was used to detect all 3 types of pathogens. The TB database consists of 630,284 patches (9.0% positives and 91% negatives). The authors used the nonmaximum suppression algorithm to exclude many overlapping bacilli that occurred on patch borders. The following results were obtained: AUC-ROC of 99% and an average accuracy of 93%.

In [17], the authors also used CNN for patch classification. The database, a set of 29,310 patches with a size of  $40 \times 40$  pixels, was fully balanced (equal number of positive and negative patches) and divided into 60% for training, 20% for validation and 20% for testing. To achieve better results, the authors used intensity normalization in the patches. Three CNNs were evaluated: the first with 1 convolutional layer, the second with 2 convolution layers and the third with 3 convolution layers. Also, 3 different inputs were evaluated: RGB patches, grayscale patches and intensity patches obtained from the subtraction of R and G channels (R-G). To achieve better results, the authors used the dropout regularization method. The best model was the one with 3 convolution layers trained with patches extracted from RGB images. The search for bacilli uses a moving window of  $40 \times 40$  pixels. At each location, this window is classified as positive or negative. The authors also used the nonmaximum suppression algorithm to exclude many overlapping bacilli occurring in patch borders. For the patch classification, an AUC-ROC of 99% and an accuracy of over 98% were obtained. For bacilli detection in the whole image, the following results were obtained: an accuracy of 73.68%, a recall of 92.63%, and an F1-score of 82.07%.

In [31], the authors performed the classification of patches, of  $64 \times 64$  pixels, using a RegNetX4 convolutional network. The database consists of tissue samples (biopsies or surgical specimens). Different data augmentation techniques were employed in the training set. For each patch, the output of the network generates a score. With this information and analyzing the context of the slide in which the patch is inserted, the pathologist decides whether the slide is positive or negative (a single bacillus is sufficient for the slide to be considered positive). The authors report an accuracy of 0.977.

In [32], the authors performed the classification of patches, of  $256 \times 256$  pixels, using a convolutional network pipeline, consisting of 2 Inception-V3 networks and a logistic regressor. Patches are classified as positive or negative. If there is a positive patch on the slide, it is classified as positive. The model achieved sensitivity of 87.13%, specificity of 87.62% and an F1-score of 80.18%.

In [33] the authors implemented a CNN model for patch classification. They used 2220 patches of  $224 \times 224 \times 3$  pixels, 50% positive patches and 50% negative patches. The data set was divided into 80/20 for training and testing, respectively. Aiming to reduce the number of network parameters, the authors reduced the number of filters in the architecture used for benchmark. To compensate for performance loss, they adopted skip connections with variable sizes. The performance achieved was close to the benchmark, but with 90% fewer parameters. Using 10-fold cross validation, the authors achieved a mean accuracy of 97.83%.

Regarding the papers of the second group that use 2 stages for bacilli detection, in Ref. [5], the authors used a supervised algorithm for pixel classification. In the first stage, pixels were classified by an RF classifier, as belonging or not to a bacillus. RGB pixel color components were used as RF inputs. Nearby pixels labeled as belonging to a bacillus were grouped in objects that are rotated, resized, and positioned in the center of  $30 \times 30$ -pixel bounding boxes, which are used for object classification such as bacillus. For object classification, the authors used another RF classifier with Hu invariant moments as input variables. The database used included 116 images with a size of  $640 \times 480$  pixels, divided into 40 images for training and 76 images for testing. The results obtained were an accuracy of 67.98%, a sensitivity of 89.34% and a specificity of 62.89%.

In [3], for pixel classification, 30 features were extracted, subtracting components from the color spaces RGB, HSI, YCbCr and Lab. Two classifiers were compared, MLP and SVM. Nearby pixels labeled as belonging to a bacillus were grouped in objects. In the sequence, 3 types of filters were used to separate objects corresponding to bacilli from artifacts: a size filter, to remove objects with very large or very small areas, a geometric filter that eliminates objects based on their eccentricity, and a rule-based filter that uses a color ratio parameter in the RGB color space. The authors used a database with 120 images of  $2816 \times 2112$  pixels. The best results were obtained with the SVM classifier: a sensitivity of 96.80% and an accuracy of 96.62%.

In [23], for pixel classification, the authors used the channel area thresholding algorithm, which is a multistage segmentation method. Nearby pixels labeled as belonging to a bacillus were grouped in objects. In the sequence, features are extracted from segmented objects using location-oriented histogram and speeded-up robust feature algorithms. A deep belief network was trained to classify an object as a bacillus or not. If the number of bacilli in one image is less than or equal to 5, it is classified as a low infection level. If the number of bacilli in one image is greater than 5, it is classified as a high infection level. The following results were obtained: an accuracy of 97.55%, a sensitivity of 97.86%, and a specificity of 98.23%.

Regarding the papers of the third group that use neural networks

specialized in detecting objects, in Ref. [30] the authors used a state-of-the-art deep learning method, the Faster Region-based Convolutional Neural Network (Faster R-CNN), which identifies TB bacilli with a bounding box. A second stage, consisting of a CNN, classifies each bounding box as a bacillus or not. The second stage was necessary because the authors identified a high false positive rate (FPR) in the experiments. The CNN reduced the FPR by 20%. The following results were obtained: a recall of 98.4%, a precision of 85.1%, and an F1-score of 91.2%. The low precision value suggests that the authors were not able to significantly reduce the number of FPs.

In [34], the authors proposed a two-stage bacillus detection system. A Faster R-CNN was also used to identify bacilli with a bounding box, followed by a CNN, that classifies each bounding box as a bacillus or not. A refinement mechanism was used in the Faster R-CNN resulting in more accurate bounding boxes. The database consists of images with  $1280 \times 960$  pixels. The input of the CNN has a size of  $48 \times 48$  pixels. They achieved a precision of 86%. The authors report that the large number of bacilli candidates made the system learning a difficult task and that a more powerful CNN should be implemented to improve false detection.

Regarding the papers of the fourth group, that use semantic segmentation, we developed a bacilli detection method using mosaic images, and a deep learning approach, a CNN [22]. Considering that images captured from microscopic fields may contain a small number of bacilli or even no bacillus, we propose to train the CNN using mosaic images, which are composed of patches extracted from microscopic fields arranged in a  $10 \times 10$  array. On average, a mosaic image contains 50% positive (containing bacillus) patches and 50% negative (without bacillus) patches. Fig. 2 shows an example of a mosaic image. In this approach, the CNN is trained with a balanced data set, preventing the classifier from specializing in just one of the classes.

Three CNN architectures were evaluated for performing the bacilli segmentation. After segmentation, an area-based filtering step is implemented to remove artifacts smaller than 25 pixels.

The best results obtained for bacilli detection were an accuracy of 99.66%, a precision of 99.35%, a sensitivity of 99.99%, a specificity of 99.34% and an F1-score of 99.67%.

Considering the FPR of 0.66% (1-specificity) and the fact that each mosaic image has an average of 50 negative patches, we have 0.33 FPs per mosaic image ( $0.66 \times 50/100$ ), which results in an average of 4.86 FP per microscopic field (as shown in section 4.2.1 the area of one microscopic field corresponds to 14.79 mosaic images). If we applied the developed method to evaluate 100 sputum smear microscopic fields of a negative diagnostic patient, it resulted in 486 FPs. According to Table 1, this number of FPs could be classified as "2+", the second most severe degree of TB diagnosis. Therefore, despite the high rates of sensitivity, specificity, and F1-score, the resulting FPR made unfeasible the proposal of automatic pulmonary diagnosis of TB, according to the WHO scale of 5 diagnostic classes.

Summarizing the reviewed papers, the papers of group 1 that use only one stage of patch classification, pay a high computational burden, due to the nonmaximum suppression algorithm. The papers of group 2 that use 2 stages for bacilli detection, involve grouping neighbor pixels to define a bacillus. However, since the regions of a bacillus can be very heterogeneous, several fragments of bacillus could be detected, which can be mistaken for artifacts and may generate false negatives. The 2 papers of group 3, that use neural networks specialized in detecting objects for bacilli detection (Faster R-CNN), showed very low precision values. The only paper of group 4, that uses semantic segmentation with mosaic images, reported a specificity of 99.34%.

In general, the reviewed papers did not propose automatic diagnosis of pulmonary TB by bright-field smear microscopy, according to the WHO recommendation (WHO, 1998) [24]. They only focus on bacilli detection or on classifying TB infection in 2 categories, low-level or high-level (Swetha et al., 2020) [23] or in classifying a slide as positive or negative [31,32].

Table 2

Description of datasets used in semantic segmentation and Tb diagnosis.

Dataset	Slide	Image size	Number of images	Application
DDS3	S01-S15	Patch Mosaic Images of $400 \times 400$ pixels and their respective binarized masks	6000 mosaic images with corresponding binarized masks: -3000 mosaic images with 50% positive patches and 50% negative patches. -1500 mosaic images with 98% negative patches and 2% positive patches. -1500 with 100% negative patches.	Training, Validation and Testing of algorithms for bacilli detection (50% for training, 25% for validation and 25% for testing)
DDS4	S16-S25	Full images of $1040 \times 1388$ pixels	250 sets of 164 DF images: -50 sets corresponding to the diagnostic negative -50 sets corresponding to a scanty diagnosis -50 sets corresponding to a positive diagnosis 1+ -50 sets corresponding to a positive diagnosis 2+ -50 sets corresponding to a positive diagnosis 3+	Evaluation of Tuberculosis Diagnosis

The main shortcomings identified in the previously reviewed papers, which prevent the automatic diagnosis of pulmonary TB using bright-field smear microscopy, is that the values of specificity for bacilli detection are below 99.34%, which leads, in semantic segmentation, as mentioned before, to a high number of FPs.

The reviewed papers do not use depth-wise convolution or channel attention mechanisms for bacilli segmentation. Aiming to improve the specificity, in this work we evaluate the use of these techniques to better explore color channels in bacilli segmentation.

## 4. Materials and methods

### 4.1. Materials

#### 4.1.1. Dataset

The datasets used in this work were made available by the Pattern Recognition and Optimization Research Group of the Federal University of Amazonas, Brazil [35]. This research was approved by the Human Research Ethics Committee of Instituto Nacional de Pesquisa da Amazônia (INPA), protocol 186/08.

The images were acquired with a Canon PowerShot A640 digital camera, which couples with a 10-megapixel CCD imager sensor with a  $4 \times$  optical zoom. The microscope used was a Zeiss Axioshop 40. A magnification of  $100 \times$  and a numerical aperture of 1.25 was used. The sputum samples from patients suspected of being pulmonary TB cases were prepared by the Kinyoun stain method [6]. The estimated microscopic visual field to be read by the microscopist corresponded to a diameter of 0.228 mm, equivalent to a circular area of  $0.04 \text{ mm}^2$ . Since 100 microscopic fields need to be read, the total area to be read corresponds to  $4 \text{ mm}^2$ . There is a c-mounted adapter between the microscope

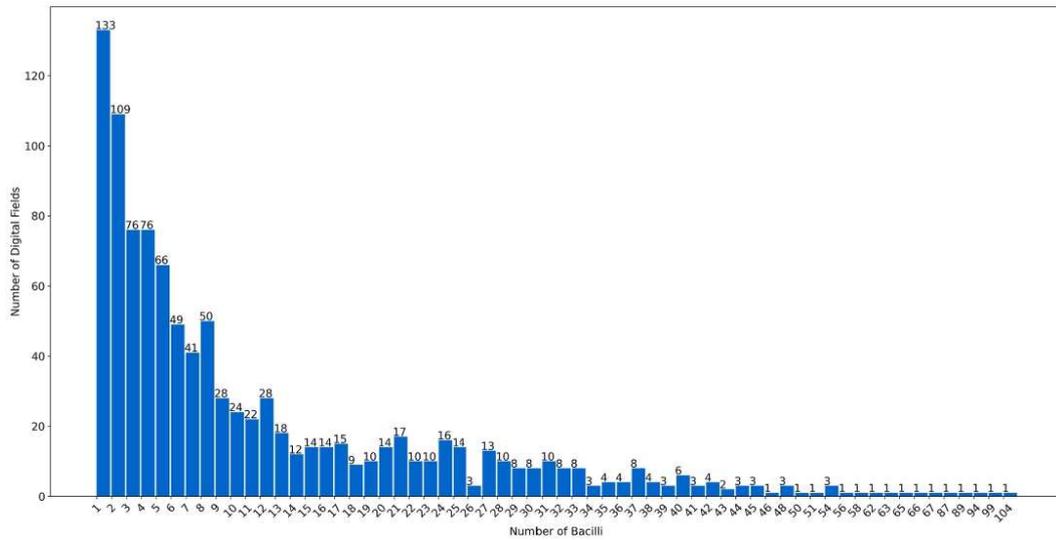


Fig. 3. Number of DF images versus number of bacilli per DF image used for building the 250 DF image sets to evaluate TB diagnosis.

and the camera, which reduces the trinocular aperture of the microscope to 61% of the microscopic field. Thus, the area of a digital field is approximately  $0.0243 \text{ mm}^2$ . To evaluate the area equivalent to 100 microscopic fields, it will be necessary to acquire approximately 164 digital fields (DF). Depth resolution of DF images is 24 bits (true color images) and the dimension is  $1388 \times 1040$  pixels. The dimension of each pixel is  $0.13 \mu\text{m} \times 0.13 \mu\text{m}$ . The conversion of the diagnostic classes shown in Table 1 to DF images is as follows: for negative, scanty and "1+", a reading of 164 DF images is required. For the diagnosis classes, "2+" and "3+", a reading of 82 and 32 DF images, respectively, is required. Each DF image was annotated by 3 experts in sputum smear analysis for TB diagnosis.

All images used in this study are extended depth focus (EDF) images. The EDF images were obtained by applying a multifocus image fusion method, developed by our research group [36]. From each microscopic field, 11 images were acquired at different focal depths, composing an image stack. The first image was acquired using the autofocus function of the microscope; then, using a focal length step of  $0.52 \mu\text{m}$ , 5 images were acquired upwards and 5 images downwards, relative to the first one. From this stack of 11 images, the developed image fusion method generated an EDF image. The aim was to bring all bacilli in a microscopic field to the same focal depth. The whole fusion process is described in Ref. [36].

Two annotated datasets were used in this work: DDS3 dataset and DDS4 dataset. The description of these datasets is shown in Table 2. Each slide corresponds to 1 patient. From each slide, 2 tiles of 164 DF images are extracted. As shown, DDS3 is composed of 6000 mosaic images of  $400 \times 400$  pixels and was used for semantic segmentation. A total of 80% of the data is used for training, while 20%, for testing. We ensured that patient slides used for training were not used for testing. The DDS4 dataset was used to evaluate TB diagnosis and consists of 250 sets of 164 DF images. To build these sets, the DF images were extracted from slides S16–S25. Fig. 3 shows the number of DF images versus the number of bacilli per DF image extracted for building these sets. The number of bacilli per DF image varies between 1 and 104. Additionally, 2270 negative DF images were extracted (DF images without any bacillus).

No cases were judged by cell culture results because the goal of this

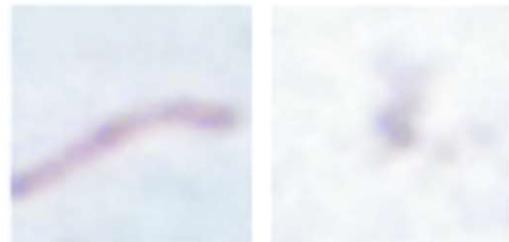


Fig. 4. Example of patches used in the color filter: (a) true positive patch; (b) false positive patch.

work is to automate the microscopic analysis of sputum smears performed by experts. Therefore, our gold standard is the dataset of digital field images that have been annotated by experts.

To train color and shape filters, we used a group of 360 FPs and 360 true positives bacilli, resulting from applying the semantic segmentation in the test set. A combination of 80% and 20% of this group were used for training and testing the filters, respectively. For the color filter, each bacillus is centralized in a  $40 \times 40$ -pixel patch. Fig. 4 shows examples of patches containing true positive and FP bacillus.

#### 4.1.2. Development environment

The simulations were performed on a computer with an Intel Core i5-9300H, 2.40 GHz CPU, 8 GB RAM, NVIDIA GeForce GTX 1650 graphics unit processor (4 GB RAM). The work was developed in the Google Colaboratory environment, using Python version 3.7.13, with TensorFlow 2.8.2. The libraries skimage version 0.18.3, sklearn version 1.0.2, and Panda's version 1.3.5 were used to develop the color and shape filters. Image manipulation was done using the libraries NumPy version 1.21.6, OpenCV version 4.1.2 and Matplotlib version 3.2.2.

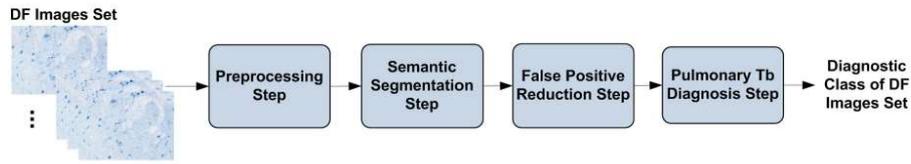


Fig. 5. The proposed methodology comprises 4 steps: preprocessing, semantic segmentation, FP reduction and TB diagnosis.

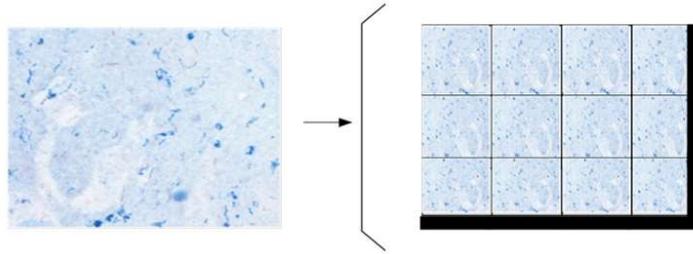


Fig. 6. Splitting a DF image of  $1040 \times 1388$  pixels into a  $3 \times 4$  grid of images of  $400 \times 400$  pixels. One image of  $400 \times 400$  pixels is used as input to the CNN that implements the semantic segmentation step. The images located in the last row and last column are padded with zeros.

4.2. Methods

The methodology used in this work for TB diagnosis, shown in the block diagram of Fig. 5, comprises 4 steps: preprocessing, semantic segmentation, filtering, and bacilli counting.

In the preprocessing step, the DF image is divided into  $400 \times 400$ -pixels images. In the semantic segmentation step, presented in Ref. [22], the bacilli are segmented using a convolutional neural network (CNN1).

In the next step, 2 filters are applied: a color filter, implemented with one CNN (CNN2) and a shape filter, implemented with one RF classifier

(RF1). These filters are implemented with the aim of reducing the number of FPs obtained in the semantic segmentation step. Each filter classifies a segmented object as a bacillus or not.

To evaluate the performance of the method presented in this study, the following metrics were employed: sensitivity (recall), specificity, precision, accuracy, and F1-score. These metrics are defined as follows.

$$sensitivity (recall) = TP / (TP + FN) \tag{1}$$

$$specificity = TN / (TN + FP) \tag{2}$$

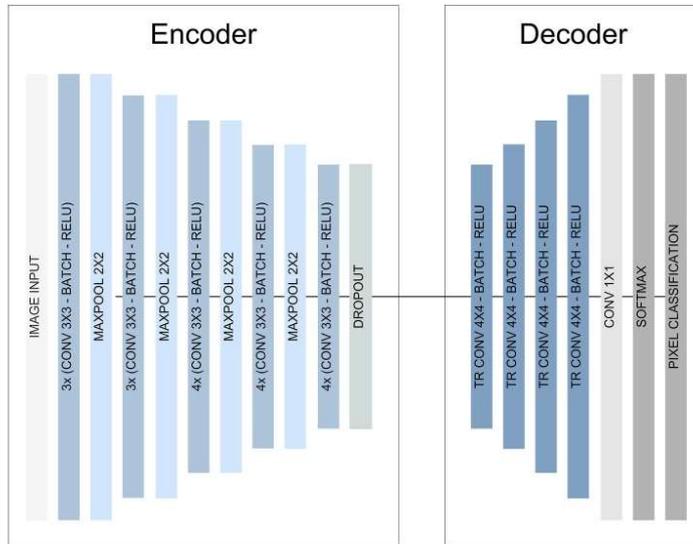


Fig. 7. CNN1 architecture used for bacilli semantic segmentation [22].



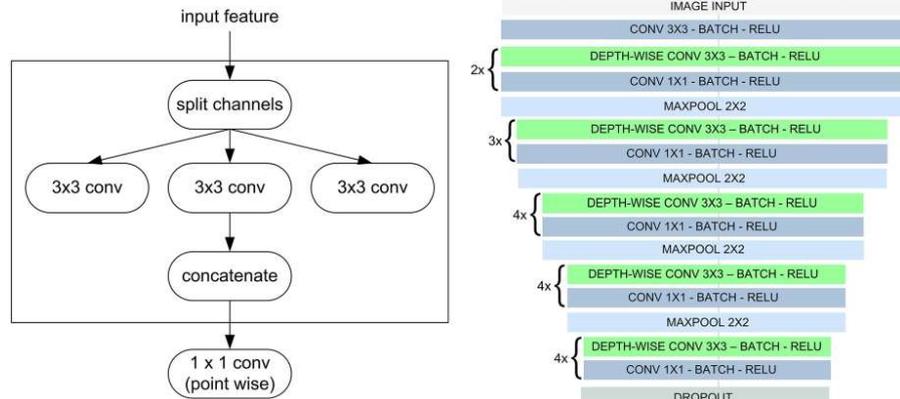


Fig. 9. (a) Depth-wise separable convolution layer; (b) Encoder using depth-wise separable convolution layer.

The encoder with channel SENet attention mechanism is illustrated in Fig. 8(b), and consists of an input layer, 4 subsampling stages and a dropout layer with dropout rate of 0.3. Each subsampling stage consists of 1 convolution layer with a  $3 \times 3$  kernel followed by sequences of a SENet block and a convolution layer with  $3 \times 3$  kernel. After these layers, a  $2 \times 2$  max pooling layer is used. All convolution layers are followed by the batch normalization layer and the ReLU activation layer.

Depth-wise separable convolution, proposed by Ref. [25], is an operation that performs a separated spatial convolution on each channel of a feature input, mixing the output via a pointwise convolution ( $1 \times 1$  convolution). Fig. 9(a) illustrates this operation. In the same way as the regular convolution, depth-wise separable convolution relies on the assumption that spatial locations in intermediate activations are highly correlated, but channels are highly independent. Compared to regular convolution, depth-wise separable convolution requires fewer computations and requires fewer parameters. Originally, it was used in the Xception classification network.

The encoder with depth-wise separable convolution, illustrated in Fig. 9(b), consists of an input layer followed by a convolution layer with a  $3 \times 3$  kernel, 4 subsampling stages and a dropout layer with a dropout rate of 0.3. Each subsampling stage consists of a depth-wise convolution layer with  $3 \times 3$  kernel and convolution layer with  $1 \times 1$  kernel. After the sequences, there is a  $2 \times 2$  max pooling layer. All convolution and depth-wise convolution layers are followed by a batch normalization layer and ReLU activation layer.

As mentioned before, the last network used in this work for evaluating bacilli segmentation was the U-Net. The U-Net architecture [27] is widely used in semantic segmentation tasks, particularly in medical image processing. It has been implemented to perform accurate segmentation even with few training images. The architecture consists of an encoder with 4 subsampling stages. Each subsampling stage is composed of 2 convolution layers with  $3 \times 3$  kernel followed by a  $2 \times 2$  max pooling layer. The decoder consists of 4 oversampling stages. Each oversampling stage is composed of a  $2 \times 2$  up-convolution layer, a concatenation with the corresponding feature map from encoder and 2 convolution layers with  $3 \times 3$  kernel. This architecture connects encoder layers with decoder layers. The authors evaluated the U-Net for segmentation of neuronal structures in electron microscopic recordings and cell segmentation in light microscopic images using 2 different datasets and achieved good performance for all the evaluated tasks.

The semantic segmentation is applied in all regions of the DF image. The method then identifies segmented objects located in areas with

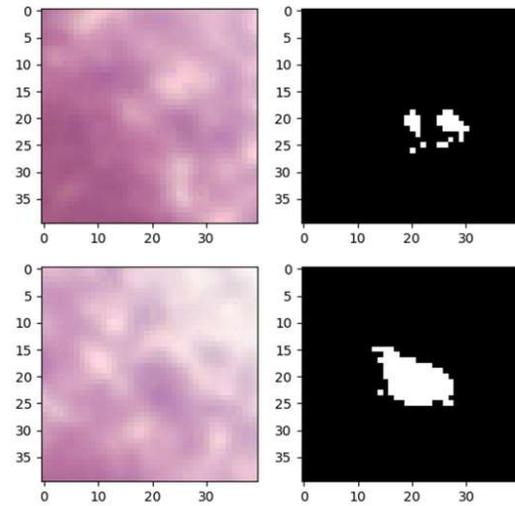


Fig. 10. Examples of objects segmented in regions with staining problems. (a) Example 1 and (b) example 2.

staining problems and discards them. To identify staining problems, we investigated the value of the hue (H) component in the neighborhood of all segmented objects. A  $40 \times 40$ -pixel window is centered in each segmented object, and a thresholding operation using the H component is applied in this window. If more than 50% of the pixels of the window have an H value higher than  $266^\circ$  (fuchsia color), the object is discarded. Fig. 10 shows 2 examples of objects segmented in areas with staining problem.

After semantic segmentation, one area filter removes objects smaller than 25 pixels or larger than 900 pixels that correspond to artifacts.

#### 4.2.3. False positive reduction

Aiming to reduce the number of FPs, each object detected in semantic segmentation is submitted to a filtering process that uses shape and color filters. As shown in Fig. 10, two filtering models were

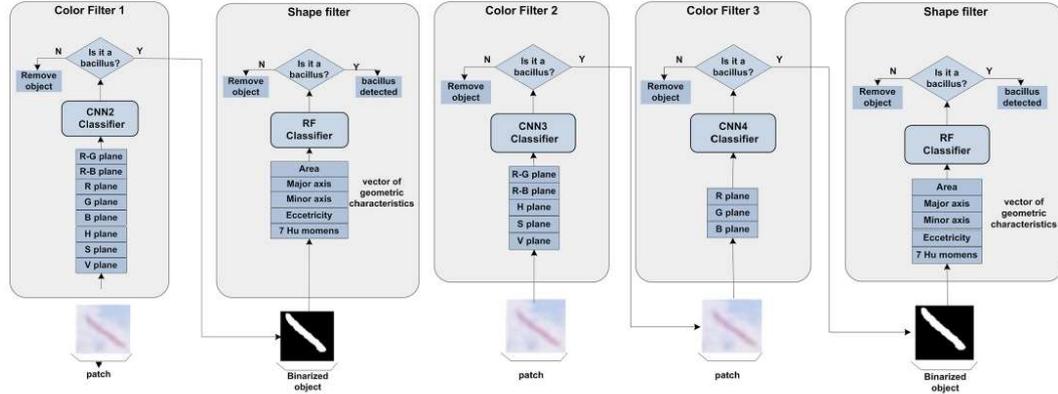


Fig. 11. Illustration of the filtering models. (a) Filtering model #1 with 1 color filter and 1 shape filter; (b) filtering model #2 with 2 color filters and 1 shape filter.

proposed. In filtering model #1, we propose a pipeline of one-color filter and one-shape filter. In filtering model #2, we propose a pipeline of two-color filters and one-shape filter.

What differentiates the color filters are the inputs. The color filter of model #1 uses 8 input planes: R-G, R-B, R, G, B, H, S and V. Color filter 1 of model #2 uses 5 input planes: R-G, R-B, H, S and V, while color filter 2 of model #2 uses 3 input planes: R, G and B.

Color filters are implemented with the same CNN2 architecture: 3 convolutional layers with a  $5 \times 5$  kernel, a  $2 \times 2$  max pooling layer, a flattened layer, a 120-neuron dense layer, 84-neuron dense layer, and the output layer, with a sigmoid function.

The RGB color space was used in this work because it is the original space in which the images are represented, and because it was used in the other studies previously published in the literature. The importance of the R-G feature for bacilli segmentation was already verified in a previous work [32]. The HSV color space was chosen because it has characteristics like the human visual system, in which intensity information (H) is uncoupled from color information (S and V).

The goal of the shape filter is to focus attention on geometric features of the segmented objects. As shown in Fig. 11, from each segmented object, a set of 11 geometric features were extracted: area, major axis, minor axis, eccentricity and the 7 Hu moments, which are invariant to translation, scale, and rotation. Yousefi et al. [11] also used Hu moments for detecting TB bacilli. The above-mentioned geometric features are described as follows.

- Major axis and minor axis: maximum and minimum distance from the bacillus boundary to the center of mass  $(\bar{x}, \bar{y})$ , corresponding to  $R_{max}$  and  $R_{min}$ .
- Eccentricity: ratio between the major axis and minor axis, as shown in equation (9).

$$eccentricity = \frac{R_{max}}{R_{min}} \quad (9)$$

- Hu moments: descriptors used to characterize the shape of a binary object in an image. There are 7 Hu moments. To calculate Hu moments, we initially define relative moments, which are centered around the centroid, as shown in equation (10).

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q I(x, y) \quad (10)$$

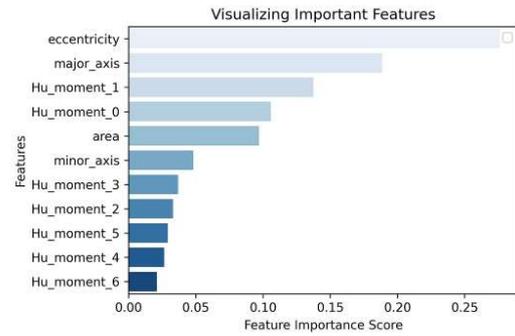


Fig. 12. Importance of geometric variables in the outcome of the RF classifier used in the shape filter. Eccentricity is the variable that most influences the classifier decision, with a feature importance score higher than 0.25, followed by major axis feature.

where:

$(\bar{x}, \bar{y})$  - centroid coordinates

$p, q$  - moment order in  $x$  and  $y$  direction, respectively

$$I(x, y) \in \{0, 1\}.$$

- The 7 Hu moments are defined in equations (11)–(17).

$$M_1 = \mu_{20} + \mu_{02} \quad (11)$$

$$M_2 = (\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2 \quad (12)$$

$$M_3 = (\mu_{30} - 3\mu_{12})^2 + (3\mu_{21} - \mu_{30})^2 \quad (13)$$

$$M_4 = (\mu_{30} + \mu_{12})^2 + (\mu_{21} + \mu_{03})^2 \quad (14)$$

$$M_5 = (\mu_{30} - 3\mu_{12})(\mu_{30} + \mu_{12})((\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2) + (3\mu_{21} - \mu_{03})(\mu_{21} + \mu_{03})(3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2) \quad (15)$$

$$M_6 = (\mu_{20} - \mu_{02})((\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2) +$$

Table 3

Results of semantic segmentations obtained in this work (boldface) compared with previously published results.

Semantic Segmentation Network	TP	FP	TN	FN	Acc.	Prec	Sens.	Spec.	F1-score
Serrao et al. [22]	49,753	167	49,657	536	0.9930	0.9967	0.9893	0.9966	0.9930
<b>Network with depth-wise separable convolution layers</b>	50,064	140	49,610	225	<b>0.9964</b>	<b>0.9972</b>	0.9955	<b>0.9972</b>	<b>0.9964</b>
<b>Network with SENet attention mechanism</b>	50,116	373	49,514	173	0.9945	0.9926	<b>0.9966</b>	0.9925	0.9946
U-Net network	50,016	2787	49,162	273	0.9701	0.9472	0.9946	0.9464	0.9703

Table 4

Reduction in the number of FPs resulting from using the 2 filtering models, on 10 sets of 164 DF images from DDS4 dataset (semantic segmentation with depth-wise convolution layers).

Step	False Positives (mean ± SD)
Semantic segmentation	546 ± 125
Semantic segmentation + filtering model #1	20 ± 7
Semantic segmentation + filtering model #2	10 ± 3

$$4\mu_{11}(\mu_{30} + 3\mu_{12})(\mu_{21} + \mu_{03}) \quad (16)$$

$$M_7 = (3\mu_{21} - \mu_{03})(\mu_{30} + \mu_{12})((\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2) - (\mu_{30} - 3\mu_{12})(\mu_{21} + \mu_{03})(3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2) \quad (17)$$

A Random Forest classifier, RF, was trained to recognize objects with bacilli shape. The parameters of the RF classifier are the following: tree depth equal to 10 and maximum number of nodes equal to 100. The accuracy in the testing set was 95.24%.

As shown in Fig. 12, eccentricity is the variable that most influences the RF classifier decision. The second most relevant feature is the major axis, which is in accordance with the fact that the bacillus is an elongated rod-like shape.

## 5. Results

Table 3 presents the results of semantic segmentations obtained in this work (boldface) and compares them with previously published results.

As shown in Table 3 the best values for accuracy, precision, specificity, and F1-score were obtained with the network using depth-wise separable convolution layers in the encoder. The best sensitivity was obtained with the network using the SENet attention mechanism in the encoder. The U-Net network showed the worst results, except for sensitivity, where it obtained a better value than the results presented in Ref. [22]. As the best semantic segmentation results were obtained with

the autoencoder using the depth-wise separable convolution layer, the results presented hereafter use it.

The minimum value of FPs, 140, obtained with the network using depth-wise separable convolution layers in the encoder, is still very high. The sequential application of the filters described in section 4.2.3 could decrease the number of FP. s.

To evaluate the performance of these filters, we took 10 sets of 164 negative DF images and detected the number of bacilli in each one. Table 4 shows the reduction in the number of FPs with filtering model #1 and #2, respectively. As shown, the color filter accounts for the largest reduction in FPs. Both methods resulted in a reduction greater than 96% in the number of FPs. Table 4 also shows that filtering model #2 was superior to model #1. Thus, the next step, the TB diagnosis, was performed using the model #2.

The confusion matrices for TB diagnosis using the DDS4 dataset are shown in Fig. 13. Fig. 13(a) shows the result of using only semantic segmentation with depth-wise convolution layers. Fig. 13(b) shows the result of using semantic segmentation with depth-wise convolution layers + filtering model #2.

Table 5 presents the metrics precision, recall and F1-score for TB diagnosis obtained with filtering model #2. Fig. 14 shows the result of bacilli detection in 5 DF images, extracted from different diagnostic set images. Table 6 shows the time performance of each methodological step, using filtering model #2, for 11 DF images with a different number of bacilli. Table 7 shows the mean time and standard deviation (SD)

Table 5

Performance of the proposed method for TB diagnosis using semantic segmentation with depth-wise convolution layer + filtering model #2.

Class	Precision	Recall	F1-score
Negative	0.771	0.740	0.755
Scanty	0.740	0.740	0.740
1+	0.961	1	0.980
2+	1	1	1
3+	1	1	1
Mean	0.894	0.896	0.895

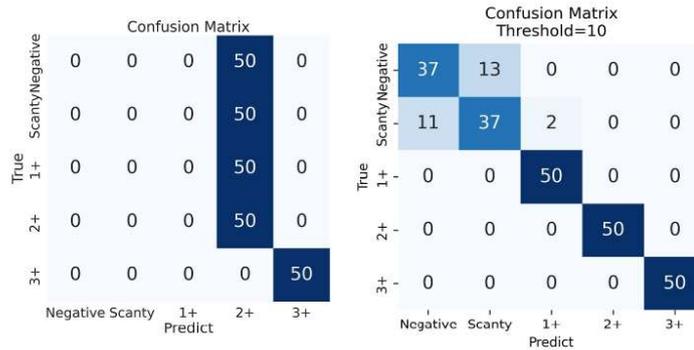
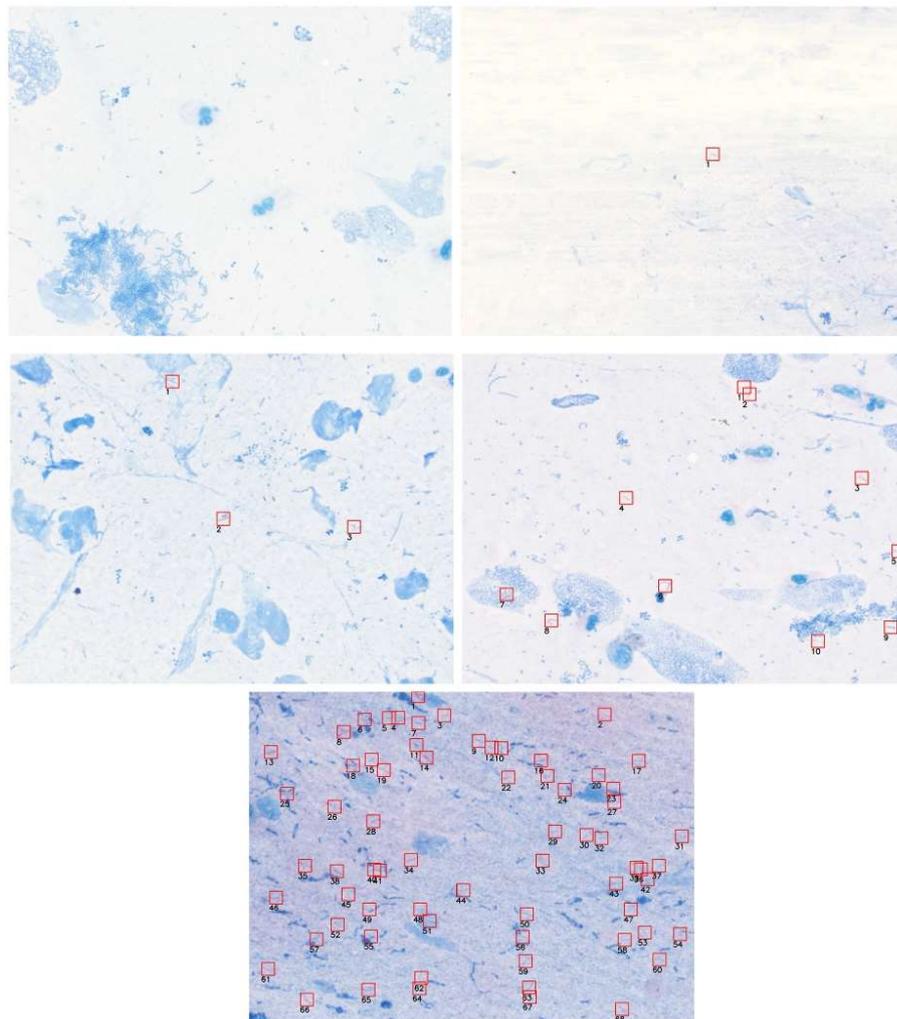


Fig. 13. Confusion matrix for TB diagnosis using (a) semantic segmentation with depth-wise convolution layer; (b) semantic segmentation with depth-wise convolution layer + filtering model #2. The horizontal axis represents the predicted class, while the vertical axis represents the gold standard.



**Fig. 14.** Bacilli detected in a DF image extracted from: (a) set of DF images of negative diagnostic class; (b) set of DF images of scanty diagnostic class; (c) set of DF images of 1+ diagnostic class; (d) set of DF images of 2+ diagnostic class; (e) set of DF images of 3+ diagnostic class. Each detected bacillus is enclosed in a red bounding box.

obtained with filtering model #2, for each diagnostic class.

## 6. Discussion

Most of the papers published in the literature did not follow the pulmonary TB diagnosis by sputum smear microscopy according to WHO recommendations (WHO, 1998) [24]. They only focus on bacilli detection, or in classifying TB infection in 2 categories, low-level or high-level [23,29] or in classifying a slide as positive or negative [31, 32]. However, counting bacilli is an intermediate step in achieving an automatic method for TB diagnosis.

Compared with the literature, the results obtained for AFB detection in this work (semantic segmentation step) were higher than the other

results previously in the same database. Costa Filho et al. [3] obtained a sensitivity and specificity of 96.80% and 96.62%, respectively. In this work, a sensitivity of 99.55% and a specificity of 99.72% were obtained. This performance was only possible due to the use of a semantic segmentation strategy using depth-wise separable convolution layers.

However, as demonstrated in this study, this specificity value was not sufficient for TB diagnosis, because, in counting the number of AFB present in at least 164 DF images, a very high number of FPs were found. The confusion matrices of Fig. 13(a) show that only the semantic segmentation step does not ensure reliable diagnoses for the negative, scanty, and 1+ classes. The 50 testing sets of negative, scanty, and 1+ classes were classified as 2+.

Two filtering models were evaluated to reduce the number of FPs.

**Table 6**  
Time performance for 11 DF images with different numbers of bacilli.

DF image	Number of bacilli (ground truth)	Time (s)			Total
		Preprocessing	Semantic segmentation with depth-wise convolution	FP reduction	
1	0	0.002	2.423	0.003	2.428
2	0	0.002	2.39	0.017	2.409
3	10	0.017	2.774	0.635	3.426
4	10	0.011	2.668	0.635	3.314
5	20	0.026	2.494	0.899	3.419
6	20	0.016	2.395	1.047	3.442
7	30	0.031	2.469	1.707	4.207
8	30	0.078	2.484	1.655	4.217
9	40	0.026	2.538	2.577	5.204
10	40	0.031	2.421	2.082	4.503
11	65	0.076	2.635	4.356	7.067

**Table 7**  
Time performance for each class of pulmonary TB diagnostic by bright-field smear microscopy.

Class	Time (s) (mean $\pm$ SD)
Negative	410.349 $\pm$ 33.96
Scanty	422.745 $\pm$ 25.289
1+	426.059 $\pm$ 29.333
2+	230.244 $\pm$ 2.22
3+	150.597 $\pm$ 6.218

Filtering model #1 uses a CNN with 8 color descriptors as inputs. Filtering model #2 comprises 2 color filters. The first one uses 1 CNN with 5 color descriptors as inputs and the second one uses another CNN with 3 color descriptors as inputs. As shown in Fig. 13(b), using filtering model #2 after the semantic segmentation, diagnostic performance improves considerably.

The main difficulty in developing an automatic pulmonary TB diagnosis by bright-field smear microscopy arises from the fact that it requires the availability of a large, robust, and annotated TB image dataset, since the diagnosis of a single patient requires counting the number of bacilli in at least 100 microscopic fields (164 DF images).

In order to evaluate the performance of the automatic TB diagnosis method based on international guidelines (WHO, 1998) [24], we built an annotated dataset, DDS4, comprised of 250 sets (50 sets of each diagnostic class) of 100 microscopic fields (164 DF images).

The main concern of the proposed TB diagnosis method was to reduce the number of FPs, by using color and shape filters. Problems with detecting FPs in pathogen examinations with microscopic tests have also been reported by other authors [30,37–39]. Since it was not possible to reduce the number of FPs to zero, the developed pulmonary TB diagnosis method was applied to 10 sets of 164 DF images with negative diagnosis. For filtering model #2, an average number of  $10 \pm 3$  FP swere obtained. Therefore, the developed method had an average positive bias of 10 FPs per set of 164 DF images.

The confusion matrix in Fig. 13(b) shows that the proposed method correctly classified all the 50 testing sets of 1+, 2+, and 3+ classes. Concerning the classes negative and scanty, as shown in Table 5, a precision of 0.771 and 0.74, respectively, were obtained. For TB diagnosis, the mean values of the metrics precision, recall and F1-score obtained with filtering model #2 were 89.4%, 89.6%, and 89.5%, respectively.

As shown in Table 7, the time performance for TB diagnosis of negative, scanty and 1+ classes is approximately 7 min. However, the time performance for TB diagnosis of class 3+, is 3.5 min. This is because, as shown in Table 1, for diagnostic class 3+, a smaller number of DF images are required. In Ref. [28] the authors state that evaluating

a TB diagnostic, the microscopy technician should take approximately 5 min. However, as stated in section 2, the basic level microscopy laboratory of a national tuberculosis control program in low-income countries have other tasks to perform besides sputum smear microscopy. Considering that a workload of 20–25 examinations per day is the maximum that can reasonably be expected from a single worker and that the daily demand may be higher, these facts may result in worker overload and, consequently, poor quality smear microscopies [24]. Therefore, the automation of the TB diagnosis will help to increase the productivity and quality of the examinations carried out.

In summary, the characteristics of the proposed method are the following.

1. Implements automatic TB diagnosis using 5 grading scales (negative, scanty, 1+, 2+ and 3+), according to the WHO recommendations.
2. Presents a diagnosis time of 7 min in the worst case.
3. Reduces false positive rate by an efficient filtering method.
4. Uses appropriate mechanisms to handle channel information in the semantic segmentation of bacilli.
5. Detects and removes areas of the slide with staining problems.
6. Presents mean precision of 89.4%, mean recall of 89.6% and mean F1-score of 89.5%

The fluorescence microscopy has, at least, a 10% higher sensitivity than bright-field smear microscopy, when the techniques are analyzed in the real-world context of tuberculosis patients with active disease. Although, in many countries, the bright-field smear microscopy is still the first diagnostic exam employed (and available) for a group of presumptive TB patients. It is simple, economic, and efficient in detecting most infectious cases. Furthermore, bright-field smear microscopy has the advantage of requiring a much simpler training, because the ability to identify the bacilli by this method is much simpler to acquire. Besides this, the Auramine-O solution used in fluorescence microscopy is carcinogenic, posing health risks to laboratory staff [40]. In our study, we observed that the proposed method has high agreement (co-positivity) with the images obtained by bright-field smear microscopy. This shows that our results, obtained with the proposed method, are very promising.

## 7. Conclusion

The main contribution of this work is proposing an automatic and rapid pulmonary TB diagnosis in bright-field smear microscopy according to the Union and WHO recommendations. The proposed method comprises a semantic segmentation step, followed by the application of color and shape filters, aiming to reduce FPs. The following performance metrics were obtained: mean precision of 89.4%, mean recall of 89.6% and mean F1-score of 89.5%. This was only possible by building a large, robust, and annotated dataset.

### CRedit authorship contribution statement

**Mikaela Kalline Maciel Serrão:** Formal analysis, Investigation, Methodology, Writing – original draft, Software. **Marly Guimarães Fernandes Costa:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Visualization, Writing – original draft, Writing – review & editing. **Luciana Botinelly Mendonça Fujimoto:** Investigation, Methodology. **Mauricio Morishi Ogusku:** Investigation. **Cícero Ferreira Fernandes Costa Filho:** Formal analysis, Investigation, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing.

### Declaration of competing interest

The authors declare the following financial interests/personal

relationships which may be considered as potential competing interests: Cicero Ferreira Fernandes Costa Filho reports financial support was provided by Samsung Electronics of Amazonia Ltda. Marly Guimaraes Fernandes Costa reports financial support was provided by Samsung Electronics of Amazonia Ltda. Mikaela Kalline Serrao reports financial support was provided by Samsung Electronics of Amazonia Ltda. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Acknowledgment

This research, carried out within the scope of the Samsung-UFAM Project for Education and Research (SUPER), according to Article 39 of Decree n°10.521/2020, and was funded by Samsung Electronics of Amazonia Ltda., under the terms of Federal Law n°8.387/1991 through agreement 001/2020, signed with UFAM and FAEPI, Brazil. The authors would also like to thank CAPES for its financial support. Jim Hesson edited the manuscript ([academicenglisholutions.com](mailto:academicenglisholutions.com)).

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.compbiomed.2024.108167>.

#### References

- [1] World Health Organization, Global Tuberculosis Report 2021, World Heal Organ, 2021.
- [2] R. Khutlang, S. Krishnan, R. Dendere, A. Whitelaw, K. Veropoulos, G. Learmonth, et al., Classification of Mycobacterium tuberculosis in images of ZN-stained sputum smears, *IEEE Trans. Inf. Technol. Biomed. Eng.* 14 (2010) 949–957, <https://doi.org/10.1109/ITTB.2009.2028339>.
- [3] C.F.F. Costa Filho, P.C. Levy, C. de Matos Xavier, L.B. Mendonça Fujimoto, M. G. Fernandes Costa, Automatic identification of tuberculosis mycobacterium, *Rev. Bras Eng. Biomed.* 31 (2015) 33–43, <https://doi.org/10.1590/2446-4740.0524>.
- [4] World Health Organization (WHO), Laboratory Services in Tuberculosis Control. Part I: Organization and Management, vols. 1–63, World Heal Organ Geneva, Switz, 1998.
- [5] S. Aayas, M. Ekinci, Random forest-based tuberculosis bacteria classification in images of ZN-stained sputum smear samples, *Signal, Image Video Process* 8 (2014) 49–61, <https://doi.org/10.1007/s11760-014-0708-6>.
- [6] K.R. Steingart, M. Henry, V. Ng, P.C. Hopewell, A. Ramsay, J. Cunningham, et al., Fluorescence versus conventional sputum smear microscopy for tuberculosis: a systematic review, *Lancet Infect. Dis.* 6 (2006) 570–581, [https://doi.org/10.1016/S1473-3099\(06\)70578-3](https://doi.org/10.1016/S1473-3099(06)70578-3).
- [7] F.M. Aziz, H.A. Nugroho, S. Wibirama, Adaptive threshold determination based on entropy in active contour without edge method for malaria parasite candidate detection, in: *Proc - 2018 4th Int Conf Sci Technol ICST 2018*, 2018, <https://doi.org/10.1109/ICSTC.2018.8528585>, 0–5.
- [8] M. Rico-García, A. Salazar, C.-A. Madrigal, L.-J. Morantes-Guzman, F.M. Cortes-Mancera, Detection of Mycobacterium Tuberculosis in Microscopic Images of Ziehl-Neelsen-Stained Sputum Smears, 6th Lat. Am. Conf. Networked Electron. Media, 2015, pp. 1–6, <https://doi.org/10.1049/ic.2015.0312>.
- [9] K.S. Mithra, W.R.S. Emmanuel, Segmentation and classification of mycobacterium from Ziehl Neelsen stained sputum images for tuberculosis diagnosis, *IEEE Int. Conf. Commun. Signal Process* (2017), <https://doi.org/10.1109/ICCSP.2017.8286676>, 1672–6.
- [10] R.S. Soans, A.G. Ramakrishnan, V.P. Shenoy, R.R. Gallikere, Classification of bacterial morphotypes from images of ZN-stained sputum-smears towards diagnosing drug-resistant TB, in: *2016 Int Conf Signal Process Commun SPCOM*, 2016, pp. 1–5, <https://doi.org/10.1109/SPCOM.2016.7746682>, 2016.
- [11] H. Yousefi, F. Mohammadi, N. Mirian, N. Amini, Tuberculosis bacilli identification: a novel feature extraction approach via statistical shape and color models, *Proc - 19th IEEE Int. Conf. Mach. Learn. Appl. ICMLA* (2020) 366–371, <https://doi.org/10.1109/ICMLA51294.2020.00065>, 2020.
- [12] E. Sugirtha, G. Muruges, Sputum smear images, Detect Tuberc Bacilli from Microsc Sputum Smear Images 16–8 (2017).
- [13] Y. Zhai, Y. Liu, D. Zhou, S. Liu, Automatic identification of mycobacterium tuberculosis from ZN-stained sputum smear: algorithm and system design, in: *2010 IEEE Int Conf Robot Biomimetics, ROBOT*, 2010, <https://doi.org/10.1109/ROBOT.2010.5723300>, 2010:41–6.
- [14] R. Rulaningtyas, A.B. Suksmono, T.L.R. Mengko, Automatic classification of tuberculosis bacteria using neural network, *Proc 2011 Int. Conf. Electr. Eng. Informatics, ICEEI* (2011) 1–4, <https://doi.org/10.1109/ICEEI.2011.6021502>, 2011.
- [15] M.I. Shah, S. Mishra, M. Sarkar, C. Rout, Automatic detection and classification of tuberculosis bacilli from ZN-stained sputum smear images using watershed segmentation, in: *International Conf. Signal Process., IET, Vidisha, India*, 2016, pp. 1–4, <https://doi.org/10.1049/cp.2016.1459>.
- [16] K.S. Mithra, W.R. Sam Emmanuel, GFNN: Gaussian-Fuzzy-Neural network for diagnosis of tuberculosis using sputum smear microscopic images, *J. King Saud Univ. - Comput. Inf. Sci.* 33 (2021) 1084–1095, <https://doi.org/10.1016/j.jksuci.2018.08.004>.
- [17] Y.P. López, C.F.F. Costa Filho, L.M.R. Aguilera, M.G.F. Costa, Automatic classification of light field smear microscopy patches using Convolutional Neural Networks for identifying, Mycobacterium Tuberculosis. 2017 Chil. Conf. Electr. Electron. Eng. Inf. Commun. Technol. (2017) 1–5, <https://doi.org/10.1109/CHILECON.2017.8229512>.
- [18] Y. Xiong, X. Ba, A. Hou, K. Zhang, L. Chen, T. Li, Automatic detection of mycobacterium tuberculosis using artificial intelligence, *J. Thorac. Dis.* 10 (2018) 1936–1940, <https://doi.org/10.21037/jtd.2018.01.91>.
- [19] J.A. Quinn, R. Nakasi, P.K.B. Mugagga, P. Byanyima, W. Lubega, A. Andama, Deep convolutional neural networks for microscopy-based point of care diagnostics, *Proc. Int. Conf. Mach. Learn. Heal. Care* 56 (2016) 1–12.
- [20] R.O. Panicker, K.S. Kalmady, J. Rajan, M.K. Sabu, Automatic detection of tuberculosis bacilli from microscopic sputum smear images using deep learning methods, *Biocybern. Biomed. Eng.* 38 (2018) 691–699, <https://doi.org/10.1016/j.bbe.2018.05.007>.
- [21] R. Dinesh Jackson Samuel, B. Rajesh Kanna, Tuberculosis (TB) detection system using deep neural networks, *Neural Comput. Appl.* 31 (2019) 1533–1545, <https://doi.org/10.1007/s00521-018-3564-4>.
- [22] M.K.M. Serrão, M.G.F. Costa, L.B. Fujimoto, M.M. Ogusku, C.F.F.C. Filho, Automatic Bacillus detection in light field microscopy images using convolutional neural networks and mosaic imaging approach, *Proc Annu Int Conf IEEE Eng Med Biol Soc EMBS* (2020), <https://doi.org/10.1109/EMBC44109.2020.9176105>, 2020-July:1903–6.
- [23] K.S. Mithra, W.R. Sam Emmanuel, Automated identification of mycobacterium bacillus from sputum images for tuberculosis diagnosis, *Signal, Image Video Process* 13 (2019) 1585–1592, <https://doi.org/10.1007/s11760-019-01509-1>.
- [24] H.L. Rieder, A. Van Deun, K. Man Kam, S. Jae Kim, T.M. Chonde, A. Trebucaq, et al., Priorities for Tuberculosis Bacteriology Services in Low-Income Countries, vol. 2, 2007.
- [25] F. Chollet, Xception: deep learning with depth-wise separable convolutions, *Proc - 30th IEEE Conf Comput Vis Pattern Recognition, CVPR* (2017), <https://doi.org/10.1109/CVPR.2017.195>, 2017:2017-Janua:1800–7.
- [26] J. Hu, L. Shen, S. Albanie, G. Sun, E. Wu, Squeeze-and-Excitation networks, *IEEE Trans. Pattern Anal. Mach. Intell.* 42 (2020) 2011–2023, <https://doi.org/10.1109/TPAMI.2019.2913372>.
- [27] W. Weng, X. Zhu, U-net: convolutional networks for biomedical image segmentation, *IEEE Access* 9 (2021) 16591–16603, <https://doi.org/10.1109/ACCESS.2021.3053408>.
- [28] R. Lumb, A. Van Deun, I. Jhastian, M.G. Fitz-Gerald, Laboratory Diagnosis of Tuberculosis by Sputum Microscopy: the Handbook, Global, Adelaide, 2013.
- [29] K. Swetha, B. Sankaragomathi, J.B. Thangamalar, Convolutional neural network based automated detection of Mycobacterium Bacillus from sputum images, in: *Proc. 5th Int. Conf. Inven. Comput. Technol.*, vol. 345, IEEE, E 47TH ST, NEW YORK, NY 10017 USA, 2020, pp. 293–300.
- [30] M. El-Melegy, D. Mohamed, T. Elmelegy, M. Abdelrahman, Identification of tuberculosis bacilli in ZN-stained sputum smear images: a deep learning approach, *IEEE Comput. Soc. Conf. Comput. Vis Pattern Recognit Work* (2019), <https://doi.org/10.1109/CVPRW.2019.00147>, 2019-June:1131–7.
- [31] S. Zuraç, C. Mogodici, T. Poncu, M. Trăscău, C. Popp, L. Nichita, et al., A new artificial intelligence-based method for identifying Mycobacterium tuberculosis in ziehl-neelsen stain on tissue, *Diagnostics* 12 (2022), <https://doi.org/10.3390/diagnostics12061484>.
- [32] M. Yang, K. Nurzynska, A.E. Walte, A. Gertych, A CNN-based active learning framework to identify mycobacteria in digitized Ziehl-Neelsen stained human tissues, *Comput. Med. Imag. Graph.* 84 (2020) 101752, <https://doi.org/10.1016/j.compmedimag.2020.101752>.
- [33] S.N. Mohanty, G. Nalinipriya, O.P. Jena, A. Sarkar, A lightweight convolutional neural network model for tuberculosis bacilli detection from microscopic sputum smear images rani oomnan panicker, in: S.J. Pawan, Jey Rajan, M.K. Sabu (Eds.), *Mach. Learn. Healthc. Appl. First*, John Wiley & Sons, 2021, p. 390.
- [34] C.P. Kuok, M.H. Horng, Y.M. Liao, N.H. Chow, Y.N. Sun, An effective and accurate identification system of Mycobacterium tuberculosis using convolution neural networks, *Microsc. Res. Tech.* 82 (2019) 709–719, <https://doi.org/10.1002/jemt.23217>.
- [35] M.G.F. Costa, C.F.F. Costa Filho. <https://tbimages.ufam.edu.br/>, 2008 [Online].
- [36] M.G.F. Costa, K.M.B. Pinto, L.B.M. Fujimoto, M.M. Ogusku, C.F.F. Costa Filho, Multi-focus image fusion for bacilli images in conventional sputum smear microscopy for tuberculosis, *Biomed. Signal Process Control* 49 (2019), <https://doi.org/10.1016/j.bspc.2018.12.018>.
- [37] E. Priya, S. Srinivasan, Automated object and image level classification of TB images using support vector neural network classifier, *Biocybern. Biomed. Eng.* 36 (2016) 670–678, <https://doi.org/10.1016/j.bbe.2016.06.008>.
- [38] C.T.N. Suzuki, J.F. Gomes, A.X. Falcão, J.P. Papa, S. Hoshino-Shimizu, Automatic segmentation and classification of human intestinal parasites from microscopy

- images, *IEEE Trans. Biomed. Eng.* 60 (2013) 803–812, <https://doi.org/10.1109/TBME.2012.2187204>.
- [39] A.U. Ibrahim, E. Guler, M. Guvenir, K. Suer, S. Serte, M. Ozsoz, Automated detection of *Mycobacterium tuberculosis* using transfer learning, *J. Infect Dev. Ctries* 15 (2021) 678–686, <https://doi.org/10.3855/JIDC.13532>.
- [40] M.S. Imaz, Handbook for the bacteriologic diagnosis of tuberculosis. Part 1: smear microscopy update. Technical coordination. Program "strengthening of the network of tuberculosis laboratories in the region of the americas, in: Lima: Unanue, OPAS, WHO, Organismo Andino de Salud – Convenio Hipólito, 2018.

**Mikaela K. Serrão** received a degree in Control and Automation Engineering from the Amazonas State University in 2017, the master's degree in electrical engineering from the Federal University of Amazonas in 2019. Doctoral student of the graduate program in electrical engineering at the federal university of Amazonas.

**Marly G. F. Costa** received the electrical engineering degree from the Federal University of Amazonas, Manaus, Brazil, in 1983, and the M.Sc. and D.Sc. degrees in electrical engineering from the State University of Campinas, Campinas, Brazil, in 1988 and 1996, respectively. She has been with the Federal University of Amazonas since 1990, where she is a Full Professor with the Electrical Engineering Program. Her research interests include biomedical imaging, artificial intelligence, and engineering education.

**Luciana B. M. Fujimoto** was graduated on Medicine from the Federal University of Amazonas, Manaus, Brazil, in 1995, attended Medical Residency in Pathology

(1996–1999) from São Paulo State University, Botucatu, Brazil, had the M.Sc. degree in Tropical Pathology (2004) and the PhD degree in Biotechnology (2010), both from the Federal University of Amazonas, Manaus, Brazil. She has been with the Department of Pathology and Legal Medicine at the Federal University of Amazonas since 2002, where is currently an Associate Professor. Her research interests include histopathology, cytopathology, and diagnosis of tuberculosis and neoplasia.

**Maurício M. Ogusku** received the Pharmacy-Biochemistry degree from the University of Sagrado Coração, Bauru, Brazil, in 1989, and M.Sc. degree in Microbiology from the University of São Paulo, São Paulo, Brazil, in 1998 and D.Sc. in Biotechnology from the Federal University of Amazonas, Manaus, Brazil, in 2012. He is a researcher at National Institute of Amazonian Research – INPA, Manaus, Brazil, since 2002. He has experience in area of Microbiology, with an emphasis on Medical Mycobacteriology. His research on following topics Tuberculosis, Mycobacteriosis, bacteriological and molecular diagnostic methods.

**Cícero F. F. Costa Filho** received the electrical engineering degree from the Federal University of Pernambuco, Recife, Brazil, in 1982, and the M.Sc. and D.Sc. degrees in electrical engineering from the State University of Campinas, Campinas, Brazil, in 1985 and 1996, respectively. He is currently a Full Professor with the Electrical Engineering Program, Federal University of Amazonas, Manaus, Brazil. His research interests include biomedical imaging, artificial intelligence, and engineering education.