

Universidade Federal do Amazonas

Faculdade de Tecnologia

Programa de Pós-Graduação em Engenharia Elétrica

YADINI PÉREZ LÓPEZ

**DETECÇÃO DO *MYCOBACTERIUM TUBERCULOSIS* EM IMAGENS DE
BACILOSCOPIA DE CAMPO CLARO UTILIZANDO REDES NEURAIIS
CONVOLUTIVAS**

MANAUS

2018

Universidade Federal do Amazonas

Faculdade de Tecnologia

Programa de Pós-Graduação em Engenharia Elétrica

YADINI PÉREZ LÓPEZ

**DETECÇÃO DO *MYCOBACTERIUM TUBERCULOSIS* EM IMAGENS DE
BACILOSCOPIA DE CAMPO CLARO UTILIZANDO REDES NEURAIAS
CONVOLUTIVAS**

**Dissertação apresentada ao Curso de
Mestrado em Engenharia Elétrica, área de
concentração Controle e Automação de
Sistemas do Programa de Pós-Graduação
em Engenharia Elétrica da Universidade
Federal do Amazonas.**

Orientadora: Prof^a. Dr^a. Marly Guimarães Fernandes Costa

Co-orientador: Prof. Dr. Cícero Ferreira Fernandes Costa Filho.

MANAUS

2018

Ficha Catalográfica

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

L864d Lopez, Yadini Pérez
Detecção do Mycobacterium tuberculosis em imagens de
baciloscopia de campo claro utilizando redes neurais convolutivas. /
Yadini Pérez Lopez. 2018
119 f.: il. color; 31 cm.

Orientadora: Profª. Drª Marly Guimarães Fernandes Costa
Coorientadora: Prof. Dr. Cícero Ferreira Fernandes Costa Filho
Dissertação (Mestrado em Engenharia Elétrica) - Universidade
Federal do Amazonas.

1. detecção. 2. identificação. 3. Mycobacterium tuberculosis. 4.
redes neurais convolutivas. 5. algoritmo Non-maximum
Suppression. I. Costa, Profª. Drª Marly Guimarães Fernandes II.
Universidade Federal do Amazonas III. Título

YADINI PÉREZ LÓPEZ

**DETECÇÃO DO MYCOBACTERIUM TUBERCULOSIS EM IMAGENS DE
BACILOSCOPIA DE CAMPO CLARO UTILIZANDO REDES NEURAIAS
CONVOLUTIVAS**

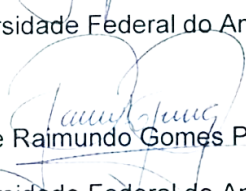
Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal do Amazonas, como requisito parcial para obtenção do título de Mestre em Engenharia Elétrica na área de concentração Controle e Automação de Sistemas.

Aprovado em 13 de abril de 2018.

BANCA EXAMINADORA



Profª. Dra. Marly Guimarães Fernandes Costa, Presidente
Universidade Federal do Amazonas



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



Profª. Dra. Luciana Botinelly Mendonça Fujimoto, Membro
Universidade Federal do Amazonas

Aos meus pais Maritza e Nicasio

Ao meu noivo Luis

AGRADECIMENTOS

Aos meus orientadores Prof^a. Dr^a Marly Guimarães Fernandes Costa e Prof. Dr. Cícero Ferreira Fernandes Costa Filho, pela confiança que depositaram em mim e por contribuir ao meu crescimento profissional;

À CAPES e à FAPEAM (processo 062.00575/2014-PROTI-PESQUISA e processo: 062.00938/2015 - PECTI/AM-PG) pelo apoio financeiro;

Aos especialistas Dr. Maurício Ogusku e Dra. Luciana Fujimoto pela grande contribuição neste trabalho;

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

Aos meus colegas Péricles, Yuri, Annie, João Pablo, Manoel, Kelly, Zulema, Patricia e Paula que sempre estiveram me incentivando e fizeram a minha estadia no Brasil mais familiar;

Aos professores cubanos Prof. Dr Luis Delfin Rojas, Prof. Dr Daniel Guzmán, Prof. Dr Reiniel Beltrán e Prof^a.Dr^a Annia Lusson pelo apoio e conselhos de vida;

À minha família, em especial aos meus queridos pais pelo amor, dedicação e apoio incondicional que me deram sempre;

À minha mãe por ser sempre o maior exemplo de força e coragem para mim;

Ao meu noivo Luis Miguel Rojas Aguilera por fazer sempre a vida ficar mais leve.

RESUMO

A Tuberculose (TB) é uma doença causada por uma bactéria de crescimento lento, *Mycobacterium tuberculosis* (MT) e desde o ano 2000 tem sido incluída entre as dez causas principais de morte a nível mundial. No ano 2015 o Brasil ocupou a 18ª posição em incidência de tuberculose, representando 9% dos casos estimados no mundo, e 33% dos estimados para as Américas. O exame de baciloscopia de campo claro é o mais utilizado em países em desenvolvimento para o diagnóstico e acompanhamento da doença. Desde o ano de 2008 têm sido desenvolvidos trabalhos pela comunidade científica focados na detecção do bacilo da TB visando a automatização da baciloscopia de campo claro. Esses estudos utilizaram bases de dados com diferentes quantidades de imagens, exploraram aspectos de cor dos bacilos e aplicaram técnicas de Processamento Digital de Imagens e/ou Aprendizado de Máquina, e mais recentemente Aprendizado de Máquina Profundo. Entretanto, verificou-se que técnicas de Aprendizado de Máquina Profundo não têm sido exploradas utilizando bases de dados de imagens de baciloscopia robustas que refletem as condições reais dos exames de baciloscopia de campo claro. O presente trabalho visa o desenvolvimento de um método de detecção automática do bacilo da TB mediante a utilização de Redes Neurais Convolutivas (RNC), utilizando conjuntos de imagens baciloscópicas de 2 pacientes nos formatos de cor RGB, R-G e Escala de Cinza. Para atingir o objetivo proposto, foi gerada uma base de dados de *patches* contendo bacilos (positivos) e *patches* sem bacilos (negativos), os quais foram utilizados no treinamento de três arquiteturas diferentes de RNC. Posteriormente, foi aplicado o algoritmo *Non-maximum Suppression* utilizando imagens completas de baciloscopia para detectar cada bacilo. Os melhores resultados na etapa de classificação de *patches* foram obtidos utilizando as imagens de baciloscopia R-G e RGB e dois modelos RNC de duas e três camadas convolutivas, alcançando uma acurácia de 99%. Na etapa final de detecção dos bacilos nas imagens completas de baciloscopia os melhores resultados foram alcançados utilizando as imagens no formato RGB atingindo um *Precision* de 56,82%, um *Recall* de 86,15% e um *F1-score* de 68,47%.

Palavras chaves: Detecção, identificação, *Mycobacterium tuberculosis*, Redes Neurais Convolutivas, algoritmo *Non-maximum Suppression*.

ABSTRACT

Tuberculosis (TB) is a disease caused by a slow-growing bacterium named *Mycobacterium tuberculosis* (MT). Since 2000 has been included among the top 10 leading causes of death worldwide. In 2015, Brazil ranked eighteenth in TB incidence, representing 9% and 33% of the estimated cases worldwide and for the Americas respectively. Light field smear microscopy is the most commonly used exam in developing countries for diagnosis and follow-up of the disease. Since 2008, several researches have been developed focused on TB bacillus detection, aiming the automation of light field smear microscopy. These studies used datasets with different amounts of images, explored different color aspects of bacilli, and applied Digital Image Processing and / or Machine Learning techniques, and more recently, Deep Learning using Grayscale images. However, Deep Learning techniques have not been explored using a robust smear microscopy image dataset that reflect real conditions of smear microscopy exams. This work presents a method for automatic detection of TB bacillus using Convolutional Neural Networks (CNN) using a dataset of images taken from 2 patients in RGB, R-G and Grayscale color formats. To reach the proposed goal, a patch dataset containing bacilli (positive patches) and without bacilli (negative patches) was generated. This patch dataset was used for training three different RNC architectures. Then, Non-maximum Suppression (NMS) algorithm was applied using CNN models and complete smear images for bacillus detection. Best results in the patch classification stage were obtained using R-G and RGB images and two CNN models, achieving an accuracy of 99% in both cases. At final stage, Bacilli detection in full smear images, best results were achieved using RGB images reaching a Precision of 56,82%, Recall of 86,15% and F1-score of 68,47%.

Keywords: Detection, identification, *Mycobacterium tuberculosis*, Convolutional Neural Networks, Non-maximum Suppression algorithm.

LISTA DE FIGURAS

Figura 1 Esquema do cubo de cores RGB. Os pontos ao longo da diagonal principal representam os valores de cinza, do preto na origem ao branco no ponto (1, 1, 1).....	11
Figura 2 Fragmento de imagem de baciloscopia de campo claro representado em: (a) no espaço de cores RGB, (b) Escala de Cinza e (c) imagem de diferença R-G.	12
Figura 3 RNA com quatro neurônios na camada de Entrada (X_1 , X_2 , X_3 e X_4), três neurônios na Camada Oculta (1, 2 e 3) e dois neurônios na camada de Saída (y_1 e y_2). As conexões de cada neurônio foram representadas com cores diferentes. Observe que cada neurônio se encontra conectado com todos os neurônios da camada anterior.	16
Figura 4 Arquitetura do <i>Perceptron</i> básico.	17
Figura 5 Hiperplano de decisão. Em um lado da linha $gx > 0$ e no outro lado $gx < 0$	19
Figura 6 Representação espacial das classes do referido problema XOR.	20
Figura 7 Hiperplano de decisão traçado utilizando <i>Perceptron</i> Multicamada para solucionar o problema XOR.....	21
Figura 8 Exemplo de arquitetura do <i>Perceptron</i> multicamada capaz de resolver o problema XOR. .	22
Figura 9 Fluxo de processamento do algoritmo <i>Backpropagation</i>	23
Figura 10 Exemplo de convolução de duas dimensões em que a saída é restringida apenas às posições onde o <i>Kernel</i> se encaixa completamente na entrada. Observe que as dimensões da entrada são de 3x4 (3 linhas e 4 colunas), o <i>Kernel</i> tem dimensões de 2x2 (2 linhas e 2 colunas). Assim a saída tem dimensões 2x3 (2 linhas e 3 colunas).....	27
Figura 11 Em (a) apresenta-se uma RN tradicional com uma camada de entrada duas camadas ocultas (Camada escondida 1 e Camada escondida 2) e uma camada de saída observe que os neurônios das Camadas escondidas 1 e 2 estão conectados a todas as unidades da camada anterior. Em (b) mostra-se uma CC (Camada Convolutiva) com (Entrada) de dimensões (32x32x3). O cubo maior de cor azul, representa a CC, o qual contém um cubo menor que representa um filtro. Observe que o primeiro neurônio do filtro, representado por um círculo de azul mais intenso, está conectado somente com uma pequena área no volume de entrada.	28
Figura 12 Exemplo de CC de profundidade, 3D, em que as matrizes da esquerda (volume de entrada), representadas em azul, são cada uma das dimensões da matriz $x : x[:, :, 0]$, $x[:, :, 1]$ e $x[:, :, 2]$. Observe que foi aplicado um <i>padding de 1</i> , $P=1$, nas bordas das matrizes. Esse <i>padding</i> é representado com valores 0 em nível de cinza. O volume de entrada resultante tem dimensões (7x7x3). As matrizes localizadas no centro da figura, w_0 e w_1 , correspondem aos pesos dos filtros w_0 e w_1 , respectivamente, aplicado com passo 1, $S=1$, sobre as entradas. As matrizes mostradas à direita da figura, em verde, representam o volume de saída o , de dimensões (3x3x2). Observe, que segundo as equações 12 e 13, a dimensão espacial da saída é (3x3). $w_2 = (5-3+2*1) / 2+1=3$ e $h_2 = ((5-3+2*1) / 2+1=3$, com profundidade $d_2 = 2$, segundo equação 14. A operação de convolução entre as matrizes $x[:, :, 0]$, $x[:, :, 1]$, $x[:, :, 2]$ e $w_1[:, :, 0]$, $w_1[:, :, 1]$ e $w_1[:, :, 2]$, têm como resultado o valor 3, contido em um quadro verde na saída $o[:, :, 1]$	31

Figura 13 Efeito de redução espacial do volume de entrada mediante uma CP. Em (a) mostra-se uma imagem de dimensões (224x224), a qual é reduzida mediante a CP para as dimensões (112x112). A forma em que ocorre essa operação é mostrada em (b), em que a matriz da esquerda representa uma porção de profundidade do volume de entrada de dimensões 4x4; sobre essa porção é aplicada	33
Figura 14 Em (a) mostra-se uma imagem de baciloscopia de campo claro com dois bacilos identificados por especialistas (objetos dentro dos quadros vermelhos). Em (b) apresenta-se o mapa de características obtido mediante a aplicação de janelas corredeiras e um classificador treinado para identificar bacilos. Observe a existência de falsos positivos em áreas incorretas, A_i (áreas não marcadas pelo especialista), assim como as detecções em áreas corretas, A_c (áreas marcadas pelo especialista e encerradas em quadrados vermelhos). Em (c) mostra-se o resultado da aplicação de um filtro gaussiano no mapa de características. Observe como as áreas A_i foram suavizadas, quase apagadas, e as áreas A_c mantiveram a forma de bacilos.	38
Figura 15 Em (a) é ilustrada a aplicação da abordagem de janelas corredeiras. Observe que foram obtidas múltiplas detecções para um mesmo bacilo (aglomerados de pontos vermelhos). Em (b) é ilustrada a aplicação do algoritmo NMS. Observam-se exatamente duas detecções (pontos vermelhos).	39
Figura 16 Imagens de baciloscopia de campo claro, (a) imagem utilizada em Quinn <i>et al</i> (2016) e (b) imagem utilizada no presente trabalho.	47
Figura 17 Diagrama de blocos que apresenta a metodologia implementada.	56
Figura 18 Fragmentos de imagens de foco estendido de baciloscopia de campo claro. Em (a) bacilos foram marcados com círculos vermelhos. Em (b) aglomerados de bacilos foram marcados com quadrados amarelos. Em (c) bacilos duvidosos foram marcados com triângulos pretos.	57
Figura 19 Em (a) apresenta-se exemplos de <i>patches</i> positivos (PP) incluídos no CO. Observe que o bacilo encontra-se centrado em cada <i>patch</i> em diferentes posições e orientações. Em (b) são mostrados exemplos de <i>patches</i> negativos (PN). Observe que o segundo <i>patch</i> , da esquerda para a direita, é um exemplo de <i>patch</i> que contém um fragmento de bacilo localizado na esquina superior esquerda.	58
Figura 20 Procedimento seguido na criação do Conjunto Original (CO) de <i>patches</i> partindo dos 11 planos focais de cada imagem de baciloscopia de campo claro.	59
Figura 21 Ilustração da arquitetura dos Modelos 1, 2 e 3.	60
Figura 22 Configuração do treinamento dos três modelos baseados em Redes Neurais Convolutivas (RNC).	61
Figura 23 Normalização dos <i>patches</i> . Em (a) mostra-se um <i>patch</i> no formato RGB e em (b) apresenta-se o <i>patch</i> mostrado em (a) normalizado. Em (c) mostra-se um <i>patch</i> no formato R-G e em (d) é apresentado o <i>patch</i> mostrado em (c) normalizado. Observe que nos <i>patches</i> normalizados o bacilo é mais visível, pois existe maior contraste entre o objeto de interesse e o fundo.	68
Figura 24 Em (a) mostra-se filtros e ativações da primeira camada convolutiva CONV1 do modelo 2 utilizando como entrada um <i>patch</i> positivo no formato R-G, com <i>passo</i> $s=1$ e em (b) utilizando <i>passo</i> $s=2$.	69

Figura 25 Filtros e ativações da primeira camada convolutiva CONV1 do modelo 2 utilizando como entrada um <i>patch</i> negativo no formato R-G com <i>passo</i> $s=2$. Observe que as ativações não apresentam manchas com forma de bacilo, diferentemente das ativações mostradas na Figura 24, onde foi utilizado um <i>patch</i> positivo como entrada do modelo.	70
Figura 26 Curvas ROC dos modelos 1, 2 e 3 para o CA nos formatos: (a) Escala de Cinzas, (b) R-G e (c) RGB.	72
Figura 27 Em (a), (d) e (g) é mostrada uma imagem de baciloscopia de campo claro nos formatos: RGB, R-G e escala de cinza respectivamente. Em (b), (e) e (h) são apresentados os mapas de características das imagens mostradas em (a), (d) e (g), utilizando <i>passo</i> $s = 5$. Finalmente em (c), (f) e (i) é possível observar o efeito da aplicação do filtro gaussiano nos mapas de características mostrados em (b), (e) e (h).	75
Figura 28 Imagem de baciloscopia de campo claro após a aplicação da metodologia proposta utilizando diferentes valores de limiar de sobreposição no algoritmo <i>Non-maximum Suppression</i> . Em (a), utilizando-se $Is = 0,2$, observa-se que as detecções coincidem com as marcações feitas pelos especialistas mostradas em (b). Em (c) apresenta-se as detecções obtidas utilizando-se $Is = 0,5$. Em (d) apresenta-se as detecções obtidas utilizando-se $Is = 0,9$. Cada ponto vermelho representa a detecção de um bacilo.	77
Figura 29 Exemplo de marcação FP (retângulo preto), onde o algoritmo identificou como sendo um bacilo um artefato com forma e cor de bacilo, mas que não foi marcado pelo especialista.	81
Figura 30 Exemplo de marcações FP, bacilo muito comprido (retângulo azul). Uma detecção correta (VP) é apresentada no retângulo preto.	81
Figura 31 Exemplo de imagem de baciloscopia com ruído avermelhado em que o algoritmo apresenta vários FP representados como pontos verdes. Uma detecção VP foi marcada com um ponto vermelho.	81
Figura 32 Um exemplo de ocorrência de bacilos gêmeos (retângulo preto). Observe na ampliação que o algoritmo identificou somente um bacilo (ponto vermelho), apresentando um FN (cruz azul).	83
Figura 33 Um exemplo de bacilo não detectado pelo algoritmo (FN) na borda da imagem de baciloscopia (retângulo preto). A cruz azul indica que nessa posição o especialista marcou como existindo um bacilo, mas o algoritmo não conseguiu identificá-lo.	83
Figura 34 Exemplo de FN do algoritmo em um bacilo de baixa intensidade (retângulo preto). Observe na ampliação que o bacilo é quase imperceptível, quando comparando com o bacilo corretamente marcado pelo algoritmo (VP), retângulo azul.	84
Figura 35 Em (a) apresenta-se uma imagem de baciloscopia de campo claro marcada por especialistas, em que os bacilos foram marcados com círculos vermelhos e bacilos duvidosos foram marcados com triângulos pretos. Em (b) apresenta-se a mesma imagem de baciloscopia com as marcações feitas pelo método proposto, em que os bacilos identificados foram marcados com pontos vermelhos. Observa-se em (b) que na posição onde os especialistas marcaram um bacilo duvidoso (marcada por um círculo preto) o algoritmo não detectou o objeto de interesse, o que pode significar que efetivamente não existe bacilo nessa posição.	85

LISTA DE TABELAS

Tabela 1 Critérios para Leitura e Interpretação dos Resultados da Baciloscopia de escarro. Fonte: Adaptado de (Campos Levi, 2012).....	9
Tabela 2 Tabela de verdade para a função XOR.	20
Tabela 3 Análise dos artigos relativos ao tema de detecção da tuberculose em imagens de baciloscopia de campo claro usando técnicas de Aprendizado de Máquina Tradicional e Profundo.	49
Tabela 4 Especificações de entrada-saída da função implementada para a obtenção do mapa de probabilidades.	64
Tabela 5 Especificações de entrada-saída da função implementada para a obtenção das BB.	64
Tabela 6 Especificações de entrada-saída do algoritmo NMS.	65
Tabela 7 Erro do modelo 2 com entradas no formato R-G com e sem aplicar as técnicas de regularização <i>dropout</i> e incremento da base de dados.	67
Tabela 8 Configurações dos parâmetros de detecção e resultados obtidos nas 6 simulações para cada combinação formato de entrada-classificador.	79
Tabela 9 Resultados obtidos utilizando as métricas <i>Precision</i> , <i>Recall</i> e <i>F1-score</i> na aplicação do método proposto com e sem descartar as imagens com ruído vermelho mediante a aplicação de um limiar de pixels vermelhos.	86

LISTA DE SIGLAS

TB	Tuberculose
MT	<i>Mycobacterium tuberculosis</i>
OMS	Organização Mundial da Saúde
MS	Ministério da Saúde
RNC	Redes Neurais Convolutivas
NMS	Supressão Não Máxima (do inglês <i>Non-maximum Suppression</i>)
AM	Aprendizado de Máquinas
AMP	Aprendizado de Máquinas Profundo
RN	Redes Neurais
RNA	Redes Neurais Artificiais
RNP	Redes Neurais Profundas
GD	Gradiente Descendente
EMQ	Erro Médio Quadrático
CC	Camadas Convolutivas
CP	Camada <i>Pool</i>
CCT	Camada Totalmente Conectada
CO	Conjunto de <i>patches</i> Original
CA	Conjunto de <i>patches</i> Aumentado
MP	Mapa de probabilidades
LP	Limiar de Probabilidade
LS	Limiar de Sobreposição
FP	Falsos Positivos
FN	Falsos Negativos
VP	Verdadeiros Positivos
LPV	Limiar de Pixels Vermelhos
AUC	Área sob a curva (do inglês <i>Area Under the Curve</i>)

SUMÁRIO

1. Introdução.....	1
1.1 Objetivo Geral	4
1.2 Objetivos Específicos.....	5
1.3 Estrutura da dissertação.....	5
2. Fundamentos teóricos	7
2.1 Introdução.....	7
2.2 Baciloscopia de campo claro	7
2.2.1 Método de coloração utilizado no exame de baciloscopia de campo claro	8
2.2.2 Interpretação dos resultados do exame de baciloscopia.....	9
2.3 Representação digital de imagens e espaços de cores.....	10
2.3.1 Espaço de cores RGB	10
2.3.2 Escala de cinza	11
2.3.3 Imagem de diferença R-G.....	11
2.4 Aprendizado de Máquina	12
2.4.1 Tarefas.....	13
2.4.2 Medida de desempenho.....	14
2.4.3 Experiência.....	15
2.4.4 Redes neurais artificiais	15
2.4.5 Aprendizado de máquina profundo (AMP)	24
2.4.6 Redes Neurais Convolutivas (RNC).....	25
2.5 Detecção de objetos	36
2.5.1 <i>Non-maximum Suppression</i> (NMS).....	37
2.5.2 Avaliação das detecções	37
3. Revisão bibliográfica.....	40
3.1 Introdução.....	40
3.2 Estado da arte das pesquisas sobre identificação automática do bacilo da TB.....	40
3.2.1 Detecção do <i>Mycobacterium tuberculosis</i> por segmentação, mediante técnicas de processamento digital de imagens	41

3.2.2 Detecção do <i>Mycobacterium tuberculosis</i> mediante a aplicação de redes neurais artificiais	41
3.2.3 Detecção do <i>Mycobacterium tuberculosis</i> mediante a aplicação de aprendizado profundo	44
3.2.4 Considerações.....	46
4.1 Introdução.....	56
4.2 Materiais	56
4.3 Métodos	57
4.3.1 Criação de uma base de dados de <i>patches</i>	57
4.3.2 Classificação de <i>patches</i>	59
4.3.3 Detecção dos bacilos nas imagens de baciloscopia	63
4.3.3.1 Predição	63
5. Resultados	67
5.1 Resultados da etapa de classificação.....	67
5.1.1 Impacto da regularização.....	67
5.1.2 Seleção dos hiperparâmetros das camadas convolutivas.....	68
5.1.3 Seleção do modelo	70
5.1.4 Influência do tipo de imagem nos resultados.....	73
5.2 Resultados da etapa de detecção dos bacilos	73
5.2.1 Mapas de probabilidade.....	73
5.2.2 Seleção do limiar de probabilidade	74
5.2.3 <i>Non-maximum Suppression</i>	76
5.2.4 Avaliação de desempenho dos modelos para os diferentes tipos de entradas	77
5.2.5 Análise dos resultados obtidos na etapa de detecção	79
5.2.6 Descarte automático de imagens com ruído vermelho	85
6. Conclusões e trabalhos futuros	87
Referências.....	90
Apêndice	94

1. INTRODUÇÃO

A Tuberculose (TB) é uma doença causada por uma bactéria de crescimento lento, *Mycobacterium tuberculosis* (MT). Essa patologia, na maior parte dos casos, acomete os pulmões, porém pode afetar qualquer parte do corpo. É uma condição curável e possível de prevenir (MS, 2014).

Desde o ano 2000 a TB tem sido incluída entre as dez causas principais de morte a nível mundial. Quinze anos depois, as estatísticas referentes a doença continuam sendo alarmantes (WHO, 2015):

- 10,4 milhões de pessoas foram infectadas com TB, das quais, 1,8 milhões morreram.
- Mais de 95% das mortes causadas pela TB ocorreram em países de média e baixa renda.
- A TB é a principal causa de morte de pessoas infectadas com o Vírus da Imunodeficiência Humana (HIV), representando 35% das mortes.
- Aproximadamente 49 milhões de vidas foram salvas através do diagnóstico e tratamento da TB entre 2000 e 2015 (WHO, 2015).

A Organização Mundial da Saúde (OMS) definiu em 2017, como um dos objetivos para o ano 2035, encerrar a epidemia global da TB. Para alcançar esta meta, foi proposto um conjunto de estratégias entre as quais destacam-se :

- Reduzir o coeficiente de incidência para menos de 10 casos para cada 100 mil habitantes.
- Reduzir o número de óbitos por TB em 95%, e o número de novos casos em 80%.

As estratégias propostas são baseadas em três pilares principais: cuidados integrados centrados no paciente e sistemas de prevenção, políticas fortes de apoio e a intensificação da investigação e inovação (Pais Ramos, 2016).

As dificuldades para superar a TB no Brasil são imensas. A presença da enfermidade está associada com a exclusão social e a marginalização de parte da população, a qual é submetida a más condições de vida, como moradia precária, desnutrição e dificuldade de acesso aos serviços e bens públicos (MS, 2014).

O Brasil ocupa a 18ª posição em incidência de tuberculose, representando 0,9% dos casos estimados no mundo, e 33% dos estimados para as Américas. Os coeficientes de mortalidade e de incidência foram reduzidos em 38,9% (3,6 para 2,2/100 mil hab.) e 34,1% (51,8 para 34,1/100 mil hab.), respectivamente, de 1990 até 2014. Com esses resultados, o país cumpriu com as metas internacionais. Apesar disso, ainda foram registrados, entre 2005 e 2014, uma média de 70 mil casos novos e 4.400 mortes por tuberculose, por ano, e entre 2012 e 2015, 840 casos novos de tuberculose droga-resistente, que são os casos que apresentam qualquer tipo de resistência aos fármacos utilizados no tratamento da doença (MS, 2016).

O estado de Amazonas apresentou os coeficientes mais altos de novos casos de tuberculose e alto índice de mortalidade em 2015, 98,3/100 mil hab. e 3,3/100 mil hab., respectivamente (MS, 2016).

A baciloscopia da TB é o tipo de exame diagnóstico mais utilizado, não somente na busca de casos infecciosos, mas também como medidor da eficácia do tratamento (WHO, 2015).

A baciloscopia consiste na pesquisa de Bacilos Álcool-Ácido Resistentes (BAAR) em um esfregaço de amostra clínica, preparado e corado segundo metodologia padronizada. Para a realização deste exame, o Ministério da Saúde (MS) recomenda a utilização do método de coloração de *Ziehl-Neelsen* ou de *Kinyoun*, pois utilizam a microscopia de campo claro, o qual é um método de rotina utilizado em países subdesenvolvidos e com alta incidência da doença. Tal fato decorre da microscopia de fluorescência exigir equipamentos de alto custo, enquanto que os microscópios óticos de luz convencional, ou luz branca, possuem baixo custo (Khutlang *et al* 2010).

Nos últimos anos, vários sistemas de diagnóstico assistido por computador (CAD) têm sido desenvolvidos. Nesse sentido, vários trabalhos presentes na literatura visam a detecção automática do MT em amostras de secreção pulmonar de pacientes, com o objetivo de desenvolver a técnica da baciloscopia automatizada, pois, foi observado que o número elevado de exames realizados pelos profissionais induz a fadiga visual, e em consequência, pode ocasionar erros de diagnósticos.

Adicionalmente, motivados pelo fato do Brasil ser um dos países reincidentes a cada ano em novos casos de TB, e o estado do Amazonas apresentar o maior índice de pessoas infectadas, e uma alta taxa de mortalidade ocasionada por essa doença. Têm sido desenvolvidos no âmbito do Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal do Amazonas (PPGEE-UFAM) trabalhos com o objetivo de desenvolver métodos para auxiliar o diagnóstico da TB, a partir de imagens de baciloscopia obtidas por meio da microscopia de campo claro.

Os resultados mais significativos até este ano de 2018, alcançados pelo mencionado grupo, foram obtidos por meio do reconhecimento e segmentação do

Mycobacterium tuberculosis em imagens de microscopia de campo claro utilizando as características de cor e redes neurais com o algoritmo *Backpropagation*. O trabalho consistiu em extrair amostras de pixels de bacilos e de fundo em imagens adquiridas no Laboratório de Micobacteriologia do Instituto Nacional de Pesquisas da Amazônia (INPA), que foram utilizadas para o treinamento de um classificador baseado na utilização de técnicas de Aprendizado de Máquina. As imagens foram automaticamente separadas em dois grupos, de acordo com o conteúdo de fundo. O método desenvolvido seleciona um conjunto ótimo de características de cor do bacilo e do fundo da imagem, empregando o método de seleção escalar de características. Essas características foram utilizadas em um classificador de pixels, um *Perceptron* multicamada, treinado pelo algoritmo *Backpropagation* (Campos Levi, 2012).

Com o objetivo de aumentar a contribuição com a comunidade científica, especialmente na área das tecnologias aplicadas à medicina, o trabalho proposto nesta dissertação tem por escopo o desenvolvimento de um novo método de detecção de bacilos da TB em imagens obtidas por baciloscopia de campo claro, utilizando a técnica de Aprendizado de Máquina Profundo: Redes Neurais Convolutivas, assim como o algoritmo *Non-maximum Suppression*.

1.1 Objetivo Geral

Contribuir para o aprimoramento da baciloscopia automatizada através do desenvolvimento de um método para a detecção do MT em imagens de baciloscopia de campo claro, baseado na utilização de técnicas de Aprendizagem de Máquina Profundo.

1.2 Objetivos Específicos

1. Avaliar se o emprego de redes neurais convolutivas (técnica de aprendizado profundo) associadas ao algoritmo *Non-maximum Suppression* pode melhorar a acurácia dos métodos de detecção de bacilos presentes na literatura.
2. Analisar o desempenho de diversas arquiteturas de Redes Neurais Convolutivas na classificação de *patches* positivos (com bacilos) ou negativos (sem bacilos).
3. Analisar o desempenho do algoritmo *Non-maximum Suppression* na detecção do bacilo da TB em imagens completas de baciloscopia, por meio da classificação de regiões utilizando um classificador de *patches* selecionado.
4. Avaliar o desempenho dos métodos referidos em três versões de imagens baciloscópicas: RGB, R-G e Escala de Cinza.

1.3 Estrutura da dissertação

Essa dissertação estrutura-se da seguinte forma:

- Capítulo 1: Introdução.
- Capítulo 2: Fundamentos Teóricos.
- Capítulo 3: Revisão bibliográfica.
- Capítulo 4: Materiais e métodos.
- Capítulo 5: Resultados.
- Capítulo 6: Conclusões.
- Referências
- Apêndice

Um *background* sobre a doença da tuberculose e sobre a importância do controle da mesma foi apresentado no Capítulo 1.

No Capítulo 2 os principais fundamentos teóricos sobre as técnicas utilizadas no trabalho são abordados: baciloscopia de campo claro; os três formatos de imagens RGB, R-G e Escala de Cinza; conceitos como: Aprendizado de Máquina Tradicional e Profundo, Redes Neurais Convolutivas e métricas de desempenho são explicados. Por último o algoritmo *Non-maximum Suppression* é abordado.

O Capítulo 3 apresenta o estudo realizado do estado da arte na detecção do MT em imagens de baciloscopia de campo claro.

No Capítulo 4 são apresentados os materiais utilizados: a base de dados de imagens de baciloscopia de campo claro. Além disso, é apresentada a metodologia proposta para a detecção do bacilo da TB. Começando pela geração de uma base de dados de *patches* positivos e negativos, são abordadas as diferentes arquiteturas de Redes Neurais Convolutivas implementadas, assim como o algoritmo *Non-maximum Suppression*. Por último, são apresentadas as métricas de desempenho utilizadas para avaliar a metodologia proposta.

No Capítulo 5 os resultados alcançados na aplicação da metodologia proposta são apresentados.

No Capítulo 6, são apresentadas as conclusões relativas à aplicação da metodologia proposta e aos resultados obtidos, assim como os trabalhos futuros.

Finalmente, no Apêndice é apresentado um artigo publicado, resultado parcial dessa dissertação.

2. FUNDAMENTOS TEÓRICOS

2.1 Introdução

Neste capítulo os principais conceitos aplicados no desenvolvimento deste trabalho serão apresentados. A Seção 2.2, abordará características relevantes acerca do processo de obtenção das imagens de baciloscopia de campo claro. Em 2.3 serão explicados os espaços de cores investigados, isto é, *RGB (Red, Green, Blue)*, *R-G* (Operação de diferença entre as componentes R e G, do espaço de cor RGB) e Escala de Cinza. A Seção 2.4 começará abordando a temática de Aprendizado de Máquina, introduzindo conceitos básicos de Redes Neurais Artificiais e do algoritmo de *Backpropagation*, para então ser abordado o tema de Aprendizado de Máquina Profundo e Redes Neurais Convolutivas. Por fim, na Seção 2.5, será abordado o conceito de detecção de objetos em imagens e o algoritmo *Non-maximum Suppression*.

2.2 Baciloscopia de campo claro

A baciloscopia de campo claro é o método principal no diagnóstico e controle do tratamento da TB pulmonar. Trata-se de um método simples, rápido e de baixo custo. O exame é solicitado aos pacientes que apresentam os sintomas da doença, com: tosse durante 3 semanas, inicialmente seca e depois com catarro, pus ou sangue; dor no peito junto ao tórax e dificuldade em respirar; produção de escarro esverdeado ou amarelado (WHO, 2017).

O primeiro passo para o diagnóstico consiste na coleta de uma amostra de escarro seguindo normas estabelecidas pelo MS. Posteriormente, a referida amostra é

processada, mediante técnicas padronizadas de coloração de lâminas, em busca de uma melhor visualização dos bacilos da TB no microscópio de campo claro. O método de coloração de *Ziehl-Neelsen* (ZN) é o método indicado pelo MS para a preparação das lâminas utilizadas no exame de baciloscopia de campo claro. As imagens baciloscópicas utilizadas neste trabalho foram obtidas mediante uma variação do referido método conhecida como método de *Kinyoun* (método de ZN a frio) (MS, 2014).

2.2.1 Método de coloração utilizado no exame de baciloscopia de campo claro

Através do método de coloração de ZN, é possível visualizar as micobactérias que possuem na sua parede celular uma grande concentração de lipídios, devido à presença de ceras e ácidos graxos, o que dificulta a sua visualização através das colorações tradicionais. O método de ZN basicamente propõe a coloração das lâminas contendo o esfregaço do paciente utilizando o corante fucsina fenicada, para a descoloração das lâminas utiliza-se o álcool-ácido, e finalmente o cobrimento das lâminas é feito com azul de metileno. A fucsina fenicada, quando aquecida, vai corar de vermelho todas as células bacterianas e outras estruturas presentes no esfregaço. O ácido diluído em álcool aplicado vai descorar todas as bactérias exceto as resistentes ao álcool-ácido (bacilo da TB), que permanecem coradas de vermelho pela fucsina. Assim, ao serem observadas após a coloração e o contraste com azul de metileno, serão visualizadas as Bactérias Resistentes ao Ácido-Álcool (BAAR), coradas de vermelho e não coradas de azul (MS, 2014).

“A diferença básica entre o *Ziehl-Neelsen* original e o *Kinyoun* é que o método de *Kinyoun* é de coloração a frio, enquanto que o método ZN é um método de coloração a

quente” (Campos Levi, 2012). Outra diferença está presente na concentração de fucsina utilizada, a qual é de 3% no método de Kinyoun, e de 0,3% no método de ZN.

2.2.2 Interpretação dos resultados do exame de baciloscopia

Na leitura e interpretação dos resultados da baciloscopia realizada com escarro espontâneo distendido e corado pelo método de ZN, ou suas versões, são lidos no mínimo 100 (cem) campos úteis do microscópio, ou seja, aqueles campos nos quais se observam elementos celulares de origem pulmonar (leucócitos, fibras mucosas e células ciliares). Os campos em que não aparecem esses elementos não devem ser contabilizados na leitura. A leitura e interpretação dos resultados da baciloscopia realizada com lâmina corada pelo método de ZN é equivalente àquelas realizadas pelo método de *Kinyoun*. Dessa forma, os critérios para leitura e interpretação dos resultados da baciloscopia servem para qualquer um dos dois métodos, conforme a padronização do MS no ano 2001 (MS, 2001) e são apresentados na Tabela 1.

Tabela 1 Critérios para Leitura e Interpretação dos Resultados da Baciloscopia de escarro.

Fonte: Adaptado de (Campos Levi, 2012).

Baciloscopia de escarro	Resultado
Não são encontrados BAAR em 100 campos	NEGATIVO
São encontrados de 1 a 9 BAAR em 100 campos	Relata-se apenas a quantidade de BAAR encontrada
São encontrados de 10 a 99 BAAR, em 100 campos	POSITIVO +
São encontrados em média de 1 a 10 BAAR por campo, nos primeiros 50 campos observados	POSITIVO ++
São encontrados em média mais de 10 BAAR por campo, nos primeiros 20 campos observados	POSITIVO +++

2.3 Representação digital de imagens e espaços de cores

A representação digital de imagens é o processo de gerar uma imagem digital utilizando como guia uma imagem óptica (Castleman e Young, 2008).

Uma imagem digital pode ser visualizada e analisada em diversos espaços de cores. O objetivo de um modelo de cores (também chamado de espaço de cores ou sistema de cores) é facilitar a especificação das cores em alguma forma padronizada e amplamente aceita. Essencialmente, um modelo de cores é uma especificação de um sistema de coordenadas e um subespaço dentro desse sistema no qual cada cor é representada por um único ponto (Gonzalez e Woods, 2008).

O modelo de cores mais utilizado é o RGB, com grande aplicação na representação em monitores coloridos e câmeras de vídeo (Gonzalez e Woods, 2008).

2.3.1 Espaço de cores RGB

No modelo RGB, cada cor aparece em suas componentes espectrais primárias de vermelho, verde e azul. Esse modelo se baseia em um sistema de coordenadas cartesianas. O subespaço de cores de interesse é o cubo, apresentado na Figura 1, no qual, os valores RGB primários estão em três vértices; as cores secundárias: ciano, magenta e amarelo estão nos outros três vértices; o preto está na origem; e o branco está no vértice mais distante da origem (Gonzalez e Woods, 2008).

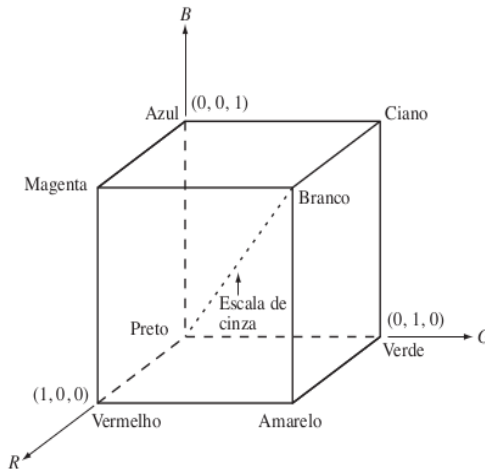


Figura 1 Esquema do cubo de cores RGB. Os pontos ao longo da diagonal principal representam os valores de cinza, do preto na origem ao branco no ponto (1, 1, 1).

Fonte: (Gonzalez e Woods, 2008)

2.3.2 Escala de cinza

Em uma imagem representada na Escala de Cinza o valor de cada pixel é uma amostra única de determinado espaço de cores e pode variar entre o preto (menor valor de intensidade) e o branco (maior valor de intensidade), podendo conter diversos tons intermediários de cinza.

2.3.3 Imagem de diferença R-G

A imagem de diferença R-G é facilmente gerada a partir do espaço RGB, realizando a subtração da componente vermelha (R) com a componente verde (G). Na Figura 2 apresenta-se três fragmentos correspondentes a uma mesma imagem de baciloscopia de campo claro representados no espaço de cores RGB, Escala de Cinza e R-G, respectivamente.

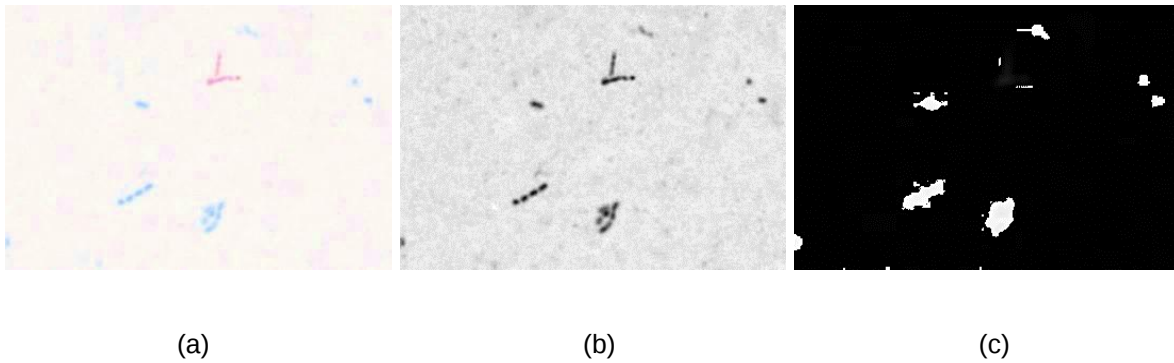


Figura 2 Fragmento de imagem de baciloscopia de campo claro representado em: (a) no espaço de cores RGB, (b) Escala de Cinza e (c) imagem de diferença R-G.

2.4 Aprendizado de Máquina

Na implementação da presente proposta foram utilizadas técnicas de Aprendizado de Máquina Profundo (AMP) que é uma área do Aprendizado de Máquina (AM). Portanto, para abordar conceitos referentes ao AMP, necessita-se de algumas definições importantes sobre o AM tradicional, as quais serão apresentadas a seguir.

“O AM é uma área da Inteligência Artificial (IA) cujo escopo de atuação inclui o estudo de algoritmos capazes de aprender automaticamente por meio da experiência, através do processamento de dados” (Sewell, 2009).

“Diz-se que um algoritmo aprende através de experiência, E , com relação a certas classes de tarefas, T , e medida de desempenho, P , ou seja, o desempenho, P , nas tarefas, T , melhora com a experiência, E ” (Mitchell, 1997).

Existem inúmeras aplicações onde diferentes tipos de experiências, tarefas e medidas de desempenho podem ser utilizados na construção de algoritmos de AM, os quais são frequentemente implementados para resolver problemas de alta complexidade para o ser humano.

2.4.1 Tarefas

Uma tarefa, T , em termos de AM, basicamente descreve a forma com que um algoritmo deve processar determinada amostra. Uma amostra é uma coleção de características, previamente medidas de forma quantitativa, que identifica um determinado objeto ou evento, o qual será processado pelo algoritmo de AM. “Uma forma frequente de representar uma amostra, é mediante um vetor, onde cada entrada do vetor é uma característica diferente da amostra” (Goodfellow, 2016). Um exemplo, aplicado no contexto desta pesquisa, é a representação de uma imagem, onde o vetor de características da mesma é representado pelos valores dos seus pixels.

A **Classificação** é uma tarefa comumente resolvida com AM. Neste tipo de tarefa o algoritmo deve determinar, entre um conjunto de categorias ou classes, a qual classe/categoria pertence determinada entrada. Para resolver tal tarefa o algoritmo é programado para produzir uma função $f \in \mathbb{R}^n$. Quando $y = f(x)$, o modelo designa à entrada representada pelo vetor x , uma categoria identificada por um código numérico y . Uma outra variante da tarefa de classificação ocorre quando a função produz como saída uma distribuição de probabilidade sobre as distintas classes (Goodfellow, 2016).

A **Deteção de anomalias** é também uma tarefa, com grande aplicação na atualidade. Nos casos de deteção, o algoritmo é preparado para processar um conjunto de objetos ou eventos e etiquetar aqueles que correspondem a um determinado padrão de interesse. Uma forma de implementar a tarefa de deteção é mediante a utilização da classificação em conjunto com métodos baseados em abordagens de Janelas Corrediças (Mitchell, 1997), isso será abordado na Seção 2.5.

2.4.2 Medida de desempenho

Com o objetivo de avaliar os resultados de um algoritmo de AM, é necessário estabelecer uma medida quantitativa do seu desempenho. A forma de calcular essa medida depende do tipo de tarefa para a qual o algoritmo foi implementado. Para tarefas de Classificação é frequentemente utilizada a curva ROC (*Receiver Operating Characteristic*). A curva ROC é um gráfico de sensibilidade (ou taxa de verdadeiros positivos) versus taxa de falsos positivos (Sewell, 2009). É uma ferramenta poderosa para medir e especificar problemas de desempenho de métodos diagnósticos em medicina, por permitir a variação da sensibilidade e especificidade para diferentes valores de corte.

Uma outra forma de medir o desempenho de modelos baseados em AM é o *Precision-Recall*. O *Recall* representa a proporção de amostras positivas corretamente identificadas pelo algoritmo, com respeito ao total de amostras positivas existentes em uma base de dados ou coleção de múltiplas amostras, equação (1). Enquanto a *Precision* é a proporção de amostras positivas corretamente identificadas do total de amostras identificadas como positivas, equação (2).

$$Recall = \frac{VP}{(VP+FN)} , \quad (1)$$

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

Em que *VP* são Verdadeiros Positivos (acertos do modelo), *FN* são Falsos Negativos e *FP* são Falsos Positivos (erros do modelo).

Utilizando as métricas de *Precision* e *Recall* é possível calcular uma outra métrica chamada de *F1-score*. O *F1-score* representa a média harmônica de *Precision* e *Recall* e é calculada mediante a equação (3).

$$F1 - score = 2 * \frac{Precision * Recall}{Precision + Recall} . \quad (3)$$

2.4.3 Experiência

Os algoritmos de AM, podem ser categorizados como Supervisionados ou Não Supervisionados, dependendo do tipo de experiência que alcança durante o processo de aprendizado (Goodfellow, 2016).

Os **algoritmos Não Supervisionados** são capazes de processar uma base de dados e aprender características úteis das amostras. São comumente utilizados em tarefas de agrupamento, que consistem em dividir a base de dados em grupos de amostras similares (Goodfellow, 2016).

Os **algoritmos Supervisionados** também são programados para aprender características de uma base de dados, mas, nestes casos, cada amostra é associada com uma etiqueta (do inglês *ground truth*). Assim, os algoritmos Supervisionados podem ser utilizados em tarefas como classificação (Goodfellow, 2016).

2.4.4 Redes neurais artificiais

Na Seção 2.4.1 apresentou-se uma definição de AM, como sendo algoritmos que são capazes de aprender a etiquetar classes a um conjunto de amostras, de forma autônoma. Dentre as ferramentas de AM utilizadas neste processo de aprendizado,

podem ser destacadas as Redes Neurais Artificiais (RNA), a qual é uma ferramenta computacional com inspiração biológica no funcionamento físico do cérebro humano (Sewell, 2009).

Uma RNA consiste em um sistema composto por unidades (neurônios ou nós) que ficam dispostos em uma ou mais camadas e se interligam por conexões que, na maioria dos modelos, estão associadas a pesos, os quais guardam o conhecimento que a rede adquiriu do ambiente. Os neurônios localizados na primeira camada da rede recebem estímulos do exterior e são chamados neurônios de entrada, os que se comunicam com o meio externo são chamados neurônios de saída e os demais são conhecidos como neurônios intermediários ou ocultos (Theodoridis e Koutroumbas, 2006). Na Figura 3, apresenta-se um exemplo de RNA com uma camada oculta.

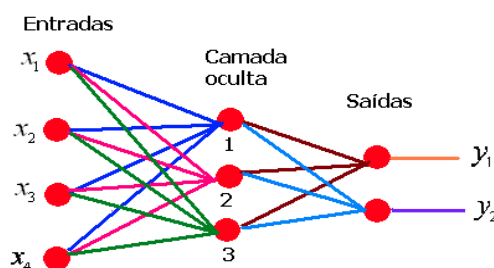


Figura 3 RNA com quatro neurônios na camada de Entrada (x_1 , x_2 , x_3 e x_4), três neurônios na Camada Oculta (1, 2 e 3) e dois neurônios na camada de Saída (y_1 e y_2). As conexões de cada neurônio foram representadas com cores diferentes. Observe que cada neurônio se encontra conectado com todos os neurônios da camada anterior.

“No processo de aprendizado, também chamado de treinamento, os pesos das conexões da rede são ajustados de forma iterativa, conforme o tipo de aprendizado que tenha sido implementado. Por exemplo, em casos de aprendizado supervisionado, os pesos serão atualizados levando em conta as saídas da rede assim como as etiquetas

de cada amostra fornecidas pelo “Supervisor” (Goodfellow, 2016). Os neurônios de cada camada são os encarregados de computar uma função, chamada de função de ativação, para produzir uma saída para a camada seguinte ou a saída final do sistema, utilizando os pesos e as entradas fornecidas pela camada anterior. A seleção da função de ativação depende do tipo de problema em questão. Geralmente são utilizadas funções matemáticas simples como: linear, sigmoide, tangente, etc. (Goodfellow, 2016).

Basicamente um algoritmo de AM, implementado com arquitetura de RNA, consiste em dispor diferentes camadas de neurônios, cujas funções de ativação devem permitir a atualização dos pesos da rede, de forma que a saída da mesma seja a mais próxima possível ao resultado esperado, sendo assim minimizada a função de custo ou erro. Na atualidade, um dos algoritmos mais utilizados com esse objetivo, é o algoritmo *Backpropagation*, que será abordado posteriormente nesta seção, o qual utiliza o algoritmo *Perceptron* para computar as saídas de cada neurônio.

Arquitetura *Perceptron*

Na arquitetura mais simples do *Perceptron*, somente um neurônio é utilizado para calcular a saída. A Figura 4 apresenta essa arquitetura.

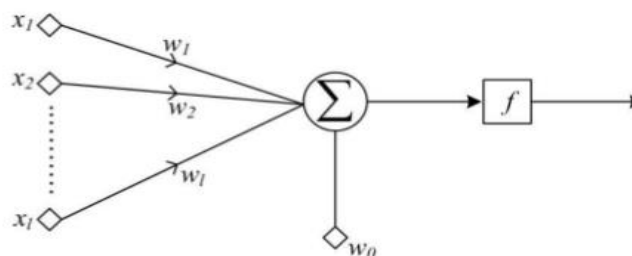


Figura 4 Arquitetura do *Perceptron* básico.

Fonte: (Theodoridis e Koutroumbas, 2006)

Em que $x_1, x_2, \dots, x_i$ representam as entradas da rede, $w_1, w_2, \dots, w_i$ são os vetores de pesos correspondentes às conexões do neurônio, w_0 é a polarização e f é a função de ativação.

O algoritmo de treinamento da arquitetura *Perceptron* utiliza a função apresentada na equação (4) para calcular as saídas dos neurônios. Essa função é linear e se aplica para classes separáveis.

$$g(x) = w'x + w_0 . \quad (4)$$

Em que $w = [w_1, w_2, \dots, w_i]'$ é o vetor de pesos, x é o vetor de características, e w_0 a polarização.

A Figura 5 apresenta o gráfico do hiperplano de separação. Observa-se que w é perpendicular ao hiperplano de decisão.

Os valores de d e z podem ser calculados mediante as equações (5) e (6) respectivamente.

$$d = \frac{|w_0|}{\sqrt{w_1^2 + w_2^2}} , \quad (5)$$

$$z = \frac{|g(x)|}{\sqrt{w_1^2 + w_2^2}} . \quad (6)$$

Em que w_0 é a distância euclidiana do ponto x no hiperplano de separação. Um lado do plano $g(x)$ é positivo (+) e o outro é negativo (-). Quando $w_0 = 0$, o hiperplano passa pela origem (Theodoridis e Koutroumbas, 2006).

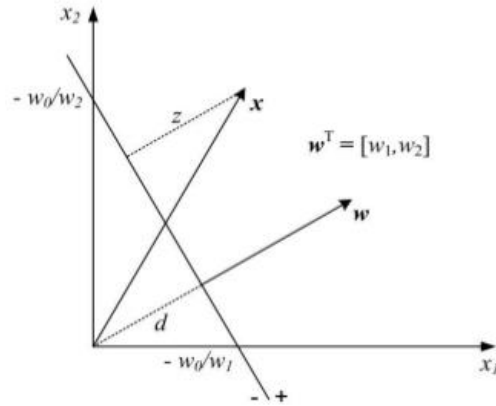


Figura 5 Hiperplano de decisão. Em um lado da linha $g(x) > 0$ e no outro lado $g(x) < 0$.

Fonte: (Campos Levi, 2012)

Os vetores de pesos são inicializados com valores arbitrários (W_0), e são corrigidos de acordo com a equação (7). O algoritmo é repetido até convergir para a solução, ou seja, até que todas as amostras sejam classificadas corretamente.

$$w(t+1) = w(t) - \eta t \sum_{x \in Y} \delta_x x \quad (7)$$

Em que $w(t)$ é o vetor de pesos estimado para a t -ésima interação, ηt é uma sequência de números reais positivos, $\sum_{x \in Y} \delta_x x$ representa o vetor de correção obtido utilizando as características classificadas erroneamente (Theodoridis e Koutroumbas, 2006).

Como já mencionado o *Perceptron* básico resolve problemas de classes linearmente separáveis. No caso de se ter classes impossíveis de serem separadas mediante uma função linear, existe uma variante do mencionado algoritmo, chamado de **Perceptron Multicamada**. O exemplo do problema do XOR, é uma forma breve, porém eficiente na hora de compreender o funcionamento deste algoritmo (Theodoridis e Koutroumbas, 2006).

A função booleana OU (OR) EXCLUSIVO (XOR), recebe como entrada dois valores binários (0 ou 1), e calcula uma saída que depende dos valores de entrada, conforme apresentado na Tabela 2.

Tabela 2 Tabela de verdade para a função XOR.

Fonte: Adaptado de (Theodoridis e Koutroumbas, 2006)

Primeira etapa				Segunda etapa
x_1	x_2	y_1	y_2	
0	0	0	0	$w_2(0)$
0	1	1	0	$w_1(1)$
1	0	1	0	$w_1(1)$
1	1	1	1	$w_2(0)$

Na Figura 6 são representadas as classes w_1 e w_2 nas respectivas posições em um espaço 2D. É possível observar que não existe uma reta ou função linear capaz de separar as duas classes.

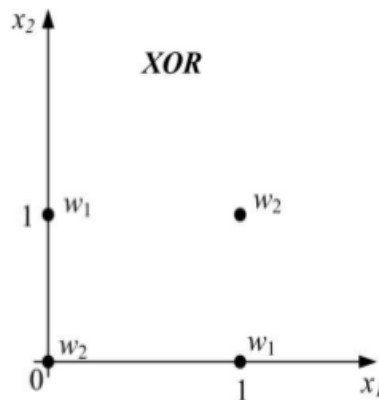


Figura 6 Representação espacial das classes do referido problema XOR.

Fonte: (Campos Levi, 2012)

Com apenas uma reta não é possível separar as classes. No entanto, traçando duas retas, dadas por duas funções lineares $g_1(x)$ e $g_2(x)$, é possível criar uma divisão entre w_1 e w_2 , conforme mostrado na Figura 7. Nessa figura, a região sombreada

representa a área pertencente à classe w_1 e a área fora da região sombreada corresponde à classe w_2 .

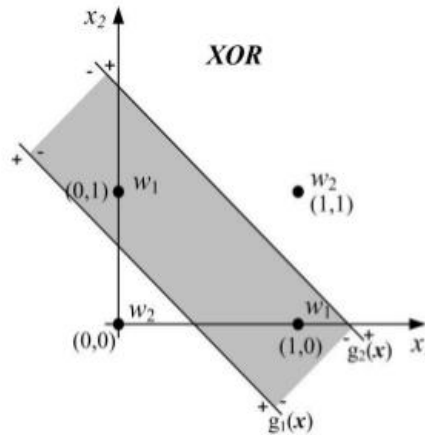


Figura 7 Hiperplano de decisão traçado utilizando *Perceptron* Multicamada para solucionar o problema XOR.

Fonte: (Campos Levi, 2012)

Para projetar uma solução, como essa mostrada na Figura 7 é necessário utilizar várias camadas de neurônios, o que dá origem ao nome da arquitetura: *Perceptron* Multicamada. A seguir mostraremos como é feita uma determinação ou projeto manual dos parâmetros dessa arquitetura. Serão executadas duas etapas para obter as duas linhas de decisão ou hiperplanos. Na primeira etapa, calcula-se a posição de um vetor de características x em relação aos hiperplanos. Na segunda etapa combinam-se os resultados da primeira etapa (Theodoridis e Koutroumbas, 2006). Um exemplo de arquitetura capaz de resolver o problema do XOR, é mostrada na Figura 8, onde os neurônios da Primeira Camada, também chamada de camada intermediária ou oculta, são encarregados de executar a primeira etapa, enquanto os neurônios da Segunda Camada ou camada de saída executam a segunda etapa (Campos Levi, 2012).

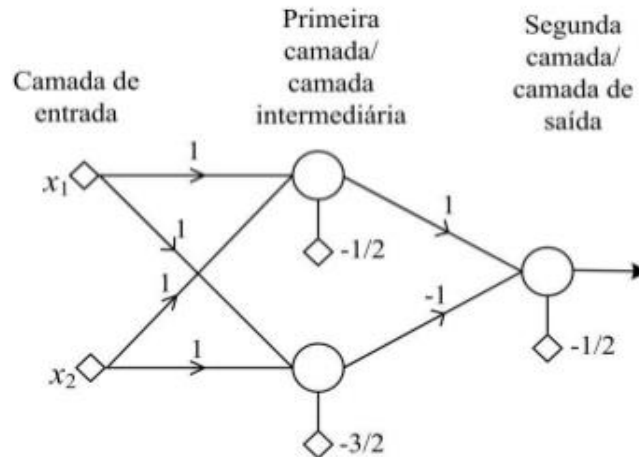


Figura 8 Exemplo de arquitetura do *Perceptron* multicamada capaz de resolver o problema XOR.

Fonte: (Campos Levi, 2012)

A determinação automática dos parâmetros dessa arquitetura é feita através do algoritmo *Backpropagation*, mostrado na sequência. Esse algoritmo é amplamente utilizado, não somente para resolver problemas de AM, mas também tem sido empregado na criação de RNA profundas, conteúdo que será abordado proximamente na Seção 2.4.5.

Algoritmo *Backpropagation*

Backpropagation é um algoritmo de AM supervisionado. O treinamento ocorre em duas fases que percorrem sentidos opostos: *forward* (para frente) e *backward* (para trás). Na Figura 9 apresenta-se o fluxo do algoritmo, onde a fase *forward* utiliza um padrão de entrada para definir a saída da rede. Os dados seguem o sentido *forward* da entrada para a saída. A fase *backward* usa a saída desejada e a saída calculada da rede para ajustar os pesos das conexões. Os erros seguem o sentido *backward* da saída para a entrada (Braga, 2000).

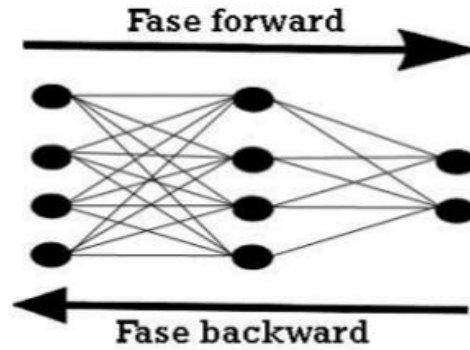


Figura 9 Fluxo de processamento do algoritmo *Backpropagation*.

Fonte: (Braga, 2000)

O algoritmo calcula os pesos da RNA, tentando minimizar a função de custo escolhida através do algoritmo do Gradiente Descendente (GD). A função de custo normalmente utilizada é o Erro Médio Quadrático (EMQ). A equação (8) expressa a forma de calcular o EMQ utilizado na implementação do presente trabalho.

$$EMQ = \frac{1}{m \sum (\hat{y} - y)^2} . \quad (8)$$

Em que m representa o número de amostras, $\hat{y}$ é a predição do modelo (valor calculado) e y é o valor esperado ou etiqueta fornecida para cada entrada (Braga, 2000).

O EMQ, assim como outras funções de custo, também é utilizado para expressar o erro de predição do modelo em amostras novas ou base de dados de teste, o qual deve ser diferente da base de dados de treinamento para garantir que os resultados das medidas de desempenho expressem realmente se o modelo é capaz de generalizar as suas predições. Generalizar significa ser capaz de classificar, de forma correta, novos casos, que não pertencem ao conjunto de treinamento, mas sim ao conjunto de teste.

O ajuste dos pesos, levando em consideração o erro obtido, é calculado segundo a equação (9).

$$w_i^n(novo) = w_i^n(antigo) + \Delta w_i^n . \quad (9)$$

Em que w_i^n representa o vetor de pesos do i -ésimo neurônio da n -ésima camada, $w_i^n(novo)$ é o novo valor do vetor de pesos, $w_i^n(antigo)$ é o valor antigo e Δw_i^n é o termo de correção para obter o próximo valor estimado de pesos, Δw_i^n é calculado mediante a equação (10) (Campos Levi, 2012).

$$\Delta w_i^n = -\mu \frac{\partial J}{\partial w_i^n} . \quad (10)$$

Em que J representa o valor da função de custo escolhida (Braga, 2000).

2.4.5 Aprendizado de máquina profundo (AMP)

O AMP é considerado um marco na área da Inteligência Artificial (IA). Diferente das RN tradicionais, que normalmente possuem uma ou duas camadas de neurônios, as Redes Neurais Profundas (RNP) são modelos mais complexos, compostos de múltiplas camadas com maior quantidade de unidades em cada camada e são capazes de representar funções cada vez mais complexas e de aprender representações de dados com diferentes níveis de abstração (Goodfellow, 2016). Estes modelos precisam de grandes bases de dados de treinamento para alcançar o melhor nível de generalização possível. Esse tema será melhor explicado na Seção 2.4.6.5. Os modelos baseados em técnicas de AMP empregam o algoritmo *Backpropagation* para guiar o aprendizado dos valores dos parâmetros internos que são utilizados para calcular os valores de ativação

ou saída de cada camada, a partir de valores semelhantes da camada anterior (Lecun *et al* 2015).

Métodos baseados em AMP têm apresentado resultados interessantes no Estado da Arte em aplicações tais como: reconhecimento e detecção de objetos em imagens (Li e Wang, 2017), descoberta de drogas (Wang *et al* 2016), etc.

2.4.6 Redes Neurais Convolutivas (RNC)

Redes Neurais Convolutivas (RNC) são um tipo especializado de RN que processam dados seguindo uma topologia em forma de *Grid*, como por exemplo, dados em forma de imagens, as quais são processadas como um *Grid* de pixels de duas ou três dimensões. “O nome de RNC advém do fato de que a rede emprega uma operação matemática chamada de Convolução, a qual é um tipo especial de operação linear”, que aplica um *kernel* em pequenas regiões do *Grid*, de forma repetida (Goodfellow, 2016). Uma definição simples de RNC é que são RN que utilizam a convolução pelo menos em uma das suas camadas.

2.4.6.1 Camada Convolutiva

A operação de Convolução aplicada em uma RNC é definida em espaços discretos, e representa a multiplicação entre duas matrizes. Vide equação 11.

$$s = x * w . \quad (11)$$

Em que o primeiro argumento, x , representa a **entrada** da rede e o segundo argumento, w , representa o **kernel**, o qual é, geralmente, de dimensões menores do que a entrada. A saída, s , é referida como **mapa de características** (Goodfellow, 2016).

A operação de convolução pode ser aplicada em várias dimensões. Para uma imagem de duas dimensões é recomendável utilizar um *kernel* bidimensional.

Na Figura 10 apresenta-se um exemplo do funcionamento da operação de convolução para duas dimensões. Nessa figura observa-se que a dimensão da saída pode mudar, dependendo das dimensões da entrada e do tamanho do *kernel*. No entanto, é possível controlar esse parâmetro da camada convolutiva, mediante a variação do hiperparâmetro passo (do inglês *stride*) (Goodfellow, 2016).

Quando se aplica RNC a imagens, o processo de extração de características não é efetuado, sendo as imagens completas passadas como entradas da rede. A própria rede encarrega-se de extrair as características de forma implícita. A utilização de RN tradicionais para o processamento de imagens de grandes dimensões implica em uma arquitetura de rede com um número muito grande de parâmetros. Por exemplo, para as imagens de baciloscopia utilizadas neste trabalho, que apresentam dimensões de $1388 \times 1040 \times 3$, a primeira camada de uma RN tradicional, com dez neurônios precisaria de 43.305.600 ($1388 \times 1040 \times 3 \times 10$) conexões, e consequentemente, pesos ligados as unidades da camada de entrada e os 4.330.560 ($1388 \times 1040 \times 3$) pixels da imagem de entrada, o qual iria comprometer o desempenho computacional da rede. As RNC empregam o conceito de conectividade local para evitar a conexão entre os neurônios das camadas convolutivas com cada elemento do volume de entrada (Goodfellow, 2016).

Vide Figura 11.

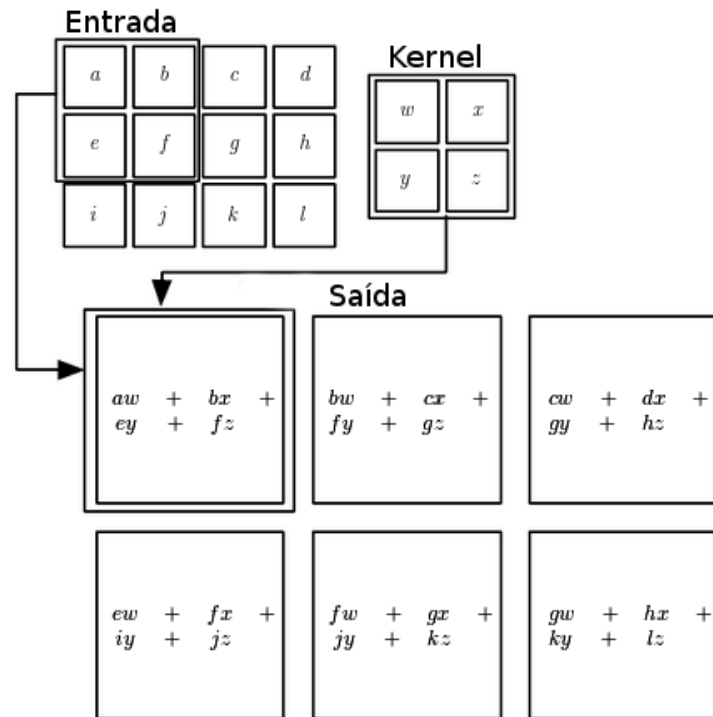


Figura 10 Exemplo de convolução de duas dimensões em que a saída é restringida apenas às posições onde o *Kernel* se encaixa completamente na entrada. Observe que as dimensões da entrada são de 3x4 (3 linhas e 4 colunas), o *Kernel* tem dimensões de 2x2 (2 linhas e 2 colunas). Assim a saída tem dimensões 2x3 (2 linhas e 3 colunas)

Fonte: (Goodfellow, 2016)

A **Conectividade Local** refere-se a conectar cada neurônio somente com uma região do volume de entrada. A extensão espacial desta região é um hiperparâmetro da rede, chamado de campo receptivo do neurônio (do inglês ***receptive field***) (isto é equivalente às dimensões espaciais dos filtros ou *kernels*). É importante destacar que existe uma diferença na forma em que são determinadas as dimensões espaciais (largura e altura) e a dimensão de profundidade: As conexões entre unidades de camadas adjacentes são locais do ponto de vista espacial (ao longo da largura e altura definidas), mas sempre são conexões completas com relação a profundidade completa do volume de entrada (Lecun *et al* 2015).

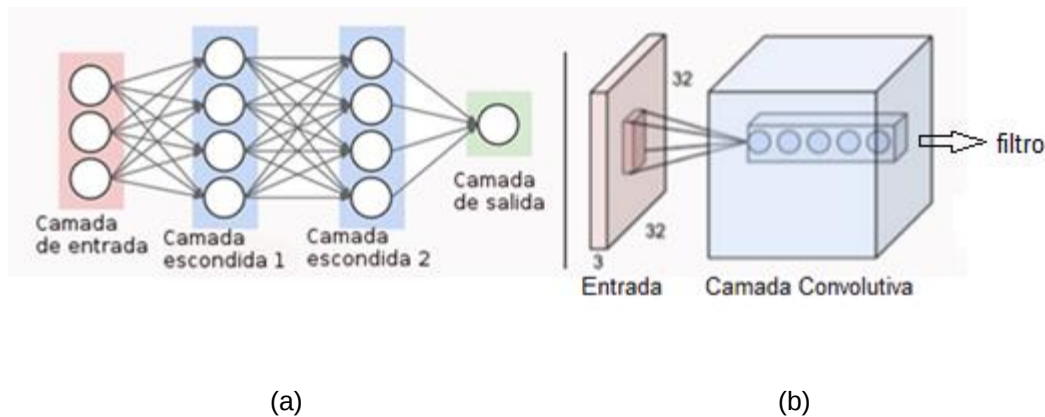


Figura 11 Em (a) apresenta-se uma RN tradicional com uma camada de entrada duas camadas ocultas (Camada escondida 1 e Camada escondida 2) e uma camada de saída observe que os neurônios das Camadas escondidas 1 e 2 estão conectados a todas as unidades da camada anterior. Em (b) mostra-se uma CC (Camada Convolutiva) com (Entrada de dimensões (32x32x3)). O cubo maior de cor azul, representa a CC, o qual contém um cubo menor que representa um filtro. Observe que o primeiro neurônio do filtro, representado por um círculo de azul mais intenso, está conectado somente com uma pequena área no volume de entrada.

Fonte: Adaptado de (Kar, 2015)

A Conectividade Local define a forma na qual são conectadas as unidades nas camadas Convolutivas com o volume de entrada. Para controlar a quantidade de neurônios no volume de saída, assim como a ordem destes, são definidos os seguintes **hiperparâmetros** de uma camada convolutiva (CC):

1. **Profundidade** (do inglês *depth*) do volume de saída representa o número de *kernels* ou filtros que serão utilizados na camada convolutiva. Cada filtro é orientado a detectar diferentes características na entrada, como por exemplo: bordas verticais, bordas horizontais, aglomeração de pixels com igual intensidade, e outras características visualmente fáceis de identificar em uma imagem.

2. **Passo** (do inglês *stride*), determina quantos pixels serão omitidos na hora de deslizar o filtro sobre o volume de entrada. Por exemplo se o passo é igual a 1, então os filtros serão deslocados 1 pixel cada vez, se o passo for igual a 2, então os filtros saltam de 2 em 2, e assim por diante. Quanto maior é o passo, menor serão as dimensões espaciais do volume de saída.
3. Além do *stride*, o **zero-padding** é utilizado para controlar as dimensões espaciais do volume de saída. Por exemplo, se é desejado preservar as dimensões da matriz ou imagem de entrada será necessário adicionar linhas e colunas ao redor desta, dependendo das dimensões de filtros definidas (Kar, 2015).

Para calcular as dimensões do volume de saída (H_2, W_2), de uma CC utiliza-se as equações 12 e 13.

$$H_2 = \frac{(H_1 - F + 2P)}{S + 1} , \quad (12)$$

$$W_2 = \frac{(W_1 - F + 2P)}{S + 1} . \quad (13)$$

Em que W_1 e H_1 representam as dimensões espaciais das entradas, F é o tamanho de campo receptivo (dimensões espaciais dos filtros), S representa o passo para convolver o volume de entrada, e por último, a quantidade de *zero-padding* usada para preencher as bordas da entrada, é representado por P (Kar, 2015). Na Figura 12 apresenta-se um exemplo do funcionamento das CC para um volume de entrada 3D.

Um outro conceito importante e representativo das CC é o de **parâmetros compartilhados** (do inglês *parameter sharing*). O compartilhamento de parâmetros

permite que se reduza o número de parâmetros (pesos e polarizações) na CC. A ideia explorada é a de que se é possível identificar uma característica em uma posição (x, y) , será possível identificar a mesma característica em uma posição (x_2, y_2) (Kar, 2015). Em outras palavras, os neurônios de cada porção de profundidade serão restringidos a utilizar os mesmos parâmetros: pesos e polarizações (do inglês *bias*).

Em resumo uma CC:

- Recebe um volume de entrada de tamanho $W_1 \times H_1 \times D_1$ (*width*, *height* e *depth*).
- São especificados quatro hiperparâmetros: número de filtros K , extensão espacial dos filtros F , o passo S e a quantidade de *zero-padding* P .
- Produz um volume de tamanho $W_2 \times H_2 \times D_2$, onde H_2 é determinado pela equação 12, W_2 pela equação 13 e D_2 , determinado mediante a equação 14.
- Mediante o Compartilhamento de Parâmetros são introduzidos a quantidade de pesos por filtro (Wf) definidos pela equação 15, o que representa um total de pesos (W) definidos pela equação 16 e K polarizações.
- No volume de saída, a d -ésima porção de profundidade, de tamanho $W_2 \times H_2$, é o resultado da aplicação da operação de convolução do d -ésimo filtro sobre o volume de entrada com passo S , utilizando-se o d -ésimo valor da polarização.

$$D_2 = K , \quad (14)$$

$$Wf = F * F * D_1 , \quad (15)$$

$$W = F * F * D_1 * K . \quad (16)$$

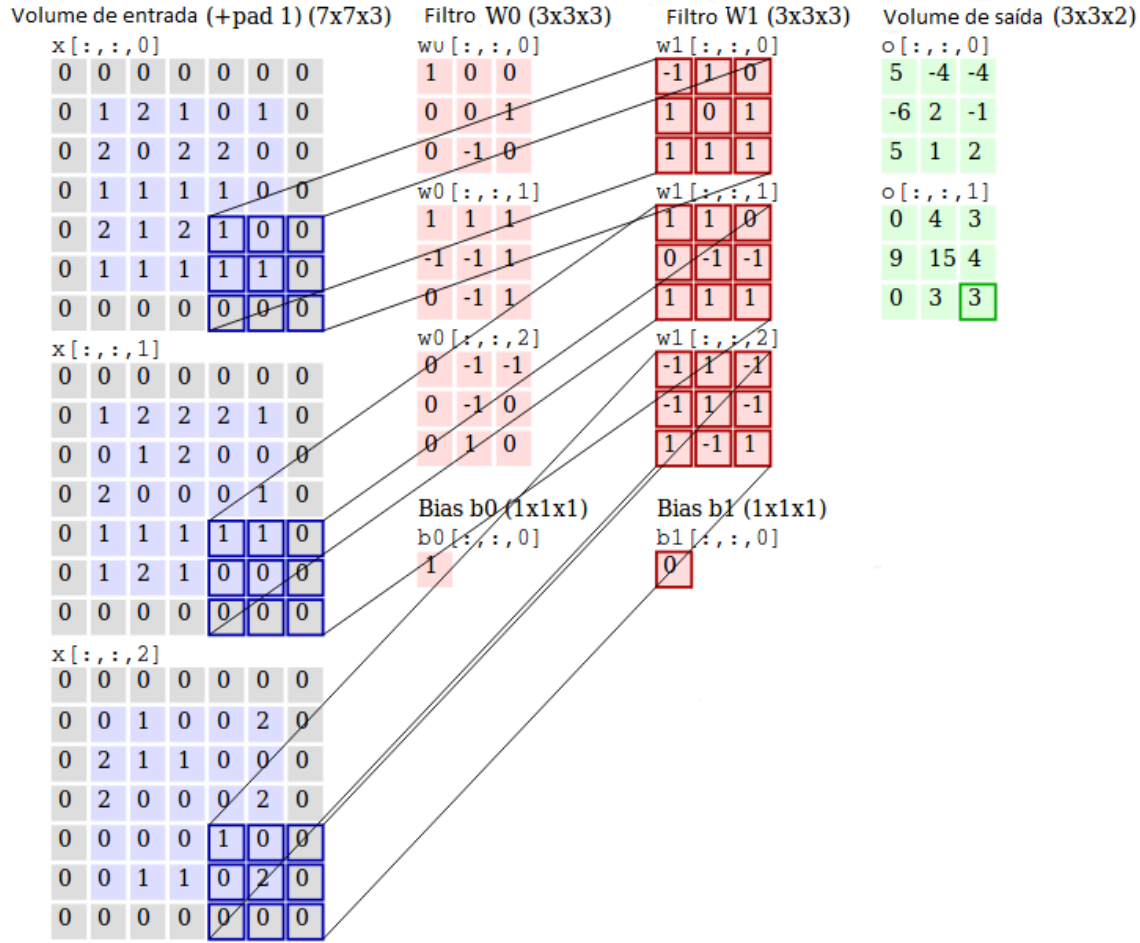


Figura 12 Exemplo de CC de profundidade, 3D, em que as matrizes da esquerda (volume de entrada), representadas em azul, são cada uma das dimensões da matriz $x : x[:, :, 0]$, $x[:, :, 1]$ e $x[:, :, 2]$. Observe que foi aplicado um *padding* de 1, $P=1$, nas bordas das matrizes. Esse *padding* é representado com valores 0 em nível de cinza. O volume de entrada resultante tem dimensões (7x7x3). As matrizes localizadas no centro da figura, w_0 e w_1 , correspondem aos pesos dos filtros w_0 e w_1 , respectivamente, aplicado com passo 1, $S=1$, sobre as entradas. As matrizes mostradas à direita da figura, em verde, representam o volume de saída o , de dimensões (3x3x2). Observe, que segundo as equações 12 e 13, a dimensão espacial da saída é (3x3). $w_2 = (5-3+2*1) / 2+1=3$ e $h_2 = ((5-3+2*1) / 2+1=3$, com profundidade $d_2 = 2$, segundo equação 14. A operação de convolução entre as matrizes $x[:, :, 0]$, $x[:, :, 1]$, $x[:, :, 2]$ e $w_1[:, :, 0]$, $w_1[:, :, 1]$ e $w_1[:, :, 2]$, têm como resultado o valor 3, contido em um quadro verde na saída $o[:, :, 1]$.

Fonte: Adaptado de (Kar, 2015)

2.4.6.2 RELU

Em arquiteturas de RNC amplamente utilizadas, como *ImageNet* (Krizhevsky *et al* 2012) cada CC é seguida pela aplicação de uma função de ativação. A seleção desta é fundamental, pois tem sido comprovado que a escolha da mesma pode levar a variações grandes no desempenho dos modelos (Sánchez, 2015).

Da mesma forma que nas RN tradicionais, nas RNC é possível utilizar funções de ativação como: *tanh* ou *sigmoid*. No entanto, no caso específico das camadas convolutivas, o estado da arte referente ao processamento de imagens com RNC, reportam os melhores resultados com a utilização da função de ativação não linear *RELU* (do inglês *Rectified Linear Units*). A camada CC, juntamente com a função *RELU*, foram inspiradas em estudos biológicos do córtex visual primária do ser humano (Goodfellow, 2016). A função *RELU*, conforme a equação 17, é aplicada pixel a pixel (do inglês *element-wise*): calcula-se o valor máximo entre o zero e o pixel x , de forma iterativa, em toda a extensão da entrada (Kar, 2015).

$$f(x) = \max(0, x) . \quad (17)$$

2.4.6.3 Camada *Pool*

Uma prática comum na hora de construir uma RNC, consiste na colocação de Camadas *Pool* (CP) entre sucessivas CC. Basicamente a função da CP é reduzir as dimensões espaciais do volume de entrada, reduzindo, por consequência, a carga de parâmetros da rede. A CP opera de forma independente em cada porção de profundidade do volume de entrada, fazendo redução das dimensões espaciais deste,

por meio de operações de Max (máximo) ou AVG (média do inglês *average*), dentre outras (Goodfellow, 2016).

A Figura 13 apresenta o efeito da CP.

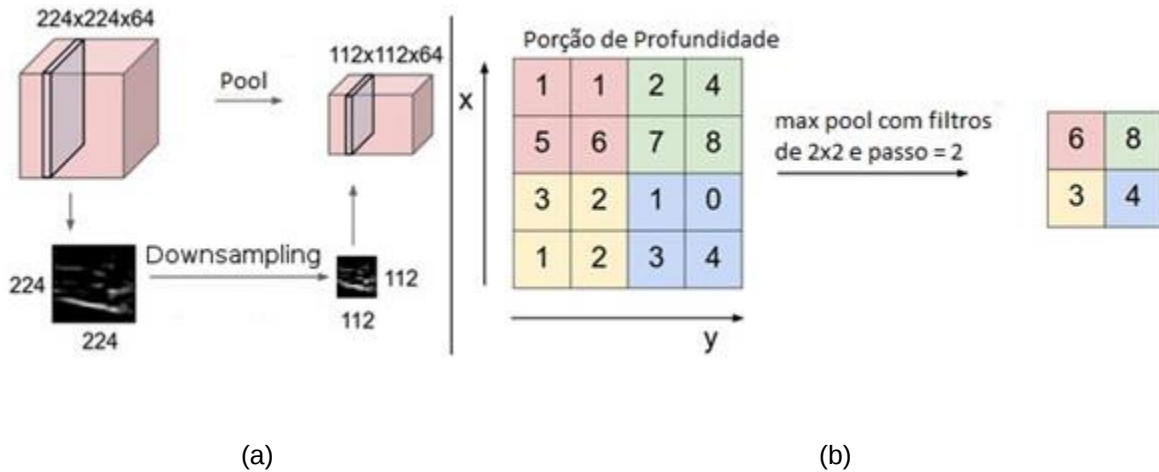


Figura 13 Efeito de redução espacial do volume de entrada mediante uma CP. Em (a) mostra-se uma imagem de dimensões (224×224) , a qual é reduzida mediante a CP para as dimensões (112×112) . A forma em que ocorre essa operação é mostrada em (b), em que a matriz da esquerda representa uma porção de profundidade do volume de entrada de dimensões 4×4 ; sobre essa porção é aplicada a operação de *Max-Pool* com filtro de (2×2) e passo $s=2$. O resultado é representado pela matriz da direita. Cada aplicação do filtro no volume de entrada e seu correspondente valor na camada de saída encontram-se representados com uma cor diferente. Observe que o volume de entrada de 4×4 foi reduzido para 2×2 . Fonte: Adaptado de (Kar, 2015)

De forma resumida, uma CP:

- Recebe um volume de entrada $W_1 \times H_1 \times D_1$.
- Necessita da especificação de dois hiperparâmetros: a dimensão espacial F e passo S .
- Produz um volume de dimensões $W_2 \times H_2 \times D_2$, em que os valores de W_2 , H_2 e D_2 , são encontrados mediante as equações 17, 18 e 19 respectivamente.

- Não introduz nenhum parâmetro, pois a função que executa depende exclusivamente do volume de entrada.

$$W_2 = \frac{(W_1 - F)}{S + 1} \quad , \quad (17)$$

$$H_2 = \frac{(H_1 - F)}{S + 1} \quad , \quad (18)$$

$$D_2 = D_1 \quad . \quad (19)$$

2.4.6.4 Camada Totalmente Conectada

Depois de várias camadas convolutivas e camadas *pool*, são comumente incluídas na arquitetura de uma RNC, as camadas Totalmente Conectadas (CTC), funcionando como um classificador tradicional. As conexões são calculadas mediante o algoritmo *Backpropagation*. Em seguida, na camada CTC é aplicada uma função não linear, conforme ocorre no caso das CC (Quinn *et al* 2016).

2.4.6.5 Regularização

Na Seção 2.4.1 foi mencionado que um modelo de AM possui uma boa capacidade de generalização quando classifica corretamente amostras ou entradas novas, além daquelas pertencentes ao conjunto de treinamento. Quando um modelo apresenta uma capacidade de generalização razoável, é possível dizer que o modelo se ajusta (do inglês *fits*) corretamente, caso contrário, pode ocorrer que o modelo tenha ajustado demasiadamente os parâmetros ao conjunto de treinamento (do inglês *overfitting*), ou pode não ter conseguido ajustar os parâmetros ao conjunto de treinamento (do inglês *underfitting*) (Goodfellow, 2016). Para evitar esses problemas,

várias técnicas, que têm sido englobadas no conceito de regularização, podem ser aplicadas.

Uma das principais técnicas aplicadas quando se trabalha com RNP é o **incremento da base de dados** (do inglês *Data Augmentation*), que consiste em aumentar artificialmente o volume de dados de entrada do modelo. Tal recurso é empregado para estudos em que os dados existentes não são suficientes, face a complexidade e a grande quantidade de parâmetros que, geralmente, apresentam os modelos profundos. Para implementar a técnica de incremento da base de dados, quando se trabalha com imagens, são aplicadas transformações não radicais nas entradas. Algumas práticas comuns são: rotação das imagens em vários ângulos, reflexão das mesmas, adição de ruídos, entre outras (Goodfellow, 2016). Assim, por exemplo, para uma base de dados original de 200 imagens, se são aplicadas: rotações de 90 e 180 graus, reflexão, adição de ruído gaussiano, em cada uma das imagens, obtém-se como resultado uma base de dados de 800 imagens, (200 imagens originais) * (4 transformações).

Uma outra técnica de regularização é o **Dropout**. Foi introduzida no ano 2012 (Srivastava *et al* 2012). O *Dropout* consiste basicamente em excluir da rede, em cada época de treinamento, neurônios individuais (escondidos ou visíveis), com probabilidade $1-p$, o que é equivalente a manter essas unidades com probabilidade p . Desta forma, em cada etapa do treinamento, a rede é reduzida de tamanho, gerando um novo modelo. Assim, vários modelos diferentes são gerados a partir da arquitetura inicial da rede. Posteriormente, as unidades removidas da rede são novamente inseridas, mantendo seus pesos originais (Srivastava *et al* 2012; Goodfellow, 2016).

Além da aplicação das técnicas mencionadas, no presente trabalho, foi efetuada a **Normalização** do volume de entrada. Em Srivastava *et. al.* (2012) recomenda-se introduzir os dados na rede com valores nos intervalos $[0,1]$ (utilizada nesse trabalho) ou $[-1,1]$, com o objetivo de minimizar a dimensão dos dados e aproximá-los em uma mesma escala.

A utilização de **filtros de dimensões pequenas** nas camadas convolutivas, geralmente (3×3) , é também uma boa prática na hora de conseguir um modelo que generaliza bem. A ideia é cobrir cada campo receptivo da imagem, utilizando uma menor quantidade de parâmetros do que seria necessário se fossem definidos filtros de dimensões maiores (Lecun *et al* 2015).

2.5 Detecção de objetos

Detecção de Objetos (DO) refere-se a detectar diferentes instâncias de uma classe particular de objeto em uma imagem. Comumente, a quantidade de objetos existentes é pequena em comparação com todas as possíveis localizações e escalas. “A abordagem mais comum de detecção de objetos reduz o problema a um Classificador Binário, onde uma imagem é percorrida por uma Janela Corrediça em cada posição, e o classificador determina se, naquela região, o objeto de interesse está presente ou não” (Amit e Felzenszwalb, 2015).

Algoritmos baseados exclusivamente na abordagem de Janelas Corrediças podem gerar detecções múltiplas em regiões onde a imagem apresenta somente um objeto de interesse, pois estes não validam a existência de detecções sobrepostas. O algoritmo **Non-maximum Suppression (NMS)** efetua a validação.

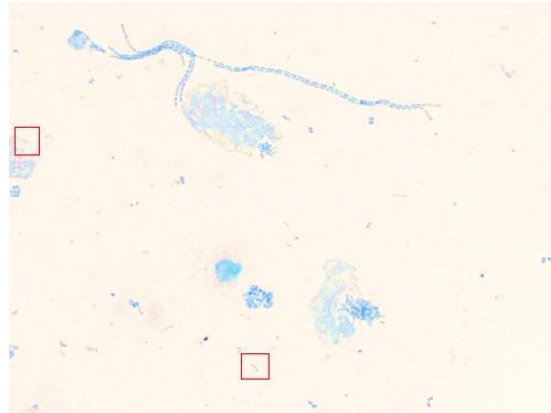
2.5.1 *Non-maximum Suppression* (NMS)

NMS é um algoritmo cuja função é evitar detecções múltiplas para um mesmo objeto (Falsos Positivos). Primeiramente, são selecionadas as dimensões espaciais (w,h) das janelas corrediças, de forma que o objeto de interesse encaixe completamente na mesma. Em seguida, essa janela é aplicada repetidamente, em todas as posições da imagem. Em cada posição, a operação gera um mapa de probabilidade para a região correspondente. Nesse mapa de saída, os pixels mais claros (com $p=1$) possuem a probabilidade mais alta de ser o objeto de interesse, enquanto os pixels pretos (com $p=0$) representam a ausência do objeto. Após este passo, pode ser aplicado um filtro gaussiano para eliminação de ruídos (áreas com baixas probabilidades de possuírem pixels do objeto de interesse), vide Figura 14. Como passo final, são eliminadas detecções superpostas. Isso é feito da seguinte forma: Em um conjunto de detecções vizinhas, é selecionada aquela com maior pontuação do classificador. Posteriormente, é calculada a sobreposição s , com respeito a cada vizinho. Finalmente, as detecções com sobreposição s maior do que um limiar de sobreposição (l_s) estabelecido, são eliminadas. Como resultado, o algoritmo suprime possíveis falsos positivos, que são produto de detecções múltiplas (Amit e Felzenszwalb, 2015). Vide Figura 15.

2.5.2 Avaliação das detecções

Na literatura, uma forma frequente de medir a precisão de um método de detecção de objetos em imagens é mediante a obtenção das métricas *Precision*, *Recall* e *F1-score* (Amit e Felzenszwalb, 2015); Quinn *et al* 2016).

Neste trabalho, uma detecção é considerada correta quando a distância entre a marcação do especialista e o ponto detectado pelo método é menor ou igual a um determinado limiar estabelecido, levando em consideração o tamanho médio dos bacilos.



(a)

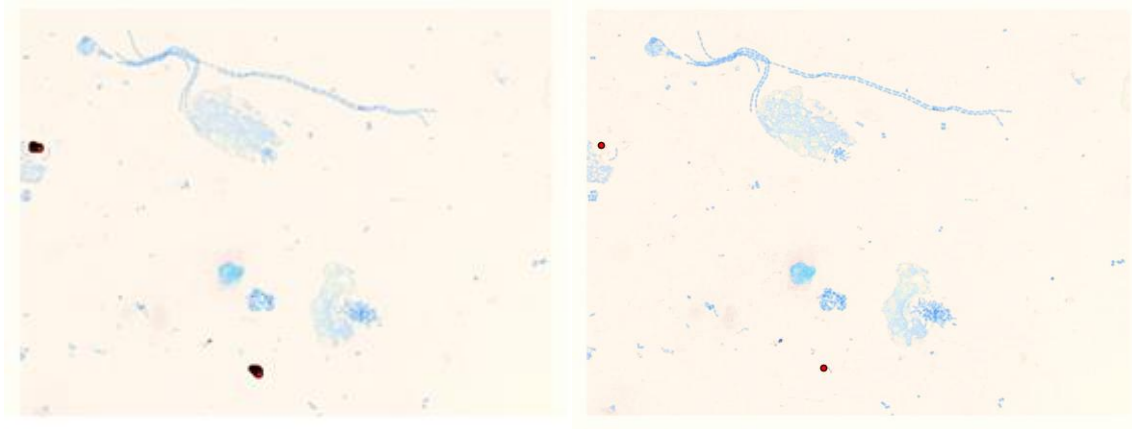


(b)



(c)

Figura 14 Em (a) mostra-se uma imagem de baciloscopia de campo claro com dois bacilos identificados por especialistas (objetos dentro dos quadros vermelhos). Em (b) apresenta-se o mapa de características obtido mediante a aplicação de janelas corredeiras e um classificador treinado para identificar bacilos. Observe a existência de falsos positivos em áreas incorretas, A_i (áreas não marcadas pelo especialista), assim como as detecções em áreas corretas, A_c (áreas marcadas pelo especialista e encerradas em quadros vermelhos). Em (c) mostra-se o resultado da aplicação de um filtro gaussiano no mapa de características. Observe como as áreas A_i foram suavizadas, quase apagadas, e as áreas A_c mantiveram a forma de bacilos.



(a)

(b)

Figura 15 Em (a) é ilustrada a aplicação da abordagem de janelas correções. Observe que foram obtidas múltiplas detecções para um mesmo bacilo (aglomerados de pontos vermelhos). Em (b) é ilustrada a aplicação do algoritmo NMS. Observam-se exatamente duas detecções (pontos vermelhos).

3. REVISÃO BIBLIOGRÁFICA

3.1 Introdução

Desde o ano 2008 têm sido publicados, em importantes bases de dados como IEEE *Xplorer*, *Web of Science* e *Scielo*, trabalhos focados na detecção automática do bacilo da TB em imagens de baciloscopia de campo claro. “Muitos grupos de pesquisa, notadamente dos países onde a TB é um problema grave de saúde (África do Sul, Índia, Malásia, Brasil) têm se debruçado em alcançar a meta da baciloscopia automatizada” (Costa *et al* 2016).

Neste capítulo é apresentada a revisão bibliográfica realizada, oferecendo informação sintetizada sobre os trabalhos citados, destacando pontos, como: os materiais e técnicas utilizadas no processo de detecção da bactéria, os resultados alcançados, assim como as métricas utilizadas para avaliar os resultados.

3.2 Estado da arte das pesquisas sobre identificação automática do bacilo da TB

Até o ano de 2016 identificamos, na literatura, a publicação de mais de 20 trabalhos cujo objetivo principal coincide com o objetivo central da presente pesquisa: detectar de forma automática o *Mycobacterium tuberculosis* em imagens de baciloscopia de campo claro. Para atingir o mencionado objetivo, 3 abordagens principais têm sido desenvolvidas:

1. Identificação do bacilo da TB mediante a aplicação de diferentes técnicas de **Processamento Digital de Imagens** para fazer segmentação dos objetos de interesse.
2. Identificação da bactéria mediante o emprego de classificadores baseados em **Aprendizado de Máquina Tradicional**.
3. E por último, detecção da bactéria mediante a aplicação de técnicas de **Aprendizado de Máquina Profundo**.

3.2.1 Detecção do *Mycobacterium tuberculosis* por segmentação, mediante técnicas de processamento digital de imagens

Trabalhos como Costa *et al* (2008), Raof *et al* (2008), Makkapati *et al* (2009), Sotaquirá *et al* (2009) e Nayak *et al* (2010), utilizaram técnicas de Processamento Digital de Imagens para identificar os bacilos em imagens de baciloscopia fazendo segmentação do objeto de interesse, mediante informação de cor tomada de diferente espaços de cores como: RGB em Costa *et al* (2008), Raof *et al* (2008) e Sotaquirá *et al* (2009), HSI em Makkapati *et al* (2009) e Nayak *et al* (2010), HSV, YIQ, YCbCr e Lab em Sotaquirá *et al* (2009).

3.2.2 Detecção do *Mycobacterium tuberculosis* mediante a aplicação de redes neurais artificiais

A partir do ano 2009 vários trabalhos começaram a utilizar Aprendizado de Máquina na detecção do bacilo da TB: Khutlang *et al* (2009), Khutlang *et al* (2010), Osman *et al* (2010a), Osman *et al* (2010b), Osman *et al* (2010c), Osman *et al* (2011), Costa *et al* (2012), García *et al* (2015), Rulaningtyas *et al* (2015) e Costa *et al* (2015). O

tamanho das bases de dados utilizadas pelos distintos autores varia de 100 para aproximadamente 1.900 imagens baciloscópias. Diferentes espaços de cor e características morfológicas foram explorados, assim como distintos tipos de classificadores foram implementados nos mencionados trabalhos.

Em Khutlang *et al* (2010) foi utilizado uma base de dados com cerca de 1.900 imagens de baciloscopia de campo claro de dimensões espaciais de 720x480 pixels. Os autores utilizaram uma combinação de um classificador bayesiano, com um classificador de regressão linear e com um classificador discriminante quadrático com o objetivo de obter um melhor resultado na tarefa de classificação de pixels. Todos os modelos foram treinados utilizando como entradas imagens de baciloscopia de campo claro RGB, gerando como saídas, imagens binárias com os potenciais bacilos segmentados. Posteriormente, a partir dessas imagens, técnicas de seleção de características como a transformada de Fisher foram aplicadas. Como último passo, com o objetivo de classificar os objetos segmentados como sendo bacilos ou não, foram aplicados os seguintes classificadores: classificador bayesiano, linear, quadrático, KNN, SVM e *Probabilistic Neural Network (PNN)*. Os resultados obtidos apresentam sensibilidade e especificidade acima de 95% para os seis classificadores, utilizando características da transformada de Fisher (Khutlang *et al* 2010).

Uma outra metodologia foi apresentada em García *et al* (2015). Os autores trabalharam com 100 imagens de baciloscopia de campo claro, as quais foram tomadas variando a iluminação em 7 capturas por imagem para um total de 700 imagens com resolução de 2048x1536 pixels, 50% positivas e 50% negativas. Como primeiro passo foram obtidas as imagens em alto contraste. Posteriormente, as imagens resultantes

foram processadas utilizando o algoritmo de clusterização *k-means* como método de segmentação, com o objetivo de separar potenciais bacilos do fundo. Posteriormente, um conjunto de características foram tomadas das imagens segmentadas e das imagens originais, em função da morfologia e textura nos espaços de cor RGB, HSV e Lab. Finalmente, essas características constituíram as entradas de um classificador baseado em Redes Neurais Artificiais, para identificar quais dos objetos segmentados eram efetivamente bacilos. Os autores apresentaram resultados de 97,59% de precisão média no passo de segmentação e 98% de acurácia após a aplicação do classificador de objetos. Uma contribuição interessante do trabalho é mostrar que, com diferentes níveis de iluminação, os resultados da segmentação, em termos de acurácia, variam de 87,88% a 99,97%.

Em Costa *et al* (2015) foi apresentado um outro método para a detecção do bacilo da TB em 120 imagens baciloscópicas de campo claro. Como primeiro passo trinta características foram geradas partindo das componentes combinadas dos espaços de cor: RGB, HSI, YCbCr e Lab. Posteriormente foi aplicada a técnica de seleção de características *Fisher's Discriminant Ratio*. As características selecionadas foram utilizadas como entrada de dois classificadores de pixels diferentes: *Support Vector Machine* (SVM) e *Feedforward Neural Network* (FNN), com o objetivo de segmentar bacilos. Para tanto, todos os pixels de uma imagem foram classificados como pertencendo ao objeto de interesse ou não. As imagens resultantes após a segmentação, foram utilizadas no segundo passo, em que combinações de três filtros morfológicos e geométricos foram aplicados para eliminar artefatos erroneamente identificados como

bacilos na etapa anterior. Os autores reportaram os seus melhores resultados com o classificador baseado em SVM+3 filtros, correspondendo a uma acurácia de 96,62%.

Tomando como exemplo as três pesquisas sumariadas, é possível afirmar ser difícil estabelecer uma comparação no desempenho dos trabalhos propostos, já que diferentes quantidades de imagens foram incluídas nas bases de dados utilizadas pelos distintos autores, diferentes características foram exploradas, assim como distintos classificadores foram implementados para atingir o objetivo de detectar os bacilos da TB em imagens de baciloscopia de campo claro.

3.2.3 Detecção do *Mycobacterium tuberculosis* mediante a aplicação de aprendizado profundo

Em Quinn *et al* (2016) focaram sua pesquisa em encontrar um modelo utilizando Redes Neurais Convolutivas para identificar três tipos de patógenos que causam diferentes doenças: a bactéria que causa a malária em imagens de amostras de sangue, o *Mycobacterium tuberculosis* em imagens de baciloscopia de campo claro e o parasito intestinal em imagens de amostras de fezes. Essas doenças são diagnosticadas mediante exames realizados empregando o microscópio ótico convencional. As imagens foram capturadas mediante a utilização de um *smartphone* acoplado ao microscópio utilizando um adaptador criado pelos próprios autores.

No caso da detecção do bacilo da TB, foco da presente pesquisa, os autores utilizaram um total de 928 imagens, nas quais especialistas identificaram 3.734 bacilos. Utilizando tais imagens foram gerados *patches* (subimagens) positivos de forma a conter os objetos marcados. Sobre os *patches* foram aplicadas sete transformações diferentes

para aumentar o número de amostras positivas. Amostras negativas foram geradas sem exceder 100 vezes o número de amostras positivas. O conjunto de *patches* positivos e negativos foi dividido em 50/50 para treinamento/teste (Quinn *et al* 2016).

Os autores projetaram um modelo baseado em RNC de quatro camadas escondidas: uma camada convolutiva de 7 filtros de tamanho 3x3, uma camada pool de fator 2, uma outra camada convolutiva com 12 filtros de tamanho 2x2 e, por último, uma camada totalmente conectada com 500 neurônios. O modelo foi treinado em 500 épocas e foi capaz de classificar um *patch* como contendo o objeto de interesse ou não (Quinn *et al* 2016).

Para detectar os bacilos em uma imagem de baciloscopia completa, os autores utilizaram o algoritmo *Non-maximum Suppression* (NMS) para descartar múltiplas detecções para um mesmo objeto.

Os autores compararam o método proposto com um outro método de visão computacional, em que características morfológicas são extraídas dos *patches*, conforme o abordado em Quinn *et al* (2014), e utilizadas como entradas de um classificador baseado em *Extremely Randomizer Trees* (ERT) (Geurts *et al* 2006) com 100 árvores (Quinn *et al* 2016).

Na avaliação de ambos modelos os autores apresentaram uma área sob a curva ROC (AUC) de 99% para o modelo RNC e de 76% para o modelo ERT. O melhor resultado alcançado não foi na classificação dos *patches* de imagens de baciloscopia, mas sim na classificação dos *patches* de imagens de malária, aparentemente devido ao fato do conjunto de amostras de sangue ser maior (Quinn *et al* 2016). O conjunto de teste

de imagens amostras de escarro não foi balanceado pois, segundo os autores, somente incluiu 9% de amostras positivas.

Na leitura do trabalho citado, não foram identificados resultados quantitativos na detecção do MT nas imagens completas de baciloscopia.

3.2.4 Considerações

De forma sumariada, a literatura consultada é apresentada na Tabela 3.

Dos trabalhos analisados 12 pesquisas conseguiram detectar o *Mycobacterium tuberculosis* mediante a utilização de técnicas de aprendizado de máquina tradicional. Vários espaços de cor e características morfológicas foram explorados. No entanto, somente um trabalho, o de Quinn *et al.* (Quinn *et al* 2016) desenvolveu uma solução utilizando aprendizado de máquina profundo, especificamente redes neurais convolutivas, porém a pesquisa não foi orientada unicamente à detecção da tuberculose em imagens de baciloscopia de campo claro. Como foi mencionado anteriormente, os autores propuseram uma arquitetura única de RNC que permite detectar três tipos diferentes de bactérias, entre as mesmas encontra-se o bacilo da TB, no entanto, os melhores resultados na aplicação do mencionado modelo não foram obtidos na identificação do *Mycobacterium tuberculosis*.

Ainda sobre o trabalho de Quinn *et al* (2016) o conjunto de teste, apesar de ter um número de amostras suficientemente grande, era desbalanceado (somente 9% das amostras eram positivas). Por outro lado, as imagens utilizadas no referido trabalho, não refletem as condições reais dos exames de baciloscopia. Na Figura 16 (a) e (b) apresenta-se um exemplo de uma imagem utilizada em Quinn *et al* (2016) com resolução

espacial 1632x1224 pixels e de uma imagem utilizada no presente trabalho com resolução espacial de 1388x1040. Observa-se que a proporção entre o tamanho dos bacilos (exemplares englobados em quadros pretos) e o fundo é diferente nas duas imagens. Aparentemente a Figura 16 (a) (exemplo de imagem usada em Quinn *et al*) é um fragmento ampliado de uma captura de um campo do microscópio óptico no exame de baciloscopia.

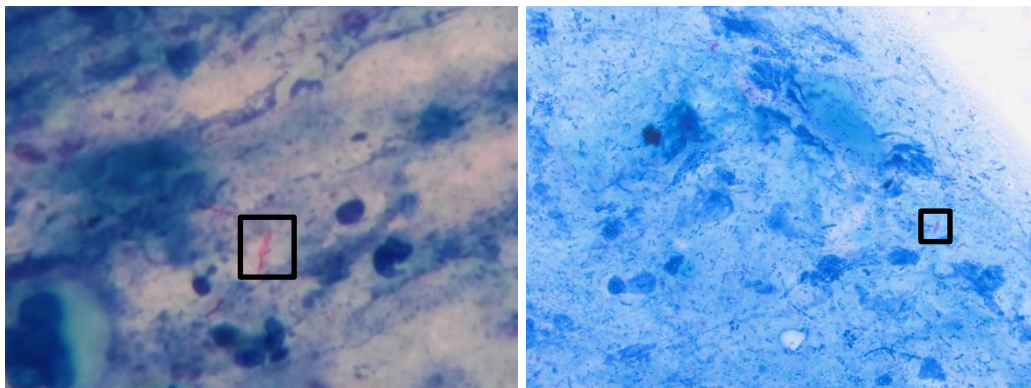


Figura 16 Imagens de baciloscopia de campo claro, (a) imagem utilizada em Quinn *et al* (2016) e (b) imagem utilizada no presente trabalho.

Atendendo ao estudo do estado da arte realizado, as principais contribuições do presente trabalho são apresentadas a seguir:

1. Geração de um conjunto de *patches* de imagens de baciloscopia de campo claro de quase 30.000 amostras totalmente balanceado, (50% das amostras serão negativas e 50% positivas). A disponibilização do referido conjunto irá facilitar o trabalho de pesquisadores cujo objetivo seja focado na detecção do *Mycobacterium tuberculosis* utilizando uma abordagem baseada em janelas correções.

2. Utilização de imagens de foco estendido (do inglês *Extended depth-of-field*) (até o momento não tem sido utilizado esse tipo de imagens em trabalhos prévios), em três formatos de cor diferentes: RGB, R-G e Escala de Cinza, as quais serão utilizadas para gerar o conjunto de *patches* mencionado no item 1. Esse conjunto será utilizado como entrada de modelos baseados em RNC, com o objetivo de explorar o comportamento dos referidos modelos em diferentes formatos de cor.
3. Para lograr êxito na identificação de um classificador de *patches* capaz de identificar corretamente o bacilo da TB em fragmentos de imagens baciloscópicas, várias arquiteturas de RNC com diferentes níveis de profundidade serão implementadas e testadas.
4. Técnicas de Regularização serão aplicadas nos modelos RNC com o objetivo de avaliar o impacto das mesmas sobre o desempenho dos modelos RNC na classificação de *patches*.

Tabela 3 Análise dos artigos relativos ao tema de detecção da tuberculose em imagens de baciloscopia de campo claro usando técnicas de Aprendizado de Máquina Tradicional e Profundo.

Referência	Materiais e métodos					Resultados
	Materiais	Distribuição da base de dados	Características	Método de Classificação	Pós Processamento	
Detection of tuberculosis in sputum smear images using two one-class classifiers. (Khutlang et al 2009)	33 imagens de 480x720 pixels	28/5 treino/teste	Coeficientes de Fourier, momentos invariantes, excentricidade, compacidade	Gaussian (G), Mixture of Gaussians (MoG) e KNN	Filtro de área após o primeiro classificador	MoG S = 90.88% E = 95.85% A = 93.47%
Classification of MT in Images of ZN-Stained Sputum Smears. (Khutlang et al 2010)	19 lâminas de 20-100 imagens por lâmina de 720x480 pixels	28/20 treino/teste	Coeficientes de Fourier, excentricidade, compacidade, média e desvio padrão dos canais G e B e valor do	Segmentação: Bayes + Regressão linear + discriminante quadrático. Classificação de objetos: KNN, Bayes, Linear,		KNN S = 97,77% E = 99,13% A = 98,55%

Materiais e métodos						
Referência	Materiais	Distribuição da base de dados	Características	Método de Classificação	Pós Processamento	Resultados
			pixel central dos canais R e G.	Quadrático, PNN e SVM.		
A genetic algorithm-neural Network Approach for MT Detection in Ziehl-Neelsen Stained Tissue Slide Images. (Osman <i>et al</i> 2010a)	120 imagens de 800×600 pixels		Sete momentos invariantes de Hu	Segmentação: <i>k-means</i> + algoritmo de crescimento de regiões		A = 89,64 %
				Classificador: baseado em Rede Neural Multicamada		
Colour Image Segmentation of Tuberculosis Bacilli in Ziehl-Neelsen-Stained Tissue Images using	De 750-1.250 imagens RGB e C-Y		Componente G do espaço RGB e componente <i>Ry</i> do espaço C-Y	Segmentação: <i>k-means</i> + filtro de mediana + algoritmo de		

Materiais e métodos						
Referência	Materiais	Distribuição da base de dados	Características	Método de Classificação	Pós Processamento	Resultados
Moving K-Mean Clustering Procedure. (Osman e Mashor <i>et al</i> 2010)	de 800x600 pixels			crescimento de regiões		
Segmentation of Tuberculosis bacilli in Zielh-Neelsen tissue slide images using hybrid Multilayered Perceptron Network. (Osman <i>et al</i> 2010b)	100 imagens HSI de 800x600 pixels		Componentes H e S do espaço HSI	Segmentação: Rede neural multicamada híbrida	Filtro de mediana	A = 99,82 %
Segmentation of Tuberculosis Bacilli in Ziehl-Neelsen-Stained Tissue Images based on K-Mean Clustering	100 C-Y imagens de 800x600 pixels		Componente de saturação do espaço de cores C-Y	Segmentação: <i>k-means</i> + algoritmo de crescimento de regiões	Filtro de mediana após <i>k-means</i>	

Materiais e métodos						
Referência	Materiais	Distribuição da base de dados	Características	Método de Classificação	Pós Processamento	Resultados
Procedure. (Osman <i>et al</i> 2010c)						
Bacilli detection in Ziehl-Neelsen-Stained Tissue using Affine Moment Invariants and Extreme Learning Machine. (Osman e Mashor <i>et al</i> 2011)	Imagens C-Y de 800×600 pixels	75/25 treino/teste	Componente de saturação do espaço de cores C-Y para Segmentação Seis momentos invariantes afins para classificação	Segmentação: <i>k-means</i> + algoritmo de crescimento de regiões Classificação: Rede neural com “ <i>Extreme Learning Machine</i> ”		A = 77,25 %
MT Recognition with Conventional Microscopy. (Costa <i>et al</i> 2012)	120 imagens RGB de 2816x2112 pixels		Componentes e diferença entre componentes dos espaços de cor	Segmentação: Rede Neural Multicamada	Filtros baseados em excentricidade e um parâmetro de cor.	A = 91,49 % S = 91,53 %

Materiais e métodos						
Referência	Materiais	Distribuição da base de dados	Características	Método de Classificação	Pós Processamento	Resultados
			RGB, HSI, YCbCr e Lab.			
Multi Patch Approach in K-Means Clustering Method for Color Image Segmentation in Pulmonary Tuberculosis Identification. (Rulaningtyas <i>et al</i> 2015)	~400 imagens de 640x360 pixels		Patches (Imagem baciloscópica subdividida)	Segmentação: <i>k-means</i>		A = 97,68 %

Materiais e métodos						
Referência	Materiais	Distribuição da base de dados	Características	Método de Classificação	Pós Processamento	Resultados
Automatic identification of MT. (Costa <i>et al</i> 2015)	120 imagens de 2816x2112 pixels		Componentes e diferença entre componentes dos espaços de cor RGB, HSI, YCbCr e Lab.	Segmentação: SVM e Rede Neural Artificial	filtro de tamanho+filtro geométrico (excentricidade) + filtro baseado em regras	SVM+3 filtros: S = 96,80 % A = 96,62 %
Detection of MT in microscopic images of Ziehl-Neelsen-stained sputum smears. (García <i>et al</i> 2015)	700 imagens de 2048x1536 pixels		Características morfológicas e de textura nos espaços de cor: RGB, HSV e Lab	Segmentação: <i>k-means</i> adaptativo Classificação: Rede Neural Artificial		A = 99,97%
Deep Convolutional Neural Networks for Microscopy-Based Point of Care	928 imagens	50/50 treino/teste		Classificação: Rede Neural Convolutiva	Filtro Gaussiano+algoritmo	AUC = 99%

Materiais e métodos						
Referência	Materiais	Distribuição da base de dados	Características	Método de Classificação	Pós Processamento	Resultados
Diagnostics. (Quinn <i>et al</i> 2016)					Non-maximum	
					Suppression	

Sensibilidade (S), Especificidade (E), Acurácia (A), Área sob a curva ROC (AUC), *Precision* (P), *Recall* (R).

4. MATERIAIS E MÉTODOS

4.1 Introdução

A presente proposta visa o desenvolvimento de uma metodologia para a detecção automática do bacilo da TB em imagens de baciloscopia de campo claro, adquiridas de lâminas contendo secreção pulmonar de pacientes adequadamente coradas através do método de *Kinyoun*. Apresenta-se, na Figura 17, um diagrama de blocos, em que se mostram as etapas da referida metodologia.

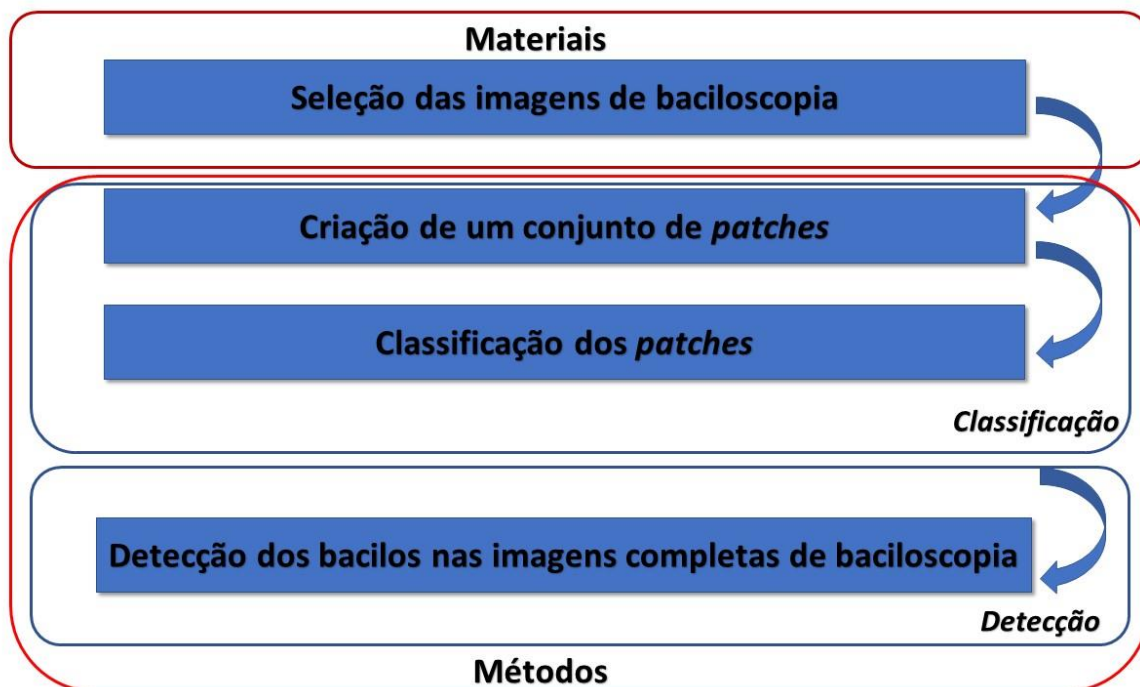


Figura 17 Diagrama de blocos que apresenta a metodologia implementada.

4.2 Materiais

No presente trabalho foram utilizadas imagens de baciloscopia de foco estendido (do inglês *extended depth-of-field (DOF)*), de dois pacientes, proporcionadas pelo grupo

de pesquisa em Reconhecimento de Padrões e Otimização da UFAM. As imagens foram obtidas pelo mencionado grupo, combinando imagens de 11 planos focais diferentes do mesmo campo de visão do microscópio ótico. Cada imagem apresenta uma resolução espacial de 1388x1040 pixels.

Para cada paciente foram gerados dois conjuntos de imagens de foco estendido com 164 campos cada um. Foram utilizados dois conjuntos do paciente 1 e um conjunto do paciente 2, totalizando 492 imagens. Três especialistas marcaram as mencionadas imagens identificando três tipos de elementos: bacilos, aglomerados de bacilos e bacilos duvidosos, conforme mostrado na Figura 18.

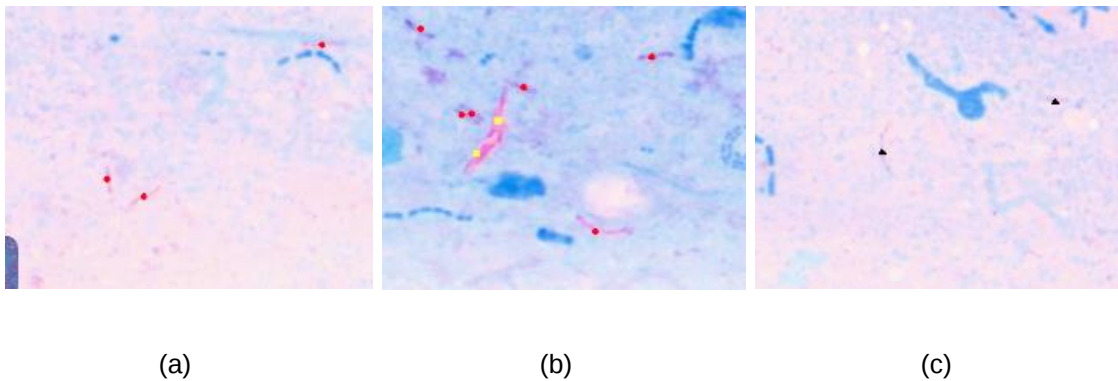


Figura 18 Fragmentos de imagens de foco estendido de baciloscopia de campo claro. Em (a) bacilos foram marcados com círculos vermelhos. Em (b) aglomerados de bacilos foram marcados com quadrados amarelos. Em (c) bacilos duvidosos foram marcados com triângulos pretos.

4.3 Métodos

4.3.1 Criação de uma base de dados de *patches*

Nesta etapa, utilizando as imagens de foco estendido de baciloscopia, foi construído uma base de dados de *patches*, denominado de conjunto original (CO), com 9.770 *patches*. O CO foi construído fazendo recortes de dimensões espaciais (40x40)

pixels nas imagens de foco estendido. Foram gerados dois tipos (classes) diferentes de *patches*: *patches* positivos (PP), e *patches* negativos (PN).

Os PP foram criados de forma a apresentar um bacilo nas coordenadas (20, 20), centro do *Patch*.

Os PN foram extraídos capturando regiões onde somente existem pequenos fragmentos de bacilos em imagens de foco estendido positivas. Desta forma, garantiu-se que bacilos completos não foram incluídos nos PN. Na Figura 19 são apresentados exemplos de PP e PN do CO.

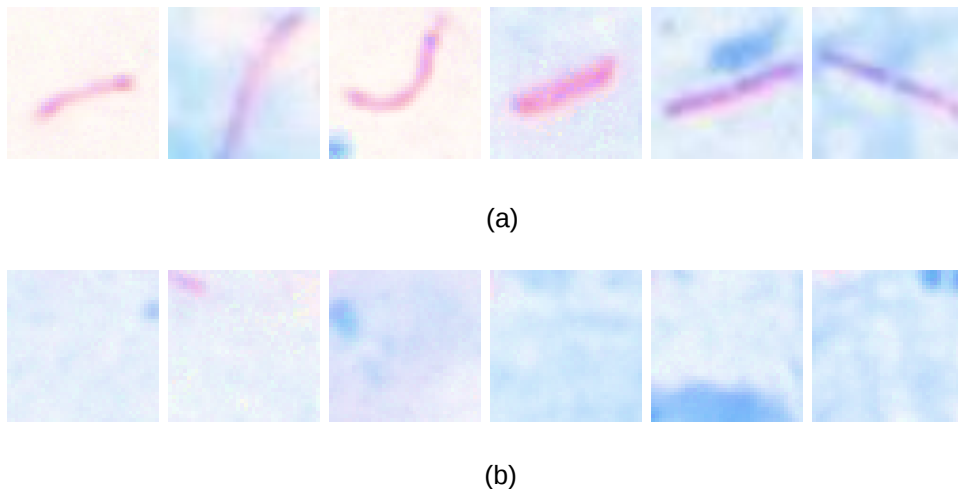


Figura 19 Em (a) apresenta-se exemplos de *patches* positivos (PP) incluídos no CO. Observe que o bacilo encontra-se centrado em cada *patch* em diferentes posições e orientações. Em (b) são mostrados exemplos de *patches* negativos (PN). Observe que o segundo *patch*, da esquerda para a direita, é um exemplo de *patch* que contém um fragmento de bacilo localizado na esquina superior esquerda.

O Conjunto original, no formato RGB, foi processado com o objetivo de obter os *patches* em mais dois formatos: Escala de Cinza e R-G. A Figura 20 ilustra o processo de criação desse conjunto.

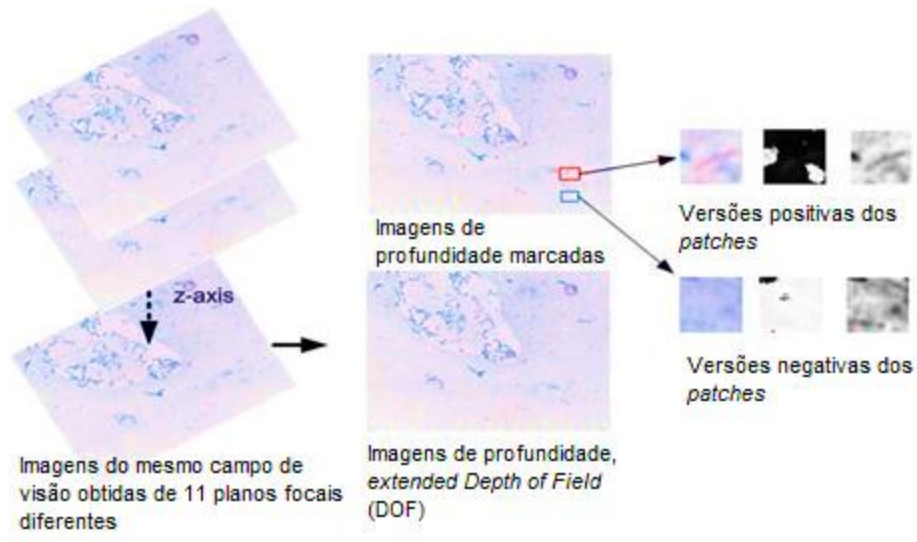


Figura 20 Procedimento seguido na criação do Conjunto Original (CO) de *patches* partindo dos 11 planos focais de cada imagem de baciloscopia de campo claro.

A partir do CO foi criado um outro conjunto de *patches* denominado de conjunto aumentado (CA), com o objetivo de usá-lo na aplicação da técnica de Regularização de incremento da base de dados. Para obter o CA foram aplicadas rotações em dois sentidos diferentes (90 e 180 graus) nos 9.770 *patches* do CO. Com tal procedimento, o conjunto aumentado foi constituído de 29.310 *patches*, 50% positivos e 50 % negativos.

4.3.2 Classificação de *patches*

4.3.2.1 Modelos RNC

Foram implementadas três arquiteturas diferentes de RNC, gerando três modelos distintos, conforme mostrado na Figura 21. O modelo 1 foi composto de uma CC, o modelo 2 por duas CC e o modelo 3 por três CC. A função de ativação utilizada nas CC foi *Relu*, enquanto que, na camada de saída, foi utilizada a função sigmoide. Nas CC experimentou-se passos $s=1$ e $s=2$. Não foi utilizado *padding*, dessa forma, após cada convolução, o volume de entrada perde duas unidades, tanto de largura como de

comprimento. Os pesos w foram inicializados seguindo a distribuição uniforme proposta em (Glorot e Bengio, 2010). Os conjuntos de *patches* criados (CO e CA) foram utilizados como entradas dos modelos nos três formatos ou versões abordadas na Secção 4.3.1.

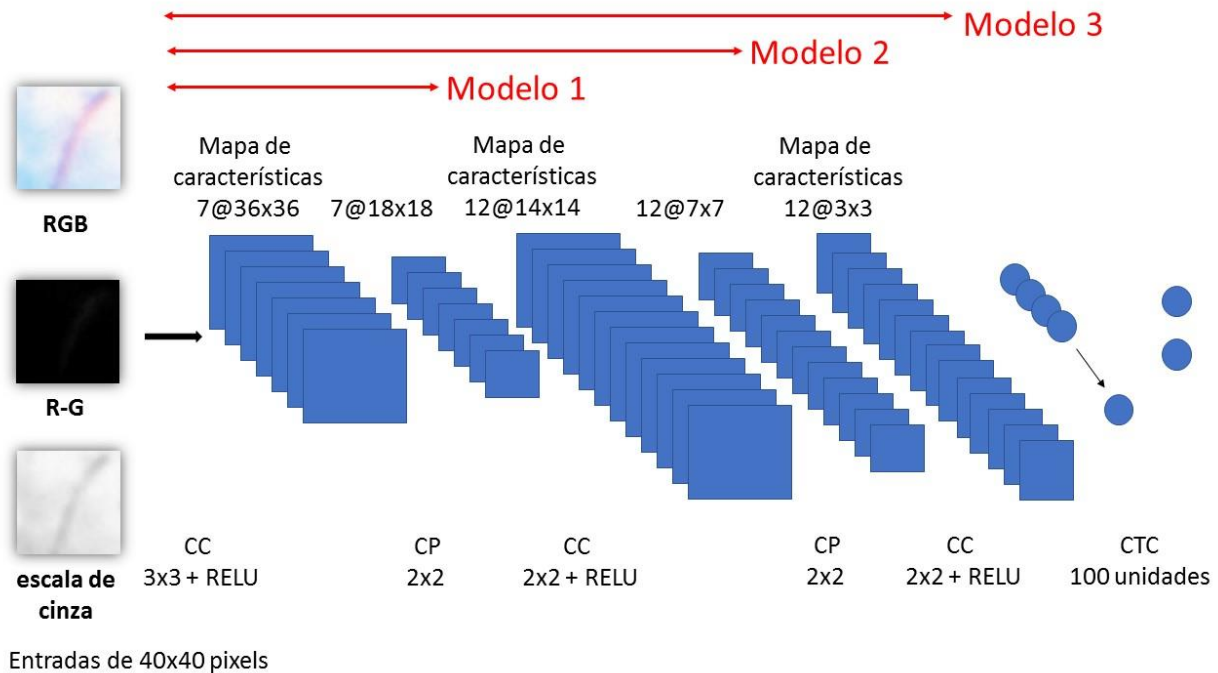


Figura 21 Ilustração da arquitetura dos Modelos 1, 2 e 3.

4.3.2.2 Detalhes do treinamento dos modelos implementados

Os modelos RNC implementados foram treinados primeiramente sem utilizar técnicas de regularização. Posteriormente, o treinamento utilizou as seguintes técnicas de regularização: incremento do conjunto de *patches* (CA), *dropout*, normalização do CA e utilização de filtros de dimensões pequenas nas CC.

Os três modelos foram treinados utilizando as três versões de *patches* (sem incluir e incluindo as técnicas de regularização mencionadas). Um total de 18 simulações foram implementadas: (3 modelos de RNC * 3 formatos de entradas * 2 (com e sem

regularização)). A Figura 22 mostra a configuração de treinamento dos modelos implementados. Observe que a técnica de normalização da base de dados foi aplicada nos dois casos, pois sem a aplicação da mencionada técnica os modelos não foram capazes de aprender, apresentando uma taxa de acerto de 50% (probabilidade do acaso). Foi aplicada uma camada *dropout* após cada CC. No Modelo 1 foi utilizada uma probabilidade $p = 0.2$. No Modelo 2 foram aplicadas probabilidades $p = 0.2$ e $p = 0.5$. Finalmente, no Modelo 3, após as CC 1, 2 e 3, foram utilizadas probabilidades $p = 0.2$, $p = 0.5$ e $p = 0.5$ respectivamente nas camadas *dropout*.

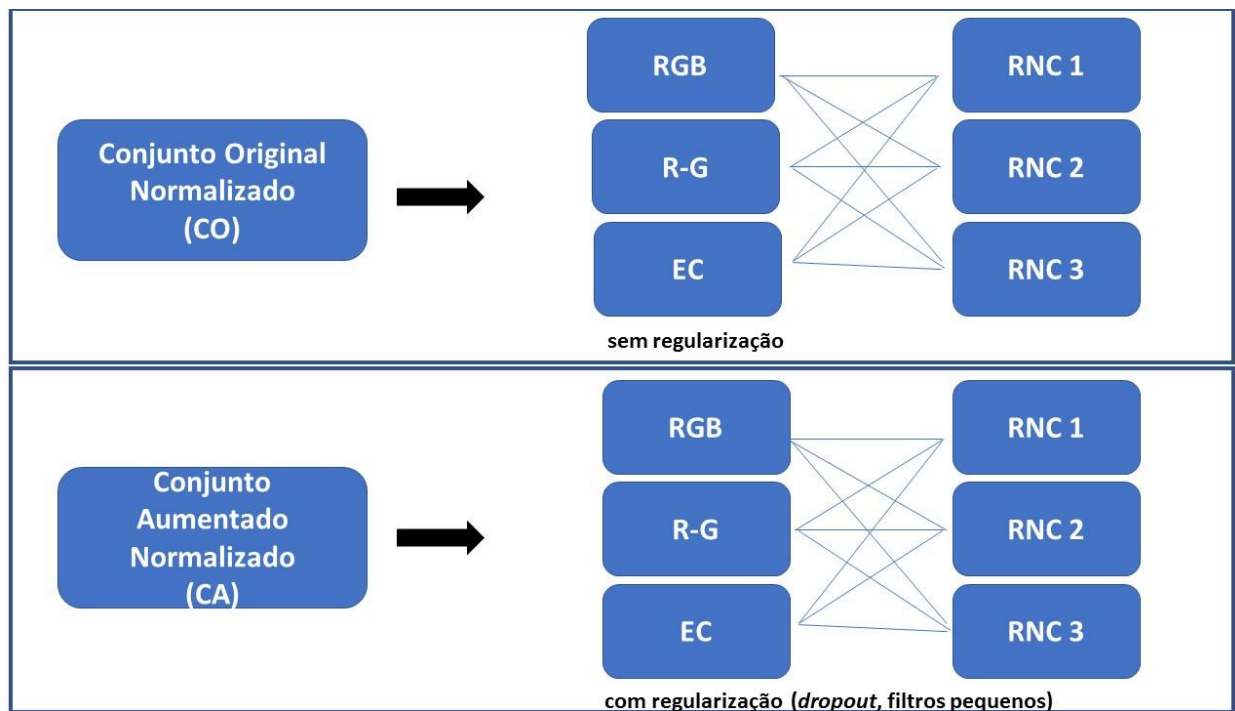


Figura 22 Configuração do treinamento dos três modelos baseados em Redes Neurais Convolucionais (RNC).

Para os processos de treinamento e teste dos modelos, o CO e o CA foram divididos como segue: 60% das amostras foram utilizadas para treinamento, 20% para validação e 20% restante para teste.

Os modelos foram treinados utilizando os seguintes hiperparâmetros de otimização incluídos na biblioteca *Lasagne*¹:

- Função de erro: erro médio quadrático (*lasagne.objectives.squared_error*)
- Atualização dos pesos: *lasagne.updates.nesterov_momentum*, fazendo ajuste do parâmetro *update_momentum*, começado por 0,9 e aumentando até 0,999.
- Taxa de aprendizado (do inglês *learning rate*), começando com valor de 0,03 e diminuindo até 0,0001.
- Número de épocas: 1.000.

Os parâmetros *update_momentum* e *update_learning_rate* são ajustados com o objetivo de diminuir o tempo de treinamento, conforme o recomendado em Sutskever *et al* (2013).

Na seção 4.4 serão mencionadas as tecnologias utilizadas na implementação da metodologia proposta.

4.3.2.3 Avaliação dos modelos

Os modelos treinados foram testados utilizando a base de dados de teste. Calculou-se o erro cometido nas classificações realizadas por cada modelo. Também foram geradas as curvas ROC para medir o desempenho dos modelos, o que facilitou a seleção da combinação modelo-formato de cor de melhor desempenho.

¹ <https://lasagne.readthedocs.io/en/latest/>

4.3.3 Detecção dos bacilos nas imagens de baciloscopia

Uma vez selecionado os melhores modelos para a classificação dos *patches* de imagens de baciloscopia, procede-se à etapa de detecção. Nessa etapa é aplicado um novo processo de classificação, em que uma janela corrediça é aplicada a cada região das imagens de baciloscopia. Como resultado, cada região é classificada como positiva (presença de bacilo) ou negativa (ausência de bacilo). Posteriormente, é aplicado o algoritmo *NMS* para eliminar detecções múltiplas para um único objeto.

4.3.3.1 Predição

O primeiro passo da etapa de detecção consiste na implementação de uma função para percorrer as imagens de foco estendido, mediante a aplicação de uma janela corrediça de dimensões (40x40). Em cada posição da janela, determina-se se a região é positiva ou negativa, mediante o uso do classificador de *patches* selecionado. A saída desse processo é um mapa de características da imagem completa. Com o objetivo de avaliar o melhor passo, foram avaliados três passos diferentes: 5, 10 e 20. Eventualmente, quanto menor o passo, maior é a quantidade de *patches* gerados e maior é a carga de processamento do algoritmo. Por exemplo, em uma imagem de baciloscopia com dimensões (1388x1040), utilizando passo $s=5$, são gerados 54.000 *patches*. Por outro lado, utilizando-se um passo $s=20$ são gerados 3.400 *patches*. As especificações da função implementada para executar este passo são mostradas na Tabela 4.

4.3.3.2. Obtendo as coordenadas das áreas positivas

O mapa de características gerado na predição foi utilizado para encontrar as coordenadas das áreas com alta probabilidade de serem positivas (do inglês *bounding*

boxes BB). Nessa etapa, foi estabelecido um limiar de probabilidade (lp) para aceitar uma predição como sendo positiva. Para reduzir os possíveis falsos positivos, opcionalmente utiliza-se um filtro gaussiano. Esta etapa retorna, para cada BB, duas coordenadas: esquerda superior e direita inferior. As especificações da função implementada para executar essa etapa são mostradas na Tabela 5.

Tabela 4 Especificações de entrada-saída da função implementada para a obtenção do mapa de probabilidades.

Função	Parâmetros de entrada	Parâmetros de Saída
Predição	1. Passo: Quantidade de pixels que são omitidos em cada salto da janela corrediça. 2. Tamanho: Dimensões espaciais da janela corrediça.	Mapa de probabilidades

Tabela 5 Especificações de entrada-saída da função implementada para a obtenção das BB.

Função	Parâmetros de entrada	Parâmetros de Saída
Obtenção das BB	1. Passo: Quantidade de pixels que são omitidos em cada salto da janela corrediça. 2. Tamanho: Dimensões espaciais da janela corrediça. 3. Gauss: Booleano (1 para executar o filtro gaussiano, 0 caso contrário). 4. Limiar de probabilidade (lp).	Coordenadas das <i>Bounding Boxes</i> que representam as áreas com maior probabilidade de serem bacilos.

4.3.3.3 Non-maximum Suppression

A última etapa do processo de detecção consistiu na aplicação do algoritmo *NMS*, abordado na Secção 2.5.1. Nessa etapa, o parâmetro limiar de sobreposição (*Is*) regula a supressão das *bounding boxes* que encerram as áreas com alta probabilidade de serem bacilos. Com o objetivo de encontrar o melhor valor de *Is*, o método foi simulado com diferentes percentuais de sobreposição, conforme o proposto em (Sánchez, 2015). Primeiramente, foram testados valores espaçados de dois em dois, desde 1% até 100%, (0,1; 0,3; 0,5; 0,7; 0,9). Posteriormente, partindo do valor de melhor desempenho encontrado, experimentou-se com espaçamento de 0,05, estabelecendo como limites do intervalo de *Is* uma unidade menor e uma unidade maior que o valor de melhor desempenho. Por exemplo, se o melhor valor com espaçamento de dois for 0,3, os valores testados com espaçamento de 0,05 serão: (0,2; 0,25; 0,3; 0,35; 0,4) (Sánchez, 2015).

Na implementação do método proposto foi utilizada a versão do algoritmo *NMS* implementada por (Rosebrock, 2015). As especificações do algoritmo são mostradas na Tabela 6.

Tabela 6 Especificações de entrada-saída do algoritmo *NMS*.

Função	Parâmetros de entrada	Parâmetros de Saída
NMS	1. BB: Coordenadas das áreas selecionadas como positivas na imagem de baciloscopia.	Coordenadas das <i>Bounding Boxes</i> que representam detecções de bacilos.
	2. Limiar de sobreposição (<i>Is</i>).	

4.4 Detalhes da implementação

Na implementação dos modelos de RNC foi utilizada a biblioteca *Lasagne*, essa biblioteca utiliza funções de *Theano*², uma outra biblioteca que facilita o trabalho com arranjos multidimensionais. *Lasagne* foi empregada no treinamento dos modelos e na visualização dos resultados. Em conjunto com *Lasagne* foi utilizado o pacote *Nolearn*³. Ambas ferramentas são baseadas na linguagem de programação *Python*⁴. A versão de *Python* utilizada foi a 2.7. A seleção desse ambiente de trabalho deveu-se ao fato do mesmo permitir realizar o treinamento dos modelos em *GPU* (*Graphics Processing Unit*), o que agiliza o processo de treinamento (Sánchez, 2015).

Também foi utilizada a biblioteca *scikit-learn*⁵, para avaliar os modelos implementados.

² <http://deeplearning.net/software/theano/index.html>

³ <https://pythonhosted.org/nolearn/>

⁴ <https://www.python.org/>

⁵ <http://scikit-learn.org/>

5. RESULTADOS

5.1 Resultados da etapa de classificação

Na etapa de classificação, com o objetivo de alcançar o erro mínimo de classificação, foram efetuadas diferentes simulações, através de variações nos modelos propostos. Nesse sentido, as técnicas de regularização desempenharam um importante papel.

5.1.1 Impacto da regularização

Nas simulações onde não foram aplicadas as técnicas de regularização, os modelos apresentaram taxas de acerto inferiores àquelas obtidas mediante a aplicação das referidas técnicas.

A aplicação das técnicas de incremento dos conjuntos de *patches* e *dropout* diminuiu o erro apresentado pelos modelos. Um exemplo disso é mostrado na Tabela 7, onde são apresentadas as taxas de erro alcançadas pelo modelo 2. No pior caso, utilizando como entradas 9.770 amostras da CO, sem a aplicação de *dropout*, alcançou-se um erro de 0,076. No melhor caso, utilizando 29.310 amostras do CA e aplicando-se *dropout*, obteve-se um erro de 0,034.

Tabela 7 Erro do modelo 2 com entradas no formato R-G com e sem aplicar as técnicas de regularização *dropout* e incremento da base de dados.

	Conjunto Original (CO)	Conjunto Aumentado (CA)
	9.770 amostras	29.310 amostras
Sem <i>Dropout</i>	0,07616	0,04879
<i>Dropout</i>	0,05788	0,03412

A normalização dos patches foi um fator chave para minimizar o erro de classificação dos modelos. Após a operação de normalização das entradas, os bacilos ficaram mais definidos e com intensidade mais forte, representando padrões mais fáceis de identificar, já que problemas como iluminação não uniforme, comum em exames de microscopia (Castleman e Young, 2008), foram diminuídos. A Figura 23 ilustra essa questão da iluminação.

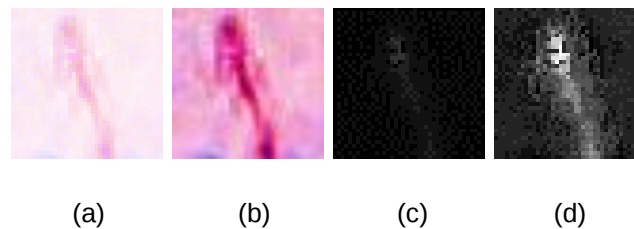


Figura 23 Normalização dos *patches*. Em (a) mostra-se um *patch* no formato RGB e em (b) apresenta-se o *patch* mostrado em (a) normalizado. Em (c) mostra-se um *patch* no formato R-G e em (d) é apresentado o *patch* mostrado em (c) normalizado. Observe que nos *patches* normalizados o bacilo é mais visível, pois existe maior contraste entre o objeto de interesse e o fundo.

5.1.2 Seleção dos hiperparâmetros das camadas convolutivas

Com o objetivo de selecionar os hiperparâmetros das camadas convolutivas de forma a potencializar o desempenho dos modelos desenhados, experimentou-se diferentes valores de passo e quantidade de filtros. Para selecionar o melhor valor de *passo*, os filtros foram aplicados com passo de um pixel ($s=1$), e com passo de dois pixels ($s=2$). Em ambos os casos, para entradas positivas, os filtros detectaram manchas que apresentavam forma de bacilo. Utilizando o *passo* igual a 2, uma maior quantidade de filtros apresentou esse tipo de resultado já a partir da primeira camada convolutiva, conforme mostrado na Figura 24. No caso dos *patches* negativos, as ativações das camadas convolutivas não apresentaram manchas com forma de bacilo, apesar de

serem utilizadas as mesmas configurações do modelo, conforme mostrado na Figura 25. Assim, os melhores resultados foram obtidos utilizando *passo* $s=1$. Por exemplo, utilizando o modelo 2 e entradas no formato R-G foi obtido um erro de teste de 0,031 utilizando *passo* $s=1$, enquanto que com *passo* $s=2$ o erro de teste foi de 0,050.

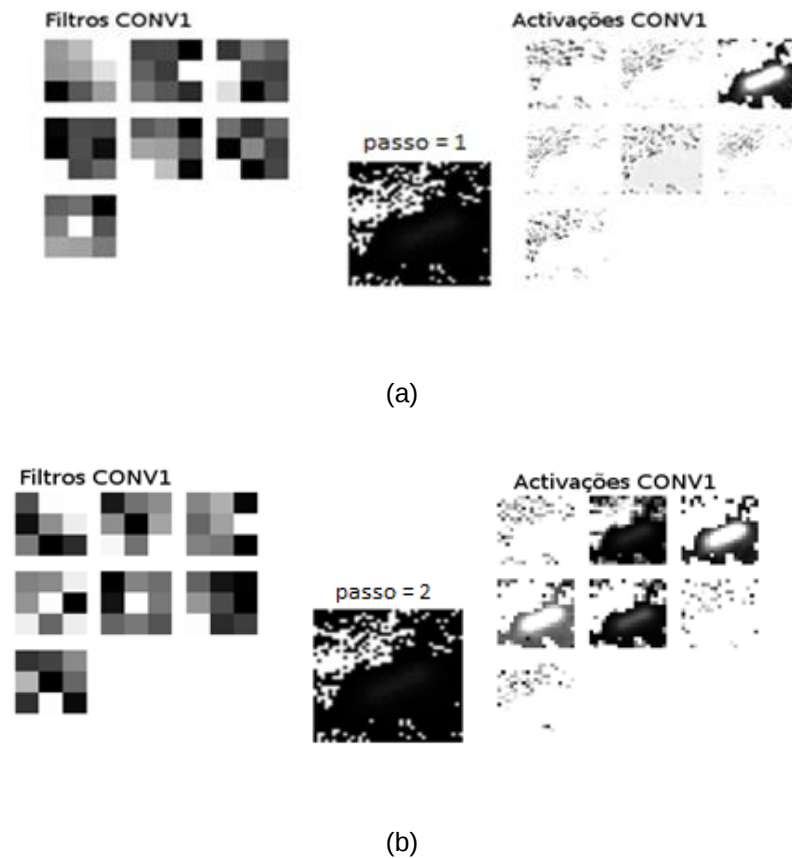


Figura 24 Em (a) mostra-se filtros e ativações da primeira camada convolutiva CONV1 do modelo 2 utilizando como entrada um *patch* positivo no formato R-G, com *passo* $s=1$ e em (b) utilizando *passo* $s=2$.

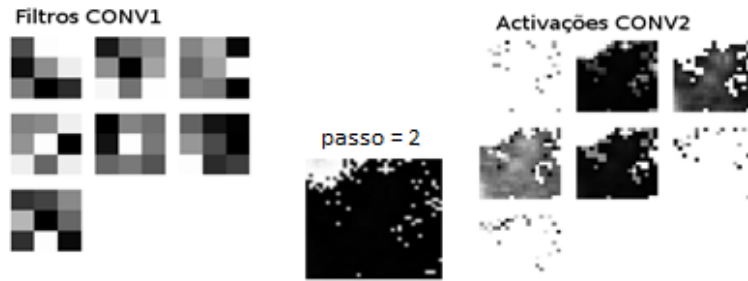


Figura 25 Filtros e ativações da primeira camada convolutiva CONV1 do modelo 2 utilizando como entrada um *patch* negativo no formato R-G com *passo* $s=2$. Observe que as ativações não apresentam manchas com forma de bacilo, diferentemente das ativações mostradas na Figura 24, onde foi utilizado um *patch* positivo como entrada do modelo.

Para selecionar a quantidade de filtros (profundidade) a ser utilizada nas camadas convolutivas, adotou-se uma estratégia de minimização da quantidade de filtros, pois, quanto mais filtros forem incluídos, maior quantidade de parâmetros apresenta o modelo, o que pode contribuir para o *overfitting* (Srivastava *et al* 2012). Inicialmente, tomando como exemplo o modelo 2 com entradas no formato R-G, foram incluídos 7 filtros na primeira camada convolutiva e 12 na segunda. Essa configuração foi inspirada no modelo proposto em Quinn *et al* (2016). Com essa configuração obteve-se um erro de 0,031. Posteriormente, aumentou-se a quantidade de filtros nas camadas convolutivas 1 e 2 para 16 e 32 respectivamente, obtendo um erro de 0,032. A diferença nos resultados não foi significativa.

Com referência ao tamanho dos filtros, trabalhou-se com filtros de dimensões pequenas (3x3) e (2x2), conforme o recomendado em Srivastava *et al* (2012).

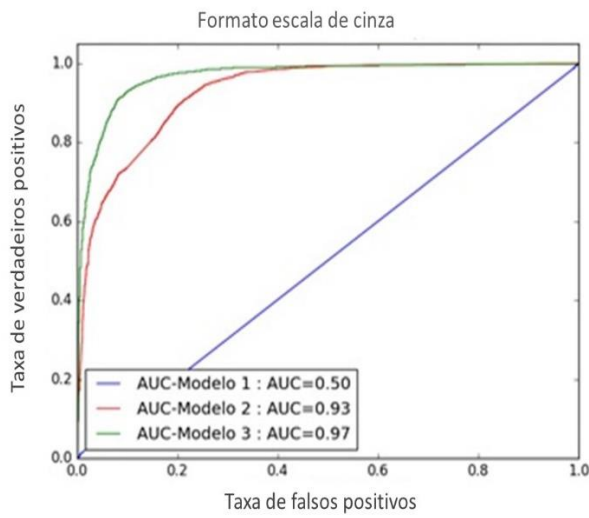
5.1.3 Seleção do modelo

Para selecionar o modelo com melhor desempenho na classificação de *patches*, os três modelos foram treinados para cada formato de entrada: RGB, R-G e Escala de

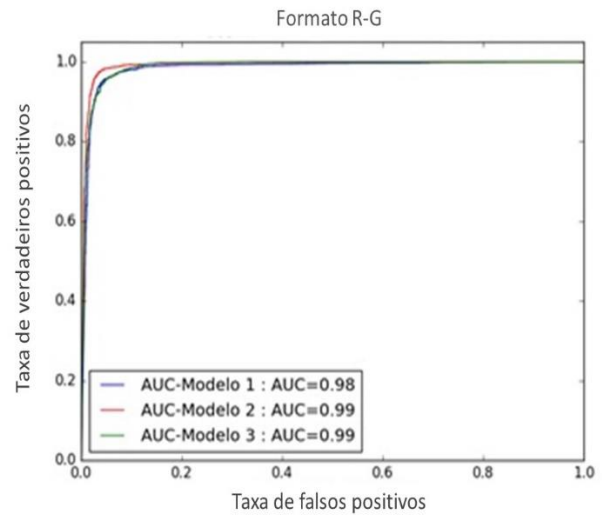
Cinza, utilizando em cada simulação, a mesma configuração de hiperparâmetros nas camadas da rede, assim como os mesmos hiperparâmetros de otimização.

A Figura 26 apresenta as áreas sob a curva ROC dos três modelos implementados utilizando o CA nos formatos: escala de cinza, R-G e RGB. Os piores resultados foram alcançados com o modelo 1, para os três formatos de cor. Esse modelo teve um comportamento significativamente variante. No formato escala de cinza, o modelo não aprendeu ($AUC = 50\%$), veja Figura 26 (a), esse comportamento é explicado pela presença de artefatos azuis com morfologia de bacilo nas imagens de baciloscopia, nesses casos, a informação de cor não é estudada, fazendo com que o classificador classifique os mencionados artefatos como sendo bacilos). Para entradas no formato R-G o modelo 1 apresentou um melhor resultado ($AUC = 98\%$), vide Figura 26 (b), devendo-se ao fato que a rede consegue processar duas informações importantes para a detecção do MT, cor e forma. No formato RGB, apesar do modelo poder processar também essas informações, o resultado foi novamente de $AUC = 50\%$, vide Figura 26 (c), esse comportamento deve-se ao número de parâmetros da rede, pois diferentemente do formato R-G e escala de cinza, o formato RGB apresenta três canais, portanto a entrada da rede é três vezes maior, aumentando assim, o número de pesos.

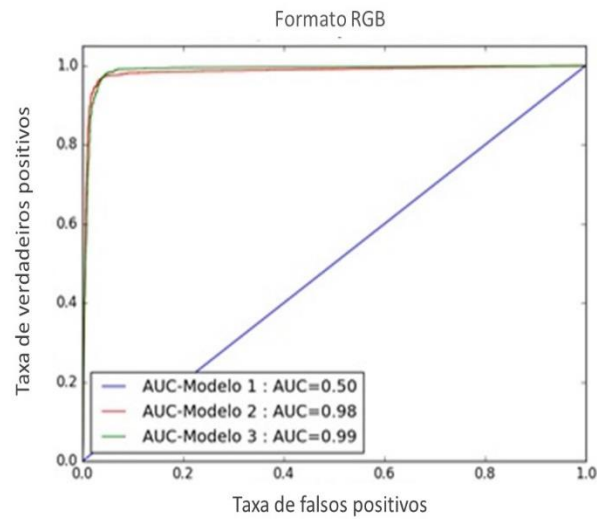
Os melhores resultados foram apresentados utilizando os modelos 2 e 3, para entradas nos formatos R-G e RGB, alcançando uma área sob a curva ROC superior a 97% em cada caso.



(a)



(b)



(c)

Figura 26 Curvas ROC dos modelos 1, 2 e 3 para o CA nos formatos: (a) Escala de Cinzas, (b) R-G e (c) RGB.

Finalmente, os modelos 2 e 3 foram escolhidos para serem utilizados na tarefa de classificação de *patches* na etapa de detecção.

5.1.4 Influência do tipo de imagem nos resultados

Os piores resultados no treinamento dos modelos foram obtidos com as bases de dados de *patches* em escala de cinza. Na verdade, esse resultado já era esperado pois, nas imagens de baciloscopia, a informação de cor é decisiva na identificação dos bacilos. Por outro lado, utilizando os *patches* RGB e R-G, foram obtidas taxas de erro menores (aproximadamente 3%).

5.2 Resultados da etapa de detecção dos bacilos

Uma vez selecionados os modelos 2 e 3 como classificadores de *patches* a serem aplicados na etapa de detecção, realizou-se os experimentos. O objetivo dessa etapa foi localizar os bacilos nas imagens de baciloscopia completas mediante a aplicação do algoritmo NMS.

5.2.1 Mapas de probabilidade

O primeiro passo na etapa de detecção foi a obtenção do mapa de probabilidades (MP). Para tanto, os classificadores de *patches* escolhidos foram aplicados em cada imagem de baciloscopia completa, com passo igual a 5, 10, e 15 pixels. Os melhores resultados foram obtidos com passo igual a 5 pixels. Com esse passo, uma maior quantidade de *patches* é analisada, o que potencializa um maior índice de detecções.

Os MP, antes e depois da aplicação do filtro gaussiano, podem ser visualizados como imagens, em que os pixels com maiores valores de probabilidade de serem bacilos aproximam-se do valor 1 (tonalidade branca). Na Figura 27 são mostrados os MP antes e depois da aplicação do filtro gaussiano, para uma imagem de baciloscopia nos

formatos RGB, R-G e escala de cinza. Observa-se, na Figura 27 (c), que as detecções fortemente marcadas em (b) foram suavizadas mediante o filtro gaussiano, o que permitiu que ruídos fossem removidos antes da aplicação do algoritmo NMS. Ressalta-se também a diferença entre os MP obtidos com os três formatos: (b) RGB, (f) R-G e (i) escala de cinza. O MP gerado utilizando o formato RGB apresenta menor quantidade de ruídos em comparação com os MP gerados utilizando os formatos R-G e escala de cinza. Uma outra observação importante é que as áreas com alta probabilidade de serem bacilos (áreas claras), preservam a forma dos bacilos presentes na imagem de entrada, conforme mostrado nas Figuras 27 (a), (b) e (c), através de bacilos circutados por elipses.

5.2.2 Seleção do limiar de probabilidade

Uma vez obtido o MP da imagem, a etapa seguinte é a aplicação do limiar de probabilidade (lp). Áreas com probabilidade de ocorrência de um bacilo menor do que lp são descartadas. O método foi validado para os valores de lp iguais a 0,4, 0,5 e 0,85. Os melhores resultados foram obtidos com $lp = 0,5$, para as imagens RGB e escala de cinza, e com $lp = 0,85$, para as imagens R-G.

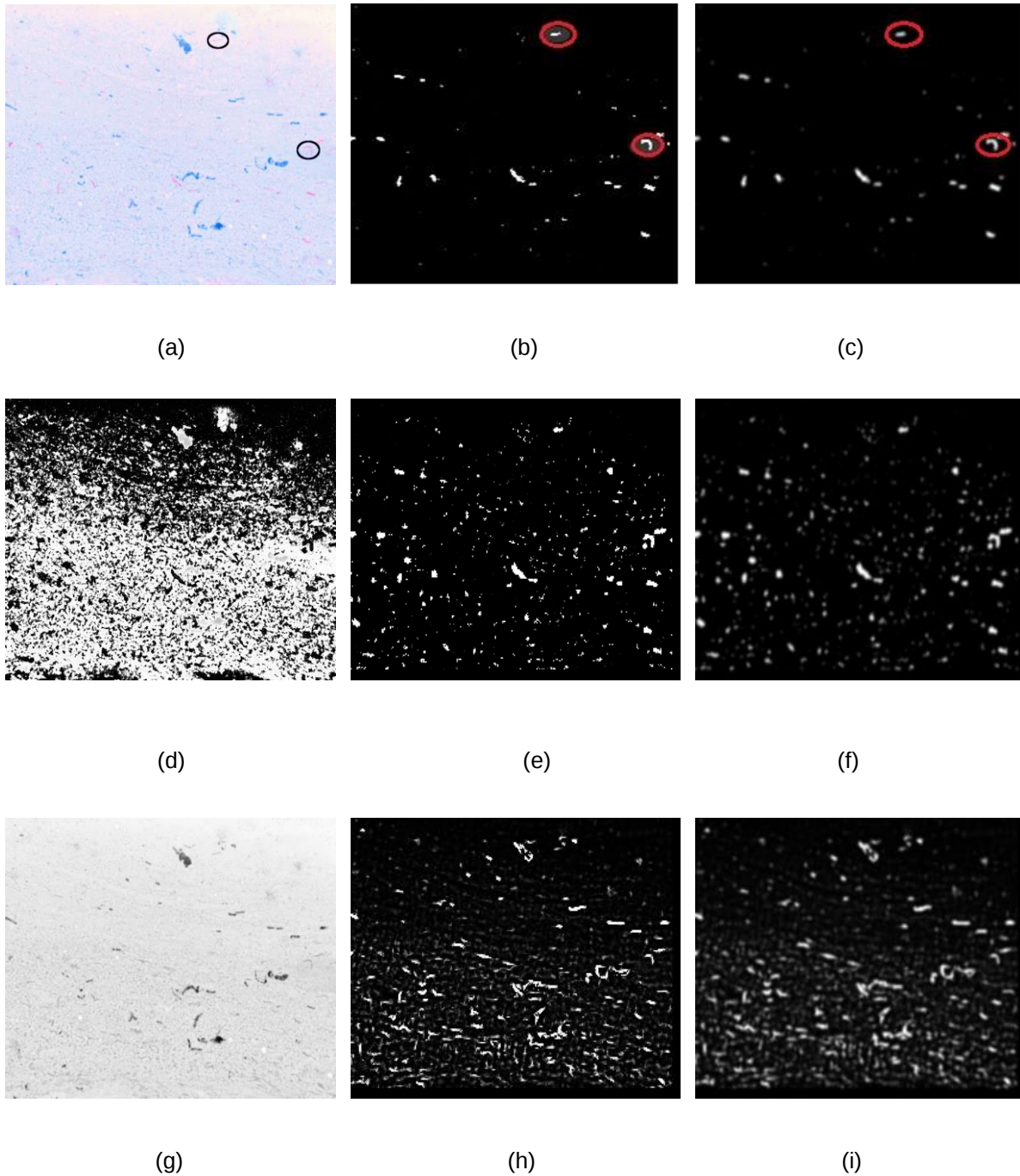


Figura 27 Em (a), (d) e (g) é mostrada uma imagem de baciloscopia de campo claro nos formatos: RGB, R-G e escala de cinza respectivamente. Em (b), (e) e (h) são apresentados os mapas de características das imagens mostradas em (a), (d) e (g), utilizando passo $s = 5$. Finalmente em (c), (f) e (i) é possível observar o efeito da aplicação do filtro gaussiano nos mapas de características mostrados em (b), (e) e (h).

5.2.3 *Non-maximum Suppression*

Conforme abordado na Secção 4.3.3.3, foram testados os valores iguais a 0,1, 0,3, 0,5, 0,7 e 0,9, como limiar de sobreposição. As taxas de acerto mais altas foram obtidas nas detecções com $Is = 0,3$. Portanto, como segunda etapa, foram testados os valores de Is iguais a 0,2, 0,25, 0,3, 0,35 e 0,4. O melhor desempenho foi obtido para $Is = 0,2$. Na Figura 28 apresenta-se uma imagem de baciloscopia resultante da aplicação do método proposto, utilizando-se os valores de Is iguais a 0,2, 0,5 e 0,9. Na Figura 28 (a) observa-se que, para o valor $Is = 0,2$, o algoritmo consegue detectar corretamente dois bacilos, coincidindo com a marcação feita pelos especialistas mostrada na Figura 28 (b). Na Figura 28 (c) mostra-se que, para $Is = 0,5$, o algoritmo detecta 4 bacilos, 2 em cada posição. Finalmente, na Figura 28 (d), utilizando um $Is = 0,9$, são detectados múltiplos bacilos em cada posição (aglomerado de pontos vermelhos).

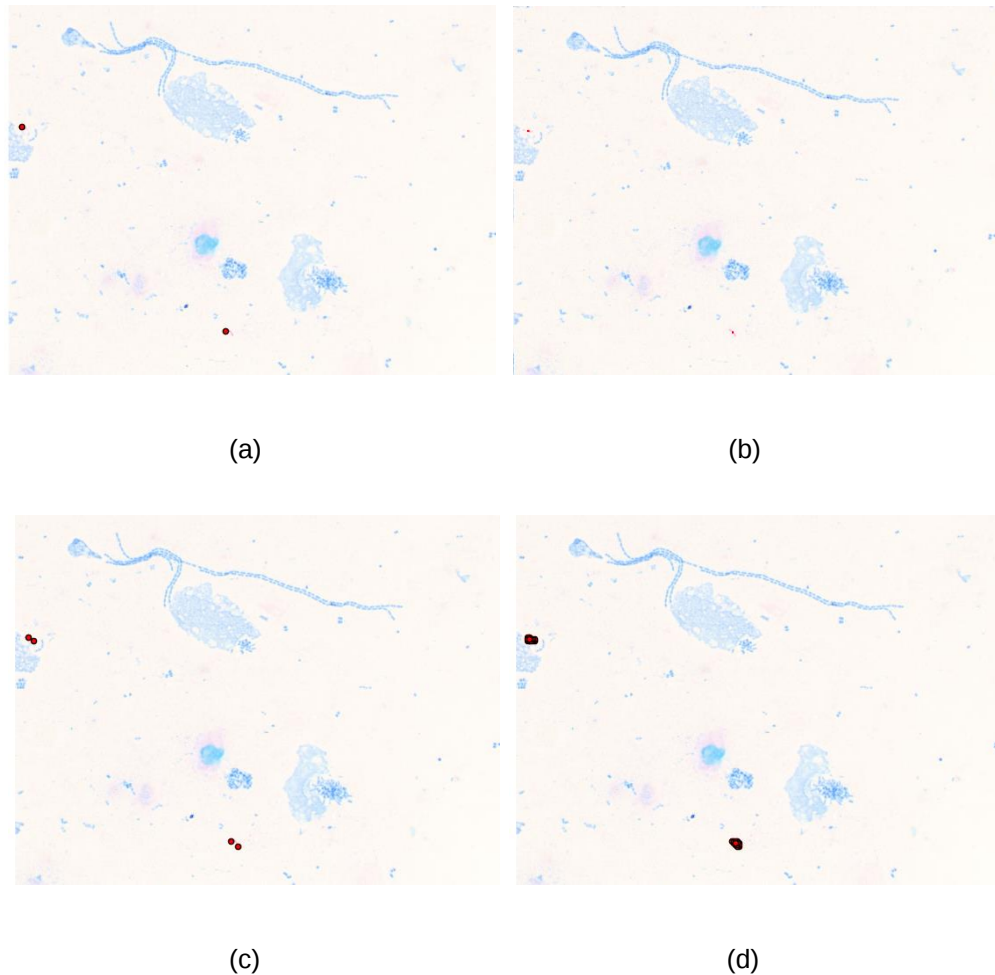


Figura 28 Imagem de baciloscopia de campo claro após a aplicação da metodologia proposta utilizando diferentes valores de limiar de sobreposição no algoritmo *Non-maximum Suppression*. Em (a), utilizando-se $Is = 0,2$, observa-se que as detecções coincidem com as marcações feitas pelos especialistas mostradas em (b). Em (c) apresenta-se as detecções obtidas utilizando-se $Is = 0,5$. Em (d) apresenta-se as detecções obtidas utilizando-se $Is = 0,9$. Cada ponto vermelho representa a detecção de um bacilo.

5.2.4 Avaliação de desempenho dos modelos para os diferentes tipos de entradas

Com o objetivo de avaliar o comportamento da metodologia implementada para cada tipo de entrada: RGB, R-G e escala de cinza, foram realizadas inicialmente 3 simulações diferentes, utilizando os classificadores de *patches* com melhores resultados obtidos na etapa anterior, os modelos 2 e 3, para um total de 6 simulações (3 tipos de

entrada * 2 classificadores). Nessas simulações, foram utilizadas 164 imagens de baciloscopia de campo claro, correspondentes ao paciente 2. Na Tabela 8 apresentam-se os detalhes dos parâmetros de detecção utilizados em cada simulação, assim como os resultados obtidos para cada combinação de formato de cor-classificador. Observe que os melhores resultados na detecção dos bacilos em função das métricas *Precision-Recall* e *F1-score*, foram obtidos com as entradas no formato de cor RGB utilizando o modelo 2. Por outro lado, os piores resultados foram obtidos utilizando o tipo de entrada em escala de cinza. Os valores de *Precision* obtidos foram muito baixos, o que significa que o número de falsos positivos nas detecções foi elevado.

É importante mencionar que existe uma margem de erro nas detecções feitas pelo método proposto com relação às marcações feitas pelos especialistas (*ground truth*). A margem de erro definida para aceitar uma detecção como sendo correta (VP) foi de 20 pixels. Na definição desse limiar, levamos em conta um dado experimental de que os bacilos, em média, podem ser envolvidos por um retângulo com dimensões de 23x22 pixels (Pinto, 2017).

Em consonância com os resultados obtidos, o modelo 2 foi assumido como o modelo a ser utilizado como classificador de *patches* e a imagem RGB como tipo a ser utilizado para representação das imagens de baciloscopia. Doravante, essa escolha será denominada de metodologia padrão.

Tabela 8 Configurações dos parâmetros de detecção e resultados obtidos nas 6 simulações para cada combinação formato de entrada-classificador.

Simulação	Classificador	Formato de cor	Parâmetros de detecção	Precision	Recall	F1
1	Modelo 2	RGB	Limiar de probabilidade (lp) = 0,5 Limiar de sobreposição (ls) = 0,2	56,52	86,15	68,47
2	Modelo 2	R-G	Limiar de probabilidade (lp) = 0,85 Limiar de sobreposição (ls) = 0,2	43,28	68,47	53,04
3	Modelo 2	Escala de Cinza	Limiar de probabilidade (lp) = 0,5 Limiar de sobreposição (ls) = 0,2	11,86	67,08	20,17
4	Modelo 3	RGB	Limiar de probabilidade (lp) = 0,5 Limiar de sobreposição (ls) = 0,2	40,9	82,9	54,79
5	Modelo 3	R-G	Limiar de probabilidade (lp) = 0,85 Limiar de sobreposição (ls) = 0,2	48,9	42,3	45,4
6	Modelo 3	Escala de Cinza	Limiar de probabilidade (lp) = 0,5 Limiar de sobreposição (ls) = 0,2	5,6	38,5	9,83

5.2.5 Análise dos resultados obtidos na etapa de detecção

As marcações dos bacilos nas imagens de baciloscopia feitas pelo algoritmo proposto foram analisadas detalhadamente. O estudo foi focado em identificar os casos

onde o algoritmo errou (falsos positivos e falsos negativos). É importante mencionar que o estudo foi realizado utilizando a metodologia padrão definida na Seção anterior.

Na sequência enumeram-se os casos principais onde observou-se a presença de falsos positivos FP nas detecções do algoritmo. Nas Figuras 29 a 34, os pontos verdes representam (FP), os pontos vermelhos representam acertos do algoritmo (VP) e as cruzes azuis são (FN):

1. **Áreas avermelhadas com forma e cor de bacilo.** Neste caso, o algoritmo identificou como sendo bacilos, áreas vermelhas com forma e cor de bacilo, mas que não foram marcadas como bacilos pelos especialistas. Na Figura 29 apresenta-se uma imagem de baciloscopia onde o algoritmo fez esse tipo de detecção. Visualmente, nestas detecções, o objeto tem a aparência de um bacilo.
2. **Bacilos muito compridos.** Bacilos com comprimento muito maior que o tamanho médio do MT geram detecções FP. Na Figura 30 apresenta-se um exemplo destas detecções (retângulo azul) em que o método detectou 2 bacilos em vez de 1.
3. **Ruído avermelhado nas imagens de baciloscopia.** Grande parte das imagens de baciloscopia utilizadas, no presente trabalho, apresentou um ruído avermelhado. Quando esse ruído tem forma alargada (forma de bacilo) o algoritmo detecta esses artefatos como sendo bacilos. Essas detecções constituem-se a maioria dos FP. A presença desses artefatos nas imagens de baciloscopia é um problema já identificado em trabalho prévio Costa *et al* (2012). Na Figura 31 é apresentado um exemplo de imagem com abundante ruído avermelhado e, por conseguinte, com alta taxa de FP.

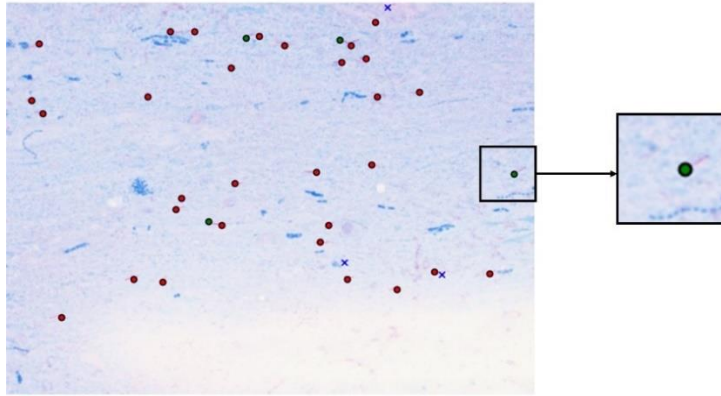


Figura 29 Exemplo de marcação FP (retângulo preto), onde o algoritmo identificou como sendo um bacilo um artefato com forma e cor de bacilo, mas que não foi marcado pelo especialista.

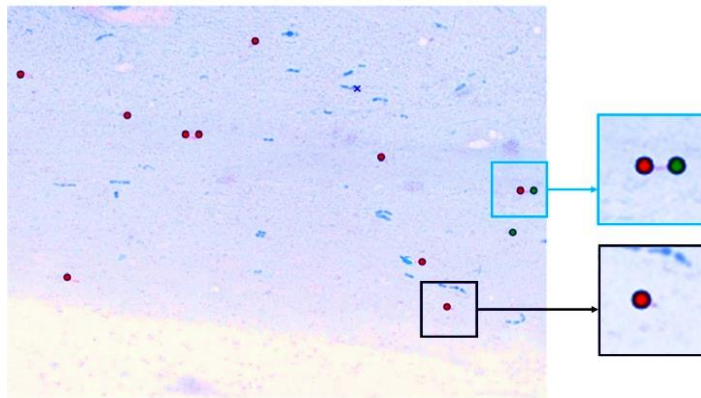


Figura 30 Exemplo de marcações FP, bacilo muito comprido (retângulo azul). Uma detecção correta (VP) é apresentada no retângulo preto.

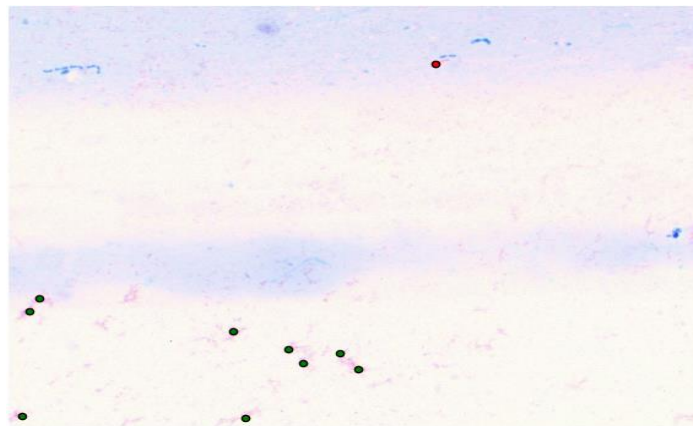


Figura 31 Exemplo de imagem de baciloscopia com ruído avermelhado em que o algoritmo apresenta vários FP representados como pontos verdes. Uma detecção VP foi marcada com um ponto vermelho.

O algoritmo também apresentou alguns casos de FN, comprometendo assim o *Recall*. Na continuidade, enumera-se os casos principais em que foram observados a presença de FN nas detecções do algoritmo:

1. **Bacilos gêmeos.** Um dos casos mais recorrentes de FN do algoritmo foi na hora de detectar bacilos muito próximos, aos quais chamamos de bacilos gêmeos. Nestes casos, o algoritmo deveria detectar dois objetos, mas somente consegue identificar um bacilo. Na Figura 32 apresenta-se um exemplo deste tipo de ocorrência.
2. **Bacilos localizados nas bordas das imagens de baciloscopia.** Quando os bacilos encontram-se nas bordas da imagem, o algoritmo não consegue identificar. Quinn *et al* (2016) também relataram o mesmo problema. Na Figura 33 apresenta-se um exemplo desse tipo de ocorrência.
3. **Bacilos com baixo contraste.** Nas imagens de baciloscopia existem bacilos que apresentam baixo contraste com respeito ao fundo da imagem, ocasionando FN nesses casos. Na Figura 34 apresenta-se um exemplo deste tipo de ocorrência.

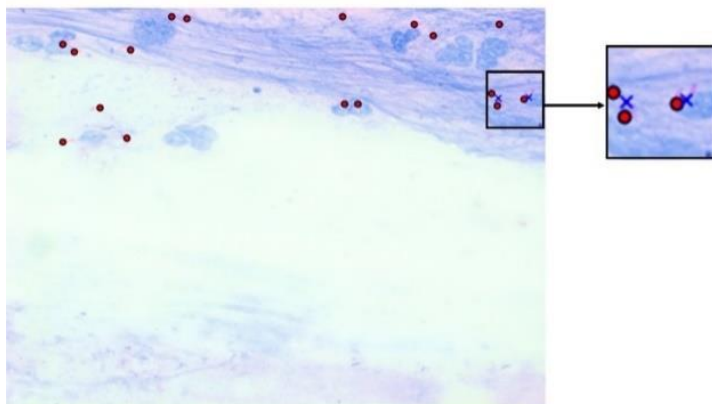


Figura 32 Um exemplo de ocorrência de bacilos gêmeos (retângulo preto). Observe na ampliação que o algoritmo identificou somente um bacilo (ponto vermelho), apresentando um FN (cruz azul).

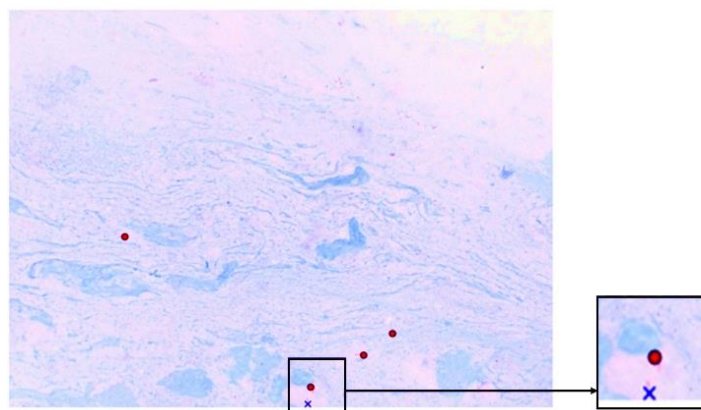


Figura 33 Um exemplo de bacilo não detectado pelo algoritmo (FN) na borda da imagem de baciloscopia (retângulo preto). A cruz azul indica que nessa posição o especialista marcou como existindo um bacilo, mas o algoritmo não conseguiu identificá-lo.

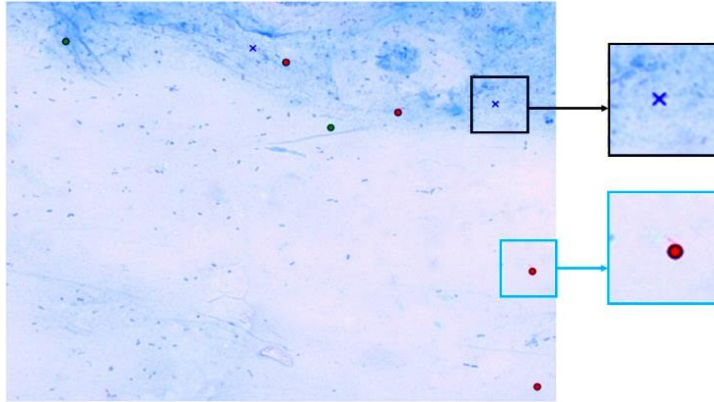


Figura 34 Exemplo de FN do algoritmo em um bacilo de baixa intensidade (retângulo preto). Observe na ampliação que o bacilo é quase imperceptível, quando comparando com o bacilo corretamente marcado pelo algoritmo (VP), retângulo azul.

Uma contribuição interessante do método proposto é a confirmação de casos positivos ou negativos em bacilos marcados como duvidosos pelo especialista. Nesses casos, a detecção feita pelo algoritmo pode ser utilizada para confirmar a presença ou ausência do bacilo. Na Figura 35 (a) apresenta-se um exemplo de marcação duvidosa do especialista delimitada por um triângulo preto, enquanto que na Figura 35 (b) são apresentadas as marcações feitas pelo algoritmo. Comparando as duas imagens é possível perceber que o algoritmo não identificou um bacilo na posição marcada como duvidosa pelo especialista, o que pode confirmar a ausência de bacilo naquela posição.

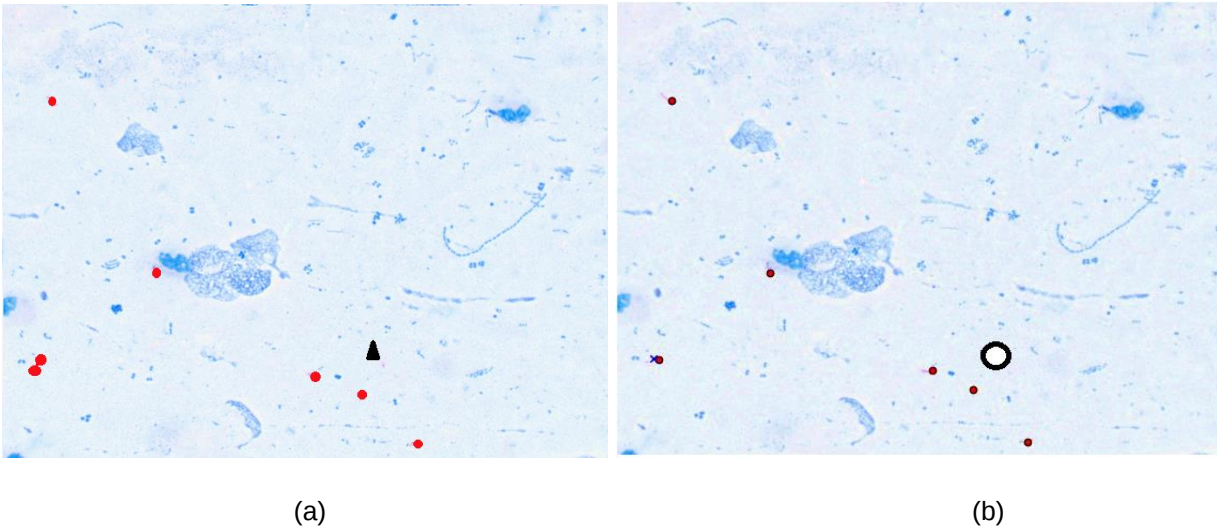


Figura 35 Em (a) apresenta-se uma imagem de baciloscopia de campo claro marcada por especialistas, em que os bacilos foram marcados com círculos vermelhos e bacilos duvidosos foram marcados com triângulos pretos. Em (b) apresenta-se a mesma imagem de baciloscopia com as marcações feitas pelo método proposto, em que os bacilos identificados foram marcados com pontos vermelhos. Observa-se em (b) que na posição onde os especialistas marcaram um bacilo duvidoso (marcada por um círculo preto) o algoritmo não detectou o objeto de interesse, o que pode significar que efetivamente não existe bacilo nessa posição.

5.2.6 Descarte automático de imagens com ruído vermelho

A presença de imagens com alto conteúdo de ruídos vermelhos no banco de imagens de baciloscopia usado, foi a principal causa de FP nas detecções do método proposto. Com o objetivo de melhorar o desempenho desse método, foram descartadas automaticamente da base de dados, as imagens que apresentaram uma porcentagem de pixels vermelhos maior do que um limiar de pixels vermelhos (*lpv*), igual a 4%. Foram excluídas 53 imagens de um total de 164. O *lpv*, foi definido de forma experimental tomando como base o proposto em Costa Filho *et al* (2012).

Com a aplicação do *lpv*, a *Precision* aumentou quase um 10%. Porém o *Recall* diminuiu aproximadamente em um 8%, devido a exclusão de imagens com alto conteúdo

de pixels vermelhos, além de imagens ruidosas. Também foram excluídas imagens com alto conteúdo de bacilos, nas quais o algoritmo tinha apresentado taxas de FN baixíssimas.

Tabela 9 Resultados obtidos utilizando as métricas *Precision*, *Recall* e *F1-score* na aplicação do método proposto com e sem descartar as imagens com ruído vermelho mediante a aplicação de um limiar de pixels vermelhos.

	<i>Precision (%)</i>	<i>Recall (%)</i>	<i>F1-score (%)</i>
Sem aplicar <i>lpv</i>	56,82	86,15	68,47
Aplicando <i>lpv</i>	65,76	78,50	71,56

6. CONCLUSÕES E TRABALHOS FUTUROS

Esta dissertação reflete um trabalho de pesquisa desenvolvido com o objetivo de propor um método de detecção automática do MT em imagens de baciloscopia de campo claro empregando Redes Neurais Convolutivas. Levando-se em consideração o estudo do estado da arte realizado, é possível ressaltar como contribuição da pesquisa a utilização de imagens de baciloscopia de foco estendido, pois até o momento nenhum outro autor empregou esse tipo de imagem na implementação de métodos visando a automatização da baciloscopia de campo claro. A base de imagens utilizada é robusta, balanceada e reflete as condições reais dos exames de baciloscopia de campo claro (presença de artefatos, problemas na preparação das lâminas, diferenças nas características das imagens, dentre outros).

Além disso, uma outra contribuição foi a construção de um conjunto de *patches* (fragmentos de imagens de baciloscopia de campo claro de foco estendido), o conjunto original CO. Também foi proposto um modelo baseado em RNC com duas camadas convolutivas incluindo técnicas de regularização, capaz de classificar um *patch* como sendo positivo ou negativo com uma AUC de 99%.

Outra contribuição importante do presente trabalho é que apresenta estatísticas dos resultados alcançados na detecção do MT em imagens completas de baciloscopia usando RNC. Segundo a interpretação do trabalho tomado como base nessa pesquisa, Quinn *et al* (2016), os autores somente apresentaram resultados quantitativos na classificação de *patches*. Resultados da aplicação de métricas na detecção do bacilo nas imagens de baciloscopia completas não foram identificados no trabalho citado.

A metodologia proposta, que utiliza um classificador de *patches* baseado em RNC em conjunto com o algoritmo NMS para detectar o MT em imagens de baciloscopia no formato RGB, apresentou um *F1-score* de 68,47%. O desempenho da metodologia é comprometido principalmente pela presença de grande quantidade de imagens com ruído avermelhado na base de dados. Esse tipo de imagem gera a maioria dos FP. Foi verificado o impacto desse problema nos resultados do método, mediante a exclusão automática de imagens com porcentagem de pixels vermelhos acima do 4%. Dessa forma o *Precision* aumentou de 56,82% para 65,76%.

Doravante, os esforços serão focados na geração de três bases de dados de *patches* com quase um milhão de amostras cada. A referida base de dados de *patches* será criada utilizando-se três conjuntos de imagens de baciloscopia de 25 pacientes: imagens de foco estendido, imagens da sexta profundidade do microscópio ótico e imagens de fusão obtidas em Pinto *et al* (2017). Posteriormente, a metodologia proposta no presente trabalho será aplicada nas novas bases de dados de *patches* e nas três bases de dados de imagens, com o objetivo de avaliar o comportamento do método nos três conjuntos referidos.

Um caminho adicional a ser explorado consiste na separação das imagens de baciloscopia em dois grupos: imagens regulares e imagens ruidosas, com o objetivo de combinar dois tipos diferentes de classificadores ou técnicas que permitam aprimorar os resultados na identificação dos bacilos nas imagens com ruídos vermelhos (principal causa de FP no método proposto).

Também é proposto trabalhar na análise por regiões nas imagens de baciloscopia, visando aplicar diferentes técnicas dependendo das características particulares de cada região.

REFERÊNCIAS

AMIT, Y.; FELZENSZWALB, P. **Object Detection**. Disponível em : <https://cs.uchicago.edu/directory/yali-amit> 2015. Acesso em 10/11/2016.

BRAGA, A. D. P.; LUDERMIR, T. B.; CARVALHO, A. P. D. L. **Redes Neurais Artificiais: Teoria e Aplicações**. Rio de Janeiro: LTC - Livros Técnicos e Científicos Editora S.A. Rio de Janeiro, 2000.

CAMPOS LEVI, P. **Reconhecimento e segmentação do mycobacterium tuberculosis em imagens de baciloscopia de campo claro utilizando as características de cor e o algoritmo Backpropagation**. 2012. Master's thesis at Federal University of Amazonas (UFAM)

CASTLEMAN, K. R.; YOUNG, I. T. **Microscope image processing**. Elsevier Inc, 2008.

CONTRIBUTORS, L. Lasagne library. 2014-2015. Disponível em: <http://lasagne.readthedocs.io/en/latest/user/installation.html>. Acesso em 18/03/2017.

COSTA FILHO, C. F. F. et al. **Automatic identification of tuberculosis mycobacterium**. Research on biomedical engineering., v. 31, 2015.

_____. **Mycobacterium Tuberculosis Recognition with Conventional Microscopy**. 34th Annual International Conference of the IEEE EMBS. San Diego, California USA, 2012.

COSTA, M. G. F. et al. **Automatic identification of mycobacterium tuberculosis with conventional light microscopy**. 30th Annual International IEEE EMBS Conference. Vancouver, British Columbia, Canada, 2008.

COSTA, M. G. F. R., L.V. et al. **Baciloscopia automatizada: uma revisão sobre os avanços tecnológicos**. XXV Congresso Brasileiro de Engenharia Biomédica – CBEB. Foz do Iguaçu-Paraná-Brasil, 2016.

GARCÍA, M. R. et al. **Detection of Mycobacterium tuberculosis in microscopic images of Ziehl-Neelsen-stained sputum smears**. (LACNEM 2015), 6th Latin-American Conference on Networked and Electronic Media, 2015.

GEURTS, P.; ERNST, D.; WEHENKEL, L. Extremely randomized trees. **Machine Learning**, v. Volume 63, 2006.

GLOROT, X.; BENGIO, Y. **Understanding the difficulty of training deep feedforward neural networks**. 13th International Conference on Artificial Intelligence and Statistics (AISTATS). Sardinia, Italy. 9 2010.

GONZALEZ, R. C.; WOODS, R. E. **Digital Image Processing**. Pearson Education, Inc, 2008.

GOODFELLOW, I. B., YOSHUA COURVILLE, AARON. **Deep Learning**: MIT Press 2016.

KAR, P. Convolutional Neural Networks (CNNs / ConvNets). 2015. Disponível em: <http://cs231n.github.io/convolutional-networks/>. Acesso em: 25/02/2017

KHUTLANG, R. et al. **Classification of Mycobacterium tuberculosis in Images of ZN-Stained Sputum Smears**. IEEE Transactions on information technology in biomedicine, vol. 14, no. 4, July 2010.

_____. **Detection of tuberculosis in sputum smear images using two one-class classifiers**. Biomedical Imaging: From Nano to Macro, 2009. ISBI '09. IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2009.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. **ImageNet classification with deep convolutional neural networks**. Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe, Nevada: Curran Associates Inc.: 1097-1105 p. 2012.

LECUN, Y.; BENGIO, Y.; HINTON, G. **Deep learning**. Nature. 521, 2015.

LI, X.; WANG, S. **Object Detection Using Convolutional Neural Networks in a Coarse-to-Fine Manner**. IEEE Geoscience and Remote Sensing Letters, 2017.

MAKKAPATI, V.; AGRAWAL, R.; ACHARYA, R. **Segmentation and Classification of Tuberculosis Bacilli from ZN-stained Sputum Smear Images**. 5th Annual IEEE Conference on Automation Science and Engineering. Bangalore, India, 2009.

MITCHELL, T. M. **Machine Learning**. McGraw-Hill Science/Engineering/Math; (March 1, 1997), 1997. ISBN 0070428077.

(MS) MINISTÉRIO DA SAÚDE. Secretaria de Vigilância em Saúde. **Tuberculose -Diagnóstico Laboratorial - Baciloscopia**, 2001.

_____. **O controle da tuberculose no Brasil: avanços, inovações e desafios**. Boletim Epidemiológico. v. 44 2014.

_____. **Perspectivas brasileiras para o fim da tuberculose como problema de saúde pública**. Boletim Epidemiológico. v. 47, n.13, 2016.

NAYAK, R.; SHENOY, V. P.; GALIGEKERE, R. R. **A New Algorithm for Automatic Assessment of the Degree of TB-infection Using Images of ZN-stained Sputum Smear**. Proceedings of 2010 International Conference on Systems in Medicine and Biology. Kharagpur, India, 2010.

OSMAN, M. K. et al. **A Genetic Algorithm-Neural Network Approach for Mycobacterium Tuberculosis Detection in Ziehl-Neelsen Stained Tissue Slide Images**. 10th International Conference on Intelligent Systems Design and Applications (ISDA), 2010a.

_____. **Segmentation of tuberculosis bacilli in ziehl-neelsen tissue slide images using hibrid multilayered perceptron network**. 10th International Conference on Information Sciences Signal Processing and their Applications (ISSPA), 2010b.

_____. **Segmentation of Tuberculosis Bacilli in Ziehl- Neelsen-Stained Tissue Images based on K-Mean Clustering Procedure.** International Conference on Intelligent and Advanced Systems (ICIAS), 2010c.

OSMAN, M. K.; MASHOR, M. Y. **Colour Image Segmentation of Tuberculosis Bacilli in Ziehl-Neelsen-Stained Tissue Images using Moving K-Mean Clustering Procedure.** Fourth Asia International Conference on Mathematical/Analytical Modelling and Computer Simulation, 2010.

_____. **Tuberculosis Bacilli Detection in Ziehl-Neelsen- Stained Tissue using Affine Moment Invariants and Extreme Learning Machine.** IEEE 7th International Colloquium on Signal Processing and its Applications: IEEE, 2011.

PAIS RAMOS, J. Fiocruz and Tuberculosis in Brazil. **Journal of Bacteriology & Mycology**, v. 2, p. 1, 2016.

PINTO, K. M. B. **Fusão de imagens multifocais e realce de nitidez: uma contribuição ao diagnóstico automático da tuberculose.** 2017. Master's thesis at Federal University of Amazonas (UFAM)

QUINN, J. et al. **Deep Convolutional Neural Networks for Microscopy-Based Point of Care Diagnostics.** Proceedings of International Conference on Machine Learning for Health Care. volume 56: 10 p. 2016.

QUINN, J. A. et al. **Automated blood smear analysis for mobile malaria diagnosis.** . Mobile Point-of-Care Monitors and Diagnostic Devices. : CRC Press 2014.

RAOF, R. A. A. et al. **Color Thresholding Method For Image Segmentation Algorithm Of Ziehl-Neelsen Sputum Slide Images.** 5th International Conference on Electrical Engineering, Computing Science and Automatic Control (CEE 2008) 2008.

ROSEBROCK, A. (Faster) Non-Maximum Suppression in Python., 2015. Blog Disponível em: <http://www.pyimagesearch.com/2015/02/16/faster-non-maximum-suppression-python/>. Acesso em: 17/05/2017.

RULANINGTYAS, R. et al. **Multi Patch Approach in K-Means Clustering Method for Color Image Segmentation in Pulmonary Tuberculosis Identification.** 4th International Conference on Instrumentation, Communications, Information Technology, and Biomedical Engineering (ICICI-BME). Bandung, 2015.

SEWELL, M. **Machine Learning** e-book, <http://machine-learning.martinsewell.com/>. Acesso em: 17/05/2017.

SOTAQUIRÁ, M.; RUEDA, L.; NARVAEZ, R. **Detection and quantification of bacilli and clusters present in sputum smear samples: a novel algorithm for pulmonary tuberculosis diagnosis.** International Conference on Digital Image Processing, 2009.

SRIVASTAVA, N. et al. **Improving neural networks by preventing co-adaptation of feature detectors.** arXiv:1207.0580 [cs.NE], 2012.

SUTSKEVER, I. et al. **On the importance of initialization and momentum in deep learning.** Proceedings of the 30 th International Conference on Machine Learning. Atlanta, Georgia, USA. 2013.

SÁNCHEZ, C. **Deep Learning for Identifying Malaria Parasites in Images**. 2015. Master's thesis at University of Edinburgh.

THEODORIDIS, S.; KOUTROUMBAS, K. **Pattern Recognition**. 3rd edition, San Diego: Elsevier Academic Press, 2006.

WANG, X. et al. **Droque detection for autonomous aerial refueling based on convolutional neural networks**. Chinese Journal of Aeronautics, 2016.

(WHO), WORLD HEALTH ORGANIZATION. **World health statistics 2015**. 2015.

_____. **Global Tuberculosis Report, Executive Summary**. 2017

APÊNDICE

As próximas páginas apresentam uma cópia de um artigo apresentando resultados parciais deste trabalho, intitulado “*Automatic classification of light field smear microscopy patches using Convolutional Neural Networks for identifying Mycobacterium tuberculosis.*”, apresentado na conferência IEEE Chilecon 2017, realizada na cidade de Pucón, Chile de 18 a 20 de outubro 2017. O artigo foi publicado na biblioteca digital *IEEE Xplore*.

Automatic classification of light field smear microscopy patches using Convolutional Neural Networks for identifying Mycobacterium Tuberculosis

Yadini Pérez López
Postgraduate Program in
Electrical Engineering

Federal University of Amazonas

Manaus, Brazil
yadini.perez@gmail.com

C. F. F Costa Filho
Postgraduate Program in
Electrical Engineering

Federal University of Amazonas

Manaus, Brazil
cfffilho@gmail.com

L. M. R. Aguilera
Computer Institute

Federal University of Amazonas

Manaus, Brazil
rojas@icom.ufam.edu.br

M. G. F. Costa
Postgraduate Program in
Electrical Engineering

Federal University of Amazonas

Manaus, Brazil
marlygfcosta@gmail.com

Abstract— Tuberculosis (TB) has been included among the top ten leading causes of death worldwide. Since 2008, several investigations have been developed by scientific community for automatic detection of mycobacterium tuberculosis (MT) in light field microscopy images. Those authors have applied techniques as digital images processing and machine learning, combined with different sizes of datasets and color spaces such as: RGB, HSI and Lab. More recently, deep learning has been used, working with grayscale, though. We present a method for automatic classification of light field smear microscopy patches using RGB, R-G and grayscale patches versions as inputs of a Convolutional Neural Networks (CNN) model for identifying MT. A dataset of negative and positive patches was created for training three CNN Models applying regularization techniques, like: dataset normalization, data augmentation and dropout. The best result in the patch classification test was reached using R-G input version and a three convolutional layers Model with regularization, getting an Area under ROC Curve of 99%.

Keywords— classification, identification, mycobacterium tuberculosis, convolutional neural network, deep learning

I. INTRODUCTION

Tuberculosis (TB) is a disease caused by a slow growing bacteria named mycobacterium tuberculosis (MT). TB has been included among the top 10 leading causes of death worldwide. More than 95% of TB deaths occurs in low and middle income countries [1]. Light field microscopy of sputum specimens is the most used medical exam to diagnose the disease in these countries, but requires expertise, including training, proficiency/competency assessments.

Since 2008, scientific community has developed several investigations for automatic detection of MT in light field smear microscopy images, aiming to support staff's work, avoiding visual fatigue and diagnostic errors. Many of the researches are from where TB disease is a serious health problem (South Africa, India, Malaysia, and Brazil) [2]. Researchers have used images datasets with different quantity of images exploring color spaces such as: RGB, HIS, and Lab. Some techniques have been applied for detecting MT: digital image processing techniques, machine learning techniques and more recently, deep learning techniques.

Many papers focused in the automatic detection of MT in light field smear microscopy images have been indexed in important databases like IEEE Xplore, Web of Science, and Scielo. To reach the referred objective the authors used one of three main approaches:

1. Identification of MT by segmentation applying different digital image processing techniques.
2. Identification of TB bacilli through classifiers based on machine learning techniques.
3. TB bacteria detection through deep learning techniques.

In [3], [4], [5], [6] and [7], authors utilized digital image processing techniques for bacilli segmentation using color information from different color spaces like: RGB in [3], [4] and [6], HSI in [5] and [7] and, HSV, YIQ, YCbCr and Lab in [6].

Around 2009 several investigations started to use machine learning techniques for MT detection: [8], [9], [10], [11], [12],

[13], [14], [15], [16], [17] and [18]. Dataset's sizes utilized by the different authors varies from 100 to approximately 1,900 samples. Different color spaces and morphologic characteristics were explored, as well as different kinds of classifiers were implemented on the mentioned works.

At 2016, Quinn et al. [19] focused on finding a Convolutional Neural Networks (CNN) model to identify the following pathogens: malaria bacteria, mycobacterium tuberculosis and intestinal parasites in blood, sputum smear and stool sample images, respectively. These diagnosis tests are performed by conventional optic microscopy. Images were captured by a smartphone coupled to a microscope. To identify the pathogens, the authors used the Non-Maximum Suppression algorithm associated with a classification model implemented by two convolutional layers. An accuracy of 99% for bacilli detection was reported. Nonetheless the best hit rates working with CNN been reported in literature using true color images as input, Quinn et al. worked with grayscale images, not exploring the color characteristics of bacilli. Additionally, although the training dataset was big enough, the testing dataset included just a 9% of positive samples.

This work presents a new method for automatic classification of light field smear microscopy patches images combining CNN with different color spaces as input, to identify MT. This combination permits exploring the color information of the bacilli.

II. MATERIALS AND METHODS

A. Patch image dataset

A patch image dataset named original dataset (OD) was created with 9,770 patches extracted from extended focus smear microscopy images dataset, named TBimages [20]. Patches images are 40x40 pixels regions. OD is balanced, i.e. 50% of patches with bacillus (positive patches-PP) and the other 50% without bacillus (negative patches-NP).

The PPs were created in a way to present a bacillus centered at the coordinates (20, 20). While the NPs were extracted in two different ways: by capturing (40x40) pixels regions from negative images and by capturing regions where there were only small fragments of bacilli, ensuring that full bacilli were not included.

Patches were extracted from extended depth-of-field (DOF) smear microscopy images. The extended DOF images were obtained combining images from 11 different focal planes on the same field of view. A total of 492 images were used. Each image has a spatial resolution of 1388x1040 pixels. Three experts labeled these images, identifying the MT, resulting in Labeled Extended DOF images. In addition, to each true color patch were generated two patches versions: grayscale and (R-G). Fig. 1 illustrate the procedure for creating the patches images Original Dataset from the full smear microscopy images.

From the OD an Augmented Dataset (AD) was obtained by applying rotations in two different directions (90 and 180 degrees) in each one of the 9,770 original patches, resulting in 29,310 patches (50% positive samples and 50% negative samples).

B. Convolutional Neural Network Models

Three different CNN architectures were implemented, generating 3 different Models as shown on Fig. 2. Model 1 was composed by one convolutional layer (CL), Model 2 by 2 CL and Model 3 by 3. The sigmoid activation function was used in the output layers, and Relu activation function in the convolutional layers. Convolutional layers use a stride $s = 1$ and $s = 2$. Zero-padding was not used, because preserving the spatial dimensions of the convolutional layer's output was not a priority. The weights w were initialized following a uniform distribution.

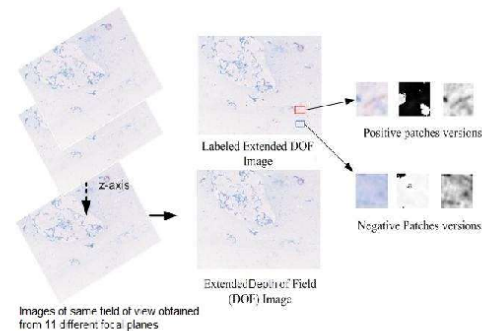


Fig. 1. Procedure for creating the Original Patches Images Dataset (OD) from a stack of smear microscopy images.

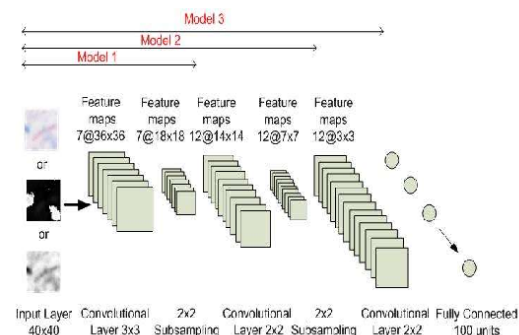


Fig. 2. Illustration of CNN architecture Models.

C. Object classification

The CNN models implemented were firstly trained without regularization techniques, and then, applying the following regularization techniques: dataset augmentation, dropout and utilization of small filters in the convolutional layers. Regularization techniques can potentiate the generalization level of artificial neural networks based models.

The CNN models were trained using three patches versions as mentioned before: RGB, R-G and grayscale. In total 18 simulations were implemented: (3 CNN Models * 3 input formats * 2 (with and without using regularization techniques)).

Fig. 3 illustrates the training configuration for the implemented models. As can see in this figure, in both cases, dataset's normalization was applied, once without this technique, CNN models were not able to learn in the training process.

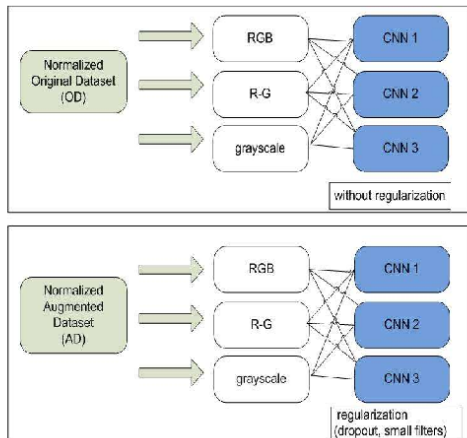


Fig. 3. Training configuration for the three convolutional neural network Models.

The CNN models were trained using the following optimization parameters: Medium Quadratic Error as *error function*, nesterov momentum as *weights update function*, with parameter *update_momentum*, starting by 0.9 and increasing up to 0.999, and *learning rate* starting at 0.03 and decreasing to 0.0001. Was defined 1,000 *epochs* for training.

III. RESULTS

A. Patch dataset

From sputum smear microscopy images, two sets of fully balanced patches data, OD and AD datasets, were designed aimed to evaluate the proposed model's performance.

B. Convolutional Neural Network models

About convolutional layers hyperparameters, it is important to highlight, that the variation of the stride hyperparameter, s , did not result in significant difference in model's performance. For example, the CNN Model 2 with R-G patches as input, had a test error of 0.031 and 0.050, for $s = 1$ and $s = 2$, respectively. Stride effect is illustrated in Fig. 4. In Fig. 4 (a) and (b) is possible to visualize, for positive patch, a bacillus-like pattern in the CL1 activations. Such pattern, on the other hand, is not observed for a negative patch, Fig. 4 (c).

C. Object classification

The use of the regularization techniques allowed each model to achieve an Area under the ROC Curve (AUC) superior to 96%.

The application of normalization technique to the inputs was a key factor in the training process, without its application, the

models were not able to learn adequately (AUC of just 50%). Normalizing the entries, a higher contrast is attained, and the object of interest, the bacillus, is easier to identify. See Fig. 5 for visualizing the normalization effect in the input patches.

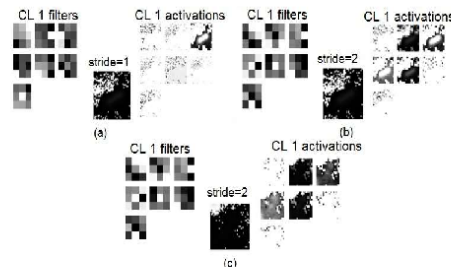


Fig. 4. Filters and activations of the first convolutional layer CL1: (a) using a positive patch input and stride $s = 1$; (b) using a positive patch input and stride $s = 2$, (c) using a negative patch input and stride $s = 2$.

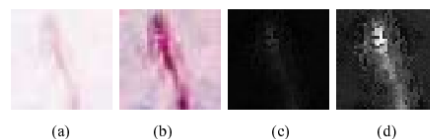


Fig. 5. Illustration of normalization effect in the input patches: (a) True color positive patch; (b) normalized true color positive patch shown in (a); (c) R-G positive patch; (d) normalized R-G positive patch shown in (c).

Table I demonstrates the contribution of other regularization techniques in order to reach lower error rates.

TABLE I. ERROR RATES COMPARISON OF CNN MODEL 2 HAVING R-G INPUTS WITHOUT APPLYING AND APPLYING DROPOUT AND DATA AUGMENTATION REGULARIZATION TECHNIQUES.

	Original Dataset	Augmented Dataset
No Dropout	0.07616	0.04879
Dropout	0.05788	0.03412

For selecting the best combination of CNN model - Input version a ROC analysis was performed. The ROC curve for every Model-Input version combination was obtained as seen in Fig. 6.

As expected, the best results were reached using the input versions that explored the color characteristics of the bacilli (R-G and RGB). In the R-G input version, all three methods obtained an accuracy superior to 97%. On the comparison of the models, the Model 3 has shown the best overall performance on the three input versions.

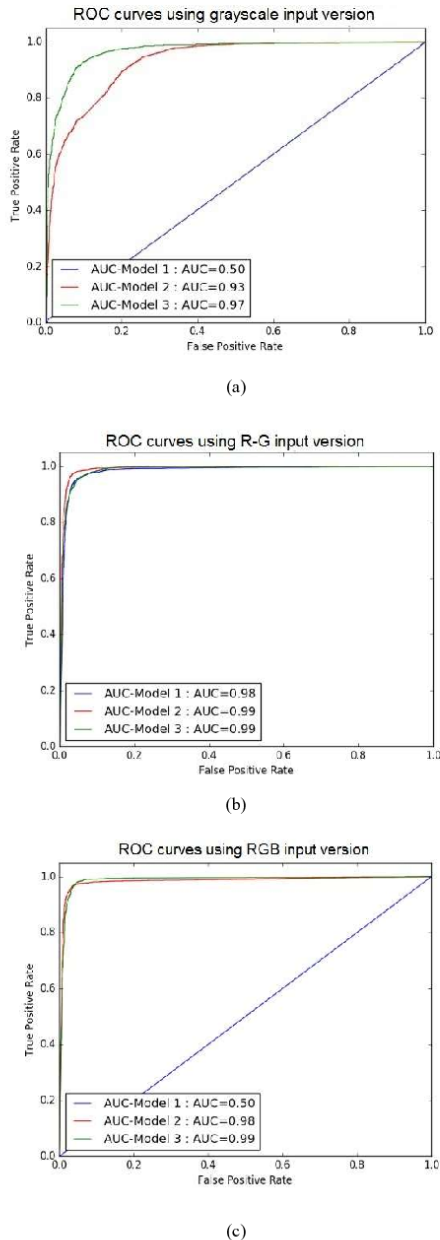


Fig. 6. ROC curves for CNN Models 1, 2 and 3 using three different patch versions: (a) grayscale, (b) R-G and (c) RGB.

IV. DISCUSSION

This work differs from the previous works developed in the literature by the following points:

- Utilization of focus extended smear microscopy images;
- Generation of a fully balanced (50/50, positive/negative) patches dataset;
- Training of CNN models using as input three different image versions: RGB, R-G and grayscale.

For comparison, the CNN Model with two convolutional layers proposed by Quinn et al [19] was implemented to do the patches classification task in our Original Patches dataset. An AUC equal to 86%, was obtained. In the other hand, our method reached an AUC of 99% for the same task.

V. CONCLUSIONS

This works presents a method for classifying light field smear microscopy patches as positive (presenting a TB bacillus) or negative (not presenting a TB bacillus). The proposed CNN model with three convolutional layers was evaluated with a robust balanced fusion image dataset and can classify a patch in positive or negative for tuberculosis bacillus presence with an accuracy of 99%. In the classifications task using three different input versions: RGB, R-G and grayscale, the worst and best results were reached using the grayscale, and the R-G format respectively. The proposed model can be used in sliding windows approach for detecting the parasite in full smear microscopy images.

REFERENCES

- [1] W. h. organization, "World health statistics 2015.," ed, 2015.
- [2] M. G. F. R. Costa, L. V., F. P. Romero, S. L. M. Silva, J. P. G. F. Costa, and C. F. F. Costa Filho, "Bacilosopia automatizada: uma revisão sobre os avanços tecnológicos," 2016.
- [3] M. G. F. Costa, C. F. F. Costa Filho, J. F. Sena, J. Salem, and M. O. de Lima, "Automatic identification of mycobacterium tuberculosis with conventional light microscopy.," 2008.
- [4] R. A. A. Raof et al., "Color Thresholding Method For Image Segmentation Algorithm Of Ziehl-Neelsen Sputum Slide Images.," presented at the 5th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE 2008) 2008.
- [5] V. Makkapati, R. Agrawal, and R. Acharya, "Segmentation and Classification of Tuberculosis Bacilli from ZN-stained Sputum Smear Images.," presented at the 5th Annual IEEE Conference on Automation Science and Engineering, Bangalore, India, 2009.
- [6] M. Sotaquirá, L. Rueda, and R. Narvaez, "Detection and quantification of bacilli and clusters present in sputum smear samples: a novel algorithm for pulmonary tuberculosis diagnosis," presented at the International Conference on Digital Image Processing, 2009.
- [7] R. Nayak, V. P. Shenoy, and R. R. Galigekere, "A New Algorithm for Automatic Assessment of the Degree of TB-infection Using Images of ZN-stained Sputum Smear.," presented at the Proceedings of 2010 International Conference on Systems in Medicine and Biology., Kharagpur, India, 2010.
- [8] R. Khutlang, S. Krishnan, A. Whitelaw, and T. S. Douglas, "Detection of tuberculosis in sputum smear images using two one-class classifiers.," ed, 2009.
- [9] R. Khutlang et al. (2010) Classification of Mycobacterium tuberculosis in Images of ZN-Stained Sputum Smears. IEEE.
- [10] M. K. Osman, F. Ahmad, Z. Saad, and M. Y. Mashor, "A Genetic Algorithm-Neural Network Approach for Mycobacterium

- Tuberculosis Detection in Ziehl-Neelsen Stained Tissue Slide Images.," ed, 2010.
- [11] M. K. Osman, M. Y. Mashor, Z. Saad, and H. Jaafar, "Segmentation of Tuberculosis Bacilli in Ziehl-Neelsen-Stained Tissue Images based on K-Mean Clustering Procedure.," ed, 2010.[12] M. K. Osman and M. Y. Mashor, "Colour Image Segmentation of Tuberculosis Bacilli in Ziehl-Neelsen-Stained Tissue Images using Moving K-Mean Clustering Procedure.," presented at the Fourth Asia International Conference on Mathematical/Analytical Modelling and Computer Simulation., 2010.
 - [13] M. K. Osman, M. Y. Mashor, H. Jaafar, and A. S. W. Wahab, "Segmentation of tuberculosis bacilli in ziehl-neelsen tissue slide images using hibrid multilayered perceptron network.," 2010.
 - [14] M. K. Osman and M. Y. Mashor, "Tuberculosis Bacilli Detection in Ziehl-Neelsen- Stained Tissue using Affine Moment Invariants and Extreme Learning Machine.," presented at the IEEE 7th International Colloquium on Signal Processing and its Applications, 2011.
 - [15] C. F. F. Costa Filho, P. C. Levy, C. M. Xavier, M. G. F. Costa, L. B. M. Fujimoto, and J. Salem, "Mycobacterium Tuberculosis Recognition with Conventional Microscopy.," presented at the 34th Annual International Conference of the IEEE EMBS, San Diego, California USA, 2012.
 - [16] M. R. García, A. Salazar, C.-A. Madrigal, L.-J. M. Guzman, and F. M. C. Mancera, "Detection of Mycobacterium tuberculosis in microscopic images of Ziehl-Neelsen-stained sputum smears.," ed, 2015.
 - [17] R. Rulaningtyas, A. B. Suksmono, T. Mengko, and P. Saptawati, "Multi Patch Approach in K-Means Clustering Method for Color Image Segmentation in Pulmonary Tuberculosis Identification.," presented at the 4th International Conference on Instrumentation, Communications, Information Technology, and Biomedical Engineering (ICICI-BME), Bandung, 2015.
 - [18] C. F. F. Costa Filho, P. C. Levy, C. d. M. Xavier, L. B. M. Fujimoto, and M. G. F. Costa, "Automatic identification of tuberculosis mycobacterium.," (in english), Research on biomedical engineering., vol. 31, 2015.
 - [19] J. Quinn, R. Nakasi, P. Mugagga, P. Byanyima, W. Lubega, and A. Andam, "Deep Convolutional Neural Networks for Microscopy-Based Point of Care Diagnostics.," 2016.
 - [20] M. G. F. Costa et al., "A sputum smear microscopy image database for automatic bacilli detection in conventional microscopy.," presented at the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, 2015.