

Detecção Automática de Macrófagos em Exame Parasitológico Utilizando Clusterização e Redes Neurais Convolucionais

Armando L. Borges¹, Viviane B. L. Dias¹, Clésio de A. Gonçalves^{2,3}, Lucas B. M. de Sousa^{1,2}, Daniel de A. Viana⁵, Ana Carolina L. Pacheco⁴, Romuere R. V. e Silva^{1,2}

¹Sistemas de Informação, Universidade Federal do Piauí, Picos, Brasil

²Ciência da Computação, Universidade Federal do Piauí, Picos, Brasil

³Departamento de Informática, Instituto Federal do Sertão Pernambucano Ouricuri, Brasil

⁴Ciências Biológicas, Universidade Federal do Piauí, Picos, Brasil

⁵Medicina Veterinária, Universidade Estadual do Ceará, Fortaleza, Brasil

{armandoluzborges, vivianebarbosa, clesio.goncalves}@ufpi.edu.br

{lucasbezerraufpi527, carolandim, romuere}@ufpi.edu.br

daniel.viana@uece.br

Abstract. *Visceral Leishmaniasis is a parasitic disease that affects the host's defense system, with dogs being its main urban reservoirs. Early diagnosis in animals is crucial to prevent transmission to humans. The parasitological examination is the gold standard for diagnosing the disease, which is repetitive and tiring work. This article presents a system for detecting and quantifying macrophages in medical images, currently assisting in diagnosis. The regions of interest were segmented using the K-Means clustering, and we used the Densenet201 architecture for detection. The methodology achieved 94.7% accuracy and 89.4% accuracy for the Kappa index. These results indicate the system's ability to aid diagnosis.*

Resumo. *A Leishmaniose Visceral é uma doença parasitária que afeta o sistema de defesa do hospedeiro, sendo os cães, seus principais reservatórios urbanos. A diagnose precoce em animais é crucial para evitar a transmissão para humanos. O padrão ouro para o diagnóstico da doença, o exame parasitológico, é um trabalho repetitivo e fadigante. Este artigo apresenta um sistema para detecção e quantificação de macrófagos em imagens médicas, visando auxiliar o diagnóstico. As regiões de interesse foram segmentadas utilizando o agrupador K-Means e na detecção foi utilizada a arquitetura DenseNet201. A metodologia atingiu 94,7% de Precisão, e 89,4% para o índice de Kappa. Isso indica a capacidade do sistema no auxílio do diagnóstico.*

1. Introdução

As Leishmanioses são um grupo de doenças parasitárias infecciosas negligenciadas, endêmicas em países tropicais e subtropicais, causadas por um protozoário do gênero *Leishmania*. A forma mais grave é a Leishmaniose Visceral (LV), que por se tratar de

uma zoonose, pode ser transmitida de animais para seres humanos através da picada de flebotomíneos fêmeas, popularmente conhecidos como mosquitos palha, sendo os cães, o principal reservatório parasitário urbano dessa doença [Ministério da Saúde 2006, Marcondes and Day 2019]. Nas Américas, a doença se encontra presente em pelo menos 12 países [PAHO 2017], sendo que 90% dos casos registrados na América Latina advém do Brasil [da Saúde]. Segundo o Ministério da Saúde, sua evolução é crônica e sistêmica, podendo levar ainda a óbito em 90% dos casos quando não tratada [Ministério da Saúde].

Levando em consideração o potencial risco à saúde humana, a identificação precoce da doença em caninos é fundamental para evitar a sua transmissão aos seres humanos, tendo em vista que os cães são animais domésticos. O padrão ouro para o diagnóstico da doença é o exame parasitológico do material aspirado da medula óssea, onde um especialista da área busca por formas amastigotas do parasito [Ministério da Saúde, Silva et al. 2014]. As leishmânias são parasitas intracelulares obrigatórios, quando inculadas na pele do hospedeiro pelos flebotomos, multiplicam-se nos fagócitos mononucleares do sistema mononuclear fagocítico [Silva 2007]. Neste contexto, a procura por macrófagos surge como uma potencial estratégia para agilizar o processo de diagnóstico da LV. As suas características morfológicas, como tamanho e forma, facilitam sua identificação em comparação com os parasitas de *Leishmania*, que apresentam dimensões menores e morfologia mais complexa. Apesar disto, a análise desses exames ainda pode ser muito difícil devido a grande quantidade de outras estruturas presentes no material biológico, que podem possuir características semelhantes às estruturas-alvo. Ademais, o grande volume de imagens geradas torna o processo longo e ainda mais complexo. Tais fatores, associados ao esforço e concentração exaustivos, necessários por parte do especialista, podem levar a incertezas e erros no diagnóstico final [e Silva et al. 2016].

Com foco neste contexto, é notável a recorrência, no cenário contemporâneo, da utilização de sistemas baseados em Inteligência Artificial (IA) para auxiliar diagnósticos e procedimentos na área da saúde. Estes sistemas possuem a capacidade de usar imagens médicas para auxiliar especialistas no diagnóstico de doenças nos mais diversos campos da medicina [Araújo et al. 2021]. Por meio da aplicação de métodos e técnicas de Visão Computacional (VC), são empregados esforços para extrair e analisar características significativas presentes nas imagens, visando identificar padrões que possam contribuir para a automação do processo de diagnóstico. Adicionalmente, podem ser empregadas técnicas de pré-processamento de imagens que modificam sua representação visual, com o intuito de realçar determinados elementos em detrimento de outros, facilitando assim, sua detecção e identificação. Essa abordagem sugere a viabilidade de desenvolver um sistema que integre tais conceitos, visando oferecer suporte a profissionais médicos e veterinários, potencializando a eficácia do diagnóstico.

Portanto, o objetivo deste artigo é apresentar o desenvolvimento de um sistema baseado em técnicas de VC e pré-processamento de imagens, capaz de identificar e quantificar macrófagos através de imagens do exame parasitológico, a fim de auxiliar os profissionais responsáveis no processo de diagnóstico da LV. Com isso, o restante do artigo está dividido da seguinte forma: A Seção 2 apresenta os trabalhos relacionados. A Seção 3 apresenta toda a metodologia utilizada no desenvolvimento do sistema. Na Seção 4 são mostrados os resultados obtidos, bem como métricas de desempenho. E por fim, na Seção 5 são realizadas algumas conclusões e são descritos brevemente os trabalhos futuros.

2. Trabalhos relacionados

Nesta seção serão apresentados estudos correlatos, que compartilham objetivos similares ao deste trabalho, analisando e discutindo eventuais limitações comuns a maioria dos métodos e técnicas empregados nas metodologias de cada um. Informações a respeito dos trabalhos relacionados podem ser observadas na Tabela 1.

Table 1. Informações a respeito das metodologias empregadas nos trabalhos relacionados.

| Autor | Base de dados | Nº de imagens | Proposta | Metologia |
|---------------------------|---|---------------|---|---|
| [Xia et al. 2009] | Privada | - | Segmentação e rotulagem de macrófagos | Distância Chamfer + segmentação watershed |
| [Liang et al. 2018] | <i>Blood Cell Count and Detection (BCCD)</i> | 364 | Classificação de células sanguíneas | CNNs + Redes Neurais Recursivas (RNNs) |
| [Bani-Hani et al. 2018] | BCCD | 364 | Classificação de leucócitos | CNNs + Algoritmos Genéticos |
| [Aziz et al. 2020] | <i>Munich AML Morphology Dataset</i> | 18.365 | Classificação de leucócitos | Segmentação baseada em Clusters + CNNs |
| [Eder et al. 2021] | Privada | 97 | Detecção de macrófagos | Mask R-CNN |
| [Trong Luong et al. 2022] | Acute Lymphoblastic Leukemia Image Database (ALL-IDB) | 1.438 | Classificação de células sanguíneas e células de leucemia | YOLOv5 |

Algumas lacunas identificadas nos trabalhos expostos na Tabela 1 remetem a adoção de arquiteturas de menor complexidade, como em [Liang et al. 2018], onde foram usados os modelos ResNet50, Xception e InceptionV3. Esta mesma escolha de redes também pôde ser observado em [Aziz et al. 2020], onde as redes AlexNet e ResNet18 foram usadas para classificar macrófagos em classes morfológicas. Além destes, o trabalho de [Eder et al. 2021] utilizou as redes ResNet50 e ResNet101 como base para detecção e segmentação de macrófagos. Em todos estes trabalhos, a decisão de utilizar arquiteturas mais simples, em comparação com redes mais profundas e complexas, impactou adversamente a eficácia das metodologias empregadas. Além disso, alguns trabalhos, como o de [Xia et al. 2009] e [Liang et al. 2018], usaram conjuntos de dados contendo imagens com tipos isolados de células, o que não reflete um contexto real, onde imagens deste tipo contém outras estruturas que podem influenciar as análises e inferências.

Para abordar estas limitações, o presente estudo propõe uma metodologia que integra arquiteturas mais avançadas e complexas, capazes de identificar e extrair um espectro mais amplo de padrões e características. Adicionalmente, emprega-se uma base de dados que representa um contexto mais próximo da realidade, e um conjunto de técnicas para a segmentação e extração dos principais componentes das imagens, a fim de minimizar a possibilidade de interferências de outras regiões nos resultados obtidos.

3. Metodologia

Nesta seção será apresentada a metodologia elaborada, bem como o detalhamento das técnicas e conceitos utilizados em sua construção.

Este trabalho utiliza o algoritmo *K-Means* para segmentação, em conjunto com a operação morfológica de abertura de área, e CNNs para detecção de macrófagos. Para isto, esta abordagem é composta de seis etapas: 1. Aquisição do conjunto de imagens; 2. Aplicação de técnicas de pré-processamento; 3. Segmentação e pós-processamento; 4. Balanceamento de classes; 5. Extração de atributos e classificação; 6. Avaliação dos resultados. Um diagrama contendo todos esses passos pode ser visualizado na Figura 1.

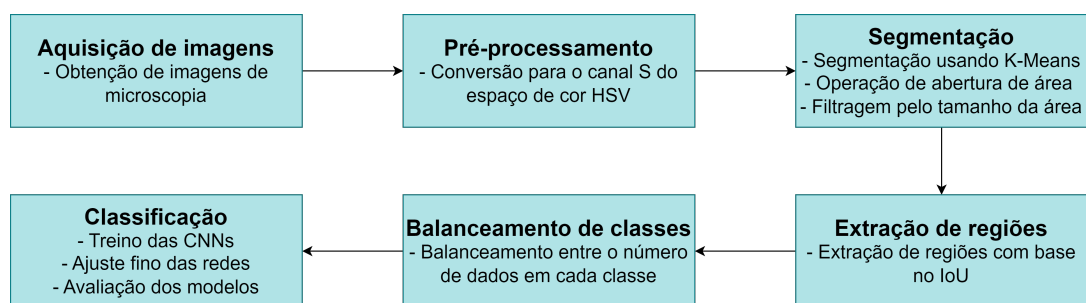


Figure 1. Diagrama representando as etapas da metodologia usada.

3.1. Aquisição do conjunto de imagens

Em colaboração com o Laboratório de Parasitologia, Ecologia e Doenças Negligenciadas (LAPEDONE), obtivemos um conjunto de 178 imagens de microscopia de material aspirado da medula óssea do fêmur de dois cães. Um especialista do LAPEDONE realizou a marcação manual de todo o conjunto de dados, fornecendo a localização dos macrófagos em cada imagem.



Figure 2. Etapa de aquisição de imagens de microscopia.

3.2. Pré-processamento

Tendo em vista que o sistema desenvolvido neste trabalho utiliza segmentação baseada em agrupamentos como uma das primeiras etapas para a detecção de macrófagos, foi necessário realizar um pré-processamento nas imagens para melhorar a separação das regiões de interesse do restante da imagem. Com isso, foi adotada uma abordagem que converte o espaço de cor original da imagem (RGB) para o espaço de cor HSV. Sendo assim, é pertinente destacar que apenas o canal de saturação desse esquema de cores, canal S, foi utilizado, visto que seu uso destacava os macrófagos do fundo da imagem. Com isto, o algoritmo de agrupamento é favorecido a segregar tais regiões em um único plano (*Cluster*). Este processo está representado na Figura 3.

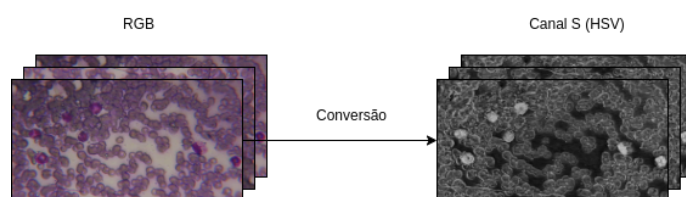


Figure 3. Etapa de pré-processamento. Conversão do espaço de cor original (RGB) para o canal S do espaço de cor HSV.

3.3. Segmentação e pós-processamento

Nesta etapa do trabalho foi utilizado o *K-Means*, um algoritmo clássico de segmentação baseada em *clusters*. De forma geral, ele é um algoritmo de agrupamento não supervisionado que divide um conjunto de dados em k grupos, com o objetivo de agrupar elementos semelhantes e segregá-los dos demais. Neste contexto, é possível observar na Figura 3 que após o pré-processamento, as imagens apresentam predominantemente três cores: preto, cinza e uma nuance próxima ao branco. Portanto, com base nesta informação, foi definido que $k = 3$. Sendo assim, foi possível segregar as regiões mais claras, correspondentes as regiões dos macrófagos, utilizando o *cluster* com maior valor, extraindo assim uma máscara de segmentação contendo a localização das estruturas de interesse.

Após a segmentação, ainda restaram estruturas finas e conectadas entre si nas máscaras obtidas. Além disso, elas também se encontravam conectadas as regiões de interesse, sendo necessário aplicar uma etapa de pós-processamento nas imagens segmentadas. Neste contexto, a primeira abordagem adotada consistiu na aplicação de uma operação de abertura morfológica para separar essas regiões dos macrófagos. Para tal, empregou-se um elemento estruturante configurado como uma matriz quadrada de dimensões 5×5 . Essa operação é composta, sequencialmente, por uma operação de erosão seguida de uma de dilatação. Esta abordagem visa aprimorar a segmentação dos macrófagos, minimizando a influência de estruturas adjacentes indesejadas.

Adicionalmente, outra etapa de pós-processamento foi implementada para eliminar os ruídos remanescentes. Para tanto, um limiar de área foi estabelecido com base nos valores extremos da área dos macrófagos. Essa estratégia visou remover ao máximo os objetos de tamanho incompatível com sua morfologia, preservando as estruturas segmentadas de interesse. Sendo assim, cada região passou por um filtro onde a sua área deveria corresponder a seguinte relação: $2.000 \leq x \leq 20.000$, onde x corresponde a área em pixels da região. Os processos realizados nesta etapa podem ser visualizados na Figura 4.

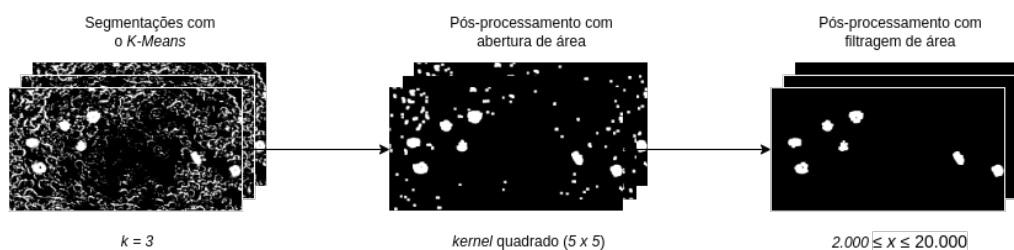


Figure 4. Etapa de segmentação e pós-processamento. As segmentações geradas pelo *K-Means* são pós-processadas com a aplicação das operações de abertura e filtragem de área.

Como é possível observar na figura 5, a partir das segmentações, recortes foram gerados e diferenciados entre macrófagos e não macrófagos, com base no cálculo do IoU (*Intersect Over Union*) entre as caixas delimitadoras fornecidas pelo especialista e as extraídas das propriedades de região. Neste contexto, foi possível contabilizar a quantidade de elementos pertencentes a cada classe. Assim, identificou-se que, do total de 137 macrófagos presentes na base de dados, 124 foram corretamente segmentados. Adicionalmente, observou-se também a identificação de 1698 regiões que pertenciam a classe dos não macrófagos. As imagens obtidas através deste processo foram subsequentemente utilizadas para compor o conjunto de dados empregado no treinamento das CNNs.

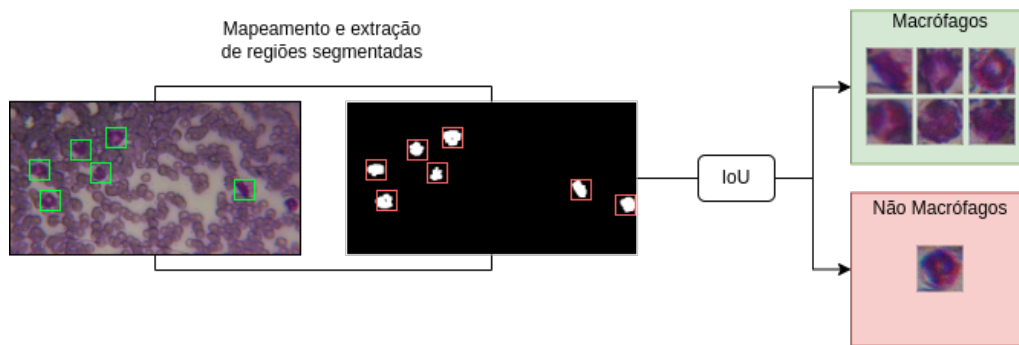


Figure 5. Processo de mapeamento e extração das regiões segmentadas. As anotações do especialista estão delineadas em polígonos verdes, e as extraídas das propriedades de região estão destacadas em vermelho.

3.4. Balanceamento de classes

Tendo em vista o desbalanceamento entre as classes dos macrófagos e dos não macrófagos, foram aplicadas técnicas de balanceamento de classes a fim de equilibrá-las, além de aumentar a variedade de seus dados, potencializando com isso, os resultados do treino das CNNs. Tais técnicas correspondem a um conjunto de operações que geram novos dados sintéticos a partir de dados originais.

Neste contexto, as operações utilizadas neste trabalho foram a rotação aleatória dentro de um intervalo de 0° a 90° , zoom aleatório de até 30%, ruído gaussiano e ISO, e desfoque gaussiano. Elas foram selecionadas devido a sua aplicação não distorcer as estruturas das regiões. O uso da rotação fará com que as redes analisem os objetos de interesse a partir de diversos ângulos. O zoom aumenta o tamanho dos objetos em comparação com o tamanho da imagem, ajudando as redes a se adaptarem a diferenças nos tamanhos das células. A aplicação de diferentes tipos de ruídos melhora a capacidade de classificar corretamente os objetos de interesse. Por fim, o desfoque ajuda as redes a identificarem melhor os objetos de interesse fora do campo de foco do microscópio. Vale mencionar que essas operações foram aplicadas de forma individual e de forma combinada, aleatoriamente. O processo descrito nesta etapa pode ser visualizado na Figura 6.

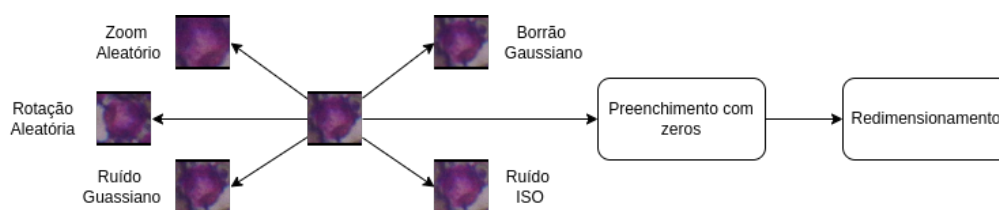


Figure 6. Etapa de balanceamento de classes. Após isso, as dimensões dos recortes são preenchidas com zeros para se igualarem. Posteriormente são redimensionadas para o tamanho de entrada padrão de cada CNN.

3.5. Extração de atributos e classificação

Neste trabalho, a extração de características e classificação das imagens foi realizada através de CNNs, pois possuem uma alta capacidade de aprendizado, devido aos milhares de parâmetros que possuem. No entanto, treinar uma arquitetura de CNN do zero

requer um grande poder computacional e uma grande quantidade de dados. Para compensar estes requisitos, foi utilizada uma técnica denominada Transferência de Aprendizado nas camadas de extração de atributos. O uso desta técnica permite adicionar a esta camada os descritores de características obtidos na resolução de outro problema semelhante [Cook 2017]. Com isso, é possível aproveitar os detectores de atributos utilizados em outra tarefa para tornar o treino mais eficiente.

Para este trabalho, foram utilizadas as arquiteturas DenseNet201 (DN201) [Huang et al. 2017], VGG16 (VGG16) [Simonyan and Zisserman 2015], ResNet152V2 (RN152V2) [He et al. 2016] e InceptionResNetV2 (IRNV2) [Szegedy et al. 2017]. Sua escolha se deu tanto pela eficácia e precisão na classificação de objetos do conjunto de dados ImageNet [Russakovsky et al. 2015], quanto pelo desempenho em outros trabalhos semelhantes de análise de imagens médicas [Gonçalves et al. 2023]. Neste contexto, durante o treino, as camadas convolucionais das redes são congeladas para evitar alterações nos descritores de atributos, mantendo inalterados os recursos obtidos com transferência de aprendizado durante o ajuste da camada totalmente conectada.

Também foi aplicada a técnica de Ajuste Fino das redes, em que as últimas camadas convolucionais das redes foram descongeladas. Esta técnica melhora o desempenho da rede na medida em que seus descritores se adaptam aos padrões do conjunto de dados. Assim, é possível aumentar a capacidade dos extratores de características de capturar informações mais relevantes e que permitam uma melhor identificação dos objetos.

Vale evidenciar que os parâmetros utilizados durante o treino impactam diretamente no desempenho das redes. Neste contexto, os parâmetros utilizados neste trabalho foram 30 épocas de treino da camada totalmente conectada, com taxa de aprendizagem de 0,01 e fator de decaimento de 0,1. Na etapa de Ajuste Fino, com as últimas camadas da rede descongeladas, o modelo foi treinado por mais 60 épocas com taxa de aprendizagem de 0,001. A taxa de aprendizado menor se justifica pela sensibilidade das camadas descongeladas aos ajustes, portanto, uma taxa menor levará o modelo a convergir com mais eficiência. Além disso, também foi utilizado *dropout* de 20% para evitar sobreajuste aos dados de treino. Ademais, esta fase pode ser visualizada na Figura 7.

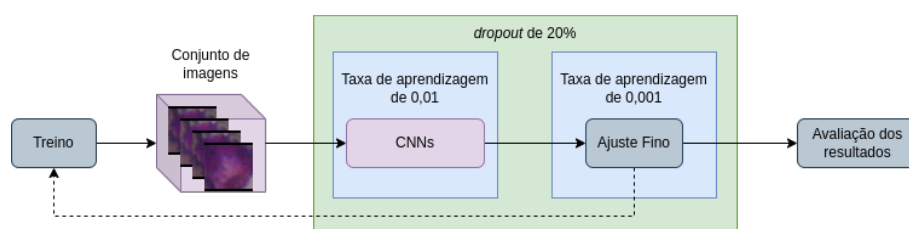


Figure 7. Processo de treino das arquiteturas de CNNs selecionadas. Inicialmente as redes são treinadas por 30 épocas com as camadas convolucionais congeladas. Posteriormente são treinadas por mais 60 épocas com essas camadas descongeladas e uma taxa de aprendizagem menor.

3.6. Avaliação dos resultados

O desempenho dos modelos foi avaliado a partir de seis métricas baseadas na matriz de confusão, sendo elas: Entropia Cruzada Binária, Acurácia, Precisão, *Recall*, F1, e índice de Kappa [Fleiss et al. 2013]. Vale destacar que a matriz de confusão é baseada em outras

quatro métricas, ou seja, ela é gerada através do número de verdadeiros positivos e negativos, que representam inferências corretas para as classes positivas e negativas, respectivamente. Da mesma forma, também há o número de falsos positivos e negativos, que representam inferências incorretas para as classes negativa e positiva, respectivamente.

Neste contexto, as métricas mencionadas são calculadas com base no conjunto de teste. É importante destacar que a desproporção significativa entre as classes pode afetar a precisão das métricas de avaliação. Para mitigar esse viés, adotou-se a técnica de *downsampling*, que envolve a remoção aleatória de amostras da classe majoritária. Esta abordagem visa equilibrar o número de instâncias entre as classes, proporcionando um cálculo mais justo e representativo das métricas de desempenho para ambas as classes.

4. Resultados e discussões

Os resultados obtidos na etapa de segmentação podem ser visualizados na Tabela 2. Neste trabalho, quatro arquiteturas de CNNs foram treinadas com a mesma metodologia e base de dados resultante da etapa anterior. Os parâmetros de treino foram uniformizados para todas as arquiteturas, e todas as redes foram avaliadas sob a mesma ótica. Portanto, as métricas de avaliação foram usadas para fins de análise e comparação de desempenho entre os modelos. Os resultados estão detalhadas na Tabela 3.

Table 2. Resultados obtidos após a etapa de segmentação.

| Animal | Nº de imagens | Total de macrófagos | Macrófagos segmentados | Não macrófagos segmentados |
|--------|---------------|---------------------|------------------------|----------------------------|
| 02 | 57 | 92 | 80 | 110 |
| 11 | 121 | 45 | 44 | 2588 |
| Total | 178 | 137 | 124 | 2698 |

Ao analisar a Tabela 2, observa-se que a segmentação foi eficaz na identificação dos macrófagos. O algoritmo *K-Means* segmentou cerca de 90% de todos os macrófagos da base de dados. Alguns deles não foram identificados devido a variações nas tonalidades das imagens, resultando na separação de suas regiões em um *cluster* distinto do comum a maioria. Este fator também explica a separação incorreta de estruturas que não são macrófagos, agrupadas juntamente com os macrófagos e submetidas aos mesmos procedimentos de pós-processamento e classificação.

Table 3. Resultados obtidos no treino das arquiteturas de CNNs selecionadas. Os melhores valores estão destacados em negrito.

| Arquitetura | Entropia Cruzada Binária | Acurácia | Precisão | Recall | F1 | Kappa |
|-------------|--------------------------|--------------|--------------|--------------|--------------|--------------|
| DN201 | 0,172 | 0,947 | 0,947 | 0,947 | 0,947 | 0,894 |
| VGG16 | 0,556 | 0,842 | 0,933 | 0,736 | 0,823 | 0,684 |
| RN152V2 | 0,146 | 0,921 | 0,944 | 0,894 | 0,918 | 0,842 |
| IRNV2 | 0,267 | 0,894 | 0,941 | 0,842 | 0,888 | 0,789 |

A Tabela 3 mostra os resultados obtidos com o treino das CNNs para classificar as regiões resultantes da etapa de segmentação. A arquitetura com melhor performance foi a DN201. Este modelo conseguiu alcançar Acurácia, Precisão, *Recall* e F1 de 0,947, além de uma pontuação de 0,894 para o índice de Kappa, indicando uma concordância quase perfeita. No entanto, a Entropia Cruzada Binária teve um valor melhor para a RN152V2, mesmo com a DN201 se saindo melhor nas demais métricas de desempenho. Isso pode ser

explicado ao se observar a profundidade de cada uma delas. O melhor modelo possui uma quantidade significativa de camadas convolucionais a mais, aproximadamente 95, proporcionando maior capacidade de capturar as nuances dos dados e obter assim, um melhor desempenho geral. Além disso, sua arquitetura permite que os mapas de atributos das camadas anteriores sejam usados como entradas em camadas subsequentes, mediando o problema da fuga do gradiente e garantindo um melhor desempenho [Huang et al. 2017].

A Figura 8 apresenta um exemplo onde o *K-Means* falhou em segmentar uma região anotada pelo especialista. Nesta imagem, as anotações do especialista são delimitadas por polígonos verdes, enquanto as detecções feitas pelo sistema são destacadas em vermelho. Observa-se que a região não detectada possui tonalidade mais fraca em comparação às demais anotações, o que contribuiu para sua não detecção, uma vez que foi separada em outro *cluster*. Em contrapartida, as outras duas regiões, caracterizadas por uma tonalidade mais intensa, foram adequadamente segmentadas e classificadas pelo sistema, alinhando-se com as marcações realizadas pelo profissional.

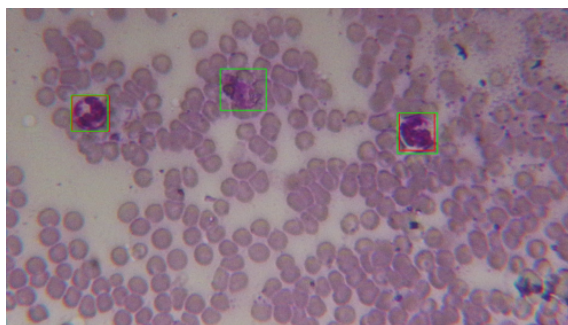


Figure 8. Detecção usando o sistema proposto. Os polígonos verdes representam as marcações do especialista, enquanto os vermelhos representam as detecções feitas pelo sistema.

Para validar o experimento, foi utilizado o *Grad-Cam* [Selvaraju et al. 2017], uma técnica que permite visualizar as zonas de ativação que influenciam a inferência das redes, assim, é possível verificar quais partes da imagem mais influenciaram no processo decisório dos modelos. Observa-se na Figura 9, que duas cores predominam os mapas de calor. As regiões mais claras, em tons de amarelo, representam as partes mais importantes consideradas pelas redes para a classificação, já as mais escuras, em tons de vermelho, representam regiões menos importantes. Sendo assim, é possível notar que a DN201 analisa as principais regiões de cada imagem, considerando-as como as partes mais relevantes para sua predição. Portanto, afirma-se que os resultados das avaliações são confiáveis, indicando que o modelo generalizou corretamente os dados de treino.

A Tabela 4 apresenta uma comparação entre a abordagem proposta neste trabalho e as empregadas nos trabalhos relacionados. O uso amplo de CNNs é algo notório, porém, as arquiteturas utilizadas nos trabalhos correlatos tem menor complexidade e número de parâmetros. Assim, elas possuem menor capacidade de detectar e extrair atributos importantes para a classificação. Isso pode ser observado na Tabela 3, onde o modelo mais simples, a VGG16, teve menor desempenho comparado aos demais. Com isso, a adoção de modelos mais complexos, com grande capacidade de aprendizado, pode trazer melhores resultados.

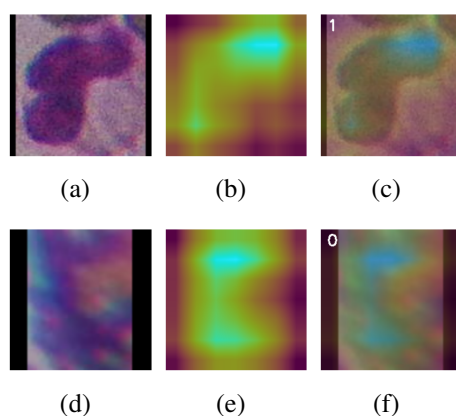


Figure 9. Zonas de ativação da última camada convolucional da rede DN201. As imagens (a) e (d) correspondem as imagens originais da classe positiva e negativa, respectivamente. As imagens (b) e (e) representam os mapas de calor gerados pelo *Grad-Cam*. Já as imagens (c) e (f) apresentam a sobreposição dos mapas de calor com as imagens originais.

Table 4. Comparação da metodologia proposta com os trabalhos relacionados. Os melhores resultados encontram-se destacados em negrito.

| Método | Acurácia | Precisão | Recall | F1 | índice de Kappa |
|---------------------------|--------------|--------------|--------------|--------------|-----------------|
| [Eder et al. 2021] | 0,935 | - | - | - | - |
| [Liang et al. 2018] | 0,908 | - | - | - | - |
| [Bani-Hani et al. 2018] | 0,910 | - | - | - | - |
| [Trong Luong et al. 2022] | 0,746 | 0,898 | 0,603 | - | - |
| [Aziz et al. 2020] | 0,933 | 0,854 | 0,914 | 0,876 | - |
| Proposto | 0,947 | 0,947 | 0,947 | 0,947 | 0,894 |

Este estudo fornece uma metodologia promissora, alcançando resultados encorajadores na aplicação de CNNs para a análise de imagens médicas. Contudo, é importante reconhecer certas limitações que podem influenciar a generalização dos resultados obtidos. Inicialmente, o volume de imagens empregadas no treinamento das redes não corresponde plenamente a diversidade e complexidade encontradas em cenários reais, podendo resultar na omissão de padrões relevantes não representados na base de dados atual. Além disso, a predominância de não macrófagos identificados durante a fase de segmentação introduz um desequilíbrio entre as classes analisadas neste estudo, afetando por consequência a precisão dos modelos. Embora a detecção automatizada de macrófagos contribua significativamente para a redução da carga de trabalho manual, ainda é necessário que os especialistas complementem o diagnóstico analisando todas as regiões detectadas.

5. Conclusão e trabalhos futuros

A metodologia empregada se mostrou robusta para a análise de imagens médicas do exame parasitológico. A utilização do canal S do espaço de cor HSV permitiu que o *K-Means* segmentasse os objetos de interesse com mais eficácia. As etapas de abertura e filtragem de área melhoraram significativamente as segmentações geradas, ao passo em que removeram boa parte dos ruídos. Apesar disto, observou-se ainda um número relevante de não macrófagos segmentados, o que contribuiu para o desbalanceamento das

classes. Ademais, a adoção de arquiteturas mais complexas e profundas resultou em melhorias notáveis nos resultados quando comparado aos trabalhos relacionados, mesmo utilizando uma base de dados que reflete melhor uma situação real, evidenciando uma possível solução para as lacunas encontradas em suas abordagens.

Em trabalhos futuros, pretende-se expandir o conjunto de dados, testar a capacidade e eficácia da metodologia proposta, visando simular um contexto mais próximo da realidade, e adotar métodos mais eficientes de segmentação. Além disso, o sistema construído neste trabalho servirá de base para um sistema maior e mais completo, onde além de detectar macrófagos, também será verificado se são saudáveis ou não quanto a presença de parasitas de LV. Ademais, também fica a aberta a possibilidade de implementar uma abordagem para procurar parasitas próximos as células do sistema mononuclear fagocitário, além de medir o grau de infecção da doença.

References

- Araújo, F. H., Silva, R. R., Medeiros, F. N., Neto, J. F. R., Oliveira, P. H. C., Bianchi, A. G. C., and Ushizima, D. (2021). Active contours for overlapping cervical cell segmentation. *International Journal of Biomedical Engineering and Technology*, 35(1):70–92.
- Aziz, S., Bilal, M., Khan, M. U., and Amjad, F. (2020). Deep learning-based automatic morphological classification of leukocytes using blood smears. In *2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)*, pages 1–5.
- Bani-Hani, D., Khan, N., Alsultan, F., Karanjkar, S., and Nagarur, N. (2018). Classification of leucocytes using convolutional neural network optimized through genetic algorithm.
- Cook, A. (2017). Using transfer learning to classify images with keras. Retrieved from <https://alexisbcook.github.io/2017/using-transfer-learning-to-classify-images-with-keras/>.
- da Saúde, M. Situação epidemiológica da leishmaniose visceral.
- e Silva, R. R. V., de Araujo, F. H. D., dos Santos, L. M. R., Veras, R. M. S., and de Medeiros, F. N. S. (2016). Optic disc detection in retinal images using algorithms committee with weighted voting. *IEEE Latin America Transactions*, 14(5):2446–2454.
- Eder, K., Kutscher, T., Marzi, A., Barroso, Á., Schnekenburger, J., and Kemper, B. (2021). Automated detection of macrophages in quantitative phase images by deep learning using a mask region-based convolutional neural network. In *Label-free Biomedical Imaging and Sensing (LBIS) 2021*, volume 11655, pages 88–94. SPIE.
- Fleiss, J. L., Levin, B., and Paik, M. C. (2013). *Statistical methods for rates and proportions*. John Wiley & Sons.
- Gonçalves, C., Andrade, N., Borges, A., Rodrigues, A., Veras, R., Aguiar, B., and Silva, R. (2023). Automatic detection of visceral leishmaniasis in humans using deep learning. *Signal, Image and Video Processing*, 17(7):3595–3601.
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). Identity mappings in deep residual networks. In *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14*, pages 630–645. Springer.

- Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708.
- Liang, G., Hong, H., Xie, W., and Zheng, L. (2018). Combining convolutional neural network with recursive neural network for blood cell image classification. *IEEE Access*, 6:36188–36197.
- Marcondes, M. and Day, M. J. (2019). Current status and management of canine leishmaniasis in latin america. *Research in Veterinary Science*, 123:261–272.
- Ministério da Saúde, B. (2006). Manual de vigilância e controle da leishmaniose visceral. *Série A. Normas e Manuais Técnicos*.
- Ministério da Saúde. Leishmaniose visceral. Retrieved from <https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/l/leishmaniose-visceral>. Accessed: 08 mar. 2024.
- PAHO (2017). Epidemiological report of the americas. leishmaniasis.
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al. (2015). Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115:211–252.
- Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 618–626.
- Silva, F. S. (2007). Patologia e patogênese da leishmaniose visceral canina. *Rev Trop Cienc Agr Biol*, 1(1):20–31.
- Silva, J. M., Zacarias, D. A., de Figueirêdo, L. C., Soares, M. R. A., Ishikawa, E. A., Costa, D. L., and Costa, C. H. (2014). Bone marrow parasite burden among patients with new world kala-azar is associated with disease severity. *The American Journal of Tropical Medicine and Hygiene*, 90(4):621.
- Simonyan, K. and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations*.
- Szegedy, C., Ioffe, S., Vanhoucke, V., and Alemi, A. (2017). Inception-v4, inception-resnet and the impact of residual connections on learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 31.
- Trong Luong, D., Duy Anh, D., Xuan Thang, T., Thi Lan Huong, H., Thuy Hanh, T., and Minh Khanh, D. (2022). Distinguish normal white blood cells from leukemia cells by detection, classification, and counting blood cells using yolov5. In *2022 7th National Scientific Conference on Applying New Technology in Green Buildings (ATiGB)*, pages 156–160.
- Xia, R., Wang, P., Chen, R., and Guo, F. (2009). One kind of macrophages images segmentation and labeling method. In *2009 2nd International Congress on Image and Signal Processing*, pages 1–5.