

A Patch-based Microscopic Image Analysis for Visceral Leishmaniasis Screening Using a Deep Metric Learning Approach.

Carlos Eduardo Ferreira Lopes¹, Eduardo Lisboa¹, Yanka Ribeiro¹, Fabiane Queiroz²

¹Instituto de Computação – Universidade Federal de Alagoas (UFAL)
Maceió – Brazil

²Laboratório de Computação Científica e Análise Numérica (LaCCAN)
Universidade Federal de Alagoas – Maceió – Brazil

{cefl,eall,yrrs}@ic.ufal.br, fabiane.queiroz@laccan.ufal.br

Abstract. *Human Visceral Leishmaniasis (VL) is a fatal disease in over 95% of untreated cases and predominantly affects populations with limited access to healthcare. Parasitological techniques are the gold standard for diagnosing VL. It involves the direct microscopic examination of the parasite amastigotes approximately 2–4 μ m in diameter. However, this process can be time-consuming and labor-intensive, necessitating a high level of expertise. We propose a novel approach to the detection of these amastigotes by combining deep metric learning with supervised classification techniques. We outperform the state-of-art for this detection problem achieving an f1-score of approximately 99% by tackling poor segmentation and class imbalance drawbacks.*

1. Introduction

Leishmaniasis is caused by species of the intracellular protozoan of the genus *Leishmania*. It is a neglected and infectious vector-borne disease. It occurs in the poorest countries and most vulnerable populations with impaired access to health services. Visceral leishmaniasis (VL) is a more severe form of leishmaniasis. In humans, it is fatal in over 95% of cases if not diagnosed and treated. According to the World Health Organization (WHO), it is characterized by irregular bouts of fever, weight loss, enlargement of the spleen and liver, and anemia. In the Americas, VL is endemic in 12 countries. South American countries such as Brazil, Argentina, Colombia, Paraguay, and Venezuela have among the highest case records. Most cases occur in Brazil [WHO TEAM 2023].

Diagnosis VL is a combination of laboratory diagnosis and clinical analysis. Laboratory examination consists of DNA-based methods (Polymerase chain reaction (PCR) and quantitative real-time PCR - qPCR) and non-DNA-based. However, these techniques are complex and expensive, and in most VL-endemic countries, they are restricted to a few teaching hospitals and research centers [Elmahallawy et al. 2014]. Among the non-DNA-based, serological methods detect antibodies or antigens (such as proteins), and parasitological methods is about the direct microscopic examination and culture from adequate samples.

Parasitological techniques are the gold standard for diagnosing VL. It includes the direct microscopic observation of the amastigote form of the parasite in the aspired/biopsied material, including bone marrow, lymph nodes, and spleen. Smears are

simple to prepare, and their direct examination is usually the best diagnostic method in more impoverished areas where PCR is not available. Still, the sensitivity of this procedure is about 60% to 85% [Elmahallawy et al. 2014].

Leishmania amastigotes are intracellular round or oval bodies, about 2–4 μm in diameter. The miniature size of these amastigotes makes this a tedious task that can be very time-consuming and require an expert skill level. Figure 1 presents an example of *Leishmania* amastigotes in a bone marrow microscopic color image. Since the VL diagnostic is time-consuming and requires costs and technical expertise, its sensitivity is relatively low. The more secure procedure is to obtain the biopsy from the bone marrow, and the sensitivity of the bone marrow stained with Giemsa¹ is about 60% to 85% [Elmahallawy et al. 2014].

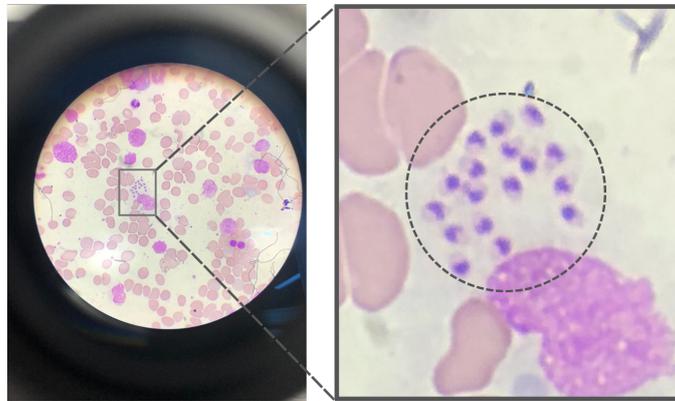


Figure 1. Example of an image captured from bone marrow smears. The zoomed circular area indicates the presence of *Leishmania* amastigotes.

Machine Learning (ML), particularly Deep Learning, has significantly impacted diagnostic fields like radiology and pathology due to its ability to recognize patterns in medical images [Lima et al. 2023, Santos et al. 2023]. Deep Convolutional Neural Networks (CNNs) are frequently employed for medical image classification tasks to overcome limitations of manual approaches. In these fields, reproducibility among physicians is often suboptimal [Van der Laak et al. 2021]. However, challenges such as inadequate training data and class imbalance can lead to overfitting and reduced performance of CNNs. To address these issues, recent approaches incorporate Deep Metric Learning (DML) and Siamese Networks [Zhang and Peng 2019] or Triplet Loss to enhance model robustness and handle class imbalance.

In this work, we propose a hybrid Deep Metric Learning and Support Machine Vector (SVM) model to classify images from bone marrow smears as either containing the amastigote or not. We undertook an image preprocessing pipeline for dealing with feature discernibility and lopsided training dataset, thus overcoming limitations encountered in previous studies.

¹Giemsa's staining solution is one of the most common microscopic stains, generally used in hematology, histology, cytology, and bacteriology for in vitro diagnostic.

2. Related Work

Most parasitic protozoans in humans range less than $50\ \mu\text{m}$ in size [Reimão et al. 2020] and present a significant challenge to diagnosis from microscopy image examination. Several approaches have been proposed for parasite examination from microscopy images in recent years [Zhang et al. 2022]. Classification methods include cell type differentiation and are typically used for object detection and segmentation.

Recent works have proposed methods for malaria parasite detection in thick blood smears using smartphones. [Yang et al. 2020], presents a method in two steps: first they applied an intensity-based Iterative Global Minimum Screening (IGMS), which performs a fast screening of a thick smear image to find parasite candidates. Afterward, a customized CNN classifies each candidate as a parasite or background. [Fuhad et al. 2020] deployed the miniaturized model in mobile phones and a server-backed web application. [Soberanis-Mukul et al. 2013] proposed a method for automatic detection *Trypanosoma cruzi* in digital microscope images obtained from peripheral blood smears treated with Wright’s stain. They propose a combination of image pre-processing algorithms with a K-Nearest Neighbors (KNN) classifier applied over a segmented region from the original image.

Few works in the state of the art implement an automated *leishmania* examination over images from bone marrow smears. In [Farahi et al. 2015], the authors utilize morphological and CV level set methods to segment *Leishmania* bodies in digital color microscopic images captured from bone marrow samples. In [Salazar et al. 2019], a semiautomatic segmentation strategy is proposed to obtain the segmentation of the evolutionary shapes of VL parasites. Smoothing filters and edge detectors enhance the optical microscopy images, and a region-growing algorithm does the segmentation. [Isaza-Jaimes et al. 2021] propose a detection method that uses image processing techniques, like low-pass filters, gradient operators, and gradient modules based on polar maps of the pixel intensities. [Górriz et al. 2018] trained a U-net model that successfully segments *leishmania* parasites and classifies them into promastigotes, amastigotes, and adhered parasites. Along the same lines, [Gonçalves et al. 2023] employed a U-Net architecture to automatically pinpoint the pixels of interest in the images, in this context, those containing *Leishmania* parasites. This process was guided by binary masks annotated by specialists. The experiments of [Farahi et al. 2015, Salazar et al. 2019, Isaza-Jaimes et al. 2021] were performed over a public dataset provided by [Farahi et al. 2015] whereas [Ronneberger et al. 2015] and [Górriz et al. 2018] conducted their experiments in non-public datasets.

These approaches were selected through a comprehensive search of relevant works on scientific databases, such as Google Scholar and Web of Science. We identified that most of the work falls short by not effectively addressing the challenge of significant class imbalance, a problem we tackle through the application of deep metric learning. Moreover, our proposed method eliminates the necessity for preprocessing assistance in the segmentation of *leishmania* amastigotes. This is achieved by employing a patch-based approach, which remains unaffected by potentially poor segmentation.

3. Datasets and Pre-processing

We analyze two different image datasets of microscopy images from bone marrow aspirates. This section discusses the collection, annotation, and pre-processing of all the datasets used to validate our experiments.

3.1. Datasets

The **Dataset I**, as described by [Farahi et al. 2015], comprises 45 light-microscope images of human patients with VL slides of bone marrow aspirates. Imaging was conducted using a Sony high-resolution digital camera (DSC-H9) coupled to an Olympus CH40RF200 microscope, yielding images at a 100x magnification with a spatial resolution of 3840×2880 pixels. The images are in 24-bit RGB color format and JPEG compression.

Dataset II consists of 76 shots from myelogram slides of 20 patients diagnosed with VL. The images have a spatial resolution of 3024×4032 pixels and are in 24-bit RGB color format. They were captured using an iPhone 8 coupled with light microscopy equipped with an oil immersion objective at 100x magnification. Each smear underwent a thorough examination by a physician for at least an hour, with only confirmed cases of VL included in the research.

An example of images from both datasets, along with its corresponding mask indicating *Leishmania* parasites, is depicted in Figure 2.

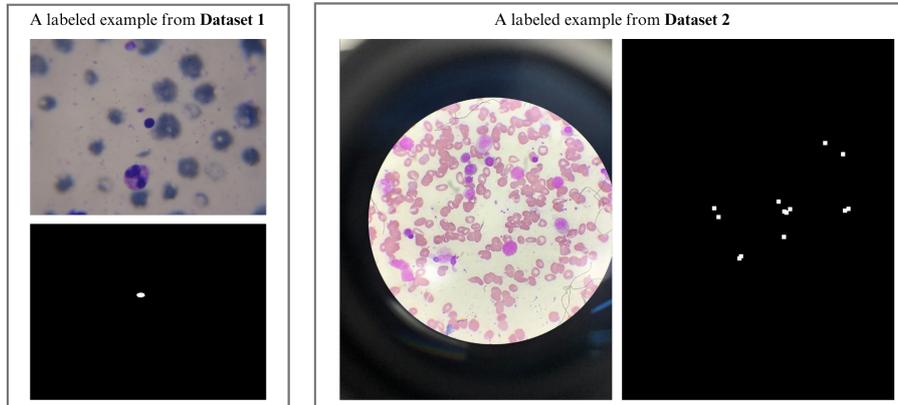


Figure 2. From top left to bottom right: Example of the original RGB color image and its associated label for the image Datasets I and II, respectively. This figure is for visual purposes only; the objects are out-of-scale.

3.2. Image Pre-processing

Considering \mathcal{I} as the union of the images from **Datasets I** and **II** and assuming that our detection method is patch-based defined, each example $x_i \in \mathcal{X}$ that feeds our algorithm will be an RGB color image patch of size 96×96 extracted from an image of \mathcal{I} . We performed some transformation over the images as shown in Figure 3. They include filtering and identification of the Region of Interest (RoI). These steps aim to segment the circular RoI, given the image smoothed with a Gaussian convolution filter with a 25×25 kernel size to remove some noise generated by the capture process. After that, we used a Circle Hough Transform (CHT) to identify arbitrary circular shapes in the blurred image. Next, we applied linear interpolation to improve contrast in both datasets.

3.3. Patches Generation

The *Leishmania* parasite represents a small dot on the image, with a proportion of 3% to 5% of the image size. Then, using images with original dimensions (set \mathcal{I}) on our model implies the problem of losing information about the amastigotes' pixels. We adopted a patch-based approach generating 96x96 patches (set \mathcal{X}) by traversing the binary masks with the locations of amastigotes in RGB images [Gonçalves et al. 2023]. When a marked area is encountered in the mask, the step of the sliding window is decreased to an eighth of its original size. Subsequently, the area of the *Leishmania* within this region is analyzed. If this area exceeds a predetermined threshold, the resulting patch is categorized as belonging to the positive class. Conversely, if the area falls below this threshold, it is classified as negative. Figure illustrates the image pre-processing steps followed by our proposed patch generation approach.

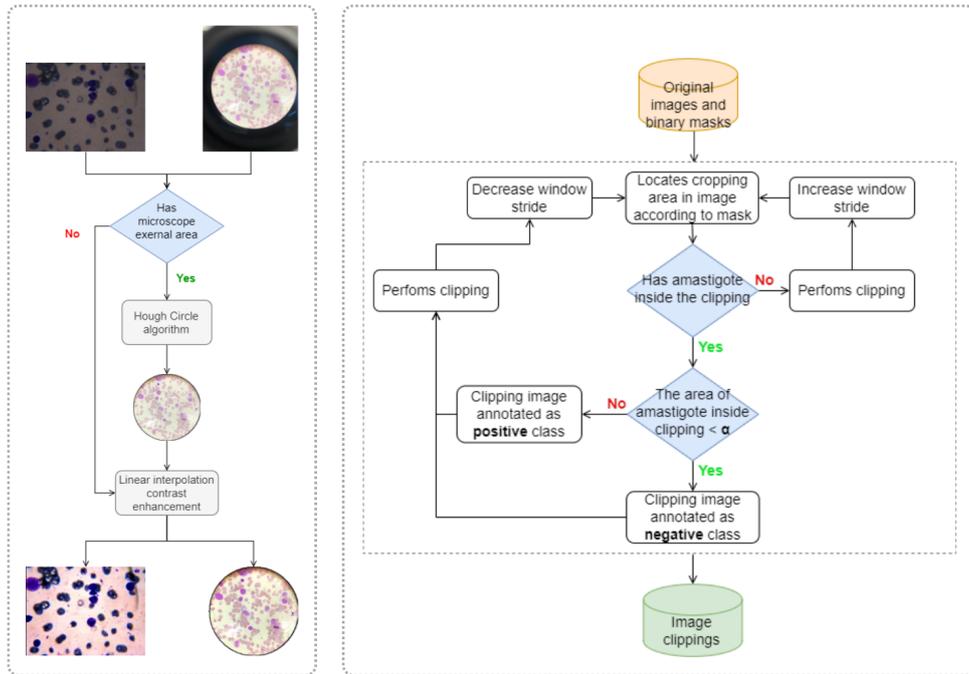


Figure 3. From left to right: A CHT is applied over the RGB images to identify our RoIs and linear interpolation is applied to contrast enhancement (left); patches of 96x96 are extracted according to the algorithm (right). This figure is for visual purposes only, and the objects are out-of-scale.

By implementing the conditional window stride size and generating as many positive patches as possible, our work innovates, avoiding relying too heavily on synthetic data as a means of resolving class imbalances. Additionally, we adapted the algorithm to preprocess images under varying conditions that may be encountered in different datasets, such as the presence of the microscope eyepiece (Figure 3).

3.4. Class Balancing

Taking into account both datasets, the set of patches denoted as \mathcal{X} comprises a total of 67,889 patches with dimensions of 96x96x3, among which 4,316 contain *leishmania* bodies. This situation elucidates that we are dealing with an imbalanced classification problem. Data augmentation is a technique used to generate additional training samples by applying transformations to existing data.

To address our class imbalance and simultaneously avoid the pitfall of overfitting, we produced synthetic data for one class and downsampled the other. For the positive class, we applied a set of specified geometric transformations, such as rotating each image by up to 120 degrees, applying both horizontal and vertical flips, and zooming up to 10%. Downsampling involves randomly removing k samples from the majority class. We set k to be a number that $k = N - 2P$, with N and P being the size of the negative and positive classes, respectively. This process achieved a dataset proportion of 1:2, with 43.468 negative patches and 21.734 positive.

4. Method

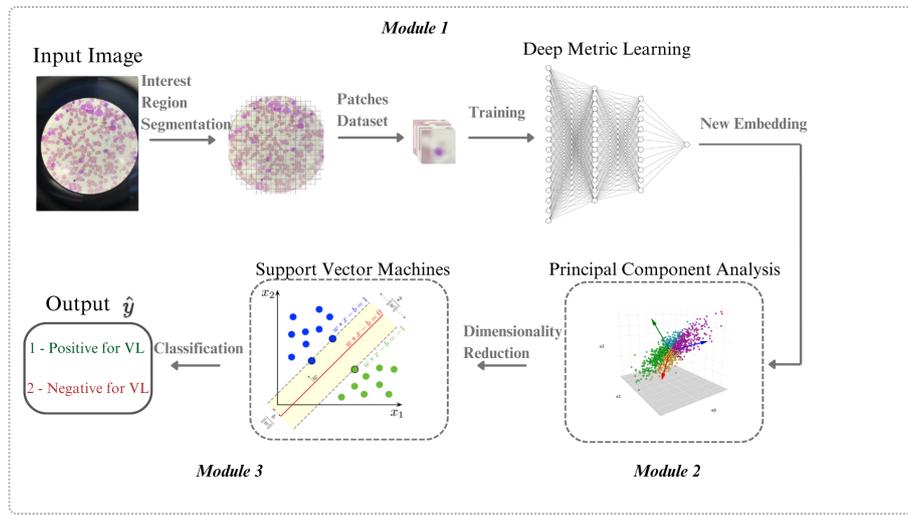


Figure 4. Schematic representation of the proposed method.

As shown in figure 4, we propose a novel hybrid patch-based classification model to assist the VL diagnostic. Our method can be divided into three modules: (i) New representation of data: a CNN with DMel model to extract embeddings containing amastigotes characteristics from bone marrow smears; (ii) Dimensionality Reduction: Performing Principal Component Analysis (PCA) on the new representation data to reduce its dimensionality and data complexity; (iii) Embeddings Classification: using classical supervised method, Support Vector Machines (SVM), to classify the embeddings and therefore assist the VL diagnostic.

4.1. Module 1: New representation of data for *Leishmania* Parasite Detection

DMeL quantifies the similarity between two elements based on the data context. These approaches find a new representation of the data so that objects with the same classes are grouped. Furthermore, unlike traditional classification model training, the training loss is computed based on the new representation space obtained from multiple patch images. We selected *Triplet*, *Circle*, *MultiSimilarity*, and *NPairs* [Schroff et al. 2015, Sun et al. 2020, Wang et al. 2019, Sohn 2016] as DMeL loss functions for the performance comparison effect. Each model is a CNN and its output is significantly influenced by the choice of the loss function during training.

4.2. Module 2: Dimensionality Reduction

The underlying principle of PCA is based on the understanding that the informative rank of the data is typically less than the number of original variables. We aimed to enhance the ability of classical models to carry out the classification task by performing dimensionality reduction on the novel embedding representation of the data. We checked the variance measurement along each principal component to identify and determine the optimal quantity of dimensions for our reduced embedding vector, ensuring the best representation of the data. Through a comparison of the relative importance of each dimension, we elected to retain only the number of principal components necessary to capture 90% of the variance in the data.

4.3. Module 3: Embeddings Classification

Finally, we conducted classification with a widely used algorithm for classification tasks. SVMs identify optimal hyperplanes in high dimensions for effective class separation and generalization. The classifiers underwent tuning via Grid Search with cross-validation to optimize regularization parameters and kernel coefficients for each CNN model. The search utilized recall macro as the scoring metric and was executed in parallel for computational efficiency. The resulting SVM classifiers efficiently categorized the transformed embeddings with high recall rates.

5. Experimental Setup and Results

Training deep learning networks requires using different sets of data, including training, validation, and testing for maximal efficacy. For this experiment, we used 70% of the data for training, 15% for validation, and 15% for testing. The training phase was conducted with a batch size of 32 images and 100 epochs monitored by an early stopping mechanism. We set an initial learning rate of 0.001 and then dynamically adjusted using a learning rate scheduler that reduced the rate by a factor of 0.1 if the validation loss reached a plateau. The output is a vector of size 128. Each CNN model was trained with the same architecture, with the only difference being the loss function utilized.

The CNN architecture consists of three 2D convolutional layers with increasing channel depths, specifically 32, 64, and 128, respectively. Each convolutional layer utilizes a kernel size of 3 and a stride of 1. Following each convolutional layer, a batch normalization layer, corresponding to its channel depth, is incorporated to stabilize the learning process by normalizing the output of each layer. Rectified Linear Unit (ReLU) activation functions are applied after each convolutional and batch normalization layer. Additionally, two max-pooling layers, each with a window size of 2x2, are implemented after the second and third convolutional layers to reduce the spatial dimensions of the feature maps. To prevent overfitting, we integrated two dropout layers, namely dropout1 with a rate of 0.25 and dropout2 with a rate of 0.5. Finally, the architecture includes two linear layers, fc1 and fc2. The former reduces the dimensionality to 512, while the latter maps these features to the specified embedding size, 128. In the context of the loss functions underlying distance metric, specifically, the Triplet, MultiSimilarity, and Circle utilized cosine similarity. While the Npairs loss function utilized the dot product.

Concerning the individual parameters within each DMel function, we decided to set a value of 0.3 for the margin hyperparameter of Triplet loss. Positive examples are

closer to the anchor than the negative examples by at least this margin. Our module followed [Sun et al. 2020] recommendation for fine-grained image retrieval setup for Circle loss. The relaxation factor, which controls the radius of the decision boundary, was set to 0.4 and 80 for the gamma parameter. For MultiSimilarity, we used an alpha of 2 and a beta of 50, enhancing the model’s focus on informative pairs. The margin parameter lambda was established at a value of 1.

As previously stated, we used recall macro as the evaluation metric of the exhaustive search for SVM parameter optimal values. Across the four loss functions analyzed, all were trained with C set to 10 and Gamma set to 1. The scores varied among the different models, with the Circle loss function achieving the highest score of 0.9994, followed by Triplet with 0.9915, MultiSimilarity with 0.9750, and NPairs with 0.9441.

To have a baseline for comparing the proposed method, we also implemented three classic feature extraction algorithms, namely SIFT, ORB, and Haralick, each of them paired with the SVM and K-Nearest Neighbors (KNN) classifiers, as well as classical CNN models (VGG19 and DeCaf).

The SIFT algorithm was configured with the number of octaves as 8, the number of scales as 3, and an upsampling of 2, meaning that before the feature detection stage, the image is upscaled by a factor of 2 and in each octave the image is repeatedly convolved with Gaussian blurs, being the number of convolutions equal to the number of scales. After each octave, the Gaussian image is downsampled by a factor of 2, repeating this process the number of octaves defined. The ORB algorithm was configured with the number of scales as 8 and the number of key points as 500, meaning that each image would be created as an image pyramid where each level of the pyramid is the image of the previous layer downsampled by a factor of 2, with the pyramid having a total number of layers equal to the number of scales, and the algorithm will try to find up to 500 key points in the images. For the Haralick algorithm, the parameter to compute the 14th feature was set to false, as this feature is considered unstable and unreliable, and the parameter to ignore zeros is also set to false, as we want to consider black pixels in the images.

The VGG19, a CNN with 16 convolutional and 3 fully connected layers, is used to extract visual features from images. Instead of relying on the final classification output, DeCaf (Deep Convolutional Activation Feature) leverages activations from earlier layers within the pre-trained VGG19. These earlier layers capture more general visual features like edges and textures, making them well-suited for feature extraction tasks.

5.1. Results

All trained models were loaded and tested against the same data set. Predictions were made using the appropriate classifier on scaled embeddings, and metrics such as **precision**, recall for positive class (**Sensibility**, which measures the proportion of true positives correctly identified), recall for negative class (**Specificity**, which measures the proportion of true negatives correctly excluded), **F1-score** and Matthews Correlation Coefficient (**MCC**). MCC is a binary classification performance indicator that provides a fair assessment even when the classes are unequal in number [Chicco and Jurman 2020]. The MCC yields a result between -1 and 1, with 1 representing a perfect prediction and achievable only if the result performed well in all four confusion matrix categories, proportionally to the number of positive and negative items in the dataset. We chose the final two metrics

Table 1. Performance results for each loss function examined.

Loss Function	Precision	Sensibility	Specificity	F1-Score	MCC
Triplet	0.963 (0.004)	0.828 (0.011)	0.984 (± 0.001)	0.890 (0.007)	0.8467
Circle	0.987 (0.002)	0.983 (0.004)	0.993 (± 0.001)	0.985 (0.002)	0.9775
Multisimilarity	0.886 (0.016)	0.961 (0.005)	0.937 (± 0.010)	0.922 (0.008)	0.8818
NPairs	0.894 (0.008)	0.950 (0.011)	0.943 (± 0.004)	0.921 (0.008)	0.8806

due to the unbalanced nature of our dataset and its property of considering the number of samples correctly classified as negative.

Table 1 presents the performance analysis for DMel for each studied loss function. Since we applied stratified cross-validation on the test set, the values are about the averages and standard deviations (in parenthesis) of these metrics then computed from the sum of all the folds' outputs.

By carefully examining the data, it is found that while the Triplet loss function demonstrated a balanced proficiency in distinguishing negative examples, as evidenced by its high recall in the negative class, it also demonstrated a tendency toward higher false negatives in the positive class, as evidenced by the lower recall shown in Table 1. This also influences its MCC score of 0.8467, which, despite being the lowest of the examined functions, still demonstrates a high level of performance.

On the other hand, MultiSimilarity stands as an improvement of Triplet and falls behind the Circle model. Despite possessing the lowest specificity, which suggests some challenges in correctly identifying all negative instances, its sensibility outperforms Triplet by 13%. A fair trade, since the nature of VL treatment is more tolerable to errors than missing a critical diagnosis. This is theoretically consistent with the MultiSimilarity function's objective of simultaneously pulling together similar examples and pushing apart dissimilar ones within the same batch, which may account for its relatively strong discriminative power.

According to these results, one can infer that the overall performance of the CNNs modified versions was excellent, with the Circle Loss model demonstrating superior performance across all classification assessments.

Notably, the proposed model, utilizing Circle loss, demonstrates superior performance with higher precision, sensitivity, specificity, and F1-score even when compared to classical methods and other CNN architectures, as shown in Table 2. Table 3 underscores the efficacy of leveraging advanced techniques such as DMel for improved classification outcomes in image processing tasks.

Table 2. Comparative classification performance for our model and other classical methods.

Type of classification	Models	Precision	Sensibility	Specificity	F1-Score
Handcrafted feature extraction	SIFT + SVM	0.8347 (± 0.009)	0.7615 (± 0.010)	0.8797 (± 0.007)	0.7964 (0.005)
	ORB + SVM	0.7709 (± 0.004)	0.7299 (± 0.006)	0.7730 (± 0.009)	0.7498 (± 0.003)
	Haralick + KNN	0.6709 (± 0.008)	0.7078 (± 0.007)	0.8267 (± 0.006)	0.6888 (± 0.005)
CNN Classification Models	VGG19	0.3750	0.5000	-	0.425
	DeCaf	0.3750	0.5000	-	0.425
Our Classification Model	DMel with Circle	0.987 (± 0.002)	0.983 (± 0.004)	0.993 (± 0.001)	0.985 (± 0.002)

Table 3. The proposed method’s performance in comparison to the state-of-the-art.

Technique	Models	Dice	Sensibility	Specificity	F1-Score
Detection Image Processing	[Isaza-Jaimes et al. 2021]	-	0.787	-	-
Segmentation U-Net	[Górriz et al. 2018] [Salazar et al. 2019]	0.777 0.850	0.823 -	- -	- -
Classification CNN + Circle loss + SVM	Our Classification Model	-	0.983 (± 0.004)	0.993 (± 0.001)	0.985 (± 0.002)

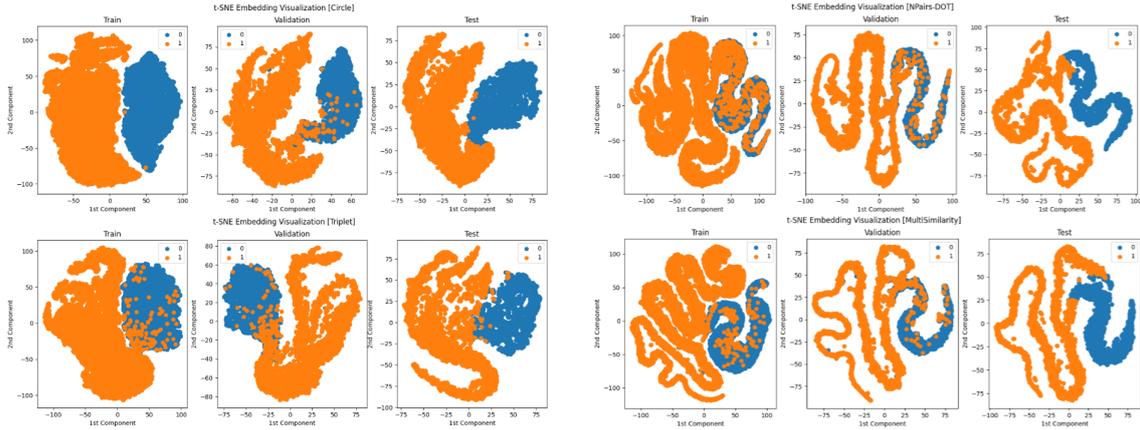


Figure 5. From top to bottom, left to right: t-SNE Embedding visualization of Circle, Triplet, NPairs and MultiSimilarity. Blue class indicates the amastigote presence, orange the absence.

For data separability analysis, we applied t-Distributed Stochastic Neighbor Embedding (t-SNE), and Figure 5 reveals that the Triplet method exhibits reasonable class separation, forming dense clusters with some overlap areas. However, inconsistencies between the validation data and other sets are observed, despite measures against overfitting. Circle Loss demonstrates exceptional performance by focusing on challenging pairs and optimizing class margins, resulting in distinct and consistent clusters in the training and test categories. In contrast, MultiSimilarity embeddings, exhibit less compact but noticeable class separation. The MultiSimilarity and NPairs loss functions, prove particularly sensitive to intraclass variance, making it more useful in scenarios where such diversity is relevant as in predicting the parasite evolutionary form.

6. Conclusion and Future Work

The evaluation of various deep metric learning methods highlighted their promising. This project was developed in partnership with the Ministry of Health, and the model has become ready for usage on smartphones. Circle Loss emerged as the clear winner, achieving impressive performance across all classification metrics, particularly sensitivity (98.3%) and specificity (99.3%). This success can be attributed to two key factors. The preprocessing of the images by segmenting them into smaller patches enhances the visibility of crucial features for the models. The SVM algorithm effectively translates the learned features into actionable diagnostic insights.

While encouraging, the evaluation also revealed areas for improvement. For instance, the Triplet loss function had a higher false negative rate. This indicates that more

exploration is required in image background pre-processing and model fine-tuning to reduce false negatives. The performance of model may be influenced by the limited size of the training dataset, which may not be representative of all possible scenarios. Additionally, it is vital to evaluate the proposed model's effectiveness in diagnosing other parasitic infections, such as malaria or Chagas disease, to understand its full potential. Building on these findings, future research will focus on optimizing the generated models for even greater effectiveness in the field of parasitological diagnostics.

References

- Chicco, D. and Jurman, G. (2020). The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation. *BMC genomics*, 21(1):1–13.
- Elmahallawy, E. K., Sampedro, A. M., Rodriguez-Granger, J., Hoyos-Mallecot, Y., Agil, A., Navarro, J. M., and Fernández, J. (2014). Diagnosis of leishmaniasis. *Journal of Infection in Developing Countries*, 8(8):961 – 972.
- Farahi, M., Rabbani, H., Talebi, A., Sarrafzadeh, O., and Ensafi, S. (2015). Automatic segmentation of Leishmania parasite in microscopic images using a modified CV level set method. In Wang, Y. and Jiang, X., editors, *Seventh International Conference on Graphic and Image Processing (ICGIP 2015)*, volume 9817, page 98170K. International Society for Optics and Photonics, SPIE.
- Fuhad, K. M. F., Tuba, J. F., Sarker, M. R. A., Momen, S., Mohammed, N., and Rahman, T. (2020). Deep learning based automatic malaria parasite detection from blood smear and its smartphone based application. *Diagnostics*, 10(5).
- Gonçalves, C., Borges, A., Dias, V., Marques, J., Aguiar, B., Costa, C., and Silva, R. (2023). Detection of human visceral leishmaniasis parasites in microscopy images from bone marrow parasitological examination. *Applied Sciences*, 13(14).
- Górriz, M., Aparicio, A., Raventós, B., Vilaplana, V., Sayrol, E., and López-Codina, D. (2018). Leishmaniasis parasite segmentation and classification using deep learning. In Perales, F. J. and Kittler, J., editors, *Articulated Motion and Deformable Objects*, pages 53–62, Cham. Springer International Publishing.
- Isaza-Jaimes, A., Bermúdez, V., Bravo, A., Castrillo, J. S., Lalinde, J. D. H., Fossi, C. A., Flórez, A., and Rodríguez, J. E. (2021). A computational approach for Leishmania genus protozoa detection in bone marrow samples from patients with visceral Leishmaniasis.
- Lima, T., Catiely, C., Sousa, V., Veras, R., and Araújo, F. (2023). Aprendizado profundo para detecção de cálculos renais em imagens de tomografia computadorizada. In *Anais do XXIII Simpósio Brasileiro de Computação Aplicada à Saúde*, pages 47–58, Porto Alegre, RS, Brasil. SBC.
- Reimão, J. Q., Coser, E. M., Lee, M. R., and Coelho, A. C. (2020). Laboratory diagnosis of cutaneous and visceral leishmaniasis: Current and future methods. *Microorganisms*, 8(11).
- Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In Navab, N., Hornegger, J., Wells, W. M., and Frangi,

- A. F., editors, *Medical Image Computing and Computer-Assisted Intervention – MIC-CAI 2015*, pages 234–241, Cham. Springer International Publishing.
- Salazar, J., Vera, M., Huérfano, Y., Vera, M. I., Gelvez-Almeida, E., and Valbuena, O. (2019). Semi-automatic detection of the evolutionary forms of visceral leishmaniasis in microscopic blood smears. *Journal of Physics: Conference Series*, 1386(1):012135.
- Santos, P., Brito, V., Filho, A. C., Sousa, A., Diniz, J., and Luz, D. (2023). Efficientbacillus: uma arquitetura profunda para detecção dos bacilos de koch. In *Anais do XXIII Simpósio Brasileiro de Computação Aplicada à Saúde*, pages 198–209, Porto Alegre, RS, Brasil. SBC.
- Schroff, F., Kalenichenko, D., and Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Soberanis-Mukul, R., Uc-Cetina, V., Brito-Loeza, C., and Ruiz-Piña, H. (2013). An automatic algorithm for the detection of trypanosoma cruzi parasites in blood sample images. *Computer Methods and Programs in Biomedicine*, 112(3):633 – 639.
- Sohn, K. (2016). Improved deep metric learning with multi-class n-pair loss objective. In Lee, D. D., Sugiyama, M., Luxburg, U. V., Guyon, I., and Garnett, R., editors, *Advances in Neural Information Processing Systems 29*, pages 1857–1865. Curran Associates, Inc.
- Sun, Y., Cheng, C., Zhang, Y., Zhang, C., Zheng, L., Wang, Z., and Wei, Y. (2020). Circle loss: A unified perspective of pair similarity optimization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6398–6407.
- Van der Laak, J., Litjens, G., and Ciompi, F. (2021). Deep learning in histopathology: the path to the clinic. *Nature Medicine*, 27(5):775–784.
- Wang, X., Han, X., Huang, W., Dong, D., and Scott, M. R. (2019). Multi-similarity loss with general pair weighting for deep metric learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5022–5030.
- WHO TEAM, C. o. N. T. D. (2023). Operational manual on leishmaniasis vector control, surveillance, monitoring and evaluation. page 119.
- Yang, F., Poostchi, M., Yu, H., Zhou, Z., Silamut, K., Yu, J., Maude, R. J., Jaeger, S., and Antani, S. (2020). Deep learning for smartphone-based malaria parasite detection in thick blood smears. *IEEE Journal of Biomedical and Health Informatics*, 24(5):1427–1438.
- Zhang, C., Jiang, H., Jiang, H., Xi, H., Chen, B., Liu, Y., Juhas, M., Li, J., and Zhang, Y. (2022). Deep learning for microscopic examination of protozoan parasites. *Computational and Structural Biotechnology Journal*, 20:1036–1043.
- Zhang, Z. and Peng, H. (2019). Deeper and wider siamese networks for real-time visual tracking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.