2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Convolutional Neural Networks for Chagas' Parasite Detection in Histopathological Images*

N. Sanchez-Patiño[1], A. Toriz-Vazquez[2], N. Hevia-Montiel[3] and J. Perez-Gonzalez[3]

***Abstract*— Chagas disease is a widely spreaded illness caused by the parasite *Trypanosoma cruzi* (T. cruzi). Most cases go unnoticed until the accumulated myocardial damage affect the patient. The endomyocardium biopsy is a tool to evaluate sustained myocardial damage, but analyzing histopathological images takes a lot of time and its prone to human error, given its subjective nature. The following work presents a deep learning method to detect T. cruzi amastigotes on histopathological images taken from a endomyocardium biopsy during an experimental murine model. A U-Net convolutional neural network architecture was implemented and trained from the ground up. An accuracy of 99.19% and Jaccard index of 49.43% were achieved. The obtained results suggest that the proposed approach can be useful for amastigotes detection in histopathological images.**

***Clinical relevance*— The proposed method can be incorporated as automatic detection tool of amastigotes nests, it can be useful for the Chagas disease analysis and diagnosis.**

I. Introduction

Chagas disease is a parasitic disease endemic from Latin America and caused by the flagellated protozoan *Trypanosoma cruzi* (T. cruzi). It can be transmitted to human beings via hematophagous insects from the subfamily *Triatominae*. There are several strains of T. cruzi, each one presenting tropism for different cells. In Mexico, 30% of symptomatic infected people develop a serious cardiomyopathy due to the accumulated myocardial damage during the years prior to diagnosis. The damage varies from minor affectations to end-stage heart failure. It is estimated that Chagas disease affects around 6 to 7 million people in Latin America; 300,000 in United States and between 80,000 to 120,000 in Europe. In México, even though the oficial records report few hundreds of cases each year, it is believed that there are at least 1.1 million people infected with T. cruzi, while 29.5 million at risk of infection. The World Health Organization has classified Chagas disease as one of the least attended tropical diseases [1, 2, 3, 4, 5].

*This work was supported by UNAM-PAPIIT programs: IT100220, IA102920 and CONACYT PDCPN-2015-102.

[1]N. Sanchez-Patiño is with Centro de Física Aplicada y Tecnología Avanzada, Universidad Nacional Autónoma de México 3001, 76230 Juriquilla, Qro., México natyangelessp@gmail.com

[2]A. Toriz-Vázquez is with the Faculty of Engineering, Universidad del Valle de México, Naranjos Punta Juriquilla 1000 Santa Rosa Jáuregui, 76230 Santiago de Querétaro, Qro., México alfonso.toriz@my.uvm.edu.mx

[3]N. Hevia-Montiel and J. Perez-Gonzalez are with the Instituto de Investigaciones en Matematicas Aplicadas y en Sistemas, Universidad Nacional Autonoma de Mexico, Km 5.5 Carretera Sierra Papacal - Chuburna, Merida Yucatan 97302, México. nidiyare.hevia@iimas.unam.mx, jorge.perez@iimas.unam.mx

T. cruzi goes through several morphological stages during its lifecyle. Inside triatomine's intestine it multiplies and develops into metacyclic tripomastigote. Then, when the triatomine bites a human and defecates into the wound, T. cruzi travels via the circulatory system, thus becoming a sanguine tripomastigote. Later on, it will transform into amastigotes to infect cells from the phagocytic system, lymphoid, muscular or nervous tissue. Amastigotes' most prominent features are their ovoid figure and ability to multiply via binary fission [1].

Comparing the 1.1 million estimated cases of Chagas disease in Mexico with the 5463 cases reported in official records across 2000-2012, a huge gap between them is easily spotted [2, 6]. The majority of these cases might have been unnoticed until the patient presented a cardiomiopathy. Chagas heart disease is characterized by T. cruzi amastigotes infecting myocardium tissue, which may induce inflammation of the four cardiac chambers, and later on, cardiomegaly and increased cardiac mass. Histological samples show amastigotes nests accumulate inside myofibers, resulting in cardiac fibrosis, degeneration and necrosis [7]. However, histological analysis of the samples by photomicrography can take a lot of time and the correct detection of nests is observer-dependent [8]. Since 1999, the interest in developing algorithms for computational assisted diagnosis (CAD) has grown with the purpose of reducing pathologists' workload. Nowadays, methods derived from deep learning are the most successful and studied ones [8]. The purpose of this work is to train a convolutional neural network (CNN) with U-Net architecture for the detection of T. cruzi amastigotes' nests in histopathological images taken during an experimental murine model.

Despite the attention that Chagas disease has received in recent years, few research has been conducted regarding machine learning or deep learning applications. That is why similar works with different diseases have been considered for comparison, e.g. malaria and leishmaniasis.
In 2013, Cetina *et al.* [9], proposed T. cruzi detection during its sanguine tripomastigote stage in blood smears using a Gaussian discriminant analysis, reporting a sensibility of 98.333% and specificity of 15.63%. In 2015, the same authors used a support vector machine (SVM) in combination with AdaBoost, obtaining a sensibility of 100% and specificity of 93.25%. In 2017, Mehanian *et al.* [10], utilized CNNs to identify and quantify the presence of malaria

parasites *Plasmodium falciparum* in blood smears, achieving a sensibility of 91.6%, specificity of 94.1% and precision of 89.7%. In 2018, Górriz *et al.* [11], applied the U-net architecture for segmentation and classification of leishmaniasis parasites, reporting a precision of 75.7% and dice score of 77.7% regarding amastigotes. Finally, in 2020, Ojeda *et al.* [12], implemented a CNN with U-net architecture for segmentation of T. cruzi sanguine trypomastigote in blood samples, reporting a F2 of 80%, recall value of 87.02%, precision of 63.04% and dice score of 68.25%.

II. Materials and Methods

A. Dataset

A set of 767 histopathological images with hematoxylin and eosin stain obtained from an experimental murine model were used. The samples were taken from 9 infected mice during the acute stage of the disease, specifically at days 25, 30 and 35 post-inoculation with 1000 blood trypomastigotes, as approved by the Ethics Committee. All images have a resolution of 2592 x 1944 pixels and were acquired with two different microscopes.

From these 767 images, 33 were reserved for the final test. MATLAB Image Processing Toolbox R2020a was used to make the manual binary mask segmentations (ROIs) for each nests of amastigotes. The manual segmentations were validated by an expert in parasitology. The resulting images were resized to 512 x 512 pixels in order to reduce the number of inputs of the neuronal network.

B. Data augmentation

The absence of medical images data is one of the challenges faced when implementing deep learning algorithms, since working with few images and low variability could incite over-fitting. Data augmentation is one solution to solve this problem. This process expands the original data by applying different transformations to create new data. Thus, the generalization of the model and precision is improved, as well as its robustness for not seen data. There is proof that geometrical, contrast and brightness operations lead to better results [13]. The following transformation were chosen for the data augmentation process: 90 and 180 degrees rotations, random rotations between -45 and 45 degrees, reflections along the vertical and horizontal axis, changes in contrast and brightness levels; and spectral changes in the RGB components. 11,508 images were obtained for training and validation split following the data augmentation procedure.

C. U-Net Architecture

The U-Net is a convolutional neural network architecture design for image segmentation or detection applications. Its structure allows to obtain results while reducing training time. It is divided into two paths; the first path is known as the contracting path or encoder, the second path is called the expansive path or decoder; which gives it a symmetrical u-shaped structure. The contraction path provides the classification information, while the expansion path allows the network to learn localized classification information.

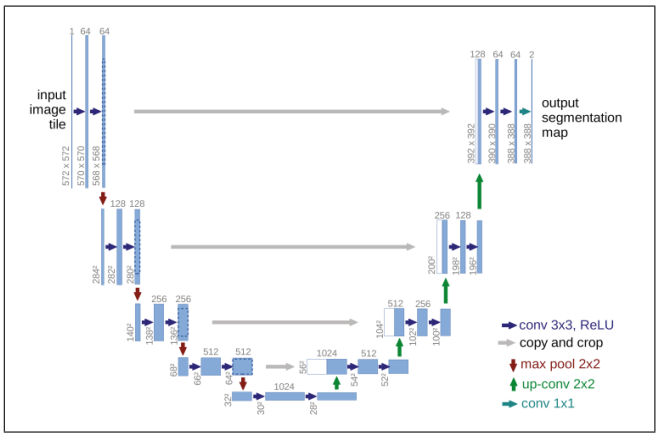

Fig. 1: U-Net architecture

There is a series of skip-connections between different layers from the contracting path and the expansive path called concatenations, they convey feature maps from one to the other. Moreover, the U-net architecture solved the persistent problem of missing pixel-level context information needed for medical image analysis, as shown in [14, 15].

The contraction path applies two 3x3 convolutions, using ReLu as the activation function. Then, a 2x2 max-pooling operation is applied. This process is repeated five times, each time increasing the kernel size as the U-Net architecture shows (Fig 1). Later, the expansion process begins, which consists of several 2x2 up-samplings along with concatenations of the contraction path map of features. The process is repeated until reaching the contraction path length. At the end, a 1x1 convolution with two filters is used to map each component feature vector to the desired number of classes in order to generate the output pixel-wise segmentation mask. Which gives us a total of 23 layers and 1,941,105 trainable parameters [14].

The computational processing was conducted in the Laboratorio Universitario de Cómputo de Alto Rendimiento (LUCAR) from the Universidad Nacional Autónoma de México (UNAM), with an Ubuntu Server 18.04.2 LTS with the following characteristics: 2x Intel Xeion E5-2640@2.5 GHz with 24 threads. A RAM memory of 64 GB along with a GPU Nvidia Tesla K20.

D. Training and Validation

To verify the reproducibility of the model and validate the results, a 5–Fold Cross–Validation method was used, where 90% of the images was used for training (n=10,357) and 10% for final validation test (n=1,151). The following parameters were chosen to train the network from the ground up: Adam gradient optimizer, binary entropy as loss function and fifty epochs. The metrics used to assess the model were binary accuracy and Jaccard Index. The Jaccard Index (1) is a measure of the relationship between the intersection of area with the area of the union of the original (A) and predicted mask (B).

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad (1)$$

Once pixels were classified, each region of the parasites can be counted considering connected component labeling. Then, two tests were carry out. The first test, named final test, was done with the original images; while the second test, named final test - Reinhard, was done using Reinhard [16] color normalization method given that Tellez *et al.* (2019) recently quantify the positive effects of color normalization on histopathological images for deep learning applications. In order to compare our results with another machine leaning algorithm, an SVM was trained to perform the same segmentation task [17]. The SVM was fed with a set of intensity features, in addition to entropy, median and mean filters with an analysis window of 11x11 pixels. The kernel used was a radial basis function; the SVM hyperparameters were optimized by a grid.

III. Results and Discussion

Table I shows the performance of U-Net and SVM classifiers. In first section (rows 2 and 3), it can be seen that the performance of the U-Net outperform the obtained results by the SVM with 98% vs. 76% for binary accuracy and 49% vs. 41% for Jaccard index. These results are consistent in the final tests, where the U-Net is the algorithm with the best performance (up to 99.7 % for accuracy and 74% according to the Jaccard index). Finally, it can be observed that the incorporation of the reinhard algorithm helps to obtain a better performance in all cases (table I section 2 vs. section 3). In Fig. 2, 2 two examples of nests detection by U-Net are shown. The case with minor error, first row, obtained a difference of nests and predicted area of 3 and 15 pixels respectively, regarding the original mask. On the other hand, the case with major error, had a difference of nests and predicted area of 15 and 3870 respectively.

Given that the aim of this work was to detect T. cruzi amastigotes nests, the U-Net gave good results predicting the masks, but since the number of nests as well as the area of each one of them differs considerably from the original and predicted mask it did not achieved a high value with Jaccard index. Which led to believe that the task of manual mask segmentation should be improved. The accuracy obtained was high, as shown on Table 1, but it can be misleading considering the great amount of pixels categorized as true negatives (background). This due to into an imbalance of classes that simplify the work of the network to recognize the background, which in turn can be seen in the normalized confusion matrix, where the network achieved 99.71% accuracy in predicting true negatives (background).

Something to highlight is that the variability of colors in images decreased the performance of the network, as shown in table I. Another problem faced by the network were the different artifacts that some images had, like smudges

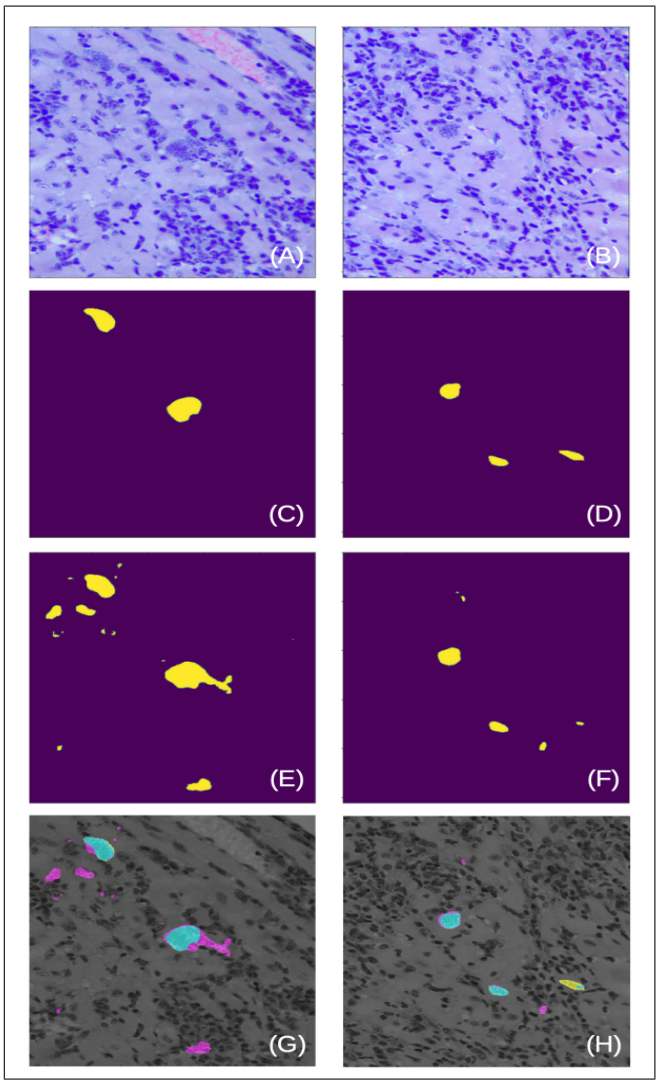

Fig. 2: Representative examples of two segmentations obtained. The worst result is shown in first column, and the best segment obtained in second column. (A-B) Original images; (C-D) Manual annotations; (E-F) Automatic segmentations and (G-H) confusion matrix overlay (Cyan: true positives, Magenta: false positives, Yellow: false negatives and Black: true negatives).

TABLE I: Metrics comparisons: training, final test, final test Reinhard

Validation	**Classifier**	**Binary Accuracy (%)**	**Jaccard Index (%)**
Cross Validation	U-Net	98.19±0.006	49.43±0.009
	SVM	76.26±4.71	41.06±3.21
Final Test	U-Net	99.63	49.57
	SVM	78.86	40.35
Final Test Reinhard	U-Net	99.70	74.30
	SVM	82.05	57.32

and blurry areas. Its important to notice that the network predicted ROIs that were not present in the original masks, but might be indeed T. cruzi's nests, although there wasn't enough certainty during the manual segmentation to mark them.

This led to believe that increasing the number of images and variability would help the network identify better the nests even when different stain colors and artifacts hinder the process. Furthermore, consulting more experts to validate the ROIs would help to include as much nests as possible in each mask. To tackle the class imbalance problem, the images could be divided into quadrants with better nests-background ratio.

IV. Conclusion

This work has proven the efficiency of the U-net architecture in detecting *Trypanosoma cruzi* amastigotes' nests to reduce pathologists workload and increase reliability. Promising results were obtained: accuracy of 99.19% ± 0.00637 and a Jaccard index of 49.43% ± 0.00986, but there are options for improvement as the challenges encountered need to be solved to improve performance. The accuracy of the U-Net at creating a mask seemed to be highly related to the manual segmentation task. Color differences and image quality also play an important role during training, as the comparison between experiments shows. Accuracy does not give too much information when evaluating masks, given the background - nests ratio. As mentioned before, there are not published works reporting on the application of machine or deep learning methods for automated detection of T. cruzi's nests in histopathological images, hopefully in the near future similar works will be found to make a thorough comparison between different architectures and methods.

ACKNOWLEDGMENT

Authors wish to acknowledge technical assistance of Adrian Duran Chavesti for LUCAR server management.

References

[1] Paz María Salazar Schettino et al. "Enfermedad de Chagas en México". In: *Revista de la Facultad de Medicina de la UNAM* 59.3 (2016), pp. 6–16.

[2] Audrey Arnal et al. "Estimating the current burden of Chagas disease in Mexico: A systematic review and meta-analysis of epidemiological surveys from 2006 to 2017". In: *PLoS Neglected Tropical Diseases* 13.4 (2019).

[3] Kárita Cláudia Freitas Lidani et al. "Chagas disease: From discovery to a worldwide health problem". In: *Journal of Physical Oceanography* 49.6 (2019).

[4] José A. De Fuentes-Vicente et al. "Trypanosoma cruzi: A review of biological and methodological factors in Mexican strains". In: *Acta Tropica* 195 (2019), pp. 51–57.

[5] Andréa M. Macedo et al. "Trypanosoma cruzi: Genetic structure of populations and relevance of genetic variability to the pathogenesis of chagas disease". In: *Memorias do Instituto Oswaldo Cruz* 99.1 (2004), pp. 1–12.

[6] Secretaría de Salud. *Prevención y Control de la Enfermedad de Chagas*. 1st. Ciudad de México, 2014.

[7] Maria Carmo Pereira Nunes et al. *Chagas Cardiomyopathy: An Update of Current Clinical Knowledge and Management: A Scientific Statement From the American Heart Association*. Sept. 2018.

[8] Oscar Jimenez-Del-Toro et al. "Analysis of histopathology images: From traditional machine learning to deep learning". In: *Biomedical Texture Analysis: Fundamentals, Tools and Challenges* (2017), pp. 281–314.

[9] Víctor Uc Cetina, Carlos Brito Loeza, and Hugo Ruiz Piña. "Chagas Parasites Detection through Gaussian Discriminant Analysis". In: *Abstraction & Application* 8 (2013), pp. 6–17.

[10] Courosh Mehanian et al. "Computer-Automated Malaria Diagnosis and Quantitation Using Convolutional Neural Networks". In: *Proceedings - 2017 IEEE International Conference on Computer Vision Workshops, ICCVW 2017* 2018-Janua (2017), pp. 116–125.

[11] Marc Górriz et al. "Leishmaniasis parasite segmentation and classification using deep learning". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 10945 LNCS (2018), pp. 53–62.

[12] Allan Ojeda-Pat, Anabel Martin-Gonzalez, and Roger Soberanis-Mukul. "Convolutional neural network u-net for trypanosoma cruzi segmentation". In: *Communications in Computer and Information Science* 1187 CCIS (2020), pp. 118–131.

[13] Connor Shorten and Taghi M. Khoshgoftaar. "A survey on Image Data Augmentation for Deep Learning". In: *Journal of Big Data* 6.1 (2019).

[14] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 9351 (2015), pp. 234–241.

[15] Nahian Siddique et al. "U-Net and its variants for medical image segmentation: theory and applications". In: (2020).

[16] Erik Reinhard et al. "Color transfer between images". In: *IEEE Computer Graphics and Applications* 21.5 (2001), pp. 34–41.

[17] David Tellez et al. "Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology". In: *Medical Image Analysis* 58 (2019).