NIH-NLM-Thin Blood Smears Pf

This dataset includes thin blood smear microscopy images from 193 malaria patients, with 5 images per patient. The data was acquired at Chittagong Medical College Hospital in Bangladesh, where Giemsa-stained, thin-blood smear images were photographed from P. falciparum-infected patients and healthy controls using a smartphone camera. The blood smear images were manually annotated by an expert, de-identified, and archived. The Institutional Review Board (IRB) at the National Library of Medicine (NLM), National Institutes of Health (NIH) granted approval to carry out the study within its facilities (IRB#12972).

All images have three color channels and image dimensions of 5312×2988 pixels. Because images have been captured through the eyepiece of the microscope, the visual region is circular.

We divided the data set into two parts: a polygon set and a point set. The difference between these two sets lies in the annotation method. In the polygon set, all red blood cells (RBCs) and white blood cells (WBCs) have been outlined manually with polygons using the Firefly annotation tool, whereas in the point set, cells have been marked by placing a point on each cell.

Images for the polygon set are stored in:

\NIH-NLM-ThinBloodSmearsPf\Polygon Set\[Patient ID]\img\[ImageName].jpg

Corresponding annotations are stored in the following subfolder:

\NIH-NLM-ThinBloodSmearsPf\Polygon Set \[Patient ID]\GT\[ImageName].txt.

Example for one entry in GT files of Polygon Set:

Patient no Cell no Cell type Annotation shape Number of points Points coordinates [width, height, width, height,...] 1-1,Uninfected,No_comment,Polygon,11,2414.05,1707,2416.4,1669.85,2431.5,1646.55,2455.9,1630.3, 2488.45,1629.05,2516.4,1644.25,2537.3,1681.45,2533.8,1715.2,2505.9,1743.05,2455.9,1745.4,2425.7,17 25.65

Images for the point set are stored in:

\NIH-NLM-ThinBloodSmearsPf\Point Set\[Patient ID]\img\[ImageName].jpg

Corresponding annotations are stored in the following subfolder:

\NIH-NLM-ThinBloodSmearsPf\Point Set \[Patient ID]\GT\[ImageName].txt.

Example for one entry in GT files of Point Set:

Patient no Cell no Cell type Annotation shape Number of points Point coordinates [width, height] 1-1,Parasitized,No Comment,Point,1,3833.3,306.0

The annotations include the positions of RBCs (uninfected + parasitized) and WBCs, in image coordinates. In each annotation file, the first line contains three numbers: The first number is the total number of entries (number of cells) in the text file. The second number is the image width

and the third number is the image height. All following lines are annotations of uninfected, parasitized and white blood cells.

If the second column in a line is "Parasitized," this line represents an annotated RBC with a parasite, whereas, if it is "Uninfected," this line represents an annotated uninfected RBC cell. Otherwise, if the second column is "White_Blood_Cell", this line is the annotation of a WBC. The fourth column indicates whether a RBC or WBC has been annotated as a "Polygon" or "Point". In case of a Polygon, column five contains the number of (x, y) pairs for the boundary annotation of the cell. Column six and seven contain the x and y coordinate for the first point of the cell boundary, respectively. Column eight and nine are the x and y coordinate for the second point, and so on. In case of a point, column six and seven are the x and y coordinate of the point.

Reference:

We request that publications resulting from the use of this data attribute the source (National Library of Medicine, National Institutes of Health, Bethesda, MD, USA) and cite the following publication, which has used the data for cell detection in an architecture called RBCNet:

Yasmin M. Kassim, Kannappan Palaniappan, Feng Yang, Mahdieh Poostchi, Nila Palaniappan, Richard J Maude, Sameer Antani, Stefan Jaeger. *Clustering-Based Dual Deep Learning Architecture for Detecting Red Blood Cells in Malaria Diagnostic Smears*. IEEE Journal of Biomedical and Health Informatics, 2020.

The code for RBCNet can be downloaded here: <u>https://github.com/nlm-malaria/RBCNet</u>

 ${}^{0}_{0} {}^{0}_{0$

Author: Dr. Yasmin M. Kassim Corresponding Author: Dr. Stefan Jaeger Lister Hill National Center for Biomedical Communications National Library of Medicine - National Institutes of Health 8600 Rockville Pike Bethesda, MD 20894,United States

For more information, please contact:

Dr. Yasmin M. Kassim Location: 38A / B1N-28N Phone Number: (301) 827-4730 E-mail: yasmin.kassim@nih.gov or

Dr. Stefan Jaeger Location: 38A / 10N1003O Phone Number: (301) 435-3198 E-mail: stefan.jaeger@nih.gov