



Haemoprotozoan Disease Classification in Cattle using Portable Device

¹ Mr. Shiva Sumanth Reddy, ² Anannya Mahapatra, ³ Akili Chakravaradhan, ⁴ Tushar Reddy, ⁵ Hemanth Reddy

¹ Professor, ² Student, ³ Student, ⁴ Student, ⁵ Student,

¹Computer Science and Engineering,

¹ Dayananda Sagar Academy of Technology and Management, Bangalore, India

Abstract: This survey aims to automate the process of identification and diagnosis of Haemoprotozoan diseases in cattle, using machine learning algorithms to build a model from a dataset of images and diagnoses collected from local labs. The project focuses on three significant Haemoprotozoan diseases that pose a significant risk to cattle all across India: Anaplasmosis, Babesiosis, and Theileriosis. The model is designed to run on a client-server web application, where a picture of the bloodstream sample is sent to the server, which then returns two parameters specifying whether the diagnosis is positive or negative, and the type of disease. This could significantly reduce mortality rates by making the process cheaper and faster. The model will be tested and evaluated several times in the training and testing modules. The evaluation process will be based on the model's accuracy, as well as other factors such as the model's speed and cost-effectiveness. The outcomes are then differentiated from the existing models already in use, to inspect the accuracy and implementation of the model. The final model will be deployed in the client-server web application.

Index Terms - Machine Learning, Haemoprotozoan, Keras R-CNN, Bi-LSTM, Classification, Client-Server.

I. INTRODUCTION

Haemoprotozoan diseases cause a threat to the livestock population about loss of cattle, reduced milk production and lowered animal power. Identification of this kind of illnesses often comes as a difficult task. The impact of these diseases on issues of productivity and well-being is significant and can be resolved with a fair economic evaluation of the underlying economic loss, which India has not yet predicted. We discuss the threat of Haemoprotozoan diseases to livestock in terms of health and productivity. It also examines the challenges of diagnosis and the potential of using deep learning and machine learning to detect and classify these diseases. We suggest that machine learning-based microscopic diagnosis could save time for doctors and reduce the rate of wrong diagnoses and missed diagnoses.

The development of deep learning and machine learning technologies has enabled the rapid diagnosis of Haemoprotozoan diseases more accurately and efficiently. Automated diagnosis using artificial intelligence technology is expected to have a far-reaching impact on disease diagnosis, as well as improved economic efficiency in the livestock industry. Additionally, the use of deep learning and machine learning technologies can contribute to an improved comprehension of the epidemiology of these diseases and improved management of their prevention and control. This improved understanding of the epidemiology of diseases can be attributed to the capacity of deep learning and machine learning technologies to process large volumes of data. Furthermore, using automated diagnosis technologies can speed up and save money on manual diagnosis. Consequently, more resources can be allocated to research and development, which can further reduce the economic burden of Haemoprotozoan diseases.

II. LITERATURE SURVEY

Segmentation for parasite images was given by Zhang, Hao, Hanlin, Xi, Chen, Yubing, Mario, Li, and Yang [1]. This paper reviews the use of identification of protozoan parasites using deep learning from microscopic images of cattle blood smears, highlighting the challenges and future trends of this technology. Two-stage methods have been developed to detect and count red blood cells in microscopy images, and CNNs have achieved high accuracy in classifying infected cells. However, the complexity and data annotation requirements limit their application in developing countries. In clinical diagnosis, it is important to identify the parasite species and the presence of mixed infections. These algorithms have achieved high accuracy and can reduce the workload of pathologists. However, some drawbacks exist, such as the performance of the network being dependent on the input data and the reliance on the shape information of macro-objects.

Segmentation using Circular Hough Transformation was given by Rahmani, Idris, Ramli and Arof [2]. This document discusses the use of the Circular Hough Transform algorithm to detect circular shapes in digital photos. It explains that the algorithm can be used to determine the centre and radius of a circle and that it is an efficient way to identify various components in an image. The Hough transform is used to detect non-circular objects, such as sickle cells, by finding intersections between circles. If the circles overlap, the thin contour of the circle of interest is monitored to produce a different circle. The intersections of the new and original circles are then noted. The RBC proposed a method for identifying crescent shapes using the circular Hough transform. After testing

with a light microscope, the tool had an overall accuracy of 92%, with 70% sensitivity and 72% specificity. The deformable detection feature can be improved in the future to identify more shapes than just crescent shapes. This technique is the best for segmenting and classifying anomalies in Haemoprotozoan cells.

The Canny edge detection method for retrieving medical images was presented by Sanaa Jadwa [3]. This research proposes an effective technique for extracting shape features from photos using the Canny edge detection operator. Edge detection is a group of mathematical methods used to identify points in digital images where the brightness abruptly changes, allowing for a reduction of data while maintaining the image's structural integrity for further processing. Canny edge detection [4] is an algorithm that finds the best-defining edges in an image by minimizing mistake rates, optimizing location, and marking edges only once when there is only one edge present. This algorithm uses the first derivative of a Gaussian function to approximate the best filter for these requirements. The Canny edge detection approach is a multi-step procedure that uses noise suppression to find edges in a picture. It has been found to be superior to other methods, according to experimental findings.

Pre-processing of blood smear images is done by using filters which are given by Saban Ozturk & Bayram Akdemir [5]. Pre-processing algorithms have a significant effect on using a Convolutional Neural Network (CNN) [6] structure for classifying histopathological images. This study tested three different pre-processing techniques and found that the normal pre-processing algorithm produced the best results, as the over-pre-processing algorithm removed many important features from the image. Pre-processing algorithms are used to highlight important features of an image and remove noise. Normal pre-processing involves subtracting the median value of the view from the original image, while over-pre-processing deletes all information about the image background texture. Results showed that normally preprocessed images produced more accurate results than the other preprocessed images, indicating that pre-processing techniques such as cell enhancement, background noise reduction, thresholding, morphological operations, and adaptive Histogram equalization are effective.

To extract features from the processed image we make use of the FSA algorithm given by Emad, Mostafa, Behrouz Naugler, and Christopher [7]. Image segmentation, feature extraction, and pattern classification are steps in the automated analysis of images of peripheral blood smears. ANN, SVM, decision trees, and K-nearest neighbour are a few examples of the supervised and unsupervised classifier learning methods that are mentioned. This document provides an overview of techniques used for blood smear image segmentation, including multispectral imaging, watershed algorithm, support vector machine, and artificial neural network. It discusses watershed clustering and features scale-space filtering for cell segmentation. It also examines the use of the Support Vector Machine (SVM) in hematopathology, specifically for acute lymphocytic and normal WBCs leukaemia detection. Several algorithms have been used to reduce the SVM training dataset and computational time, but further improvement is still needed. Feature Selection Algorithms (FSA) are used to make the provided data's dimensions lower for a classifier. Filter-type FSA ranks features according to their predictive power, while wrapper-type FSA evaluates feature subsets in light of how well they work with a particular classifier. Two algorithms, Sequential Forward Selection (SFS) and Sequential Backward Selection (SBS), are discussed.

To classify the given three Haemoprotozoan diseases we need to train a model which is given by Ümit, Zafer, and Zryan [8]. This presents an end-to-end design for breast cancer detection based on a Fully Convolutional Network (FCN) and a Bidirectional Long Short-Term Memory (Bi-LSTM). The BreakHis database was used to assess the outcome of the proposed approach, which was discovered to be more efficient than findings from earlier studies. This study explored the use of pre-trained CNN architectures to classify breast cancer histological image data. The combination of CNN and logistic regression produced the best results with a classification with a greater success rate. Additionally, the study proposed an end-to-end BC diagnosis approach based on FCN and Bi-LSTM, which deviates from traditional deep learning techniques. The FCN was used to process large-scale images and enable varying input sizes, while the Bi-LSTM was employed to solve time series problems.

To implement hybrid feature selection, we need to use an algorithm called Relief-PCA which is described by Divya Jain, and Vijendra Singh [9]. Feature selection is a pre-processing method used to decrease the number of features in the dataset. It is being used in the medical field to identify signs related to a particular disease and to aid medical practitioners in making faster and easier decisions with the help of high-throughput technologies and machine learning methods. Classification techniques and feature reduction techniques are being used for the effective diagnosis of diseases. This article presents a hybrid feature selection approach combining ReliefF and Principal Component Analysis (PCA) methods to decrease the number of features in a dataset and improve the performance of a classifier. ReliefF is an ordering method that updates weights on nearest hits and misses, while PCA is a dimensionality reduction method that extracts the maximum variance from the variables. The ReliefF-PCA feature selection method was applied to a chronic disease dataset and evaluated by comparing the performance of a classification model with and without feature selection. The results of the comparison were then analyzed to determine the effectiveness of this method compared to other feature selection methods. The comparison yielded the best results this method, making it the most effective feature selection approach for the given dataset and classification model.

To understand the classification of images, R-CNN is one of the methods that can be implemented given by Jane, Allen, and Deepali [10]. To quickly process large image sets for cell identification we make use of Keras R-CNN. This approach is more effective than conventional segmentation techniques because it can deal with noisy pictures, changes in light, and clumped cells more effectively and doesn't require as much preparation, algorithm selection, or parameter adjustment. Object detection algorithms, which use deep learning to identify features and combinations of features in unprocessed pixels of images, can quickly and accurately identify individual cells and their phenotypes in a single step. This is much faster than pixel-level segmentation, making it an ideal solution for tasks such as digital pathology, high-throughput drug screening, and basic biology research. R-CNN is a method of creating bounding boxes and classifications for an image. The model is used for prediction of the image after being trained on a batch of images with known bounding box coordinates and labels. By contrasting the anticipated outcomes with the recognised ground truth annotations, the model's accuracy is assessed.

Frame-based implementation for the segmentation of images is given by Aiche chian, Bin Dong, and Zuowei Shen [11]. Medical picture segmentation is a challenging issue that has been tried to solve in this paper. It tells us about a frame-based model and a quick implementation for generic medical image segmentation issues in this study. The suggested frame-based model performs better than the total variation-based model in terms of capturing important characteristics of biological structures, according to numerical studies. Additionally presented is the effective segmentation of blood arteries and aneurysms in 3D CT angiography pictures. demand a significant investment in time and computing power. The widespread use of deep learning for parasite identification is constrained by the dearth of publicly accessible datasets. Tight frame redundancy typically results in a sparse approximation of images, which is recognised to be a desirable quality for challenges involving image restoration, like denoising, inpainting, deblurring, etc. Segmentation of medical images allows us to reduce the complexity of images thereby enhancing the image for further analysis.

Understanding of segmentation which is a part of image pre-processing is given by Salwa-Abdul Latif, and Mohanad Salman [12]. The majority of object identification and image recognition tasks require the use of frame segmentation, and the effectiveness of this process determines how well these tasks are classified. The process of segmenting an image into distinct, homogenous segments is known as image segmentation; as a result, the efficacy of pattern recognition is increased when an image is represented in straightforward, easy-to-understand forms. Pointing out the various image segmentation techniques available, along with their various advantages and disadvantages. Analyzing different segmentation techniques according to their use case, along with their particular niche. Pointing out various real-life applications for the techniques and operations. Not enough analysis was done to compare data and performance differences between multiple models for a particular use case. Showcases that edge & region-based segmentation analysis is the way to go to differentiate and identify different Haemoprotozoan cells.

III. METHODOLOGY

The proposed model consists of six phases, data collection-real time blood smear samples of cows, data pre-processing, segmentation, feature selection, classification and real-time client-server implementation. The first five stages belong to image processing and machine-learning solution and the last step is real-time application implementation. The block diagram is given below for the proposed system:

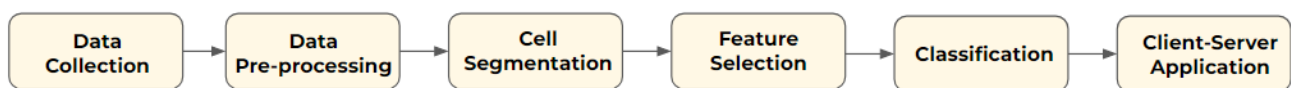


Figure. 1 Block diagram of proposed metadata Collection

3.1 Data Collection

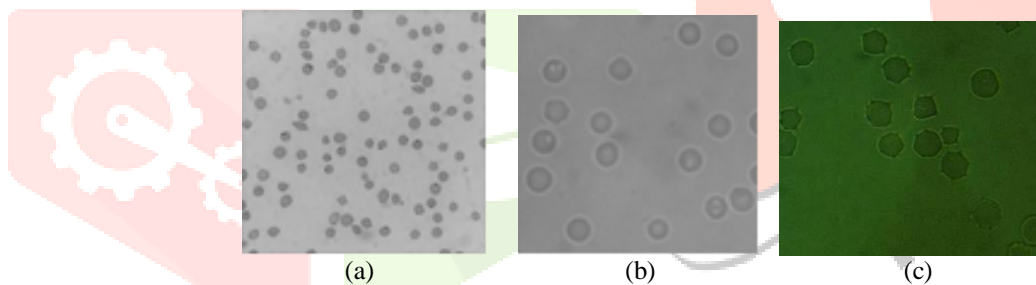


Figure. 2 Collected blood smears: (a) Anaplasmosis
(b) Babesiosis (c) Theileriosis

Data collection is an important step in the machine learning process, as it determines the size and quality of the dataset used to train the model. In this study, cell images were gathered from labs that analyse blood samples, and used to detect the three Haemoprotozoan illnesses. Gathering and organising data takes longer than simply training a model on it. The process of obtaining pertinent information and organising it to provide datasets for machine learning is known as data collection. In this study, we obtain samples of cell images from the labs that analyse and gather bovine blood samples. Anaplasmosis, Babesiosis, and Theileriosis are the three Haemoprotozoan illnesses represented in these samples.

3.2 Data Pre-Processing

Preprocessing of image data transforms it into a format that makes it simple for machine learning algorithms to solve and detect features. It is employed to lessen the complexity and increase the accuracy of a model. Preprocessing allows us to reduce unwanted distortions and increase application-specific characteristics.

3.2.1 Color Normalization

In images taken, the distribution of color depends on the lighting, which changes greatly in real-world settings. It is expected that object recognition based on pixel information will account for these impacts. There have been several documented algorithms for color constancy or color normalizing. We quickly go through a few of these: whitening, color normalization across all color spaces, and color rotation.

3.2.2 Gaussian Filter

A Gaussian filter is combined with the relevant image in order to even it. In order to decrease the visibility of clear noise present on the edge detection method being used, this step will slightly smooth the image. The equation for a size-dependent Gaussian filter is given as $(2k+1) (2k+1)$. It is needed to use the two-dimensional Gaussian function provided by (3.1)

$$g(x, y) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/(2\sigma^2)} \quad (3.1)$$

3.2.3 Sobel Filter

A pair of 3x3 convolution masks, G_x and G_y , are used by the Sobel edge detector; one estimates the gradient in the x-direction and the other in the y-direction. The Sobel operator accentuates areas of high spatial frequency that correlate to edges by performing a 2-D spatial gradient measurement on an image [13]. Combining these will then reveal the absolute gradient magnitude and orientation at each point. The given magnitude of the acclivity can be calculated by (3.2) as shown below:

$$|G| = \sqrt{G_x^2 + G_y^2} \quad (3.2)$$

3.2.4 Histogram Equalization

A graph that displays the distribution of intensity in a picture is called a histogram. It basically indicates the number of pixels for each intensity value that is taken into account. To increase contrast in images, a technique known as histogram equalisation is utilised [14].

The transformation equation is given as (3.3):

$$S_k = T(r_k) = (L - 1) \sum_{j=0}^k p_r(r_j) \quad (3.3)$$

Where S_k and r_k are output and input pixel intensities. L is the maximum intensity value and $p_r(r_j)$ is the probability of intensity level r_j in the image.

3.3 Cell Segmentation

Identification of the shapes of distinct components in an image is the main use of the Circular Hough Transform method. In general, it is an extended form that can distinguish between several circle forms in an image. The edges can be found using the Canny Edge Detector, which is utilised as an input for the Circular Hough Transform's extraction of the nucleus cells.

Using a simple feature extraction method, circles may be located in flawed images. The circle candidates are generated by "polling" in the Hough parameter space and selecting the local maximum from an accumulator matrix. The following is an example of how to express a circle in two dimensions.

3.3.1 Circle Parameters

Parametric Equation for a circle is (3.4):

$$(x - a)^2 + (y - b)^2 = r^2 \quad (3.4)$$

The equation will primarily discuss the three parameters a , b , and r . Here the centres of the x and y directional circles are a and b . Using the following Equations, the parameter circle is represented: (3.5), (3.6)

$$x = a + r \times \cos(\theta) \quad (3.5)$$

$$y = a + r \times \sin(\theta) \quad (3.6)$$

3.3.2 Canny Edge Detection

The amount of data that needs to be processed can be significantly decreased while still obtaining valuable structural information from various visual objects with the help of the Canny edge detection technology. To recognise fundamental shapes, a technique known as Canny Edge detection is employed. The basic standards for edge detection are as follows:

- Edge detection with a low error rate, which means that every edge that can be seen in the image should be accurately detected.
- The centre of the edge should be accurately localised by the operator's edge point detection.
- Picture noise should, whenever feasible, prevent the creation of fake edges, and an image edge should only be marked once.

The blurred image's edges might point in a variety of directions; therefore, the Canny algorithm uses four filters to distinguish between the horizontal, vertical, and diagonal edges. The edge detection operator (G_y) returns values for the first derivative in both the horizontal (G_x) and vertical (G_y) directions. This makes it possible to identify the edge gradient and direction (3.7):

$$\theta = \text{atan2}(G_y, G_x) \tag{3.7}$$

Where atan2 is the arctangent function with two inputs, and G can be calculated using the hypotenuse function.

3.4 Feature Selection

Prior to using any machine learning technique, features are chosen. This removes attributes from the dataset that are unnecessary and have no bearing on the categorization process. Relief is an algorithm that chooses features using a filter-method technique. Using the segmentation resultant one-dimensional vector generates a feature score that is then used to rank and choose the highest-ranking attributes for feature selection.

Table. 1 Features considered for processed image

Feature	Description
Area	Total number of pixels in a given boundary region
Perimeter	the radius of the area of interest
Eccentricity	An ellipse whose eccentricity is specified by a scalar.
Equivalent Diameter	the diameter of a circle whose area is equal to that of the area
Extent	scalar that represents the ratio of the pixels in the region to all of the pixels in the bounding box.
Convex Area	scalar that indicates the convex image's pixel count
Filled Area	Scalar indicating how many pixels are on in the filled image

Here, 20% of the dataset is used for testing while 80% is used for training. [15] We make use of the ReliefF algorithm that makes use of vector data from the previous step. The following is ReliefF algorithm equations (3.8):

$$W[x] = M + H_q \tag{3.8}$$

where,

$$H = - \sum_{i=1}^k D((x, r_i, H_k) / k \tag{3.9}$$

$$M = \sum_{C \neq cl(r_i)} [\left(\frac{P(C)}{1 - P(cl(r_i))} \right) \sum_{i=1}^k D(x, r_i, M_i(C))] / k \tag{3.10}$$

Their benefits include applying binary or continuous data, executing in low-order polynomial time [16], and being noise-tolerant and resilient to feature interactions. There aren't enough training instances to fool the algorithm, and they are unable to discern between redundant features.

3.5 Classification

The practice of classifying and labelling groups of pixels within an image according to predetermined rules is known as image classification. The two classification techniques are "supervised" and "unsupervised."

Bi-LSTM is used to classify images in order to distinguish between three diseases. It differs from a standard LSTM where input flows in two directions when bidirectional. We may make input flow in one way, either backwards or forward, using the standard LSTM [17]. Bi-directional input can be configured to flow in both directions in order to keep all current and historical data.

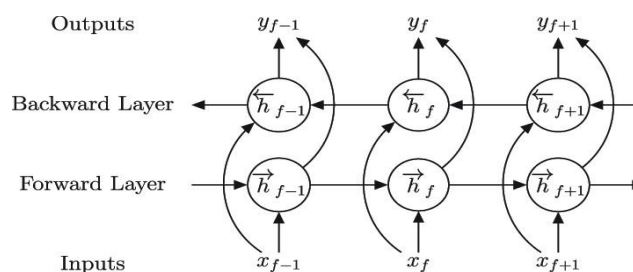


Figure. 3 bi-LSTM process

3.6 Client–Server Implementation

A client-server application is a software application distributed among two or more computers connected through a network. The client computer sends requests to the server, which performs some action and returns the results to the client.

The client-server model is a common architecture for many types of software applications, including web applications, database applications, and file-sharing systems. In this model, the client computer initiates a request for information or a service, and the server responds to the request.

There are several advantages to using a client-server architecture. One advantage is that it allows for better resource management, as the server can perform tasks and provide services to multiple clients simultaneously. Another advantage is that it allows for easier maintenance and updates, as the server can be updated independently of the client.

A phone camera or a web application is used to capture the image captured by the microscope. The suggested machine learning model processes and classifies the image after which the server provides the response back to the application. This integration of the machine learning model to the application is done using Flask from the Python library

IV. METHODOLOGY

Haemoprotozoan diseases are transmitted by protozoan parasites that infect animal blood cells, including cattle. These diseases can have serious consequences for the health and productivity of infected animals, and it is important to accurately diagnose and treat them promptly.

The diagnosis and treatment of Haemoprotozoan infections in cattle can benefit from the use of portable instruments for disease classification. Such devices can allow for rapid and accurate diagnosis in the field, which can be particularly important in large herds or in situations where access to laboratory facilities is limited. However, the effectiveness of a portable device for Haemoprotozoan disease classification in cattle will depend on the specific characteristics of the device and the accuracy of its diagnostic algorithms. Further research may be needed to assess the possibility of portable technology for this application.

With the help of the portable device, it outlines a system for identifying Haemoprotozoan diseases in cattle using image processing and machine learning. The system makes use of algorithms to identify circular edges and anomalies caused by the infection, allowing for accurate classification even with a limited dataset.

REFERENCES

- [1] Zhang C, Jiang H, Jiang H, Xi H, Chen B, Liu Y, Juhas M, Li J, Zhang Y. Deep learning for microscopic examination of protozoan parasites. *Comput Struct Biotechnol J*. 2022 Feb 11; 20:1036-1043. DOI: 10.1016/j.csbj.2022.02.005. PMID: 35284048; PMCID: PMC8886013.
- [2] A M Rahmani, Mohammad & Idris, Mohd & Ramli, R. & Arof, Hamzah. (2015). Segmentation of Crescent Shape Blood Cells using Circular Hough Transform. *Indian Journal of Science and Technology*. 8. 10.17485/ijst/2015/v8i12/70659
- [3] Dr. Sana'a khudayer Jadwa, "Canny Edge Detection Method for Medical Image Retrieval" *International Journal of Scientific Engineering and Applied Science (IJSEAS)- Volume-2, Issue-8, August 2016*
- [4] "Canny Edge Detector". Wikipedia. https://en.wikipedia.org/wiki/Canny_edge_detector/ (accessed Dec. 23, 2022)
- [5] Şaban Öztürk, Bayram Akdemir, Effects of Histopathological Image Pre-processing on Convolutional Neural Networks, *Procedia Computer Science*, Volume 132, 2018, Pages 396-403, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2018.05.166>.
- [6] Hung, Jane & Goodman, Allen & Ravel, Deepali & Lopes, Stefanie & Rangel, Gabriel & Nery, Odailton & Malleret, Benoît & Nosten, Francois & Lacerda, Marcus & Ferreira, Marcelo & Renia, Laurent & Duraisingh, Manoj & Costa, Fabio & Marti, Matthias & Carpenter, Anne. (2020). Keras R-CNN: library for cell detection in biological images using deep neural networks. *BMC Bioinformatics*. 21. 10.1186/s12859-020-03635-x.
- [7] Emad A. Mohammed, Mostafa M.A. Mohamed, Behrouz H. Far, Christopher Naugler. "Peripheral blood smear image analysis: A comprehensive review", *Journal of Pathology Informatics*, 2014. 5. 9. 10.4103/2153-3539.129442.
- [8] Ümit Budak, Zafer Cömert, Zryan Najat Rashid, Abdulkadir Şengür, Musa Çıbuk. "Computer-aided diagnosis system combining 8 1% 9 1% 10 1% 11 1% FCN and Bi-LSTM model for efficient breast cancer detection from histopathological images", *Applied Soft Computing*, 2019, 105765. doi:10.1016/j.asoc.2019.105765
- [9] Jain, Divya & Singh, Vijendra. (2018). Diagnosis of Breast Cancer and Diabetes using Hybrid Feature Selection Method. 64-69. 10.1109/PDGC.2018.8745830.
- [10] Hung, Jane & Goodman, Allen & Ravel, Deepali & Lopes, Stefanie & Rangel, Gabriel & Nery, Odailton & Malleret, Benoît & Nosten, Francois & Lacerda, Marcus
- [11] Dong, Bin & Chien, Aichi & Shen, Zuwei. (2010). Frame-based segmentation for medical images. *Communications in mathematical sciences*. 9. 551-559. 10.4310/CMS.2011.v9.n2.a10.
- [12] Abdulateef, Salwa & Salman, Mohanad. (2021). A Comprehensive Review of Image Segmentation Techniques. *Iraqi Journal for Electrical and Electronic Engineering*. 17. 166-175. 10.37917/ijeee.17.2.18.
- [13] "Sobel Filter" <https://homepages.inf.ed.ac.uk/rbf/HIPR2/sobel.htm> (accessed Dec. 23, 2022)
- [14] Kang atul, "Histogram Equalization," *TheAILearner*, Apr.2019. <https://theailearner.com/2019/04/01/histogram-equalization/> (accessed Dec. 21, 2022).

- [15] S. Reddy and N. Channewgoda, "Edge Boost Curve Transform and Modified ReliefF Algorithm for Communicable and Non-Communicable Disease Detection Using Pathology Images," *International Journal of Intelligent Engineering and Systems*, vol. 14, no. 2, pp. 463–473, Apr. 2021, DOI: 10.22266/ijies2021.0430.42.
- [16] S.S. Blessy Trencia Lincy, N. Suresh Kumar. "An enhanced pre-processing model for big data processing: A quality framework" , 2017 International Conference on Innovations in Green Energy and Healthcare Technologies (IGEHT), 2017
- [17] "Long short-term memory". Wikipedia. https://en.wikipedia.org/wiki/Long_short-term_memory/ (accessed Dec. 23, 2022)

