

Measuring the Performance of Computer Vision Systems in Malaria Studies

<https://doi.org/10.3991/ijoe.v16i11.16035>

Juan V. Lorenzo-Ginori ^(✉), Lyanett China-Valdés
Universidad Central “Marta Abreu” de Las Villas, Santa Clara, Cuba
juanl@uclv.edu.cu

Niurka Mollineda-Diogo, Alfredo Meneses-Marcel
Centro de Bioactivos Químicos, Santa Clara, Cuba

Abstract—Digital image processing-computer vision (DIP-CV) systems are used to automate malaria diagnosis through microscopy analysis of thin blood smears. Some variability is observed in the experimental design to evaluate the statistical measures of performance (SMP) of such systems. The objective of this work is assessing good practices when using SMP to evaluate DIP-CV systems for malaria diagnosis. A mathematical model was built to characterize diagnosis using DIP-CV systems and used to obtain curve families showing the relationships among various SMP of these systems, both using theoretical equations and computer simulation. Curve families showing (a) the relationships among the minimum number of positive erythrocytes (RBCs) to be observed, the per object (RBC) sensitivity and the probability to detect at least one positive, (b) per specimen sensitivity vs. total number of RBCs observed for a typical per object sensitivity and a range of parasite densities (c) per object positive predictive value vs. per object specificity for a typical per object sensitivity and various parasite densities. When determining the per specimen sensitivity, the parasite density p showed to have more influence on the number of RBCs that must be analyzed than the per object sensitivity. Measuring p accurately depends heavily upon the per object positive predictive value of the classifier. For low p values, this would require very high per object specificity and a high enough value of observed RBCs to measure this accurately.

Keywords—Malaria, Plasmodium, Digital Image Processing, Computer Vision, Statistical Measures.

1 Introduction

Malaria continues being one of the largest health problems faced by our planet, where in 2018 228 million cases of malaria occurred worldwide, with 405 000 deaths, mostly children under 5 years old and with a high prevalence in the Sub-Saharan Africa region [1]. An effective way to diagnose malaria and to determine the infection rate is the analysis of thin blood smears in a microscope, during which the infected erythrocytes (red blood cells, RBCs) are detected and can be counted to determine the parasite

density p which will designate in this work the proportion of infected RBCs. This procedure constitutes, for the human experts, a time-consuming task, prone to intra- and inter-analysts' errors due to tiredness, subjectivity and lack of experience. As a consequence, research on the application of digital image processing-computer vision (DIP-CV) systems to analyze the thin blood smears during malaria studies, constitutes a current topic that has produced in the last few years numerous scientific publications. However, the issue of the practical meaning of the statistical measures of performance (SMP) of these systems and the influence on them of the errors due to various sources usually present, has not been considered with enough interest and deserves a more detailed analysis, which is the object of this work. At this point, we emphasize that the performance of binary detection of the parasites is of paramount importance in any system, both due to its intrinsic importance when calculating the parasite density and to its role as a usual first stage in systems devoted to classify also the *Plasmodium* species and their life stages. Analyzing a representative set of published papers reveals the lack of uniformity with which the system's effectiveness is evaluated. This problem is increased due to the lack of available, public annotated databases that could be used for this purpose.

An early work [2] uses covariance, correlation coefficients and distances to compare their DIP-CV system results with those from two human analysts, with a total number of RBCs in the order of 180; however, the standard SMP measures were not used there. Linear regression was used in Ref. [3] to determine the correlation between the outcomes of a proposed system and human analysts on sets having up to 16000 RBCs, reporting good results in terms of correlation with some dependence upon the experimental conditions and also without expressing the results in terms of SMP. Most references, however, use the standard SMP like sensitivity (S_e), specificity (S_p) and predictive positive value (PPV), for which a relatively wide range of values has been obtained. An example using a neural network classifier in Ref. [4] reports $S_e = 0.8513$ and $PPV = 0.8084$ in a set of RBCs that comprised 481 true positives. This work recognizes the influence of p on the evaluation of S_e as well as of the false positives (FP) in the accuracy (ACC) when evaluating p , and used an elementary simulation to assess the system's performance. A DIP-CV system that uses a neural networks classifier in Ref. [5] obtained average $S_e = 0.76$ and $PPV = 0.82$ with a limited number of RBCs. Another work [6] reports values of $S_e = 0.724$, $S_p = 0.976$, $PPV = 0.858$ and $ACC = 0.933$ for their experiments using 4100 RBCs (669 infected) to compare three classifier systems for binary detection. This work also addresses the classification of species and stages of the parasite and includes a thorough analysis of the importance and the effects of the parasite density value p on the estimations of the SMP and p itself. However, as will be shown later, this analysis deserves an extension to obtain new insights on these relationships. Ref. [7] worked with 1000 RBCs showing a maximum $S_e = 0.98$ and presumably $PPV = 0.98$ (the definition given there for precision is unclear), without considering the influence of p . Ref. [8] reports $S_e = 1$ and $S_p = 0.5$ to 0.88 without providing enough data on the experimental conditions of this evaluation.

Experiments with 888 RBCs using 148 in each class (normal and infected) for training five different classifiers reported in [9] shows as the best results obtained (using

neural networks) $S_e = 1$, $S_p = 0.9874$, $PPV = 0.9973$ and $ACC = 0.9673$. This paper also tabulates for comparison the SMP of a number of works where the problem that gives rise to our work can be appreciated. Ref. [10] made experiments with 12,577 erythrocytes, out of which 713 were infected ($p=0.0568$), with the final purpose of classifying the species of the parasite and comparing three classifiers, obtaining values of $S_e = 0.94$ and $S_p = 0.997$, and calculating the F-measure which reflects the balance between S_e and PPV . Research using a K-means classifier and 118 RBC images reported in [11] obtained $S_e = 0.93$ and $PPV = 0.95$. Ref. [12] reports a Support Vector Machine (SVM) classifier using samples from patients with p estimated in the range 0.07-0.16 and various numbers of RBCs (up to 45096), obtaining $S_e = 0.849$, $S_p = 0.999$, $PPV = 0.742$ and a 95% confidence interval for these estimations. A test for a parasite detector classifier in Ref. [13] used 7000 RBCs, with 359 out of them contaminated ($p=0.05$); this work reports $S_e = 0.9972$, $S_p = 0.9994$ and $PPV = 0.989$, which are notoriously high values that could deserve further tests. Another research [14] which used an SVM classifier obtained $S_e = 0.9894$, $S_p = 0.9612$ and $ACC = 0.9866$. However, information on the number of RBCs tested or p value is not provided. The problem of determining the parasite density was addressed in Ref. [15] and in this case the authors provide only the accuracy obtained, $ACC = 0.9646$, as well as regression curves and tables of parasite counts per slide showing the degree of agreement with human-made counts.

An approach using an image analysis system [16] was developed based in a 16-layer convolutional neural network (CNN) model. The binary classification performance of this system was evaluated by means of a ten-fold cross-validation using 27578 single cell images with a 1:1 ratio of infected cells to uninfected cells. The results reported in this work were $S_e = 0.9699$, $S_p = 0.9775$, $PPV = 0.9773$ and F1 score = 0.9736.

A number of 138 quantitative features related to color, morphology and texture from segmented erythrocytes are proposed to detect malaria in Ref. [17] and tested using four feature selection methods and three classifiers, among which the best results were obtained using Correlation-based Feature Selection (CFS) together with a C4.5 classifier. These were $S_e = 0.992$, $S_p = 0.996$. This work also used a balanced erythrocyte database with 500 normal and 500 infected (divided in five groups of 100 infected with two species: 200 *Plasmodium falciparum* and 300 *Plasmodium vivax*) in two and three sets of 100 erythrocytes respectively at different stages (rings, gametocytes and schizonts). This work presents a comparison to other image processing algorithms for malaria detection in a table in which the variability in the use of different SMP and datasets can be appreciated.

A comprehensive review is presented in Ref. 18 on malaria parasites detection systems based in morphological image processing. In this article the authors summarized the results of 27 different approaches, which in regard of erythrocyte classification were expressed in terms of the standard SMP. However, it is to be noticed that in the various works analyzed, these measures of performance have been obtained under different experimental conditions in terms of the number of erythrocytes involved, the level of class balance and the variability of image acquisition conditions, which limits the value of the corresponding SMP values to allow a fair comparison of their effectiveness.

SVM with different kernels were employed as classifiers in Ref. [19], using as features the histogram statistics as well as the gray level co-occurrence matrix (GLCM) and the gray-level run-length matrix (GLRLM) as texture features. SVM with cubic kernel obtained the best results reporting $S_e = 0.951$, $S_p = 1$, $ACC = 0.974$, $PPV = 1$ and $NPV = 0.949$. The dataset employed consisted in 975 isolated erythrocytes segmented out of 46 digital microscopy images of thin blood smears.

A custom convolutional neural network (CNN) and pre-trained models: VGG-19, Squeeze-net, Inception ResNet-V2 and All-ensemble, were implemented and evaluated in Ref. [20] to classify RBCs into normal and parasitized in thin blood smears. The models in this case were evaluated in terms of SMP using accuracy, area under ROC curve (AUC), MS error, positive predictive value, F score and Matthews correlation coefficient. A dataset composed of 27558 RBCs was evenly divided into normal and infected, obtaining as best results: $ACC = 0.9932$, $AUC = 0.9931$, $MSE = 0.9$, $PPV = 0.9956$ and $F = 0.9908$.

Table 1 summarizes in chronological order the results reported in the previous representative sample of articles reporting the results of DIP-CV systems in the analysis of microscopy images of thin blood smears during malaria studies. In some cases, the SMP have been extended by calculating, from the data provided, measures not explicitly given in the corresponding article.

A review article [21] analyzes a large number of papers (173 references) describing techniques involved in machine learning and image analysis for the detection of malaria. Here the authors recognize explicitly the difficulty that gives rise to our article: they consider as a very difficult task to compare the performance of the published systems, due to the variability of blood slides coming from different origins as well as of the methods to prepare slides and to acquire the images. They mention also the fact that the evaluation sets are often too small or too limited as well as the lack of publicly available image benchmark datasets. Having all these in mind the authors considered that there was not a reliable basis to include in their article any formal comparative evaluation based in SMP. In this short review, the papers analyzed were chosen both to illustrate some of the most recent work and previous works of noticeable relevance as well. The values of the SMP reported in these articles reflect that there is also variability in the number of decimal places with which these values have been given.

It is worth to mention that the laboratory assays made to develop anti-malarial drugs, using the rodent species *Plasmodium berghei*, usually generate large numbers of images in which determining accurately the parasite density is of crucial importance. This makes them an important target for the application of DIP-CV techniques.

The purpose of this work is to emphasize the importance of making an appropriate evaluation of the DIP-CV algorithms used in malaria studies, to provide further insights about the practical meaning of the SMP when making such evaluations and to provide some clues to set the limits of validity of the measured parameters. The approach followed was to develop a mathematical model that allowed extensive simulations of the RBCs classification process. This model encompasses all the sources of error in a typical DIP-CV system, making abstraction of their physical nature. Then representative values of the SMP were obtained and an analysis was made in order to clarify their actual meaning when assessing the limitations of a given system.

Table 1. Statistical measures of performance frequently used, as they appear reported in various articles showing results of DIP-CV systems in the analysis of microscopy images of thin blood smears to evaluate the presence of plasmodium.

Ref./date	RBCs	p	Seo	Spo	PPVo	ACC
2/2002	180	N/A	N/A	N/A	N/A	N/A
4/2006	N/A	N/A	0.8513	N/A	0.8084	N/A
3/2007	16000	N/A	N/A	N/A	N/A	N/A
10/2009	12577	0.0568	0.94	0.997	0.949	0.994
6/2010	4100	N/A	0.724	0.976	0.858	0.933
13/2013	7000	0.051	0.9972	0.9994	0.989	N/A
11/2014	118	N/A	0.93	0.95	N/A	N/A
12/2014	N/A	0.07-0.16	0.849	0.999	0.742	N/A
5/2014	N/A	N/A	0.76	N/A	0.82	N/A
7/2015	1000	N/A	0.998	N/A	0.98	N/A
9/2015	888	N/A	0.9986	0.9594	0.9919	0.9605
14/2015	N/A	N/A	0.9894	0.9612	N/A	0.9866
15/2015	N/A	N/A	N/A	N/A	N/A	0.9646
8/2016	N/A	N/A	1	0.5-0.8	N/A	N/A
16/2016	25578	0.5	0.9699	0.9775	0.9773	N/A
17/2017	1000	0.5	0.992	0.996	N/A	N/A
19/2019	975	0.525	0.951	1	1	0.974
20/2019	25578	0.5	N/A	N/A	0.9956	0.9932

At this point, some distinctive properties of DIP-CV systems in the analysis of malaria are to be mentioned:

When the microscopy slides are analyzed by human experts, it is assumed that the per-specimen sensitivity depends only on the parasite density, and the per-object (per individual RBCs) sensitivity and specificity are deemed as 100 percent. On the contrary, a DIP-CV system has inherently various sources of false positives and false negatives: noise and artifacts due to imperfections in preparing the slides and in the image acquisition processes, limitations of image processing algorithms like filtering and segmentation, limited discriminating power of the features used and the effectiveness of the classifier algorithms. Studying the effects of these errors is one of the main purposes of this work.

In the visual analysis by human experts, it is very laborious to make an exact accounting of the RBCs. This determines the use of approximate methods like counting only the infected RBCs and the leukocytes (WBCs, white blood cells), estimating a proportion between the amounts of total infected RBCs (n_p) and observed WBCs (n_w) and calculating the parasite density p in infected RBCs/ μ l from these data using:

$$p = \left(n_p / n_w \right) d_w \tag{1}$$

where the WBC density is considered approximately equal to $d_w = 6000$ WBC/ μ l of blood. A DIP-CV system usually does not have this limitation: it can perform a complete and fairly exact accounting of the RBCs within its limits of precision, allowing using just the necessary RBCs to have a good evaluation of p as the fraction of infected

RBCs. This makes important to know what would be actually the minimum numbers of RBCs necessary to be analyzed for evaluating fairly an expected range of p values.

In laboratory experiments with rodents, the per-specimen SMP are very important because they indicate the margins within which an infected specimen would have a satisfactorily low probability of being considered free of the parasite and conversely a non-infected specimen would not be deemed as infected. This is of course also important in diagnose tasks with humans.

In this work, the following issues were addressed: (a) determining the minimum number of positive objects (infected RBCs) to be analyzed for evaluating p , (b) the relationship between per-object and per-specimen sensitivity and (c) the influence of false positives on the precision with which p can be measured.

The core of the analysis presented here is using the Bernoulli process as a model to describe the behavior one RBC classification in a DIP-CV system used for malaria studies, from which analyzing a set of M RBCs follows the binomial probability distribution. This can be justified by looking at the properties of such processes in Ref. [22] which are summarized below:

1. The experiment consists of repeated trials which are the individual analyses of independent, previously segmented, erythrocytes.
2. Each trial results in a binary outcome: an RBC is infected (positive) or non-infected (negative).
3. The RBCs are analyzed at random, which justifies the assumption that the probability p of finding a positive is constant for any observed RBC.
4. The repeated trials are fully independent processes: a given result does not have influence on the rest.

The remainder of this article is organized as follows. In Section “Materials and methods” the models and experimental design used in the simulation experiments are explained in detail. Section “Results” shows the results obtained from the equations describing the theoretical model and those resulting from the simulation experiments. Then section “Discussion” refers to the interpretation of the curve families obtained and analyzes the influence of this information upon the evaluation of the SMP in DIP-CV systems for the analysis of malaria. Finally, in Section “Conclusion” the main findings are summarized and their expected influence upon the design of experiments to test the DIP-CV systems that have been the object of this study.

2 Materials and Methods

The problem of determining the relationship between the sensitivity per object (i.e. the proportion of infected RBCs that the system is capable to detect) and the sensitivity per specimen, (which corresponds to the proportion of positive specimens that the system is capable to detect when analyzing a large number of individual RBCs from their blood samples), was addressed here. This issue has been treated before in some detail in Ref. [6], whereas a different approach has been introduced in the present work.

2.1 Basic definitions

Starting from the definition of sensitivity

$$S_e = TP / (TP + FN) \quad (2)$$

where TP stands for the amount of true positives and FN means the amount false negatives, we will call per object sensitivity S_{eo} , that associated to the detection of individual infected RBCs (objects in what follows), and per specimen sensitivity S_{es} , that associated to the detection of an infected specimen after having analyzed the images of its blood smears. In analogous way, the per object (S_{po}) and per specimen (S_{ps}) specificity can be defined.

The following definitions will be used in this work:

S_{eo} = per object sensitivity.

S_{po} = per object specificity.

F_{fno} = 1- S_{eo} = false negative fraction in reference to objects.

F_{fpo} = 1- S_{po} = false positive fraction in reference to objects.

PPV_o = positive predictive value in reference to objects.

S_{es} = per specimen sensitivity.

S_{ps} = per specimen specificity.

p = parasite density, or proportion of infected RBCs (objects) in a given specimen.

$pperc$ = p expressed as a percentage.

P_o = probability of detecting at least one positive (infected) RBC.

P_{omin} = minimum desired value of P_o in a given experiment.

M = total number of objects (RBCs) analyzed.

N = number of positive (infected) objects in a sample of M objects (RBCs).

It is worth to mention that one μ l of blood contains approximately 5×10^6 RBCs as stated in Ref. [23].

2.2 Mathematical model

In order to perform the simulation study proposed here, the process of classifying a set of M objects (RBCs) as infected (positive) or not (negative), was modeled using the binomial probability function.

- Relationship among S_{eo} , P_o and N :

The probability of detecting at least one positive object for a given F_{fno} when analyzing N positive objects is given by

$$P_o = 1 - (F_{fno})^N \quad (3)$$

From this expression, the number of positive objects N that should be analyzed to detect at least one, with probability at least equal to P_{omin} , is

$$N \geq \frac{\log(1-P_{omin})}{\log(F_{fno})} \quad (4)$$

This expression will be used to calculate N_{min} as a function of S_{eo} . Notice that this result lead to the necessity of finding the number M of RBCs to be examined to find at least N positives, which would allow detecting a positive specimen with a desired probability.

- Relationships among S_{es} , S_{eo} , p and M :

The per-specimen sensitivity S_{es} is affected both by the presence of positive objects -which depends upon the parasite density- in the sample analyzed by the DIP-CV system, and by the per-object sensitivity of the latter. The problem of the presence of i positive objects in a sample of M objects (RBCs) when p is the probability that a given object be positive (parasite density) is characterized by the binomial probability distribution [22]

$$P_i = \binom{M}{i} p^i (1-p)^{M-i}, \quad (5)$$

where

$$\binom{M}{i} = \frac{M!}{i!(M-i)!}. \quad (6)$$

The probability of classifying a positive specimen as such when analyzing M RBCs pertaining this specimen can be calculated by adding for all i ($1 \leq i \leq M$) the probability of finding i positives in M , weighted by the probability that at least one of them be detected, for a given S_{eo} , which is

$$S_{es} = \sum_{i=1}^M (1 - F_{fno}^i) \binom{M}{i} p^i (1-p)^{M-i} \quad (7)$$

Equation (7) allows determining iteratively the number of objects M that should be analyzed to obtain a desired value of S_{es} given S_{eo} and an estimated p . However usually $M \gg 1$, which is necessary for a precise calculation of p whose values are usually low. Therefore, evaluating equation (7) requires calculating large factorials whose direct computation might even not be feasible. Then the well-known approximation of the binomial probability function by the Poisson distribution [22] for $p \ll 1$ can be used to obtain

$$S_{es} = \sum_{i=1}^M (1 - F_{fno}^i) \frac{e^{-\mu} \mu^i}{i!}, \quad \mu = pM \quad (8)$$

This expression was used to calculate curve families that reveal useful information about the interdependence between S_{es} , S_{eo} , M and p . When the condition $p \ll 1$ does not hold (*e.g.* $p \geq 0.01$) the normal approximation of the binomial distribution was used instead. This is given by

$$S_{es} = \sum_{i=1}^M (1 - F_{fno}^i) N(i, \mu, \sigma) \quad (9)$$

where N is the normal (Gaussian) probability calculated in the interval $(i \pm 0.5)$ with parameters $\mu = pM$ and $\sigma = \sqrt{p(1-p)M}$.

- Model testing by computer simulation:

Curve families were obtained by means of equations (8) and (9) and also calculated by means of computer simulation. Simulating the detection of positive RBCs consisted here in generating a first random binary sequence associated to the M analyzed RBCs, with $P(1) = p$, where 1 count as a positive object (RBC). Then a second binary random sequence with $P(1) = S_{eo}$ is generated to simulate the appearance of FN. The second sequence has the purpose to change the state of a positive outcome in the previous sequence, according to the considered value of S_{eo} , whenever this element corresponds to a false negative. The value of S_{es} was determined by repeating the process just described and considering that the outcome was a positive specimen whenever at least one TP was found in a run of the simulation experiment. The number n of times that this experiment had to be repeated for each value of M to evaluate S_{es} was determined [22] by the expression

$$n = \frac{z_{\alpha/2}^2}{4e^2}, \tag{10}$$

where z is the normalized normal variable, $\alpha/2$ the significance level (5%) and e (0.02) the maximum absolute error that was set when estimating the proportion of detected positive specimens. This led to the value $n \cong 2400$.

- Assessing the effect of false positives:

The rate of false positive detections is a very important factor to be considered in this analysis, given that especially for low values of p the occurrence of false positives can severely distort the results. Here the positive predictive value PPV_o was used as the statistical measure of performance which characterizes the effect of false positives.

A first approach to analyze this problem is to study the dependence of PPV_o as a function of the specificity and the parasite density. Consider a sample of M RBCs and a parasite density p and assume that in the mean the number of positive objects in the sample is pM , then

$$TP + FN = pM \tag{11a}$$

$$TN + FP = (1 - p)M \tag{11b}$$

Now from equations (11) and the definitions of S_{eo} and S_{po} , a system of four equations can be formed from which the dependence of PPV_o upon S_{eo} , S_{po} and p can be found as

$$PPV_o = \frac{S_{eo}p}{S_{eo}p + (1-p)(1-S_{po})} \tag{12}$$

This equation was used to obtain curves that illustrate how the proportion of true positives depends on the specificity for a given parasite density and sensitivity.

In order to validate the analysis made, these curves were also obtained by computer simulation. In addition to the random sequences used previously to compare with results of equations (8) and (9), a third binary random sequence with $P(1) = F_{fpo}$ was used to simulate the occurrence of false positives by changing the state of the negatives in the first binary sequence. Consider that when measuring the parasite density an apparent value of p will be obtained, affected both by the presence of false positives and false negatives. It is easy to prove the relationship

$$p_{\text{apparent}} = p_{\text{real}} \frac{S_{eo}}{PPV_o} \tag{13}$$

Some implications of this relationship will be discussed later.

3 Results

- Relationship among S_{eo} , P_o and N :

The relationship obtained through equation (4) is shown in Fig. 1, where it is seen that a lower S_{eo} implies the need of observing more positive RBCs to detect at least one with a high value of P_{omin} . The actual values obtained for P_o for a set of values of S_{eo} and their corresponding minimum value of N (N_{min}) are also shown.

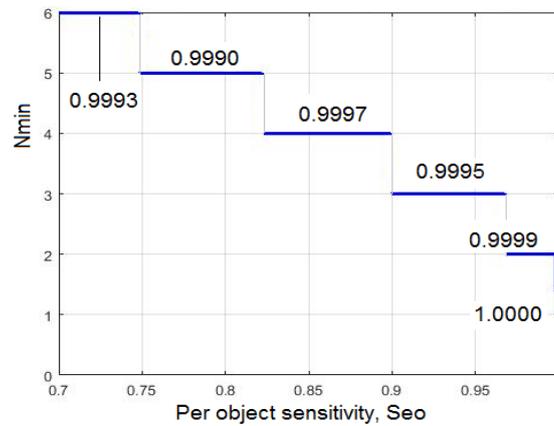


Fig. 1. Minimum number of positive RBCs to be analyzed to detect at least one with probability $P_o \geq 0.999$ as a function of the per object sensitivity S_{eo} . Actual values of P_o for selected values of S_{eo} and the corresponding N_{min} are indicated on the segments.

- Relationships among S_{es} , S_{eo} , M and p :

The interdependence among S_{es} , S_{eo} , M and p for low and high values of p respectively (the borderline was considered around $pperc=0.5\%$), was obtained using equations (8) and (9). Figures 2 and 3 show respectively the correspondence between the

results obtained using equations (8) and (9) and those obtained by computer simulation. In Figure 2a S_{es} was calculated for $S_{eo} = 0.95$ (a reasonable practical value according to those reported in the literature) and a wide range of M, for low values of p as a parameter expressed in percentage as pperc, obtained using the Poisson's approximation of equation (8) and the same conditions are set in Fig. 2b, in this case using computer simulation.

An analogous procedure was followed to obtain the curve families shown in Figures 3a and 3b, in this case employing the Gaussian approximation of equation (9) for Fig. 3a and using computer simulation for Fig. 3b.

Figures 4a and 4b show the dependence of S_{es} upon the number M of RBCs analyzed, for $S_{eo} = 0.95$ and several values of parasite density expressed in percentage, for a range of high S_{es} values (above 0.99). Here, in the case of Figure 4a, the parasite density values are compatible with the Gaussian approximation and Figure 4b corresponds to the range of low values of p, compatible with the Poisson's approximation.

- Assessing the effect of false positives:

Evaluation of per object positive predictive values PPV_o in terms of S_{eo} , S_{po} and p were made using equation (12) and plotted in Fig. 5a. Figure 5b shows the same curves obtained by means of computer simulation. Notice the correspondence between the curves obtained using the analysis that led to equation (12) and the computer simulation results.

4 Discussion

- Relationship among F_{fno} , P_o and N:

Figure 1 portrays the information about the minimum number of positive objects (infected RBCs) that have to be observed for a given per object sensitivity to have a very high certainty (P_o) of detecting at least one of them. This graph illustrates the effect of an insufficient per object sensitivity on the per specimen sensitivity.

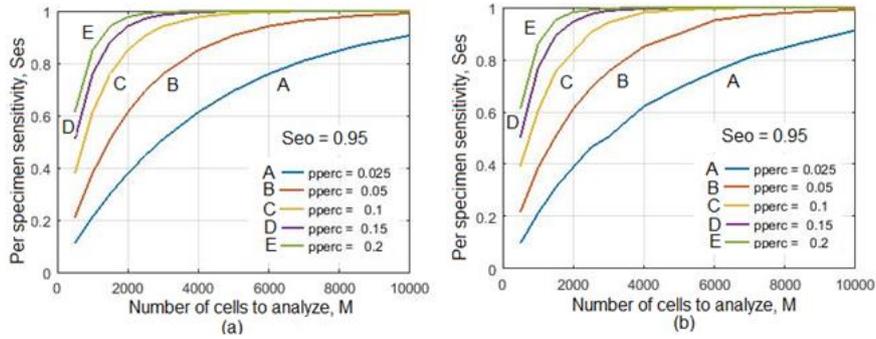


Fig. 2. Per specimen sensitivity value S_{es} as a function of M for $S_{eo} = 0.95$ and a set of low values of p expressed in percentage as pperc, obtained (a) using the Poisson approximation with equation (8) and (b) by means of computer simulation.

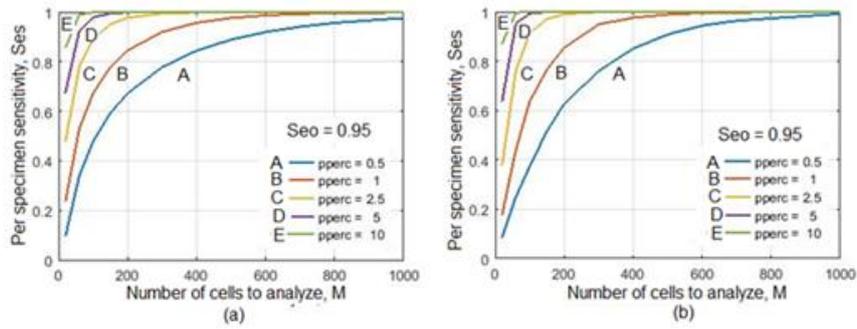


Fig. 3. Per specimen sensitivity value S_{es} as a function of M for $S_{eo} = 0.95$ and a set of high values of p expressed in percentage as pperc, obtained (a) using the Gaussian approximation with equation (9) and (b) by means of computer simulation.

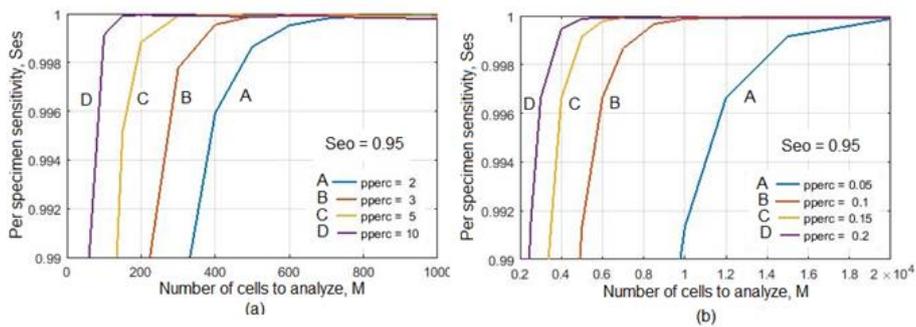


Fig. 4. Per specimen sensitivity for values above 0.99 as a function of the number of RBCs analyzed, at a per object sensitivity $S_{eo}=0.95$. The parasite density p is expressed in percentage pperc. (a) for high values of p (Gaussian approximation, equation 9) and (b) for low values of p (Poisson approximation, equation 8).

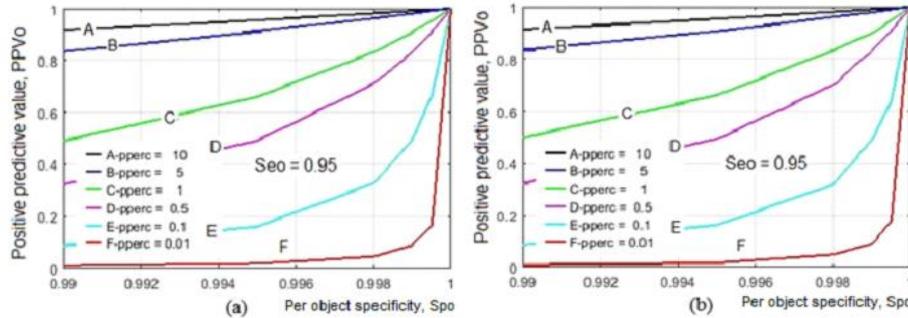


Fig. 5. Combined effect on the positive predictive value PPV_o of the per object specificity value S_{po} and the parasite density p expressed in percentage for a typical value of S_{eo} , (a) using equation (12) and (b) by means of computer simulation.

when p is low and very few positives appear in the samples under analysis if the goal is having a high per specimen sensitivity. For example, a per object sensitivity below 0.95 would require to observe at least three positives to meet the goal of certainty in infected specimen detection and if the infection rate is below 0.1 %, this would demand in the mean analyzing at least 3000 RBCs.

- Relationships among S_{es} , S_{eo} , M and p :

Figures 2a and its counterpart 2b obtained through simulation, both using the Poisson model, indicate that for a practical value above which the per object sensitivity can be obtained in DIP-CV systems ($S_{eo} = 0.95$, see Table 1) and an expected value of infection rate expressed in percentage $pperc$, the required number M of observed erythrocytes in order to obtain a reasonably (very) high per specimen sensitivity S_{es} , tends to be very high for relatively low values of $pperc$. For higher values of $pperc$, Figures 3 (a and b) obtained using the Gaussian model show that a rather high value of S_{es} can be obtained for much lower values of M .

It is worth to mention that the mean square error was calculated between S_{es} vs. M curve families obtained by the theoretical approximation and by computer simulation and very low values (magnitude order of 10⁻⁵) were obtained for the range ($0.6 \leq S_{es} \leq 1$).

Figures 4 a and b are intended to illustrate the same relationships as figures 2 and 3 but in this case for very high values of S_{es} , (e. g. above 0.99), which could be desired in a practical situation. These figures differ in their corresponding ranges of parasite density p ; notice that obtaining high enough values of S_{es} for low values of $pperc$ as those used in Figure 4b might require analyzing much higher numbers of RBCs, reaching the range above ten thousand. As will be seen later, this is effect is worsening by the fact that with low p the negative effect of the presence of false positives increases significantly.

Curves S_{es} vs. M obtained for a low p with S_{eo} as the parameter, not included here due to space reasons, showed a low relative dependence of S_{es} on S_{eo} when high values of S_{es} are sought.

The obtained relationships illustrate how the existing parasite density tends to be a very significant parameter in the process of detecting infected RBCs in regard of the value of M needed; in this case the number of RBCs to be observed in order to find the required minimum number of positives to ensure at least one detection tends to be very high, particularly for low values of p .

- Assessing the effect of false positives:

Figure 5 is related to consequences derived from equation (13), which suggests that measuring p with perfect accuracy would be theoretically possible in a detection system which produces equal numbers of FN and FP , in other words PPV_o should be ideally equal to S_{eo} . This condition cannot of course be attained in practice, because the mechanisms that give rise to FP and FN are distinct and produce them at random.

Failures in correct classification of RBCs in a DIP-CV system depend upon the specific attributes of the observed cell: in fact, it is impossible to ensure that $FP=FN$. Therefore, the objective should be to obtain both S_{eo} and PPV_o as close to unity as possible. Figure 5 reveals that PPV_o drops very fast when S_{po} decreases, even for a range of high values of the latter, when p has a moderately low value. This means that obtaining a relatively high PPV_o when p tends to be low, would require very high values of S_{po} , an objective that cannot be overlooked in applications like drug development, where low infection rates are monitored in the evolution of the response of laboratory animals to drugs under testing. The interest in obtaining very high specificity, however, is not clearly sought in many works.

It is important to recognize also that to evaluate with acceptable accuracy such high values of S_{po} , M must be also high enough, in order to have estimation errors in the order of the third decimal place or lower. Evaluating S_{po} and S_{eo} , on the other hand, is a matter of evaluating proportions and equation (10) should be applied to determine the minimum M needed in order to evaluate them with controlled error for an expected range of p .

5 Conclusion

This work has been motivated by the need to establish a common framework to perform the evaluation of DIP-CV systems employed in the analysis of microscopy images during malaria studies. The approach followed in this work could be valuable to address similar analysis in other DIP-CV current applications related to microscopy images, examples of which can be found in [24] and [25]. In the work developed here a mathematical model of the analysis of thin blood smears employing DIP-CV in the diagnosis of malaria was introduced and it allowed revealing some limitations that can be observed in many works when determining the SMP of DIP-CV algorithms for this application, especially the sensitivity per specimen and the parasite density of a sample. It was found that when determining the per specimen sensitivity, the parasite density has a much larger influence on the number of RBCs that must be analyzed than the per object sensitivity.

A second factor to take into account is that measuring p with adequate accuracy depends strongly upon the positive predictive value of the algorithm. For low expected values of p , this would require very high values of specificity per object S_{po} , which consequently demands a high enough value of M in order to measure this accurately.

The fact that the number of RBCs to be analyzed is closely related to the accuracy with which the SMP can be calculated should be taken into account when evaluating a DIP-CV system. In this case, a minimum M is to be defined for (a) measuring S_{eo} for an expected p , (b) measuring the parasite density p with a desired accuracy, (c) obtaining S_{es} for some S_{eo} and an expected p , the latter being the dominating factor. As a final remark, especial attention should be taken to provide high enough S_{po} in conditions of low parasite density in order to obtain reasonably good PPV_o .

6 Acknowledgement

The authors wish to gratefully acknowledge Prof. Dr. Edmund N. Biden and Mrs. Kristel Desjardins from the University of New Brunswick, Canada, for their support and assistance that made possible the publication of this article.

7 References

- [1] World Health Organization (WHO). (2019). World malaria report 2019. <https://www.who.int/news-room/feature-stories/detail/world-malaria-report-2019>
- [2] Di Ruberto C., Dempster A., Khan S., Jarra B. (2002). Analysis of infected blood cell images using morphological operators. *Image Vis. Comput.* 20: 133–146. [https://doi.org/10.1016/S0262-8856\(01\)00092-0](https://doi.org/10.1016/S0262-8856(01)00092-0)
- [3] Sio S.W., Sun W., Kumar S., Bin W.Z., Tan S.S., Ong S.H., et. al. (2007). MalariaCount: an image analysis-based program for the accurate determination of parasitemia. *J. Microbiol. Methods.* 68(1): 11–18. <https://doi.org/10.1016/j.mimet.2006.05.017>
- [4] Ross N.E. (2006). Automated image processing method for the diagnosis and classification of malaria on thin blood smears. *Med. Biol. Eng. Comput.* 44: 427–36. <https://doi.org/10.1007/s11517-006-0044-2>.
- [5] Tomari R., Zakaria W.N.W., Jamil M.M.A., Nor F.M., Fuad N.F.N. (2014). Computer Aided System for Red Blood Cell Classification in Blood Smear Image. *Procedia Comput. Sci.* 42: 206–13. <https://doi.org/10.1016/j.procs.2014.11.053>
- [6] Tek F.B., Dempster A.G., Kale I. (2010). Parasite detection and identification for automated thin blood film malaria diagnosis. *Comput. Vis. Image Underst.* 114(1): 21–32. <https://doi.org/10.1016/j.cviu.2009.08.003>
- [7] Alkrimi J.A., Jalab H.A., George L.E., Ahmad A.R., Suliman A., Al-Jashamy K. (2015). Comparative Study Using Weka for Red Blood Cells Classification. *Inter-national Journal of Medical, Health, Biomedical and Pharmaceutical Engineering*, 9(1):19-22. <http://repository.um.edu.my/97268/1/red-blood-cells-classification.pdf>
- [8] Chavan S., Nagmode M. (2016). Malaria Disease Identification and Analysis Using Image Processing. *IJCAT International Journal of Computing and Technology*; 1(6): 263-269. <http://ijcat.org/IJCAT-2014/1-6/Malaria-Disease-Identification-and-Analysis-Using-Image-Processing.pdf>

- [9] Das D.K., Maiti A.K., Chakraborty C. (2015). Automated system for characterization and classification of malaria-infected stages using light microscopic images of thin blood smears. *J. Microsc.* 257(3):238–252. <http://doi.org/10.1111/jmi.12206>.
- [10] Díaz G., González F.A., Romero E. (2009). A semi-automatic method for quantification and classification of erythrocytes infected with malaria parasites in microscopic images. *J. Biomed. Inform.* 42(2): 296–307. <http://doi.org/10.1016/j.jbi.2008.11.005>.
- [11] Khan N.A., Pervaz H. Latif A.K., Musharraf A. (2014). Unsupervised identification of malaria parasites using computer vision. *IEEE 11th Int. Jt. Conf. on Comput. Sci. Softw. Eng. JCSSE 2014.* 263–267. <http://ieeexplore.ieee.org/abstract/document/6841878>. <https://doi.org/10.1109/jcsse.2014.6841878>
- [12] Linder N., Turkki R., Walliander M., Mårtensson A., Diwan V., Rahtu E., et. al. (2014). A malaria diagnostic tool based on computer vision screening and visualization of *Plasmodium falciparum* candidate areas in digitized blood smears. *PLoS ONE* 9(8): e104855. <http://dx.plos.org/10.1371/journal.pone.0104855>. <https://doi.org/10.1371/journal.pone.0104855>
- [13] May Z., Aziz A.M., Sarah S., Salamat R. (2013). Automated quantification and classification of malaria parasites in thin blood smears. *IEEE Int. Conf. on Signal Image Process. Appl. ICSIPA 2013.* 369–373. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6708035. <https://doi.org/10.1109/icsipa.2013.6708035>
- [14] Savkare S.S., Narote S.P. (2015). Automated system for malaria parasite identification. *IEEE Int. Conf. on Communication & Computing Technology (ICCICT)*, 1–4. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=7045660. <https://doi.org/10.1109/icict.2015.7045660>
- [15] Arco J.E., Górriz J.M., Ramírez J., Álvarez I., Puntonet C.G. (2015). Digital image analysis for automatic enumeration of malaria parasites using morphological operations. *Expert Systems with Applications.* 42(6): 3041–3047. <http://dx.doi.org/10.1016/j.eswa.2014.11.037>. <https://doi.org/10.1016/j.eswa.2014.11.037>
- [16] Liang Z., Powell A., Ensoy I., Poostchi M., Silamut K, Palaniappan K. et. al. (2016). CNN-based image analysis for malaria diagnosis. *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 493–496. <https://ieeexplore.ieee.org/abstract/document/7822567>. <https://doi.org/10.1109/bibm.2016.7822567>
- [17] Maiti M., Dhane D., Mungle T., Maiti A.K., Chakraborty C. (2017). Web-Enabled Distributed Health-Care Framework for Automated Malaria Parasite Classification: an E-Health Approach. *Journal of Medical Systems.* 41 (192): 1-18. <https://doi.org/10.1007/s10916-017-0834-0>
- [18] Loddo A., Di Ruberto C., Kocher M. (2018). Recent Advances of Malaria Parasites Detection Systems Based on Mathematical Morphology. *Sensors* 18(513): 1-21. <http://doi.org/10.3390/s18020513>.
- [19] Prananda A.R., Nugroho H.A., Ardiyanto I. (2019). Enumeration of *Plasmodium* Parasites on Thin Blood Smear Digital Microscopic Images. *5th International Conference on Science in Information Technology (ICSITech)*: 223-228. <https://ieeexplore.ieee.org/abstract/document/8987492>. <https://doi.org/10.1109/icsitech46713.2019.8987492>
- [20] Rajaraman S., Jaeger S., Antani S.K. (2019). Performance evaluation of deep neural ensembles toward malaria parasite detection in thin-blood smear images. *PeerJ.* 7:e6977. <https://doi.org/10.7717/peerj.6977>.
- [21] Poostchi M., Silamut K., Maude R.J., Jaeger S., Thoma G. (2018). Image analysis and machine learning for detecting malaria. *Translational Research* 194: 36-55. <https://doi.org/10.1016/j.trsl.2017.12.004>

- [22] Walpole RE, Myers RH, Myers SL, Keying EY. (2013). Probability and Statistics for Engineers and Scientists. Pearson New International Edition. Pearson Higher Ed. 2013.
- [23] Tek F.B., Dempster A.G., Kale I. Computer vision for microscopy diagnosis of malaria. Malar. J. 8:1–14. <https://doi.org/10.1186/1475-2875-8-153>
- [24] Amer HM, Abou-Chadi FEZ, Kishk SS, Obayya MI. (201X). A CAD System for the Early Detection of Lung Nodules Using Computed Tomography Scan Images. International Journal of Online and Biomedical Engineering (IJOE). Vol 15, No 04 (2019). <https://www.online-journals.org/index.php/i-joe/article/view/9837>. <https://doi.org/10.3991/ijoe.v15i04.9837>
- [25] Dewi EM, Purwanti E, Apsari R. (201X). Cervical Cell Classification using Learning Vector Quantization (LVQ) Based on Shape and Statistical Features International Journal of Online and Biomedical Engineering (IJOE). Vol 15, No 02 (2019) <https://www.online-journals.org/index.php/i-joe/article/view/9796>. <https://doi.org/10.3991/ijoe.v15i02.9796>

8 Authors

Juan V. Lorenzo-Ginori obtained the Ph.D. degree in Technical Sciences at Universidad Central “Marta Abreu” de Las Villas (UCLV), Cuba. He is Consultant and Emeritus Professor at the Informatics Research Center (IRC), Faculty of Mathematics, Physics and Computing, UCLV, where he was the founder and former Director of the Signal and Image Processing Laboratory. His current research interests are in Digital Image Processing and its applications. Prof. Lorenzo-Ginori has authored numerous papers in international journals as well as taught undergraduate and graduate courses in various countries. He is also a Member of Merit of the Academy of Sciences of the Republic of Cuba.

Lyanett China-Valdés obtained the B.Sc. degree in Biomedical Engineering and the M.Sc. degree in Signals and Systems at the Faculty of Electrical Engineering, UCLV. She worked as a researcher and Instructor at UCLV in the area of Digital Image Processing. M.Sc. China-Valdés is currently working as Biomedical Engineer at the Electro-medicine services of the Cuban Ministry of Public Health in the Villa Clara Province.

Niurka Mollineda-Diogo obtained a M.Sc. degree in Preventive Medicine and works currently as Associate Professor and Researcher at the Parasitology Group, Center for the Study of Chemical Bioactives at UCLV. Prof. Mollineda-Diogo teaches courses in Physiology and Medical Biophysics and her research interests are related to the discovery of new drugs to treat parasite diseases like malaria and leishmania.

Alfredo Meneses-Marcel obtained the Dr. in Veterinary Medicine degree at UCLV, an M. Sc. Degree at Instituto de Medicina Tropical “Pedro Kouri”, Cuba and the Ph. D. degree (with award) in Parasitology and Microbiology at Universidad Complutense de Madrid, Spain. He works as Associate Professor and Researcher at the Center for the Study of Chemical Bioactives at UCLV, where he is currently Head of the Biological Department.

Article submitted 2020-06-04. Resubmitted 2020-06-28. Final acceptance 2020-07-02. Final version published as submitted by the authors.