Efforts to Eliminate Malaria: Developing a Malaria Screening Capability Using Machine Learning on Blood Test Data

NAKATANI Hitoshi

Malaria remains endemic in many countries. In 2023, it affected 263 million people and caused 597,000 deaths[1]. The Sustainable Development Goals (SDGs) include malaria elimination by 2030 as one of their key targets[2]. Early detection and treatment are essential to prevent severe disease progression and further transmission. However, because initial symptoms are often mild, diagnosis may be delayed. To address this, we have developed a screening function—intended to suggest the possibility of disease rather than provide a definitive diagnosis—using complete blood count (CBC) and C-reactive protein (CRP) measurements obtained from the Microsemi LC-667G CRP, an automated hematology analyzer. This paper describes the hematological effects of malaria and the development of a machine learning model to identify malaria-related patterns from CBC and CRP data.

Introduction

Malaria is an infectious disease that is prevalent primarily in tropical and subtropical regions. After a person is bitten by an Anopheles mosquito, a unicellular protozoan parasite known as the malaria parasite enters the bloodstream and invades hepatocytes, leading to their destruction. Subsequently, the parasite invades red blood cells, where it multiplies and divides, eventually leading to their destruction. The parasites are then released into the bloodstream and repeat this cycle by invading new red blood cells[3]. Five species of malaria parasites are known to cause malaria in humans, among which Plasmodium falciparum and Plasmodium vivax pose the greatest threat. P. falciparum is the most widespread and life-threatening species in sub-Saharan Africa, while P. vivax is the predominant species in countries outside sub-Saharan Africa[4]. Common clinical features include fever, anemia, and splenomegaly (enlargement of the spleen)[5]. The initial symptoms are often mild and difficult to distinguish from those of other febrile illnesses. If left untreated, the disease can rapidly progress to a severe form and may be fatal within 24 hours, particularly in infections caused by P. falciparum[4]. Therefore, early detection and prompt initiation of appropriate treatment are crucial.

Typical diagnosis of malaria involves microscopic examination of stained blood smears and rapid diagnostic tests (RDTs) that detect malaria parasite–derived antigens or enzymes[6]. However, in cases of relatively mild infection,

or depending on the physician's clinical judgment, malaria may not be suspected and thus not tested for, raising concerns about missed diagnoses. Furthermore, although microscopic examination is the most accurate method, it is time-consuming, making it impractical to perform this test on all specimens. To address these issues, screening for malaria using only routine blood tests that are commonly performed during febrile episodes has been proposed[7]. From the perspective of reducing diagnostic workload and dependence on individual expertise, it is desirable for automated hematology analyzers to be equipped with automated malaria screening functions that process blood samples after collection.

Research on malaria diagnosis using automated hematology analyzers has reported methods utilizing CBC and a 5-part white blood cell differential (neutrophils, lymphocytes, monocytes, eosinophils, and basophils; hereafter referred to as 5Diff)[7]. The HORIBA Group also markets products that provide malaria detection flags based on CBC and 5Diff parameters. However, studies using CBC and a 3-part white blood cell differential (lymphocytes, monocytes, and granulocytes; hereafter referred to as 3Diff) are limited[8]. In regions with a high prevalence of malaria, there is a demand to minimize measurement costs, making CBC- and 3Diff-based products, which are reasonably priced, more suitable. Therefore, we developed a function to flag suspected malaria cases based on CBC and 3Diff measurements (hereafter referred to as the malaria screening flag).

Microsemi LC-667G CRP

The Microsemi LC-667G CRP (hereafter LC-667G) is an automated hematology analyzer suitable for routine clinical use. In addition to CBC, it can measure 3Diff and CRP. The device features a compact design and reasonable cost, delivering CBC and 3Diff results in approximately one minute, and CRP results within approximately four minutes when CRP measurement is included.

The LC-667G measures white blood cell count (WBC), red blood cell count (RBC), hematocrit (Hct), and platelet count (PLT) using the electrical impedance method (Figure 1). In this method, blood cells suspended in saline pass through a micro-aperture while a constant current is applied, and the resulting changes in electrical resistance are measured as changes in voltage. The magnitude of the voltage change is proportional to the volume of each blood cell, allowing both cell count and cell volume to be determined from the voltage pulses. Because of its high accuracy in measuring particle size and volume, the electrical impedance method remains a standard technique for blood cell counting.

CRP concentration is measured using the latex immunoturbidimetric method. After hemolyzing whole blood with a lysing reagent, latex reagents sensitized with anti-human C-reactive protein antibodies are allowed to react in the presence of a stabilizer. CRP in the specimen and latex particles in the reagent undergo antigen–antibody reactions, leading to latex particle aggregation. The rate of change in turbidity due to this aggregation is measured using red light, and the CRP concentration in the hemolyzed specimen is determined using a polynomial calibration curve prepared from standard serum. Simultaneously, the measured Hct value is used to convert the result to the plasma CRP concentration.

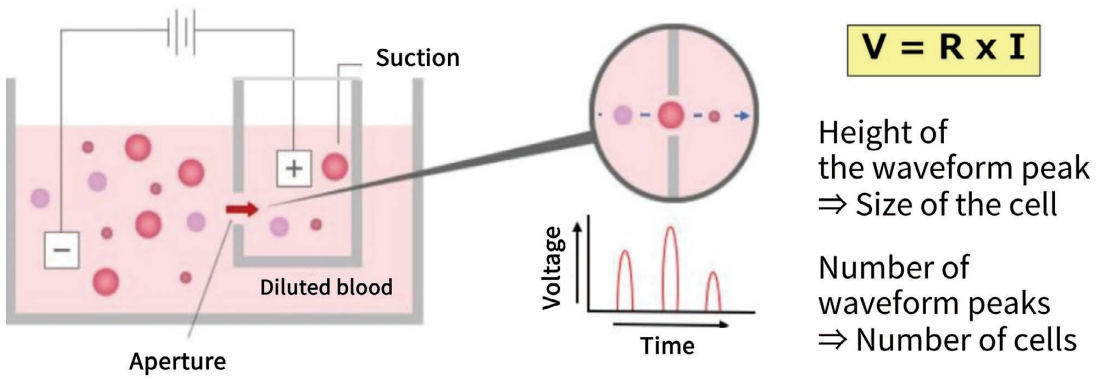

Figure 1 Impedance method

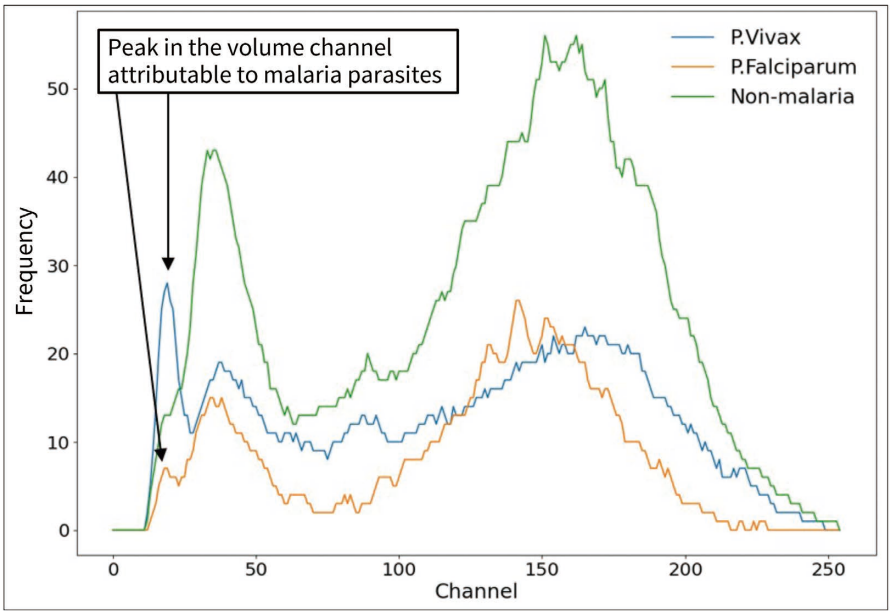

Figure 2 Volume distribution of white blood cells in samples of P. Vivax, P. Falciparum and Non-malaria. The unit of the horizontal axis, Channel, takes values from 0 to 255 that are proportional to the volume of white blood cells.

Examination of Predictive Factors

We examined which measurement parameters should be used as predictors of malaria infection, considering both data-analytic and hematological perspectives on how malaria affects the blood. First, data collection for developing the malaria screening flag was conducted at clinics in India during the monsoon seasons from 2018 to 2019 using the LC-667G. Blood tests were performed on all febrile patients, and CBC, 3Diff, and CRP results were collected. Simultaneously, malaria positivity or negativity was determined by microscopic examination of blood smears and RDTs.

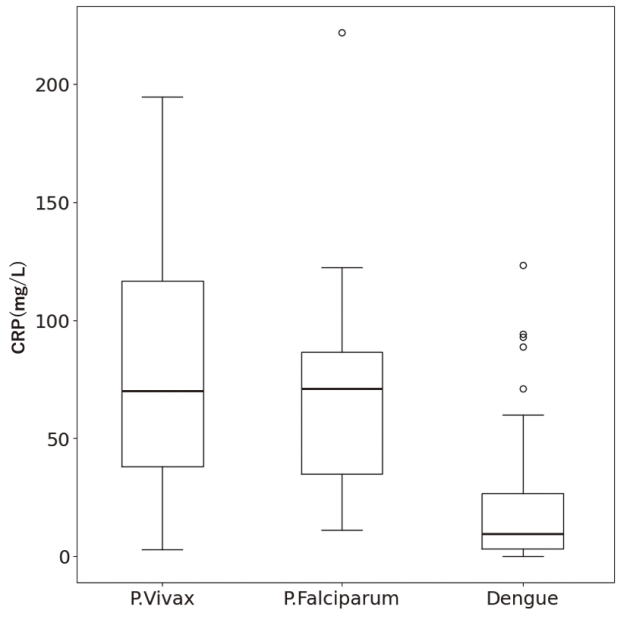

Figure 3 Comparison of CRP concentrations in samples of P. Vivax, P. Falciparum and dengue fever.

Analysis of the collected data revealed several trends in malaria-positive samples. One such trend is the appearance of an abnormal peak in the WBC volume–frequency distribution, which is more common in P. vivax malaria (Figure 2). This peak is believed to represent aggregates of malaria parasites rather than white blood cells. In contrast, this peak is smaller in P. falciparum malaria, likely because the larger P. falciparum parasites are more readily sequestered in the spleen and are less frequently present in peripheral blood[8].

An increase in CRP concentration was also observed. CRP is a biomarker of inflammation that is known to increase during infections. Although not specific to malaria, CRP has been reported as a useful marker for assessing malaria severity[9]. CRP is also useful for differentiating malaria from dengue fever, as these diseases often overlap geographically and present with similar symptoms and CBC results, complicating differential diagnosis. As shown in Figure 3, CRP concentrations tend to be higher in malaria than in dengue fever, because CRP elevation is generally mild in viral infections such as dengue, which is caused by the dengue virus[10].

A decrease in PLT was also noted. Multiple mechanisms are thought to contribute to thrombocytopenia, such as increased platelet sequestration in the spleen due to splenomegaly and aggregation of platelets into larger particles that are counted as single events following red blood cell destruction by malaria parasites[11].

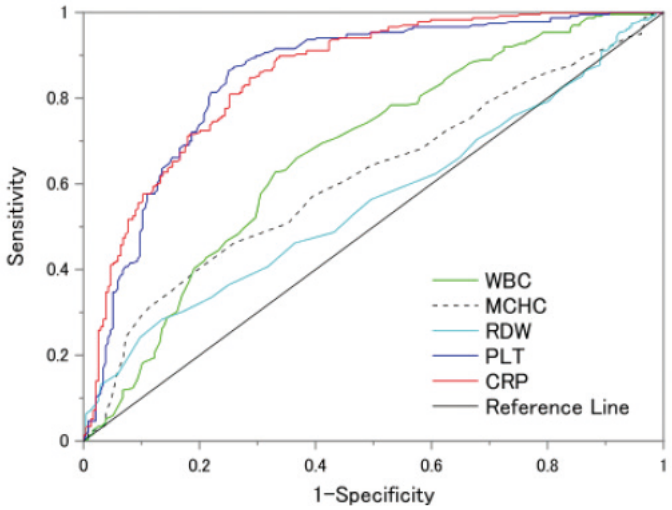

Figure 4 Comparison of ROC Curves for hematological parameters in the malaria diagnosis

Table 1 Comparison of AUCs for hematological parameters in the malaria diagnosis

	AUC
WBC	0.672
MCHC	0.613
RDW	0.558
PLT	0.849
CRP	0.852

To confirm the superior discriminatory power of CRP and PLT for malaria infection compared with other parameters, Figure 4 and Table 1 are presented. Figure 4 shows a receiver operating characteristic (ROC) curve for distinguishing malaria-positive from malaria-negative cases. The ROC curve plots "sensitivity" (the proportion of true positives correctly identified) against "1 − specificity" (the proportion of false positives among negatives) for varying cutoff values. Higher sensitivity and lower 1 − specificity indicate better discrimination, with superior performance represented by curves approaching the upper left corner of the graph. The area under the curve (AUC) quantifies this performance, representing the area beneath the ROC curve. From Figure 4 and Table 1, CRP and PLT demonstrate better discriminatory power than WBC, MCHC (mean corpuscular hemoglobin concentration), and RDW (red cell distribution width).

Modeling

The core of the malaria screening flag calculation method is a malaria screening model constructed using machine learning algorithms (Figure 5). Two key considerations were addressed in the machine learning process. The first is ensuring model interpretability and explainability. Interpretability refers to the extent to which it is possible to explain how the model derives outputs from inputs, whereas explainability refers to the extent to which the reasons for the outputs can be explained. Without sufficient interpretability and explainability, clinicians would

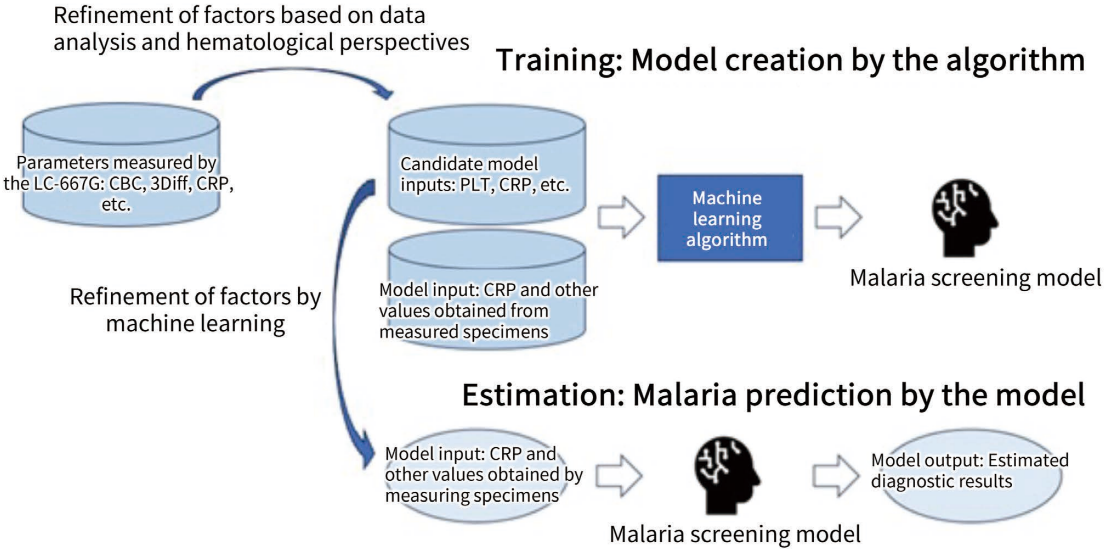

Figure 5 Workflow for the development and application of a machine learning model for malaria screening based on hematological parameters

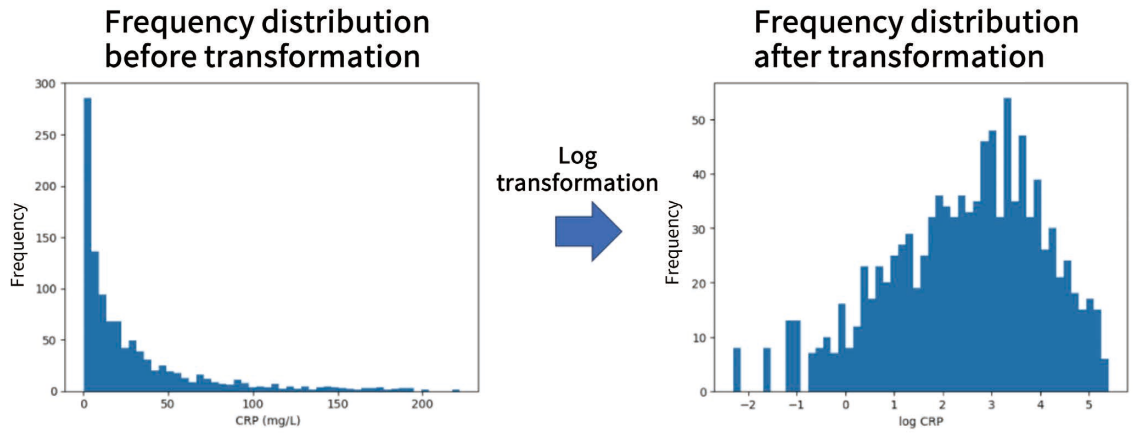

Figure 6 Comparison of CRP value distributions before and after log transformation, showing improved normalization

be unable to explain the diagnostic rationale to patients, making such models unsuitable for clinical adoption. In this study, we employed linear sparse modeling as the machine learning algorithm. This algorithm produces models in which the outputs are calculated by summing the products of the inputs and their corresponding weights (Equation 1), allowing the contribution of each input to be quantified and thereby ensuring interpretability. Additionally, the algorithm automatically selects important factors from the inputs and uses only those selected, thereby improving model explainability. As discussed in the previous section, the input factors are designed based on the mechanisms by which malaria affects the blood, and each factor correlates with the degree of suspicion of malaria infection, facilitating understanding of their impact on the outputs.

$$Y = \sum_{i=1}^{n} W_i X_i \qquad (1)$$

Here, X_i denotes the i-th input, W_i the weight corresponding to X_i, and Y the output representing the degree of suspected malaria infection.

The second consideration is achieving sufficient accuracy. For the screening flag to be reliable, its correlation with the gold standard—microscopic diagnosis of blood smears—must be high. To achieve this, we made several adjustments. The algorithm assumes that the input factors follow a normal distribution, but the actual data distributions may differ. For example, CRP exhibits a right-skewed distribution, so a logarithmic transformation was applied to approximate a normal distribution (Figure 6). Also, although data were collected during the monsoon seasons, when malaria incidence peaks, the number of malaria-positive samples was smaller than the number of negative samples. In machine learning, models tend to learn more from the majority class, so we adjusted the learning weights for each sample to control this bias. Since high sensitivity is required for screening purposes, we adjusted the learning procedure to optimize the decision threshold, based on ROC curve analysis, to meet the required level of accuracy. Furthermore, because blood characteristics are significantly influenced by race, age, and sex, we considered the applicable range of the model to reduce the risk of erroneous judgments. These measures contributed to improving the accuracy of the model.

Conclusion

We have developed a malaria screening flag based on CBC, 3Diff, and CRP measurements. The model underlying this flag utilizes machine learning, enabling automated, highly accurate, and explainable judgments that are independent of individual expertise. Labor savings allow for the testing of more patients than was previously possible, and screening of patients without overt symptoms contributes to the early detection of malaria. The HORIBA Group will continue to contribute to global health through measurement and analysis utilizing digital transformation (DX) and artificial intelligence (AI) technologies.

* Editorial note: This content is based on HORIBA's investigation at the year of publication unless otherwise stated.

References

[1] World Health Organization, "World malaria report 2024," https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2024, accessed 2026-03-31.

[2] 日本SDGs協会, "SDGsについて" https://japansdgs.net/about/sdgs/, accessed 2026-03-31. (in Japanese)

[3] K. Baer, C. Klotz, S. H. I. Kappe, T. Schnieder, and U. Frevert, "Release of Hepatic Plasmodium yoelii Merozoites into the Pulmonary Microvasculature," *PLoS Pathogens*, vol. 3, no. 11, e171, 2007, https://doi.org/10.1371/journal.ppat.0030171

[4] World Health Organization, "Malaria," https://www.who.int/news-room/fact-sheets/detail/malaria, accessed 2026-03-31.

[5] 久枝 一, "マラリアにおける腸内細菌叢と病態の関連性", *Journal of Japanese Biochemical Society*, vol. 88, no. 5, pp. 649–653, 2016. (in Japanese)

[6] MSD Manuals (Professional Version), "マラリア", https://www.msdmanuals.com/ja-jp/professional/13-%E6%84%9F%E6%9F%93%E6%80%A7%E7%96%BE%E6%82%A3/%E8%85%B8%E7%AE%A1%E5%A4%96%E5%AF%84%E7%94%9F%E5%8E%9F%E8%99%AB/%E3%83%9E%E3%83%A9%E3%83%AA%E3%82%A2, accessed 2026-03-31. (in Japanese)

[7] G. Campuzano-Zuluaga, T. Hänscheid, and M. P. Grobusch, "Automated haematology analysis to diagnose malaria," *Malaria Journal*, vol. 9, art. 346, 2010, https://doi.org/10.1186/1475-2875-9-346

[8] J. Nishimura, P. Dharap, and S. Raimbault, "The utility of basic blood counts, WBC histogram and C-reactive protein in detecting malaria," *BMC Infectious Diseases*, vol. 21, no. 1, art. 1006, 2021, https://doi.org/10.1186/s12879-021-06704-5

[9] A. K. Mu, P. C. Bee, Y. L. Lau, and Y. Chen, "Identification of protein markers in patients infected with Plasmodium knowlesi, Plasmodium falciparum and Plasmodium vivax," *International Journal of Molecular Sciences*, vol. 15, no. 11, pp. 19952–19961, 2014, https://doi.org/10.3390/ijms151119952.

[10] 中西雅樹, 稲葉 亨, 藤田直久, "シリーズ：CRP-基礎と臨床応用 1. CRP(C reactive protein; C反応性蛋白)" 医療と検査機器・試薬, vol. 35, pp. 725–730, 2012. (in Japanese)

[11] M. V. Lacerda, M. P. Mourão, H. C. Coelho, and J. B. Santos, "Thrombocytopenia in malaria: who cares?," *Memórias do Instituto Oswaldo Cruz*, vol. 106 (Suppl. 1), pp. 52–63, 2011, https://doi.org/10.1590/S0074-02762011000900007

NAKATANI Hitoshi

Assistant Section Leader
IoT & Data Analytics Dept.
Process Engineering Center
R&D Division
HORIBA, Ltd.