Machine learning approaches classify clinical malaria outcomes based on haematological parameters

Collins M. Morang, Lucas Amenga–Etego, Saikou Y. Bah, Vincent Appiah, Dominic S. Y. Amuzu, Nicholas Amoako, James Abugri, Abraham R. Oduro, Aubrey J. Cunnington, Gordon A. Awandare and Thomas D. Otto

Abstract

Background: Malaria is still a major global health burden, with more than 3.2 billion people in 91 countries remaining at risk of the disease. Accurately distinguishing malaria from other diseases, especially uncomplicated malaria (UM) from non-malarial infections (nMI), remains a challenge. Furthermore, the success of rapid diagnostic tests (RDTs) is threatened by Pfhrp2/3 deletions and decreased sensitivity at low parasitaemia. Analysis of haematological indices can be used to support the identification of possible malaria cases for further diagnosis, especially in travellers returning from endemic areas. As a new application for precision medicine, we aimed to evaluate machine learning (ML) approaches that can accurately classify nMI, UM, and severe malaria (SM) using haematological parameters.

Methods: We obtained haematological data from 2,207 participants collected in Ghana: nMI (n = 978), SM (n = 526), and UM (n = 703). Six different ML approaches were tested, to select the best approach. An artificial neural network (ANN) with three hidden layers was used for multi-classification of UM, SM, and UM. Binary classifiers were developed to further identify the parameters that can distinguish UM or SM from nMI. Local interpretable model-agnostic explanations (LIME) were used to explain the binary classifiers.

Results: The multi-classification model had greater than 85% training and testing accuracy to distinguish clinical malaria from nMI. To distinguish UM from nMI, our approach identified platelet counts, red blood cell (RBC) counts, lymphocyte counts, and percentages as the top classifiers of UM with 0.801 test accuracy (AUC = 0.866 and F1 score = 0.747). To distinguish SM from nMI, the classifier had a test accuracy of 0.96 (AUC = 0.983 and F1 score = 0.944) with mean platelet volume and mean cell volume being the unique classifiers of SM. Random forest was used to confirm the classifications, and it showed that platelet and RBC counts were the major classifiers of UM, regardless of possible confounders such as patient age and sampling location.

Conclusion: The study provides proof of concept methods that classify UM and SM from nMI, showing that the ML approach is a feasible tool for clinical decision support. In the future, ML approaches could be incorporated into clinical decision-support algorithms for the diagnosis of acute febrile illness and monitoring response to acute SM treatment particularly in endemic settings.

Keywords: Machine learning, Uncomplicated Malaria, Severe Malaria, Haematological parameters, Classification
Background
In 2018, there were 228 million cases of malaria worldwide, 93% of which occurred in the African region [1]. Furthermore, approximately 450,000 deaths were reported, of which 61% were children under 5 years old [1]. According to WHO 2018 report, over 2.7 billion US dollars were spent towards various control and elimination efforts to address the global burden of malaria [1]. This includes over 2.74 billion doses of artemisinin-based combination therapies, procured in 2017 [1]. Unfortunately, incorrect diagnosis leads to incorrect treatment. It can increase the chances of antimalarial drug resistance, or for false negative diagnosis, it may result in misdiagnosis of malaria, inappropriate treatment, and progress to severe disease or death [2–4]. The gold standard for malaria diagnosis is microscopy, which requires extensive training, but rapid diagnostic tests (RDTs) have become the frontline diagnostic tools for malaria because of their ease of use at point-of-care [5].

One drawback of RDTs is the emergence of gene deletions of the target antigen, histidine-rich protein (Pfhrp2/3) in the parasite genome, which render parasites undetectable by the most common RDTs [6]. Other challenges include insufficient sensitivity to detect low-level parasitaemia and the number of tests which need to be performed per positive result in settings with declining or low transmission [2, 6, 7]. Different problems are faced in non-endemic countries, where imported malaria must be suspected as a possible cause of fever before an RDT or microscopy would be performed in the first place, and failure to identify cases at first contact with health services often results in worse clinical outcomes [8, 9]. Therefore, improved and complementary malaria diagnostic techniques are required, which can overcome some or all of these limitations.

Complete blood counts (CBCs) are the most commonly performed laboratory test in most hospitals in both developing and developed countries. The CBC is usually relied upon to provide clues for the diagnosis of patients where advanced methods for detection of specific diseases are lacking, with a parameter such as decreased platelet counts often associated with severe malaria (SM) [10, 11]. In addition, haemoglobin (Hb) levels are very important for the classification of SM cases [12]. Indeed, the changes in haematological parameters during clinical malaria have been studied extensively to aid in the understanding of disease pathogenesis [13–19]. However, the potential and diagnostic value of haematological parameters measured by commonly available automated haematology analysers has not been fully studied using unbiased approaches such as machine learning (ML) techniques. These haematological parameters have the potential to be used in differentiating clinical malaria from other febrile illnesses, especially in areas where the reliability of RDTs is challenged by the high prevalence of Pfhrp2/3 deletion mutant parasites.

ML approaches use algorithms based on statistical assumptions and mathematical rules to learn patterns and produce meaningful classifications based on the association of each variable with the disease outcome [20–24]. These classifications can then be applied to new disease cases to make classifications on the most probable cause. This classification capability of ML has not been extensively implemented in the diagnosis of clinical malaria. To date, only a single study has reported the use of ML to diagnose malaria using clinical history and symptoms captured verbally and visually [25]. The sample size (n = 376) was very small to deduce meaningful classifications, and the author concluded that more work would be needed [25]. Despite this, there have been far reaching studies on the application of ML in other areas of malaria research [26–30]. The diagnosis of malaria using ML on clinical datasets has been impaired by the lack of large data, as well as difficulty in data curation. Moreover, classical modelling is prone to over-fitting or under-fitting of data [31], but recent approaches such as imputation, encoding, centering and scaling of variables, and model optimization [24] enable augmented use of ML in malaria classification.

We hypothesized that we can classify clinical malaria and non-malarial infections (nMI) with an ML approach. We first collected and curated data from 2,207 patients including nMI (n = 978), uncomplicated malaria (UM) (n = 703), and SM (n = 526). We generated ML models to classify clinical malaria (UM and SM) from nMI using haematological parameters.

Methods
Study population and sample collection
Standards for Reporting Diagnostic Accuracy Studies (STARD) guidelines [32, 33] were followed in this study. The current study utilizes unpublished data of 526 patients from a previous case-control study of SM conducted by the Navrongo Health Research Centre (NHRC) located in the Kassena-Nankana Districts (KNDs) in the Upper East Region of Northern Ghana. In the original study, children with acute febrile symptoms admitted to the Navrongo War Memorial Hospital (NWMH), the only referral facility in the KNDs, were evaluated for inclusion into the study from August to December 2002 and May 2003 to April 2004. Full details of the study procedure, inclusion criteria, and demographic and clinical characteristics of SM cases may be reviewed in Oduro et al. [34].
In brief, the inclusion criteria for SM cases were (1) all children between 1 and 15 years old, (2) presenting with fever (or history of fever in the past 24 h) and were admitted to the NHRC-IRB, (2) residence in the Navrongo Health and Demographic Surveillance System area [34], and (3) willingness of parents/caregivers to offer informed consent. Criteria for SM diagnosis and enrollment into the original study were classified as having SM by the WHO standard guidelines that include haemoglobin <5 g/dL or haematocrit <15% [34, 35]. Ethical approval for the SM study was obtained from the NHRC Institutional Review Board (IRB), Noguchi Memorial Institute of Medical Research (NMIMR) IRB, Naval Medical Research Center IRB, and Ghana Health Service (GHS) Ethics Review Committee (ERC). Informed consent was obtained and documented, followed by administration of a questionnaire about the presenting symptoms and clinical examinations. Participants who did not consent and meet the study inclusion criteria and those who had reported taking antimalarial treatment in the past 2 weeks were excluded from the study, while those who turned out to be malaria negative by standard microscopy were withdrawn from the study. All study samples were taken prior to initiation of treatment except for samples taken for clinical monitoring during admission or for follow-up after discharge from the hospital.

The nMI and UM participants were recruited in a hospital-based cross-sectional study involving two hospitals: Kintampo North-Municipal Hospital, Kintampo, and Ledzokuku Krowor Municipal Assembly Hospital (LEKMA), Accra. The inclusion criteria were (1) outpatient children 1–15 years old, (2) presenting with fever or history of fever in the past 24 h or axillary temperature ≥38°C, (3) and (4) signed informed consent by self (adolescents) and parent/guardian. The exclusion criterion was participants with known chronic disease or history of antimalarial drug use in the past 2 weeks. Ethical approval was also obtained from NMIMR-IRB, GHS-ERC, and Kintampo Health Research Centre (KHRC) ERC. A case was defined as nMI if the individual presenting to the hospital was malaria negative by either RDTs, Taqman array, or microscopy. Clinical data such as age, sex, and body temperature and symptoms such as fever were collected on recruitment.

**Sample collection procedure**

Venous blood was collected in the ante-cubital fossa. Tourniquet was not applied beyond 1 min during venesection to avoid haemo-concentration, which could give erroneous results for all parameters measured. Samples were taken mostly between 8 am and 12 pm to avoid variations due to individuals’ activity (such as rehydration and food intake). Samples (5 mL) were taken into K3 EDTA tubes (BD Vacutainer; Becton Dickinson, NJ, USA). Samples that could not get analysed within 2 h from the time of collection were stabilized at 2–8 °C to avoid changes that could occur in some haematological parameters should the sample be left on the bench for more than 3 h. Samples were analysed not later than 24 h from the time of sample storage at 2–8 °C. No capillary blood sample was taken during the study as it presents with subtle variations from venous blood parameters. CBC analysis was performed using the automated ABX Micros 60 haematology analyser, which measures white blood cell parameters, red blood parameters, and platelet parameters (Additional file 1: Table S1). Data were manually cross-referenced twice for accuracy to ensure consistency in sample collection procedures.

**Statistical classifier: median split**

Kernel density estimation, which is a non-parametric technique, was used to estimate the probability density function of each haematological parameter and kernel distribution for each parameter between nMI, UM, and SM and visualized using density plots in R (R version 4.0.2). The median value within each diagnostic group (nMI, UM, and SM) was computed, and the mean of any two group medians was used for ‘median split’ to generate a dichotomous variable for each parameter (low and high levels representing below and above median, respectively) [36]. Contingency tables were used to summarize the relationship between clinical diagnosis (nMI, UM, and SM), and each dichotomous parameter. The generalized linear models for predictive analysis were used to explain the relationship between the clinical diagnosis and the dichotomous parameter. Odds ratios were computed through the exponent of the regression coefficients (logits) to estimate the strength of the relationship. Any OR with 95% confidence interval (CI) that includes a null value (1.0) indicated that the parameter was not significantly associated with clinical diagnosis. ANOVA was used to compare the model with the null model and chi-square test used to compute the degree of significance. All the analyses were done in R-software (R version 4.0.2).

**Data pre-processing and normalization**

A multivariate imputation via chained equations (MICE) plot was used to visualize the missing observations in the data. It was difficult to determine whether the missing values were missing ‘completely at random’ or ‘missing at random’ or ‘not at random’ to enable selection of the imputation method. Therefore, the demographic/clinical data and microscopy results were not imputed and were not used for modelling. The majority of the haematological parameters had less than 5% missing data, and the missing values were imputed using MICE
package in R. Each variable in the training and test data was transformed using the Yeo-Johnson function, centred to have a mean of 0, and scaled to have a standard deviation of 1. The original dataset (before pre-processing and normalization) is available in Additional file 2: Table S2.

**Machine learning**

Six ML algorithms were evaluated to identify the best algorithm that can classify the binary data. These include partial least squares (PLS) logistic regression, multiple adaptive regression splines (MARS), random forest, decision trees, support vector machine, and artificial neural networks (ANN). PLS logistic regression was implemented by reducing the dimension of haematological parameters so as to increase accuracy. We used 10-fold cross-validation while tuning through 16 principal components (PC), whereby the optimal model used 2 PC. The optimal hyperparameters for MARS (with cross-validation) were determined in a grid search of 30 different combinations of 3rd degree and sampling 1000 terms to retain the final model [37]. Decision tree was implemented with the rpart function, which performs auto-tuning with an optimal subtree of 10 total tree splits. Random forest and support vector machines were implemented by first performing a grid search to identify the optimal hyperparameters followed by classification analysis. Three ANN were developed, one multi-classification ANN (nMI vs UM vs SM) and two binary classifications denoted as ANN (UM and nMI) and ANN (SM and nMI). For each ML model, the data were split into 80% training and 20% testing. The outcome was the clinical diagnosis of the participant (as concluded by the attending clinicians) having either UM or nMI or SM. Haemoglobin and haematocrit levels were not included in the modelling because they are used to support the diagnosis of malaria [10, 19, 35, 38].

**Hyperparameter tuning for artificial neural networks**

The ANN was composed of an input layer of 15 haematological parameters. The loss was computed using categorical cross-entropy for the multi-classifier and binary cross-entropy for binary classifiers, while accuracy was used as the main evaluation metric. During training, the 80% training data was further split into 70% training and 30% validation with randomization (Fig. 1). Tensor board visualizations were used to check the dynamic graphs of our training and test metrics. Hyperparameters were tuned to identify the optimal model parameters for each classification. A hyper-grid was developed that adjusts the model capacity, normalization term, kernel regularization, and learning rate. To maximize the validation error performance, we tuned 12, 32, 64, 128, 256, and 512 rectified linear units (ReLU) in three hidden layers. We used batch normalization on each hidden layer for gradient propagation and performance improvement. We varied the dropout rate from 0.1, 0.2, 0.3, and 0.4 in all the three layers to identify the best dropout regularization that prevents the model from latching to happenstance patterns that are not significant. We used ‘Adam’ as the optimizer, but we varied the learning rate (0.1, 0.05, 0.001, and 0.0001) to find a global minimum. The tfruns R package was used to implement the hyper-grid in R-software, using 500 epochs, batch size of 64, and validation split of 0.3. These Keras models were initialized for all the three classifications to select the optimal model.

**Model evaluations**

Yardstick package was used to perform classifications on the test data as well as compute the performance of the model. The confusion matrix, accuracy, area under the receiver operating characteristic curve (AUC), precision and recall, and F1 score were the metrics used to evaluate performance. The F1 score is a measure of test data accuracy, which is a weighted average between precision and recall. To explain the model, we used local interpretable model-agnostic explanations (LIME Package in R) [39]. The classification model was set up, and an ‘explainer’ of the classifying model was initiated using the training data and the model output classifications. The explainer was used to explain the results of the test data-set as classification explanations (feature weights). The feature weights were used to build a heatmap for each ANN indicating how each feature explains the model.

**Effect of patient age and sampling location on the model predictions**

To test if patient age and sampling location significantly affected the models, we used three models: (1) a model for all the UM and nMI cases (n = 1681), (2) a model for UM and nMI from Kintampo cases only (n = 726), and (3) a model for only Kintampo cases and ages under 4 years (n = 416). We tested the possibility of using the ANN to evaluate the models but there was some level of over-fitting and under-fitting of the 2nd and 3rd models, due to sample size limitation. Therefore, random forest was subsequently used, because of (1) its robustness to smaller sample size with minimal over-fitting of the data and (2) its ability to reduce the high variance from decision trees by combining several trees into one ensemble tree [40].

**Statistical analysis**

The clinical categorical data was analysed using Pearson’s chi-square while the continuous data such as the haematological parameters were analysed using the Kruskal-Wallis test with Dunn’s post hoc tests across
the three groups (UM, SM, and nMI). All tests were two sided, and statistical significance was set at \( P < 0.05 \) for all analyses with adjustment for multiple testing. Data analyses were performed using R-software (R version 4.0.2), R-studio (version 1.1), and Python (version 2.7). The R codes with the methods, including the curated data files, can be found on github: https://github.com/misita-falcon/Machine-Learning-in-Clinical-Malaria.

**Results**

**Characteristics of the study participants**

Participants were recruited as follows: 38.8% (857/2,207) from Accra, 32.9% (726/2,207) from Kintampo, and 28.3% (624/2,207) from Navrongo (Fig. 1). These participants from all the three areas constitute 44.3% (978/2,207) nMI, 31.8% (703/2,207) for UM, and 23.8% (526/2,207) for SM cases (Fig. 1). The median age was 3 years (range 2–6 years) for nMI, 4 years (range 2–7 years) for UM, and 1 year (range 1–2 years) for SM. The median ages were significantly different as determined by the Kruskal-Wallis test \( (P < 0.001) \) (Table 1). The SM cases had a significantly higher median body temperature (38.3; range = 37.5–39.2), compared to the nMI (37.2; range = 36.5–38.4) and UM (38.1; range = 37–39) \( (P < 0.001) \). There was a significant difference in proportions of individuals \( (P < 0.001) \) among nMI, UM, and SM from different locations (Kintampo, Navrongo, and Accra) as determined by the chi-square analysis (Table 1). There was no association between sex and clinical diagnosis, although the number of females was higher than males in all three groups \( (P = 0.247) \); nMI was 51.2% (501/978), UM was 54.9% (386/703), and SM was 55.1% (290/526) (Table 1). Fever was more common in SM (99.2%, 522/526) compared to UM (85.5%, 601/703) and lowest in nMI (59.4%, 581/978), and the chi-square analysis shows that there was an association between fever and clinical diagnosis \( (P < 0.001) \) (Table 1).

Participants with UM had a higher geometric mean parasite density (27,467.59 parasites/μL, SD = 8.44) compared to SM individuals (16,674.41 parasites/μL, SD = 8.61). But, the median levels did not vary significantly between the two groups \( (P = 0.592) \) (Table 1). Participants with nMI were negative by microscopy, RDT, and Taqman array. There were 212 different suspected infections in the nMI group, and the top 10 include upper respiratory tract infections (17%, 167/978), malaria (9.5%, 93/978), gastroenteritis (7.6%, 75/978), sepsis (6.1%, 60/978), otitis media (5.9%, 58/978), enteric fever (2.6%, 26/
978), fever (2.1%, 23/978), tonsillitis (2.3%, 23/978), pneumonia (2.1%, 21/978), and anaemia (1.9%, 19/978) (Additional file 1: Fig. S1). Laboratory results indicated that majority of the samples were undetermined/not available/not known (96%, 937/978), with only 4% having accurate laboratory results (41/978). Some of the organisms that were laboratory confirmed include Streptococcus pneumonia, Staphylococcus aureus, Salmonella typhi, Coxiella burnetti, and dengue virus (Fig. 2). Only 2 UM participants had co-infections (laboratory confirmed) with P. falciparum, and these individuals had Salmonella typhi and group D streptococcus.

Since the sample size of laboratory-confirmed nMI cases was low, all the samples were grouped as nMI, instead of individual diseases during ML classifications.

Haematological parameters vary between nMI, UM, and SM

Median values for all the haematological parameters were significantly different among nMI, UM, and SM (P < 0.001) (Table 2), but most of the parameters do not show distinct distributions between the different clinical diagnosis groups (Fig. 3). More so, Dunn’s post hoc tests indicated that platelet distribution width, percentage neutrophils, and percentage lymphocytes were not significantly different between the nMI, UM, and SM except median parasite density and patient sex

| Characteristic       | Non-malaria infections (nMI) | Uncomplicated malaria (UM) | Severe malaria (SM) | P value |
|----------------------|-----------------------------|-----------------------------|---------------------|---------|
| N = 2,207            | N = 978 (44.3%)             | N = 703 (31.8%)             | N = 526 (23.8%)     |         |
| Patient age          |                             |                             |                     |         |
| Mean (SD)            | 4.23 (3.57)                 | 4.95 (3.57)                 | 1.66 (0.93)         | < 0.001 |
| Median (range)       | 3.0 (2–6)                   | 4.0 (2–7)                   | 1.0 (1–2)           |         |
| Body temperature     |                             |                             |                     |         |
| Mean (SD)            | 37.4 (1.18)                 | 38.1 (1.23)                 | 38.4 (1.15)         | < 0.001 |
| Median (range)       | 37.2 (36.5–38.4)            | 38.1 (37–39)                | 38.3 (37.5–39.2)    |         |
| Parasite density     |                             |                             |                     |         |
| Geometric mean (SD)  | 0 (0–0)                     | 27,467.59 (8.44)            | 16,674.41 (8.61)    | 0.592   |
| Median (range)       | 0 (0–0)                     | 29,426 (3,144–105,351)      | 25,160 (3,560–86,560)|         |
| Location             |                             |                             |                     |         |
| Accra (n, %)         | 657 (67.2%)                 | 200 (28.4%)                 | 0 (0%)              | < 0.001 |
| Kintampo (n, %)      | 321 (32.8%)                 | 405 (57.6%)                 | 0 (0%)              |         |
| Navrongo (n, %)      | 0 (0.0%)                    | 98 (13.9%)                  | 526 (100.0%)        |         |
| Sex                  |                             |                             |                     |         |
| Female (n, %)        | 477 (48.8%)                 | 317 (45.1%)                 | 236 (44.9%)         | 0.209   |
| Male (n, %)          | 501 (51.2%)                 | 386 (54.9%)                 | 290 (55.1%)         |         |
| Fever symptom        |                             |                             |                     |         |
| No (n, %)            | 395 (40.4%)                 | 97 (13.8%)                  | 4 (0.8%)            | < 0.001 |
| Yes (n, %)           | 581 (59.4%)                 | 601 (85.5%)                 | 522 (99.2%)         |         |
| Missing (n, %)       | 2 (0.2%)                    | 5 (0.7%)                    | 0 (0%)              |         |

Patient age, body temperature, and parasite density were analysed using the Kruskal-Wallis test while recruitment location, sex, and fever were analysed using the chi-square test at 95% CI. All the participant characteristics were significantly different between the nMI, UM, and SM except median parasite density and patient sex

aKruskal-Wallis test
bChi-square test
cDunn (1964) Kruskal-Wallis multiple comparison—UM vs SM only

To further confirm this hypothesis, the median was used to split the variables into categorical variables (low and high levels). The relationship or predictive value of the categorical parameters to accurately classify the clinical diagnosis was determined using contingency tables (Additional file 2: Table S3). The percentage number of individuals who had low levels of each parameter cannot be used to confidently classify the disease outcomes.
Fig. 2 (See legend on next page.)
parameter and were classified with either nMI, UM, or SM (Fig. 4c, d).

Additionally, we determined whether the levels could predict whether an individual has UM or SM using odds ratios. First, we predicted UM, and majority of the parameters were associated with clinical diagnosis of UM and nMI ($P < 0.001$), except mean cell volume, lymphocyte percentage, mixed cell counts, and neutrophil counts (Additional file 1: Table S4). The parameters that were not associated for nMI–SM category were lymphocyte counts, mean corpuscular Hb, lymphocyte percentage, mixed cell counts, and neutrophil counts (Additional file 1: Table S4). Furthermore, some of the haematological parameters had a 95% confidence interval that included the null value (1) when evaluating the odds ratios, which signifies that they are not significantly associated with clinical diagnosis (Additional file 1: Table S4).

**Table 2** Comparison of median and interquartile ranges in haematology values measured in nMI, UM, and SM cases

| Parameters                  | Non-malaria infections (a), $N = 978$ | Uncomplicated malaria (b), $N = 703$ | $a$ vs $b$ | Severe malaria (c), $N = 526$ | $b$ vs $c$ | $a$ vs $c$ | $a$ vs $b$ vs $c$ |
|-----------------------------|--------------------------------------|--------------------------------------|------------|-------------------------------|------------|------------|-----------------|
| WBC indices                 | Median IQR                           | Median IQR                           | $P$ value  | Median IQR                    | $P$ value  | $P$ value  | $P$ value       |
| WBC count ($10^{3}/μL$)     | 9.3 7.0–12.8                         | 8.3 6.3–10.8                         | < 0.001    | 11.6 8.3–16.6                 | < 0.001    | < 0.001    | < 0.001         |
| Lymphocyte count ($10^{3}/μL$) | 3.0 2.0–4.5                          | 1.9 1.3–3.0                         | < 0.001    | 3.8 2.4–6.0                   | < 0.001    | < 0.001    | < 0.001         |
| Mixed cell count ($10^{3}/μL$) | 0.8 0.5–1.1                          | 0.5 0.3–0.8                         | < 0.001    | 0.9 0.5–1.4                   | < 0.001    | 0.004      | < 0.001         |
| Neutrophil count ($10^{3}/μL$) | 4.8 3.3–7.6                          | 5.4 3.7–7.6                         | 0.115      | 6.5 4.4–9.4                   | < 0.001    | < 0.001    | < 0.001         |
| Lymphocyte percent (%)      | 35.8 22.6–47.8                       | 24.7 16.8–36.8                      | < 0.001    | 33.9 26.5–44.4                | < 0.001    | 0.964      | < 0.001         |
| Mixed cell percent (%)      | 8.6 6.7–11.0                         | 6.9 5.0–9.2                         | < 0.001    | 8.2 5.5–11.3                  | < 0.001    | 0.012      | < 0.001         |
| Neutrophil percent (%)      | 54.4 41.7–69.0                       | 67.8 53.7–77.1                      | < 0.001    | 55.8 46.6–66.2                | < 0.001    | 0.568      | < 0.001         |
| RBC indices                 | Median IQR                           | Median IQR                           | $P$ value  | Median IQR                    | $P$ value  | $P$ value  | $P$ value       |
| RBC count ($10^{6}/μL$)     | 4.5 4.2–5.0                          | 4.1 3.6–4.5                         | < 0.001    | 2.4 1.7–3.2                   | < 0.001    | < 0.001    | < 0.001         |
| Hb level (g/dL)             | 11.0 10.1–11.8                       | 10.1 8.8–11.2                       | < 0.001    | 5.6 4.1–7.4                   | < 0.001    | < 0.001    | < 0.001         |
| Haematocrit (%)             | 34.5 32–37.1                         | 31.1 27.2–34.8                      | < 0.001    | 16.7 12.0–21.1                | < 0.001    | < 0.001    | < 0.001         |
| RBC distribution width (%)  | 15.1 14.0–16.6                       | 15.7 14.7–17.1                      | < 0.001    | 18.1 16.2–20.1                | < 0.001    | < 0.001    | < 0.001         |
| Mean cell volume (fl)       | 76.0 71.2–80.3                       | 76.0 72.0–81.0                      | 0.510      | 70.0 64.7–75.4                | < 0.001    | < 0.001    | < 0.001         |
| Mean corpuscular Hb (pg)    | 23.7 21.8–25.6                       | 24.9 23.0–26.4                      | < 0.001    | 23.8 21.5–26.7                | 0.006      | < 0.001    | < 0.001         |
| Mean cell Hb concentration (g/dL) | 31.6 29.6–32.5                     | 32.3 31.5–33.3                      | < 0.001    | 35.1 31.2–37.4                | < 0.001    | < 0.001    | < 0.001         |
| Platelet indices            | Median IQR                           | Median IQR                           | $P$ value  | Median IQR                    | $P$ value  | $P$ value  | $P$ value       |
| Platelet count ($10^{3}/μL$) | 292.0 226.0–360.0                    | 140.0 92.0–216.0                    | < 0.001    | 98.0 61.0–156.0               | < 0.001    | < 0.001    | < 0.001         |
| Mean platelet volume (fl)   | 8.2 7.6–8.9                          | 8.1 7.5–8.9                         | < 0.001    | 6.9 6.4–7.8                   | < 0.001    | < 0.001    | < 0.001         |
| Platelet distribution width (fl) | 15.0 13.9–15.4                  | 14.5 12.4–15.6                      | 0.005      | 15 12.0–17.3                  | < 0.001    | < 0.001    | < 0.001         |

$P$ value—Kruskal-Wallis test with Dunn’s post hoc tests

$P$ values were analysed using the Kruskal-Wallis test with post hoc tests. The parameters include WBC indices, RBC indices, and platelet indices. All the haematological parameters were significantly different between the nMI, UM, and SM ($P < 0.001$).
malaria from nMI. The UM vs nMI model was trained on 942 samples, validated on 403 samples, and tested on 336 samples for each ML approach. The SM vs nMI model was trained on 843 samples, validated on 361 samples, and tested on 300 samples for each ML approach (Fig. 1). Among the six ML approaches, the training accuracies ranged between 0.794 and 0.856 to classify UM while the training accuracies ranged between 0.937 and 0.985 in classifying SM. The test accuracies ranged from 0.777 to 0.857 for the UM model and
for the SM model (Additional file 1: Table S5). The SVM approach and the ANN generated the overall best classification outcome.

Hyperparameter tuning for the ANN ($n = 55,290$ combinations) showed that the optimal model for multi-classification had 0.862 training accuracy with a model capacity of 3 layers (128, 64, and 16), with dropouts of 0.4 for layer 1, 0.3 for layer 2, and 0.4 for layer 3, and learning rate of 0.001 (as represented in Additional file 1: Fig. S2). The optimal model ($n = 55,290$ combinations) for ANN (nMI vs SM) with 0.985 accuracy had a model capacity of 3 layers (16, 128, and 256 RELU units, respectively), the dropout rate was 0.2 and 0.4 for the first two layers and the last layer had 0.1, and a learning rate of 0.0001. The optimal model ($n = 55,290$ combinations) for ANN (nMI vs UM) with 0.856 training accuracy had a model capacity of 3 hidden layers of 256, 64, and 16 RELU units, respectively; the dropout rate was 0.1 for the first and last layer and 0.3 for the second layer, and a learning rate of 0.0001. Training and validation history plots for the ANN showed good levelling off for accuracy and loss, as well as acceptable divergence between training loss/accuracy and validation loss/accuracy for all the three models (Additional file 1: Fig. S3).

Also, the history plots suggest that there was near zero over-fitting or under-fitting of the data as indicated by closeness of the training and validation curves (Additional file 1: Fig. S3). The ANN (UM vs. nMI) achieved 0.856 training accuracy and 0.842 validation accuracy, while the testing accuracy of the model was 0.801 (kappa 0.583) (Table 3). The training and testing accuracies demonstrate the confidence of the networks in classifying UM. The ANN (SM vs nMI) achieved a higher accuracy ($\geq 0.96$) for training, validation, and testing accuracy (Table 3). Both ANN had an F1 score of above 0.747, which means the model can be used for the

Fig. 4 Non-symmetrical predictive values of clinical diagnosis using median split (high vs low levels) of each haematological parameter. A ‘median split’ was used to divide each quantitative parameter into categorical variables by the median value (calculated as a mean of nMI and UM or SM median value shown in Table 2). The predictive values are calculated from contingency tables (Additional file 2: Table S3). a The percentage predictive value in predicting nMI from low levels. b Percentage predictive value in predicting SM or UM using the low levels. c Percentage predictive value in predicting nMI using high levels. d Predictive values of UM or SM using high levels.
classification of clinical malaria (Table 3). Since the binary classifiers had the best performance, we also performed a multi-classification analysis to assess the ability of the ANN to differentiate among UM, SM, and nMI. The data available for the multi-classification model was 2,207 samples, which were split to 80% training (n = 1,766) and 20% testing (n = 441). The training data was further split to 70% (n = 1,236) training and 30% (n = 530) cross-validation with accuracies of 0.862 and 0.828, respectively. The test accuracy was 0.853; kappa, precision, recall, and F1 score of the model was > 0.768 (Table 3). The accuracy of multi-classification model provides confidence in the binary classifications.

### Diagnostic value of the models using ROC curves

Having shown the accuracy of the models, we determined the ROC curves of ANN (UM vs nMI) and ANN (SM vs nMI) to show the diagnostic ability of these binary classifiers. Both classifiers had very good performance with an AUC of 0.866 for ANN (UM vs nMI) and AUC of 0.983 for ANN (SM vs nMI) (Fig. 5 and Table 3). This showed that the models could be used to distinguish individuals with SM or UM from those with nMI. The cut-offs for UM show that there is a trade-off in sensitivity and specificity as the cut-off increases or decreases, which is not the case for SM. These results could frame the clinical utility of the models and provide a benchmark for future studies.

### Platelet and RBC counts classify clinical malaria from non-malaria infections

The models were investigated to identify which haematological parameters were classified to be important for either SM, UM, or nMI using local interpretable model-agnostic explanations (LIME). Case by case analysis of the individuals showed that some haematological parameters are important classifiers of UM (Additional file 1: Fig. S4). Case by case analysis was merged into heatmap to generate a consolidated picture of useful parameters for classification (Fig. 6). The top three parameters that had low feature weights for UM are platelet counts, RBC counts, and lymphocyte percentages (Fig. 6a). Based on the order of importance, the top three parameters that were important for SM classification include RBC counts, platelet counts, and mean platelet volume (Fig. 6b). This shows that both platelet and RBC counts are important parameters for clinical malaria while the lymphocyte percentages were unique for UM. These parameters might be used to classify clinical malaria cases from nMI, with a very good diagnostic ability as shown by the ROC analysis (Fig. 5).

### Patient age and sampling location do not affect the model classifications

We further tested if the models are agnostic to age and location variance. There was a significant difference in patient age between nMI and UM (P < 0.001), but there...
**Fig. 5** ROC curve for classification of SM was near perfect. The ROC curve plots sensitivity versus specificity for all possible cut-offs. Each point on the curve represents a different cut-off value, which is connected to form a curve. The diagonal line is a reference line for the ROC curve. 

- **a** ROC for the ANN (UM vs nMI) with an area under the curve (AUC) of 0.866 which is basically an average of true positive rate across all possible false positive rates.
- **b** ROC for the ANN (SM vs nMI) is right angled which means its near perfect with an AUC of 0.983. The levels of AUC indicate a good performance of the models in classifying UM and SM.

**Fig. 6** Platelet and RBC counts identified as classifiers of both UM and SM. The Keras model was explained using local interpretable model-agnostic explanations (LIME Package in R-software). The explainer results of the test data, which are represented as feature weights, were extracted from the explainer and used to plot the heatmaps to show a consolidated picture of the importance of each haematological parameter. The weights that are $<-0.1$ indicate that they are low during UM or SM. 

- **a** The heatmap shows that platelet, RBC, and lymphocyte percentages/counts can classify UM and 
- **b** shows the haematological parameters that can classify SM, and they include RBC counts, platelet counts, mean platelet volume, and mean cell volume.
was no significant difference in samples within Kin-
tampo as well as children under the age of 4 years
(Fig. 7a, c, e; Additional file 1: Fig. S5 & S6). The per-
formance accuracy of the random forest models was
0.806, 0.767, and 0.768 for models 1, 2, and 3, respec-
tively (Fig. 7b, d, f). The most important parameters
that were featured across the three models were platelet
and RBC counts, which are similar to the top two parameters
identified by the ANN. Therefore, the data illustrates
that age and location do not affect model classifications,
and the platelet or RBC counts determined by ANN can
be used to reliably classify clinical malaria from nMI in
these datasets.

Discussion
Automated CBC is one of the blood tests routinely per-
formed for children presenting to health facilities with
fever. However, CBC analysis generates a significant
amount of data on a range of haematological parameters,
and the data is underutilized with only Hb and Hct
levels being routinely used as an indicator of clinical
malaria. Thus, an automated algorithm to detect malaria
based on the haematological parameters as outlined in
this study could have great value as a complementary
malaria diagnostic strategy, particularly at frontline
health centres where CBC is routinely performed. Such
an algorithm also has the added value of enabling the
monitoring of treatment outcomes for in-patients.

In malaria-endemic settings, malaria rapid diagnostic
tests (mRDTs) have revolutionized diagnosis and signifi-
cantly reduced presumptive treatment, particularly in
rural settings where trained microscopists are lacking
[3]. However, reports of emerging Pfhrp2/3 gene dele-
tions threaten the future reliability of the RDTs. False
negative RDT results are also known to occur in low-
density infections [2, 6, 7]. Thus, an approach that is au-
tomated and agnostic to parasite genetic variation is crit-
ical both as a fail-safe and a surveillance strategy for
false negative mRDTs (which might occur due to supply
chain mismanagement or gene variation) [6]. In very
low transmission settings, ML models have the poten-
tial to replace the primary use of mRDTs when the
diagnostic yield of mRDTs becomes very low (i.e.
many mRDTs needed to detect a single case of mal-
aria). In non-endemic settings where malaria may
occur in immigrants and non-immune travellers, the

Fig. 7 Classification of haematological parameters using random forest shows that patient age and sampling location do not affect the ML
models. Three models were generated: a a model for all the UM and nMI cases (n = 1681), which show a significant difference in patient age, while b shows the impurity-based measurement of the feature importance of the model; c a model for UM and nMI from Kintampo cases only
(n = 756), which do not show any significant difference between the patient age, and d shows the feature importance of the model; and e a
model for only Kintampo cases and ages under 4 years, whereby there was no significant difference between the nMI and UM (n = 416) and
f shows the feature importance of the model. The samples for each model were split 80% for training and 20% for testing. The accuracy of the
models was 0.806, 0.767, and 0.768, respectively. The most important feature across the three models was platelet and RBC counts.
models may allow another fail-safe mechanism in case the diagnosis of malaria was not suspected by clinicians and malaria RDT or microscopy was not performed. Despite these advantages, there would be a little extra cost associated with incorporating the algorithm and an automated message into haematology analyser output, a message that can prompt clinicians to consider malaria in the presence of suggestive haematological features.

Previous ML studies have looked into haematological parameters more generally and to classify sickle cell anaemia using deep convolutional networks [41, 42], but did not classify clinical diagnosis. For the first time, ML approaches that can classify infections in children based on haematological parameters have been generated. Six different ML methods were evaluated, and they were all shown to classify clinical malaria from nMI with high accuracy especially the SVM and the ANN. We used the ANN to deconvolute the results: it identified platelet and RBC counts as the top features in classifying both UM or SM from nMI. Low RBC counts can be attributed to extensively parasitized RBCs, which are sequestered during SM [43]. This highlights the significance of RBC counts during Plasmodium falciparum malaria infections. In most occasions except cerebral malaria, SM is associated with anaemia due to RBC lysis during parasite invasion as well as many other RBC abnormalities [44]. This makes the diagnosis of SM much easier than UM, whereby one parameter, such as Hb level of < 5 g/dL, can diagnose or classify the disease.

Cohen et al. analysed data from 680,964 individuals with fever and confirmed that majority of antimalarial drugs are given to malaria-negative individuals [45]. Overtreatment indicates that most nMI can go without being treated, for their true cause, which is also possible for UM and this can lead to drug resistance. Therefore, the difference between febrile outpatient infections is far more challenging, especially between nMI and UM due to similarity in clinical presentations. In large population studies, values of studied parameters more generally and to classify clinical malaria based on haematological parameters may limit the generalizability of the models to other countries; this is also the case for the limited range of SM manifestations in our dataset and the spectrum of laboratory-confirmed nMI. The few nMI cases that were clearly diagnosed and still grouped/retained as nMI may also present minimal bias to the models. Lastly, the study did not have adults > 15 years to comparatively understand the role of age in differentiating clinical malaria based on haematological parameters. Therefore, we recommend that more studies are needed to inform the broader utility of this work. Despite that only 4.6% (75/1645) of the cases were discordant between microscopy and RDT, probably due to hrp2/hrp3 deletions, although there is an insignificant chance that misclassification of malaria could have had an impact on our study. These limitations will be taken into account for further studies to inform the broader clinical utility of this work.

**Conclusions**

Fever is the most common symptom reported in sSA, and correct diagnosis of the implicated pathogen is of high importance for precision medicine. Personalized treatment reduces overtreatment, decreases malaria mortality and antimalarial resistance. This report demonstrates proof-of-principle that ML can be used to distinguish clinical malaria from nMI using routine haematological data. Case by case analysis showed that the models can make classifications based on
combination of three parameters: platelet and RBC counts, lymphocyte counts/percentage, and mean platelet volume. These could be used for precision diagnosis of an individual’s risk of having malaria, to inform the need for confirmatory diagnosis by microscopy or to prompt investigation for other diagnoses when malaria is unlikely. Further work is to calibrate and improve the classification capability of the model using more data from other geographical and transmission settings, demographic groups, co-infections, and different disease severities. Our findings hold promise for the design of clinical software to support the diagnosis of malaria in the WHO African region and might also prove useful for the diagnosis of malaria in returning travellers from non-endemic countries.

Supplementary Information

Supplementary information accompanies this paper at https://doi.org/10.1186/s12916-020-01823-3.

Additional file 1: Table S1. The list of haematological parameters adopted from laboratory procedure manual by the CDC [50]. Table S4. The odds ratio of median categories providing the odd of being diagnosed with either nMI, UM, and SM. The median categories were; low and high levels. Table S5. Performance evaluation of six machine learning models to classify clinical malaria outcomes. Fig. S1. Word cloud of clinical manifestations using clinicians/doctors notes or suspected infections. Fig. S2. Artificial Neural Network Schematic. Fig. S3. Plot for the training and validation history of the ANN. Fig. S4. Case by case analysis of the classification capability of the ML models. Fig. S5. Density estimates of the haematological parameters between nMI and UM cases for sub-sampled data from Kintampo only, as well limit of children under 4 years of age.

Additional file 2: Table S2. Clinical and raw haematological data of study participants. Table S3. The median splits predictive values on the clinical diagnostic categories (nMI, UM, and SM) of the study participants.

Acknowledgements

We acknowledge the study participants as well as the staff of Kintampo Municipal, who provided support for this study. We are grateful to Emmanuel Allotey, Prince Nyarko, Henrietta Mensa-Brown, Felix Ansah, Jersley Chiawuruah, Jonas Kengne, Nsoh Godwin Anabire, Reuben Ayoiv-Djanie, and Nancy Nyarko for their contributions on the data collection and constructive criticism of the work. We also express our sincere gratitude to Dorothy Annan, Deborah Mettle, Bright Yemi, Rachel Abban, Samirah Saiid, Joyceiline Kwaro, and Israel Ossei for assisting in cross-referencing the data. We acknowledge the University of Ghana for providing the high-performance computing resources (the ZUPUTO) used for this work.

Authors’ contributions

CM, GA, LA, TO, SB, DA, AC, and VA contributed to the design and conceptualization of the work, as well as editing and critique of the manuscript drafts. LA, JA, DA, AO, and NA contributed to the acquisition of data. CM performed the data analysis, including building the models, model interpretation, and drafting the manuscript. All authors read and approved the final manuscript.

Funding

The study was supported by a DELTAS Africa grant (DEL-15-007: Awandare). The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS)’s Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa’s Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust (107755/Z/15/Z; Awandare) and the UK government. TO is supported by the Wellcome Trust grant 104111/Z/14/Z. The funder had no role in the study design and where to publish. The views expressed in this publication are those of the authors and not necessarily those of AAS, NEPAD Agency, Wellcome Trust, or the UK government.

Availability of data and materials

All data generated or analysed during this study are included in this published article and its supplementary information files.

Ethics approval and consent to participate

This study was approved by the Ethics Committees of the Ghana Health Service (GHS-ERC: 12/05/12 & GHS-ERC 12/06/16), the Kintampo Health Research Centre (KHRCEC/FEA/2011-13), the Navrongo Health Research Centre (NHRC-IRB135/08/2012), and the Noguchi Memorial Institute for Medical Research (NMIMR) (NMIMR-IRB CPN 004/11-12). All participants were provided with written informed consent prior to inclusion in the study.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

1. West African Centre for Cell Biology of Infectious Pathogens (WACCBIP), Department of Biochemistry, Cell and Molecular Biology, University of Ghana, Legon, Accra, Ghana. 2. Florey Institute, Molecular Biology and Biotechnology, University of Sheffield, Sheffield, UK. 3. Department of Applied Chemistry and Biochemistry, C. K Tedam University of Technology and Applied Sciences, Navrongo, Ghana. 4. Ministry of Health, Navrongo Health Research Centre (NHRC), Navrongo, Ghana. 5. Section of Pediatric Infectious Disease, Department of Infectious Disease, Imperial College London, London, UK. 6. Institute of Infection, Immunity & Inflammation, M.S. University of Glasgow, Glasgow, UK.

Received: 24 June 2020 Accepted: 22 October 2020

Published online: 30 November 2020

References

1. WHO. World malaria report; World Health Organization 2018;4:186. http:// apps.who.int/iris/bitstream/10665/254912/1/WHO-HTM-GMP-2017-4-eng. pdf?ua=1. Accessed 15 Oct 2017.

2. Watson OJ, Sumner KM, Janko M, Goel V, Winskill P, Slater HC, et al. False-negative malaria rapid diagnostic test results and their impact on community-based malaria surveys in sub-Saharan Africa. BMJ Glob Heal. 2019;4:e001582. https://doi.org/10.1136/bmjgh-2019-001582.

3. Mouatcho JC, Dean Goldberg JP. Malaria rapid diagnostic tests: challenges and prospects. J Med Microbiol. 2013;62:1491–505. https://doi.org/10.1099/ jmm.0.052906-0.

4. WHO. False-negative RDT results and P. falciparum histidine-rich protein 2/3 gene deletions. World Heal Organ. 2017. dohttps://doi.org/10.1186/1475-2875-10-166.

5. Wanja EW, Kuya N, Moranga C, Hickman M, Johnson JD, Moseti C, et al. Field evaluation of diagnostic performance of malaria rapid diagnostic tests in western Kenya. Malar J. 2015;14:456. https://doi.org/10.1186/s12936-015-1508-y.

6. Agaba BB, Yeka A, Nsobya S, Arinaitwe E, Nankabirwa J, Opigo J, et al. Systematic review of the status of pHr2p and pHr3p gene deletion, approaches and methods used for its estimation and reporting in Plasmodium falciparum populations in Africa review of published studies 2010-2019. Malar J. 2019;18:355. https://doi.org/10.1186/s12936-019-2987-4.

7. Ranadive N, Kunene S, Darthe S, Ntsalihlali N, Nhlabathi N, Dlamini N, et al. Limitations of rapid diagnostic testing in patients with suspected malaria: a diagnostic accuracy evaluation from Swaziland, a low-endemicity country aiming for malaria elimination. Clin Infect Dis. 2017;64:1221–7. https://doi.org/10.1093/cid/cix131.
8. Angelo KM, Libman M, Caumes E, Hamer DH, Kain KC, Leder K, et al. Malaria after international travel: a GeoSentinel analysis. 2003–2016. Malar J. 2017;16:293. https://doi.org/10.1186/s12936-017-1936-3.

9. Grobush MP, Schlangenhaus P. Self-diagnosis and self-treatment of malaria by the traveler. Travel Med. 2019;5:169–78. doi:https://doi.org/10.1097/01.TR.0000558785.323-54696-6.00016-1.

10. Lampah DA, Yeo TW, Mallyo M, Kenangalem E, Douglas NM, Ronaldon D, et al. Severe malarial thrombocytopenia: a risk factor for mortality in Papua, Indonesia. J Infect Dis. 2015;211:623–34.

11. Hanson J, Phu NH, Hasan MU, Charunwatthana P, Plewes K, Maude RJ, et al. The clinical implications of thrombocytopenia in adults with severe falciparum malaria: a retrospective analysis. BMC Med. 2015;13:97. https://doi.org/10.1186/s12936-015-0334-5.

12. White NJ. Anaemia and malaria. Malar J. 2018;17:371. https://doi.org/10.1186/s12936-018-0259-9.

13. Squire DS, Asmah RH, Brown CA, Adjei DN, Obeng-Nkrumah N, Ayeh-Kumi PF. Effect of Plasmodium falciparum malaria parasites on haematological parameters in Ghanaian children. J Parasit Dis. 2010;40:303–11.

14. Mukovge H, Kikomeko S, Sembajwe LF, Seguya A, Namugwanya C. How reliable are haematological parameters in predicting uncomplicated Plasmodium falciparum malaria in an endemic region? ISRN Trop Med. 2012:2013:1–9.

15. Anibre NG, Arzum A, Francis A, Frank A, Osman A, Kanwugu N, et al. Evaluation of haematological indices of childhood illnesses in Tamale Metropolis of Ghana. 2018.

16. Warimwe GM, Recker M, Kiragu EW, Buckee CO, Wambua J, M'B_dirix U, et al. Plasmodium falciparum var gene expression homogeneity as a marker of the host-parasite relationship under different levels of naturally acquired immunity to malaria. PLoS One. 2013;8:e70467. https://doi.org/10.1371/journal.pone.0070467.

17. Kotepe M, Phiwkham D, Phiwkham N, Chupeerach C, Duangman S. Effect of malarial infection on haematological parameters in population near Thailand-Myanmar border. Malar J. 2014;13:15.

18. Lee SJ, Stepniewska K, Anstey N, Ashley E, Barnes K, Binh TQ, et al. The relationship between the hemoglobin concentration and the haematocrit of Plasmodium falciparum malaria. Malar J. 2008;7:11.

19. Quinto L, Aponte JJ, Menéndez C, Sacarlal J, Aide P, Espasa M, et al. Severe falciparum malaria in young children of the Kassena-Nankana district of northern Ghana. Malar J. 2007;6:96.

20. World Health Organization (WHO). Management of severe malaria: a practical handbook. 3rd edition. WHO Library Cataloguing-in-Publication Data; 2013.

21. Iacobucci D, Posavac SS, Kardes PR, Schneider MJ, Popovich DL. The median split: robust, refined, and revived. J Consum Psychol. 2015;25:690–704.

22. White NJ, Pukrittayakamee S, Hien TT, Faiz MA, Mokuolu OA, Dondorp AM. The median split: robust, refined, and revived. J Consum Psychol. 2015;25:690–704.

23. Schmidhuber J. Deep learning in neural networks: an overview. Neural Netw. 2015;61:85–117. https://doi.org/10.1016/j.neunet.2014.09.003.

24. Mok victorious? K, Anstey N, Ashley E, Barnes K, Binh TQ, et al. The relationship between the hemoglobin concentration and the haematocrit in Plasmodium falciparum malaria. Malar J. 2008;7:11.

25. Cunin G, Kukar M, Notar N, Bivric M, Čelmečić P, Notar N, et al. An application of machine learning to haematological diagnosis. Sci Rep. 2018;8:411. https://doi.org/10.1038/s41598-017-18564-8.

26. White NJ, Pukrittayakamee S, Hien TT, Faiz MA, Mokuolu OA, Dondorp AM. Malaria. Lancet (London, England). 2012;379:106–107. https://doi.org/10.1016/S0140-6736(11)61197-3.

27. Mooney SJ, Stepniewska K, Anstey N, Ashley E, Barnes K, Binh TQ, et al. The relationship between the hemoglobin concentration and the haematocrit in Plasmodium falciparum malaria. Malar J. 2008;7:11.

28. Lecun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521:436–44. https://doi.org/10.1038/nature14539.

29. Schmidhuber J. Deep learning in neural networks: an overview. Neural Netw. 2015;61:85–117. https://doi.org/10.1016/j.neunet.2014.09.003.

30. Iacobucci D, Posavac SS, Kardes PR, Schneider MJ, Popovich DL. The median split: robust, refined, and revived. J Consum Psychol. 2015;25:690–704.

31. Butler KT, Davies DW, Cartwright H, Iyayev O, Walsh A. Machine learning for molecular and materials science. Nature. 2018;559:547–55. https://doi.org/10.1038/s41586-018-0337-3.

32. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Canadian Med J. 2003;164:463–8.

33. Cohen JF, Korevaar DA, Altman DG, Bruns DE, Gatsonis CA, Hooft L, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. BMJ Open. 2016;6:e012799.

34. Oduro AR, Koram KA, Rogers W, Atuquba F, Ansaah P, Anyoringiya T, et al. Severe falciparum malaria in young children of the Kassena-Nankana district of northern Ghana. Malar J. 2007;6:96.

35. World Health Organization (WHO). Management of severe malaria: a practical handbook. 3rd edition. WHO Library Cataloguing-in-Publication Data; 2013.

36. Gunčar G, Kukar M, Notar N, Bivric M, Čelmečić P, Notar N, et al. An application of machine learning to haematological diagnosis. Sci Rep. 2018;8:411. https://doi.org/10.1038/s41598-017-18564-8.

37. Wambua J, M'Bdirix UC, Musa K, Adero T, Adero J, Adu P, Espasa M, et al. Relationship between hemoglobin and haematocrit in the definition of anemia. Trop Med Int Health. 2006;11:295–302.

38. Ribeiro MT, Singh S, Guestrin C. “Why should I trust you?” Explaining the predictions of any classifier. Scand J Infect Dis. 2016;48:1135–44.

39. Law A, Wiener M. Classification and regression by randomForest. 2002. http://www.stat.berkeley.edu/ Accessed 12 May 2020.

40. Xu M, Papageorgiou DP, Abidi SZ, Dao M, Zhao H, Karniadakis GE. A deep convolutional neural network for classification of red blood cells in sickle cell anemia. PLoS Comput Biol. 2017;13:e1005746. https://doi.org/10.1126/sciadv.1700574.