Pre-Analytical Quality Validation for Malaria Diagnosis: A Multi-Stage Deep Learning and Computer Vision Framework

Chamuditha Wanninayake
Department of Computer Science,
Faculty of Computing,
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka.
Email: it22114358@my.sliit.lk

Sayumi Devasurendra
Department of Computer Science,
Faculty of Computing,
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka.
Email: it22099204@my.sliit.lk

Pubuth Prajeshwara
Department of Computer Science,
Faculty of Computing,
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka.
Email: it22587824@my.sliit.lk

Anuja Uthsara
Department of Computer Science,
Faculty of Computing,
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka.
Email: it22101310@my.sliit.lk

Dr. Lakmini Abeywarshana
Department of Computer Science,
Faculty of Computing,
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka.
Email: lakmini.d@sliit.lk

Dr. Mahima Weerasinghe
Department of Computer Science,
Faculty of Computing,
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka.
Email: mahima.w@sliit.lk

Abstract*—Malaria continues to pose a significant global health burden, where accurate and timely diagnosis is critical for effective treatment. Microscopic examination of Giemsa-stained blood smear slides remains the gold standard; however, its reliability depends heavily on staining quality, smear preparation, and operator expertise. These factors introduce variability, increase workload, and can lead to inconsistent diagnostic outcomes. This study presents an integrated system for Malaria diagnosis that incorporates staining quality assessment, blood smear quality analysis, and automated parasite detection within a sequential diagnostic framework. A multi-task deep learning model evaluates staining quality and recommends optimal staining duration, while a computer vision-based approach analyzes smear quality from mobile-captured images using features including smear dimensions, feather edge formation, spacing and positioning, and stain density. Only slides satisfying predefined quality criteria are forwarded to the parasite detection stage, where deep learning models perform infection classification, parasite localization, and species identification. The staining model achieved its highest per-class accuracy of 55.2% on Grade III, representing optimal staining conditions, while time prediction achieved a mean absolute error of approximately 4 minutes. The smear quality assessment module produced an average measurement error of 1.76 mm against expert-recorded ground truth values. The parasite detection and species identification models achieved mAP@0.5 scores of 0.812 and 0.897 respectively. By validating slide quality before microscopic examination, the system limits the exposure of detection models to substandard inputs, reducing both analytical overhead and the risk of misdiagnosis in routine laboratory settings.

Keywords— Blood smear analysis, computer vision, deep learning, Malaria diagnosis, parasite detection, staining quality assessment

I. Introduction

Malaria remains a major global health concern, particularly in regions where timely and accurate diagnosis is essential for effective treatment and disease control [1]-[4]. Microscopic examination of Giemsa-stained blood smear slides continues to be the gold standard for malaria diagnosis, as it enables parasite detection, species identification, and estimation of parasite density [5]-[7]. Despite its widespread use, the reliability of this method depends heavily on several pre-analytical and analytical factors, including staining quality, smear preparation, and the expertise of laboratory personnel [5], [8], [9].

In routine clinical settings, these processes are largely performed manually and evaluated through visual inspection [10]. As a result, the diagnostic workflow is often subject to variability, depending on the skill level of technicians and local laboratory conditions [1], [9]. This can lead to inconsistencies in slide preparation and interpretation, ultimately affecting diagnostic accuracy. Furthermore, the absence of early-stage quality validation means that slides failing to meet preparation standards are often identified only during microscopic examination, increasing both the workload on laboratory personnel and the risk of diagnostic errors [11], [2], [6].

A. Literature Review

1) Staining Quality Assessment

Staining quality plays a fundamental role in malaria diagnosis, as it directly affects the visibility of parasite morphology and cellular structures. A properly stained smear is characterized by distinct parasite coloration and a clear background, enabling accurate identification of malaria parasites. In current practice, staining quality is typically assessed through visual inspection based on biological and morphological criteria, including parasite coloration and background consistency [1], [5], [8], [12], [13].

Several studies have explored automated approaches for evaluating staining quality, primarily using color-based features and histogram analysis techniques [1]. While these methods have shown promising results in distinguishing between well-stained and poorly stained samples, they are generally limited to basic classification tasks. More importantly, they do not provide actionable guidance for correcting staining errors or determining the optimal staining duration. Furthermore, commonly used fixed staining protocols do not account for variations in environmental conditions, reagent quality, and laboratory

practices, leading to inconsistencies across different settings [1], [12], [14].

2) Blood Smear Quality Assessment

Blood smear preparation is another critical factor influencing the reliability of malaria diagnosis. Conventional evaluation methods rely on manual inspection of smear characteristics such as size, shape, feather edge formation, and cell distribution, based on established laboratory guidelines [9], [11], [15]. A properly prepared smear should exhibit a gradual transition from thick to thin regions, a well-defined feather edge, and an appropriate monolayer region suitable for microscopic examination.

Recent research has introduced automated image analysis techniques for smear evaluation, including segmentation-based methods and quantitative analysis of cell morphology [11]. These approaches aim to provide more objective measures of smear quality by analyzing features such as red blood cell distribution and aggregation. However, most existing methods focus primarily on microscopic characteristics and do not fully capture macroscopic aspects such as smear positioning, spacing, and overall slide usability. Additionally, factors such as stain density, which significantly affect the visibility of cells and parasites, are often not explicitly considered. Manual smear preparation further introduces variability due to differences in technique and environmental conditions, leading to inconsistent slide quality [9], [11].

3) Malaria Parasite Detection

Advancements in machine learning and deep learning have significantly improved automated malaria parasite detection. Techniques based on convolutional neural networks, object detection frameworks such as YOLO, and transformer-based architectures have been widely applied to tasks such as parasite detection, classification, and counting [2]-[4], [6], [7], [10], [17]. These models are capable of identifying infected cells and estimating parasite density with high accuracy under controlled conditions.

Despite these advancements, several challenges remain when deploying these systems in real-world clinical environments. Detection models often struggle with small or overlapping parasites, variations in staining quality, and inconsistencies in image acquisition conditions [3], [16], [17]. Additionally, many datasets used for training contain annotation inconsistencies and class imbalance, which can affect model performance. More importantly, most existing approaches focus primarily on parasite detection without incorporating mechanisms to assess whether the input slides are suitable for reliable analysis, limiting their practical applicability in routine diagnostic workflows [3], [9].

B. Research Gap

Although considerable progress has been made in individual areas of malaria diagnosis, existing research remains fragmented and does not address the diagnostic workflow as a unified process. Current approaches predominantly focus on parasite detection, while pre-analytical stages such as staining quality assessment and smear preparation are often overlooked or treated independently. As a result, diagnostic systems frequently rely on high-quality input data without verifying whether the prepared slides meet the necessary quality standards.

In addition, staining quality assessment remains largely subjective and lacks automated methods for recommending optimal staining duration. Similarly, smear quality evaluation is still dependent on manual inspection and does not comprehensively assess key factors such as stain density, smear positioning, and spacing. The absence of integrated quality control mechanisms means that poorly prepared slides are often identified only after microscopic analysis, leading to inefficiencies and potential diagnostic errors.

Therefore, there is a clear need for an integrated system that combines staining quality assessment, blood smear quality analysis, and parasite detection within a single framework. Such a system should enable early-stage validation of slide quality, provide actionable feedback to improve slide preparation, and ensure that only suitable slides are used for parasite detection. Addressing these limitations can significantly improve the efficiency, consistency, and reliability of malaria diagnostic workflows.

II. Methodology

A. System Overview and Architecture

The proposed system follows a multi-stage diagnostic workflow in which staining quality assessment, blood smear quality analysis, and automated parasite detection are applied as successive validation stages. At each stage, slides that do not meet predefined quality criteria are identified and excluded from further processing. The overall architecture of the proposed system is illustrated in Fig. 1.

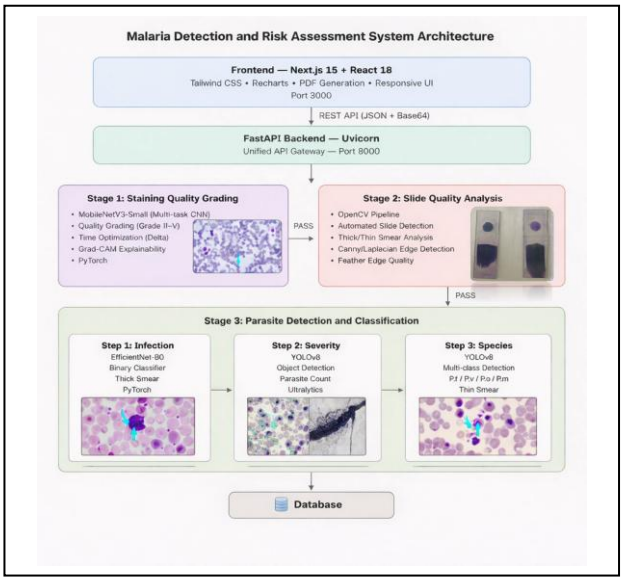

Fig. 1. Overall architecture of the proposed integrated malaria diagnostic system

Initially, microscopic images of thick and thin blood smears are used to evaluate staining quality and determine the optimal staining duration. This stage ensures that staining conditions are suitable for accurate interpretation. Subsequently, the entire slide is captured using a mobile

device, and the resulting image is analyzed to assess smear quality based on macroscopic features such as smear dimensions, feather edge formation, spacing, and stain density. Only slides satisfying predefined quality criteria are forwarded to the final stage, where microscopic images are used for parasite detection, species classification, and severity estimation.

This staged structure reflects a practical constraint of routine laboratory settings: preparation deficiencies that go undetected before microscopic examination add to technician workload and increase the risk of misdiagnosis. By validating slide quality at each stage, the system limits the exposure of downstream detection models to inputs that do not meet established preparation standards

B. Dataset Description

The system is developed and evaluated using a combination of locally collected and publicly available datasets, all of which were obtained under the supervision or validation of the Anti-Malaria Campaign (AMC), Sri Lanka. This ensures that the datasets reflect real-world laboratory conditions and maintain clinical reliability.

1) Staining Quality Dataset

The staining quality assessment component is trained and evaluated on a dataset of 466 microscopic images collected in collaboration with laboratory experts at AMC. The dataset includes both thick and thin smears prepared under varying Giemsa stain concentrations spanning staining grades II through V.

To account for the relatively small dataset size and prevent data leakage, a batch-aware five-fold cross-validation strategy is employed. Images originating from the same staining batch are assigned to the same fold, ensuring model evaluation on staining conditions not represented in the training set.

2) Smear Quality Assessment Dataset

The smear quality assessment component is developed and validated using slide images collected during AMC field and laboratory activities. The dataset includes images captured from two commonly used slide types (25 mm × 75 mm and 26 mm × 76 mm), reflecting variations encountered in practical laboratory settings.

An initial dataset of 293 slide images was used during system development. These images were categorized into three quality levels—Good, Average, and Non-accurate (indicating insufficient quality for diagnostic use)—based on assessments provided by AMC laboratory personnel.

A separate validation dataset of 140 slide images was collected to assess the robustness of the proposed approach. Each image in this dataset is accompanied by ground truth measurements provided by experienced laboratory personnel, enabling quantitative comparison between system outputs and manual measurements.

3) Parasite Detection Dataset

Data for the parasite detection component was sourced from publicly available malaria datasets and microscopy images collected in collaboration with AMC. All datasets were reviewed and validated by AMC experts to ensure annotation consistency and clinical relevance.

The thick smear dataset consists of 1182 images annotated at the image level (infected and uninfected) and at the parasite level, where individual parasite instances are marked using bounding boxes. The thin smear dataset contains approximately 230 images encompassing multiple *Plasmodium* species.

Both datasets are partitioned into training, validation, and testing subsets following a 70:15:15 ratio to enable reliable model evaluation.

Representative input images for each system component are illustrated in Fig. 2. A consolidated summary of the datasets used is provided in Table I.

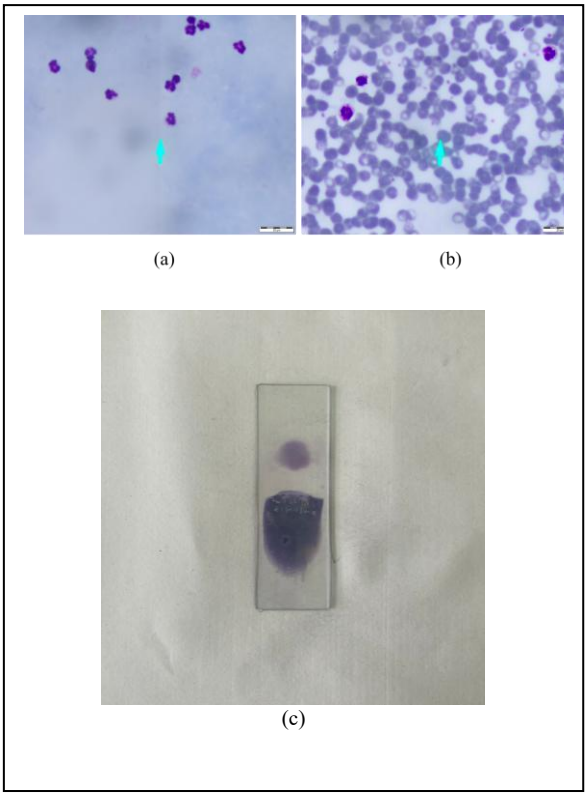

Fig. 2. Representative input images used in the proposed system, including (a) thick smear microscopic image, (b) thin smear microscopic image, and (c) full slide image captured using a mobile device

Table I. Dataset Summary

Component	Dataset Type	Number of Images	Annotation
Staining	Cross-validation	466	Grade + Time
Smear Analysis	Development	293	Quality Labels
	Validation	140	Labels + Measurements
Parasite Detection (Thick)	Train/Validation/Test	1182	Infection + Bounding boxes
Parasite Detection (Thin)	Train/Validation/Test	~230	Species + Segmentation

C. Staining Quality Assessment Method

The staining quality assessment component is implemented using a multi-task deep learning model based on the MobileNetV3-Small architecture. The model jointly predicts the staining quality grade and the corresponding staining duration adjustment, providing both a quality assessment and a quantitative basis for corrective action.

Microscopic images are resized and normalized prior to model input. Data augmentation techniques, including rotation, flipping, and controlled color variations, are applied during training to account for variability in staining intensity and illumination conditions commonly observed in laboratory environments.

The staining quality prediction task is formulated as an ordinal classification problem to preserve the inherent ordering between staining grades, reducing misclassification between non-adjacent classes. In parallel, a regression branch estimates the required staining duration adjustment, providing technicians with a quantitative basis for adjusting staining duration.

Training proceeds in two stages: in the first, the feature extraction layers are frozen to preserve pretrained representations while task-specific layers adapt to the staining dataset; in the second, the entire network is fine-tuned using differential learning rates.

Model performance is evaluated using accuracy, precision, recall, and F1-score across all staining quality grades.

D. Smear Quality Assessment Method

The smear quality assessment component is implemented using classical computer vision techniques, as this allows each measurement to be traced back to explicit algorithmic steps rather than learned feature representations, producing outputs that can be directly verified against established laboratory criteria.

Input images undergo preprocessing to reduce noise and normalize illumination conditions. The microscope slide is automatically detected and isolated using contour-based analysis, ensuring that all subsequent measurements are confined to the slide region.

To obtain physically meaningful measurements, pixel-based distances are converted into real-world units using pixel-to-millimeter calibration. The calibration factor is derived from the detected slide width and the known physical dimensions of standard microscope slides. Two commonly used slide sizes are supported—25 mm × 75 mm and 26 mm × 76 mm—with the calibration factor adapting dynamically based on the detected slide width to maintain consistent measurement accuracy across both formats. The calibration factor is computed as:

$$pixels\_per\_mm = \frac{W_{pixels}}{W_{mm}} \quad (1)$$

where W_{pixels} represents the detected slide width in pixels and W_{mm} corresponds to the actual slide width in millimeters. Measurements are then converted as:

$$Measurement_{mm} = \frac{Measurement_{pixels}}{pixels\_per\_mm} \quad (2)$$

Following calibration, the thick smear region is detected using thresholding and contour-based analysis, and its diameter is computed in millimeter units. The thin smear region is identified using intensity-based segmentation to capture its full extent, with length estimated using bounding geometry.

Feather edge detection is performed using gradient and texture-based analysis to identify the tapered region of the thin smear where cells are distributed in a monolayer. This region is critical for accurate microscopic examination. Spatial relationships between smear components are evaluated by measuring distances between key regions of the slide.

In addition to geometric and spatial features, stain density is evaluated using region-based intensity analysis within the thin smear. The smear is divided into body and feather regions, and mean intensity values are computed to characterize cell overlap. Lower intensity values indicate excessive staining density, where red blood cells overlap and obscure microscopic visibility, while higher intensity values correspond to well-distributed monolayer regions. Predefined intensity thresholds are used to determine whether the smear provides sufficient clarity for parasite examination—a distinction particularly relevant for slides that pass geometric criteria but remain unsuitable for reliable diagnostic use.

Slides are classified as Good, Average, or Non-accurate according to predefined laboratory standards, with the classification derived from the combined evaluation of geometric, spatial, and intensity-based criteria described above. Measurement accuracy is evaluated using Mean Absolute Error (MAE), which quantifies the difference between system-generated measurements and expert-provided ground truth values.

E. Parasite Detection and Classification Method

The parasite detection component is designed as a multi-stage deep learning pipeline covering infection screening, parasite localization, and species identification.

Microscopic images are resized and normalized prior to model input. Data augmentation is applied during training to account for variability in imaging conditions and parasite morphology. An EfficientNet-based classification model is applied to thick smear images to determine infection status, with transfer learning used to compensate for limited training data availability.

For samples identified as infected, a YOLOv8-based object detection model is applied to localize parasite instances within thick smear images. The total parasite count derived from detection output is used to estimate infection severity. The detection model is optimized for small object detection in complex backgrounds, which is a key challenge in malaria microscopy.

Species identification is performed using a YOLOv8-based segmentation model applied to thin smear images. Unlike bounding box detection, instance segmentation preserves the morphological boundaries of parasite

structures, which is particularly relevant for distinguishing between *Plasmodium* species with overlapping visual characteristics.

Together, the three stages produce infection status, parasite count, severity estimation, and species identification as discrete diagnostic outputs. Model performance is evaluated using precision, recall, and mean Average Precision at IoU threshold 0.5 (mAP@0.5).

III. Results and Discussion

This section reports the experimental results of the proposed system across its three primary components: staining quality assessment, smear quality analysis, and parasite detection. The performance of each component is examined quantitatively and qualitatively, followed by a discussion of the integrated system behavior.

A. Staining Quality Assessment Results

The staining quality assessment model was evaluated using a batch-aware five-fold cross-validation strategy to ensure that model performance reflects generalization to unseen staining batches rather than memorization of batch-specific characteristics.

The classification performance across staining grades is summarized in Table II.

Table II. Staining Classification Performance

Grade	Accuracy (%)
Grade II	45.9
Grade III	55.2
Grade IV	42.6
Grade V	50.1

Grade III, representing optimal staining conditions, achieved the highest per-class accuracy at 55.2%, which is the most clinically significant result as it reflects the model's ability to correctly identify slides suitable for diagnostic use. Grade IV exhibited the lowest accuracy at 42.6%, attributable to two compounding factors: its visual similarity to adjacent grades and its disproportionate representation in the training dataset relative to other classes.

In addition to classification performance, the system generates structured outputs that translate model predictions into actionable laboratory guidance. A representative example of the system output is summarized in Table III, illustrating how predicted staining grades, confidence scores, and time adjustment recommendations are combined to support staining optimization.

Table III. Example System Output Demonstrating Automated Staining Quality Assessment and Time Optimization

Feature	Measured Value	Clinical Interpretation	Assessment
Predicted Grade	Grade III	Optimal staining range	Acceptable
Model Confidence	52.4%	Moderate confidence prediction	Reliable
Staining Time (Input)	27 minutes	Within operational range	Valid input
Recommended Time Adjustment	-2.4 minutes	Slight over-staining detected	Reduce time
Optimal Time Estimate	5 minutes	Time achieving Grade III	Optimal
Grade II Probability	24.5%	Under-stained likelihood	Low
Grade III Probability	52.4%	Highest likelihood (optimal)	Selected
Grade IV Probability	10.6%	Slight over-staining possibility	Low
Grade V Probability	12.4%	Slight over-staining possibility	Low
Final Decision		**Meets AMC optimal staining criteria**	**Pass**

The table demonstrates that the system not only classifies staining quality but also provides interpretable decision criteria aligned with laboratory standards, including identification of optimal staining conditions and corrective recommendations.

The time prediction branch achieved a Mean Absolute Error (MAE) of approximately 3.75-4.25 minutes across validation folds. Given that staining durations range from 5 to 41 minutes depending on dilution concentration, this margin corresponds to less than 10% of the full protocol range, placing it within a practically acceptable tolerance for laboratory correction guidance.

Qualitative analysis using Grad-CAM visualizations further indicates that the model consistently attends to regions of staining intensity variation and cell contrast — features that laboratory personnel prioritize during manual staining assessment.

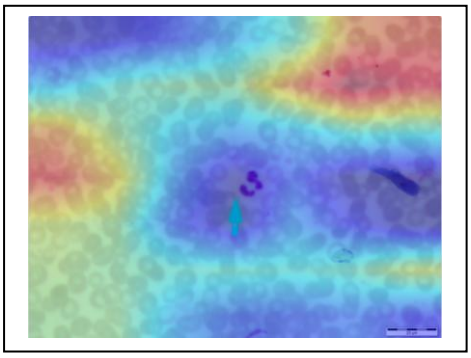

Fig. 3. Grad-CAM visualization of staining quality classification highlighting regions influencing the model's decision

As shown in Fig. 3, the model focuses on diagnostically relevant regions, particularly areas of cell distribution and stain intensity variation, demonstrating alignment with expert-defined visual assessment criteria. This correspondence provides additional confidence in the reliability and interpretability of the model beyond aggregate performance metrics.

B. Smear Quality Assessment Results

The smear quality assessment component was evaluated using an independent validation dataset of 140 slide images,

each annotated with expert-provided ground truth measurements. The evaluation focused on the agreement between system-generated measurements and manual measurements.

Measurement accuracy was assessed using Mean Absolute Error (MAE) across multiple geometric features, including thick smear diameter, thin smear length, and spatial distances between smear regions. A comparison between expert measurements and system outputs for representative slides is presented in Table IV.

Table IV. Measurement Comparison Between Manual and System

Slide	Measurement	Manual (mm)	System (mm)	Error (mm)
Slide_001	Thick Smear	11.0	10.47	0.53
	Thin Smear	33.0	27.95	5.05
	Top-Thick Gap	10.0	8.80	1.20
	Thick-Thin Gap	8.0	7.00	1.00
Slide_002	Thick Smear	13.0	11.34	1.66
	Thin Smear	23.0	21.63	1.37
	Top-Thick Gap	12.0	10.20	1.80
	Thick-Thin Gap	10.0	8.20	1.80
Slide_003	Thick Smear	10.0	9.50	0.50
	Thin Smear	22.0	25.85	3.85
	Top-Thick Gap	20.0	17.90	2.10
	Thick-Thin Gap	13.0	12.70	0.30

The system achieved an average error of approximately **1.76 mm** across all measurement types, a margin that falls within the tolerance range considered acceptable for laboratory slide evaluation.

The largest individual errors were observed in thin smear length measurements, particularly in cases where smear boundaries were not clearly defined or where manual reference values themselves carried estimation uncertainty. The system nonetheless maintained consistent proportional relationships between smear components across all evaluated slides, which is the functionally critical criterion for determining slide suitability.

In addition to geometric measurements, the system produced structured feedback identifying common preparation deficiencies including excessive smear thickness, improper spacing, and inadequate feather edge formation. Because each output is derived from explicit algorithmic steps rather than learned representations, the basis for each quality determination can be directly examined—a property that distinguishes this approach from black-box classification models and is particularly relevant in laboratory settings where preparation feedback must be actionable.

Representative outputs of the smear quality assessment component are shown in Fig. 4, illustrating smear region detection—including thick, thin, and feather edge identification—spatial measurement of smear positioning, and stain density analysis. A representative example of the system-generated quantitative output and final decision is summarized in Table V, demonstrating how individual measurements and thresholds are combined to produce an overall smear quality assessment.

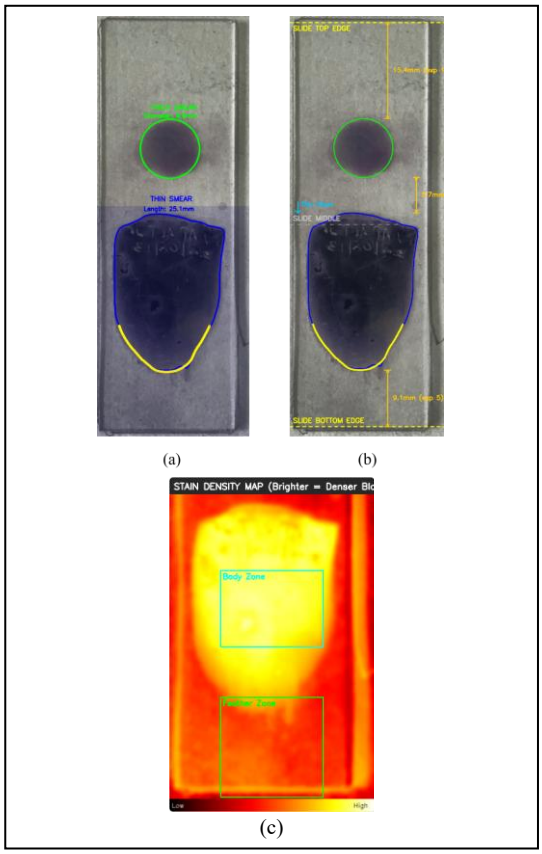

Fig. 4. Smear quality assessment outputs. (a) Detection of thick (green) and thin smear (blue) regions with feather edge (yellow) identification. (b) Spacing and positioning analysis. (c) Stain density heatmap visualization

Table V. Example of system-generated smear quality assessment output, showing measured features, corresponding expected ranges, and rule-based evaluations used to derive the final slide quality decision

Feature	Measured Value	Expected Range	Assessment
Thick Smear Diameter	9.56 mm	9–11 mm	Good
Thin Smear Length	25.10 mm	20–30 mm	Good
Slide Top → Thick Smear	15.4mm	~10 mm (7-13 mm acceptable)	Average
Slide Top → Thick Smear	5.7mm	~10 mm (7-13 mm acceptable)	Average
Thin Tip → Slide Bottom	9.1mm	~5 mm (3-7 mm acceptable)	Average
Thin Smear Center Offset	1.9mm	≤5 mm	Good
Body Region Intensity	37	High intensity threshold	Too Dense
Feather Region Intensity	122	Optimal range	Optimal
Feather Edge Quality	-	Structural criteria	Ideal
Overall Assessment	**-**	**-**	**Average**

C. Parasite Detection Results

The parasite detection component was evaluated using a standard 70:15:15 train–validation–test split across classification, detection, and segmentation tasks.

Performance metrics for the detection and species identification models are summarized in Table VI.

Table VI. Parasite Detection and Classification Performance

Component	Precision	Recall	mAP@0.5
Detection Model	0.783	0.771	0.812
Species Identification	0.866	0.824	0.897

Confusion matrices for the infection classification model on the validation and test sets are shown in Fig. 5. The final model achieved 33 true negatives and 135–140 true positives on the test set, with false negatives reduced to 3–8 cases — reflecting the prioritization of infected case sensitivity during training.

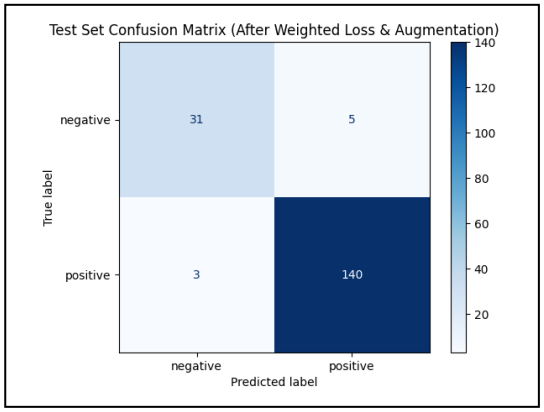

Fig. 5. Confusion matrix for infection classification model on the test dataset

The infection classification model achieved a test accuracy of 95.6%, with precision of 0.97 and recall of 0.98 for infected cases, reflecting the prioritization of sensitivity during training to minimize false negatives in clinical use. Specific classification metrics are reported in Table VI.

The detection model achieved a mAP@0.5 of 0.812, demonstrating reliable localization of parasite instances across the test set. The primary sources of error were very small parasites and instances where parasite morphology overlapped with background artifacts, both of which contributed to the observed false positive and missed detection cases.

The segmentation-based species identification model achieved a mAP@0.5 of 0.897, with strong performance for dominant species such as *Plasmodium falciparum*, while performance for minority species was comparatively lower, reflecting the combined effect of class imbalance and morphological similarity between certain Plasmodium species.

Representative detection and segmentation outputs are shown in Fig. 6, illustrating bounding box-based parasite localization in thick smear images and segmentation-based species identification in thin smear images. In addition to visual outputs, the system generates structured quantitative results including infection status, parasite count, severity estimation, and species classification.

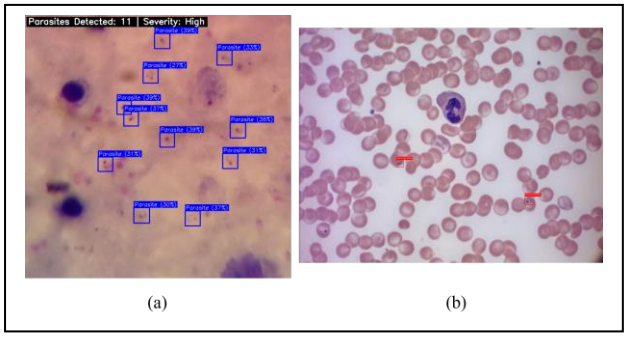

Fig. 6. Parasite detection and segmentation outputs. (a) Thick smear with bounding box detections. (b) Thin smear with segmentation-based species identification

A representative example of these system-generated outputs is summarized in Table VII, demonstrating how model predictions are translated into clinically interpretable decisions, including infection confirmation, parasite burden estimation, and risk-level classification.

Table VII. Example of system-generated parasite detection and classification output

Component	Measured Value	Expected/ Interpretation	Assessment
Infection Status	Positive	Parasites detected	Infected
Confidence Score	100.0%	High confidence threshold	Reliable
Parasite Count	11	1-4 → Low 5-10 → Medium >10 → High	High Severity
Detection Output	11 bounding boxes	Localization of parasite instances in image	Accurate
Species Identified	Plasmodium falciparum	Localization of parasite	High Risk
Species Confidence	56.0%	Model prediction confidence	Moderate
Overall Assessment	-	**Combined decision**	**High Severity Case**

Collectively, the three-stage pipeline produces infection status, parasite count, severity estimation, and species identification as discrete outputs. The primary limitations—small parasite detection and minority species classification—represent the most direct targets for improvement through targeted dataset expansion

D. Integrated System Discussion

Collectively, the results confirm that staining assessment, smear quality evaluation, and parasite detection can function as a coherent sequential framework, with upstream quality

checks consistently limiting the exposure of detection models to substandard inputs.

The sequential filtering approach addresses a practical inefficiency in current laboratory workflows, where preparation errors are frequently identified only after microscopic examination has been completed. By contrast, the proposed system surfaces preparation deficiencies before the computationally intensive detection stage, reducing both the analytical burden on laboratory personnel and the volume of slides requiring repeated preparation.

The combination of rule-based computer vision and deep learning components means that outputs from the smear analysis module can be directly traced to explicit algorithmic steps and verified against laboratory criteria, while the learning-based components handle classification and detection tasks where handcrafted rules would be insufficient. This distinction is practically relevant in clinical laboratory settings, where the auditability of automated decisions carries as much weight as predictive accuracy.

The findings suggest that incorporating structured quality validation as a prerequisite to parasite analysis is both technically feasible and operationally relevant to routine diagnostic workflows—a conclusion supported by the close agreement between system outputs and expert assessments across all three components.

IV. Conclusion and Future Work

This study presented a sequential diagnostic framework for malaria in which staining quality assessment, smear quality evaluation, and automated parasite detection are applied as successive validation stages. The proposed system addresses a key limitation in conventional workflows, where slide preparation quality is typically evaluated only after microscopic examination has begun, often leading to unnecessary analytical effort on substandard slides.

Each component demonstrated performance consistent with expert assessment. The staining model provided quality grading and time adjustment predictions, while Grad-CAM analysis indicated that the model focuses on visually relevant regions used in manual evaluation. The smear quality assessment module achieved an average measurement error of approximately 1.76 mm when compared with expert-provided ground truth values. The parasite detection and species identification components achieved mAP@0.5 scores of 0.812 and 0.897, respectively, demonstrating strong performance when operating on quality-filtered inputs.

The sequential design ensures that slide preparation deficiencies are identified prior to microscopic analysis, reducing unnecessary workload and improving diagnostic reliability. The use of explicit algorithmic steps in the smear assessment module ensures that quality determinations can be directly audited, complementing the learning-based detection components where rule formulation would be impractical.

Despite these strengths, several limitations remain. The staining model was trained on data collected from a single laboratory, which may limit generalizability across different environments. The parasite detection component is affected by class imbalance and limited annotated data for certain Plasmodium species. In addition, the smear quality module is sensitive to variations in image acquisition conditions, including lighting and camera positioning.

Future work will focus on expanding the dataset across multiple laboratory settings to improve robustness and generalization. Enhancements to the staining model will target improved time prediction accuracy and better representation of underrepresented staining conditions. For parasite detection, efforts will focus on improving performance in challenging scenarios such as low parasite density and overlapping cells, as well as improving species classification through more balanced datasets.

In the longer term, deployment across multiple laboratory sites would enable systematic collection of staining and preparation data at scale, providing an empirical basis for standardizing laboratory procedures across diverse clinical settings.

Acknowledgment

The authors gratefully acknowledge Dr. Kumudu Gunasekara, Parasitologist at the Anti-Malaria Campaign (AMC), Sri Lanka, for her expert guidance and valuable insights throughout this study, and Ms. Shanika Wickramasinghe at AMC for her continuous support during data collection and contributions to domain knowledge. The authors also thank the laboratory technicians at AMC for their assistance during data collection and validation. Their practical expertise was instrumental in ensuring the clinical relevance and reliability of the datasets used in this research.

References

[1] W. M. F. Amaris, C. Martinez, L. J. Cortés-Cortés, and D. R. Suárez, “Image features for quality analysis of thick blood smears employed in malaria diagnosis,” Malar. J., vol. 21, no. 1, 2022, doi: 10.1186/s12936-022-04064-2.

[2] S. Boit and R. Patil, “An efficient deep learning approach for malaria parasite detection in microscopic images,” Diagnostics, vol. 14, no. 23, 2024, doi: 10.3390/diagnostics14232738.

[3] B. Lufyagila, B. Mgawe, and A. Sam, “Fine-tuned YOLO-based deep learning model for detecting malaria parasites and leukocytes in thick smear images: A Tanzanian case study,” Mach. Learn. Appl., vol. 21, 2025, doi: 10.1016/j.mlwa.2025.100687.

[4] S. Sawant and A. Singh, “Malaria cell detection using deep neural networks,” arXiv preprint arXiv:2406.20005, 2024.

[5] “Giemsa staining of malaria blood films: Malaria microscopy standard operating procedure MM-SOP-07A 1. Purpose and scope.”

[6] F. Abdurahman, K. A. Fante, and M. Aliy, “Malaria parasite detection in thick blood smear microscopic images using modified YOLOv3 and YOLOv4 models,” BMC Bioinformatics, vol. 22, no. 1, 2021, doi: 10.1186/s12859-021-04036-4.

[7] Y. Hirimutugoda and G. Wijayarathna, “Image analysis system for detection of red cell disorders using artificial neural networks,” Sri Lanka J. Bio-Medical Informatics, vol. 1, no. 1, p. 35, 2010, doi: 10.4038/sljbmi. v1i1.1484.

[8] T. Sungkapong, N. Chaianantakul, T. Dangsompong, and N. Weaingnak, “The effect of different Giemsa staining conditions on thin blood film malaria identification,” 2018.

[9] H. Kim, M. Hur, G. d’Onofrio, and G. Zini, “Real-world application of digital morphology analyzers: Practical issues and challenges in clinical laboratories,” Diagnostics, 2025, doi: 10.3390/diagnostics15060677.

[10] A. Tao and B. Han, “Deep unsupervised learning for microscopy-based malaria detection,” arXiv preprint arXiv:2009.00197, 2020.

[11] S. McDermott et al., “Autohaem: 3D printed devices for automated preparation of blood smears,” Rev. Sci. Instrum., vol. 93, no. 1, 2022, doi: 10.1063/5.0076901.

[12] R. C. L. Hadiana, A. K. Setya, S. T. I. K. Nasional, and J. R. Solo-Baki, “Optimisation of 3% Giemsa staining time in malaria microscopic examination,” Int. J. Global Health Res., doi: 10.37287/ijghr. v7i3.6229.

[13] “Quality control of Giemsa stock solution and buffered water: Malaria microscopy standard operating procedure MM-SOP-03C 1. Purpose and scope.”

[14] “Preparation of Giemsa working solution: Malaria microscopy standard operating procedure MM-SOP-04.”

[15] E. Simson, M. G. Gascon-Lema, and D. L. Brown, “Performance of automated slidemakers and stainers in a working laboratory environment—routine operation and quality control,” Int. J. Lab. Hematol., vol. 32, no. 1, 2010, doi: 10.1111/j.1751-553x.2009.01141. x.

[16] H. B. Baydargil and T. Bocklitz, “Unstained blood smear analysis: A review of rule-based, machine learning, and deep learning techniques,” J. Biophotonics, 2025, doi: 10.1002/jbio.202500121.

[17] A. Koirala et al., “Deep learning for real-time malaria parasite detection and counting using YOLO-mp,” IEEE Access, vol. 10, pp. 102157–102172, 2022, doi: 10.1109/ACCESS.2022. 3208270.