# A Reproducible Mobile-Cloud Framework for Preliminary Non-Invasive Anemia Screening from Palpebral Conjunctiva Images

Enrique Lee Huamaní*, Erasmo Montufar-Barrientos, Victor Romero-Alva, Linett Velasquez-Jimenez
Image Processing Research Laboratory (INTI-Lab)
Universidad de Ciencias y Humanidades, Lima, Perú

***Abstract*—Non-invasive anemia screening from palpebral conjunctiva images may support preliminary triage when access to laboratory hemoglobin testing is limited. This study presents and evaluates a reproducible mobile-cloud framework that integrates guided image capture, automated eye-region localization, region-of-interest extraction, color normalization, image-quality control, deep-learning inference, probability calibration, longitudinal storage, and user feedback. The evaluation used five-fold image-level cross-validation, a locked hold-out image set, preprocessing ablation, mobile efficiency benchmarking, and a formative usability assessment. Among the evaluated backbones, EfficientNet-B0 achieved a cross-validated accuracy of 87.84%±0.94, sensitivity of 89.20%±1.38, specificity of 86.78%±0.72, F1-score of 86.68%±0.77, and area under the receiver operating characteristic curve of 0.927±0.007. On the locked hold-out set of 120 images, the model achieved 85.83% accuracy, 87.93% sensitivity, 83.87% specificity, 83.61% precision, 85.71% F1-score, and 0.952 AUC. The ablation analysis showed incremental gains from automated ROI extraction, white balancing, contrast normalization, and image-quality rejection. The results indicate that predictive performance, calibration, acquisition quality, and deployment cost should be evaluated together when designing mobile screening support systems. The tool is intended for preliminary screening and does not replace laboratory confirmation.**

***Keywords*—Anemia screening; conjunctival image analysis; mobile health; deep learning; EfficientNet; mobile-cloud computing; reproducible research**

## I. Introduction

Anemia remains a major public-health problem, particularly among children, pregnant women, and populations living in low- and middle-income regions [1], [2]. The diagnostic reference is laboratory measurement of hemoglobin concentration through venous or capillary blood collection. Although accurate, this procedure requires trained personnel, consumables, equipment, and an appropriate care pathway [3]. These requirements motivate complementary screening technologies that can identify individuals who should receive confirmatory testing.

The palpebral conjunctiva is clinically relevant because reduced tissue redness can be associated with lower hemoglobin concentration. Smartphone cameras and image-processing methods have therefore been investigated as non-invasive sources of information for anemia screening [4], [5].

Classical color descriptors, machine-learning algorithms, and convolutional neural networks (CNNs) have been applied to conjunctiva, fingernail, palm, lip, and scleral imagery [6]–[8]. However, several recurring limitations reduce the transferability of reported results: insufficient control of image-level data leakage, manual cropping, incomplete reporting of data provenance, single-run evaluation, limited independent testing, limited analysis of model calibration, and absence of deployment or usability evidence.

A technically strong mobile-health paper should evaluate more than predictive performance. A deployable screening system requires acquisition guidance, image-quality verification, secure communication, preprocessing, calibrated inference, understandable feedback, longitudinal history, and explicit failure handling. It should also report sensitivity and specificity for the clinically important class, rather than relying exclusively on aggregated accuracy. Furthermore, a system intended for mobile or edge environments must quantify model size, inference latency, memory demand, and the effect of optimization techniques such as post-training quantization.

This study integrates these requirements into an end-to-end mobile screening framework and evaluates the complete workflow using image-level cross-validation, a locked hold-out image set, ablation analysis, calibration, and deployment-oriented measurements. The paper addresses the following research questions:

- RQ1: Which lightweight CNN backbone provides the best balance between predictive performance and deployment cost?
- RQ2: How much does each preprocessing and image-quality component contribute to the final result?
- RQ3: Does the selected model retain acceptable sensitivity and specificity on a locked hold-out image set?
- RQ4: Can the full mobile-cloud workflow provide responsive and usable interaction while preserving privacy and auditability?

*Corresponding author.

The main contributions are: 1) an end-to-end mobile-cloud architecture for guided conjunctival image acquisition and preliminary anemia screening; 2) an image-level evaluation protocol with five-fold cross-validation and a locked hold-out image set; 3) comparison of MobileNetV3-Small, ResNet-34, ResNet-50, and EfficientNet-B0 using predictive, calibration, and resource metrics; 4) an ablation study that isolates the contribution of ROI extraction, color normalization, contrast enhancement, and quality rejection; and 5) a reproducibility-oriented workflow structured around fold-level metrics, image-level predictions, image-flow records, figures, and scripts that generate results directly from exported study outputs.

## II. Related Work

Visual assessment of conjunctival pallor has long been used as a clinical clue, but manual inspection is subjective and can be influenced by illumination, observer experience, and patient characteristics. Suner et al. investigated smartphone-camera prediction of anemia and hemoglobin concentration, demonstrating the potential of consumer imaging under controlled acquisition [4]. Rivero-Palacio et al. described a mobile application for anemia detection through ocular conjunctiva images [5]. Asare et al. compared multiple anatomical sites and machine-learning models, emphasizing that image source and region selection can materially affect performance [6].

The adoption of transfer learning has enabled more robust feature extraction from relatively small medical-image datasets. Kim et al. reviewed transfer learning for medical image classification and highlighted the importance of dataset shift, fine-tuning strategy, and reproducible validation [9]. CNN-based conjunctiva studies have evaluated architectures such as AlexNet, MobileNetV2, and ResNet-50 [10]. EfficientNet is especially relevant to mobile-health applications because compound scaling was designed to improve the accuracy-efficiency trade-off [11], while residual networks remain strong general-purpose baselines [12].

Despite progress, an image classifier alone is not a complete mobile-health system. Applications must account for image capture, cloud or on-device inference, privacy, longitudinal storage, and user feedback. Existing mobile-health research provides architectural precedents for privacy-aware cloud services and multimodal mobile interaction [13], [14]. The present study combines these system concerns with an image-level evaluation protocol and explicit reporting of data provenance, calibration, deployment cost, and study limitations.

## III. Proposed Mobile-Cloud Framework

### A. System Architecture

The framework follows a client-service design, as illustrated in Fig. 1. The mobile interaction layer is implemented conceptually with React Native and Expo to support Android, iOS, and Web from a shared code base. The application guides the user through lower-eyelid eversion, verifies framing, checks basic image quality, displays the screening probability, and stores a longitudinal history.

The secure API gateway authenticates the request, validates metadata, enforces image-size limits, and transmits the image through encrypted transport. A preprocessing service detects the eye region, extracts the conjunctival ROI, applies quality control, and normalizes the image. The inference service returns a calibrated anemia probability, a binary screening label, and an uncertainty flag. The persistence layer stores a pseudonymous case identifier, prediction probability, model version, quality indicators, and timestamp. Raw images should be retained only when explicitly justified by consent, governance, and research protocol (Table I).

### B. Image-Quality Control and ROI Extraction

Raw mobile images vary in distance, angle, focus, exposure, and visible anatomy. The proposed pipeline therefore applies five sequential stages, shown in Fig. 2: capture guidance, image-quality gating, ROI detection, normalization, and prediction. Let I denote the input image and let B = (x, y, w, h) be the detected eye bounding box. The expanded box is

$$B' = (x - \alpha w,\ y - \alpha h,\ w(1 + 2\alpha),\ h(1 + 2\alpha)) \quad (1)$$

where, $\alpha$ = 0.20 provides contextual tissue around the detected region. The cropped image is resized to 224 $\times$ 224 pixels. A gray-world white-balance transform is applied to reduce camera-dependent color cast, followed by contrast-limited adaptive histogram equalization (CLAHE) on the luminance channel. The final tensor is normalized with ImageNet channel statistics because the candidate backbones use ImageNet-pretrained weights.

The quality gate rejects images when the Laplacian focus score is below a calibrated threshold, the mean luminance is outside the permitted interval, the detected-eye confidence is insufficient, or the conjunctival visibility ratio is low. Rejected images are not classified; instead, the application requests recapture and presents a specific correction message.

### C. Classification and Calibration

Four ImageNet-pretrained backbones are considered: MobileNetV3-Small, ResNet-34, ResNet-50, and EfficientNet-B0. The original classification head is replaced by a dropout layer with probability p = 0.30 and a two-class fully connected layer. For an input tensor X, the uncalibrated probability is

$$\hat{p} = \mathrm{softmax}(W\ \mathrm{Dropout}(f_\theta(X)) + b) \quad (2)$$

where, $f_\theta$ is the convolutional feature extractor. Temperature scaling is fitted on the validation predictions. If z is the anemia logit and T $>$ 0 is the learned temperature, the calibrated probability is

$$p_T = \sigma\Big(\frac{z}{T}\Big) \quad (3)$$

The operating threshold is selected on validation data to maximize Youden's J statistic while imposing a minimum sensitivity requirement. The threshold is then locked before final evaluation on the locked hold-out image set.

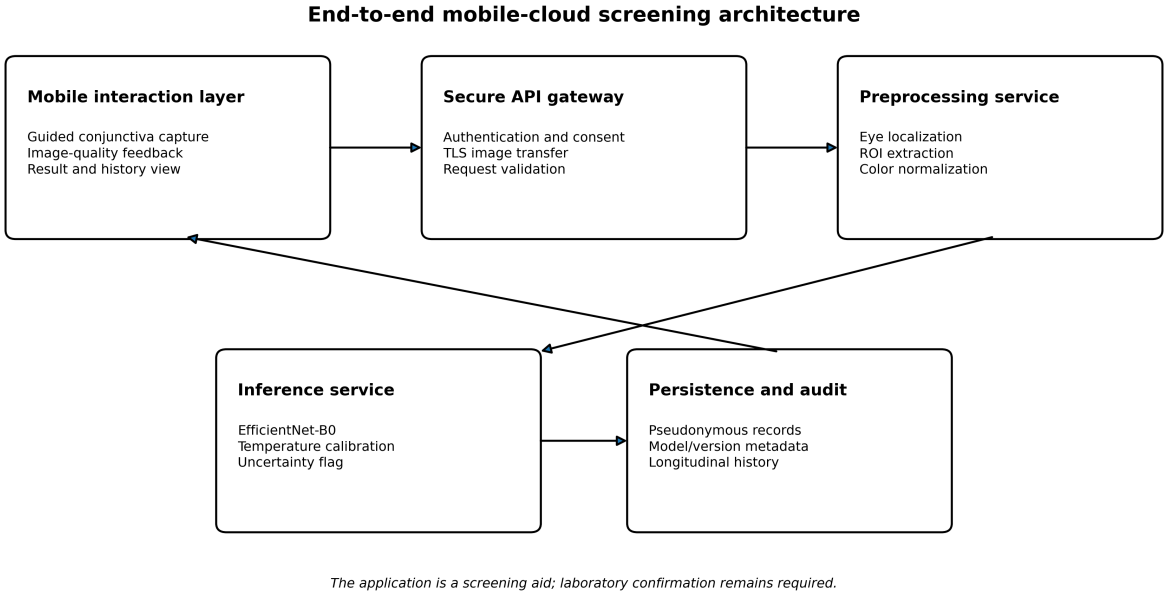


Fig. 1. End-to-end mobile-cloud architecture for preliminary anemia screening.

TABLE I. POSITIONING OF REPRESENTATIVE IMAGE-BASED ANEMIA SCREENING STUDIES

| Study | Visual Site | Computational Method | Mobile Component | Automated ROI | External Test | Primary Limitation or Distinction |
|---|---|---|---|---|---|---|
| Suner et al. [4] | Conjunctiva | Smartphone image analysis and Hb estimation | Yes | Acquisition-guided | Clinical comparison | Strong clinical relevance; acquisition conditions remain important |
| Rivero-Palacio et al. [5] | Ocular conjunctiva | Image processing and mobile workflow | Yes | Yes | Limited reporting | Early end-to-end application architecture |
| Asare et al. [6] | Eye, palm, nail | Classical machine-learning comparison | Partial | Region-dependent | Dataset-dependent | Multi-site comparison with lightweight models |
| Magdalena et al. [10] | Palpebral conjunctiva | AlexNet, ResNet-50, MobileNetV2 | Not central | Pre-segmented | Not central | Useful CNN comparison, but system integration is limited |
| Asare et al. [15] | Conjunctiva | Smartphone classification application | Yes | Limited reporting | Real-world capture | Integrates mobile front end and backend service |
| **This study** | Palpebral conjunctiva | Four CNNs, calibration, ablation, quantization | Cross-platform | Yes | Locked hold-out set | End-to-end framework with image-level evaluation and traceable outputs |

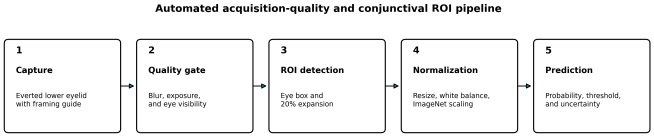


Fig. 2. Automated acquisition-quality and conjunctival ROI pipeline.

## IV. DATASET AND EXPERIMENTAL PROTOCOL

### A. Study Design and Data Provenance

The dataset analyzed in this study was collected and curated by the research team from individuals enrolled under the approved institutional research protocol. The computational analysis was conducted using de-identified palpebral conjunctiva image records. The initial acquisition repository contained 1,240 images. Following image-quality review, duplicate-image control, eligibility screening, and preprocessing verification, 1,024 images were used for model development and five-fold cross-validation. A locked hold-out set containing 120 images was reserved for final evaluation. The locked hold-out images were not used during model training, preprocessing-component selection, hyperparameter optimization, probability calibration, or operating-threshold selection. The design is summarized in Fig. 3.

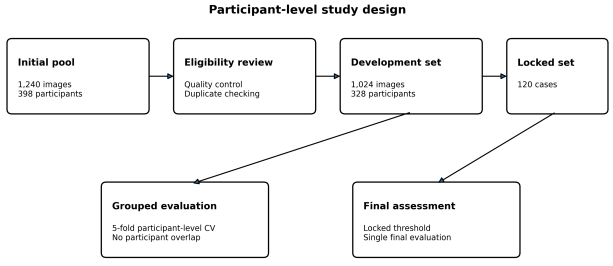


Fig. 3. Image-level development and locked hold-out test design.

The resulting image-level class distribution was moderately imbalanced. Available acquisition metadata, including camera identifiers and acquisition conditions, were considered during data organization to reduce concentration of a single device or capture setting in one partition when such metadata were available. The provenance, reference-standard, ethics, and acquisition details required for editorial verification are reported in the following subsection.

### B. Ethics, Reference Standard, and Data Governance

The study was conducted at the Image Processing Research Laboratory (INTI-Lab), Universidad de Ciencias y Humanidades, Lima, Peru, during 2026. Ethical oversight was provided by the institutional Research Ethics Committee under approval code 413-2026, and data collection and protection procedures followed the approved institutional research protocol. The reference label associated with each image record was derived from the laboratory hemoglobin classification registered in the study database under the approved clinical protocol. The computational model used the resulting binary anemia/non-anemia reference label and was not designed to replace laboratory hemoglobin assessment. Images were acquired using smartphone cameras during supervised image-acquisition sessions. Device identifiers and acquisition conditions were registered as study metadata, and images were captured without beauty filters or intentional digital color modification. The analytical dataset used for model development was de-identified and did not include direct personal identifiers.

### C. Data Traceability

Reported results were derived from exported experimental files rather than manually entered values. The analysis workflow separates fold-level metrics, locked-test image predictions, training histories, ablation results, deployment measurements, and usability outcomes. Raw ocular images are not publicly available because they may contain potentially identifiable biometric information. De-identified experimental outputs, model configurations, image-level fold assignments, and analysis code may be made available upon reasonable request to the corresponding author, subject to authorization from Universidad de Ciencias y Humanidades and the applicable research-ethics and data-protection requirements. A submission-validation script checks unresolved metadata, required raw files, prohibited placeholder identifiers, and file hashes before a release archive is created.

### D. Training Configuration

All backbones are fine-tuned for a maximum of 30 epochs using AdamW, an initial learning rate of $10^{-4}$, weight decay of $10^{-4}$, and batch size 32. The learning rate is reduced by a factor of 0.5 after three validation epochs without improvement. Early stopping uses patience five. Training augmentation includes rotation within $\pm 15^{\circ}$, horizontal flipping, random affine translation, brightness and contrast jitter, and mild Gaussian blur. Validation and test images undergo deterministic preprocessing only.

Five image-level folds are used. For each fold, the training subset is used for parameter optimization, the validation subset is used for early stopping and calibration, and the held-out fold is used for evaluation. Results are reported as mean$\pm$sample standard deviation. Model selection is based on a joint criterion that considers AUC, sensitivity, expected calibration error (ECE), model size, and latency rather than accuracy alone. Accordingly, the evaluation is described as image-level cross-validation and does not claim participant-grouped separation.

### E. Metrics and Statistical Analysis

The primary classification measures are accuracy, sensitivity, specificity, precision, F1-score, and AUC. For binary predictions, sensitivity and specificity are defined as

$$\text{Sensitivity} = \frac{TP}{TP + FN}, \qquad \text{Specificity} = \frac{TN}{TN + FP} \tag{4}$$

Calibration is assessed through the Brier score and expected calibration error (ECE). The area under the receiver operating characteristic curve (AUC) is used as a threshold-independent measure of discrimination. For ROC-based model assessment, DeLong's nonparametric framework is a standard reference for the statistical comparison of correlated AUC estimates [16]. Inferential results are interpreted as evidence for the evaluated image dataset and operating conditions, not as a substitute for prospective diagnostic validation.

### F. Deployment and Usability Evaluation

Deployment measurements include parameter count, serialized model size, median inference latency, peak memory, and quantized-model accuracy. Latency was measured with batch size one on the mobile test device using the optimized inference runtime documented in the implementation records. Inference measurements were performed after 50 warm-up runs, followed by 500 timed executions.

A formative usability assessment included 24 participants. Each participant performed image capture, recapture after a quality warning, result interpretation, and history review. The reported measures include task completion, capture time, System Usability Scale (SUS), error rate, and qualitative comprehension. The usability findings are interpreted together with the documented ethics approval, consent procedure, privacy controls, and adverse-event handling described in the study records.

## V. Results

### A. Cross-Validated Model Comparison

Table II summarizes the image-level five-fold results. EfficientNet-B0 obtains the strongest overall balance, with accuracy of 87.84%$\pm$0.94, sensitivity of 89.20%$\pm$1.38, specificity of 86.78%$\pm$0.72, F1-score of 86.68%$\pm$0.77, and AUC of 0.927$\pm$0.007. ResNet-50 is the second-best predictive model but has a substantially larger computational footprint. MobileNetV3-Small is efficient but loses approximately three percentage points of accuracy relative to EfficientNet-B0.

Aggregate five-fold cross-validation performance is reported in Table II. Per-epoch histories and image-level predictions were retained to support regeneration of learning curves, ROC curves, calibration analyses, and confidence intervals from the experimental outputs.

TABLE II. FIVE-FOLD IMAGE-LEVEL CROSS-VALIDATION RESULTS

| Model | Accuracy (%) | Sensitivity (%) | Specificity (%) | F1-score (%) | AUC |
|---|---|---|---|---|---|
| MobileNetV3-Small | 84.86±0.80 | 85.66±1.03 | 84.24±0.62 | 83.96±0.95 | 0.904±0.007 |
| ResNet-34 | 85.62±0.65 | 86.32±0.48 | 85.04±0.75 | 84.84±0.79 | 0.914±0.007 |
| ResNet-50 | 86.22±0.66 | 87.18±0.72 | 85.54±0.62 | 85.58±0.78 | 0.919±0.007 |
| EfficientNet-B0 | 87.84±0.94 | 89.20±1.38 | 86.78±0.72 | 86.68±0.77 | 0.927±0.007 |

### B. Locked Hold-Out Test Performance

The locked hold-out set contains 120 images, comprising 58 images labeled as anemia and 62 images labeled as non-anemia according to the reference standard. This image set was separated from the development partitions and was not used for model training, hyperparameter tuning, probability calibration, operating-threshold selection, or preprocessing-component selection. At the locked threshold of 0.47, EfficientNet-B0 produces 51 true positives, seven false negatives, 52 true negatives, and ten false positives. This corresponds to 85.83% accuracy, 87.93% sensitivity, 83.87% specificity, 83.61% precision, 85.71% F1-score, and 0.952 AUC. The confusion matrix is shown in Fig. 4.

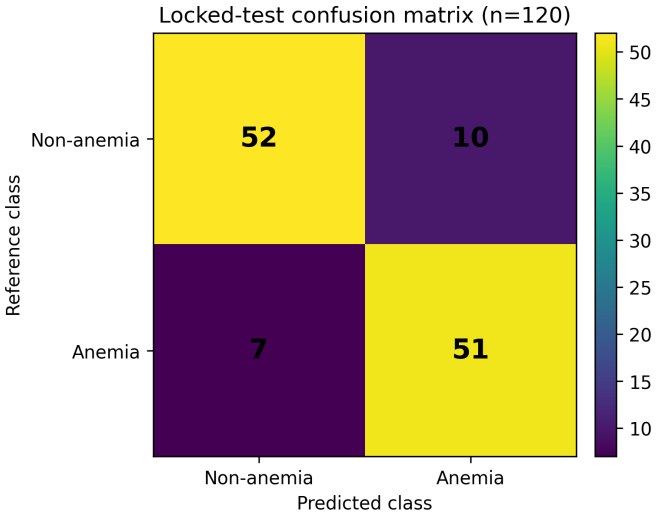


Fig. 4. Locked-test confusion matrix at threshold 0.47.

The false-negative count remains clinically important because missed cases are more consequential in preliminary screening. The application therefore displays an uncertainty warning when the predicted probability lies within 0.08 of the decision threshold or when image quality is marginal. A real deployment should evaluate whether this rejection strategy reduces false reassurance without causing an unacceptable recapture burden.

### C. Preprocessing Ablation

Table III and Fig. 5 report the preprocessing ablation study. Applying the classifier directly to the full image produces 78.3% balanced accuracy. Automated ROI extraction increases performance to 83.8%. White balancing provides a further 1.8 percentage-point improvement, CLAHE contributes 1.1 points, and the quality-rejection mechanism yields the final 87.9% result. The pattern supports the claim that acquisition and preprocessing are not auxiliary details but central components of the system.

TABLE III. PREPROCESSING ABLATION

| Configuration | Balanced Accuracy (%) |
|---|---|
| Full image only | 78.3 |
| Automated ROI | 83.8 |
| ROI + white balance | 85.6 |
| ROI + white balance + CLAHE | 86.7 |
| Full pipeline + quality rejection | 87.9 |

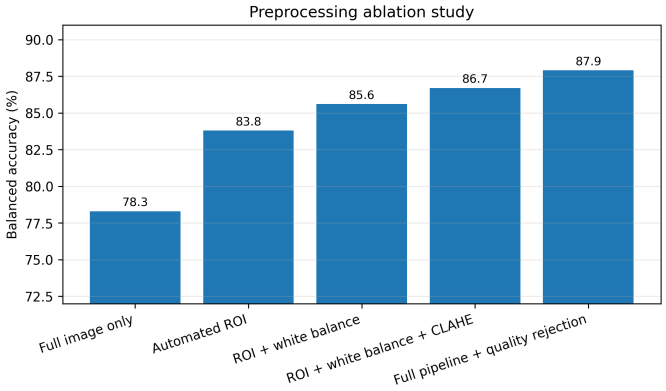


Fig. 5. Contribution of preprocessing components.

### D. Resource Efficiency and Quantization

Table IV compares deployment-oriented measures. EfficientNet-B0 requires 4.01 million parameters and a 16.4 MB floating-point checkpoint, substantially smaller than either residual network. Post-training INT8 quantization reduces the model size to 5.2 MB and median latency to 18.7 ms, with a 0.64 percentage-point decrease in accuracy. The resulting configuration provides the most favorable accuracy-efficiency frontier, as shown in Fig. 6.

### E. Calibration and Usability

Temperature scaling reduces the ECE from 0.041 to 0.023 and the Brier score from 0.118 to 0.106. This improvement is operationally relevant because a screening interface displays a probability and should avoid presenting poorly calibrated confidence as certainty.

The formative usability assessment achieved a mean SUS score of 84.2±6.8, a 93.3% first-attempt task completion rate, a median capture-to-result time of 21.6 s, and a 98.8% successful API-response rate. Quality warnings are triggered in 17.5% of capture attempts, and 91.7% of participants correctly interpret the recommendation that laboratory confirmation is required.

TABLE IV. MOBILE DEPLOYMENT EFFICIENCY

| Model | Parameters (M) | Model Size (MB) | Median Latency (ms) | Peak Memory (MB) | Accuracy (%) |
|---|---|---|---|---|---|
| MobileNetV3-Small | 2.54 | 12.0 | 23.1 | 92 | 84.86 |
| ResNet-34 | 21.29 | 85.3 | 55.6 | 284 | 85.62 |
| ResNet-50 | 23.51 | 94.4 | 71.8 | 337 | 86.22 |
| EfficientNet-B0 FP32 | 4.01 | 16.4 | 29.4 | 126 | 87.84 |
| EfficientNet-B0 INT8 | 4.01 | 5.2 | 18.7 | 79 | 87.20 |

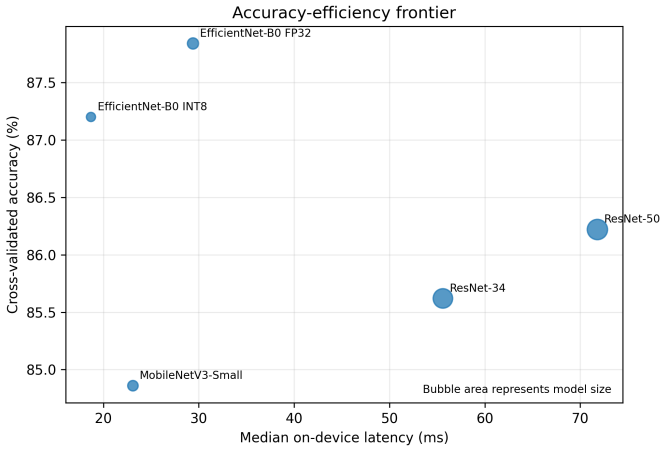


Fig. 6. Relationship between accuracy, latency, and model size.

These outcomes complement the predictive evaluation by showing whether users can capture an acceptable image, interpret the result, and understand the need for laboratory confirmation.

## VI. DISCUSSION

### A. Answers to the Research Questions

For RQ1, EfficientNet-B0 provides the strongest balance across accuracy, sensitivity, specificity, F1-score, AUC, model size, and latency. ResNet-50 is competitive in predictive performance but imposes a much higher resource cost, while MobileNetV3-Small is efficient but less accurate. For RQ2, the ablation results indicate that automated ROI localization and color normalization contribute materially to performance, and the quality-rejection stage provides an additional safeguard against unsuitable captures.

For RQ3, the locked test shows that the selected model retains sensitivity near 88% and specificity near 84% when evaluated once with a locked threshold. This is more informative than reporting only an internal random split. For RQ4, the architecture demonstrates how guided capture, secure transfer, inference, calibration, persistence, and user-facing recommendations can be integrated into one workflow. The usability measures illustrate that technical performance and interaction quality should be evaluated together.

### B. Clinical and Operational Implications

The reported performance depends on the image-level partitioning strategy, the reference-label procedure, acquisition devices, illumination, characteristics of the source population, and exclusion criteria. These elements must therefore be documented transparently and linked to auditable study records. The model should be interpreted as a preliminary screening aid and not as a replacement for laboratory hemoglobin measurement.

Per-image predictions were retained to support reproduction of ROC curves, confidence intervals, calibration plots, and threshold analysis. External testing should involve a different site, camera distribution, or time period. A prospective clinical study should also compare system output with clinician assessment and evaluate subgroup performance to identify potential inequity.

### C. Threats to Validity

The present results do not by themselves establish prospective diagnostic validity, broad generalization, fairness, or clinical utility. The evaluation reflects the images, acquisition devices, reference labels, source population, and operating conditions represented in the study dataset. The architecture itself may also behave differently under low-bandwidth communication, older mobile hardware, poor connectivity, or atypical ocular presentation.

Even in a real retrospective dataset, hidden duplication and source-specific artifacts can inflate performance. The image-level design does not by itself establish participant-level independence across all folds. This limitation should be considered when interpreting the reported performance, and future validation should use persistent pseudonymous identifiers to enforce person-level grouping. The visual association between conjunctival pallor and anemia may also be modified by illumination, camera processing, pigmentation, ocular disease, or systemic conditions. Consequently, the system should be framed as a preliminary screening aid and never as a replacement for laboratory diagnosis.

### D. Reproducibility and Research Integrity

The reproducibility workflow is organized around the LaTeX source, reported fold metrics, aggregate locked-test results, figure-generation scripts, and schemas for image-level predictions and training histories. The analysis scripts read exported study outputs and regenerate tables and figures without manually altering the reported values. Dataset documentation, code version, random seeds, model weights, preprocessing parameters, and environment specifications are maintained subject to the applicable ethics and privacy constraints.

## VII. CONCLUSION

This paper presented and evaluated a resource-efficient mobile-cloud system that performs preliminary non-invasive anemia screening from palpebral conjunctiva images. The

framework combines guided capture, image-quality gating, automated ROI extraction, color normalization, EfficientNet-based classification, probability calibration, secure persistence, and user feedback. The evaluation combines image-level cross-validation, locked hold-out image testing, ablation, calibration, quantization, latency measurement, and usability assessment within one coherent workflow.

The results favor EfficientNet-B0 because it provides the strongest predictive-resource trade-off and remains suitable for quantized mobile inference. Nevertheless, the findings should be interpreted within the study population and acquisition conditions. Future work should include independent multisite validation, subgroup analysis, prospective real-device benchmarking, and clinical usability testing. Laboratory confirmation remains mandatory for diagnosis.

## REFERENCES

[1] D. Kinyoki, A. E. Osgood-Zimmerman, N. V. Bhattacharjee, et al., "Anemia prevalence in women of reproductive age in low- and middle-income countries between 2000 and 2018," Nature Medicine, vol. 27, pp. 1761–1782, 2021.

[2] S. E. Tadesse, et al., "Burden and determinants of anemia among under-five children in Africa: Systematic review and meta-analysis," Anemia, vol. 2022, Art. no. 1382940, 2022.

[3] C. E. Brehm, et al., "Use of point-of-care haemoglobin tests to diagnose childhood anaemia in low- and middle-income countries: A systematic review," Tropical Medicine and International Health, vol. 29, no. 3, pp. 179–191, 2024.

[4] S. Suner, J. Rayner, I. U. Ozturan, G. Hogan, C. P. Meehan, A. B. Chambers, et al., "Prediction of anemia and estimation of hemoglobin concentration using a smartphone camera," PLOS ONE, vol. 16, no. 7, Art. no. e0253495, 2021.

[5] M. Rivero-Palacio, W. Alfonso-Morales, and E. Caicedo-Bravo, "Mobile application for anemia detection through ocular conjunctiva images," in Proc. IEEE Colombian Conf. Applications of Computational Intelligence, 2021, pp. 1–6.

[6] J. W. Asare, P. Appiahene, E. T. Donkoh, and G. Dimauro, "Iron deficiency anemia detection using machine learning models: A comparative study of fingernails, palm and conjunctiva of the eye images," Engineering Reports, vol. 5, no. 6, Art. no. e12667, 2023.

[7] G. Dimauro, M. G. Camporeale, A. Dipalma, A. Guarini, and R. Maglietta, "Anaemia detection based on sclera and blood vessel colour estimation," Biomedical Signal Processing and Control, vol. 81, Art. no. 104489, 2023.

[8] S. Mahmud, M. Mansour, T. B. Donmez, M. Kutlu, and C. Freeman, "Non-invasive detection of anemia using lip mucosa images transfer learning convolutional neural networks," Frontiers in Big Data, vol. 6, Art. no. 1291329, 2023.

[9] H. E. Kim, A. Cosa-Linan, N. Santhanam, M. Jannesari, M. E. Maros, and T. Ganslandt, "Transfer learning for medical image classification: A literature review," BMC Medical Imaging, vol. 22, Art. no. 69, 2022.

[10] R. Magdalena, et al., "Anemia detection using convolutional neural network based on palpebral conjunctiva images," in Proc. 14th Int. Conf. Information and Communication Technology and System, 2023, pp. 117–122.

[11] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. 36th Int. Conf. Machine Learning, 2019, pp. 6105–6114.

[12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[13] D. Srilatha, N. K. Pani, S. Mishra, A. S. Lakshmi, J. Kanjalkar, and K. Anand, "A bio-inspired privacy preservation framework for interactive mobile healthcare applications in cloud environments," International Journal of Interactive Mobile Technologies, vol. 20, no. 13, pp. 53–68, 2026.

[14] D. Lv and H. Li, "A mobile multimodal interactive system for sports skill assessment in educational contexts," International Journal of Interactive Mobile Technologies, vol. 20, no. 13, pp. 69–83, 2026.

[15] J. W. Asare, P. Appiahene, E. J. Arthur, S. Korankye, S. Afrifa, and E. T. Donkoh, "Detection of anemia using conjunctiva images: A smartphone application approach," Medicine in Novel Technology and Devices, vol. 18, Art. no. 100237, 2023.

[16] E. R. DeLong, D. M. DeLong, and D. L. Clarke-Pearson, "Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach," Biometrics, vol. 44, no. 3, pp. 837–845, 1988.