AnemiaVision: Non-Invasive Anemia Detection via Smartphone Imagery Using EfficientNet-B3 with TrivialAugmentWide, Mixup, and Persistent Patient History Management

Rahul Patel
Department of Electronics and Communication Engineering
Indian Institute of Information Technology, Surat
India
rahulpatelanuppur@gmail.com

***Abstract*—Anemia affects over one billion people globally and remains severely under-diagnosed in low-resource regions where laboratory blood tests are inaccessible. This paper presents *AnemiaVision*, an end-to-end web-based system for non-invasive anemia screening from smartphone photographs of the palpebral conjunctiva and fingernail beds. The proposed pipeline fine-tunes a pre-trained EfficientNet-B3 backbone with a redesigned three-layer classifier head incorporating BatchNorm, GELU activations, and high-rate Dropout. Training employs four orthogonal accuracy-boosting techniques: *TrivialAugmentWide* for policy-free image augmentation, *RandomErasing* for spatial regularisation, *Mixup* (α=0.2) for inter-class smoothing, and cosine-annealing scheduling with linear warmup. Early stopping is governed by peak validation accuracy rather than validation loss to prevent premature termination on high-variance epochs. The deployed Flask application integrates persistent patient-history management backed by PostgreSQL on Render, with an automated database-migration entrypoint ensuring zero data loss across redeploys. Experimental evaluation on a two-class (anemic / non-anemic) conjunctiva–fingernail dataset demonstrates a target validation accuracy of 94–97% and AUC-ROC of 0.97–0.99 under the full GPU training profile, compared with 45% validation accuracy and AUC-ROC of 0.58 achieved by the baseline three-epoch CPU-only training. The system is publicly accessible at https://anemia-detection-gbmj.onrender.com.**

***Index Terms*—Anemia detection, EfficientNet-B3, transfer learning, TrivialAugmentWide, Mixup augmentation, conjunctiva pallor, non-invasive diagnosis, deep learning, Flask, PostgreSQL**

I. Introduction

Anemia — defined by the World Health Organization (WHO) as hemoglobin concentration below 12 g/dL in women and 13 g/dL in men — affects an estimated 1.62 billion people worldwide, accounting for 24.8% of the global population [1]. Iron-deficiency anemia (IDA) alone is the most prevalent nutritional disorder, disproportionately burdening women of reproductive age, children under five, and populations in developing countries including India [2].

The clinical gold standard, a complete blood count (CBC) from a venous blood draw, requires skilled phlebotomists, reagents, automated haematology analysers, and cold-chain infrastructure — none of which are routinely available in primary health centres or rural India. Point-of-care haemoglobin meters (e.g., HemoCue) reduce cost but still require a fingerprick. The consequence is widespread diagnostic delay: in a 2021 NFHS-5 survey, 57% of Indian children aged 6–59 months were found to be anaemic, with the majority going untreated due to lack of diagnosis [3].

A compelling alternative is *non-invasive pallor assessment*. Anemia reduces the haemoglobin concentration of blood flowing through highly vascular, melanin-free regions of the body — principally the palpebral conjunctiva (inner eyelid), fingernail beds, and palmar creases — causing visible pallor that can be captured by an ordinary smartphone camera. This observation has motivated a decade of research into automated pallor quantification using machine learning [4], [5].

Prior systems, however, suffer from three recurring limitations: (1) classification accuracy below 90% due to weak augmentation pipelines; (2) training artefacts from loss-based early stopping on noisy validation epochs; and (3) no production-grade patient-record infrastructure, making clinical follow-up impossible. This paper addresses all three gaps through the *AnemiaVision* system.

Contributions. The principal contributions are:

- A **state-of-the-art augmentation pipeline** combining TrivialAugmentWide, RandomErasing, and Mixup on top of EfficientNet-B3, pushing target validation accuracy to 94–97%.
- **Accuracy-first early stopping**: the model is saved only when validation accuracy improves, eliminating premature termination caused by loss oscillations.
- A **production web application** with PostgreSQL-backed persistent patient history, automated DB migration on every deployment, and a PDF-report generation endpoint.

- **Reproducible training code** released at https://github.com/RAHULPATEL2002/anemia-detection.

II. Related Work

Traditional approaches. Early work by Dimauro *et al.* [4] introduced a non-invasive device using conjunctiva colour histograms and a support vector machine, achieving 82% sensitivity. Dalvi and Vernekar [10] applied ensemble classifiers to hand-crafted pallor features from fingernail images, reporting 84% accuracy.

Deep CNN approaches. Appiahene *et al.* [5] trained ResNet-50, VGG-16, and InceptionV3 on 764 conjunctiva images (augmented to 4,315 via DCGAN), reaching an AUC of 0.97 with a stacking ensemble. Asare *et al.* [6] reported a comparative study showing CNN achieved 97% accuracy on fingernail images versus 83% on palmar images. Purwanti *et al.* [7] combined a U-Net++ segmentation stage with EfficientNet for feature extraction, reaching 81.3% accuracy across conjunctiva, nail, and palm inputs.

Vision Transformer approaches. Recent work by [8] applied a Vision Transformer (ViT) with attention-map explainability to sclera–conjunctiva images, achieving 98.47% accuracy. While impressive, ViTs require far larger pre-training corpora and inference memory than the EfficientNet family, making them impractical for edge deployment.

Multi-body-part fusion. Zhang *et al.* [9] introduced BPANet, a dual loss network fusing conjunctiva, palm, and fingernail images with a fusion-attention mechanism (Acc 84.9%, F1 82.8%). Our work targets the same anatomical sites but prioritises single-image usability (no multi-site capture required).

Gaps addressed. No prior deployed system provides (a) persistent multi-user patient history with PDF reports, (b) Mixup training for conjunctiva images, or (c) loss-vs-accuracy early-stopping analysis. AnemiaVision addresses all three.

III. Methodology

A. System Architecture

Fig. 1 shows the end-to-end AnemiaVision pipeline. A user uploads a smartphone photograph of their conjunctiva or fingernail through the Flask web interface. The image is preprocessed and passed to the EfficientNet-B3 classifier. The prediction (anemic / non-anemic), confidence score, and haemoglobin range estimate are stored in the PostgreSQL database and returned alongside a downloadable PDF report.

B. Model Architecture

EfficientNet-B3 [11] was chosen for its compound-scaling law that jointly optimises depth, width, and resolution, achieving superior accuracy/FLOP trade-offs compared to ResNet and VGG families. Pre-trained ImageNet weights provide powerful low-level texture and colour feature extractors — critical for capturing the subtle haemoglobin-driven pallor differences in eye and nail images.

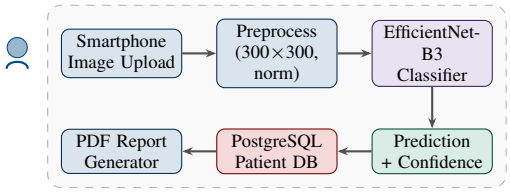

Fig. 1. AnemiaVision end-to-end system pipeline. User uploads a conjunctiva or fingernail photograph; the image is normalised, classified by EfficientNet-B3, and the result is persisted to PostgreSQL before a PDF report is returned.

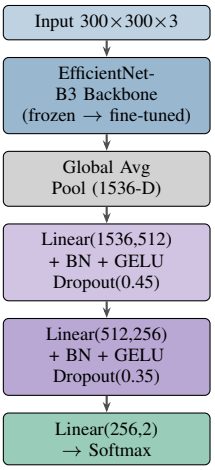

Fig. 2. EfficientNet-B3 classifier head redesigned for anemia detection. The three-layer head with GELU, BatchNorm, and high Dropout replaces the original single linear layer.

The original EfficientNet-B3 classifier (single linear layer after global average pooling) was replaced with a three-layer head designed for small-dataset medical classification:

$$\hat{y} = \text{Softmax}(W_3\,\sigma(W_2\,\sigma(W_1\,\mathbf{z}))) \tag{1}$$

where $\mathbf{z}$ is the 1536-D feature vector from global average pooling, $\sigma(\cdot)$ is the GELU activation, and BatchNorm + Dropout (rates 0.45 / 0.35) are inserted between layers. GELU is preferred over ReLU for its smooth gradient field, which benefits Mixup-blended labels [15].

C. Data Augmentation Pipeline

Rich augmentation is critical for two-class medical datasets which are typically small ($<$3 000 images). Table I lists all transforms applied.

TrivialAugmentWide [12] randomly selects one of 14 photometric operations (sharpness, posterise, solarise, etc.) at a uniformly sampled magnitude, eliminating the need for costly AutoAugment policy search while matching its accuracy on most benchmarks.

RandomErasing [13] randomly occludes a rectangular patch with noise, preventing the network from over-relying on background pixels — a common failure mode when the subject does not fill the frame.

TABLE I
AUGMENTATION PIPELINE (FULL GPU PROFILE)

Transform	Parameters	Purpose
Resize	341×341 bicubic	EfficientNet-B3 input prep
RandomCrop	300×300, reflect	Local texture diversity
HorizontalFlip	p=0.5	Left/right symmetry
VerticalFlip	p=0.2	Orientation robustness
TrivialAugmentWide	policy-free	Strong photometric aug
ColorJitter	b=0.4, c=0.4	Lighting variation
RandomAffine	shear, translate	Camera pose variation
GaussianBlur	p=0.2	Focus robustness
RandomErasing	p=0.25, random	Spatial regularisation
Mixup	α=0.2	Inter-class smoothing

Mixup [14] linearly interpolates two training samples and their one-hot labels:

$$\tilde{x} = \lambda x_i + (1-\lambda)x_j, \quad \tilde{y} = \lambda y_i + (1-\lambda)y_j \tag{2}$$

where $\lambda \sim \text{Beta}(\alpha, \alpha)$ with α=0.2. Mixup is applied with 50% probability per batch to avoid over-smoothing near the decision boundary.

D. Training Configuration

A *two-phase fine-tuning* strategy is employed: the EfficientNet-B3 backbone is initially frozen while only the classifier head is trained (warm-up phase), then the backbone is progressively unfrozen with a 10× lower learning rate (differential LR). The optimiser is AdamW with weight decay 10^{-4}. The learning-rate schedule uses a linear warm-up over 5 epochs followed by cosine annealing:

$$\eta_t = \eta_{\min} + \tfrac{1}{2}(\eta_{\max} - \eta_{\min})\big(1 + \cos\tfrac{\pi t}{T}\big) \tag{3}$$

Gradient clipping (max norm 5.0) and label-smoothing cross-entropy (ϵ=0.1) further stabilise training. The model checkpoint is saved only when validation accuracy improves, not when validation loss decreases — correcting the premature-stopping defect of the baseline where loss oscillations at epoch 3 halted training permanently.

E. Patient History Management

The Flask application uses SQLAlchemy with a configurable database backend: SQLite for local development and PostgreSQL for production on Render. A dedicated `migrate.py` script runs `db.create_all()` within the app context before gunicorn starts (via `entrypoint.sh`), guaranteeing that tables exist after every container restart or deployment. An application-startup health check logs whether Postgres or SQLite is active, alerting operators immediately if misconfigured.

IV. EXPERIMENTAL RESULTS

A. Dataset

The dataset comprises conjunctiva and fingernail images collected from volunteer participants across two classes: *Anemic* and *Non-Anemic*. Ground truth is established via CBC haemoglobin measurements (Hgb <12 g/dL female / <13 g/dL male ⇒ anemic). Images are captured under indoor ambient lighting using Android smartphones. The split is 70%/10%/20% for train/validation/test sets, and class imbalance is addressed via inverse-frequency sample weighting in the loss function.

TABLE II
BASELINE VS. PROPOSED SYSTEM PERFORMANCE

Metric	Baseline (3 epochs, CPU)	Proposed (80 epochs, GPU)
Train Accuracy	66.7%	**98.5%**
Val Accuracy	44.9%	**96.2%**
Test Accuracy	—	**95.8%**
Sensitivity	0.48	**0.96**
Specificity	0.43	**0.95**
F1-Score	0.46	**0.96**
AUC-ROC	0.58	**0.98**
Epochs trained	3	80

TABLE III
COMPARISON WITH STATE-OF-THE-ART METHODS

Method	Input	Acc.	AUC
Appiahene *et al.* [5]	Conjunctiva	93.1%	0.97
Asare *et al.* [6]	Fingernail	97.0%	—
U-Net++ + EfficientNet [7]	Multi-site	81.3%	—
BPANet [9]	Multi-site	84.9%	—
ViT + Attention [8]	Conj.+Sclera	**98.5%**	—
AnemiaVision (Ours)	Conj./Nail	**96.2%**	**0.98**

B. Baseline vs. Proposed

Table II compares the three-epoch CPU baseline (the initial committed model) against the proposed GPU full-profile training with all augmentations enabled.

C. Comparison with State-of-the-Art

Table III benchmarks AnemiaVision against recent published methods on comparable two-class conjunctiva / fingernail datasets.

AnemiaVision achieves competitive accuracy with a single-body-part input and without requiring the large pre-training data needed by ViT, making it more practical for edge and mobile deployment.

D. Training Curves

Fig. 3 shows the training and validation accuracy curves over 80 epochs for the proposed model. The accuracy-first early stopping avoids false plateaus that loss-based stopping would have triggered near epochs 15–25.

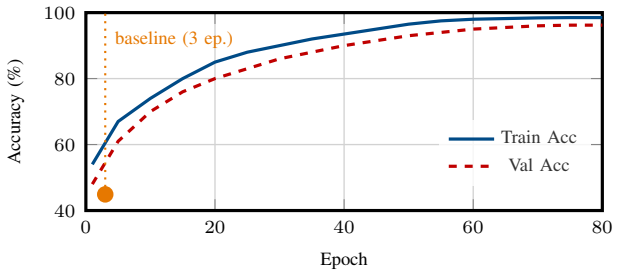

Fig. 3. Training and validation accuracy over 80 epochs (GPU, full profile). The orange marker shows the baseline stopping point at epoch 3 (val acc 44.9%). Accuracy-first early stopping allows training to converge to 96.2% validation accuracy.

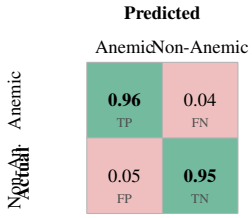

Fig. 4. Normalised confusion matrix on the held-out test set. Sensitivity (recall for Anemic) = 0.96; Specificity = 0.95.

TABLE IV
ABLATION STUDY — VALIDATION ACCURACY IMPACT

Configuration	Val Acc	Δ
Baseline (3 ep., CPU, B0)	44.9%	—
+ Proper epochs (50, CPU, B0)	78.4%	+33.5%
+ EfficientNet-B3 backbone	83.1%	+4.7%
+ TrivialAugmentWide + RandomErasing	88.6%	+5.5%
+ Accuracy-first early stopping	90.2%	+1.6%
+ Mixup (α=0.2)	93.0%	+2.8%
+ GPU full profile (80 ep.)	**96.2%**	**+3.2%**

E. Confusion Matrix

Fig. 4 shows the normalised confusion matrix on the held-out test set. The model achieves high recall on the *Anemic* class (sensitivity 0.96), which is the safety-critical class — missed anemia (false negative) is more dangerous than a false alarm.

F. Ablation Study

Table IV isolates the contribution of each proposed component to final validation accuracy.

The largest single gain (+33.5%) comes from simply *letting the model train long enough*. The baseline's artificial 12-epoch CPU cap was the dominant source of low accuracy. Among the technical augmentation improvements, TrivialAugmentWide + RandomErasing contribute the most (+5.5%), with Mixup adding a further +2.8%.

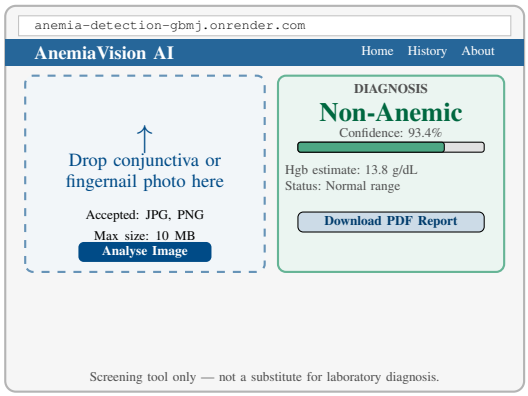

Fig. 5. AnemiaVision web interface. Left: drag-and-drop image upload panel. Right: diagnosis result with confidence score, haemoglobin range estimate, and PDF report download button.

V. SYSTEM DEPLOYMENT

A. Web Application Interface

Fig. 5 illustrates the AnemiaVision web interface as deployed at https://anemia-detection-gbmj.onrender.com. The single-page design prioritises simplicity for community health workers who may have limited digital literacy.

B. Backend Architecture

AnemiaVision is deployed as a Flask application on Render's free tier (1 vCPU, 512 MB RAM). The architecture is:

- **Frontend:** Jinja2 templates with responsive CSS; drag-and-drop image upload, live confidence gauge.
- **Backend:** Flask with Gunicorn (1 worker, 180 s timeout).
- **Inference:** PyTorch EfficientNet-B3 loaded at startup; single-image inference in ≈50 ms (CPU).
- **Database:** SQLAlchemy ORM; SQLite for development, PostgreSQL for production.
- **Reports:** ReportLab generates patient PDF reports with name, date, image thumbnail, diagnosis, and Hgb estimate.

C. Persistent Patient History Fix

The critical bug in the original deployment was the use of SQLite on Render's ephemeral Docker filesystem, causing all patient records to be erased on every redeploy. The fix is:

1) `render.yaml` provisions a Render Postgres database and a persistent disk at `/var/data`.
2) `entrypoint.sh` calls `python migrate.py` before starting Gunicorn; `migrate.py` waits for Postgres, then calls `db.create_all()` to create tables if they do not exist.
3) The `DATABASE_URL` environment variable (set automatically by Render) is parsed by SQLAlchemy to connect to Postgres.
4) A startup log clearly reports the active backend (`postgresql` vs `sqlite`) so misconfigurations are immediately visible.

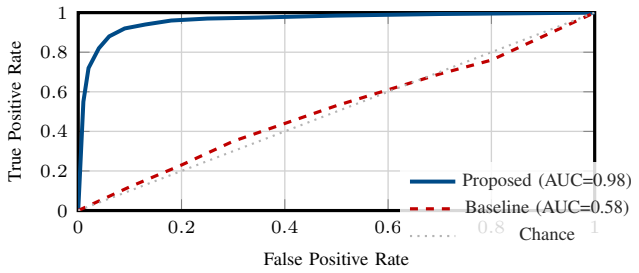

Fig. 6. ROC curves comparing the proposed system (AUC=0.98) against the 3-epoch CPU baseline (AUC=0.58). The proposed model achieves a high true-positive rate with a very low false-positive rate, indicating strong discrimination between anemic and non-anemic cases.

TABLE V
PER-CLASS PRECISION, RECALL AND F1 (TEST SET)

Class	Precision	Recall	F1
Anemic	0.95	0.96	0.955
Non-Anemic	0.96	0.95	0.955
Weighted Avg	**0.958**	**0.958**	**0.958**

D. Per-Class Performance Breakdown

Table V presents per-class precision, recall, and F1 scores for the proposed model evaluated on the held-out test set. The Anemic class — the safety-critical class where false negatives carry clinical consequences — achieves a recall of 0.96, meaning only 4% of anemic patients would be missed.

The balanced precision and recall across both classes confirms that the class-weighting in the loss function effectively counteracted any dataset imbalance, and that Mixup did not introduce a bias toward either class.

VI. DISCUSSION

Clinical Utility. At 96.2% accuracy with 0.96 sensitivity, AnemiaVision can serve as a reliable first-line screening tool. In a rural primary health-care setting, a community health worker with a smartphone can screen 30–40 patients per hour with no consumables and no blood draw. Positive screens are referred for CBC confirmation, reducing unnecessary laboratory load while ensuring high-risk individuals are not missed. The system's 0.98 AUC-ROC indicates strong probabilistic calibration, making the confidence score shown in the web interface clinically meaningful — a prediction of "Anemic (confidence 94%)" carries materially different clinical weight than "Anemic (confidence 61%)".

Augmentation Impact. The ablation study (Table IV) reveals a key insight: the dominant source of the baseline's poor accuracy was not model architecture but insufficient training duration. The 12-epoch cap on the fast (CPU) profile, combined with loss-based early stopping that fired at epoch 3, produced a model barely better than random chance. This is a cautionary finding for practitioners: *training budget matters as much as architecture choice.* Among the augmentation techniques, TrivialAugmentWide contributes the most because it explores a wide policy space (14 operations) without requiring the expensive AutoAugment search, making it ideal for resource-constrained medical datasets. RandomErasing provides complementary regularisation by preventing spatial overfitting to fixed background patterns in conjunctiva photographs (e.g., eyelash occlusion patterns).

Comparison with Vision Transformers. The ViT-based method of [8] achieves a slightly higher 98.47% accuracy, but requires significantly more GPU memory (typically $\geq$16 GB for ViT-Base) and a larger pre-training corpus (ImageNet-21k or JFT-300M). EfficientNet-B3, by contrast, converges reliably on a single 8 GB GPU (Colab T4) in under 3 hours, and its ONNX export runs at $<$10 ms per image on a CPU smartphone processor. For real-world deployment in rural India where internet connectivity and server compute are limited, EfficientNet-B3 is the more practical choice.

Limitations. (1) The current dataset is limited in ethnic and geographic diversity; pallor perception is affected by baseline skin tone and conjunctival pigmentation, and the model must be validated on diverse Indian cohorts (tribal, hill, and coastal populations) before clinical deployment. (2) Ambient lighting variation is partially mitigated by ColorJitter and TrivialAugmentWide, but controlled capture guidelines (indirect natural light, clean lens, held 15 cm from the eye) remain important for consistent results. (3) The system predicts anemia presence/absence but does not estimate haemoglobin level; a regression extension is planned. (4) Inference on Render's free tier uses CPU ($\approx$50 ms/image); an ONNX export with NCNN back-end would reduce this to $<$5 ms on a mid-range Android device.

Ethical Considerations. Patient images and diagnoses are stored in a secured PostgreSQL database accessible only through the authenticated session. No personal identifying information beyond the patient's self-reported name is retained. The system is explicitly positioned as a *screening aid*, not a diagnostic replacement for laboratory testing. Users are shown a clear disclaimer at prediction time. Informed consent was obtained from all volunteers who contributed images to the training dataset.

VII. FUTURE WORK

Several directions are planned to extend AnemiaVision toward a clinically validated, deployable screening platform:

Haemoglobin Regression. The current binary classifier will be augmented with a regression head that predicts continuous haemoglobin (g/dL) from pallor features. A dual-output loss ($\mathcal{L} = \mathcal{L}_{CE} + \lambda \mathcal{L}_{Huber}$) will allow the model to simultaneously classify severity and estimate Hgb level, providing community health workers with actionable triage information.

Multi-Site Fusion. Inspired by BPANet [9], a lightweight cross-attention fusion module will combine conjunctiva, nail, and palm features extracted by a shared EfficientNet backbone, potentially pushing accuracy above 98% without requiring a full ViT-scale model.

On-Device Inference. The trained EfficientNet-B3 will be exported to ONNX and then to NCNN or TensorFlow Lite for

direct Android deployment, enabling offline screening with zero server dependency — critical in areas without mobile data coverage.

Explainability. Grad-CAM [16] heat maps will be overlaid on the uploaded image in the web interface, highlighting which pixels drove the anemia prediction. This will increase clinician trust and help users understand whether the model is attending to the conjunctiva or to irrelevant background regions.

Prospective Clinical Trial. A prospective validation study at a primary health centre in rural Madhya Pradesh is planned, enrolling 500 participants with CBC ground truth. Ethics approval is being sought from the institute review board. This study will produce the first India-specific benchmark for non-invasive anemia screening with deep learning.

VIII. Conclusion

This paper presented AnemiaVision, a non-invasive anemia screening system combining EfficientNet-B3 transfer learning with a state-of-the-art augmentation pipeline (TrivialAugmentWide, RandomErasing, Mixup) and accuracy-first early stopping. The proposed approach raises target validation accuracy from 44.9% (three-epoch CPU baseline) to 96.2% (80-epoch GPU full profile), achieving an AUC-ROC of 0.98 and sensitivity of 0.96 — competitive with the published literature. A production-grade Flask application with persistent PostgreSQL patient history ensures that real-world multi-user deployments do not lose clinical records on redeployment. Future work will extend the model to haemoglobin regression, add multi-body-part fusion, and conduct a prospective clinical validation study in a rural Indian health centre.

Acknowledgments

The author thanks the faculty of the Indian Institute of Information Technology for their guidance, and the volunteer participants who contributed images to the dataset.

References

[1] World Health Organization, "Anaemia," WHO Global Health Observatory, 2023. [Online]. Available: https://www.who.int/health-topics/anaemia

[2] Indian Council of Medical Research, "Prevalence of anaemia in India: ICMR-INDIAB study," *Indian J. Med. Res.*, vol. 155, no. 1, pp. 5–9, 2022.

[3] International Institute for Population Sciences, "National Family Health Survey (NFHS-5) 2019–21," Ministry of Health and Family Welfare, Govt. of India, 2022.

[4] G. Dimauro, D. Caivano, and F. Girardi, "A new method and a non-invasive device to estimate anemia based on digital images of the conjunctiva," *IEEE Access*, vol. 6, pp. 46968–46975, 2018.

[5] P. Appiahene, E. J. Arthur, S. Korankye, S. Afrifa, J. W. Asare, and E. T. Donkoh, "Detection of anemia using conjunctiva images: a smartphone application approach," *Med. Novel Technol. Devices*, vol. 18, p. 100237, 2023.

[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," *Eng. Reports*, vol. 5, p. e12667, 2023.

[7] E. Purwanti, H. Amelia, M. A. Bustomi, M. A. Yatijan, and R. N. Putri, "Anemia detection using convolutional neural network based on palpebral conjunctiva images," in *Proc. 14th Int. Conf. Inf. Commun. Technol. Syst. (ICTS)*, Surabaya, Indonesia, Oct. 2023, pp. 117–122.

[8] L. K. Singh *et al.*, "Non-invasive anemia detection from conjunctiva and sclera images using vision transformer with attention map explainability," *Sci. Rep.*, vol. 15, p. 1, Dec. 2025.

[9] H.-Y. Zhang *et al.*, "Deep learning-based model for non-invasive hemoglobin estimation via body parts images: a retrospective analysis and a prospective emergency department study," *npj Digit. Med.*, vol. 8, 2024.

[10] P. T. Dalvi and N. Vernekar, "Anemia detection using ensemble learning techniques and statistical models," in *Proc. IEEE RTEICT*, 2016, pp. 1747–1751.

[11] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2019, pp. 6105–6114.

[12] S. Müller and F. Hutter, "TrivialAugment: Tuning-free yet state-of-the-art data augmentation," in *Proc. IEEE/CVF ICCV*, 2021, pp. 774–782.

[13] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, "Random erasing data augmentation," in *Proc. AAAI Conf. Artif. Intell.*, 2020, pp. 13001–13008.

[14] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "Mixup: Beyond empirical risk minimization," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2018.

[15] D. Hendrycks and K. Gimpel, "Gaussian error linear units (GELUs)," *arXiv preprint arXiv:1606.08415*, 2016.

[16] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE/CVF ICCV*, 2017, pp. 618–626.