A Reproducible Deep Learning Pipeline for Augmentation-Enhanced Classification of Anemia-Related Blood-Cell Images

R. Remya
Providence College of Engineering, Kerala, India
remya.r@providence.edu.in

M. Janani
St. Joseph's College of Engineering, Tamil Nadu, India
jananim@stjosephs.ac.in

P. Thilakavathy
Vels Institute of Science Technology and Advanced Studies, Tamil Nadu, India
thilakavathy.se@vistas.ac.in

Mitha Rachel Jose
Laurea University of Applied Sciences, Espoo, Finland
mitha.jose@laurea.fi

R. Reji
Thangal Kunju Musaliar Institute of Technology, Kerala, India
rejir@tkmit.ac.in

L. P. Supriya
Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Tamil Nadu, India
supriyabinnyb@gmail.com

Mohammad Asim
Sharda University, Uttar Pradesh, India
er.mohdasim@gmail.com

Lekshmi R. Nair
Cochin University of Science and Technology, Kerala, India
lekshmi.r.nair@cusat.ac.in

Dheeraj Tiger
The NorthCap University, Gurugram, Haryana, India
dheerajtiger@gmail.com

Divya Saleela
University of Southampton, United Kingdom
d.saleela@soton.ac.uk (corresponding author)

Received: 24 April 2026 | Revised: 13 May 2026, 24 May 2026, 25 May 2026, 3 June 2026, and 4 June 2026 | Accepted: 5 June 2026

 DOI: https://doi.org/10.48084/etasr.19533

ABSTRACT

This study presents a reproducible end-to-end deep learning pipeline for proof-of-concept classification of anemia-related and control-like blood cell images using a publicly accessible online dataset. The workflow integrates online data download, automatic extraction, dataset restructuring, class balancing, augmentation-enhanced preprocessing, transfer-learning-based model training, visualization, and structured metric generation within a single executable framework. A balanced binary dataset was constructed to reduce the effect of severe class imbalance, and augmentation strategies, including flipping, rotation, color jitter, and affine transformation, were applied only during training. A ResNet18-based classifier showed strong convergence under a fixed random seed and single train-validation-test split. On a small held-out test subset of 32 images, all images were correctly classified. However, this result should be interpreted cautiously because the test subset was small, no cross-validation or external validation was performed, and performance on limited balanced data may overestimate generalizability. Class-distribution plots, representative sample panels, learning curves, confusion matrices, prediction files, and metric summaries support transparent inspection and reproducibility. The main contribution of this study is a reproducible engineering workflow for blood-cell image classification, rather than a claim of algorithmic novelty or clinical readiness.

Keywords-blood cell image classification; deep learning; reproducible pipeline; data augmentation; class balancing; ResNet18; hematology imaging; medical image analysis

I. INTRODUCTION

Anemia and related blood disorders continue to represent a major global health burden. Recent estimates indicate that anemia affects a substantial proportion of the world population, particularly children, women of reproductive age, and pregnant women, while inherited red blood cell disorders such as sickle cell disease also remain highly prevalent in many regions [1]. These trends highlight the need for scalable and reliable computational tools that can support blood-cell image analysis and enable earlier, more consistent screening workflows.

Microscopic blood-smear examination remains one of the most widely used approaches for assessing red blood cell morphology and identifying abnormal cellular patterns [2, 3]. Despite its diagnostic importance, manual interpretation is time-consuming, labor-intensive, and subject to inter-observer variability. These limitations become more pronounced when large volumes of images must be screened or when specialist expertise is limited. As a result, automated image-based methods have become increasingly attractive for assisting hematological analysis.

Deep learning is effective in medical image analysis because it can learn discriminative features directly from raw images [4]. For blood-cell morphology assessment, Convolutional Neural Networks (CNNs) can capture subtle differences in shape, contour, texture, and structure, making them suitable for anemia-related red blood cell classification. Prior studies have used both conventional machine learning and deep learning methods for automated blood-cell classification [5, 6]. However, earlier approaches often depended on hand-crafted features, dataset-specific preprocessing, and limited validation designs.

With the increasing adoption of deep learning, CNNs have become a dominant approach for blood-cell image classification [7, 8]. Recent anemia-screening work has also explored adaptive filtering with deep CNNs to improve red blood cell image analysis and support automated anemia detection [9]. Architectures such as ResNet, MobileNet, AlexNet, VGG, EfficientNet, and transformer-based variants have been applied to erythrocyte morphology recognition and related medical image classification tasks. These models can learn discriminative spatial and texture features directly from raw images, although limited dataset size and restricted sample diversity remain common constraints.

Another important direction in the literature is the use of hybrid and ensemble classification systems, feature selection, and deep feature fusion methods [10-13]. Some studies combine CNN-based feature extraction with traditional classifiers [14], while others explore ensemble strategies to improve robustness and predictive accuracy. Lightweight models are also important for computationally constrained environments. However, many studies remain focused on final model accuracy rather than on the transparency and reproducibility of the full experimental pipeline.

More recent research has moved beyond simple binary classification toward multi-class settings, anomaly-aware prediction, explainable AI-based interpretation, and healthcare-monitoring frameworks [15-17]. Techniques such as Grad-CAM, LIME, and feature visualization can highlight image regions influencing model predictions and may improve interpretability. Nevertheless, several recurring limitations remain across the literature: inaccessible or limited datasets, severe class imbalance, insufficient evaluation uncertainty, lack of external validation, and insufficient reporting of preprocessing and split reproducibility.

This work aimed to address these issues by developing a reproducible end-to-end deep learning pipeline for the classification of anemia-related and control-like blood-cell images using publicly accessible online data. The main contribution of this paper is system-level integration and reproducibility. First, the workflow integrates data download, preprocessing, augmentation, training, evaluation, visualization, and export in one implementation. Second, it introduces a balanced experimental setting to reduce the direct effect of class imbalance, while acknowledging that this balanced setting is not equivalent to real-world clinical prevalence. Third, it reports a broad set of evaluation metrics and visual outputs. Fourth, it explicitly discusses limitations

related to small test size, lack of cross-validation, possible sampling bias, and the need for external validation.

II. METHODOLOGY

A. Proposed Framework

The proposed framework was developed as a unified pipeline for automated classification of blood-cell images into two categories: anemia-like and control-like. Rather than treating model training as an isolated step, the framework integrates all major stages of the experimental process, including data download, restructuring, balancing, preprocessing, training, validation, testing, and result export. This design was chosen to ensure that the workflow remains consistent, traceable, and easy to reproduce. A key objective was to improve methodological transparency. For this reason, the pipeline was designed to generate not only numerical outputs but also visual outputs at different stages of the experiment. These include representative images from the training, validation, and test subsets, class-distribution plots, loss and accuracy curves, confusion matrices, and metric bar charts. By combining classification with visual inspection and structured reporting, the framework supports a more interpretable and credibility-oriented evaluation process.

The overall workflow, shown in Figure 1, begins with online dataset download and automatic extraction. The acquired images are then reorganized into task-specific classes and balanced to reduce class skew. The processed dataset is partitioned into training, validation, and test subsets, followed by deterministic preprocessing for all subsets and random augmentation only for the training subset. A transfer-learning model is then trained, validated, and evaluated, and the final outputs are exported as image files and structured metric reports. In this way, the proposed framework functions as a complete computational pipeline rather than a model-only solution.

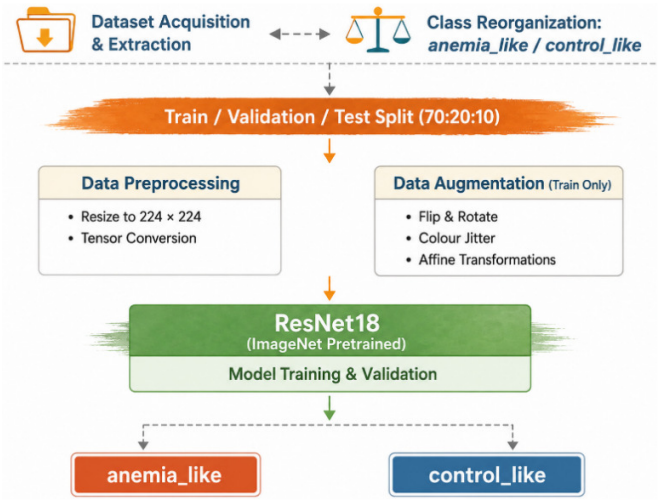

Fig. 1. Overview of proposed blood-classification pipeline.

B. Dataset Acquisition and Construction

The dataset used in this study was obtained from a publicly accessible online source [18] and downloaded directly within the computational workflow. After the download stage, the compressed archives were extracted automatically, and all valid image files were identified for further processing. This automated acquisition step reduced manual intervention and ensured that the experimental pipeline could be rerun from the original downloadable sources.

The raw downloaded data were not immediately suitable for direct model training because the images originated from different collections and showed substantial class imbalance. The extracted images were reorganized into two task-specific categories, denoted as anemia-like and control-like. The class distribution was inspected during preprocessing to identify imbalance before constructing the final balanced dataset. Since the control-like category contained more images than the anemia-like category, the majority class was downsampled to match the minority class. After balancing, the final dataset contained 158 anemia-like images and 158 control-like images, having a total of 316 images for binary classification.

Downsampling was selected for this proof-of-concept experiment because it created a simple and controlled balanced setting in which both classes contributed equally to training and evaluation metrics. However, this choice also reduces majority-class diversity and may not reflect real clinical prevalence. Alternative strategies such as class weighting, focal loss, oversampling, and synthetic augmentation may preserve more information and should be assessed in future work.

To reduce the risk of data leakage, partitioning was performed at the image level before training augmentation was applied. Random augmentations were applied only to the training subset, while validation and test images were processed using deterministic resizing and tensor conversion only. Filename-level checking was performed during dataset restructuring. Duplicate and near-duplicate microscopy images remain an important threat to validity because visually similar images may arise from the same acquisition setting and could inflate performance estimates if present across training, validation, and test subsets.

C. Data Partitioning Strategy

After balancing the dataset, the images were partitioned into training, validation, and test subsets using a 70:20:10 split. This produced 221 training samples, 63 validation samples, and 32 test samples. The test subset contained 15 anemia-like images and 17 control-like images. Because the held-out test set contained only 32 images, the results should be interpreted cautiously; even perfect test-set classification would not by itself confirm robust generalization. Training images were used for weight optimization, validation images were used to monitor model behavior and select the best epoch, and the test subset was reserved for final performance assessment. The held-out test set was used only once after validation-based model selection and was not used during training, augmentation, and hyperparameter tuning.

Random shuffling was performed using a fixed seed before partitioning so that the same train-validation-test split could be recreated across repeated runs of the workflow. The reported results correspond to one fixed-seed run and one fixed split. Cross-validation, bootstrapping, and multiple independent training runs were not conducted in the current version.

The validation subset was used for model selection. The network state achieving the highest validation accuracy was retained as the best model. This separation reduces direct test-set leakage, but it does not remove the broader limitations associated with small test-set size and single-split evaluation.

D. Preprocessing and Data Augmentation

All images were resized to 224×224 pixels before model training to standardize input dimensions and ensure compatibility with the selected architecture. Training-only augmentation was applied to improve generalization and reduce overfitting, including horizontal flipping, controlled vertical flipping, rotation, color jitter, and affine transformation. Validation and test images received only deterministic resizing and tensor conversion, with no random augmentation, ensuring stable and fair evaluation.

E. Network Architecture

A ResNet18 network, pretrained on ImageNet, was used as the classification backbone. ResNet18 was selected because it provides a suitable balance between feature representation and computational efficiency for a relatively small blood-cell image dataset. Transfer learning was used instead of training a model from scratch, allowing the network to reuse visual representations learned from large-scale image data. This improved optimization efficiency and reduced the amount of task-specific data required.

The original fully connected layer was replaced with a two-class output layer for predicting anemia-like and control-like images. The pretrained convolutional feature extractor was retained, while the final classification layer was adapted to the target task. This choice also supported reproducibility because ResNet18 is widely available, computationally accessible, and commonly used in image classification research.

F. Training Procedure

The network was trained using cross-entropy loss, which is appropriate for supervised two-class classification. Optimization was performed using the Adam optimizer with a learning rate of 1×10^{-4}, supporting stable convergence during transfer learning. Training was conducted over multiple epochs, with training loss, training accuracy, validation loss, and validation accuracy recorded at each epoch to monitor convergence behavior and identify possible overfitting. The final model was selected based on validation performance, where the network state with the highest validation accuracy was retained as the best model and used for final test evaluation. This avoided selecting the model based on test-set performance. Learning curves for loss and accuracy were also saved automatically to provide a visual summary of training stability, validation behavior, and overall convergence.

G. Visualization and Output Generation

The proposed workflow integrates visualization and output generation to support transparency and reproducibility. Representative image panels are generated from the training, validation, and test subsets to inspect image appearance, class characteristics, and subset composition. Class-distribution plots are also produced to verify the balanced structure of the dataset. At the model level, the framework generates and automatically saves learning curves, confusion matrices, and performance bar charts to summarize convergence behavior, class-level prediction behavior, and overall classifier performance. The workflow also supports visualization of incorrectly predicted test samples where misclassifications occurred, enabling inspection of potential failure cases. These visual and structured outputs make the pipeline easier to audit, interpret, and reproduce.

H. Evaluation Metrics

The proposed framework was evaluated using multiple performance metrics to provide a more informative assessment than accuracy alone. Standard measures include accuracy, precision, recall, and F1-score, supplemented with balanced accuracy, specificity, sensitivity, Matthews correlation coefficient, Cohen's kappa, and ROC-AUC to better characterize class-level performance, agreement, and threshold-based separability. Confusion-matrix analysis is also used to show the distribution of true positives, true negatives, false positives, and false negatives. In addition, all predictions, summary metrics, classification reports, and visual figures are automatically exported as structured files. This ensures that the evaluation outputs could be inspected, reused, compared, and reproduced in future experiments.

III. EXPERIMENTAL SETUP AND EVALUATION METRICS

Baseline models, including InceptionV3, MobileNetV2, ShuffleNetV2, and ResNet50, were compared using the same split, preprocessing, augmentation policy, training procedure, and evaluation metrics. These models were selected to represent commonly used convolutional image-classification backbones with different characteristics: InceptionV3 as a multi-scale convolutional architecture, MobileNetV2 and ShuffleNetV2 as lightweight efficient models, and ResNet50 as a deeper residual model to compare with the compact ResNet18 framework. The comparison was treated as indicative rather than a definitive architecture-ranking study. All models used the same fixed 70:20:10 split, 12 epochs, Adam optimizer, learning rate of 1×10^{-4}, batch size of 16, 224 × 224 input size, training-only augmentation, validation-based model selection, and the same held-out test set.

TABLE I. TRAINING CONDITIONS OF THE PROPOSED AND BASELINE MODELS

Model	Training conditions	Evaluation
InceptionV3	Same as the proposed framework	Same held-out test set
MobileNetV2	Same as the proposed framework	Same held-out test set
ShuffleNetV2	Same as the proposed framework	Same held-out test set
ResNet50	Same as the proposed framework	Same held-out test set
Proposed ResNet18	Same fixed protocol	Same held-out test set

IV. RESULTS AND DISCUSSION

The proposed framework showed stable convergence, with training accuracy of 94.12% and validation accuracy of 80.95% in epoch 1. Validation accuracy increased to 85.71% in epoch 2 and 98.41% in epoch 3, with later epochs reaching full validation accuracy, while validation loss decreased from 0.3581 to 0.0072. These trends indicate stable optimization, improved prediction confidence, and progressive learning of discriminative blood-cell features, as shown in Figure 2.

Fig. 2. Training and validation accuracy curves.

The final performance on the held-out test subset was strong in this fixed-split proof-of-concept evaluation. All 32 held-out test images were correctly classified, corresponding to 100.0% accuracy. Other evaluation measures, including balanced accuracy, precision, recall, F1-score, specificity, sensitivity, Matthews correlation coefficient, Cohen's kappa, and ROC-AUC, also reached their maximum values in this run. However, since the held-out test set contained only 32 images, uncertainty remains substantial. For the observed held-out test accuracy of 32/32 correct classifications, the approximate 95% exact binomial confidence interval is 89.1–100.0%. This wide interval shows that the fixed-split result remains statistically uncertain and should be interpreted cautiously. Table II presents a comparison of classification accuracy between the proposed framework and baseline deep learning models, obtained from the final predictions on the same held-out test set after validation-based model selection. The proposed ResNet18-based framework achieved the highest accuracy among the compared models in this experimental setting. This may be because ResNet18 is less prone to overfitting than deeper models in this small-dataset setting. However, this result is dataset-specific and should not be interpreted as general superiority over deeper or more recent architectures.

TABLE II. ACCURACY COMPARISON OF THE PROPOSED AND BASELINE MODELS

Model	Accuracy
InceptionV3	0.8304
MobileNetV2	0.8371
ShuffleNetV2	0.8594
ResNet50	0.8728
Proposed ResNet18	1.0000

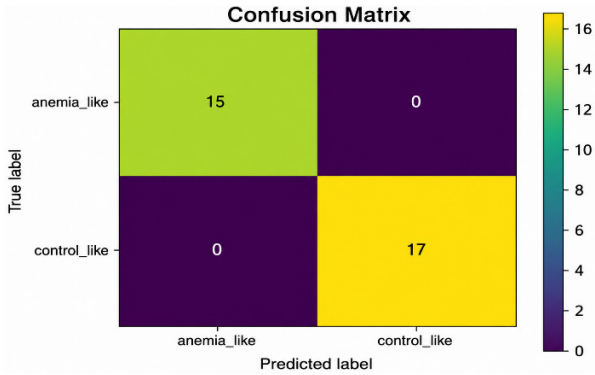

Fig. 3. Confusion matrix.

The confusion matrix further confirmed the observed test-set performance. As shown in Figure 3, all test images were correctly classified and no false positives or false negatives were observed in this held-out subset. Specifically, all 15 anemia-like images and all 17 control-like images were assigned to their correct classes. This explains the perfect values obtained for specificity, sensitivity, Matthews correlation coefficient, Cohen's kappa, and ROC-AUC in the reported run. Figure 4 shows representative correctly classified held-out test samples, where each image is displayed with its ground-truth and predicted label. These examples provide qualitative support for the reported test-set results and allow visual inspection of the test images used in the final evaluation.

A strength of the proposed framework is that the reported performance was obtained after explicit class balancing and training-only augmentation. This is important because the original dataset was skewed toward the control-like class, which could have produced biased performance estimates if left unaddressed. The balanced setting allowed both classes to contribute equally to training and evaluation metrics.

Despite these promising findings, the results should be interpreted with caution. The final test subset contained only 32 images, including 15 anemia-like and 17 control-like samples. Therefore, despite the perfect test performance is encouraging, it should be viewed as proof-of-concept evidence rather than definitive evidence of broad clinical generalization.

V. LIMITATIONS AND THREATS TO VALIDITY

This study has some important limitations. The held-out test set contained only 32 images, so the reported fixed-split performance should be interpreted as proof-of-concept evidence rather than robust clinical validation. The approximate 95% exact binomial confidence interval for the observed 32/32 held-out test accuracy is 89.1–100.0%, further highlighting the uncertainty caused by the small test subset. The evaluation was based on one fixed seed and one train-validation-test split, without cross-validation, bootstrapping, or repeated independent runs, meaning that the results may be sensitive to data partitioning and training variability. The dataset was also artificially balanced through downsampling, which reduced majority-class diversity and may not reflect real clinical prevalence. In addition, no external validation was performed across independent datasets, acquisition protocols, staining conditions, scanners, or clinical populations.

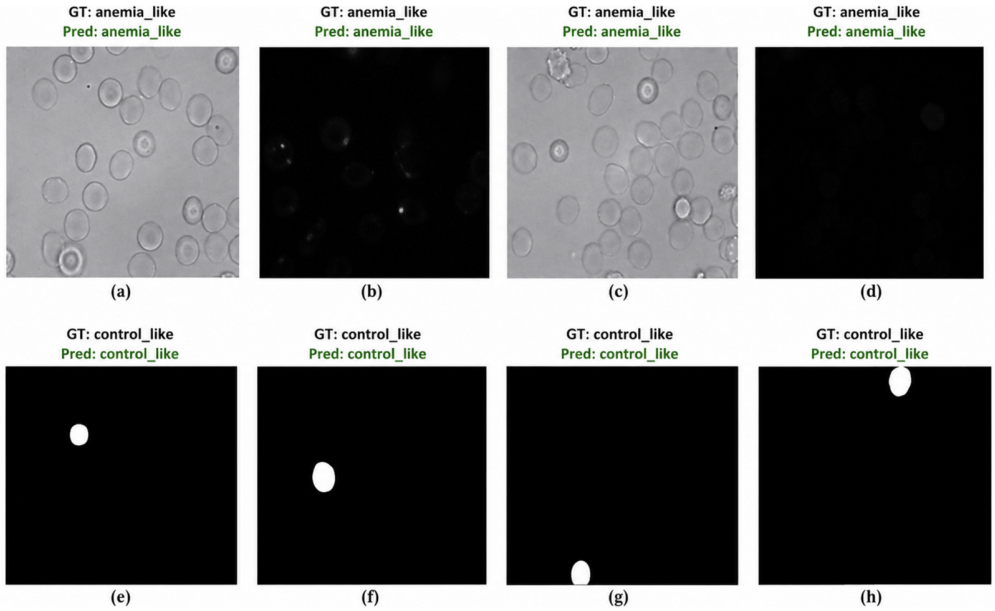

Fig. 4. Representative correctly classified held-out test images with Ground-Truth (GT) and Predicted (Pred.) labels. Panels (a–d) show anemia-like samples, and panels (e–h) show control-like samples.

Although training augmentation was separated from validation and test preprocessing, possible duplicate or near-duplicate images remain an important threat to validity in small public microscopy datasets. The current workflow also lacks augmentation ablation, Grad-CAM, feature visualization, and detailed interpretability analysis. Finally, the baseline comparison should be considered indicative unless all models are trained under identical splits, augmentation policies, training schedules, and hyperparameter settings, ideally with repeated runs and statistical testing.

VI. CONCLUSION AND FUTURE RESEARCH

This paper presented a reproducible end-to-end deep learning framework for proof-of-concept blood-cell image classification using publicly accessible data. This study addressed the need for transparent and reproducible workflows in anemia-related image analysis by integrating dataset download, extraction, class balancing, augmentation-enhanced preprocessing, transfer-learning-based classification, visualization, and metric generation within a single executable pipeline. Under a fixed-seed single-split evaluation, the ResNet18-based model correctly classified all 32 held-out test images, demonstrating feasibility in a controlled proof-of-concept setting. The main contribution is a reproducible engineering workflow rather than a new classification architecture or evidence of clinical readiness. The findings should be interpreted cautiously because the test set was small, no cross-validation or external validation was performed, and rigorous duplicate screening remains future work. Future studies should validate the workflow using larger and more diverse datasets, perform repeated-run and cross-validation analysis, compare class-balancing and augmentation strategies, evaluate external datasets, incorporate Grad-CAM or related explainability methods, strengthen duplicate screening using image-hash and similarity-based approaches, and release verification files to support independent reproducibility.

DECLARATION OF COMPETING INTERESTS

The authors declare no known competing financial interests or personal relationships that could have influenced this work.

ACKNOWLEDGMENT

The authors acknowledge the dataset providers, open-source software developers, and their respective institutions for supporting this work.

DATA AVAILABILITY AND REPRODUCIBILITY

The study used the publicly available Red Blood Cell RedTell Dataset in [18] and provides the code and the 32-image held-out test panel through the project repository at [19].

AI USE AND DECLARATION OF GENERATIVE AI USE

During the preparation of this work, the authors used generative AI tools only for language refinement and editing support. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

REFERENCES

[1] M. A. Warner and A. C. Weyand, "The Global Burden of Anemia," in *Blood Substitutes and Oxygen Biotherapeutics*, H. Liu, A. D. Kaye, and J. S. Jahr, Eds. Springer International Publishing, 2022, pp. 53–59.

[2] K. T. Navya, K. Prasad, and B. M. K. Singh, "Analysis of red blood cells from peripheral blood smear images for anemia detection: a methodological review," *Medical & Biological Engineering & Computing*, vol. 60, no. 9, pp. 2445–2462, Sept. 2022, https://doi.org/10.1007/s11517-022-02614-z.

[3] A. Husham, M. Hazim Alkawaz, T. Saba, A. Rehman, and J. Saleh Alghamdi, "Automated nuclei segmentation of malignant using level sets," *Microscopy Research and Technique*, vol. 79, no. 10, pp. 993–997, Oct. 2016, https://doi.org/10.1002/jemt.22733.

[4] S. Divya, L. P. Suresh, and A. John, "A Deep Transfer Learning framework for Multi Class Brain Tumor Classification using MRI," in

2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), Dec. 2020, pp. 283–290, https://doi.org/10.1109/ICACCCN51052.2020.9362908.

[5] L. Alzubaidi, M. A. Fadhel, O. Al-Shamma, J. Zhang, and Y. Duan, "Deep Learning Models for Classification of Red Blood Cells in Microscopy Images to Aid in Sickle Cell Anemia Diagnosis," *Electronics*, vol. 9, no. 3, Mar. 2020, Art. no. 427, https://doi.org/10.3390/electronics9030427.

[6] M. M. Alam and M. T. Islam, "Machine learning approach of automatic identification and counting of blood cells," *Healthcare Technology Letters*, vol. 6, no. 4, pp. 103–108, Aug. 2019, https://doi.org/10.1049/htl.2018.5098.

[7] S. Sharma and D. Dudeja, "Towards Intelligent Anemia Detection: An Empirical Analysis of Clinical and Imaging-Based Machine Learning Approaches," in *2025 IEEE 7th International Conference on Computing, Communication and Automation (ICCCA)*, Nov. 2025, pp. 1–6, https://doi.org/10.1109/ICCCA66364.2025.11325119.

[8] K. K. Mohammed, N. Dahmani, R. Ahmed, A. Darwish, and A. E. Hassanien, "An Explainable AI and Optimized Multi-Branch Convolutional Neural Network Model for Eye Anemia Diagnosis," *IEEE Access*, vol. 13, pp. 71840–71857, 2025, https://doi.org/10.1109/ACCESS.2025.3560689.

[9] J. B. Lazaro, J. C. Dela Cruz, and J. F. Villaverde, "An adaptive filter for anemia screening using deep convolutional neural network," *Franklin Open*, vol. 12, Sept. 2025, Art. no. 100345, https://doi.org/10.1016/j.fraope.2025.100345.

[10] A. A. Mahmud, P. Chowdhury, M. B. Uddin, K. E. Delowar, T. R. Talha, and B. Dewanjee, "AI-Driven anemia diagnosis: A review of advanced models and techniques." arXiv, 2025, https://doi.org/10.48550/ARXIV.2510.11380.

[11] Y. Tewachew, D. Dagneb, M. Andualem, and Y. Mekuriaw, "Predicting the Level of Anemia among Ethiopian Neonatal Using Ensemble Machine Learning Algorithms." In Review, Feb. 09, 2026, https://doi.org/10.21203/rs.3.rs-8548736/v1.

[12] K. Malarkodi, "AnemiaX: A Predictive Intelligence Model for Hematological Health," in *2025 3rd International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI)*, Sept. 2025, pp. 162–168, https://doi.org/10.1109/ICoICI65217.2025.11254642.

[13] M. Karimi *et al.*, "Feature Selection Methods in Big Medical Databases:A Comprehensive Survey," *International Journal of Theoritical & Applied Computational Intelligence*, pp. 181–209, 2025, https://doi.org/10.65278/IJTACI.2025.21.

[14] M. A. Khan, M. I. Sharif, M. Raza, A. Anjum, T. Saba, and S. A. Shad, "Skin lesion segmentation and classification: A unified framework of deep neural network features fusion and selection," *Expert Systems*, vol. 39, no. 7, Aug. 2022, Art. no. e12497, https://doi.org/10.1111/exsy.12497.

[15] N. H. Q. Nguyen and A. C. Phan, "Explainable AI-Based Approach for Automated Blood Cell Classification," in *Advances in Information and Communication Technology*, vol. 1831, T. X. Tu, P. T. Nghia, N. T. Thuy, V. D. Thai, L. H. Son, and N. Van Nui, Eds. Springer Nature Switzerland, 2026, pp. 268–278.

[16] N. H. Q. Nguyen, T. T. Nguyen, and A. C. Phan, "A Lightweight Explainable Deep Learning for Blood Cell Classification," *Computer Modeling in Engineering & Sciences*, vol. 145, no. 2, pp. 2435–2456, 2025, https://doi.org/10.32604/cmes.2025.070419.

[17] N. Ayesha and H. Khalidi, "A Vision Transformer with a Self-Attention Mechanism for High-Accuracy Blood Cell Classification," *Engineering, Technology & Applied Science Research*, vol. 16, no. 1, pp. 32557–32563, Feb. 2026, https://doi.org/10.48084/etasr.16600.

[18] A. Makhro *et al.*, "Red Blood Cell RedTell Dataset." Zenodo, Apr. 05, 2023, https://doi.org/10.5281/ZENODO.7801430.

[19] D. Saleela, "divsal009/Anemia." May 13, 2026, [Online]. Available: https://github.com/divsal009/Anemia.