A Multimodal Framework For Anaemia Screening Using Images And Clinical Features: A Comprehensive Survey And Methodological Proposal

Aswathy S
Dept. of Artificial Intelligence and Data Science
Rajagiri School of Engineering & Technology
Kakkanad, Kerala, India
aswathysanthosh773@gmail.com

Liyan Grace Shaji
Dept. of Artificial Intelligence and Data Science
Rajagiri School of Engineering & Technology
Kakkanad, Kerala, India
liyang@rajagiritech.edu.in

***Abstract*—Anaemia is a major global health burden that urgently demands accessible and non-invasive screening solutions. Traditional single modality diagnostic methods often fail to capture subtle physiological cues, leading to reduced sensitivity and limiting their reliability in real world clinical settings. To overcome these limitations, we survey recent deep-learning approaches for non-invasive anaemia detection and propose a high efficiency m ultimodal f ramework t hat c ombines visual features extracted by EfficientNetV2 with structured clinical data such as age, hemoglobin level, and gender. The model addresses class imbalance using SMOTE and leverages a Multi Head Self Attention fusion layer to dynamically weight and integrate information across modalities, capturing complex interdependencies that simple early or late-fusion strategies often miss. Moreover, explainable AI tools, including Grad-CAM and SHAP, are embedded within the pipeline to provide both visual and quantitative interpretability, enabling clinicians to understand the basis of predictions and fostering trust in the system. This unified, a ttention d riven m ultimodal a pproach o ffers a scalable and robust framework for non-invasive anaemia screening, with potential to enhance early detection and support clinical decision making.**

***Index Terms*—Anaemia Screening, Multimodal Fusion, EfficientNetV2, SMOTE, Multi-Head Self-Attention, Explainable AI**

I. Introduction

Anaemia remains one of the most widespread global health challenges, affecting billions of individuals and disproportionately burdening low- and middle-income countries. Despite being highly preventable and treatable, its persistence is driven by structural barriers such as poor access to diagnostic facilities, nutritional deficiencies, i nfectious d iseases, a nd chronic health conditions. Accurate detection traditionally relies on invasive venous blood sampling for haemoglobin estimation, a gold-standard method that requires laboratory infrastructure, trained personnel, and consumables. These constraints limit its scalability and feasibility for large community-level screening initiatives, emphasizing the urgent need for robust, low-cost, and non-invasive diagnostic alternatives.

Recent developments in computer vision and machine learning have catalyzed significant progress in non-invasive anaemia screening. Early systems utilized simple color-based features extracted from conjunctiva, palm, or fingernail images, enabling rapid estimation of pallor-related biomarkers through classical machine learning techniques. Although these approaches achieved promising accuracy in controlled environments, they were hindered by sensitivity to illumination, camera variability, and demographic differences. The class imbalance common in anaemia datasets further reduced their robustness, necessitating methods such as SMOTE, ADASYN, and GAN-based augmentation to improve representativeness and generalization.

Deep learning methods have since emerged as the dominant paradigm, offering superior feature extraction and higher diagnostic performance across imaging modalities. Convolutional Neural Networks applied to conjunctival and retinal fundus images have shown strong correlation with haemoglobin levels, while transformer-based architectures further enhance global contextual understanding. In addition, studies on peripheral blood smear images have demonstrated the relevance of micro-level cell morphology for anaemia detection. Nevertheless, most existing solutions remain unimodal, relying solely on images or clinical records, which limits their ability to capture the complex, multifactorial nature of anaemia and reduces their robustness across diverse populations.

To address these limitations, multimodal learning has gained increasing traction. By combining imaging data with structured clinical features such as age, gender, and haemoglobin measurements, multimodal frameworks provide a more holistic representation of patient health. Recent literature reports notable improvements in diagnostic accuracy, stability, and clinical relevance through late fusion, feature-level fusion, and attention-based modeling. However, many modern architectures remain computationally heavy, data-hungry, or insufficiently interpretable factors that hinder their deploy-

DOI: 10.5281/zenodo.19827041

ment in real-world clinical settings, particularly in resource-constrained environments.

In response to these challenges, this paper proposes a high efficiency multimodal anaemia screening framework that integrates visual features extracted via EfficientNetV2 with key clinical attributes. Class imbalance is addressed using SMOTE to ensure balanced learning, while a Multi Head Self Attention fusion mechanism enables dynamic weighting across modalities to enhance predictive power. To ensure clinical transparency, the model incorporates Explainable AI tools Grad-CAM for spatial interpretability and SHAP for feature attribution providing both anatomical and quantitative insights into model decisions. Through this unified and interpretable design, the framework aims to advance scalable, accurate, and trustworthy non-invasive anaemia screening for diverse global healthcare settings.

II. Diagnostic Modalities and Data Integration

A comprehensive understanding of anaemia screening requires familiarity with the core biomarkers and data modalities relevant to computational diagnosis. Anaemia manifests both physiologically and morphologically, producing measurable changes in systemic blood parameters and retinal vascular characteristics. Modern machine learning approaches leverage these biomarkers to construct more reliable, non-invasive diagnostic systems, particularly when combining clinical indicators with image derived features. This integration forms the basis of multimodal learning, enabling models to capture complementary aspects of disease presentation and overcome the limitations of single-modality methods.

A. Clinical Biomarkers

Clinical biomarkers provide the quantitative foundation for anaemia assessment. The primary indicator is haemoglobin (Hb) concentration, which directly reflects the oxygen carrying capacity of blood. Additional features such as age, gender, red blood cell (RBC) indices, mean corpuscular volume (MCV), mean corpuscular haemoglobin concentration (MCHC), and total RBC count offer deeper insights into the subtype and severity of anaemia. These clinical attributes are widely used in machine learning based diagnostic systems because they encode physiological information that is not always visually apparent in images. Deep models that incorporate structured clinical variables exhibit improved predictive stability and interpretability, especially when training datasets are augmented using class imbalance techniques such as SMOTE, which ensures equal representation of anaemic and non-anaemic groups and prevents biased learning.

B. Retinal Imaging as a Non-Invasive Modality

Retinal fundus imaging has emerged as a promising non-invasive modality for anaemia detection due to the rich microvascular information it provides. Changes in vessel caliber, tortuosity, oxygenation patterns, and optic disc coloration can reflect systemic haemoglobin levels. Convolutional Neural Networks (CNNs) have demonstrated strong capability in extracting discriminative vascular and textural features from fundus photographs, enabling high accuracy predictions of anaemia status. Compared with external pallor based imaging, retinal imaging captures deeper physiological cues linked to blood health, making it a reliable modality for deep learning–based screening pipelines.

C. Morphological and Vascular Feature Extraction

Retinal images also contain subtle morphological biomarkers correlated with chronic and acute haemoglobin deficiencies. Decreased arterial venous contrast, reduced vessel density, and altered microvascular branching patterns can be quantified using modern image-analysis techniques. Deep learning models such as EfficientNet, ResNet, and transformer based encoders have shown superior performance in learning vascular morphology directly from raw images, eliminating the need for handcrafted feature engineering. The incorporation of such vascular markers significantly enhances model sensitivity to early stage anaemia, especially when combined with clinical features.

D. Motivation for Multimodal Fusion

Despite the strong predictive power of individual modalities, neither clinical data nor retinal images alone capture the full spectrum of anaemia's physiological effects. Clinical biomarkers provide numerical precision, while retinal images offer a visual manifestation of systemic blood status. Multimodal fusion leverages the complementary nature of these data streams by jointly learning from structured clinical information and deep visual features. Advanced fusion techniques including Multi Head Self Attention, feature level integration, and adaptive weighting enable models to dynamically prioritize the most informative patterns across modalities. This results in improved diagnostic accuracy, robustness, and generalizability across different populations and imaging conditions.

E. Explainability and Computational Needs

As multimodal systems grow in complexity, clinical adoption requires interpretable and trustworthy predictions. Explainable AI (XAI) techniques such as Grad-CAM provide spatial heatmaps that highlight retinal regions contributing to the model's prediction, while SHAP assigns quantitative importance scores to clinical features (e.g., Hb, MCV, age). These tools ensure transparency and enable clinicians to validate whether the model's reasoning aligns with established physiological understanding. Combined with class-balancing strategies like SMOTE and efficient architectures such as EfficientNetV2, these methods enable the development of accurate, scalable, and clinically interpretable anaemia screening systems suitable for real world deployment.

III. Modelling Approaches for Anaemia Detection

To contextualize modern computational approaches for anaemia detection, it is essential to define a structured taxonomy. This classification highlights the evolution from early handcrafted methods to advanced multimodal deep learning

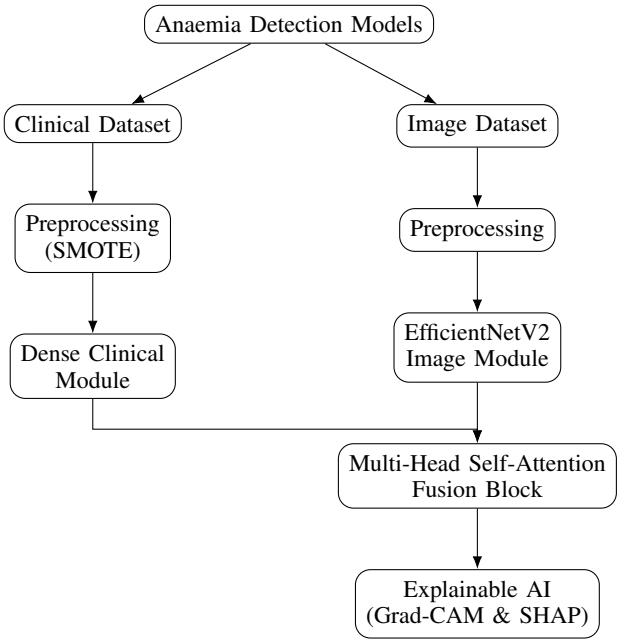

Fig. 1. Revised taxonomy of methods in multimodal anaemia detection.

systems capable of fusing heterogeneous data. We categorize the literature and methodological trends into four overarching families: (1) classical machine learning models using engineered clinical features, (2) unimodal deep learning systems applied to retinal images, (3) deep tabular models leveraging structured clinical biomarkers, and (4) multimodal fusion architectures that integrate visual and clinical data using sophisticated attention mechanisms. This progression reflects the broader shift from manual feature engineering toward end to end learning and cross-modal representation extraction. The high-level taxonomy illustrated in Fig. 1 captures these categories and situates the proposed model within this landscape.

A. Classical Machine Learning Approaches

Historically, anaemia prediction relied heavily on classical machine learning (ML) applied to structured clinical features such as haemoglobin (Hb), mean corpuscular volume (MCV), RBC count, age, and gender. Methods such as SVMs, logistic regression, decision trees, and Random Forests extracted insights from tabular biomarkers with strong interpretability and low computational overhead. These approaches required extensive feature engineering and domain expertise, limiting their adaptability and sensitivity to subtle disease patterns. Moreover, anaemia datasets are typically imbalanced, with fewer positive cases; thus, oversampling techniques like SMOTE were frequently employed to mitigate biased learning. Although these classical systems laid the groundwork for computational anaemia detection, they lacked the representational power needed for complex real-world images.

B. Deep Learning Models (Retinal Images)

The introduction of deep learning transformed anaemia screening by enabling CNN-based systems to learn directly from raw retinal images. Retinal fundus photographs contain rich microvascular cuescsuch as vessel caliber, tortuosity, and optic disc coloration that correlate strongly with haemoglobin levels. Architectures like ResNet, Inception, and EfficientNet automated the extraction of hierarchical visual features, surpassing classical ML approaches. In this domain, EfficientNetV2 has emerged as a state of the art choice: its compound scaling strategy balances depth, width, and resolution to optimize accuracy and efficiency. Transfer learning from ImageNet further boosts performance by leveraging preclearned representations of shapes, textures, and edges. Although highly effective, unimodal image models are limited because they lack access to physiological information encoded in clinical biomarkers.

C. Deep Learning Models for Clinical Biomarkers

Another unimodal direction involves deep neural networks trained solely on clinical data. These models typically employ multiblayer perceptrons (MLPs) or dense networks to learn nonlinear relationships between clinical variables most critically, haemoglobin concentration, along with demographic attributes such as age and gender. Despite the small dimensionality of clinical data, dense architectures with dropout regularization effectively capture complex patterns and generate meaningful embeddings (e.g., a 1280-dimensional clinical feature vector). These models are computationally lightweight and perform well on structured data; however, they lack the visual context carried by retinal images and therefore cannot capture microvascular manifestations of anaemia.

D. Multimodal Fusion Hybrids

Recognizing that anaemia presents through both physiological biomarkers and retinal vascular signatures, the most advanced systems integrate both modalities. Modern multimodal frameworks employ parallel encoders such as EfficientNetV2 for retinal images and a dense clinical network for haemoglobin and demographic features—to extract high level embeddings from each data type. While simple concatenation represents an early fusion approach, more sophisticated methods have emerged. The most powerful among them use Multi Head Self Attention within the fusion block. This mechanism enables the model to learn cross modal dependencies, dynamically weigh informative feature interactions, and capture subtle relationships between clinical signals (e.g., Hb level) and retinal patterns (e.g., vessel contrast). Such deep fusion approaches consistently outperform both unimodal systems and naive concatenation by producing a unified 2560-dimensional representation rich in multimodal context.

E. Explainability and Trustworthy Deep Models

As multimodal models grow increasingly complex, clinical deployment requires robust interpretability. To address this, XAI techniques are integrated directly into the modeling

pipeline. Grad-CAM provides spatial heatmaps that highlight retinal regions driving the CNN's decision typically vascular structures or disc-related features—while SHAP explains the influence of clinical features by quantifying their contribution to the final prediction. Together, they deliver both visual and numerical transparency, allowing clinicians to evaluate whether the model's reasoning aligns with medical knowledge. These tools also support fairness and bias detection by revealing whether the model relies on clinically appropriate cues. When combined with SMOTE for balanced training and an efficient backbone like EfficientNetV2, the resulting multimodal architecture becomes accurate, interpretable, and suitable for real-world anaemia screening at scale.

IV. Methodological Review of Anaemia Detection

This section provides a comprehensive review of representative anaemia detection methodologies across the major modeling families identified in our taxonomy. We highlight the architectural strategies adopted, the modalities they leverage, and the limitations that motivate the development of advanced multimodal systems.

A. Classical Machine Learning and Feature Engineering

Early computational approaches for anaemia detection predominantly relied on classical machine learning (ML) applied to structured clinical biomarkers. These pipelines typically constructed feature vectors comprising haemoglobin level, red blood cell indices (MCV, MCHC, RBC count), and demographic variables such as age and gender. Classifiers including Support Vector Machines, Decision Trees, Logistic Regression, Random Forests, and Naïve Bayes were widely adopted due to their interpretability and low computational cost. Despite their utility, these models were constrained by their dependence on manually engineered features, which limited their ability to capture complex interactions and subtle physiological patterns. Furthermore, the inherent class imbalance in anaemia datasets often resulted in biased predictions, requiring resampling techniques such as SMOTE to restore distributional balance. While these classical systems established important baselines, their predictive capability remained inferior to modern deep learning approaches.

B. Unimodal Deep Learning Models

1) Deep Learning for Retinal Image Analysis

Deep learning marked a significant shift by enabling end to end learning from retinal fundus images. Convolutional Neural Networks (CNNs) such as ResNet, Inception, and EfficientNet extracted hierarchical vascular features associated with haemoglobin deficiency, including alterations in vessel calibre, tortuosity, and optic disc reflectance. EfficientNetV2, in particular, demonstrated strong performance due to its optimized compound scaling and transfer learning from ImageNet. These models achieved high diagnostic accuracy but remained limited by their reliance on a single visual modality, lacking access to critical physiological biomarkers.

2) Deep Learning for Clinical Biomarkers

Another unimodal line of research employed multi layer perceptrons to learn complex relationships within structured clinical data. Dense networks with dropout regularization produced compact clinical embeddings capable of representing haemoglobin values and associated haematological indices. Although computationally lightweight and effective for numerical data, clinical-only models failed to integrate retinal vascular cues and therefore lacked sensitivity to early visual manifestations of anaemia.

3) Limitations of Unimodal Approaches Both image only and clinical only models suffer from inherent modality blindness: CNNs cannot interpret numerical biomarkers, and dense networks cannot perceive retinal abnormalities. This inability to leverage complementary information motivated the emergence of multimodal frameworks.

C. Attention-Enhanced and Hybrid Models

Recent studies have introduced attention mechanisms to enhance both unimodal and hybrid architectures. In particular, spatial and channel attention modules have been integrated into convolutional neural networks (CNNs) to emphasize salient retinal regions and refine intermediate feature maps. While these extensions improve the quality and discriminative power of visual representations, the resulting architectures remain predominantly unimodal and are therefore limited in their ability to capture cross modal dependencies between visual and clinical information. Nevertheless, these works provide important evidence for the effectiveness of attention mechanisms in medical representation learning and motivate their integration into more expressive multimodal frameworks. and laid the foundation for cross-modal attention techniques.

D. Multimodal Fusion Strategies

Multimodal architectures represent the most advanced direction in anaemia detection, integrating heterogeneous data sources into a unified predictive model. Typical designs employ parallel encoder such as EfficientNetV2 for retinal images and a dense network for clinical biomarkers—to generate modality specific embeddings.

1) Early and Late Fusion

Early fusion integrates features at the input level, while late fusion merges the independent predictions of modality specific models. Although easy to implement, these strategies do not capture meaningful interactions between modalities and therefore exhibit suboptimal performance.

2) Deep Fusion with Multi Head Self Attention

State of the art multimodal systems employ deep fusion layers incorporating Multi Head Self Attention (MHSA). This mechanism allows the model to dynamically weigh image and clinical embeddings, capturing non-linear cross modal relationships and enabling individualized feature prioritization. By learning interaction patterns between retinal features and biomarkers such as haemoglobin, MHSA based fusion consistently outperforms naïve concatenation approaches.

3) Explainability Driven Multimodal Systems

To enhance transparency and clinical trust, modern multimodal frameworks integrate Explainable AI (XAI) modules. Grad-CAM provides spatial heatmaps identifying retinal regions influencing CNN predictions, while SHAP assigns quantitative contribution values to clinical features. These dual modality explanations offer both visual and numerical interpretability, facilitating reliability assessment, bias detection, and clinical validation.

V. Comparative Tables

To synthesize the surveyed literature, we present a comprehensive table highlighting the key methodological characteristics of recent influential works, contrasting unimodal and multimodal approaches. Direct comparison of performance metrics is often misleading due to variations in datasets, cohorts, and validation strategies. Therefore, these tables focus on modalities, architectural choices, and key innovations.

VI. Comparative Analysis and Motivation

The survey highlights a necessity for a system that simultaneously balances computational efficiency, deep feature interaction, and integrated interpretability to address the shortcomings of existing models. Our framework is conceived as a holistic solution to these challenges, drawing motivation from the limitations observed across the literature (Table I).

A. Addressing the Efficiency Accuracy Trade Off

Models achieving the highest reported accuracy often rely on complex architectures or resource intensive feature engineering (e.g., RexNet, which uses 181 engineered features), severely limiting their deployment in resource constrained settings. The integration of proprietary attention mechanisms (like RC-BAM) also adds high computational demands.

- **Motivation → Solution:** We select **EfficientNetV2** for the Image Module. This architecture is intrinsically optimized using compound scaling, yielding superior accuracy to latency performance, thus making the system practical and scalable for edge deployment and real time screening.

B. Overcoming Suboptimal Feature Fusion

Existing multimodal models primarily rely on early fusion via simple concatenation or employ proprietary attention mechanisms that are complex and computationally heavy. These approaches often fail to capture the subtle, non-linear dependencies between heterogeneous data, particularly when the image quality is variable or the clinical data is sparse. This limits the model's ability to dynamically weigh modality importance.

- **Motivation → Solution:** The framework uses a **Transformer-based Multi Head Self Attention Layer** for deep feature fusion. This mechanism dynamically learns complex, non-linear relationships between the image and clinical features, maximizing the synergistic signal and resulting in a robust, intelligently weighted fused embedding. This is superior to static weighting schemes.

C. Integrating Clinical Interpretability

The "black box" nature of deep learning is a major barrier to clinical adoption. While some studies offer fragmented explanations (e.g., Darshan et al. use SHAP/LIME but exclude the image modality), providing a unified, modality-specific explanation for the final fused decision is crucial for clinician trust.

- **Motivation → Solution:** We integrate **Grad-CAM** (for visual insight) and **SHAP** (for quantitative clinical insight) directly into the pipeline. This dual-modality XAI approach ensures transparency, providing both the anatomical basis for the prediction (via retinal/facial region localization) and the quantitative influence of structured patient data (Hb, Age, Gender).

D. Mitigating Data Imbalance and Training Robustness

Anaemia datasets are often inherently class imbalanced (e.g., fewer severe cases, or fewer minority ethnicity data points), which restricts model generalization and can lead to biased prediction, such as overestimating low Hb and underestimating high Hb levels.

- **Motivation → Solution:** Data imbalance is rigorously addressed using **SMOTE** (**Synthetic Minority Oversampling Technique**) and the use of class weighted loss functions during training. This comprehensive strategy ensures balanced learning, minimizes bias, and improves predictive robustness across all diagnostic categories.

E. Model Training and Evaluation

The entire model is trained end-to-end using the Adam optimizer with early stopping and learning-rate scheduling for optimized stability and convergence. Performance is rigorously evaluated using a comprehensive suite of metrics on the test dataset: accuracy, precision, recall, F1-score, and ROC-AUC. The integrated XAI modules play a dual role in both supporting the final diagnosis and serving as a mechanism to validate the model's reasoning against established clinical pathology.

VII. Conclusion

This comparative study highlights the rapid evolution of anaemia detection methodologies from early handcrafted machine-learning approaches to modern deep learning and multimodal fusion frameworks. Image only models, whether based on conjunctiva, palm, nail bed, retinal, or blood smear images, demonstrate strong potential for non-invasive screening but remain sensitive to illumination, device variability, and population differences. Clinical data based models provide stable, hematology driven assessment but lack the ability to capture visual biomarkers such as pallor or erythrocyte

TABLE I
COMPREHENSIVE SURVEY OF RECENT APPROACHES FOR ANAEMIA SCREENING (EXPANDED TO 20 ENTRIES)

Reference	Year	Modalities	Key Method / Techniques Used	Key Strengths / Limitations
I. Multimodal and Advanced Fusion Architectures				
Ramzan et al. [1]	2024	Conjunctiva + EHR	Multimodal fusion using MobileNetV2 enhanced with **RC-BAM**; Grad-CAM for explainability.	Non-invasive, high accuracy (up to 95%) by fusing clinical and image data. High computational cost, limited generalizability to other populations.
Ramzan et al. [10]	2024	Palm Image + Hematology Data	Multimodal fusion via **AMSA model** (Attention enriched AlexNet) and embedding layers.	Achieved superior performance ($\sim$ 99.58% accuracy) through integrated attention. Computationally intensive; restricted to binary classification.
Berghout [15]	2024	Palm, Conjunctiva, Nail (Multisite)	**RexNet (LSTM-based)** on 181 engineered features; SMOTE.	Extremely high accuracy ($\sim$ 99.83%), robust features. High model complexity and synthetic bias risk from SMOTE.
Said & Tunga [21]	2025	Hematological + Demographic	Hybrid stacking model (RF+ANN base, XGBoost meta) with Youden's J threshold optimization.	Achieved balanced and clinically reliable prediction via flexible threshold optimization. No image modality used.
II. Deep Learning and Regression Models (Image-Centric)				
Khan et al. [2], [24]	2025	Retinal Images + Clinical (Age, Gender)	Dual-task model (classification + Hb estimation) using **InceptionV3**; Grad-CAM.	High accuracy (98%) and low error (MAE $\sim$ 0.58 g/dL). Trained on a South Indian diabetic population, limiting generalizability.
Sehar et al. [13]	2025	Conjunctiva Images	**DCGAN** augmentation and stacking ensemble (GoogLeNet, VGG16, ResNet50).	Non-invasive, high AUC (AUC 0.97). Heavy reliance on GAN-generated data and RGB-derived features.
Navya et al. [14]	2025	PBS Images (Smear)	Dual model: ResNet50 classification + **YOLOv7-tiny** object detection.	Provides granular morphological insight ($\sim$ 98.4% accuracy). Modest dataset size (435 images) and annotation burden.
Suner et al. [5]	2021	Conjunctiva (Smartphone RAW)	Stepwise regression on 26 extracted color/texture features via k-fold cross validation.	Accessible using consumer-grade smartphones. Sensitive to image quality and operator involvement in ROI selection.
Bhusham et al. [25]	2023	Palpebral Conjunctiva	Deep CNN using GLCM (texture) features for classification and Hb prediction.	Automated classification and prediction using textural features. Limited to conventional palpebral images, requires ROI segmentation.
Appiahene et al. [19]	2023	Conjunctival Pallor	Dual-output deep learning framework (ViT, CNNs) under a joint loss function.	Introduced **CP-AnemiC** dataset; unified multi task architecture. High computational cost of transformer models.
III. Classical ML and Deep Tabular Models (Clinical/Hematological)				
Darshan et al. [16]	2025	Blood Attributes (24 params)	Differential diagnosis (IDA vs. AA) using Stacked Ensemble + multi tool XAI (**SHAP**, LIME).	Strong explainability using standard hematology tests. Modest sample size, retrospective data, no image modality.
Mojumdar et al. [11]	2024	Hematological Dataset	Statistical analysis and benchmark using **AnaDetect** (1,000 records).	Open access and ML ready dataset; gender-specific insights. No image context, single-center origin.
Appiahene et al. [7]	2023	Palm Images	Comparative study of ML (Naïve Bayes, k-NN, SVM) on CIELAB features.	Very high accuracy (Naïve Bayes $\sim$ 99.96%). Sensitive to lighting variations.
Dhakal et al. [18]	2023	CBC Parameters	Optimized XGB boosting ensemble strategies for pediatric prediction.	Optimized XGB boosting achieved $\sim$ 99% accuracy. Single center data, computationally heavy ensembles.
Zemariam et al. [23]	2024	Demographic + Risk Factors	Supervised ML algorithms (RF, SVM) for youth girls in Ethiopia.	Domain- specific prediction in a high risk group. Relies on survey/risk factors rather than lab-confirmed data.
Aiwale et al. [12]	2024	Nail bed images	Decision tree based RGB feature system + Expert System (ESDD).	Low cost, portable, simple GUI. Limited model expressiveness, binary output only.
Siddiqa et al. [22]	2025	Demographic + Risk Factors	Machine Learning (RF, KNN) for early childhood anemia.	Focus on prevalence and predictors in a specific region (Ghana). Risk factor analysis, not direct diagnostic imaging.
IV. Reviews, Diagnostics, and Global Burden Analysis				
Neogi et al. [3]	2020	POC Devices (HemoCue, TrueHb)	Diagnostic accuracy assessment across Indian community settings.	Rigorous multi-site, real world comparison. Showed non-invasive devices performed poorly.
Safiri et al. [4]	2021	Global Burden Data	Global assessment of anemia burden (GBD 2019) using ST GPR.	Vast global scope, high demographic resolution. Limited type specific anemia analysis.

morphology. The most promising direction observed across the literature is the shift toward multimodal systems that integrate image features with structured clinical inputs, leveraging techniques such as deep CNN encoders, attention based fusion, and data-balancing methods like SMOTE. These hybrid frameworks consistently show improved robustness, generalizability, and interpretability, especially when combined with explainable AI tools such as Grad-CAM and SHAP. Overall, the comparative analysis underscores that future anaemia screening solutions should prioritize multimodal integration, computational efficiency, and clinical transparency to ensure scalable, equitable, and trustworthy deployment in diverse healthcare settings.

VIII. Future Direction

Based on the comprehensive review and the specific limitations of the proposed multimodal framework, future research should focus primarily on enhancing the system's robustness, generalizability, and clinical utility across dynamic environments. Key future directions include transitioning the static diagnostic model to a temporal prediction system by incorporating longitudinal data analysis (i.e., multiple patient visits over time) to model disease progression trajectories, rather than just providing a single, static diagnosis. Furthermore, addressing the critical real world problem of privacy and data fragmentation requires developing a federated learning version of the framework, which would enable collaborative training across multiple institutions and diverse demographics (e.g., different continents) without requiring sensitive clinical or image data to leave local servers. Finally, the framework's biological grounding can be deepened by expanding the prediction task beyond binary classification to include the estimation of severity levels and the differential diagnosis of specific anaemia subtypes (like Iron Deficiency Anaemia vs. Aplastic Anaemia), thereby increasing its practical value for clinical management.

References

[1] M. Ramzan, M. U. Saeed, and G. Ali, "Enhancing anemia detection through multimodal data fusion: a non-invasive approach using EHRs and conjunctiva images," Discover Artificial Intelligence, vol. 4, no. 100, 2024.

[2] R. Khan, V. Maseedupally, K. A. Thakoor, R. Raman, and M. Roy, "Noninvasive anemia detection and hemoglobin estimation from retinal images using deep learning: A scalable solution for resource-limited settings," Translational Vision Science & Technology, vol. 14, no. 1, pp. 20, 2025.

[3] S. Neogi, J. Sharma, S. Pandey et al., "Diagnostic accuracy of point-of-care devices for detection of anemia in community settings in India," BMC Health Serv Res, vol. 20, no. 468, 2020.

[4] S. Safiri, A. A. Kolahi, M. Noori et al., "Burden of anemia and its underlying causes in 204 countries and territories, 1990–2019: results from the Global Burden of Disease Study 2019," J Hematol Oncol, vol. 14, no. 185, 2021.

[5] S. Suner, J. Rayner, I. U. Ozturan et al., "Prediction of anemia and estimation of hemoglobin concentration using a smartphone camera," PLoS ONE, vol. 16, no. 7, p. e0253495, 2021.

[6] 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," Med Biol Eng Comput, vol. 60, pp. 2445–2462, 2022.

[7] P. Appiahene, J. W. Asare, E. T. Donkoh et al., "Detection of iron deficiency anemia by medical images: a comparative study of machine learning algorithms," BioData Mining, vol. 16, no. 2, 2023.

[8] J. W. Asare, P. Appiahene, E. T. Donkoh et al., "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. 11, p. e12667, 2023.

[9] M. N. Garcia-Casal, et al., "Diagnosing anemia: Challenges selecting methods, addressing underlying causes, and implementing actions at the public health level," Annals of the New York Academy of Sciences, vol. 1524, no. 1, pp. 37-50, 2023.

[10] M. Ramzan, J. Sheng, M. U. Saeed et al., "Revolutionizing anemia detection: integrative machine learning models and advanced attention mechanisms," Vis. Comput. Ind. Biomed. Art, vol. 7, no. 18, 2024.

[11] M. U. Mojumdar, D. Sarker, M. Assaduzzaman et al., "AnaDetect: An extensive dataset for advancing anemia detection, diagnostic methods, and predictive analytics in healthcare," Data Brief, vol. 58, p. 111195, 2024.

[12] S. Aiwale, M. T. Kolte, V. Harpale et al., "Non-invasive Anemia Detection and Prediagnosis," Journal of Pharmacology and Pharmacotherapeutics, vol. 15, no. 4, pp. 408-416, 2024.

[13] N. Sehar, N. Krishnamoorthi, and C. V. Kumar, "Deep Learning Model-Based Detection of Anemia from Conjunctiva Images," Healthc Inform Res, vol. 31, no. 1, pp. 57-65, Jan. 2025.

[14] N. K. T, S. Verma, K. Prasad, and B. M. K. Singh, "Efficient diagnostic model for iron deficiency anaemia detection: a comparison of CNN and object detection algorithms in peripheral blood smear images," Automatika, vol. 66, no. 1, pp. 1-15, 2025.

[15] T. Berghout, "Joint Image Processing with Learning-Driven Data Representation and Model Behavior for Non-Intrusive Anemia Diagnosis in Pediatric Patients," J. Imaging, vol. 10, no. 10, p. 245, Oct. 2024.

[16] B. S. D. Darshan, N. Sampathila, G. M. Bairy et al., "Differential diagnosis of iron deficiency anemia from aplastic anemia using machine learning and explainable Artificial Intelligence utilizing blood attributes," Scientific Reports, vol. 15, p. 505, 2025.

[17] V. Geetha, et al., "Diagnostic approach to anemia in adults using machine learning," Journal of Pharmaceutical Negative Results, vol. 13, no. 9, 2022.

[18] P. Dhakal, S. Khanal, and R. Bista, "Prediction of Anemia Using Machine Learning Algorithms," International Journal of Computer Science and Information Technology, vol. 15, no. 1, pp. 15-30, 2023.

[19] P. Appiahene, J. W. Asare, E. T. Donkoh et al., "CP-AnemiC: a conjunctival pallor dataset and benchmark for anemia detection in children," Medicine in Novel Technology and Devices, vol. 18, p. 100244, 2023.

[20] P. Appiahene, J. W. Asare, E. T. Donkoh et al., "Application of Ensemble Models Approach in Anemia Detection Using Images of the Palpable Palm," Medicine in Novel Technology and Devices, vol. 20, p. 100269, 2023.

[21] R. O. Said and M. Tunga, "Hybrid Machine Learning Model for the Prediction of Anaemia," Machine Learning with Applications, p. 100741, 2025.

[22] M. Siddiqa, G. Shah, M. S. Butt et al., "Early Childhood Anemia in Ghana: Prevalence and Predictors Using Machine Learning Techniques," Children, vol. 12, no. 924, 2025.

[23] A. B. Zemariam, A. Yimer, G. K. Abebe et al., "Employing supervised machine learning algorithms for classification and prediction of anemia among youth girls in Ethiopia," Sci Rep, vol. 14, no. 9080, 2024.

[24] R. Khan, V. Maseedupally, K. A. Thakoor, R. Raman, and M. Roy, "Noninvasive Anemia Detection and Hemoglobin Estimation from Retinal Images Using Deep Learning: A Scalable Solution for Resource-Limited Settings," Translational Vision Science & Technology, vol. 14, no. 1, pp. 20, 2025.

[25] C. Bhusham, A. K. R. Poreddy, T. B. Krishna and P. Kokil, "Automated Anemia Classification and Hemoglobin Level Prediction using Deep CNN and GLCM Features of Palpebral Conjunctiva Images," in 2023 IEEE 7th Conference on Information and Communication Technology (CICT), 2023, pp. 1-6.