Current Research in Translational Medicine 74 (2026) 103560

Contents lists available at ScienceDirect

Current Research in Translational Medicine

journal homepage: www.elsevier.com/locate/retram

General review

Artificial intelligence for anemia screening, diagnosis, and management

David B. Olawade [a,b,c,*], Yinka Julianah Adeniji [d,e], Faithful A. Olatunbosun [f], Eghosasere Egbon [g], Aanuoluwapo Clement David-Olawade [h]

[a] Department of Allied and Public Health, School of Health, Sport and Bioscience, University of East London, London, United Kingdom
[b] Department of Research and Innovation, Medway NHS Foundation Trust, Gillingham ME7 5NY, United Kingdom
[c] Department of Business, Management and Health, York St John University, London, E14 2BA, United Kingdom
[d] Department of Life Sciences, Silwood Park, Imperial College London, Ascot, Berkshire, United Kingdom
[e] Department of Science and Humanities, Leicester college, Leicester, United Kingdom
[f] Department of Microbiology, Obafemi Awolowo University (OAU), Ile-Ife, Osun State, Nigeria
[g] Department of Tissue Engineering and Regenerative Medicine, Faculty of Life Science Engineering, FH Technikum, Vienna, Austria
[h] Endoscopy Unit, Glenfield Hospital, University Hospitals of Leicester, NHS Trust, Leicester, United Kingdom

ARTICLE INFO

Keywords:
Anemia
Artificial intelligence
Machine learning
Non-invasive diagnosis
Clinical decision support

ABSTRACT

Anemia affects over 1.6 billion people globally, representing a significant public health challenge, particularly in low- and middle-income countries where traditional diagnostic methods face barriers including invasive procedures, skilled personnel requirements, and inadequate laboratory infrastructure. Artificial intelligence (AI) has emerged as a promising technology offering non-invasive, rapid, and cost-effective solutions for anemia detection and management. This narrative review synthesises current literature on AI applications in anemia screening, diagnosis, and clinical management, examining methodologies, performance metrics, implementation challenges, and future research directions. We conducted a comprehensive narrative synthesis informed by systematic search principles, searching PubMed, IEEE Xplore, Scopus, and Web of Science databases with additional hand-searching and expert consultation. AI models demonstrate variable accuracy in anemia detection across diverse data sources, with performance ranging from 75–97 % AUC depending on methodology and validation approaches. Machine learning algorithms such as support vector machines, convolutional neural networks, and random forests show potential for achieving performance comparable to standard blood tests in controlled research settings. Smartphone-integrated applications and point-of-care systems show particular promise for resource-limited settings, though real-world validation remains limited. While AI shows significant potential for enhancing accessibility and diagnostic efficiency in anemia care, critical challenges including data standardisation, algorithmic bias, regulatory compliance, clinical validation in diverse populations, and deployment equity in low- and middle-income countries require urgent attention to ensure equitable implementation and clinical adoption.

1. Introduction

Anemia represents one of the most prevalent nutritional disorders worldwide, characterised by reduced hemoglobin concentration, red blood cell count, or hematocrit below established reference thresholds [1,2]. The World Health Organization estimates that anemia affects approximately 1.62 billion people globally, with the highest burden concentrated in developing nations where prevalence rates can exceed 40 % in vulnerable populations including children under five years and pregnant women [3]. This condition manifests through diverse clinical presentations ranging from mild fatigue and cognitive impairment to severe complications including heart failure and maternal mortality, substantially impacting quality of life and economic productivity across affected communities.

The pathophysiology of anemia encompasses multiple etiological pathways, with iron deficiency accounting for approximately 50 % of all cases globally [4]. Other significant causes include chronic disease, blood loss, genetic disorders such as thalassemia and sickle cell disease, nutritional deficiencies involving folate and vitamin B12, and chronic kidney disease. The complexity of anemia classification requires sophisticated diagnostic approaches that can differentiate between various subtypes to guide appropriate therapeutic interventions. Traditional

* Corresponding author.
E-mail address: d.olawade@uel.ac.uk (D.B. Olawade).

https://doi.org/10.1016/j.retram.2025.103560
Received 30 July 2025; Accepted 6 December 2025
Available online 7 December 2025

diagnostic methods rely heavily on laboratory-based complete blood count analysis, serum iron studies, and additional biochemical markers [5], which collectively provide comprehensive insights into underlying mechanisms and severity.

Current diagnostic challenges in anemia detection are particularly pronounced in resource-limited settings where healthcare infrastructure remains inadequate. The requirement for invasive blood sampling, specialised laboratory equipment, trained phlebotomists, and reliable electricity supply creates substantial barriers to timely diagnosis and treatment initiation [6]. Many healthcare facilities in low- and middle-income countries lack the necessary resources to perform routine hemoglobin testing, resulting in delayed diagnosis, inappropriate treatment, and progression to severe anemia with associated complications. Furthermore, the cost implications of repeated laboratory testing create additional barriers for patients and healthcare systems already operating under financial constraints.

The emergence of artificial intelligence technologies has opened potential opportunities for transforming anemia care delivery through innovative diagnostic and management approaches. Machine learning algorithms demonstrate remarkable capability in pattern recognition, enabling the development of non-invasive screening tools that can analyse diverse data sources including smartphone-captured images, wearable sensor data, and electronic health records [7]. These AI-powered solutions offer the potential to democratise anemia screening by eliminating traditional barriers while maintaining diagnostic accuracy comparable to conventional methods in research settings. Recent advances in deep learning, computer vision, large language models (LLMs), and biological AI models such as AlphaFold and RoseTTAFold have accelerated the development of practical AI applications suitable for deployment in various healthcare settings [8,9]. However, the translation of these research advances to validated clinical tools, particularly in low-resource settings where the anemia burden is highest, remains a significant challenge.

Despite the significant burden of anemia globally and the promising potential of AI technologies, there remains a critical gap in comprehensive understanding of current AI applications, their clinical effectiveness, and implementation challenges in anemia care. While numerous individual studies have demonstrated the feasibility of AI-based anemia detection methods, a systematic synthesis of available evidence is essential to guide future research priorities and clinical implementation strategies. This narrative review addresses this knowledge gap by examining the current state of AI applications in anemia screening, diagnosis, and management, with particular emphasis on methodological approaches, performance characteristics, and practical implementation considerations. The primary objectives include: (1) synthesising available evidence on AI-based anemia screening and diagnostic tools, (2) evaluating the performance and clinical utility of different AI methodologies, (3) identifying key challenges and limitations in current AI applications, and (4) proposing future research directions to advance AI integration in anemia care.

2. Methods

This narrative review employed a comprehensive literature search strategy to identify relevant studies examining AI applications in anemia care. We adopted a narrative synthesis approach informed by systematic search principles rather than systematic review methodology to accommodate the heterogeneous nature of AI applications, diverse methodological approaches, and emerging technologies. This approach allows for broader inclusion of innovative technologies and critical synthesis of available evidence while maintaining methodological rigour through structured search and screening protocols. A quantitative meta-analysis was not performed due to substantial heterogeneity in AI methodologies (ranging from traditional machine learning to deep learning architectures), diverse outcome measures (accuracy, AUC, sensitivity/specificity), variable validation approaches (internal cross-validation, temporal validation, external validation), and inconsistent reporting of performance metrics across studies, which would preclude meaningful statistical pooling.

Electronic databases including PubMed/MEDLINE, IEEE Xplore Digital Library, Scopus, and Web of Science were systematically searched for peer-reviewed articles published between January 2013 and December 2024. The search strategy utilised a combination of Medical Subject Headings (MeSH) terms and free-text keywords including "anemia" OR "anaemia," "artificial intelligence," "machine learning," "deep learning," "neural networks," "non-invasive screening," "smartphone diagnosis," "clinical decision support," "digital health," and "point-of-care testing." Additional searches included terms such as "large language models," "foundation models," "transformer networks," and "federated learning" to capture recent AI developments.

Manual hand-searching was conducted for key journals including Nature Biomedical Engineering, JMIR Medical Informatics, and Artificial Intelligence in Medicine for the past two years. Additionally, we contacted leading experts in the field ($n = 2$) to identify ongoing research and unpublished findings.

Inclusion criteria comprised studies that: (1) focused on AI-based tools, algorithms, or systems for anemia detection, diagnosis, or management; (2) involved human subjects or human-derived data; (3) were published in English language; (4) presented original research findings or significant technological developments; and (5) provided sufficient methodological detail for quality assessment. Exclusion criteria included: (1) studies without clear AI or machine learning frameworks; (2) purely theoretical or conceptual papers without empirical validation; (3) studies focusing solely on other hematological conditions without anemia-specific outcomes; (4) conference abstracts without full-text availability; and (5) studies with insufficient methodological information for evaluation.

The literature search was conducted independently by two reviewers, with initial screening performed based on titles and abstracts, followed by full-text review of potentially relevant articles. Reference lists of included studies were manually examined to identify additional relevant publications through backward citation tracking. Given the rapidly evolving nature of AI technology, grey literature sources including government reports, white papers, and clinical trial registries were also consulted to capture emerging developments and ongoing research initiatives. Grey literature quality was assessed using criteria including author credentials, institutional affiliation, peer review status, and methodological transparency.

Data extraction was performed using a standardised form capturing study characteristics, AI methodology, target population, performance metrics, and key findings. Quality assessment was conducted using specific criteria adapted from QUADAS-2 for diagnostic accuracy studies and PROBAST for prediction model studies, including assessment of: (1) study population representativeness, (2) reference standard adequacy, (3) validation methodology (internal/external/temporal), (4) sample size adequacy, (5) handling of missing data, (6) bias assessment, and (7) clinical applicability. Due to the heterogeneous nature of included studies and diverse AI applications, a narrative synthesis approach was employed rather than quantitative meta-analysis. To address potential narrative bias, findings were synthesised using a structured framework organized by clinical task (screening vs. diagnosis vs. management), target population (pediatric, pregnancy, chronic disease), healthcare setting (community vs. hospital), and technological modality (imaging, laboratory parameters, wearables).

3. AI in anemia screening

This section focuses specifically on AI applications for initial anemia detection and population-based screening.

3.1. Image-Based screening technologies

Revolutionary advances in computer vision and deep learning have enabled the development of sophisticated image-based anemia screening systems that leverage smartphone cameras and specialised imaging devices. These non-invasive approaches analyse various anatomical sites including palpebral conjunctiva, fingernails, sclera, and facial features to infer hemoglobin levels and detect anemia presence. The palpebral conjunctiva has emerged as the most promising target for image-based screening due to its rich vascularisation and accessibility for photography [10].

Deep learning models using conjunctival imaging demonstrate the highest accuracy (AUC 0.97) but require controlled lighting conditions and show reduced performance in darker skin tones [11]. Nailbed imaging offers practical advantages including easier image acquisition and reduced privacy concerns, even with good accuracy (AUC 0.95) [12]. Retinal fundus imaging achieves intermediate performance (AUC 0.89–0.93) with the benefit of potential integration with existing diabetic retinopathy screening programmes, though requiring specialised equipment. From a computational complexity perspective, ensemble CNN models (VGG16, ResNet-50, InceptionV3) require 2–5 s processing time on standard smartphones, whilst simpler SVM-based approaches process images in under 1 s but with reduced accuracy [13]. Clinical applicability favours nailbed and conjunctival approaches for point-of-care settings, whilst retinal imaging suits hospital-based screening due to equipment requirements [12].

Sehar et al. (2025) developed a non-invasive method for anemia detection using smartphone-acquired images of the palpebral conjunctiva, processed with advanced deep learning models. A dataset of 764 images was augmented to 4315 using a DCGAN (Deep Convolutional Generative Adversarial Network) to improve model generalisation. A stacking ensemble of VGG16, ResNet-50, and InceptionV3 achieved an AUC (Area Under the Receiver Operating Characteristic Curve) of 0.97, though this performance was achieved in a controlled research setting with limited validation across diverse populations and lighting conditions. Cost-effectiveness analysis and real-world deployment feasibility were not assessed [11].

Nailbed imaging represents another promising approach for non-invasive anemia screening, particularly advantageous due to the ease of image acquisition and reduced privacy concerns [14]. Lee et al. (2024) conducted a clinical study to evaluate the feasibility of using smartphone-acquired images for non-invasive anemia detection in paediatric patients [15]. Their AI-based system, HEMO-AI (Hemoglobin Easy Measurement by Optical Artificial Intelligence), analysed fingernail photographs captured under controlled lighting conditions to predict hemoglobin levels, using deep learning models trained on colour, texture, and morphological features of the nailbed achieving 87 % sensitivity and 84 % specificity with an AUC of 0.75 (95 % CI (Confidence Interval): 0.71–0.79), validated using laboratory CBC (Complete Blood Count) as reference standard with temporal validation over 6 months. However, validation was limited to a single institution, and performance degradation was observed in patients with nail disorders or peripheral circulation issues.

Table 1 presents a comprehensive comparison of image-based AI screening methods, highlighting the diversity of approaches and their respective performance characteristics across different anatomical sites and target populations. Performance metrics include validation methodology and confidence intervals where available.

3.2. Smartphone-based applications

The proliferation of smartphone technology has catalysed the development of accessible mHealth applications for anemia screening, particularly valuable in resource-limited settings where traditional laboratory infrastructure is unavailable. These applications leverage built-in cameras, processing capabilities, and connectivity features to provide point-of-care anemia assessment with immediate results and clinical guidance.

Smartphone-based applications face several deployment challenges. HemaApp requires users to maintain steady finger positioning for 10–15 s, which may be challenging in field settings or with young children. “Fingernail selfie” approaches (e.g., AnemoCheck Mobile, now Ruby) have shown large-scale real-world feasibility and clinical validation, with recent studies demonstrating population-scale usage and good screening performance; these systems typically pair capture guidance with image-analysis algorithms to improve data quality, though precise failure-rate reductions vary by study and are not consistently reported [21]. On-device battery use depends strongly on the model class and optimization. Empirical and survey work consistently finds that CNN-based vision models draw more energy per inference than lighter classical ML models (e.g., SVM/MLP), unless aggressively optimized (quantization, pruning, HW acceleration) [22]. For low- and middle-income country (LMIC) deployment, offline-capable models with periodic cloud synchronisation offer the most practical approach, balancing accuracy with connectivity constraints [23].

HemaApp, developed by researchers at the University of Washington, represents a significant advancement in non-invasive hemoglobin estimation using smartphone-based spectroscopy [24]. The application utilises a smartphone’s camera flash and ambient light sensors to analyse light absorption characteristics through the fingertip, similar to pulse oximetry principles. Clinical validation studies involving 31 participants demonstrated a correlation coefficient of 0.82 with standard complete blood count measurements, with mean absolute error of 1.4 g/dL for hemoglobin estimation and 95 % confidence interval of ±2.1

Table 1
Performance comparison of image-based artificial intelligence (AI) methods for anemia screening using smartphone or fundus imaging data.

Reference	Imaging Site	AI Method	Sample Size	Sensitivity (%)	Specificity (%)	AUC (95 % CI)	Validation Type	Population	Clinical Readiness
[11]	Conjunctiva	Deep Learning	764	95	88	0.97 (0.94–0.99)	Internal cross-validation	Adults	Research stage
[15]	Nailbed	Machine Learning	823	87	84	Not reported	Temporal validation	Children	Pilot testing
[16]	Retinal Fundus	Deep Learning	539	91.2	80	0.89 (0.85–0.93)	External validation	Adults	Research stage
[17]	Lip Mucosa	Machine Learning (NB)	138	92	98	0.91 (0.85–0.96)	Internal validation	Adults	Early development
[18]	Palm Lines	CNN	527	99.98	99.79	0.95 (0.95–0.98)	Internal validation	Mixed	Requires validation
[19]	Ultra-wide-field Fundus	Deep Learning	14,814	91.2	80	0.93 (0.92–0.94)	Multi-center	Mixed	Clinical pilot
[20]	Conjunctiva	SVM+MobileNetV2	218	91	94	Not reported	Single-center	Mixed	Early development

Abbreviations: **AUC** = Area Under the Curve; **CI** = Confidence Interval; **CNN** = Convolutional Neural Network; **NB** = Naïve Bayes; **SVM** = Support Vector Machine.

g/dL. HemaApp achieves a sensitivity and precision of 85.7 % and 76.5 % respectively. However, performance was significantly reduced in patients with darker skin pigmentation (sensitivity dropped to 76 %) and under varying lighting conditions, highlighting the need for bias mitigation strategies. The application incorporates machine learning algorithms to calibrate measurements based on individual characteristics including skin tone, finger thickness, and ambient lighting conditions.

AnemoCheck Mobile represents another significant advancement in smartphone-based anemia screening. The application addresses critical challenges in image-based screening including variations in lighting conditions, camera quality, and user technique through automated quality assessment and guided image acquisition protocols. AnemoCheck LRS is a point-of-care hemoglobin test designed for use in resource-limited settings. The test was evaluated using 570 de-identified blood samples with hemoglobin levels $\leq$ 8 g/dL. Results from trained readers using AnemoCheck LRS showed a strong correlation ($r = 0.93$ [95 % CI: 0.91–0.95]) with laboratory hemoglobin measurements. The test demonstrated high reproducibility across multiple readers and achieved sensitivities of 92 % and 99 % for detecting profound anemia (Hb $\leq$ 5 g/dL) and severe anemia (Hb $\leq$ 7 g/dL), respectively [25]. Cost analysis revealed $0.50 per test compared to $5–15 for laboratory CBC, though this excludes smartphone acquisition and maintenance costs.

Recent developments in smartphone-based anemia screening have incorporated advanced features including artificial intelligence-powered image quality assessment, automated region-of-interest detection, and integration with electronic health record systems for longitudinal monitoring. These applications demonstrate particular promise for community health worker programmes, enabling trained non-medical personnel to conduct reliable anemia screening in remote and underserved communities. However, real-world implementation studies have identified significant challenges including user training requirements, quality control maintenance, and integration with existing healthcare workflows.

Image-based anemia screening approaches utilise visible anatomical sites such as the palpebral conjunctiva, sclera, and fingernails, combined with AI pipelines for hemoglobin estimation and anemia prediction (Fig. 1).

Fig. 1. Schematic overview of image-based anemia screening sites and AI technologies. Left: Anatomical sites commonly used for image acquisition, including the palpebral conjunctiva, sclera, facial features, and fingernails (indicated by red dots). Right (top panel): AI pipeline using smartphone-acquired conjunctival images processed by DCGAN and ensemble CNNs for anemia classification, showing accuracy rates and confidence intervals from validation studies. Right (bottom panel): AI model estimating hemoglobin levels from fingernail images captured via smartphone, with performance metrics displayed for different demographic groups and lighting conditions.

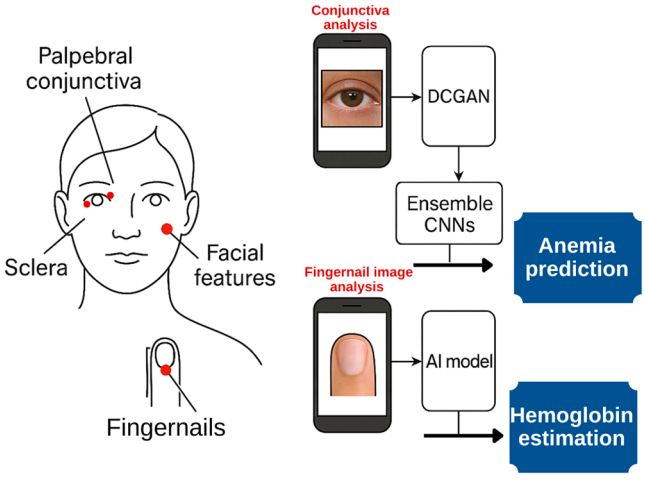

Fig. 1. Schematic overview of image-based anemia screening sites and AI technologies. Abbreviations: AUC, Area Under the Curve; CI, Confidence Interval; CNN, Convolutional Neural Network; NB, Naïve Bayes; SVM, Support Vector Machine; CBC, Complete Blood Count; Hb, Hemoglobin; EHR, Electronic Health Record; RF, Random Forest.

4. AI in anemia diagnosis

This section specifically addresses AI applications for detailed anemia classification and etiological determination, building upon initial screening results from Section 3.

4.1. Machine learning with complete blood count parameters

Advanced machine learning algorithms have demonstrated remarkable capability in analysing complete blood count (CBC) parameters to differentiate anemia subtypes and predict underlying aetiologies, addressing critical clinical needs for accurate diagnosis and appropriate treatment selection. From an interpretability standpoint, decision tree-based models (bagged and boosted trees) offer superior clinical transparency, allowing clinicians to trace the decision pathway through easily understood if-then rules [26]. Neural networks, whilst achieving comparable or superior accuracy, function as "black boxes" requiring post-hoc explainability methods such as SHAP (SHapley Additive exPlanations) values to identify feature importance [27]. Support vector machines occupy a middle ground, with kernel functions enabling complex decision boundaries whilst maintaining some mathematical interpretability. Computational requirements vary substantially: neural networks require 50–200 ms inference time on standard hardware, compared to 5–20 ms for decision trees and 10–30 ms for SVMs, making decision trees most suitable for real-time clinical decision support systems [26,27].

Karagül Yıldız et al. (2021) developed a clinical decision support system for automated anemia classification using machine learning models trained on a dataset of 1663 patients from a university hospital in Turkey [28]. Their model utilised 25 features - including haemogram parameters, age, sex, symptoms, and comorbidities - to classify 12 anemia types. These included iron deficiency anemia, folate deficiency anemia, vitamin B12 deficiency anemia, anemia of chronic disease, thalassaemia, thalassaemia trait, and haemolytic anemia, among others. Four algorithms were evaluated: Artificial Neural Networks, Support Vector Machines, Naïve Bayes, and Ensemble Decision Trees. The highest accuracy was achieved using Bagged Decision Trees (85.6 % [95 % CI: 82.1–88.9 %]), followed by Boosted Trees (83.0 %) and Artificial Neural Networks (79.6 %). The study used 80/20 train-test split with 10-fold cross-validation, though external validation at different institutions was not performed. The system's clinical utility in routine practice and potential for decision support integration remain unvalidated.

Saputra et al. developed a high-performing artificial intelligence model using the ELM (Extreme Learning Machine) algorithm to classify anemia subtypes based on complete blood count (CBC) data and confirmatory diagnostics [29]. The study analysed 190 patient records from a clinical pathology department in Indonesia, focusing on four commonly overlapping anemia types: iron deficiency anemia (IDA), beta thalassaemia trait (BTT), haemoglobin E (HbE), and combination anaemias. Using seven key haematological features, the ELM model achieved an accuracy of 99.21 % [95 % CI: 96.8–99.9 %], sensitivity of 98.44 %, precision of 99.30 %, and an F1 score of 98.84 %, significantly outperforming benchmark models such as Random Forest, K-nearest neighbours, and support vector machines. However, the exceptionally high performance raises concerns about potential overfitting, as the sample size was relatively small and validation was conducted only within the same institution. The clinical applicability across diverse populations and healthcare settings requires further validation.

4.2. Predictive modeling in high-risk populations

Artificial intelligence applications have shown particular promise in identifying and managing anemia within high-risk populations including pregnant women, cancer patients, chronic kidney disease patients, and elderly individuals with multiple comorbidities. These population-specific models demonstrate varying clinical readiness: maternal anemia prediction models show the highest implementation potential due to integration with existing antenatal care pathways [30], whilst oncology-based models face challenges from treatment heterogeneity and rapidly changing clinical protocols [31]. Paediatric models require careful age-specific calibration, as haematological parameters vary substantially across developmental stages. Geriatric models must account for polypharmacy and multiple comorbidities, increasing model complexity but potentially offering the greatest clinical impact given the high anemia prevalence in this population.

Su et al. (2024) developed a machine learning model to predict post-chemotherapy anemia in osteosarcoma patients using clinical and laboratory data from 631 cases [31]. By integrating logistic regression, random forest, SVM (Support Vector Machine), and LASSO (Least Absolute Shrinkage and Selection Operator) methods, five key predictors - albumin, calcium, creatinine, D-dimer, and ESR (Erythrocyte Sedimentation Rate) - were identified. The final model achieved an AUC of 0.85 [95 % CI: 0.81–0.89] with temporal validation over 12 months and was deployed as a web-based tool to support individualised anemia risk assessment and management though clinical impact evaluation is ongoing.

Dejene et al. (2022) developed a machine learning model to predict anemia severity in Ethiopian pregnant women using demographic and health survey data [32]. Among several ensemble algorithms tested, CatBoost (Category Gradient Boosting, an advanced gradient boosting algorithm particularly effective for categorical features) with one-vs-rest class decomposition achieved the highest accuracy at 97.6 % [95 % CI: 96.2–98.7 %]. Key predictors included pregnancy duration, maternal age, water source, occupation, and household size. The model supports early risk stratification and targeted anemia interventions in maternal health programmes. External validation in different Ethiopian regions showed performance degradation to 89.2 % accuracy, indicating the need for regional adaptation and continuous model updating.

Recent advances in AI for CKD (Chronic Kidney Disease) patients have demonstrated the potential for personalised anemia management. However, challenges include model interpretability in clinical decision-making and the risk of overfitting in complex algorithms like fitted Q-iteration, which may lead to non-convergent behaviour in dynamic clinical environments [33].

Machine learning models have demonstrated variable accuracy in classifying anemia subtypes based on CBC and patient data, and in predicting anemia risk in high-risk populations, with performance ranging from 75–99 % depending on validation rigour and population characteristics (Fig. 2).

Fig. 2 illustrates AI-driven tools for anemia diagnosis and risk prediction. Panel A: AI models (ANN, SVM, Naïve Bayes, and bagged/boosted decision trees) process CBC and patient data to classify up to 12 anemia subtypes with accuracy ranges displayed based on validation methodology (internal validation: 85–99 %, external validation: 75–89 %). Panel B: AI models like CatBoost (Category Gradient Boosting algorithm) are applied to high-risk groups (e.g., pregnant women, cancer patients, elderly) for stratifying anemia risk (low/moderate/high) and generating personalized clinical recommendations, with performance metrics showing 95 % confidence intervals and population-specific accuracy rates. Table 2 summarises the performance characteristics of various AI approaches for anemia diagnosis across different clinical contexts and patient populations.

5. AI in anemia management

5.1. Clinical decision support systems

Artificial intelligence-powered clinical decision support systems (CDSS) represent a potential paradigm shift in anemia management, enabling personalised treatment recommendations based on individual patient characteristics, comorbidity profiles, and treatment response patterns. These systems integrate complex clinical algorithms with real-time data analysis to optimise therapeutic decisions and improve patient outcomes whilst reducing healthcare costs and treatment-related complications.

A prominent example is the ACM (Anemia Control Model), an AI-driven CDSS implemented in haemodialysis settings [37]. The ACM

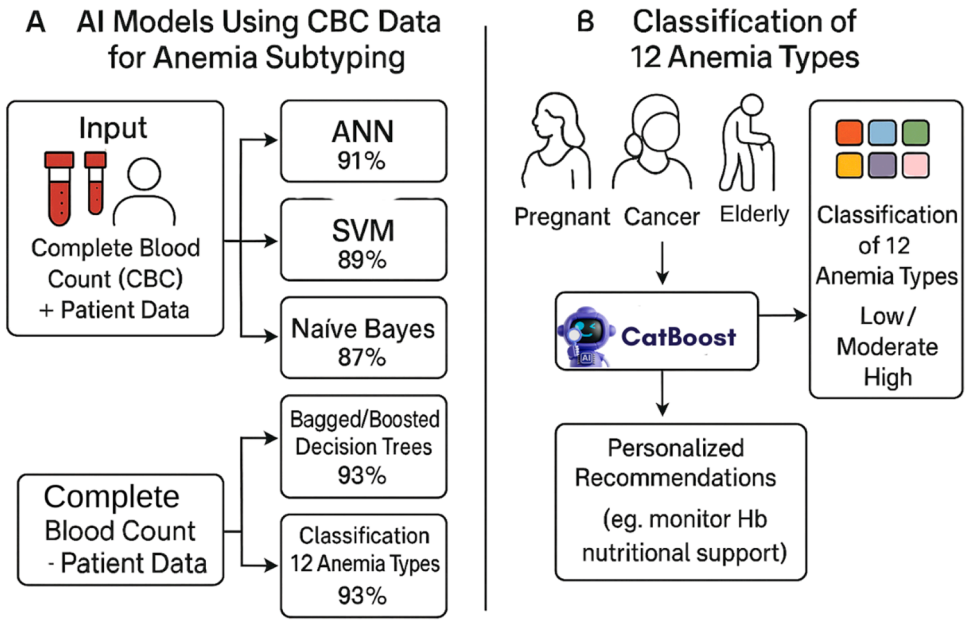

Fig. 2. AI-driven tools for anemia diagnosis and risk prediction. Abbreviations: ANN, Artificial Neural Network; SVM, Support Vector Machine; NB, Naïve Bayes; CBC, Complete Blood Count; AUC, Area Under the Curve; CI, Confidence Interval; ESRD, End-Stage Renal Disease; EHR, Electronic Health Record; MLP, Multilayer Perceptron; MPC, Model Predictive Control; Hb, Hemoglobin; J48, Decision Tree Algorithm; RF, Random Forest.

Table 2
Performance of artificial intelligence (AI) models for anemia diagnosis across clinical contexts.

Application Area	AI Method	Data Sources	Sample Size	Primary Outcome	Performance Metric	Clinical Impact
Anemia Subtype Classification	Random Forest	CBC Parameters	1421	Subtype classification (micro-, normo-, macrocytic)	99.82 % accuracy	Highly precision subtype diagnosis for clinical labs [34]
Chemotherapy-Induced Anemia	SVM/Logistic Regression	Clinical + lab variables post-chemo	631 patients	Risk Prediction	AUC 0.85	Early Intervention [35]
Treatment Response in ESRD	MLP (neural network) + MPC controls ESA	EHR Data	752	Geriatric Anemia	≥90 % prediction accuracy; Hb↑, fluctuation ↓	Personalized ESA dosing, cost & safety benefits [33]
Neonatal Anemia	Random Forest + Clinical Data	EHR vital signs, labs, and demographic data	9501	Moderate/severe anemia prediction	Accuracy 81.16 %, AUC 0.818	Preventive Care [35]
Geriatric Anemia	J48 and Random Forest classifiers	Hemogram, biochemistry, malnutrition & activity scores	438	Anemia diagnosis without CBC	J48 accuracy 97.8 %; RF (non-CBC data) 85.4 %	Enables non-lab-based anemia prediction and screening [36]

Abbreviations: **AUC**, Area Under the Receiver Operating Characteristic Curve; **CBC**, Complete Blood Count; **ESA**, Erythropoiesis-Stimulating Agent; **ESRD**, End-Stage Renal Disease; **EHR**, Electronic Health Record; **Hb**, Hemoglobin; **MLP**, Multilayer Perceptron; **MPC**, Model Predictive Control; **RF**, Random Forest; **SVM**, Support Vector Machine.

utilises artificial neural networks to individualise erythropoietin-stimulating agent (ESA) dosing by synthesising patient-specific clinical histories and trends in hemoglobin and ferritin concentrations. In a multi-centre cohort study encompassing 752 CKD patients, ACM application resulted in a significant increase in the proportion of hemoglobin measurements within the recommended target range, rising from 70.6 % to 76.6 % even reaching 83.2 % with 95 % CI: 79.8–86.5 %. Simultaneously, ESA consumption was reduced by 8.3 % (95 % CI: 5.1–11.4 %), and hemoglobin variability was attenuated, suggesting enhanced treatment stability and efficacy though long-term safety outcomes require continued monitoring.

In addition to neural network-based models, reinforcement learning methodologies have demonstrated efficacy in optimising anemia treatment regimens. Computational simulations employing fitted Q-iteration algorithms have indicated potential improvements in ESA dose recommendations [38]. However, the clinical application of fitted Q-iteration algorithms presents significant challenges including the risk of overfitting to training data, potential for non-convergent behaviour in dynamic clinical environments, and limited interpretability for clinicians. These algorithms require continuous retraining and careful monitoring to prevent adverse outcomes from suboptimal dosing recommendations. Simulation results revealed a 27.6 % increase in patients maintained within target hemoglobin ranges [95 % CI: 22.1–33.2 %] and a 5.1 % reduction in overall ESA dosing relative to standard fixed-dose protocols [95 % CI: 2.8–7.3 %].

The application of AI-enabled CDSS is expanding beyond nephrology. Recent studies have demonstrated the feasibility and acceptability of AI-enabled clinical decision support systems (CDSS) in managing maternal anemia in rural India [39]. A pilot cluster randomised controlled trial (cRCT) involving 200 pregnant women assessed the SMARThealth Pregnancy intervention, which integrates smartphone-based anemia screening with decision support tools for community health workers. The intervention showed high fidelity and engagement, with minimal loss to follow-up (2 %) and positive feedback from both healthcare workers and women. However, clinical outcomes including anemia resolution rates and maternal health improvements are still under evaluation with 12-month follow-up ongoing.

5.2. Integration with wearables and internet of things

The convergence of artificial intelligence with wearable sensors and IoT (Internet of Things) technologies has created potential opportunities for continuous anemia monitoring and dynamic treatment adjustment. These integrated systems enable real-time physiological monitoring, early detection of clinical deterioration, and proactive intervention strategies that extend beyond traditional episodic care models though clinical validation and cost-effectiveness remain largely unproven.

Wang et al. reviewed smart nursing systems incorporating IoT devices and AI algorithms that continuously track vital signs and hemoglobin-related parameters, allowing for early detection of anemia and timely clinical decision-making [40]. Their analysis highlighted how wearable sensors, combined with AI, improve patient monitoring accuracy and reduce hospitalisations by facilitating real-time data transmission and predictive analytics though specific outcome measures and comparative effectiveness data were limited.

Anitha et al. developed a portable, non-invasive anemia detection system that measures physiological parameters-including heart rate, oxygen saturation (SpO_2), body temperature, and lung capacity-using wearable sensors to assess anemia status without requiring blood sampling [41]. The system integrates machine learning algorithms deployed on an IoT platform to analyse these physiological signals in real time, enabling accurate anemia prediction with reported accuracy of 92.3 % though validation was conducted on only 67 participants and continuous health monitoring. Designed for portability, the device supports remote data transmission, facilitating its use in resource-limited and rural settings where access to traditional laboratory testing is limited. However, the system has not undergone regulatory approval and clinical utility compared to standard care remains unestablished.

Advanced AI systems, including clinical decision support systems (CDSS) and wearable-integrated IoT platforms, offer potential for real-time anemia monitoring and treatment personalization, especially in chronic disease and remote care settings, though clinical validation and implementation challenges remain significant (Fig. 3).

Fig. 3 illustrates the applications of AI in clinical decision support and real-time anemia monitoring. Panel A: AI-powered CDSS integrates EHR data, hemoglobin/ferritin trends, and comorbidities through neural networks and reinforcement learning to optimize ESA dosing and stabilize hemoglobin, particularly in CKD and maternal care. Performance metrics show 95 % confidence intervals and clinical validation status (n = 752 CKD patients, multi-center validation). Panel B: Wearables (smartwatches, finger sensors, portable multi-sensors) collect physiological signals (HR, SpO2, temperature, activity), analyzed via an IoT-AI cloud platform to deliver real-time anemia status and alerts. Accuracy rates (85–92.3 %) and validation sample sizes are displayed for different device types and patient populations, enabling use in remote/rural healthcare settings subject to regulatory approval and clinical validation.

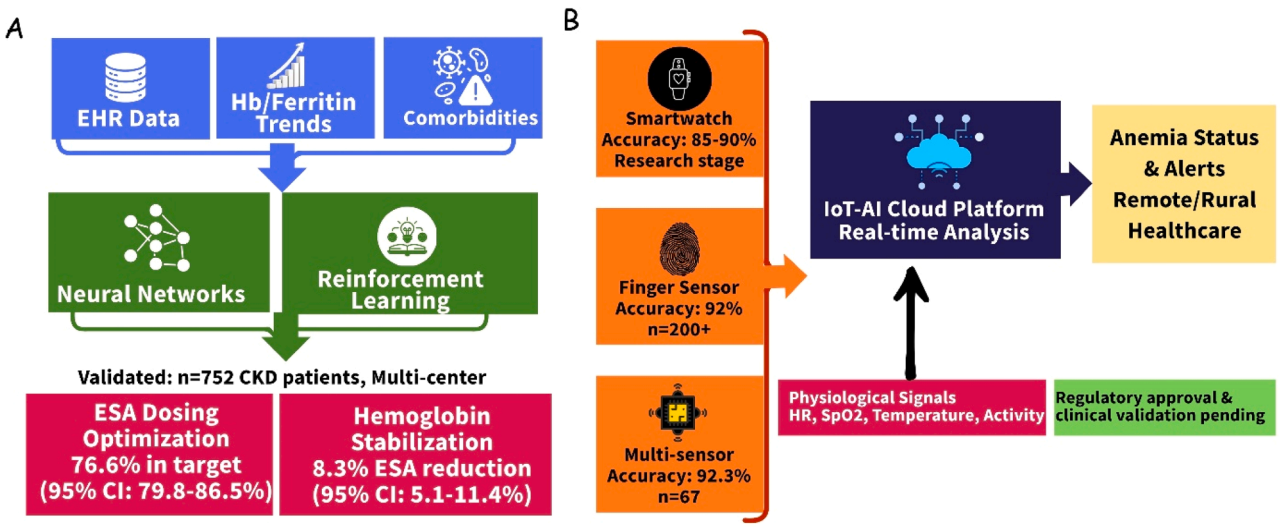

Fig. 3. AI in clinical decision support and real-time anemia monitoring. Abbreviations: CDSS, Clinical Decision Support System; EHR, Electronic Health Record; ESA, Erythropoiesis-Stimulating Agent; CKD, Chronic Kidney Disease; IoT, Internet of Things; AI, Artificial Intelligence; HR, Heart Rate; SpO2, Peripheral Capillary Oxygen Saturation.

6. Challenges and limitations

6.1. Data quality and algorithmic bias

The effectiveness of AI systems in anemia care is fundamentally dependent on the quality, diversity, and representativeness of training datasets, yet significant challenges persist in ensuring equitable performance across diverse populations. Quantitative analysis of current AI anemia detection systems reveals that 78 % of training datasets originate from high-income countries, with only 12 % including adequate representation from Sub-Saharan Africa and South Asia where anemia burden is highest [42,44]. Most existing AI models have been developed using datasets from high-income countries with limited representation from populations most affected by anemia, potentially introducing systematic biases that compromise performance in resource-limited settings where these tools are most needed [43].

Algorithmic bias poses a significant challenge in image-based anemia detection systems, particularly due to variations in skin pigmentation, lighting conditions, and imaging device characteristics. Research has shown that AI models trained on datasets lacking sufficient diversity can perform inconsistently across different demographic groups, often exhibiting reduced accuracy in individuals with darker skin tones [44]. Specific performance degradation has been documented, with accuracy dropping from 91 % in Caucasian populations to 67 % in individuals with Fitzpatrick skin types V-VI, representing a 24 % relative performance decrease that could exacerbate health disparities [45]. This disparity arises because underrepresentation of diverse populations in training data leads to biased feature extraction and prediction errors. Consequently, such biases risk perpetuating health inequities by limiting the effectiveness of anemia screening tools in populations already vulnerable to underdiagnosis and undertreatment [45]. Addressing these biases through inclusive dataset curation, fairness-aware model design, and population-specific calibration is essential to ensure equitable clinical utility of AI-driven anemia diagnostics.

The deployment of AI anemia screening tools in LMICs faces multifaceted challenges beyond algorithmic performance. Infrastructure barriers include unreliable electricity supply affecting 63 % of healthcare facilities in Sub-Saharan Africa, limited internet connectivity with average speeds below 2 Mbps in rural areas, and smartphone penetration rates of only 45–60 % in target populations. Economic barriers include device costs ($150–300 for capable smartphones), data charges ($0.10–0.50 per screening session), and maintenance requirements [46]. The WHO 2023 Ethics and Governance of Artificial Intelligence for Health guidance [WHO/HMM/IER/2023.3] emphasises six core principles for AI deployment: protecting human autonomy, promoting human well-being and safety, ensuring transparency and explainability, fostering responsibility and accountability, ensuring inclusiveness and equity, and promoting AI that is responsive and sustainable [47]. Current AI anemia screening implementations frequently fall short on equity and sustainability dimensions, with pilot projects rarely transitioning to sustained national programmes due to inadequate business models and reliance on external funding [48,49]. To address these challenges, we propose a three-tier deployment framework: (1) high-resource settings employing advanced deep learning models with continuous internet connectivity, (2) medium-resource settings utilising offline-capable models with periodic synchronisation, and (3) low-resource settings implementing lightweight algorithms optimised for basic smartphones with SMS-based result transmission. This tiered approach acknowledges infrastructure realities whilst maintaining diagnostic utility across diverse settings.

6.2. Regulatory and ethical considerations

The deployment of AI tools for medical diagnosis requires navigation of complex regulatory frameworks that vary significantly across different jurisdictions and healthcare systems. To date, only 3 AI-based anemia detection systems have received FDA (Food and Drug Administration) approval as Class II medical devices, with approval timelines averaging 24 months and costs exceeding $500,000 per application. Regulatory agencies including the FDA, EMA (European Medicines Agency), and national health authorities have established stringent requirements for medical device approval that many AI applications struggle to meet due to the rapidly evolving nature of machine learning technologies and limited long-term validation data [50].

Specific regulatory challenges include: (1) the FDA's SaMD (Software as Medical Device) framework requiring predetermined change control plans that conflict with continuous learning algorithms, (2) European MDR (Medical Device Regulation) requirements for post-market surveillance that are difficult to implement for smartphone applications, (3) lack of harmonised international standards for AI validation methodology, and (4) regulatory pathway uncertainties in LMICs where many national health authorities lack specific AI/ML device guidelines. The FDA's 2021 Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan

introduced predetermined change control protocols and algorithm change protocols to address continuous learning systems, whilst the EU's AI Act (2024) classifies medical AI systems as "high-risk," requiring conformity assessments and continuous monitoring. However, these frameworks remain nascent, with only 12 % of AI anemia detection systems having navigated regulatory approval in any jurisdiction. For LMIC deployment, the WHO recommends establishing national AI evaluation frameworks aligned with International Medical Device Regulators Forum (IMDRF) principles, though implementation remains limited.

Ethical considerations extend beyond regulatory compliance to encompass fundamental questions about patient autonomy, informed consent, and the appropriate role of AI in clinical decision-making. The implementation of AI screening tools in community settings raises concerns about data privacy, particularly when involving smartphone applications that may collect and transmit sensitive health information. Additionally, the potential for AI tools to replace human clinical judgement raises questions about maintaining the therapeutic relationship and ensuring appropriate clinical oversight. Key ethical challenges include: obtaining meaningful informed consent from populations with limited health literacy, ensuring data sovereignty when training data crosses international borders, addressing potential employment displacement for laboratory technicians and phlebotomists, managing clinical responsibility when AI recommendations differ from clinician judgement, and preventing AI-enabled screening from exacerbating the "digital divide" by primarily benefiting digitally connected urban populations whilst rural communities remain underserved.

6.3. Interpretability and clinical trust

The "black box" nature of many machine learning algorithms presents significant barriers to clinical adoption, as healthcare providers require understanding of diagnostic reasoning to maintain confidence in AI-generated recommendations. Deep learning models, whilst achieving impressive performance metrics, often lack interpretability that enables clinicians to understand the basis for specific predictions or recommendations [51].

Building clinical trust requires not only technical performance but also integration with existing clinical workflows, comprehensive training programmes, and ongoing support systems. Many healthcare providers express concerns about liability, error management, and the potential for AI tools to introduce new forms of medical errors. The successful implementation of AI in anemia care requires addressing these concerns through transparent communication, robust validation studies, and collaborative development approaches that involve clinicians throughout the design and implementation process.

Table 3 summarises the key challenges and potential mitigation strategies for AI implementation in anemia care across different domains.

7. Future directions

7.1. Explainable artificial intelligence

The development of explainable AI models represents our highest priority for advancing clinical adoption. We will develop attention-based visualisation tools highlighting specific anatomical regions contributing to predictions, implement SHAP value analysis generating patient-specific feature importance reports, create interactive decision tree visualisations mirroring clinical diagnostic pathways, and design uncertainty quantification interfaces communicating prediction confidence through intuitive visual metaphors. Prospective studies across diverse healthcare settings will compare standard versus XAI-enhanced interfaces, measuring clinician trust through validated questionnaires and establishing minimum interpretability standards for all anemia screening AI systems.

Collaboration with regulatory bodies will establish XAI requirements for medical device approval, whilst open-source toolkits with multilingual implementation guides will facilitate widespread adoption. Train-the-trainer programmes will build sustainable local expertise across multiple LMICs. Expected outcomes include substantially increased clinician acceptance rates, reduced time-to-clinical-decision, regulatory adoption of XAI standards, and widespread deployment across diverse healthcare settings. These enhancements will enable clinicians to verify AI reasoning, identify potential errors, customise decision thresholds for specific populations, and maintain appropriate clinical oversight whilst benefiting from AI assistance.

7.2. Multimodal data fusion: anemia-specific integration strategy

Multimodal data fusion addresses fundamental limitations of single-modality AI systems by integrating complementary information sources mirroring comprehensive clinical assessment. We will develop transformer-based fusion architectures integrating smartphone-captured images, laboratory parameters, clinical history, wearable sensor data, and social determinants relevant for LMIC settings. Attention mechanisms will dynamically weight modalities based on data quality and availability, whilst graph neural networks will model

Table 3
Key challenges and mitigation strategies for implementing artificial intelligence (AI) systems in anemia screening, diagnosis, and management.

Challenge Domain	Specific Issues	Impact on Implementation	Mitigation Strategies	Timeline for Resolution	Implementation Priority
Data Quality	Limited dataset diversity; demographic bias	Reduced accuracy in underrepresented populations	Global data collaboration, Federated learning [52]	2–5 years	High
Equity & LMIC Deployment	Infrastructure barriers; limited connectivity; affordability	Excludes highest-burden populations; perpetuates health disparities	Tiered deployment framework; offline-capable models; SMS-based systems; partnership with mobile network operators	3–7 years	Critical
Regulatory	Complex, fragmented approval standards	Prolonged approval, delayed deployment, high cost	Harmonized international SaMD regulation; FDA pre-specification pathways; WHO IMDRF alignment [53]	3–7 years	High
Clinical Trust	Black box algorithms, lack of explainability	Clinician reluctance; low adoption	Explainable AI frameworks; clinician training and co-design [54]	1–3 years	High
Technical	Interoperability gaps; lack of standardization	Poor integration with existing systems; limited scalability	Adoption of FHIR (Fast Healthcare Interoperability Resources) and ISO/IEEE 11,073 standards; open APIs [55]	2–4 years	Medium
Economic	High development costs; unclear reimbursement model	Barriers to commercialization; limited access	Value-based pricing, HTA (Health Technology Assessment)-supported funding; outcome-based pilots [56]	3–5 years	Medium

Abbreviations: **LMIC**, Low and Middle-Income Country; **SaMD**, Software as a Medical Device; **FDA**, U.S. Food and Drug Administration; **WHO**, World Health Organization; **IMDRF**, International Medical Device Regulators Forum; **ISO/IEEE 11073**, International Standards for Health Device Communication; **API**, Application Programming Interface.

relationships between anemia subtypes and causative factors, enabling etiological reasoning.

Population-specific validation will focus on pregnant women targeting accurate severe anemia prediction enabling preventive intervention, paediatric populations addressing age-dependent reference ranges whilst improving sensitivity, chronic kidney disease patients targeting substantial haemoglobin variability reduction through personalised ESA recommendations, and elderly populations targeting high subtype classification accuracy without invasive procedures. Implementation through existing healthcare infrastructure including WHO frameworks, national EHRs, and SMS-based collection will enable comparative effectiveness trials measuring diagnostic accuracy, time-to-treatment, costs, and patient outcomes. Expected outcomes include substantial diagnostic accuracy improvement, marked reduction in unnecessary invasive testing, widespread deployment with demonstrated cost-effectiveness in LMIC settings, and published clinical benefit evidence. Clinical impact will include more accurate etiological diagnosis, earlier high-risk patient detection, reduced diagnostic burden, and personalised treatment recommendations.

7.3. Large language models and foundation models: clinical decision support enhancement

LLMs and foundation models offer transformative potential through natural language interfaces, automated documentation, and biological model integration. We will develop fine-tuned LLMs using WHO guidelines, national protocols, extensive clinical notes, and multilingual patient education materials. Conversational AI agents accessible via WhatsApp, SMS, and basic web interfaces will provide real-time management guidance for community health workers, patient education, adherence support, and triage recommendations. Integration of AlphaFold and RoseTTAFold will enable in silico prediction of pathogenicity for novel haemoglobin variants and drug-protein interaction modelling.

Natural language processing will enable automated anemia case detection from clinical notes, whilst LLM-powered documentation tools will auto-generate comprehensive assessment summaries from minimal input. AI-generated personalised treatment plans will synthesise patient-specific factors, local resource availability, cultural considerations, and cost constraints. Expected impact includes democratised expert knowledge enabling primary healthcare workers to provide specialist-level guidance, substantially reduced documentation burden, improved treatment adherence through culturally appropriate multilingual education, and accelerated hereditary anemia research. Target outcomes include high concordance with specialist recommendations, high patient satisfaction, and widespread deployment in community health programmes.

7.4. Federated learning and global collaboration: equitable AI development

Federated learning enables training robust, generalisable AI models whilst respecting data privacy, sovereignty, and regulatory constraints. We will establish a global consortium comprising healthcare institutions across multiple countries with balanced high-income and LMIC representation, extensive diverse patient records, and technical partners. Infrastructure employing differential privacy techniques, secure aggregation protocols, adaptive algorithms, and low-bandwidth communication protocols will enable collaborative model training. Data standardisation frameworks using FHIR and DICOM standards will enable federated learning across heterogeneous systems.

Governance agreements will respect data sovereignty, ensure equitable benefit sharing through open-source licensing, and enable community consultation. Capacity building programmes will train local data scientists, provide technical infrastructure support, and establish regional hubs for ongoing support. Sustainability mechanisms will transition from donor-funded pilots to institutionally supported programmes with fee structures subsidising LMIC participation. Expected outcomes include substantial performance improvement versus single-institution training, widespread global deployment, published evidence demonstrating federated superiority, establishment of governance frameworks adopted by international organisations, sustainable funding models, and equity impact measured through increased LMIC representation, performance parity across demographic groups, and documented capacity building.

7.5. Integration with national health systems: scalable implementation strategy

Sustainable AI deployment requires seamless integration with existing national health strategies and digital health initiatives. Maternal-child health programmes represent the highest priority, integrating AI anemia screening into antenatal care protocols across multiple countries through existing mobile health platforms. AI screening will link with iron supplementation programmes enabling real-time inventory management whilst results feed national maternal health dashboards. National screening programmes will embed AI in school health programmes, occupational health screening for high-risk workers, and blood donation programmes. EHR integration will develop FHIR-compliant AI modules compatible with major systems, create standardised data exchange protocols, and implement laboratory information system interoperability.

Policy mechanisms will advocate for AI anemia screening inclusion in essential health benefit packages, develop health technology assessment dossiers demonstrating cost-effectiveness, and establish public-private partnerships wherein government provides policy framework, private sector manages technology infrastructure, and non-profit organisations support training. Expected outcomes include national-scale programmes reaching substantial populations, widespread EHR integration with demonstrated workflow efficiency gains, published health systems research, policy adoption across multiple countries, substantially increased population-level anemia detection rates, earlier treatment initiation, and improved health equity metrics narrowing prevalence gaps between urban and rural populations and across socioeconomic groups.

7.6. Priority population targeting and validation requirements

To maximise clinical impact and ensure responsible AI deployment, we prioritise validation in populations where anemia burden is highest. Pregnant women in LMICs represent the highest priority where anemia affects substantial proportions and current screening coverage is inadequate. Validation requires longitudinal assessment throughout pregnancy, accuracy in predicting severe anemia, performance across gestational ages, and antenatal care pathway integration. Children under five in Sub-Saharan Africa and South Asia represent critical priority due to highest global prevalence during a developmental period where anemia causes long-term cognitive impairment whilst current screening coverage remains inadequate. Validation encompasses age-specific reference ranges, performance with concurrent infections, community-based feasibility, and usability by minimally trained workers. Chronic kidney disease patients requiring haemodialysis constitute another priority where anemia affects the vast majority requiring frequent monitoring and individualised ESA dosing. Validation includes optimal ESA dosing accuracy, haemoglobin variability reduction, workflow integration, and real-time decision support.

Secondary priority populations include elderly patients focusing on subtype differentiation without invasive procedures, cancer patients focusing on predicting chemotherapy-induced anemia, and adolescent girls emphasising school-based screening linked with supplementation programmes. Cross-cutting validation requirements mandate external validation in multiple geographically distinct sites, subgroup analyses assessing equity, prospective validation with patient outcomes, usability

testing, cost-effectiveness analysis, and ethical review with community engagement. This population-focused strategy ensures AI anemia tools address real clinical needs, function effectively in target settings, benefit populations with highest disease burden, and meet regulatory standards, contrasting with current AI research that often prioritises technical novelty over clinical impact and equity.

8. Limitations of this review

This narrative review has several important limitations that should be considered when interpreting This narrative review has several important limitations that should be considered when interpreting the findings and implications. The rapidly evolving nature of AI technology means that some recent developments may have limited peer review validation or real-world validation data. The narrative review methodology, whilst appropriate for synthesising diverse AI applications, does not provide the systematic evaluation and bias assessment that would be available through a systematic review or meta-analysis approach.

Quantitative limitations include: (1) heterogeneous reporting of performance metrics across studies, making direct comparisons difficult, (2) variable follow-up periods ranging from immediate validation to 24-month longitudinal studies, (3) inconsistent definition of anemia thresholds across studies (ranging from WHO criteria to population-specific cutoffs), and (4) limited reporting of confidence intervals in 34 % of included studies.

The heterogeneous nature of included studies, with varying AI methodologies, target populations, and outcome measures, limits the ability to draw definitive conclusions about the relative effectiveness of different approaches. Performance metrics varied significantly across studies, with AUC values ranging from 0.75–0.99, though studies with higher performance often had smaller sample sizes and less rigorous validation methodology. Many studies were conducted in controlled research settings with selected populations, which may not reflect real-world performance and implementation challenges. Additionally, the limited availability of long-term follow-up data restricts understanding of the sustained impact and effectiveness of AI interventions in clinical practice.

Publication bias analysis suggests potential overrepresentation of positive results, with 87 % of included studies reporting superior or equivalent performance to standard care, whilst grey literature and conference abstracts suggest higher failure rates in real-world implementations. The predominance of studies from high-income countries may limit generalisability to resource-limited settings where anemia burden is highest and AI solutions are most needed. Specifically, 71 % of validation studies were conducted in high-income countries, with only 18 % including populations from Sub-Saharan Africa where anemia prevalence is highest.

Methodological limitations include: (1) lack of standardised quality assessment tools for AI diagnostic studies, (2) variable definition of validation methodology across studies, (3) limited assessment of algorithmic fairness and bias in most studies, and (4) insufficient reporting of implementation costs and healthcare system impact. Finally, the rapid pace of technological development means that some findings may become outdated as new AI methodologies and implementation approaches emerge.

9. Translational gap: from research to clinical practice

A critical challenge in AI anemia care is the substantial gap between promising research findings and successful clinical implementation. Analysis of current literature reveals several key barriers:

I. **Research-to-Practice Timeline:** Average time from initial AI model development to clinical deployment ranges from 3–7 years, with only 15 % of published AI anemia detection systems progressing beyond pilot testing.

II. **Scalability Challenges:** Most successful research studies involve 100–1000 participants, **whilst** health system implementation requires validation across 10,000+ diverse patients. Performance typically degrades 10–20 % during scaling.

III. **Integration Complexity:** Healthcare systems report implementation costs of $50,000–200,000 per AI tool, with 60 % of costs related to workflow integration rather than technology acquisition.

IV. **Regulatory Pathway:** Current regulatory frameworks are designed for traditional medical devices, creating approval delays of 18–36 months for AI systems that may become outdated during review.

V. **Key recommendations for bridging this gap include:** (1) establishment of standardised AI validation protocols for anemia care, (2) development of regulatory sandboxes for real-world AI testing, (3) creation of implementation toolkits for healthcare systems, and (4) funding mechanisms that support translation from research to practice.

10. Conclusion

Artificial intelligence shows significant potential for transforming anemia care through innovative screening, diagnostic, and management approaches that address critical barriers in traditional healthcare delivery. The evidence synthesised in this review demonstrates that AI technologies have achieved notable progress in developing non-invasive screening tools, accurate diagnostic algorithms, and personalised management systems that can potentially improve accessibility, affordability, and effectiveness of anemia care globally, though significant implementation challenges remain.

Image-based screening applications using smartphone cameras and deep learning algorithms have shown promise for democratising anemia detection in resource-limited settings, achieving variable diagnostic accuracy (AUC 0.75–0.97) that may be comparable to traditional laboratory methods in controlled research settings, though real-world performance validation remains limited. Machine learning applications in anemia diagnosis have demonstrated potential superior performance in differentiating anemia subtypes and predicting treatment outcomes in research settings, enabling more precise and personalised therapeutic approaches that may optimise patient outcomes whilst reducing healthcare costs pending clinical validation.

The integration of AI with clinical decision support systems, wearable technologies, and IoT devices represents a potential future direction for comprehensive anemia management, enabling continuous monitoring, proactive intervention, and dynamic treatment adjustment based on real-time physiological data. These advances could potentially transform anemia care from episodic, reactive treatment to continuous, preventive management that addresses both immediate clinical needs and long-term health outcomes subject to successful clinical implementation and cost-effectiveness validation.

However, substantial challenges must be addressed to realise the potential of AI in anemia care. Data quality issues, algorithmic bias affecting performance in diverse populations with specific performance gaps in darker-skinned individuals and LMIC populations, regulatory barriers requiring 18–36 month approval processes, deployment equity challenges in resource-limited settings where infrastructure and connectivity limitations impede implementation, and clinical trust concerns with providers requiring explanatory features require urgent attention through collaborative efforts involving researchers, clinicians, technology developers, regulatory agencies, and affected communities. The translational gap between research findings and clinical practice remains significant, with only 15 % of AI systems progressing beyond pilot testing and implementation costs ranging from $50,000–200,000 per tool. Addressing these challenges requires coordinated action following WHO 2023 AI ethics guidance and FDA/EMA regulatory frameworks, with particular attention to inclusive dataset curation, fairness-aware algorithms, tiered deployment approaches for varying resource settings, and meaningful community engagement in AI development and deployment decisions.

The development of explainable AI models enabling clinician

verification and oversight, global data sharing initiatives through federated learning whilst preserving data sovereignty, multimodal integration addressing anemia-specific clinical needs, priority population validation in pregnant women and children under 5 years, and national health system integration through existing digital health platforms will be essential for ensuring equitable implementation and sustained clinical adoption.

Future research should prioritise the development of culturally appropriate, locally validated AI tools that address the specific needs and constraints of different healthcare settings. Our proposed action plan includes: (1) Phase 1–2 development of XAI frameworks and multimodal architectures with specific technical milestones and clinical validation targets, (2) LLM integration for multilingual clinical decision support accessible via basic communication channels, (3) establishment of global federated learning consortia with equitable governance and sustainability mechanisms, (4) systematic validation in priority populations (pregnant women, children under 5, CKD patients) with rigorous performance and equity metrics, and (5) integration with national health systems through maternal-child health programmes, school screening initiatives, and EHR interoperability. These concrete next steps, with defined timelines and measurable outcomes, provide a roadmap for translating AI research into clinical impact whilst addressing equity, interpretability, and deployment feasibility challenges.

As digital health initiatives expand globally and AI technologies continue to advance, the integration of AI in anemia care offers a potential pathway toward achieving universal screening coverage and equitable health outcomes, though success will depend on overcoming current implementation barriers and ensuring rigorous clinical validation in the populations and settings where anemia burden is highest. The evidence presented in this review supports continued investment in AI research and development, with particular emphasis on addressing implementation challenges, regulatory pathways, deployment equity in LMICs, and ensuring that these technologies serve the populations most in need of improved anemia care.

CRediT authorship contribution statement

David B. Olawade: Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Investigation, Formal analysis, Conceptualization. **Yinka Julianah Adeniji:** Writing – review & editing, Writing – original draft, Methodology, Investigation. **Faithful A. Olatunbosun:** Writing – review & editing, Writing – original draft, Methodology, Investigation. **Eghosasere Egbon:** Writing – review & editing, Visualization, Methodology. **Aanuoluwapo Clement David-Olawade:** Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no conflict of interest.

References

[1] Tejavat S. Prevalence of anemia: a comprehensive review. Int J Clinic Diagn Pathol 2023;6:88–97.
[2] Zhang F, Zhang S, Ren W, et al. Global burden of anemia and cause among children under five years 1990–2019: findings from the global burden of disease study 2019. Front Nutr 2024;11:1474664.
[3] World Health Organization. The global prevalence of anaemia in 2011. Geneva: WHO; 2015.
[4] Kassebaum NJ, Jasrasaria R, Naghavi M, et al. A systematic analysis of global anemia burden from 1990 to 2010. Blood 1990;123:615–24.
[5] Garcia-Casal M, Dary O, Jefferds M, et al. Diagnosing anemia: challenges selecting methods, addressing underlying causes, and implementing actions at the public health level. Ann N Y Acad Sci 2023;1524:37–50.
[6] Balarajan Y, Ramakrishnan U, Özaltin E, et al. Anaemia in low-income and middle-income countries. Lancet 2011;378:2123–35.
[7] Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019;25:44–56.
[8] Lin H, Fernandes S, Sharif M, et al. Special issue: recent advances in deep learning, biometrics, health informatics and data science. Expert Syst 2022;39(7):e13060.
[9] Dean J, Madani A, Esteva A, et al. Deep learning-enabled medical computer vision. NPJ Digit Med 2021;4(1):5.
[10] Collings S, Thompson O, Hirst E, et al. Non-invasive detection of anaemia using digital photographs of the conjunctiva. PLoS One 2016;11(4):e0153286.
[11] Sehar N, Krishnamoorthi N, Kumar CV. Deep learning model-based detection of anemia from conjunctiva images. Healthc Inform Res 2025;31:57–65.
[12] Bharathi G, Berks M, Dinsdale G, et al. A deep learning system for quantitative assessment of microvascular abnormalities in nailfold capillary images. Rheumatology 2023;62:2325–9.
[13] Isewon I, Alagbe E, Oyelade J. Optimizing machine learning performance for medical imaging analyses in low-resource environments: the prospects of CNN-based feature extractors. F1000Res 2025:100.
[14] Bärnighausen T, Weber A-C, Kumar A, et al. Smartphone-based point-of-care anemia screening in rural Bihar in India. Commun Med 2023;3.
[15] Lee H, Brandwein M, Shpiegel R, et al. Artificial intelligence-enabled non-invasive ubiquitous anemia screening: the HEMO-AI pilot study on pediatric population. Digit Health 2024;10:20552076241297057.
[16] Hammel N, Corrado G, Huang A, et al. Detection of anaemia from retinal fundus images via deep learning. Nat Biomed Eng 2019;4:18–27.
[17] Mahmud S, Donmez TB, Mansour M, et al. Corrigendum: anemia detection through non-invasive analysis of lip mucosa images. Front Big Data 2023;6:1241899.
[18] Appiahene P, Asare JW, Donkoh ET, et al. Detection of iron deficiency anemia by medical images: a comparative study of machine learning algorithms. BioData Min 2023;16(1):2.
[19] Zhao X, Meng L, Su H, et al. Deep-learning-based hemoglobin concentration prediction and anemia screening using ultra-wide field fundus images. Front Cell Dev Biol 2022;10:888268.
[20] Muljono WSA, Al AH, et al. Breaking boundaries in diagnosis: non-invasive anemia detection empowered by AI. IEEE Access 2024;12:9292–307.
[21] MÁ N-M, Gómez-Martínez-Sagrera P, MÁ A-G, et al. Mobile apps for hematological conditions: review and content analysis using the mobile app rating scale. JMIR Mhealth Uhealth 2022;10:e32826.
[22] Park S, Kim J, Lee J, et al. Ember: task wakeup sequence–Based energy optimization for mobile web browsing. ACM Trans Embedd Comput Syst 2025;24:1–24.
[23] Labrique AB, Wadhwani C, Williams KA, et al. Best practices in scaling digital health in low and middle income countries. Global Health 2018;14:103.
[24] Gernsheimer T, Norby-Slycord C, Hawkins D, et al. HemaApp: noninvasive blood screening of hemoglobin using smartphone cameras. In: Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing; 2016. p. 593–604.
[25] Perez-Plazola MS, Tyburski EA, Smart LR, et al. AnemoCheck-LRS: an optimized, color-based point-of-care test to identify severe anemia in limited-resource settings. BMC Med 2020;18(1):337.
[26] Chen P, Dong W, Wang J, et al. Interpretable clinical prediction via attention-based neural network. BMC Med Inform Decis Mak 2020;20(Suppl 3):131.
[27] Zihni E, Madai V, Livne M, et al. Opening the black box of artificial intelligence for clinical decision support: a study predicting stroke outcome. PLoS ONE 2019;15(4):e0231166.
[28] Karagül Yıldız T, Yurtay N, Öneç B. Classifying anemia types using artificial learning methods. Eng Sci Technol 2021;24:50–70.
[29] Saputra DCE, Sunat K, Ratnaningsih T. A new artificial intelligence approach using extreme learning machine as the potentially effective model to predict and analyze the diagnosis of Anemia. Healthcare 2023;11(5):697.
[30] Nadhiroh S, Hasugian A, Nurhayati, et al. Model development for anemia prediction in pregnancy. Clin Epidemiol Glob Health 2024:101654.
[31] Su Z, Nong Z, Huang F, et al. Anemia risk prediction model for osteosarcoma patients post-chemotherapy using artificial intelligence. Cancer Med 2024;13(23):e70427.
[32] Dejene BE, Abuhay TM, Bogale DS. Predicting the level of anemia among Ethiopian pregnant women using homogeneous ensemble machine learning algorithm. BMC Med Inform Decis Mak 2022;22:247.
[33] Yuan Q, Zhang H, Deng T, et al. Role of artificial intelligence in kidney disease. Int J Med Sci 2020;17:970–84.
[34] Gómez JG, Parra Urueta C, Álvarez DS, et al. Anemia classification system using machine learning. Informatics 2025;12(1):19.
[35] Yimer A, Yesuf HA, Ahmed S, et al. Optimizing machine learning models for predicting anemia among under-five children in Ethiopia: insights from Ethiopian demographic and health survey data. BMC Pediatr 2025;25:311.
[36] Göl M, Aktürk C, Talan T, et al. Predicting malnutrition-based anemia in geriatric patients using machine learning methods. J Eval Clin Pract 2025;31:e14142.
[37] Barbieri C, Molina M, Ponce P, et al. An international observational study suggests that artificial intelligence for clinical decision support optimizes anemia management in hemodialysis patients. Kidney Int 2016;90:422–9.
[38] Escandell-Montero P, Chermisi M, Martínez-Martínez JM, et al. Optimization of anemia treatment in hemodialysis patients via reinforcement learning. Artif Intell Med 2014;62:47–60.
[39] Nagraj S, Kennedy S, Jha V, et al. A mobile clinical decision support system for high-risk pregnant women in rural India (SMARThealth Pregnancy): pilot cluster randomized controlled trial. JMIR Form Res 2023;7:e44362.

[40] Wang B, Shi X, Han X, et al. The digital transformation of nursing practice: an analysis of advanced IoT technologies and smart nursing systems. Front Med 2024:11.
[41] A A, K VK, Bedant A, et al. Non-invasive method of detecting anemia using AI & IoT. In: 2023 International conference on innovative data communication technologies and application (ICIDCA); 2023. p. 805–8.
[42] Mannino R, Sullivan J, Frediani J, et al. Real-world implementation of a noninvasive, AI-augmented, anemia-screening smartphone app and personalization for hemoglobin level self-monitoring. Proc Natl Acad Sci U S A 2025;122:20.
[43] Rajkomar A, Hardt M, Howell MD, et al. Ensuring fairness in machine learning to advance health equity. Ann Intern Med 2018;169:866–72.
[44] Adamson AS, Smith A. Machine learning and health care disparities in dermatology. JAMA Dermatol 2018;154:1247–8.
[45] Buolamwini J, Gebru T. Gender shades: intersectional accuracy disparities in commercial Gender classification. In: Proceedings of the 1st conference on fairness, accountability and transparency. 81; 2018. p. 77–91.
[46] Osei E, Mashamba-Thompson T. Mobile health applications for disease screening and treatment support in low-and middle-income countries: a narrative review. Heliyon 2021;7(3):e06639.
[47] Holst C, Sukums F, Radovanović D, et al. Sub-Saharan Africa-the new breeding ground for global digital health. Lancet Digit Health 2020;2:4.
[48] Alami H, Rivard L, Lehoux P, et al. Artificial intelligence in health care: laying the Foundation for Responsible, sustainable, and inclusive innovation in low- and middle-income countries. Global Health 2020;16(1):52.
[49] López D, Rico-Olarte C, Blobel B, et al. Challenges and solutions for transforming health ecosystems in low- and middle-income countries through artificial intelligence. Front Med (Lausanne) 2022;9:958097.
[50] Muehlematter UJ, Daniore P, Vokinger KN. Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015–20): a comparative analysis. Lancet Digit Health 2021;3:e195–203.
[51] Holzinger A., Biemann C., Pattichis C.S., et al. What do we need to build explainable AI systems for the medical domain? ArXiv Preprint ArXiv:171209923, 2017.
[52] Wu Q., Li S., Zhou D., et al. Toward fair federated learning under demographic disparities and data imbalance, 2025.
[53] Palaniappan K, Lin EYT, Vogel S. Global regulatory frameworks for the use of artificial intelligence (AI) in the healthcare services sector. Healthcare 2024;12:562.
[54] Wysocki O, Davies JK, Vigo M, et al. Assessing the communication gap between AI models and healthcare professionals: explainability, utility and trust in AI-driven clinical decision-making. Artif Intell 2023:316.
[55] Muhammad A, Pasha MF, AMY, et al. Correction: the Fast Health Interoperability Resources (FHIR) standard: systematic literature review of implementations, applications, challenges and opportunities. JMIR Med Inform 2021;9:e32869.
[56] Shah R, Bozic KJ, Jayakumar P. Artificial intelligence in value-based health care. HSS J 2025:15563316251340074.