

Artificial intelligence-enabled non-invasive ubiquitous anemia screening: The HEMO-AI pilot study on pediatric population

Daniel Gordon^{1,*}, Jason Hoffman^{1,2,*}, Keren Gamrasni^{1,*}, Yotam Barlev^{1,3}, Alex Levine^{1,3}, Tamar Landau¹, Ronen Shpiegel⁴, Avishai Lahad⁴, Ariel Koren⁴, Carina Levin⁴, Osnat Naor⁴, Hannah Lee², Xin Liu², Shwetak Patel², Gilad Chayen^{4,**} and Michael Brandwein^{1,5,**} 

Abstract

Objective: Determine whether data collected from a smartphone camera can be used to detect anemia in a pediatric population.

Methods: HEMO-AI (Hemoglobin Easy Measurement by Optical Artificial Intelligence), a clinical study carried out from December 2020 to February 2023, recruited patients from the Pediatric Emergency Department, Pediatric Inpatient Department and Pediatric Hematology Unit of the Haemek Medical Center, Afula, Israel. A population-based sample of 823 patients aged 6 months to 18 years who had undergone a venous blood draw for a complete blood count since being admitted to the hospital were enrolled. Patients with total leukonychia, nailbed darkening or discoloration due to medication, nail clubbing, clinically indicated jaundice, subungual hematoma, nailbed lacerations, avulsion injuries, or nail polish applied on fingernails were not eligible for study recruitment. Video and images of the patients' hand placed in a collection chamber were collected using a smartphone camera.

Results: About 823 samples, 531 from a 12.2 megapixel camera and 256 from a 12.2 megapixel camera, were collected. About 26 samples were excluded by the study coordinator for irregularities. About 97% of fingernails and 68% of skin samples were successfully identified by a post-trained machine learning model. Separate models built to detect anemia using images taken from the Pixel 3 had an average precision of 0.64 and an average recall of 0.4, whereas models built using the Pixel 6 had an average precision of 0.8 and an average recall of 0.84. Further supplementation of training data with synthetic data boosted the precision of the latter to 0.84 and the average recall to 0.87.

Conclusions: This study lays the groundwork for the future evolution of non-invasive, pain-free, and accessible anemia screening tools tailored specifically for pediatric patients. It identifies important sample collection parameters and design, provides critical algorithms for the pre-processing of fingernail data, and reports an initial capability to detect anemia with 87% sensitivity and 84% specificity.

Trial Registration: Prospectively registered on www.clinicaltrials.gov (Identifier: NCT04573244) on 15 September 2020, prior to subject recruitment.

¹Department of Research & Development, MyOr Diagnostics Ltd., Zichron Yaakov, Israel

²Paul G. Allen School of Computer Science and Engineering, University of Washington, Seattle, WA, USA

³Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel

⁴Pediatric Department, HaEmek Medical Center, Afula, Israel

⁵Department of Molecular Biology, Ariel University, Ariel, Israel

*These authors contributed equally to this work.

**These authors are co-corresponding authors.

Corresponding author:

Michael Brandwein, Department of Research & Development, MyOr Diagnostics Ltd., Bareket 29, Zichron Yaakov, Israel.
Email: michael@myorcare.com



Keywords

Telemedicine general, adolescent medicine, pediatrics medicine, remote patient monitoring personalized medicine, connected devices personalized medicine

Submission date: 23 January 2024; Acceptance date: 17 October 2024

Key points

- Question: Can smartphone cameras be used to detect anemia in the pediatric population?
- Findings: In this clinical study that included 823 children aged 6 months to 18 years, we developed a machine learning model to classify anemia with 87% sensitivity and 84% specificity.
- Meaning: Machine learning algorithms have the potential to detect anemia in pediatric patients using only a smartphone camera.

Introduction

The emergence of artificial intelligence (AI) in healthcare has begun to transform diagnostics, and as of October 2022, the FDA has approved 521 artificial intelligence and machine learning-enabled medical devices.¹ Nearly 3% of these approved devices are hematology devices and include several devices for the analysis of blood taken for a complete blood count (CBC). To date, an entirely non-invasive AI-enabled device for blood component monitoring has yet to receive FDA approval.

Anemia is defined as either a reduced absolute number of circulating red blood cells (RBCs), a diminution in the volume of occupied RBCs, or an insufficiency of the RBCs to meet physiological oxygen-carrying capacity needs.² It is the most widespread hematological condition globally and affected 42% of children under 5 years of age in 2016.² Anemia is associated with negative health outcomes in the pediatric population, including delayed cognitive and behavioral development³ and functioning.⁴ Clinical symptoms of anemia include fatigue, shortness of breath, palpitations, and clinical pallor.⁵

Anemia is most often diagnosed through a venous blood draw in the clinical setting, relying on several blood parameters, including hemoglobin (Hb), hematocrit, RBC count, mean corpuscular volume, and others. Whereas Hb often provides the first clue as to an individual's status, the other measurements, coupled with a detailed clinical history, are used in diagnostic algorithms to pinpoint the cause of anemia.^{6,7} The World Health Organization's recommendations on the use of Hb concentration for diagnosing anemia, which has been widely accepted since its publication, provide a summary table outlining blood Hb levels to diagnose anemia stratified

by age, sex, and pregnancy. Importantly, it states that children from 6 months to 14 years of age are non-anemic if their blood Hg level is higher than 11–12 grams per deciliter.⁸

The bulk of anemia screening is carried out in the health-care setting through the CBC, a routine medical procedure requiring a venous blood draw and specialized machinery. At scale, the CBC is the most accurate and cost-effective anemia screening methodology. However, it necessitates an invasive and painful blood draw by a trained professional, does not provide instantaneous results, and relies on expensive machinery and trained technicians.⁹ This makes it particularly inaccessible in underdeveloped countries. The HemoCue, introduced in 1980, attenuates, but does not alleviate, these pain points, as it requires blood from a finger prick and specialized machinery, yet can provide results in about 1 minute.¹⁰ A further advancement in attenuating some of the pain points associated with the CBC came with the introduction of the Masimo Radical-7 SpHb Station which allows for entirely non-invasive blood measurement, albeit at the cost of precision.¹¹ Importantly, the unique absorption spectra of Hb and its derivatives were characterized nearly a century ago.¹²

Recognizing the need for a noninvasive and accessible tool for anemia detection and taking into account the known colorimetric element of anemia (i.e., pallor at distinctive anatomical sites), a flurry of studies have explored the feasibility of using a smartphone for blood Hb measurements in the past few years (summarized in Wemyss et al.⁹). Many of these studies have employed machine learning because it can create an objective function out of previously subjective measures in multifactorial scenarios, such as that in anemia screening where blood coloration in fingernails, noise from ambient lighting, skin tone, and age can all play a role. Each study addresses the inherent combined challenges of ambient lighting, feature selection, color channels, and model input slightly differently, and together paint a promising picture of the future of anemia detection using smartphones. However, studies suffer from small sample sizes, ranging from 29 to 344 individuals, and all but one include data solely from adults. Therefore, there remains a need to develop a smartphone-based anemia screening technology suited for the pediatric population using a robust and representative dataset with broad applicability to the greater population.

Materials and methods

Participant recruitment and selection criteria

Patients were recruited from the Pediatric Emergency Department (PED), Pediatric Inpatient Department (PD), and Pediatric Hematology Unit (PHU) of the Haemek Medical Center, Afula, Israel from December 2020 until February 2023. Written informed consent was obtained by parents/guardians of enrolled participants according to the approved protocols and procedures received from the Helsinki Committee of the Haemek Medical Center (study name: HEMO-AI, approval number 0150-20-EMC). The study was prospectively registered on www.clinicaltrials.gov (Identifier: NCT04573244) on 15 September 2020,

prior to subject recruitment. Enrollment criteria can be found in this article's Supplementary Methods Section.

Data collection

Patients were instructed to place their hands in a customized chamber, described in the Supplementary Methods Section and in Supplementary Figure 1, in a splayed position for data collection. An Android-native application (app) was designed and developed for the purpose of the study data collection, described further in the Supplementary Methods Section. The app was loaded onto a phone (GP3) with a 12.2 megapixel rear-facing camera and on a separate phone (GP6) with a 50 megapixel rear-facing camera.

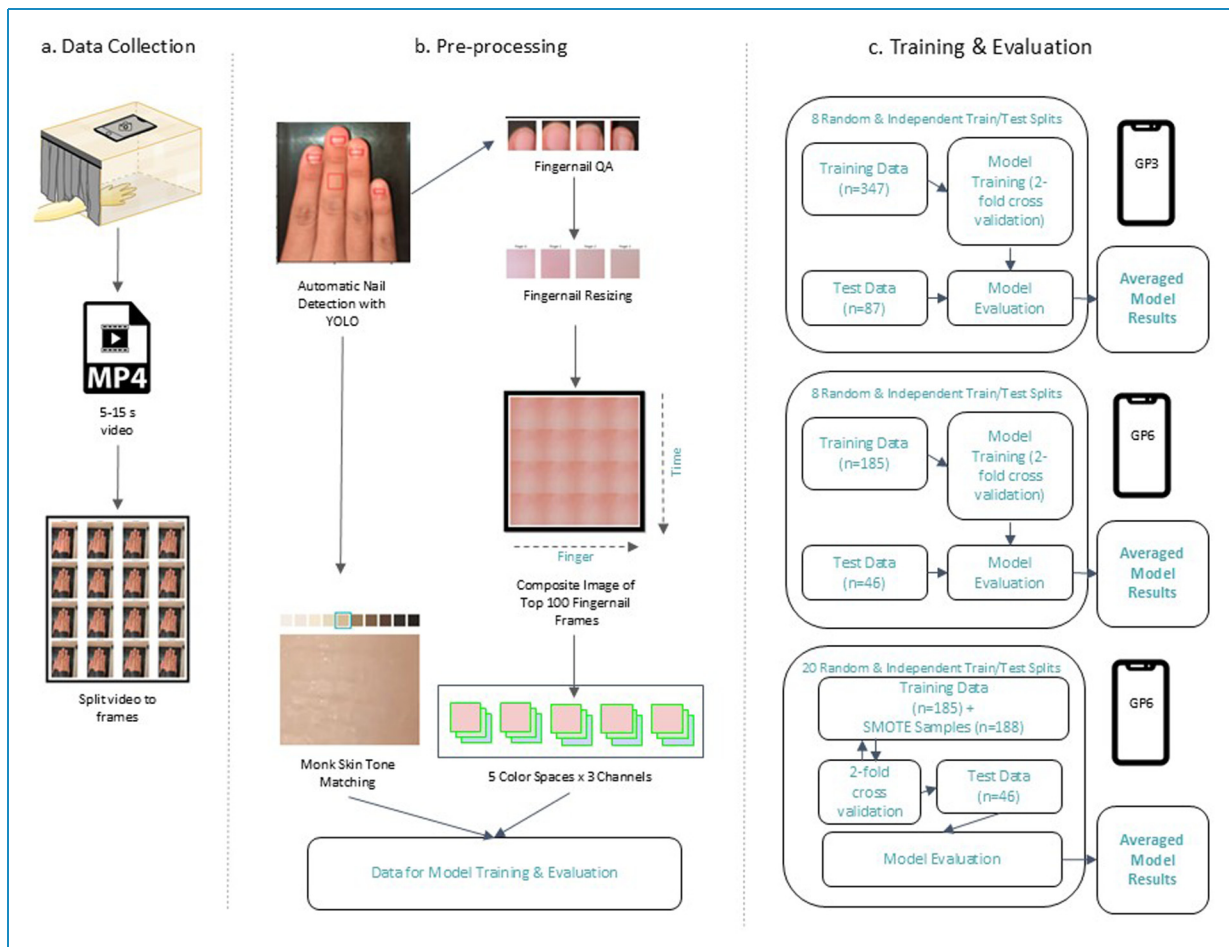


Figure 1. Overview of data collection and processing pipeline. (a) Signal extraction occurs by collecting 5–15 s of video on a smartphone camera from the patient's hand situated in a collection chamber. The video is then parsed to individual frames. (b) Individual frames are put through a set of pre-processing steps. The post-trained YOLO model detects a swatch of skin from the middle finger and the four fingernails from the index, middle, ring, and little fingers. The skin swatch is matched to its corresponding monk skin tone. The fingernail images undergo a quality scoring and resizing pipeline, at the end of which a composite image of the top 100 fingernail frames is generated. The gain from 15 different color channels is extracted. (c) Three distinct models were trained and evaluated. The upper panel depicts the development of a model using images collected with a GP3, the middle panel depicts the development of a model using images collected with a GP6, and the lower panel depicts the development of a model using images collected with a GP6, which was supplemented with SMOTE-generated images for the training of the model.

Data preprocessing

Fingernails from patient images were manually tagged and used to post-train an object-detection classifier, YOLOv8 (You Only Look Once), to detect patient fingernails from a set of images.¹³ A histogram representing the color intensity was constructed for 15 color channels and was input into the machine learning model. Additionally, a 50×50 pixel area of skin from the middle finger was detected and automatically matched to its corresponding monk skin tone (MST) swatch which was input into the machine learning model along with the relative error of detected MST to account for inter-level variability.¹⁴ Figure 1(a) and (b) depicts the data preprocessing workflow and further information on the data preprocessing pipeline can be found in this article's Supplementary Methods Section.

Anemia classification algorithms

XGBoost, a scalable gradient-boosted decision tree machine learning algorithm¹⁵ was trained on 80% of the clinical dataset and then tested on the remaining 20% of the dataset. A twofold cross-validation was performed for each model iteration whereby the training set was split in half to optimize the model hyperparameters. Eight independent test-train splits were conducted and eight separate classification models were built to observe the reproducibility and consistency of the model when trained with different samples. The Synthetic Minority Over-sampling Technique (SMOTE) was used to generate additional artificial anemia examples for the train set.¹⁶ Figure 1(c) depicts the training and evaluation pipelines. Model hyperparameters included a maximum tree depth of 3, a learning rate of 0.5, a Logloss loss function, and grid search for hyperparameter tuning. Model development was performed using Sklearn in the Python 3.8 object-oriented programming language. Performance evaluation metrics included recall (sensitivity), accuracy, precision (positive predictive value), f1 score, and area under the curve (AUC) of the receiver operating characteristic (ROC) curve.

Results

Clinical study results

We collected 823 clinical samples from pediatric patients who visited the PED, and PHU from patients who were hospitalized in the PD at HaEmek Medical Center. About 36 samples were excluded by the study coordinator for irregularities, including fingernails that were photographed with fingernail polish. Samples were collected using a GP3 from December 2020 until June 2022 ($n=531$) and then using a GP6 from June 2022 until February 2023 ($n=256$). Figure 2 depicts a CONSORT diagram presenting sample collection. Most samples were collected in the PED from patients as they underwent a CBC, with an additional $\sim 10\%$ of samples gathered in the PHU and PD from patients whose CBC had been previously recorded and were found to be anemic. Anemic patients were significantly older than non-anemic patients in the first round of sample collection ($p < 0.01$, Table 1), whereas anemic patients were significantly younger than non-anemic patients in the second round of sample collection ($p < 0.0001$, Table 1). No significant differences were noted for sex between anemic and non-anemic patients in either round of sample collection. Histograms depicting the distribution of both data sets are presented in Supplementary Figure 2.

Fingernail detection

We developed a model to autonomously distinguish fingernails and skin sections from other sections of the parsed images, serving as essential inputs for the subsequent algorithm designed to quantify blood Hb concentrations. A pre-trained machine learning YOLO model was post-trained using 329 tagged images of the patients' fingernails and representative skin section. The model's performance was then tested on 44 images. The model successfully identified 97% of fingernails and 68% of skin samples. Figure 3(c) and (d) depicts the confusion matrix and the associated precision-

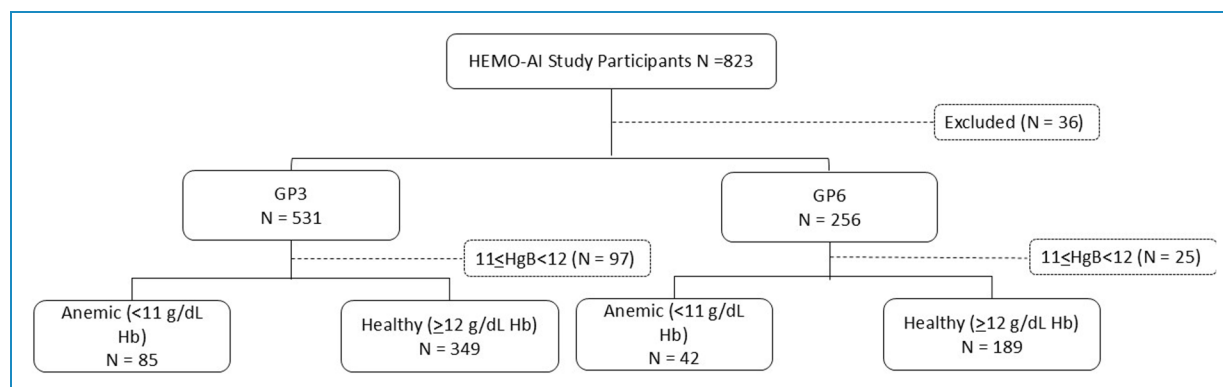


Figure 2. Flow diagram of patients enrolled in the Hemoglobin Easy Measurement by Optical Artificial Intelligence (HEMO-AI) study.

Table 1. Demographic and clinical characteristics of anemic and non-anemic patients.

	GP3		GP6			
	>12 g/dL Hb, n = 349	<11 g/dL Hb, n = 82	p-value	>12 g/dL Hb, n = 189	<11 g/dL Hb, n = 42	p-value
Age (median [IQR])	12 [8.00, 15.00]	8 [2.25, 18.50]	0.207	12 [8.00, 14.00]	5 [2.00, 10.75]	<0.001
Sex = male (%)	213 (61.0)	49 (59.8)	0.931	109 (57.7)	20 (47.6)	0.31
WBC (median [IQR])	8.77 [6.81, 12.09]	9.54 [6.51, 14.46]	0.211	9.11 [6.78, 11.97]	10.52 [6.54, 15.83]	0.418
NEUT (median [IQR])	66.2 [54.00, 79.60]	55.6 [48.43, 69.02]	<0.001	66.4 [54.25, 81.30]	65.9 [53.23, 77.57]	0.505
LYM (median [IQR])	24.4 [12.75, 33.70]	29.9 [18.32, 39.40]	0.001	24 [11.40, 32.95]	22.9 [13.10, 31.30]	0.699
MPX1 (median [IQR])	0.8 [−1.80, 3.40]	3.6 [−1.93, 5.82]	0.004	9.65 [6.40, 12.57]	10.15 [6.62, 12.03]	0.966
MONO (median [IQR])	4.9 [3.85, 5.90]	5.1 [4.20, 6.70]	0.039	5.3 [4.20, 6.30]	4.5 [3.60, 5.80]	0.069
EOS (median [IQR])	1.4 [0.60, 2.60]	1.5 [0.60, 2.87]	0.674	1.5 [0.50, 3.10]	0.95 [0.40, 1.58]	0.02
BASO (median [IQR])	0.4 [0.25, 0.60]	0.6 [0.40, 0.80]	<0.001	0.4 [0.30, 0.60]	0.4 [0.20, 0.60]	0.505
LUC (median [IQR])	1.8 [1.20, 2.50]	2.3 [1.67, 3.30]	<0.001	1.5 [1.00, 2.10]	1.7 [1.30, 2.70]	0.06
RBC (median [IQR])	4.86 [4.59, 5.06]	3.69 [3.27, 4.45]	<0.001	4.89 [4.65, 5.17]	4.26 [3.54, 4.71]	<0.001
HCT (median [IQR])	39.3 [37.70, 41.62]	29.4 [25.70, 32.00]	<0.001	39.7 [38.20, 42.10]	31.35 [28.08, 32.27]	<0.001
HB (median [IQR])	13.2 [12.60, 13.90]	9.7 [8.43, 10.40]	<0.001	13.1 [12.50, 13.90]	10.3 [9.00, 10.70]	<0.001
MCV (median [IQR])	82.5 [79.20, 85.40]	77.2 [70.90, 83.00]	<0.001	82.3 [79.10, 85.50]	74.9 [67.17, 76.95]	<0.001
MCH (median [IQR])	27.7 [26.60, 28.90]	25.3 [22.50, 27.30]	<0.001	27.1 [26.20, 28.40]	23.7 [21.08, 26.05]	<0.001
MCHC (median [IQR])	33.5 [32.90, 34.10]	32.4 [31.40, 33.30]	<0.001	33.1 [32.30, 33.80]	32.75 [31.42, 33.30]	0.007
RDW (median [IQR])	13.6 [13.10, 14.10]	16.6 [15.20, 18.40]	<0.001	14 [13.50, 14.60]	15.5 [15.03, 16.40]	<0.001
HDW (median [IQR])	2.54 [2.43, 2.73]	2.99 [2.74, 3.26]	<0.001	2.52 [2.40, 2.73]	2.82 [2.69, 3.35]	<0.001

(continued)

Table 1. Continued.

	GP3		GP6		p-value	<11 g/dL Hb, n = 82	<11 g/dL Hb, n = 189	<11 g/dL Hb, n = 42	p-value
	>12 g/dL Hb, n = 349	<11 g/dL Hb, n = 82	>12 g/dL Hb, n = 189	<11 g/dL Hb, n = 42					
MICRO (median [IQR])	1.7 [1.00, 3.80]	9.45 [4.25, 20.30]	2.1 [1.10, 3.85]	9.95 [6.83, 27.08]	<0.001				<0.001
MACRO (median [IQR])	0.1 [0.00, 0.20]	0 [0.00, 0.20]	0.1 [0.00, 0.30]	0 [0.00, 0.10]	0.017				<0.001
HYPO (median [IQR])	0.8 [0.40, 1.60]	4.8 [2.22, 11.20]	2.3 [0.88, 5.80]	6.4 [2.35, 18.65]	<0.001				<0.001
HYPHER (median [IQR])	0.7 [0.45, 1.10]	0.9 [0.50, 1.55]	0.4 [0.20, 0.80]	0.65 [0.40, 1.33]	0.062				0.008
PLT (median [IQR])	256 [216.00, 306.50]	324 [242.00, 487.00]	263 [216.00, 313.00]	282.5 [190.25, 385.00]	<0.001				0.305
MPV (median [IQR])	7.5 [7.00, 8.00]	7.4 [6.75, 7.95]	7.9 [7.40, 8.47]	7.3 [6.50, 8.15]	0.232				0.004

Wilcoxon rank-sum test was used to compare the differences between the groups. Proportions in sex variables were compared using chi-square test. IQR: interquartile range; WBC: white blood cell count; NEUT: neutrophils; LYM: lymphocytes; MPX: myeloperoxidase index; MONO: monocytes; EOS: eosinophils; BASO: basophils; LUC: large unstained cells; RBC: red blood cell count; HCT: hematocrit; HB: hemoglobin; MCV: mean corpuscular volume; MCH: mean corpuscular hemoglobin; MCHC: mean corpuscular hemoglobin concentration; RDW: red cell distribution width; HDW: hemoglobin distribution width; MICR: microcytosis; MACRO: macrocytosis; HYPO: hypochromia; HYPHER: hyperchromia; PLT: platelet count; MPV: mean platelet volume.

confidence curve for the fingernail detection process along with a set of representative training (Figure 3(a)) and representative test samples (Figure 3(b)). Fingernails detected from images or videos taken with the GP3 were each 20×20 pixels (400 pixels total), whereas fingernails detected from images or video taken with the GP6 were 80×80 pixels (6400 pixels total).

Anemia classification model

About 8500 fingernail images from 85 anemic patients and 34,900 fingernail images from 349 non-anemic patients were taken with a GP3 and separately, 4200 fingernail images from 42 anemic patients and 18,900 fingernail images from 189 non-anemic patients taken with a GP6 were used to train and validate eight independent iterations of the XGBoost model. Importantly, all 100 images from each individual were pooled together and presented as one composite image to either the train or test dataset once per model iteration. Samples with a blood Hb level below 11 g/dL were classified as anemic, whereas samples above 12 g/dL were classified as non-anemic. The mean accuracy for the model using images taken from the GP3 phone was 0.51 with an average precision of 0.64, average recall of 0.4 and an average ROC AUC of 0.53. The mean accuracy for the model using images taken from the GP6 phone was 0.79 with an average precision of 0.8, average recall of 0.84 and an average ROC AUC of 0.81. Supplementary Table 1 presents the metrics of each of the eight independent test-train iterations. To train the model on a more balanced anemia/non-anemia sample set, we used the SMOTE computational technique to generate new artificial anemia samples for the training set ($n = 118$). We trained 20 independent iterations with corresponding random and independent train-test splits of the XGBoost model. The mean accuracy for the model using images taken from the GP6 phone was 0.85 with an average precision of 0.84, average recall of 0.87, and an average ROC AUC of 0.93. Supplementary Table 2 presents the independent metrics of each of the 20 test-train iterations and Figure 4 presents the average ROC curve with the corresponding mean AUC and CI. The mean sensitivity of these models is 87% and the mean specificity is 84%.

Discussion

The results of our study provide a framework for the continued development of a non-invasive, pain-free, and accessible anemia screening tool for the pediatric patient that does not require specialized and costly machinery or expertise. We honed machine learning models that automatically detect fingernails of all sizes while in motion as a crucial preprocessing step for widespread adoptability of any smartphone-based anemia classification algorithm. The pipeline parses video to images and extracts color channel

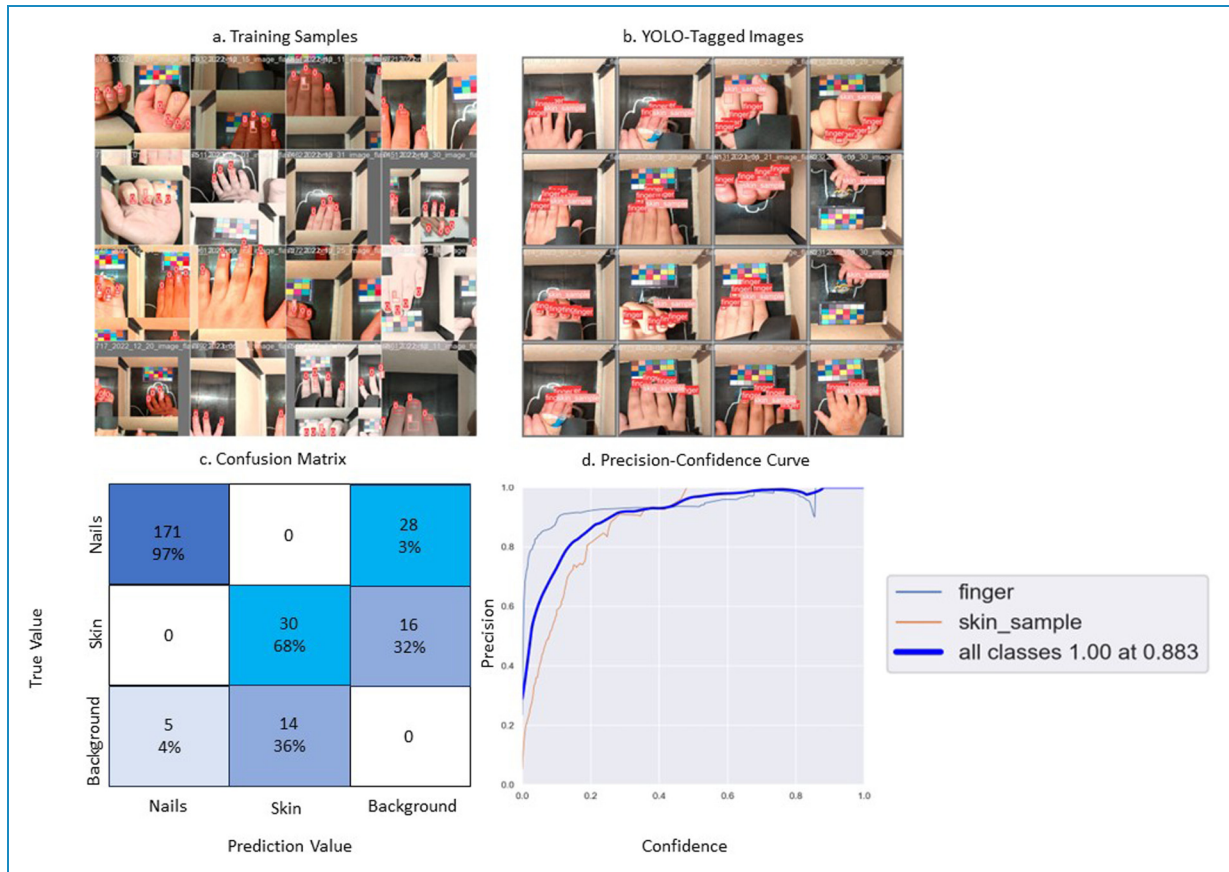


Figure 3. Automatic fingernail detection. (a) A set of representative training samples for the post-training of the YOLO fingernail and skin detection model. (b) A set of representative labeled samples tagged by the post-trained YOLO fingernail and skin detection model. (c) Confusion matrix for the fingernail, skin, and background detection process. (d) Precision recall curve for the model is shown. Precision represents the positive predictive value of the algorithm, whereas recall represents the sensitivity of the model.

histogram data and statistics from each finger in each frame of the video. Anemia detection algorithms were sensitive to image quality, as measured by pixel density, and performed well only when trained and tested on images taken with the GP6 (released in October 2021) and not from the GP3 (released in October 2018). The sensitivity (87%) and specificity (84%) of our anemia classification pipeline provide an attractive alternative for anemia screening in the context of routine well-baby visits, annual checkups, or in underserved populations.

Whereas our anemia screening algorithms are first developed on a large dataset of pediatric patients, other tools developed for the non-pediatric population provided important benchmarks for the development of our pipeline. The vast majority of studies undertaken to date use photographs of the eye, including the sclera,¹⁷ retinal fundus,¹⁸ or lower eyelid.¹⁹ Others used fingernail²⁰ or fingertip images.²¹ Many also grapple with the effect of ambient light on the quality of the photograph, a feature more easily controlled when photographing the hand as opposed to the facial region. The design and employment of a specialized photo collection chamber for our study

(Supplementary Figure 1) was essential for controlling for ambient lighting conditions, thereby ensuring consistency and minimizing extraneous variables, which enhances the accuracy and repeatability of the image data obtained. The materials used to construct the specialized photo collection chamber—wood, masking tape, and black cloth—are economical and portable, making it an optimal tool for utilization in underserved populations and adaptable to a myriad of settings, thereby democratizing access to this technology.

Previous studies count anemic patients with inherited forms of anemia (i.e., sickle cell anemia or thalassemia) as the majority of anemic patients in their cohort. This restricts the model's generalizability, as the idiosyncrasies and unique characteristics of this cohort may not accurately reflect the heterogeneity and diverse manifestations of the condition in the broader population. Furthermore, this limits the potential intended purpose of the technology, as regulatory bodies will most likely restrict the intended purpose of the device to match that of the study population.

In our anemia detection pipeline, we made the deliberate choice to use videos over static images, a decision that

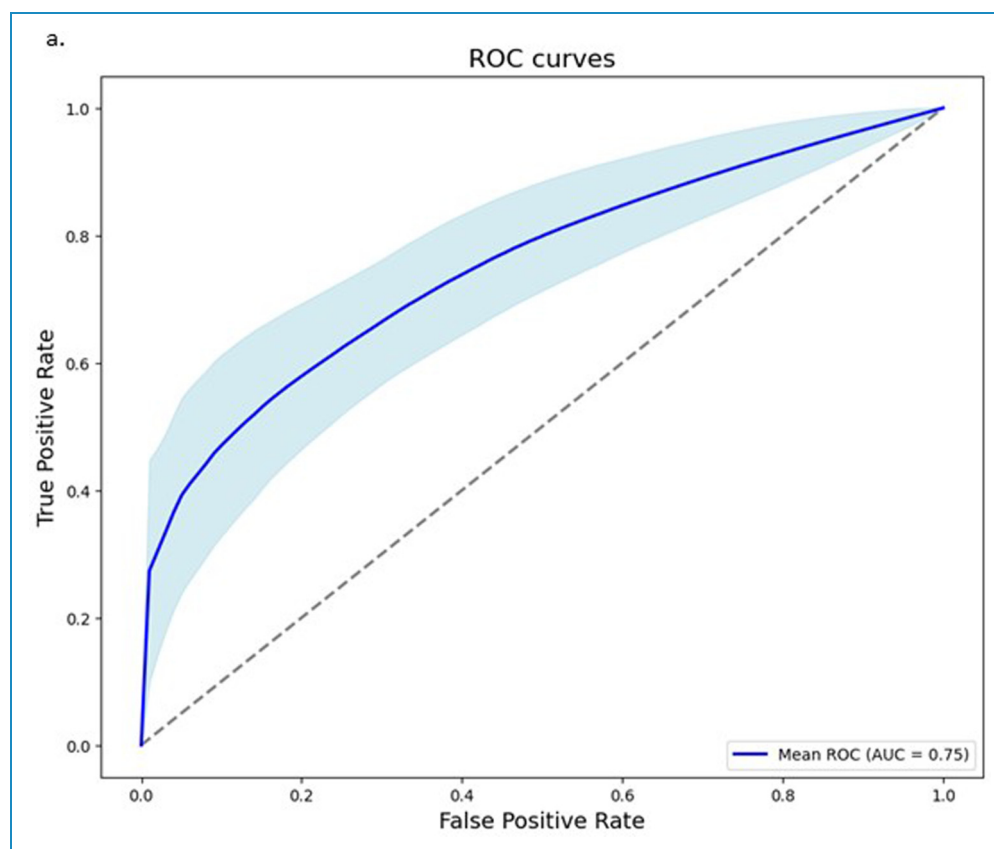


Figure 4. Anemia detection model performance. (a) Mean area under the receiver operating characteristic curve of 0.75 for the XGBoost model with 95% CIs representing the variation of 20 independent model runs.

significantly enriched our data pool for machine learning models. By parsing each video into hundreds of individual frames, we were able to extract a multitude of data points from a single sample. Crucially, we implemented a strict policy of single representation for each individual's samples: the images parsed from a particular person's video were pooled and utilized only once per k-split model, in either the training set or test set, but not both. This approach ensured that each individual's samples were exclusively represented once in either the test set or the training set, thereby preserving the integrity and variability of the datasets. This shift to video data was enabled through the post-training of the YOLO machine learning model to specifically identify fingernails and a skin swatch from the finger in each frame. This approach not only expanded our data repository but also helped to enhance the overall performance of our anemia detection models by enabling more nuanced and varied data for training. In general, drawbacks of using videos over images can include poorer image quality and noise due to compression, which were outweighed by the advantages of videos in our sample set and pipeline.

Our pipeline recognizes the need to account for skin tone when utilizing images to screen for anemia, given the

known racial bias in pulse oximetry measurements.²² Variations in skin pigmentation can significantly influence the accuracy of color-based assessments, introducing potential bias and disparity in diagnosis. Neglecting to account for this variable could inadvertently perpetuate the issues experienced with pulse oximetry, wherein darker skin pigmentation has been linked to less accurate measurements. The introduction of the MST¹⁴ and the matching of an individual's skin tone to a correlative MST (Figure 1) provides the groundwork for an integrative approach to ensure equitable and accurate anemia detection across diverse populations.

Limitations of our study include the relatively small sample size of the data set used to train the algorithms and the non-normal distribution of ages and significant differences of ages between anemic and non-anemic patients. The former stems from our decision to recruit patients from the general population, mostly in the PED, and not from more anemia-prone clinical settings. This decision was made considering the notable phenotypic abnormalities of RBCs in sickle-cell anemic and thalassemic patients, which we hypothesized would affect the colorimetric properties of the blood and the notable pallor of cancer patients undergoing chemotherapy, a phenomenon not directly related to blood

Hb levels. Algorithms trained with such patient samples may learn to classify individuals based on the unique physiological presentations mentioned, and not by blood Hb levels as mentioned above. To address the limitations of sample size and age differences, we trained eight separate and independent algorithms using eight random train-test splits. The results and metrics of the algorithms were consistent and indicative of a robust methodology for classifying anemia using the input data. Furthermore, we utilized the SMOTE, a computational tool often used in the field of health informatics to address the issue of imbalance in datasets by generating artificial examples from the minority class (anemic patients). This is achieved by pinpointing characteristics within the minority data points and creating new instances that, while similar, exhibit minor variations, resulting in a more evenly distributed dataset for better medical research outcomes. Importantly, SMOTE was used only on images in the training set, thereby improving the training capabilities of the model without hampering its evaluation.

Another limitation of our study stems from the deliberate omission of samples with Hgb values between 11 and 12 g/dL (GP3 = 97/531, GP6 = 25/256). While this exclusion limits the clinical application of the algorithm, it was taken in consideration of the WHO recommendations, which demarcate anemia at 11.0 g/dL, 11.5 g/dL, or 12.0 g/dL for different pediatric populations. As there were not enough data to create age- and gender-specific algorithms, the 11–12 g/dL range was excluded. With additional data collection, we could establish models divided by age and gender, and their corresponding clinical cutoffs for anemia. We could also build and train independent models for different skin tones. This would facilitate the inclusion of samples within the aforementioned Hb range, thereby enhancing the model's inclusivity and accuracy.

Conclusions

Our research offers promising insights that lay the groundwork for the future evolution of non-invasive, pain-free, and accessible anemia screening tools tailored specifically for pediatric patients. Our approach harnesses state-of-the-art technology to potentially enable care equity regardless of gender, skin tone, or age. Given the sensitivity of this population to pain and the profound impact anemia can have on their development, the advantages of a non-invasive classification pipeline are multifaceted. This method mitigates the discomfort and fear often associated with invasive CBCs, enhancing the patient experience, adherence to screening protocols, and, consequently, early detection rates. Our results also indicated that the performance of anemia detection algorithms was sensitive to image quality, as defined by pixel density. Therefore, a potential avenue for future research lies in optimizing these algorithms for diverse imaging capabilities of various smartphone models, as current models performed well exclusively with images

captured by GP6, but not GP3. As we move forward, we aim to address the limitations of our current model, broaden its applicability, and thereby democratize access to efficient and painless anemia screening, potentially transforming the prognosis for children worldwide.

Acknowledgements: The authors would like to thank all patient volunteers and their parents/guardians for taking part in the study. We would also like to acknowledge the clinical staff and study coordinator, Mrs. Rotem Shvartzman, for their role in gathering the clinical samples.

Contributorship: DS and MB had full access to all the data in the study and took responsibility for the integrity of the data and the accuracy of the data analysis. DS, JH, KG, YB, AL, TL, HL, XL, SP, and MB all took part in the data analysis.

Declaration of conflicting interests: The authors declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: DS, JH, KG, YB, AL, TL, and MB report personal fees from MyOr Diagnostics Ltd. during the conduct of the study.

Ethical approval: Written informed consent was obtained by parents/guardians of enrolled participants according to the approved protocols and procedures received from the Helsinki Committee of the Haemek Medical Center (study name: HEMO-AI, approval number 0150-20-EMC). The study was prospectively registered on www.clinicaltrials.gov (Identifier: NCT04573244) on 15 September 2020, prior to subject recruitment.

Funding: The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was funded by MyOr Diagnostics Ltd. (Grant No. 04573244), which played an active role in the design and analysis of the study.

ORCID iD: Michael Brandwein  <https://orcid.org/0000-0003-4271-3837>

Supplemental material: Supplemental material for this article is available online.

Reference

1. U.S. Food and Administration. *Artificial intelligence and machine learning (AI/ML)-enabled medical devices*, 2022, <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices>
2. Chaparro CM and Suchdev PS. Anemia epidemiology, pathophysiology, and etiology in low- and middle-income countries. *Ann N Y Acad Sci* 2019;1450:15–31.
3. McCann JC and Ames BN. An overview of evidence for a causal relation between iron deficiency during development and deficits in cognitive or behavioral function. *Am J Clin Nutr* 2007; 85: 931–945.

4. Beard J. Iron deficiency alters brain development and functioning. *J Nutr* 2003; 133: 1468S–1472S.
5. Gallagher PG. Anemia in the pediatric patient. *Blood* 2022; 140: 571–593.
6. Tefferi A (2004, July). Practical algorithms in anemia diagnosis. In *Mayo Clinic Proceedings*. Elsevier, 2004, Vol. 79, No. 7, pp. 955–956.
7. Moscheo C, Licciardello M, Samperi P, et al. New insights into iron deficiency anemia in children: a practical review. *Metabolites* 2022; 12: 289.
8. World Health Organization. *Haemoglobin concentrations for the diagnosis of anaemia and assessment of severity*. 2011. World Health Organization.
9. Wemyss TA, Nixon-Hill M, Outlaw F, et al. Feasibility of smartphone colorimetry of the face as an anaemia screening tool for infants and young children in Ghana. *PLoS One* 2023; 18: e0281736.
10. Cohen AR and Seidl-Friedman J. Hemocue[®] system for hemoglobin measurement: evaluation in anemic and nonanemic children. *Am J Clin Pathol* 1988; 90: 302–305.
11. Causey MW, Miller S, Foster A, et al. Validation of non-invasive hemoglobin measurements using the Masimo Radical-7 SpHb Station. *Am J Surg* 2011; 201: 592–598.
12. Horecker BL. The absorption spectra of hemoglobin and its derivatives in the visible and near infra-red regions. *J Biol Chem* 1943; 148: 173–183.
13. Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016, pp. 779–788.
14. Schumann C, Olanubi GO, Wright A, et al. Consensus and subjectivity of skin tone annotation for ML fairness. arXiv preprint arXiv:230509073. 2023. <https://doi.org/10.48550/arXiv.2305.09073>
15. Chen T and Guestrin C. XGBoost: a scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery, San Francisco, CA, USA, 13–17 August 2016, pp. 785–794.
16. Chawla NV, Bowyer KW, Hall LO, et al. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res* 2002; 16: 321–357.
17. Lobbes H, Dehos J, Pereira B, et al. Computed and subjective blue scleral color analysis as a diagnostic tool for iron deficiency: a pilot study. *J Clin Med* 2019; 8: 1876.
18. Mitani A, Huang A, Venugopalan S, et al. Detection of anaemia from retinal fundus images via deep learning. *Nat Biomed Eng* 2020; 4: 18–27.
19. Dimauro G, Caivano D and Girardi F. A new method and a non-invasive device to estimate anemia based on digital images of the conjunctiva. *IEEE Access* 2018; 6: 46968–46975.
20. Mannino RG, Myers DR, Tyburski EA, et al. Smartphone app for non-invasive detection of anemia using only patient-sourced photos. *Nat Commun* 2018; 9: 4924.
21. Hasan MK, Sakib N, Love RR, et al. RGB pixel analysis of fingertip video image captured from sickle cell patient with low and high level of hemoglobin. Proceedings of the 2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON), New York, NY, USA, 19–21 October 2017, pp. 499–505.
22. Sjoding MW, Dickson RP, Iwashyna TJ, et al. Racial bias in pulse oximetry measurement. *N Engl J Med* 2020; 383: 2477–2478.