Rev. Cient. Sist. Inform. 5(2), e955, doi: 10.51252/rcsi.v5i2.955

UNIVERSIDAD NACIONAL DE SAN MARTIN

RCSI

https://revistas.unsm.edu.pe/index.php/rcsi

e-ISSN: 2709-992X

Data Paper

Artículo de datos

Jul-Dec, 2025

Non-invasive multimodal dataset for the detection of iron deficiency anemia in young adults: fingertip videos, palm videos, and nail photographs

Conjunto de datos multimodal no invasivo para la detección de anemia por deficiencia de hierro en jóvenes adultos: videos de yema del dedo, palma de la mano y fotografías de uñas

Valles-Coral, Miguel Angel[1*]
Injante, Richard[1]
Navarro-Cabrera, Jorge Raul[1]
Pinedo, Lloy[1]
Salazar-Ramírez, Luis[1]
Farro-Roque, María Elena[1]
Quintanilla-Morales, Luz Karen[1]

[1]Universidad Nacional de San Martín, Peru

Received: 14 Apr. 2025 | **Accepted:** 10 Jul. 2025 | **Published:** 20 Jul. 2025

Corresponding author*: mavalles@unsm.edu.pe

How to cite this article: Valles-Coral, M.A., Injante, R., Navarro-Cabrera, J.R., Pinedo, L., Salazar-Ramírez, L., Farro-Roque, M.E., Quintanilla-Morales, L.K. (2025). Non-invasive multimodal dataset for the detection of iron deficiency anemia in young adults: fingertip videos, palm videos, and nail photographs. *Revista Científica de Sistemas e Informática, 5*(2), e955. https://doi.org/10.51252/rcsi.v5i2.955

ABSTRACT

Iron deficiency anemia affects a significant proportion of the young population in both rural and urban areas of Peru. In response to the need for non-invasive, accessible, and reproducible methods for its detection, we developed this dataset as part of a research project funded by the Universidad Nacional de San Martín, which applies computer vision techniques to automatically classify patients as anemic or non-anemic. The aim is to provide a standardized base of videos and images that supports the development and validation of classification and regression models to estimate hemoglobin levels without the need for blood extraction. This data paper presents a multimodal dataset composed of non-invasive visual records collected to facilitate the detection of iron deficiency anemia in young adults through machine learning models. The dataset includes 909 fingertip videos, 909 palm videos (with controlled hand opening), and 909 nail photographs, all linked to individual clinical data such as age, sex, hemoglobin level, and symptomatology.

Keywords: artificial intelligence; biomedical videos; clinical dataset; computer vision; hemoglobin; machine learning; non-invasive detection

RESUMEN

La anemia por deficiencia de hierro afecta a una proporción considerable de la población joven en zonas rurales y urbanas del Perú. En respuesta a la necesidad de métodos no invasivos, accesibles y reproducibles para su detección, desarrollamos este dataset como parte de un proyecto de investigación financiado por la Universidad Nacional de San Martín, el cual aplica técnicas de visión artificial para clasificar automáticamente a los pacientes como anémicos o no anémicos. El objetivo es proveer una base estandarizada de videos e imágenes que permita desarrollar y validar modelos de clasificación y regresión para estimar el nivel de hemoglobina sin necesidad de extracción sanguínea. Este data paper presenta un dataset multimodal compuesto por registros visuales no invasivos, recolectados con el propósito de facilitar la detección de anemia por deficiencia de hierro en jóvenes adultos mediante modelos de aprendizaje automático. El conjunto incluye 909 videos de la yema del dedo, 909 videos de la palma de la mano (con apertura controlada del puño) y 909 fotografías de las uñas, todos vinculados a datos clínicos individuales como edad, sexo, nivel de hemoglobina y sintomatología.

Palabras clave: inteligencia artificial; vídeos biomédicos; conjunto de datos clínicos; visión artificial; hemoglobina; aprendizaje automático; detección no invasiva

1. INTRODUCTION

Globally, 29.9% of women and 39.8% of children are affected by anemia (World Health Organization, 2021), primarily caused by iron deficiency (Del Castillo et al., 2023). This condition leads to serious health and economic consequences, especially in developing countries (Prieto-Patron et al., 2020). Another susceptible and less studied population is university students, who, due to high academic and social demands, experience reduced productivity and physical health (Alkhaldy et al., 2020; Khani Jeihooni et al., 2021; Quiliche Castañeda et al., 2021). Early detection of anemia is therefore crucial to safeguard quality of life. However, the lack of access to rapid, accurate, affordable, and low-cost diagnosis hinders timely care (Perez-Plazola et al., 2020), and in some cases, the invasive nature of blood tests (finger pricks) leads to rejection (An et al., 2021).

In response, non-invasive detection of anemia using techniques such as spectrophotometry, colorimetric analysis with digital cameras, and pulse oximetry—combined with machine learning algorithms—has shown acceptable accuracy levels. For instance, (Williams Asare et al., 2023) reported a 99.79% accuracy using Convolutional Neural Networks (CNN) to analyze the palms of children aged 5 to 59 months.

Nevertheless, current scientific literature reveals a lack of studies focused on non-invasive detection of iron deficiency anemia in university students. Therefore, we propose this dataset as a resource for developing smartphone-based solutions—using mid-range devices—that enable non-invasive anemia detection in this population. The dataset was collected as part of a research project funded by the Universidad Nacional de San Martín. It applies computer vision techniques to automatically classify patients as anemic or non-anemic and provides a standardized base of videos and images for developing and validating classification and regression models to estimate hemoglobin levels without the need for blood extraction.

2. MATERIALS AND METHODS

2.1. Data collection

The data collection process was conducted at the University Medical Center of the Universidad Nacional de San Martín, Peru. A total of 909 voluntary participants, aged between 18 and 25 years (540 women and 369 men), were evaluated by trained professionals (licensed nurses, nutritionists, and general physicians). Hemoglobin (Hb) levels were measured using the non-invasive Rad-67 device (Figure 1a), while visual samples were simultaneously recorded using a Samsung Galaxy A73 5G smartphone.

2.2. Types of Visual Records

For each participant, the following records were obtained:

- A 30-second video of the right index fingertip, captured at 60 frames per second with fixed parameters: ISO 250, shutter speed 1/60, focus 0.0, and white balance at 4400K (Figure 1b).
- A video of the palm of the hand, in which the participant begins with a closed fist and progressively opens it (Figure 1c).
- A photograph of the fingernails taken under controlled lighting conditions inside a light box, using specific technical parameters: aperture f/1.8 to allow adequate light entry, shutter speed

1/180 to prevent blurring, ISO 50 to minimize digital noise, and a 5.06 mm lens to capture the nail surface in detail (Figure 1d).

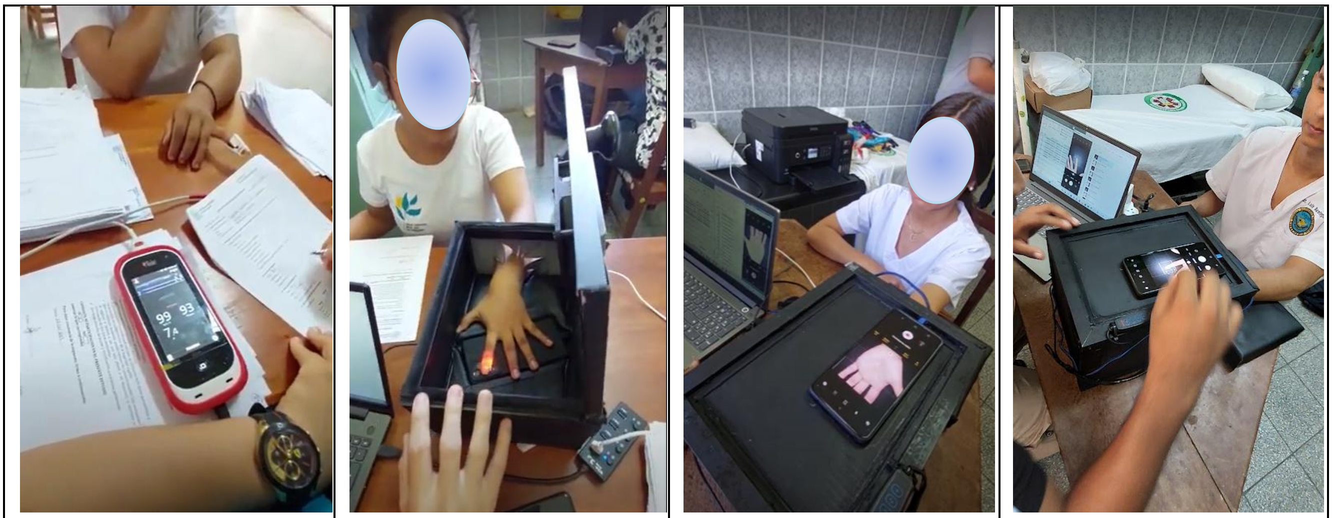
In all cases, the camera was remotely controlled using scrcpy software, which ensured standardized lighting and capture angles.

2.3. Data Organization and Storage

All samples were stored in a centralized database. The information corresponding to each file and participant was initially organized in a spreadsheet (`.xlsx` format). To optimize access and data management, a dictionary-type data structure named `data_by_dni` was created, where each participant's National Identity Document (DNI) served as the primary key. Five elements were associated with each key: (1) video files, (2) sex, (3) reported symptoms, (4) weight-to-height ratio, and (5) classification label (anemic/non-anemic).

2.4. Clinical Category Assignmen

Regarding the category attribute, it is important to clarify that the class labels assigned (Normal, Mild, Moderate, Severe) were not arbitrarily defined. These categories were established through a formal consultation with the Blood Bank of the city of Tarapoto, which took into account contextual factors such as local altitude and other region-specific physiological conditions. Therefore, the classification thresholds used reflect clinically validated standards that are specifically adapted to our target population.



1(a) 1(b) 1(c) 1(d)

Figure 1. Illustrative panel of the visual records included in the dataset: Rad-67 device (1a), fingertip video (1b), palm video (1c), and nail photograph (1d).

3. STRUCTURE AND CONTENT OF THE DATASET

The dataset is organized as follows:

```
dataset/
├── palma/
│   └── leve
│   └── moderada
│   └── normal
├── unas/
│   └── leve
```

```
│    └── moderada
│    └── normal
├── yemas/
│    └── leve
│    └── moderada
│    └── normal
└── metadata.csv
```

- Each subfolder contains multiple image or video files named with an anonymous identifier (e.g., ID001.jpg, ID002.jpg, ..., IDNNN.mp4, ID001.mp4, ID002.mp4, ..., IDNNN.mp4).

- The metadata.csv file includes the following columns: [ID, Sexo, Edad, Fatiga, Debilidad, Latidos Irregulares, Dificultad Para Respirar, Mareos O Aturdimiento, Dolor En El Pecho, Manos Y Pies Fríos, Dolores De Cabeza, Normal, Leve Moderado, Frecuencia Cardiaca, Oxigeno, Peso, Talla, Yemas, Palmas, Unas]. It should be noted that all column names are in Spanish. In the columns related to symptoms or clinical categories, values are encoded in binary format: 0 indicates absence and 1 indicates presence of the recorded condition.

Dataset size and characteristics: A total of 909 participants were recorded, distributed as follows:

Table 1. Distribution of the collected samples

	Normal	Mild	Moderate	Severe	Anemia (%)
Femenine	287	179	74	0	46.85
Masculine	335	33	1	0	9.21

In the articles published by our team (Navarro-Cabrera et al., 2025; Valles-Coral et al., 2024), this same dataset was used to train and validate models for non-invasive anemia detection. However, due to the complete absence of samples in the "Severe" category, only the "Normal," "Mild," and "Moderate" classes were considered. This imbalance in the number of samples prompted the application of a subsampling technique, whose rationale and results are detailed in each article. The class with the fewest samples was used as the reference for proportional selection (Chakraborty et al., 2021)

Important: The published dataset includes only the original videos and photographs collected during the data acquisition process, without any preprocessing. Frame extraction, anatomical region segmentation, or implementation of machine learning models were conducted in subsequent methodological studies and are not part of the publicly available dataset.

4. DATA ACCESS

The dataset is available for download from the data archive on Google Drive managed by the research team.

5. STUDIES DEVELOPED WITH THIS DATASET

This dataset has been used in various previous studies aimed at exploring innovative approaches to the non-invasive detection of iron deficiency anemia. Below is a brief summary of two relevant studies:

- **Fingertip video analysis**: In *Valles-Coral et al. (2024)*, a convolutional neural network (CNN)-based model was implemented using fingertip videos. The model achieved over 90% accuracy in binary classification (anemia/non-anemia), demonstrating the feasibility of optical analysis of peripheral microcirculation.
- **Classification based on nail images**: In *Navarro-Cabrera et al. (2025)*, a computer vision model was trained on nail images. The study reached an average accuracy of 87.5% in multi-level classification (normal, mild, moderate), highlighting the clinical relevance of visible nail alterations.

These results validate the usefulness of the dataset for developing non-invasive diagnostic tools applicable in resource-limited settings. Both articles are available in indexed scientific journals (*iJOE and Frontiers in Big Data*) and cite this dataset as a primary source.

6. ETHICAL CONSIDERATION

Although the national identification number (DNI) was initially used to systematically organize the files, an anonymization process was implemented by renaming the files after their initial classification. This practice reflects a responsible approach that balances data integrity with the protection of sensitive personal information.

The management of clinical data —including videos, biometric information (weight and height), reported symptoms, and hemoglobin levels— was carried out following protocols that ensure participant confidentiality. These measures, combined with the removal of personal identifiers in later processing stages, ensure compliance with ethical standards for research in health and technological development, safeguarding the privacy and anonymity of participants.

FUNDING

This work was funded by the Universidad Nacional de San Martín as part of the research project titled "HemoTupunaApp: Non-invasive clinical detection of iron deficiency anemia using smartphones", approved by University Council Resolution No. 1063-2022-UNSM/CU-R.

CONFLICT OF INTERES

The authors declare that they have no conflicts of interest related to the development of the present study.

REFERENCES

Alkhaldy, H., Hadi, R., Alghamdi, K., Alqahtani, S., Al Jabbar, I. H., Al Ghamdi, I., Bakheet, O. E., Saleh, R. M., Shehata, S., & Aziz, S. (2020). The pattern of iron deficiency with and without anemia among medical college girl students in high altitude southern Saudi Arabia. *Journal of Family Medicine and Primary Care*, *9*(9), 5018.

https://doi.org/10.4103/jfmpc.jfmpc_730_20

An, R., Huang, Y., Man, Y., Valentine, R. W., Kucukal, E., Goreke, U., Sekyonda, Z., Piccone, C., Owusu-Ansah, A., Ahuja, S., Little, J. A., & Gurkan, U. A. (2021). Emerging point-of-care technologies for anemia detection. *Lab on a Chip*, *21*(10), 1843–1865. https://doi.org/10.1039/D0LC01235A

Chakraborty, J., Majumder, S., & Menzies, T. (2021). Bias in machine learning software: why? how? what to do? *Proceedings of the 29th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, 429–440. https://doi.org/10.1145/3468264.3468537

Del Castillo, L., Cardona-Castro, N., Whelan, D. R., Builes, J. P., Serrano-Coll, H., Arboleda, M., & Leon, J. S. (2023). Prevalence and risk factors of anemia in the mother–child population from a region of the Colombian Caribbean. *BMC Public Health*, *23*(1), 1533. https://doi.org/10.1186/s12889-023-16475-0

Khani Jeihooni, A., Hoshyar, S., Afzali Harsini, P., & Rakhshani, T. (2021). The effect of nutrition education based on PRECEDE model on iron deficiency anemia among female students. *BMC Women's Health*, *21*(1), 256. https://doi.org/10.1186/s12905-021-01394-2

Navarro-Cabrera, J. R., Valles-Coral, M. A., Farro-Roque, M. E., Reátegui-Lozano, N., & Arévalo-Fasanando, L. (2025). Machine vision model using nail images for non-invasive detection of iron deficiency anemia in university students. *Frontiers in Big Data*, *8*. https://doi.org/10.3389/fdata.2025.1557600

Perez-Plazola, M. S., Tyburski, E. A., Smart, L. R., Howard, T. A., Pfeiffer, A., Ware, R. E., Lam, W. A., & McGann, P. T. (2020). AnemoCheck-LRS: an optimized, color-based point-of-care test to identify severe anemia in limited-resource settings. *BMC Medicine*, *18*(1), 337. https://doi.org/10.1186/s12916-020-01793-6

Prieto-Patron, A., V. Hutton, Z., Fattore, G., Sabatier, M., & Detzel, P. (2020). Reducing the burden of iron deficiency anemia in Cote D'Ivoire through fortification. *Journal of Health, Population and Nutrition*, *39*(1), 1. https://doi.org/10.1186/s41043-020-0209-x

Quiliche Castañeda, R. B., Turpo-Chaparro, J., Torres, J. H., Saintila, J., & Ruiz Mamani, P. G. (2021). Overweight and Obesity, Body Fat, Waist Circumference, and Anemia in Peruvian University Students: A Cross-Sectional Study. *Journal of Nutrition and Metabolism*, *2021*, 1–9. https://doi.org/10.1155/2021/5049037

Valles-Coral, M. A., Navarro-Cabrera, J. R., Pinedo, L., Injante, R., Quintanilla-Morales, L. K., & Farro-Roque, M. E. (2024). Non-Invasive Detection of Iron Deficiency Anemia in Young Adults Through Finger-Tip Video Image Analysis. *International Journal of Online and Biomedical Engineering (IJOE)*, *20*(14), 53–70. https://doi.org/10.3991/ijoe.v20i14.50141

Williams Asare, J., Appiahene, P., Timmy Donkoh, E., & Dimauro, G. (2023). Iron deficiency anemia detection using machine learning models: A comparative study of fingernails, palm and conjunctiva of the eye images. *Engineering Reports*, e12667. https://doi.org/10.1002/ENG2.12667

World Health Organization. (2021). *Anaemia in women and children. The Global Healt Observatory*.