



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

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

Valles-Coral, Miguel Angel^{1*}

Injante, Richard¹

Navarro-Cabrera, Jorge Raúl¹

Pinedo, Lloy¹

Salazar-Ramírez, Luis¹

Farro-Roque, María Elena¹

Quintanilla-Morales, Luz Karen¹

¹Universidad Nacional de San Martín, Perú

Recibido: 14 Abr. 2025 | **Aceptado:** 10 Jul. 2025 | **Publicado:** 20 Jul. 2025

Autor de correspondencia*: mavalles@unsm.edu.pe

Cómo citar este artículo: Valles-Coral, M.A., Injante, R., Navarro-Cabrera, J.R., Pinedo, L., Salazar-Ramírez, L., Farro-Roque, M.E., Quintanilla-Morales, L.K. (2025). 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. *Revista Científica de Sistemas e Informática*, 5(2), e955. <https://doi.org/10.51252/rcsi.v5i2.955>

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; videos biomédicos; conjunto de datos clínicos; visión artificial; hemoglobina; aprendizaje automático; detección no invasiva

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



1. INTRODUCCIÓN

A nivel mundial el 29.9% de las mujeres y el 39.8% de los niños tiene anemia (World Health Organization, 2021) ocasionada principalmente por deficiencia de hierro (Del Castillo et al., 2023), lo cual genera graves consecuencias sanitarias y económicas, sobre todo en los países en desarrollo (Prieto-Patron et al., 2020). Otra población susceptible y menos estudiada son los estudiantes universitarios quienes al experimentar alta carga académica y social reducen su productividad y salud física (Alkhalidy et al., 2020; Khani Jeehooni et al., 2021; Quiliche Castañeda et al., 2021). Es entonces que la detección temprana de anemia es crucial para salvaguardar la calidad de vida de la población; sin embargo, la falta de acceso a un diagnóstico rápido, preciso, accesible y de bajo costo impide la atención adecuada (Perez-Plazola et al., 2020), y en otros casos el método invasivo (pinchazos) genera rechazo (An et al., 2021).

Frente a ello, la detección no invasiva de anemia utilizando técnicas como espectrofotometría, análisis colorimétrico con cámara digital y oximetría de pulso, complementado con el uso de algoritmos de aprendizaje automático, son técnicas actuales que han demostrado niveles aceptables de precisión, como evidencian (Williams Asare et al., 2023) al reportar 99.79% de precisión empleando CNN (Red Neuronal Convolucional) en la palma palpable de niños entre 5 y 59 meses.

No obstante, en la literatura científica actual no se identifican estudios orientados a la detección no invasiva de anemia ferropénica en estudiantes universitarios. Por ello, proponemos este conjunto de datos como insumo para el desarrollo de soluciones basadas en smartphone de gama media, que permitan detectar anemia de manera no invasiva en esta población. Este dataset fue obtenido como parte de un proyecto de investigación financiado por la Universidad Nacional de San Martín. Aplica técnicas de visión artificial para clasificar automáticamente a los pacientes como anémicos o no anémicos, y provee una base estandarizada de videos e imágenes que permitirá desarrollar y validar modelos de clasificación y regresión para estimar el nivel de hemoglobina sin necesidad de extracción sanguínea.

2. MATERIALES Y MÉTODOS

2.1. Recolección de datos

La recolección de datos se llevó a cabo en el Centro Médico Universitario de la Universidad Nacional de San Martín, Perú. Un total de 909 participantes voluntarios, entre 18 y 25 años de edad (540 mujeres y 369 varones), fueron evaluados por personal profesional capacitado (Licenciados en enfermería, bromatólogos y médicos generales). La medición de hemoglobina (Hb) se realizó mediante el dispositivo no invasivo Rad-67 (Figura 1a), y de forma paralela se registraron muestras visuales utilizando un smartphone Samsung Galaxy A73 5G.

2.2. Tipos de registros visuales

Para cada participante se obtuvieron:

- Un video de **30 segundos de la yema del dedo índice derecho**, capturado a **60 fotogramas por segundo** con parámetros fijos: ISO 250, velocidad de obturación 1/60, enfoque 0.0 y balance de blancos 4400K (Figura 1b).

- Un video de **30 segundos palma de la mano** en el que el participante inicia con el puño cerrado y luego lo abre progresivamente (Figura 1c).
- Una fotografía de las uñas capturada en condiciones controladas dentro de una caja con luz, utilizando parámetros técnicos específicos: apertura $f/1.8$ para permitir una entrada adecuada de luz, velocidad de obturación $1/180$ para evitar desenfoques, ISO 50 para reducir el ruido digital, y una lente de 5.06 mm que permitió captar con precisión la superficie ungueal (Figura 1d).

Para todos los casos, la cámara fue controlada remotamente mediante el software `scrpy`, lo que garantizó la estandarización en iluminación y ángulo de captura.

2.3. Organización y almacenamiento de datos

Todas las muestras fueron almacenadas en una base de datos centralizada. La información correspondiente al archivo y al individuo fue organizada inicialmente en un archivo en formato `.xlsx`. Para optimizar el acceso y organización, se construyó una estructura de datos tipo diccionario denominada `data_by_dni`, donde el DNI de cada participante funcionó como clave principal. A cada clave se asociaron cinco elementos: (1) archivos de video, (2) sexo, (3) síntomas reportados, (4) relación peso-talla, y (5) etiqueta de clasificación (anémico/no anémico).

2.4. Asignación de categorías clínicas

En cuanto al atributo categoría, cabe precisar que las etiquetas de clase asignadas (Normal, Leve, Moderada, Severa) no fueron definidas arbitrariamente. Estas fueron establecidas tras una consulta formal con el Banco de Sangre de la ciudad de Tarapoto, que consideró factores contextuales como la altitud local y otras condiciones fisiológicas propias de la región. Por tanto, los umbrales de clasificación utilizados reflejan estándares clínicos validados y adaptados específicamente a nuestra población objetivo.

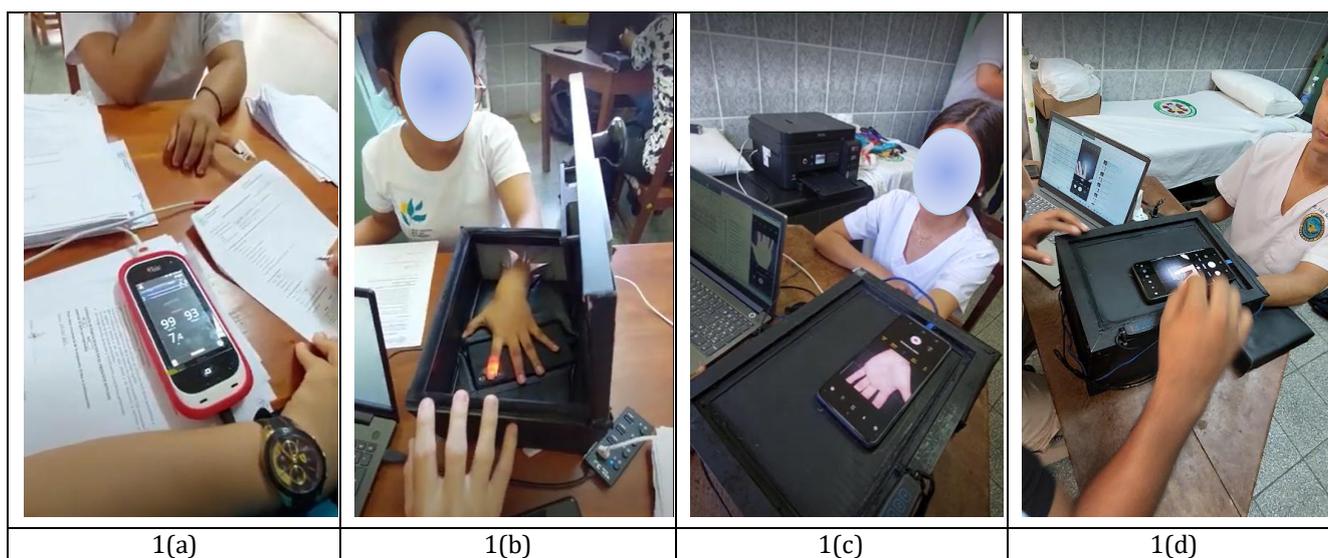


Figura 1. Panel ilustrativo de los registros visuales incluidos en el dataset: Rad-67 (1a), video de yema del dedo (1b), video de palma de la mano (1c) y fotografía de uñas (1d)

3. ESTRUCTURA Y CONTENIDO DEL CONJUNTO DE DATOS

El dataset está organizado de la siguiente manera:

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

- Cada subcarpeta contiene varios archivos de imagen o vídeo nombrados con una identificación anónima (e.g., ID001.jpg, ID002.jpg, ..., IDNNN.mp4, ID001.mp4, ID002.mp4, ..., IDNNN.mp4).
- El archivo metadata.csv contiene las siguientes columnas: **[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]**. Cabe precisar que todos los nombres de columnas están en español. En las columnas correspondientes a síntomas o categorías clínicas, los valores están codificados de forma binaria: 0 indica ausencia y 1 indica presencia de la condición registrada.

Tamaño y características del dataset: En total, se recopilaron datos de 909 participantes, distribuidos de la siguiente manera:

Tabla 1. Distribución de las muestras obtenidas

	Normal	Leve	Moderada	Severa	Anemia (%)
Femenino	287	179	74	0	46.85
Masculino	335	33	1	0	9.21

En los artículos publicados por nuestro equipo (Navarro-Cabrera et al., 2025; Valles-Coral et al., 2024), se empleó este mismo dataset para entrenar y validar modelos de detección no invasiva de anemia. Sin embargo, debido a la escasez total de muestras en la categoría “Severa”, se optó por trabajar únicamente con las clases “Normal”, “Leve” y “Moderada”. Este desbalance en la cantidad de muestras motivó la aplicación de una técnica de submuestreo, cuya lógica y resultados se detallan en cada artículo. La clase con menor número de elementos fue tomada como referencia para la selección proporcional (Chakraborty et al., 2021).

Importante: El dataset publicado contiene exclusivamente los videos originales y las fotografías obtenidas durante el proceso de recolección de datos, sin ningún tipo de preprocesamiento. La extracción de frames, segmentación de regiones anatómicas o implementación de modelos de aprendizaje automático fue desarrollada posteriormente en estudios metodológicos asociados, y no forma parte del dataset disponibilizado.

4. ACCESO A LOS DATOS

El conjunto de datos está disponible para descarga en el archivo de datos en [Google Drive gestionado por el equipo de investigación](#).

Licencia: CC BY 4.0 International.

5. ESTUDIOS DESARROLLADOS CON ESTE DATASET

Este conjunto de datos ha sido utilizado en diversas investigaciones previas con el objetivo de explorar enfoques innovadores en la detección no invasiva de anemia ferropénica. A continuación, se resumen brevemente dos estudios relevantes:

- **Análisis de video de yema del dedo:** En (Valles-Coral et al., 2024), se implementó un modelo basado en redes neuronales convolucionales (CNN) utilizando los videos de la yema del dedo. Se obtuvo una precisión superior al 90% en la clasificación binaria (anemia/no anemia), demostrando la viabilidad del análisis óptico de microcirculación periférica.
- **Clasificación a partir de imágenes de uñas:** En (Navarro-Cabrera et al., 2025), se entrenó un modelo de visión por computadora con imágenes de uñas. El estudio alcanzó una precisión promedio del 87.5% en la clasificación multinivel (normal, leve, moderada), destacando el valor clínico de las alteraciones ungueales visibles.

Estos resultados validan la utilidad del dataset para el desarrollo de herramientas diagnósticas no invasivas aplicables en contextos con recursos limitados. Ambos artículos se encuentran disponibles en revistas científicas indexadas (*ijOE* y *Frontiers in Big Data*) y citan este dataset como fuente primaria.

6. CONSIDERACIONES ÉTICAS

Aunque inicialmente se utilizó el DNI como identificador para organizar los archivos de manera sistemática, se implementó un proceso de anonimización mediante el renombramiento de los archivos tras su clasificación inicial. Esta práctica demuestra un enfoque responsable que equilibra la integridad de los datos con la protección de información personal sensible.

El manejo de datos clínicos —incluyendo videos, información biométrica (peso y talla), síntomas reportados y niveles de hemoglobina— se realizó siguiendo protocolos que aseguran la confidencialidad de los participantes. Estas medidas, combinadas con la eliminación de identificadores personales en las etapas posteriores del procesamiento, permiten cumplir con los estándares de ética en investigación en salud y desarrollo tecnológico, garantizando la privacidad y el anonimato de los participantes.

FINANCIAMIENTO

Este trabajo fue financiado por la Universidad Nacional de San Martín en el marco del proyecto de investigación titulado “HemoTupunaApp: Detección clínica no invasiva de anemia ferropénica utilizando smartphones”, aprobado mediante Resolución del Consejo Universitario N.º 1063-2022-UNSM/CU-R.

CONFLICTO DE INTERESES

Los autores declaran que no tienen conflictos de interés relacionados con el desarrollo del presente estudio.

REFERENCIAS

- Alkhalidy, 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 Health Observatory*.