

Ciencia Latina Revista Científica Multidisciplinar, Ciudad de México, México.
ISSN 2707-2207 / ISSN 2707-2215 (en línea), enero-febrero 2025,
Volumen 9, Número 1.

https://doi.org/10.37811/cl_rcm.v9i1

DESARROLLO DE UNA HERRAMIENTA PARA ESTIMAR MACRONUTRIENTES EN LECHE MATERNA BASADO EN APRENDIZAJE DE MÁQUINA

DEVELOPMENT OF A TOOL TO ESTIMATE BREAST MILK MACRONUTRIENTS BASED ON MACHINE LEARNING

Samara Acosta Jimenez

Universidad Autónoma de Zacatecas, México

Miguel Mauricio Mendoza Mendoza

Universidad Autónoma de Zacatecas, México

Carlos Eric Galván Tejada

Universidad Autónoma de Zacatecas, México

Carlos Mario Cázarez de Lira

Hospital General Fresnillo Dr. José Haro Ávila, México

Irma E. González-Curiel

Universidad Autónoma de Zacatecas, México

DOI: https://doi.org/10.37811/cl_rcm.v9i1.16418

Desarrollo de una Herramienta para Estimar Macronutrientes en Leche Materna Basado en Aprendizaje de Máquina

Samara Acosta Jimenez¹samaracosta@uaz.edu.mx<https://orcid.org/0000-0002-1724-1172>Universidad Autónoma de Zacatecas
Zacatecas, México**Miguel Mauricio Mendoza Mendoza**mauricio.mendoza@uaz.edu.mx<https://orcid.org/0009-0009-2192-3350>Universidad Autónoma de Zacatecas
Zacatecas, México**Carlos Eric Galván Tejada**ericgalvan@uaz.edu.mx<https://orcid.org/0000-0002-7635-4687>Universidad Autónoma de Zacatecas
Zacatecas, México**Carlos Mario Cázarez de Lira**cazali0606@gmail.comHospital General Fresnillo Dr. José Haro Ávliá
Fresnillo, México**Irma E. González Curiel**irmacuriel@uaz.edu.mx<https://orcid.org/0000-0002-0044-9196>Universidad Autónoma de Zacatecas
Zacatecas, México

RESUMEN

Este estudio se enfoca en el desarrollo de un sistema basado en Machine Learning para estimar la composición de macronutrientes en leche materna, dirigiéndose a hospitales con limitación de recursos. La leche materna es esencial para la nutrición neonatal, y su composición varía de acuerdo a las diferentes fases de lactancia. Sin embargo, la falta de equipo especializado en medición de estos nutrientes que podrían impactar negativamente en la nutrición personalizada, especialmente en infantes prematuros o nacidos con pesos bajos. Los datos fueron recolectados de 333 muestras, incluyendo variables como grados Dornic, pH, y niveles macronutrientales. Se implementan varios modelos de regresión como, regresión lineal, regresión Ridge, regresión Lasso, ElasticNet, árboles de decisión, bosques aleatorios y XGBoost, y mediante validación cruzada, identificar el modelo más efectivo. Asimismo, se desarrolla una interfaz gráfica creada con PyQt5, la cual permite entrada y visualización clara de resultados, facilitando el uso por el personal médico, incluyendo un sistema de recomendación básico para identificar al tipo de paciente al que se estima brindar esa leche. A pesar de algunas limitaciones, como la mayor accesibilidad y la incorporación de datos adicionales, el sistema de atención neonatal promete mejorar la nutrición en clínicas con recursos limitados.

Palabras clave: leche materna, macronutrientes, machine learning

¹ Autor principal

Correspondencia: cazali0606@gmail.com

Development of a tool to estimate breast milk macronutrients based on Machine Learning

ABSTRACT

This study focuses on the development of a machine learning-based system to estimate macronutrients composition in breast milk, targeting resource-limited hospital settings. Breast milk is essential for neonatal nutrition, and its composition varies throughout the different stages of lactation. However, the lack of specialized equipment to measure these nutrients can negatively impact personalized nutrition, especially in premature or low birth weight infants. Data are collected from 333 patients, including variables such as Dornic grades, pH, and macronutrient levels. Various regression models are implemented, such as Linear Regression, Ridge Regression, Lasso Regression, ElasticNet, Decision trees, Random forests, and XGBoost, using cross-validation to identify the most effective model. In addition, a graphical interface is developed, and created in PyQt5, which allows data entry and clear visualization of results, facilitating its use by medical staff, it also includes a basic recommendation system to identify the type of patient as estimated. Despite some limitations, such as the need for greater accessibility and the incorporation of additional data, the system represents an important advance in neonatal care, promising to improve nutrition in resource-limited clinics

Keywords: breast milk, macronutrients, machine learning

*Artículo recibido 10 diciembre 2025
Aceptado para publicación: 15 enero 2024*



INTRODUCCIÓN

La leche materna es la mejor fuente de nutrición para infantes y recién nacidos, dando una mezcla balanceada de macronutrientes esenciales (proteínas, grasas y carbohidratos), así como inmunoglobulinas y otros factores de protección (Victora et al., 2016). De acuerdo con la Organización Mundial de la Salud (OMS), se recomienda alimentar solo con leche materna durante los primeros seis meses de vida, enfocándose en la importancia del desarrollo del infante y la prevención de padecimientos (WHO, 2019). En temas hospitalarios, especialmente en unidades neonatales, la leche materna es un recurso crucial que no solo nutre a los recién nacidos, sino que promueve su desarrollo y su salud a largo plazo. Sin embargo, en entornos hospitalarios, medir con precisión los macronutrientes en leche materna donada puede ser un reto debido a la falta de equipo médico especializado. Esta limitación puede impactar directamente en la nutrición personalizada de los neonatos, particularmente en candidatos a cuidados intensivos (Vázquez-Román et al., 2014) (Oysal & Özkaya, 2022). La composición de la leche materna varía conforme al tiempo, desde pre-calostro a leche madura, convirtiéndose en un aspecto crítico el identificar y cuantificar las concentraciones de nutrientes en cada etapa (Hassiotou et al., 2013). Esto es vital para los infantes prematuros o de peso bajo al nacer, los cuales son dependientes de ingesta nutricional específica de acuerdo a sus necesidades (Ramel & Belfort, 2022) (Fenton et al., 2022). La ausencia de dispositivos médicos avanzados para el análisis macronutricional de la leche materna es un reto para muchos hospitales. Estos análisis suelen requerir equipo sofisticado y costoso como puede ser el espectrofotómetro, el cual no siempre está disponible en centros de salud con limitaciones, sin embargo la recolección y uso de datos está siendo emergente para superar estas limitantes (Shaikh et al., 2022). Esta situación crea una necesidad urgente de alternativas que puedan estimar los macronutrientes utilizando variables accesibles y datos clínicos como el tipo de leche, los grados Dornic y el pH. El desarrollo de una herramienta basada en Machine Learning (ML) mejora la precisión en la administración de leche materna, contribuyendo a la optimización de la nutrición del neonato y facilitando un control en su crecimiento y desarrollo en esta crítica etapa de su vida (Mari Soria, 2023). Para el manejo de este problema, la investigación se dirige al desarrollo de una herramienta basada en ML la cual estima la composición macronutricional de la leche materna, utilizando datos de fácil acceso como el tipo de leche (calostro, leche transicional, leche



madura y leche madura tardía), grados Dornic y pH. Este sistema contará con una interfaz de diseño simple para su facilidad en el uso por el personal médico. El desarrollo de esta herramienta no solo presenta un avance en la estimación en tiempo real de los macronutrientes presentes en una muestra de leche, sino que añade una solución que puede ser aplicada a través de varios entornos hospitalarios.

METODOLOGÍA

Este estudio se enfoca en el desarrollo de un sistema de estimación para macronutrientes en leche materna utilizando técnicas de ML. Los pasos metodológicos que se siguieron durante la investigación se muestran en la figura 1 y se detallan debajo de ella.

1. Colección de los datos

Los datos recolectados se presentan en 333 muestras, distribuidos en 4 categorías: 100 muestras en calostro, 100 en leche transicional, 100 en leche madura y 33 en leche madura tardía. Para cada muestra, las variables que se recolectaron fueron: grados Dornic, pH, grasa (g/dL), proteína en crudo (g/dL), carbohidratos (CHO)(g/dL), sólidos totales (T), energía (kcal/dL) y proteína verdadera. Esta aproximación asegura una variedad representativa en los datos mostrando las diferentes etapas de la lactancia.

2. Preprocesamiento de los datos

Antes del modelado de los datos, las instancias con información nula o "NA" son removidas para asegurar la integridad de los datos. Asimismo, se categorizó una variable en el proceso de programación, lo cual es esencial para la preparación de los datos. Los algoritmos de ML requieren entradas numéricas; las categorías por tipo de leche (calostro, leche transicional, leche madura y leche madura tardía) son transformadas a un formato numérico utilizando la técnica de "One-Hot encoding", convirtiendo cada categoría en una columna binaria.

3. Desarrollo de la herramienta

En esta fase, se implementan varios algoritmos de ML dirigidos a la regresión, los cuales incluyen regresión lineal, regresión Ridge, regresión Lasso, ElasticNet, árboles de decisión, bosques aleatorios y XGBoost. Cada uno de estos modelos tiene características únicas de regresión y predicción, que dependiendo de la naturaleza de los datos de entrada pueden ser más o menos ideales.

Para evaluar el rendimiento de los modelos propuestos, el conjunto de datos se dividió en conjunto de

entrenamiento y conjunto de prueba, con una proporción de 80% de los datos para entrenamiento y el 20% restante para la validación. Esta división se hace de manera aleatoria, asegurando que ambos sean representativos de la población total, ayudando a evitar sesgos. Seguido, a cada modelo se aplica la validación cruzada, lo que divide el set de entrenamiento en "5-kfold" subconjuntos. Cada modelo es entrenado con un subconjunto y validado con los restantes, rotando su entrenamiento con los subconjuntos para asegurar que todos los datos se utilicen tanto en entrenamiento como en validación.

4. Métricas de evaluación

Las métricas como el coeficiente de determinación de R cuadrada (R^2) son utilizadas para determinar la precisión en las predicciones. Esta fase permite la identificación del modelo más efectivo para la estimación de cada macronutriente, asegurando que el sistema brinde predicciones confiables y precisas. R^2 es una métrica estadística utilizada en modelos de regresión que brinda información sobre la capacidad de las variables independientes de explicar las variables dependientes. R^2 se presenta en fracciones entre 0 y 1; cuando el valor se aproxima a 0 significa que la relación es no lineal entre variables, y mientras se aproxima su valor a 1 significa que los puntos de mapeo de los datos son más exactos a la regresión lineal.

5. Interfaz de usuario

Se desarrolla una interfaz gráfica intuitiva con PyQt5, permitiendo al usuario el ingreso de datos de la leche como pH, grados Dornic y tipo de leche. Esta interfaz es diseñada con una aproximación al diseño centrado en el usuario, priorizando simplicidad y accesibilidad. Esto incluye campos de ingreso de datos claros y botones de acción de fácil interacción. Asimismo, se añade un botón informativo que provee una guía sobre a qué tipo de paciente se recomienda brindarle la leche que se ingresó, ayudando al usuario a interpretar mejor los resultados de las predicciones en el contexto clínico. La interfaz no solo busca la usabilidad del sistema, sino que brinde claridad y accesibilidad en la visualización de los datos de las estimaciones macronutrientales.

RESULTADOS

En la tabla 1 se muestran los valores de R^2 de cada modelo utilizado en el estudio. Estos valores reflejan la capacidad de los modelos para explicar la variabilidad en las predicciones de los macronutrientes de la leche materna.



Se implementa una recomendación básica utilizando una función de puntaje para identificar a pacientes cuyas características se alineen con las necesidades nutricionales. Este sistema evalúa distintos factores, como los niveles de pH y grados Dornic, para determinar qué pacientes tienen características óptimas para la leche. El sistema recibe datos de entrada, creando una nueva instancia, y aplica el modelo más efectivo identificado en la fase de desarrollo.

Con la entrada de datos de pH, grado Dornic y tipo de leche, el sistema genera un conjunto de características que alimentan al modelo para realizar predicciones. Para cada macronutriente, se brindan estimaciones precisas obtenidas con el conjunto de entrenamiento. Al presentar los resultados, se permite una visualización más completa sobre la composición de la leche materna, facilitando las decisiones basadas en información sobre su uso y administración en entornos clínicos.

La figura 3 muestra la interfaz con recomendaciones basadas en el tipo de paciente neonatal. Esta interfaz permite al usuario no solo ver la estimación macronutricional, sino también guías específicas sobre la nutrición neonatal. Con base en los macronutrientes obtenidos, se ofrecen sugerencias personalizadas que toman en cuenta las necesidades energéticas y nutricionales de cada paciente, facilitando la toma de decisiones en el uso y administración de la leche materna en el contexto clínico.

Figura 1 Flujo de trabajo seguido en la metodología

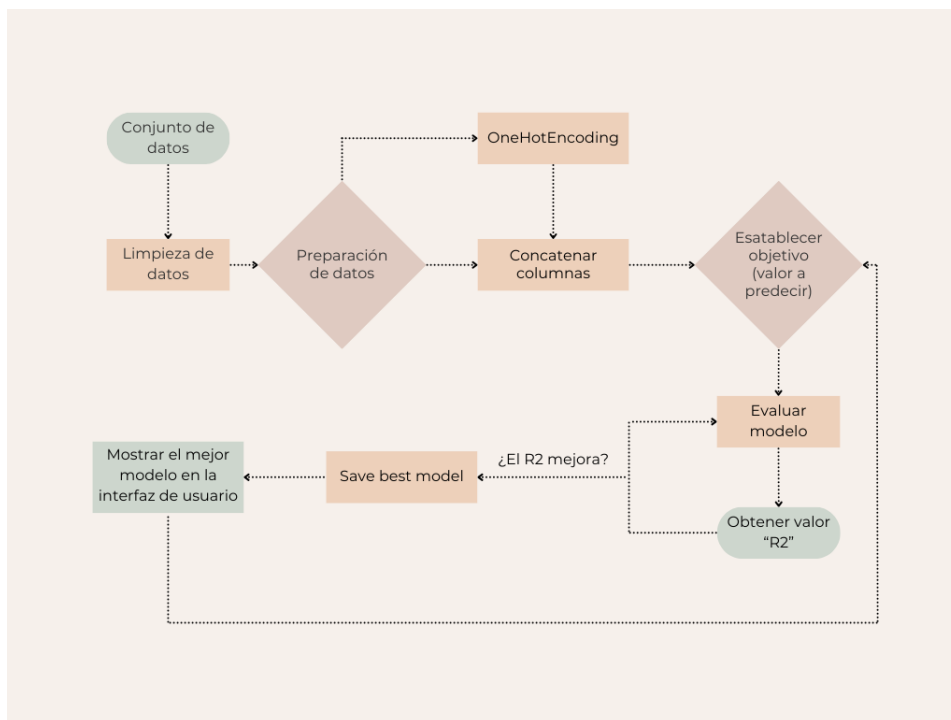


Tabla 1: Coeficiente de determinación R². Valores de cada modelo de regresión entrenado en la estimación de macronutrientes en leche materna

Valores de R ²					
Modelo	Proteína verdadera	Colesterol	Solidos	Energía	Grasa
Regresion lineal	0.4406	-0.0291	0.0277	0.072	0.0503
Regresion Ridge	0.4371	-0.0286	0.0250	0.0265	0.0499
Regresion Lasso	-0.0090	-0.0193	-0.0274	-0.0013	-0.0054
Regresion polinomial	0.1903	-0.0360	0.6735	-0.2812	-0.3029
ElasticNet	-0.009	-0.0197	-0.0274	-0.0074	-0.0054
Árboles de decisión	0.2596	-0.0282	-0.1767	0.1966	0.1515
Bósques aleatorios	0.3902	-0.0273	0.1942	0.2815	0.1530
XGBoost	0.1338	-0.0263	-0.0146	0.3229	-0.0043

Figura 2

Interfaz de usuario con recomendaciones nutrimentales de acuerdo con el tipo de paciente dirigido

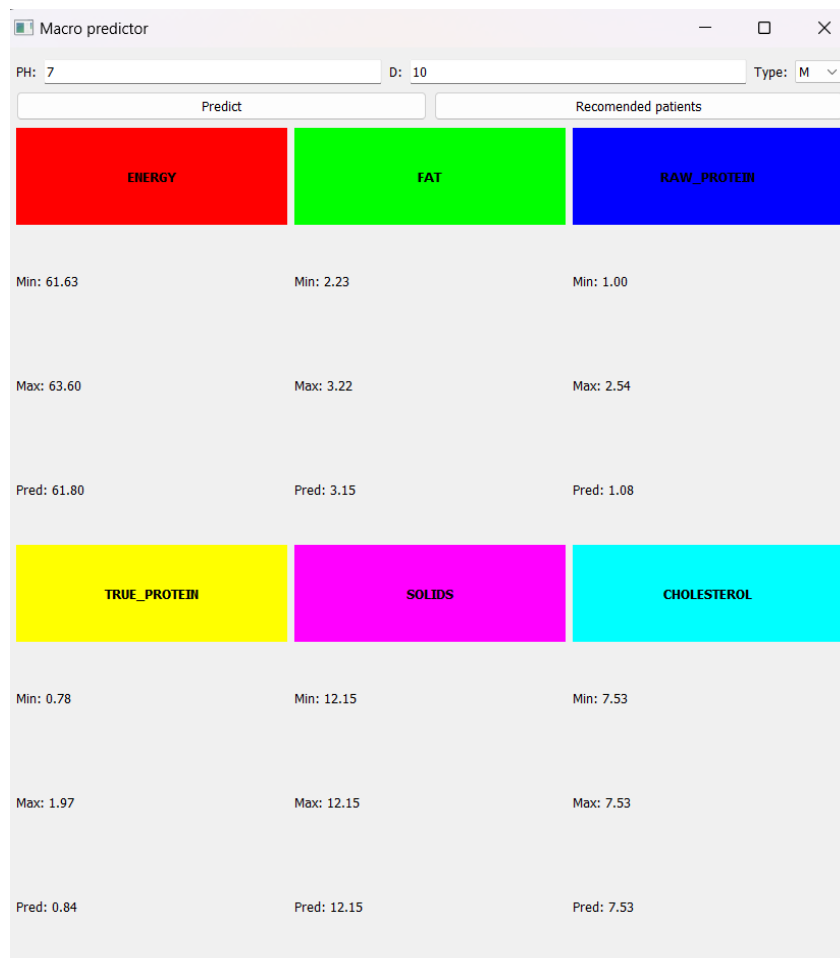
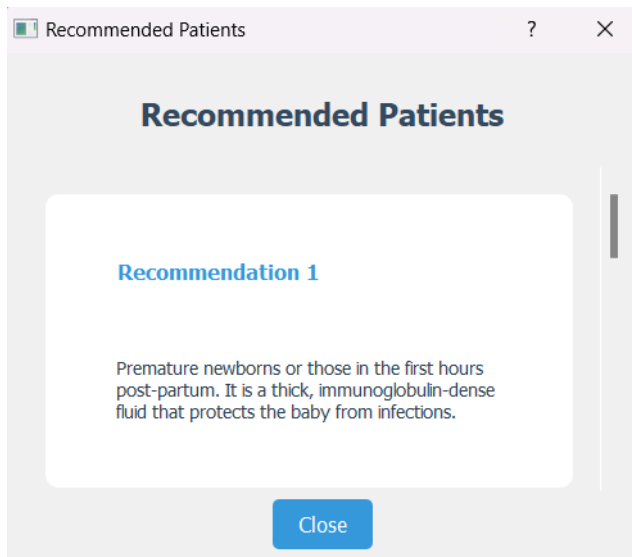


Figura 3 Interfaz gráfica desarrollada para la estimación de macronutrientes en leche materna



DISCUSIONES

La implementación del sistema de estimación de macronutrientes en leche materna representa un avance significativo en el cuidado neonatal, especialmente en ámbitos donde la falta de equipo médico especializado limita la capacidad de análisis. Sin embargo, este desarrollo enfrenta algunas limitaciones que deben ser consideradas.

Primero, a pesar de la métrica de desempeño en el coeficiente de determinación, debe notarse que estos valores pueden variar dependiendo específicamente del modelo escogido. Automatizar el proceso de selección del modelo apropiado asegura que las predicciones sean más eficaces, pero también es importante considerar que añadir métricas adicionales puede darnos una visión más completa sobre la calidad de las predicciones. Los mejores modelos se escogen con base a su desempeño durante la validación cruzada, específicamente obteniendo el promedio de R^2 en cada “k-fold”.

También se considera que el sistema puede beneficiarse con la integración de características adicionales sobre los donantes de leche materna, como historial clínico, antecedentes nutricionales y otros factores de salud, lo cual podría enriquecer el análisis y mejorar la precisión de las predicciones. En versiones futuras, es ideal incorporar estos datos para ofrecer recomendaciones personalizadas y optimizar la nutrición neonatal.

Por otra parte, varios autores exponen las limitaciones de equipo médico en hospitales con recursos limitados (Murthy & Adhikari, 2013; Muttalib et al., 2021; Stevenson et al., 2021), similar a lo que se

expone en la presente investigación. La necesidad de alternativas, como se propone, es vital para asegurar que todos los neonatos, especialmente aquellos en cuidados intensivos, reciban una nutrición adecuada. Además, el estudio de Wong et al. desarrolla modelos de ML para predecir la composición de la leche materna (Wong et al., 2021), lo cual se alinea con nuestro objetivo. Sin embargo, el trabajo que se presenta busca ser más accesible, facilitando el uso en el ámbito clínico a través de una interfaz gráfica intuitiva desarrollada en PyQt5. Aunque esta interfaz es un primer paso, se planea el desarrollo de este software para que pueda ser utilizado tanto en dispositivos móviles como en computadoras, expandiendo su accesibilidad y usabilidad.

Finalmente, el trabajo de Kumar et al. resalta cómo el tipo de leche y otros factores influyen en la nutrición de los neonatos de bajo peso. Estos puntos subrayan la importancia de contar con herramientas que puedan guiar la nutrición de acuerdo con las necesidades específicas de los neonatos, algo que nuestra herramienta busca lograr al ofrecer estimaciones personalizadas. Con el potencial de seguir mejorando, se planea que esta herramienta evolucione para satisfacer las necesidades clínicas y mejorar la salud de los neonatos, especialmente en clínicas con recursos limitados.

CONCLUSIONES

El desarrollo de este programa basado en ML ha demostrado ser una solución viable para la estimación de los macronutrientes de la leche materna en hospitales que no cuentan con equipo de medición directa. La aplicación de este sistema puede mejorar la precisión en la nutrición neonatal al proporcionar datos que optimicen el uso de la leche donada. Aunque este sistema es básico, puede mejorarse en versiones posteriores del software, permitiendo una personalización más avanzada en cuanto al cuidado y uso de la leche donada. Además, la incorporación de una interfaz amigable para el usuario facilitaría su uso con fines educativos, ayudando al personal de salud a identificar el tipo de paciente específico en función de la información ingresada. Esta aproximación tiene el potencial de ser aplicable en una amplia gama de entornos clínicos.

REFERENCIAS BIBLIOGRAFICAS

Fenton, T. R., Elmraged, S., & Alshaikh, B. (2022). Nutrición, crecimiento y resultados clínicos a largo plazo. *Atención Nutricional de Lactantes Prematuros*, 122(2), 12–33.
<https://doi.org/10.1159/000526463>



- Hassiotou, F., Hepworth, A. R., Williams, T. M., Twigger, A.-J., Perrella, S., Lai, C. T., Filgueira, L., Geddes, D. T., & Hartmann, P. E. (2013). Breastmilk Cell and Fat Contents Respond Similarly to Removal of Breastmilk by the Infant. *PLOS ONE*, 8(11), e78232. <https://doi.org/10.1371/journal.pone.0078232>
- Mari Soria, R. (2023, April 1). *Cuidados orientados a favorecer la lactancia en neonatos de alto riesgo*. | EBSCOhost. <https://openurl.ebsco.com/contentitem/gcd:163439280?sid=ebsco:plink:crawler&id=ebsco:gcd:163439280>
- Murthy, S., & Adhikari, N. K. (2013). Global Health Care of the Critically Ill in Low-Resource Settings. *Annals of the American Thoracic Society*, 10(5), 509–513. <https://doi.org/10.1513/AnnalsATS.201307-246OT>
- Muttalib, F., González-Dambrauskas, S., Lee, J. H., Steere, M., Agulnik, A., Murthy, S., & Adhikari, N. K. J. (2021). Pediatric Emergency and Critical Care Resources and Infrastructure in Resource-Limited Settings: A Multicountry Survey*. *Critical Care Medicine*, 49(4), 671. <https://doi.org/10.1097/CCM.0000000000004769>
- Oysal, G., & Özkaya, I. (2022). Nutrición, ganancia de peso y estado de estancia hospitalaria de los lactantes de la Unidad de Cuidados Intensivos Neonatales que sin ser colocados en brazos de sus madres. *Revista española de nutrición comunitaria = Spanish journal of community nutrition*, 28(4), 1.
- Ramel, S. E., & Belfort, M. B. (2022). Nutrición del prematuro y su cerebro. *Atención Nutricional de Lactantes Prematuros*, 122(2), 49–63.
- Shaikh, Z. A., Khan, A. A., Teng, L., Wagan, A. A., & Laghari, A. A. (2022). BIoMT Modular Infrastructure: The Recent Challenges, Issues, and Limitations in Blockchain Hyperledger-Enabled E-Healthcare Application. *Wireless Communications and Mobile Computing*, 2022(1), 3813841. <https://doi.org/10.1155/2022/3813841>
- Stevenson, A. G., Tooke, L., Edwards, E. M., Mangiza, M., Horn, D., Heys, M., Abayneh, M., Chimhuya, S., & Ehret, D. E. Y. (2021). The use of data in resource limited settings to improve

quality of care. *Seminars in Fetal and Neonatal Medicine*, 26(1), 101204.

<https://doi.org/10.1016/j.siny.2021.101204>

Vázquez-Román, S., Bustos-Lozano, G., López-Maestro, M., Rodríguez-López, J., Orbea-Gallardo, C., Samaniego-Fernández, M., & Pallás-Alonso, C. R. (2014). Impacto en la práctica clínica de la apertura de un banco de leche en una unidad neonatal. *Anales de Pediatría*, 81(3), 155–160.

<https://doi.org/10.1016/j.anpedi.2013.11.011>

Victora, C. G., Bahl, R., Barros, A. J. D., França, G. V. A., Horton, S., Krasevec, J., Murch, S., Sankar, M. J., Walker, N., & Rollins, N. C. (2016). Breastfeeding in the 21st century: Epidemiology, mechanisms, and lifelong effect. *The Lancet*, 387(10017), 475–490.

[https://doi.org/10.1016/S0140-6736\(15\)01024-7](https://doi.org/10.1016/S0140-6736(15)01024-7)

WHO, Y. S. (2019, November 11). *Breastfeeding*. <https://www.who.int/health-topics/breastfeeding>

Wong, R. K., Pitino, M. A., Mahmood, R., Zhu, I. Y., Stone, D., O'Connor, D. L., Unger, S., & Chan, T. C. Y. (2021). Predicting Protein and Fat Content in Human Donor Milk Using Machine Learning. *The Journal of Nutrition*, 151(7), 2075–2083. <https://doi.org/10.1093/jn/nxab069>

