International Journal of Science Academic Research
Vol. 06, Issue 08, pp.10485-10487, August 2025
Available online at http://www.scienceijsar.com

ISSN: 2582-6425

Research Article

APPLICATION OF AI AND MACHINE LEARNING TO ANALYZE PROTEIN CONTENT IN U.S. COMMERCIAL BABY FOODS

Malick T. Yedjou, Kevine M. Makoudjou, Ingrid K. Tchakoua, Solange S. Tchounwou and *Clement G. Yedjou

Department of Biological Sciences, College of Science and Technology, Florida Agricultural and Mechanical University, 1610 S. Martin Luther King Blvd, Tallahassee, FL 32307, USA

Received 10th June 2025; **Accepted** 13th July 2025; **Published online** 29th August 2025

Abstract

Proteins are essential macronutrients that support the growth, development, and maintenance of tissues in children. Nutrient requirements vary with age, weight, and physiological needs, making age-specific dietary planning critical. Adequate protein intake promotes both physical growth and cognitive development, while diverse sources such as lean meats, dairy, legumes, and nuts help meet varying nutritional needs and encourage lifelong healthy eating habits. This study analyzed a nutritional dataset of 244 baby foods using artificial intelligence (AI) and machine learning to assess protein content, categorizing items into three groups based on protein content: low (0.0–5.9 g/day), moderate (6.0–10.9 g/day), and high (11.0–15.0 g/day). The majority (n = 202) fell into the low-protein range, followed by 22 in the moderate range and 20 in the high range. Age-specific protein requirements, expressed in grams per kilogram of body weight (g/kg), were assessed for four age groups: 0–6 months (1.52 g/kg; 12.6–15.8 g/day; 5.5–6.0 kg), 7–9 months (1.20 g/kg; 9.0–10.2 g/day; 7.5–8.5 kg), 10–12 months (1.00 g/kg; 8.5–9.5 g/day; 8.5–9.5 kg), and 1–3 years (1.05 g/kg; 12.6–15.8 g/day; 12.0–15.0 kg). Low-protein foods may be insufficient for infants with reduced breastmilk or formula intake, while high-protein foods often rich in meat, dairy, or fortified products can help meet upper-range requirements. These findings underscore the need for careful alignment of complementary food protein levels with age-specific nutritional guidelines to support optimal growth and development in early childhood.

Keywords: Baby foods, Protein content, AI, machine learning, Infants, Young children.

INTRODUCTION

Protein is a critical macronutrient for infants and young children, serving as a foundational element in the development and maintenance of body tissues, immune function, and cognitive processes[1]. Dietary protein requirements are dynamic, changing with age, weight, and physiological status. For infants transitioning from exclusive milk feeding to complementary feeding, ensuring adequate protein intake is essential for preventing growth faltering and promoting healthy development[2,3].Emerging research suggests that plant proteins such as those derived from soy, legumes, and other seeds- may contribute to improved lipid and glucose metabolism via bioactive peptides. In vitro and animal studies highlight that certain soy-derived peptides can reduce total cholesterol and triglycerides (e.g., by ~5.7% and up to 12.4%, respectively), possibly through mechanisms like bile acid binding or inhibition of HMG-CoA reductase[4,5]. Moreover, legume-based bioactive peptides (from pulses like beans, peas, and lentils) exhibit antioxidant, anti-inflammatory, antihypertensive, and hypocholesterolemic functions in vitro, which may offer metabolic benefits relevant to metabolic syndrome. Yet, human clinical trials remain scarce[6]. Conversely, higher consumption of sulfur amino acids common in animal proteins is associated with elevated cardio metabolic risk. NHANES III data show that high SAA intake is linked to worsened cholesterol, glucose, insulin, and inflammation markers[7,8].

Corresponding Author:** ***Clement G. Yedjou,
Department of Biological Sciences, College of Science and Technology, Florida Agricultural and Mechanical University, 1610 S. Martin Luther King Blvd, Tallahassee, FL 32307, USA.

High-protein diets particularly those where protein constitutes 25–32% of total energy, have demonstrated notable benefits in individuals with type 2 diabetes, including modest weight loss (~2 kg), a reduction in glycated hemoglobin ($HbA_{1}c$) by roughly 0.5 %, and slight improvements in systolic and diastolic blood pressure (–3.1 mmHg and –1.9 mmHg, respectively). However, these interventions did not significantly alter fasting blood glucose or lipid profiles, underscoring that the metabolic benefits may not extend uniformly across all markers[9]. Advancements in artificial intelligence (AI) and machine learning (ML) offer new opportunities to analyze large-scale nutritional datasets, detect patterns, and generate actionable insights in infant nutrition research[10–12]. This study leverages AI and ML techniques to assess protein content across a range of commercially available baby foods, classify them into practical intake categories, and compare these findings against age-specific protein requirements.

Approaches

1. Data Sources: The data used for our study were sourced from the United States Department of Agriculture.
2. Dataset: A nutritional dataset comprising 244 commercially available baby foods was analyzed. Protein content per serving was extracted from the database and standardized to daily intake estimates.
3. Classification: Using AI-driven clustering and supervised classification algorithms (k-means and decision tree models), foods were grouped into three protein intake categories:

- **Low protein**: 0.0–5.9 g/day
- **Moderate protein**: 6.0–10.9 g/day
- **High protein**: 11.0–15.0 g/day

The classification thresholds were informed by complementary feeding guidelines from the World Health Organization [1].

4. Age-specific protein requirements: Protein needs were calculated in grams per kilogram of body weight (g/kg) for four age groups based on average body weights (Table 1).

Table 1. summarizes the age range, protein requirement, estimated daily protein intake, and average body weight for infants and young children

Age Group	g/kg Requirement	Estimated Daily Protein (g/day)	Avg. Body Weight (kg)
0–6 months	1.52	12.6–15.8	5.5–6.0
7–9 months	1.20	9.0–10.2	7.5–8.5
10–12 months	1.00	8.5–9.5	8.5–9.5
1–3 years	1.05	12.6–15.8	12.0–15.0

5. Selection and exclusion Criteria:

a. Selection Criteria: We selected 244 out of 245 commercial baby foods based on protein content ranging from 0 to 15 grams and categorized them into three groups (low, moderate, and high protein content).
b. Exclusion Criteria: We eliminated one commercial baby food with high protein content, with a value of 25 grams.

RESULTS

Our study revealed that the majority of baby foods contain low protein, and a small number of foods contain moderate and high protein (Figure 1). Figure 1 shows the distribution of baby foods by protein category (low, moderate, and high protein). Of the 244 baby food products, 202 (82.8%) were low protein, 22 (9.0%) were moderate protein, and 20 (8.2%) were high protein.

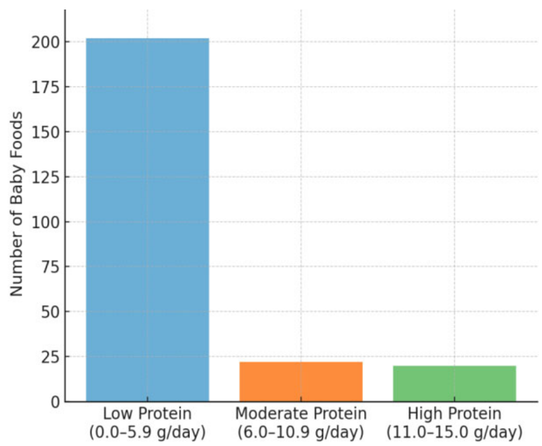

Figure 1. Distribution of baby foods by protein category. All values are based on WHO and Food and Nutrition Board guidelines

Figure 2 presents recommended protein intakes by age group in g/kg, highlighting that complementary food contributions must be carefully balanced with ongoing milk feedings to avoid both insufficiency and excess.

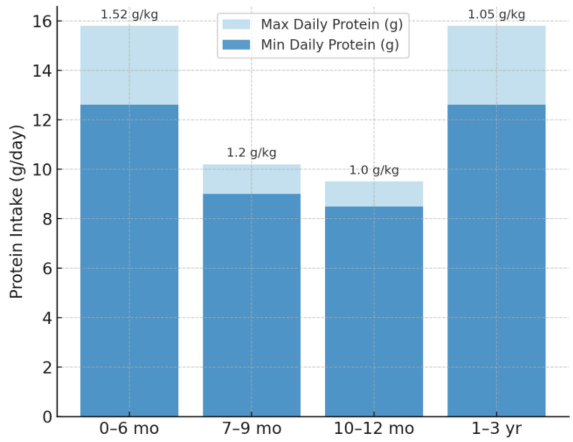

Figure 2. Recommended protein intake by infants and young children expressed in g/kg. All values are based on WHO and Food and Nutrition Board guidelines

DISCUSSION

Proteins are fundamental macronutrients that serve as the primary building blocks for growth and play a vital anabolic role in the development and maintenance of muscles and tissues. On average, proteins account for approximately 15% of the body's total energy intake. Structurally, they are large biomolecules composed of amino acid chains linked by peptide bonds and can be classified as dipeptides, oligopeptides, or polypeptides based on chain length. In this study, we applied artificial intelligence (AI) and machine learning (ML) techniques to evaluate protein levels across a range of commercial baby foods. Our analysis revealed a disproportionate representation of low-protein products within the market. These findings are consistent with broader concerns regarding nutritional adequacy in infant diets. Previous work in our laboratory has demonstrated the versatility of AI and ML in healthcare research, including breast cancer screening and classification [13,14].as well as investigations into the role of fruit and vegetable consumption in reducing COVID-19 incidence and mortality[15,16]. From a nutritional standpoint, low-protein baby foods are generally suitable for infants who receive sufficient breastmilk or formula, as these provide a complete amino acid profile and meet baseline protein requirements. However, for infants with reduced milk intake due to weaning, limited maternal supply, or feeding difficulties these low-protein options may fall short of recommended levels. In contrast, high-protein products, often derived from meat or fortified with protein, can help meet the upper threshold of protein needs but also carry the potential risk of exceeding requirements, particularly when combined with high volumes of milk consumption. Figure 3 below emphasizes how protein content in a meal can vary depending on both the type and quantity of protein sources. Animal proteins (fish, chicken, steak) typically provide more protein per serving compared to plant-based sources, though plant proteins (legumes, chickpeas) can contribute significantly when consumed in adequate amounts. This visual aid helps illustrate that dietary protein can be adjusted to meet different nutritional needs, from minimal intake to high-protein diets,

Low Protein: Depicted with a small portion of fish accompanied by broccoli and a side of cottage cheese. This represents meals with relatively smaller amounts of protein-rich foods, suitable for individuals with low protein requirements or dietary restrictions.

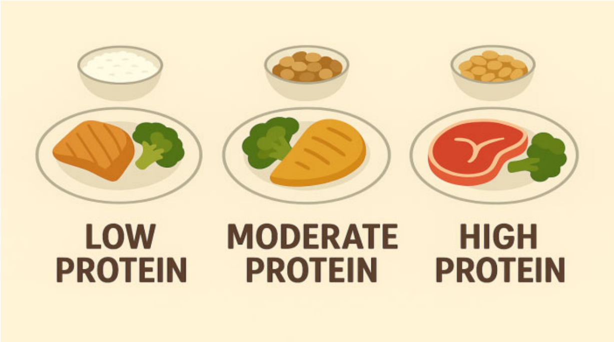

Figure 3. Protein content in foods ranging from low to high. The diagram visually compares meals containing low, moderate, and high protein content

Moderate Protein: Illustrated with a grilled chicken breast and broccoli, alongside a side of legumes. This portion contains a balanced amount of protein, appropriate for maintaining general health in most adults.

High Protein: Shown with a large steak, broccoli, and a side of chickpeas. This meal represents a protein-dense option, often used in high-protein diets for muscle growth, weight management, or specific health goals.

Conclusion

By applying AI and ML to a large-scale baby food dataset, this study provides actionable insights into the distribution of protein content and its alignment with age-specific needs. These findings can inform dietary recommendations, public health policies, and product innovation to better support healthy growth and development in infants and toddlers. The practical implication for caregivers and health professionals is the need to tailor complementary feeding choices to the individual child's age, growth trajectory, and milk intake. AI-powered nutrient profiling offers a scalable tool to guide both product development and parental selection toward meeting optimal protein targets in early life.

Author Contributions: Conceptualization, M.T.Y., K.M.M., I.K.T., S.S.T., and C.G.Y.; Methodology, M.T.Y., K.M.M., I.K.T., S.S.T., and C.G.Y.; formal analysis, M.T.Y., K.M.M., I.K.T., S.S.T., and C.G.Y.; investigation, M.T.Y., K.M.M., I.K.T., and S.S.T.; funding, C.G.Y.; supervision, C.G.Y.; writing-original draft preparation, all authors; writing-reviewing and editing, all authors. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported in part by the National Science Foundation (NSF) under Grants #2142465 and #2417643 at Florida Agricultural and Mechanical University, Tallahassee, FL, United States. Additional support was provided by the National Institutes of Health (NIH) under Grant #U54MD013376-05 at Morgan State University, Baltimore, MD, United States. The views expressed in this publication are solely those of the authors and do not necessarily represent the official positions of the NSF or NIH.

Institutional Review Board Statement: Not applicable.
Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the present review article are included in the article.

Conflicts of Interest: The authors declare no conflicts of interest.

REFERENCES

1. WHO. WHO Guideline for Complementary Feeding of Infants and Young Children 6–23 Months of Age. *World Health Organization,* (2023).
2. Grummer-Strawn, L. Time to revise the complementary feeding guidelines? *Ann Nutr Metab.,* 71, (2017).
3. Cattaneo, A. et al. ESPGHAN's 2008 recommendation for early introduction of complementary foods: How good is the evidence? Maternal and Child *Nutrition*, vol. 7 Preprint at https://doi.org/10.1111/j.1740-8709.2011.00363.x (2011).
4. Peighambardoust, S. H., Karami, Z., Pateiro, M. & Lorenzo, J. M. A review on health-promoting, biological, and functional aspects of bioactive peptides in food applications. *Biomolecules*, vol. 11 Preprint at https://doi.org/10.3390/biom11050631 (2021).
5. Korhonen, H. & Pihlanto, A. Technological Options for the Production of Health-Promoting Proteins and Peptides Derived from Milk and Colostrum. Curr Pharm Des., 13, (2007).
6. Garcés-Rimón, M., Morales, D. & Miguel-Castro, M. Potential Role of Bioactive Proteins and Peptides Derived from Legumes towards Metabolic Syndrome. Nutrients vol. 14 Preprint at https://doi.org/10.3390/nu14245271 (2022).
7. Zhang, X. et al. Identification of a leucine-mediated threshold effect governing macrophage mTOR signalling and cardiovascular risk. Nat Metab 6, (2024).
8. Dong, Z. et al. Association of sulfur amino acid consumption with cardiometabolic risk factors: Cross-sectional findings from NHANES III. *EClinical Medicine,* 19, (2020).
9. Chiavaroli, L. et al. Effect of low glycaemic index or load dietary patterns on glycaemic control and cardiometabolic risk factors in diabetes: Systematic review and meta-analysis of randomised controlled trials. *The BMJ,* vol. 374 Preprint at https://doi.org/10.1136/bmj.n1651 (2021).
10. Haq, K. T., Howell, S. J. & Tereshchenko, L. G. Applying Artificial Intelligence to ECG Analysis. *Circ Arrhythm Electrophysiol,* 13, (2020).
11. Hassoun, S. et al. Artificial Intelligence for Biology. *Integr Comp Biol.,* 61, (2021).
12. Topol, E. J. High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine,* vol. 25 Preprint at https://doi.org/10.1038/s41591-018-0300-7 (2019).
13. Clement G. Yedjou, Solange S. Tchounwou, Richard A. Aló, Rashid Elhag, BereKetMochona, and L. L. Application of Machine Learning Algorithms in Breast Cancer Diagnosis and Classification. *Int J Sci Acad Res.,* 2, 3081–3086 (2021).
14. Wei, Z. et al. Quantitative DCE Dynamics on Transformed MR Imaging Discriminates Clinically Significant Prostate Cancer. *Cancer Control.,* 31, (2024).
15. Yedjou, C. G. et al. COVID-19 Prevalence in Nations with Normal Body Mass Index (BMI): Implications of Artificial Intelligence (AI) in Healthcare. *J Nutr Food Sci.,* 15, (2025).
16. Yedjou, C. G. et al. Chemo-Preventive Effect of Vegetables and Fruits Consumption on the COVID-19 Pandemic. *J Nutr Food Sci.,* 4, (2021).

****