

Identificación de Aminoácidos de Cadena Ramificada en Proteína de Quinoa mediante Modelo de Regresión Lineal con Aprendizaje de Máquinas

Identification of Branched-Chain Amino Acids in Quinoa Using Linear Regression Modelling with Machine Learning.

DOI: <http://dx.doi.org/10.17981/cesta.03.02.2022.04>

Artículo de investigación científica. Fecha de recepción: 29/04/2022. Fecha de aceptación: 17/06/2022.

Nicolas Caicedo Cerón 

Natural Company. Cali (Colombia)
Universidad Santiago de Cali. Cali (Colombia)
nicolas.caicedo01@usc.edu.co

Juan Caicedo Cerón 

Natural Company. Cali (Colombia)
jjcc1421@gmail.com

José Oñate Garzón 

Universidad Santiago de Cali. Cali (Colombia)
jose.onate00@usc.edu.co

Para citar:

N. Caicedo, J. Caicedo y J. Oñate, "Identificación de Aminoácidos de Cadena Ramificada en Proteína de Quinoa mediante Modelo de Regresión Lineal con Aprendizaje de Máquinas", *J. Comput. Electron. Sci.: Theory Appl.*, vol. 3, no. 2, pp. 24–32, 2022. <https://doi.org/10.17981/cesta.03.02.2022.04>

Resumen

Introducción— La quinoa, se clasifica como pseudocereal y contiene componentes bioactivos, tales como proteínas con un alto valor nutricional. El consumo de proteína y especialmente aminoácidos esenciales de cadena ramificada juegan un papel fundamental en la dieta, ya que favorece al mantenimiento de estructuras corporales.

Objetivos— El presente artículo busca realizar un análisis y cuantificación de los aminoácidos de la proteína de Quinoa usando la base de datos del Departamento de Agricultura de los Estados Unidos (USDA).

Metodología— Se utiliza el lenguaje de programación Python con el fin de desarrollar modelos de aprendizaje de máquinas para pronosticar la presencia de aminoácidos ramificados en proteína de Quinoa y mediante regresión lineal obtener datos estadísticos que suministren una cantidad aproximada de estos nutrientes. Asimismo se puede encontrar información de artículos importantes donde se presentan técnicas de aprendizaje automático para obtener información nutricional.

Resultados— Se determinan los aminoácidos valina, leucina e isoleucina en la proteína aislada de Quinoa y a su vez se utilizan como controles la quinoa cruda, quinoa cocida, garbanzo, arroz y brócoli. Los valores de concentración p/p obtenidos en g/100g fueron 15.9, 2.1, 0.66, 3.43, 1.7 y 0.37 respectivamente.

Conclusiones— Estos resultados evidencian la presencia de aminoácidos de cadena ramificada en la Quinoa, constatando la importancia de su valor nutricional y sus posibles efectos en la salud muscular.

Palabras clave— Aprendizaje de máquinas; regresión lineal; Sklearn; Chenopodium quinoa; aminoácidos

Abstract

Introduction— Quinoa is classified as a pseudocereal and contains bioactive components, such as proteins with high nutritional value. The consumption of protein and especially branched-chain essential amino acids play a fundamental role in the diet, since it favors the maintenance of body structures.

Objectives— The present article seeks to perform an analysis and quantification of amino acids in Quinoa protein using the United States Department of Agriculture database (USDA). Python programming language was used to develop machine learning models to predict the presence of branched amino acids in Quinoa protein and through linear regression to obtain statistical data that provide an approximate amount of these nutrients. Important articles can be found where information on machine learning techniques to obtain nutritional information is presented.

Results— The amino acids valine, leucine and isoleucine were identified in isolated Quinoa protein and, in turn, raw quinoa, cooked quinoa, chickpea, rice and broccoli were used as controls. The w/w concentration values obtained in g/100g were 15.9, 2.1, 0.66, 3.43, 1.7 and 0.37, respectively.

Conclusions— These results show the presence of branched-chain amino acids in quinoa, confirming the importance of its nutritional value and its possible effects on muscle health.

Keywords— Machine learning; linear regression; Sklearn; Chenopodium quinoa; amino acids



I. INTRODUCCIÓN

De la región Andina proviene la Quinoa, una planta herbácea del género *Chenopodium*, la cual tiene un ciclo de vida anual, cuenta con más de tres mil variedades y una sorprendente adaptabilidad a diversos pisos térmicos, resiste temperaturas desde -4°C hasta los 38°C , no necesita mucha agua para su desarrollo, alcanza una altura de 0.5 m a 3.0 m según su especie, tolera humedades relativas desde 40% hasta 88% y puede ser cultivada a nivel del mar o hasta los 4000 metros de altitud [1]. Sus semillas son cilíndricas con diámetros entre 1.8 mm y 2.2 mm, pueden presentarse de colores variados de tonalidades blancas, rojas, rosadas, negras, grises, cafés y amarillas. En las semillas se localizan sus componentes bioactivos y su mayor valor nutricional, conteniendo un alto valor proteico entre 13.81% y 21.9% acorde a su variedad [2], una adecuada disposición de oligoelementos, vitaminas, minerales y de la presencia de todos los aminoácidos esenciales [3].

Las proteínas son moléculas presentes en los seres vivos y tienen múltiples funciones, generalmente comparten una característica estructural común, ser polímeros lineales de aminoácidos. Hasta ahora, se han descrito más de 300 aminoácidos diferentes en la naturaleza, pero sólo 20 se encuentran normalmente como constituyentes de las proteínas de mamíferos. Cada aminoácido tiene un grupo carboxilo ($-\text{COO}-$), un grupo amino ($-\text{NH}_2+$) primario (excepto la prolina, que tiene un grupo amino secundario) y una cadena lateral distintiva (grupo R) unido al átomo de carbono α [4].

Durante el curso de la evolución, los animales desde los protozoos hasta los mamíferos perdieron la capacidad enzimática para sintetizar ciertos aminoácidos y deben ser requeridos en la dieta. Estas sustancias se denominan aminoácidos esenciales, los cuales son nueve: histidina (His), isoleucina (Ile), leucina (Leu), lisina (Lys), metionina (Met), fenilalanina (Phe), treonina (Thr), triptófano (Trp) y valina (Val). Cumplen con dos funciones distintas: la primera actuar como bloques de construcción de proteínas y la segunda actuar como señales nutricionales para provocar reacciones fisiológicas [5].

El mantenimiento de la masa muscular esquelética a lo largo de la vida depende de la ingesta de proteínas dietéticas adecuadas, las proteínas de origen vegetal tienen una menor calidad que las proteínas de origen animal. Sin embargo, proteínas de diferentes fuentes vegetales pueden combinarse de tal modo que el resultado sea equivalente en valor nutricional a la proteína animal [4], [6]. Los aminoácidos dentro de las proteínas dietéticas estimulan de forma aguda la síntesis de proteínas musculares [7]. Como tal, la ingesta de la cantidad y calidad adecuadas de proteínas dietéticas es un requisito previo para la salud del sistema músculo esquelético. Los aminoácidos esenciales son cruciales para la estimulación de síntesis de proteínas musculares [8], [9]. Un subgrupo de aminoácidos esenciales, los Aminoácidos de Cadena Ramificada (BCAA por sus siglas en inglés), los cuales son Ile, Leu y Val, están asociados con el metabolismo en tejidos como el músculo esquelético [10]. La suplementación de BCAA en periodo de recuperación genera un menor grado de dolor y daño muscular, menor percepción del esfuerzo y mayor respuesta anabólica [11]. Existe evidencia clínica que sugiere que la suplementación con BCAA tiene efectos beneficiosos sobre el peso corporal, la grasa y la masa corporales magra [5].

Un análisis estadístico del valor nutricional, particularmente de la cantidad de BCAA presentes en la proteína de Quinoa, integrando el procesamiento de muchos datos mediante cálculos matemáticos, se puede realizar con un algoritmo de regresión lineal, el cual entrena un conjunto de datos con funciones lineales. La regresión lineal múltiple trata de ajustar modelos lineales entre una variable dependiente y más de dos variables independientes [12], expresando magnitudes de correlación entre variables. También la regresión lineal se utiliza para predecir el valor de una variable según el valor de otra. La variable que desea predecir se denomina *variable dependiente*. La variable que está utilizando para predecir el valor de la otra variable se denomina *variable independiente* [13]. El análisis de regresión lineal considera los coeficientes de la ecuación lineal, implicando una o más variables independientes las cuales predicen el valor de la variable dependiente [12]. Dicha relación puede considerarse significativa o fuerte (distinta de una relación normal) o no significativa o débil (similar a una relación normal). La correlación es una medida de la similitud de la variabilidad de dos magnitudes estocásticas, la cual tiene el propósito de averiguar en qué medida la variable dependiente puede explicarse por la variable independiente [13].

El uso de herramientas digitales como el Aprendizaje de Máquinas (ML por sus siglas en inglés), ha facilitado el procesamiento de datos de una manera y en una escala sin igual para los humanos, ya que integra matrices complejas de factores para proporcionar un valor preciso o estratificado, y soporta un proceso de predicción [14]. El ML aplica inteligencia artificial en la que los sistemas aprenden interactivamente de los datos para investigar patrones estadísticos y predecir relaciones previamente desconocidas, es capaz de manejar datos más complejos que las técnicas estadísticas utilizadas tradicionalmente [14]. Las técnicas de ML están diseñadas para comprender y representar matemáticamente los patrones presentes en los datos. Como resultado, la clave para crear algoritmos de ML precisos y aplicables radica tanto en el tamaño como en la calidad de los datos utilizados [12].

En este escenario, el uso de técnicas de ML como la regresión lineal, usan el lenguaje de programación Python, el cual posee una comunidad amplia en la que es posible apoyarse para el uso adecuado de librerías y bases de datos. Además su gran popularidad radica en la simplicidad del lenguaje y la fácil lectura de los programas, lo que permite al desarrollador enfocarse más en resolver el problema para el cual se escribe el código, que en los detalles técnicos relativos al lenguaje de programación [15]. Esto facilita interpretar múltiples factores, como la información nutricional de la base de datos del Departamento de Agricultura de los Estados Unidos (USDA por sus siglas en inglés) [16], estableciendo una correlación de los aminoácidos esenciales no ramificados, los cuales son His, Lys, Met, Phe, Thr y Trp frente a los BCAA, además de un algoritmo para usarlo en el lenguaje de programación Python y así crear un modelo de predicción que determine la cantidad BCAA presentes en alimentos que contengan todos los aminoácidos esenciales. Para el análisis de la concentración proteica de aminoácidos presentes en la proteína de quinua, se tomaron datos reportados en un artículo de la revista Centro Azúcar de la Universidad Central “Marta Abreu” de Las Villas (UCLV) localizada en Santa Clara (Cuba) [17]. Estos datos se usaron en el modelo predictivo probabilístico de ML y permitieron identificar concentraciones específicas de BCAA en la proteína de quinua, apoyando la toma de decisiones al considerar un nutriente con alta cantidad proteica como parte esencial en la dieta.

II. TRABAJOS RELACIONADOS

El grupo de investigación en tecnología de la información, de la Universidad de Wageningen en Países Bajos presenta información de técnicas de aprendizaje automático para nutrición de precisión, llevando a cabo una Revisión Sistemática de la Literatura (SLR), en el cual se recuperaron 4930 artículos de bases de datos electrónicas y se seleccionaron 60 estudios primarios, donde se ven cuatro tareas de aprendizaje automático en forma de regresión, clasificación, recomendación y agrupamiento, los autores consideran fundamental incorporar ML en los métodos de investigación nutricional, para facilitar la integración de muchas características complejas, lo que permite el desarrollo de enfoques de Precisión en Nutrición de alto rendimiento [18], [19]. Por otro lado, estudios realizados en Tailandia presentan un estudio comparativo con modelos matemáticos, regresión lineal múltiple y redes neuronales artificiales para predecir la relación de humedad y la tasa de secado de cubos de piña deshidratada; los modelos predictores fueron los óptimos para estimar el proceso de secado de frutas recubiertas y sin recubrir con precisión y beneficios satisfactorios para la industria alimentaria [20], [21]. En otro artículo de Filipinas se compara la precisión de cuatro algoritmos de ML, utilizando dos esquemas de clasificación, para predecir la desnutrición en función de los factores de riesgo individuales y domésticos. El riesgo nutricional de los niños se clasificó según los Rangos Aceptables de Distribución de Macronutrientes (AMDR) desarrollados por la Academia Nacional de Medicina (NAM) y las Ingestas Dietéticas de Referencia de Filipinas (PDRI). La prevalencia de desnutrición fue mayor con NAM AMDR (82.67%) en comparación con PDRI AMDR (78.71%). Este artículo concluye que Random forest fue el algoritmo de ML más preciso (78.55%), capaz de predecir la desnutrición en función de los gastos del hogar, la edad de los niños y del hogar, la inseguridad alimentaria y la diversidad dietética [22].

III. MATERIALES Y MÉTODOS

A. Base de datos

Inicialmente se obtuvieron datos extraídos por medio de Web Scraping de la base de datos FoodData Central [23], presentados en forma tabular en 8789 filas, en las cuales se exponen datos de alimentos como arroz, quinua cruda, quinua cocida, garbanzos, brócoli entre otros y 77 columnas las cuales tienen información nutricional de macronutrientes (carbohidratos, fibra, azúcar, proteína, alanina, arginina, ácido aspártico, cisteína, ácido glutámico, glicina, histidina, isoleucina, leucina, lisina, metionina, fenilalanina, prolina, serina, treonina, triptófano, tirosina, valina), vitaminas (niacina, riboflavina) y minerales (magnesio, fosforo potasio, hierro, zinc) en concentraciones p/p obtenidos en g/100g. Esta base de información [23] es integrada a Python para la creación de una muestra de datos supervisados, permitiendo así definir un modelo de predicción que calcule los BCAA respecto a otras variables.

En el presente análisis, la información utilizada tiene la finalidad de predecir mediante un método de regresión lineal con ML, la cantidad de BCAA presente en diversos alimentos, tomando los datos de FoodData Central del USDA. FoodData Central es un sistema de datos integrados que suministra datos nutricionales, los cuales, a partir de abril del 2020 son actualizados mensualmente y proporcionan información sobre el perfil de nutrientes y enlaces a investigaciones agrícolas y experimentales [16]. Incluye cinco tipos de datos distintos

que proveen información sobre perfiles de alimentos y nutrientes: Foundation Foods, Food and Nutrient Database for Dietary Studies 2017-2018 (FNDDS 2017-2018), National Nutrient Database for Standard Reference Legacy Release (SR Legacy), USDA Global Branded Food Products Database (Branded Foods) y Experimental Foods [16].

La base de datos cuenta con un registro de aproximadamente 8 800 alimentos, con información de sus macronutrientes, vitaminas y minerales como grasas saturadas, no saturada, agua, proteína, aminoácidos esenciales, vitamina C, magnesio, zinc entre otros nutrientes [23]. Posteriormente, la información actualizada a marzo de 2022 fue analizada usando el lenguaje de programación Python (versión 3). Se eligió el lenguaje de programación Python debido a que tiene un extenso catálogo de librerías para aprendizaje de máquinas, que son piezas de código previamente construidas y comprobadas lo cual facilita el desarrollo del programa [24].

Adicional a los datos que contenía la base anteriormente mencionada se integró a la base, información nutricional referente a la proteína aislada de Quinoa [17], la incorporación de esta información es necesaria para comparar los nutrientes de los alimentos presentes en la base de datos con la proteína aislada de Quinoa. Esto debido a que USDA no contenía la información específica de la cantidad de aminoácidos esenciales (His, Ile, Leu, Lys, Met, Phe, Thr, Trp y Val) disponibles en la proteína aislada de Quinoa, los cuales están presentes en el artículo y son fundamentales para usar del modelo de predicción de ML.

Luego se realizó una normalización de la información mediante la asignación numérica para que el lenguaje programático estuviera acorde con la interpretación de Python y pudiera ser analizado por el mismo y de este modo generar un proceso decisorio mediante un gráfico de calor, identificando los coeficientes de correlación, aquellos coeficientes que estuvieran por encima de 0.8 fueron seleccionados, para reducir la alta cantidad de datos proporcionada por USDA [17]. Uno de los factores que llevó a la utilización del modelo de regresión lineal múltiple, fue la identificación de tendencias lineales en las gráficas de aminoácidos esenciales no ramificados His, Lys, Met, Phe, Thr y Trp contra los BCAA.

B. Descripción del Modelo

Se interpretó la base de datos mediante un diagrama de dispersión entre los datos correlacionados de los BCAA vs los aminoácidos esenciales no ramificados de todos los datos ingresados en Python. Se evidencia en la Fig. 1 una alta linealidad en la correlación entre los BCAA que exhiben concentraciones de 0 g/100g a 15 g/100g con presencia de Ile, Leu y Val vs los aminoácidos esenciales no ramificados, His, Lys, Met, Phe, Thr y Trp. Este parámetro soporta el uso del modelo de regresión lineal para realizar un análisis predictivo de la concentración de BCAA.

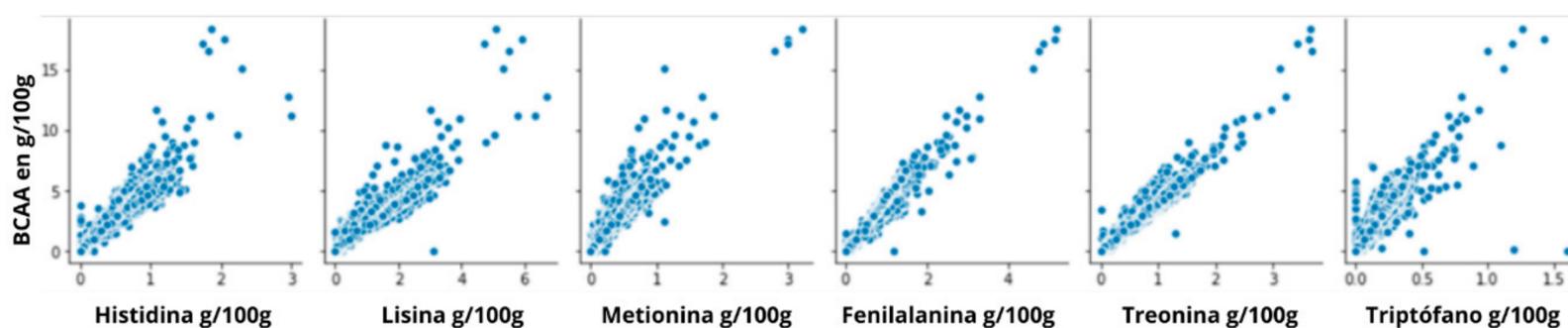


Fig. 1. Diagrama de dispersión entre los datos correlacionados Aminoácidos esenciales no ramificados vs BCAA.

Fuente: Elaboración propia.

La regresión lineal permitió crear un modelo de predicción basado en las variables de aminoácidos no ramificados altamente correlacionados vs la variable independiente (BCAA) ingresados en la base de datos. La regresión lineal se ajustó mediante una superficie que minimiza la diferencia entre los valores estimados y reales [25].

Posteriormente se genera un modelo de regresión lineal mediante técnicas de ML utilizando el paquete “Sklearn” [23] para estimar las relaciones entre los valores de análisis matriz Y (BCAA) y matriz X (Total de datos ingresados) en términos de predicción de variables Y respecto a X .

Por último, se realizó el entrenamiento del modelo de ML utilizando el 80% de los datos y mediante el 20% de los datos restantes se comprueba el modelo, se selecciona esta distribución con base al principio de Pareto [26]. La calidad de los modelos de regresión lineal se evaluó mediante el coeficiente de determinación (R²) y el error-cuadrático-medio (MSE) para calibración y validación [27]. La Fig. 2 describe el flujo de información y el desarrollo final del modelo a partir de ML.

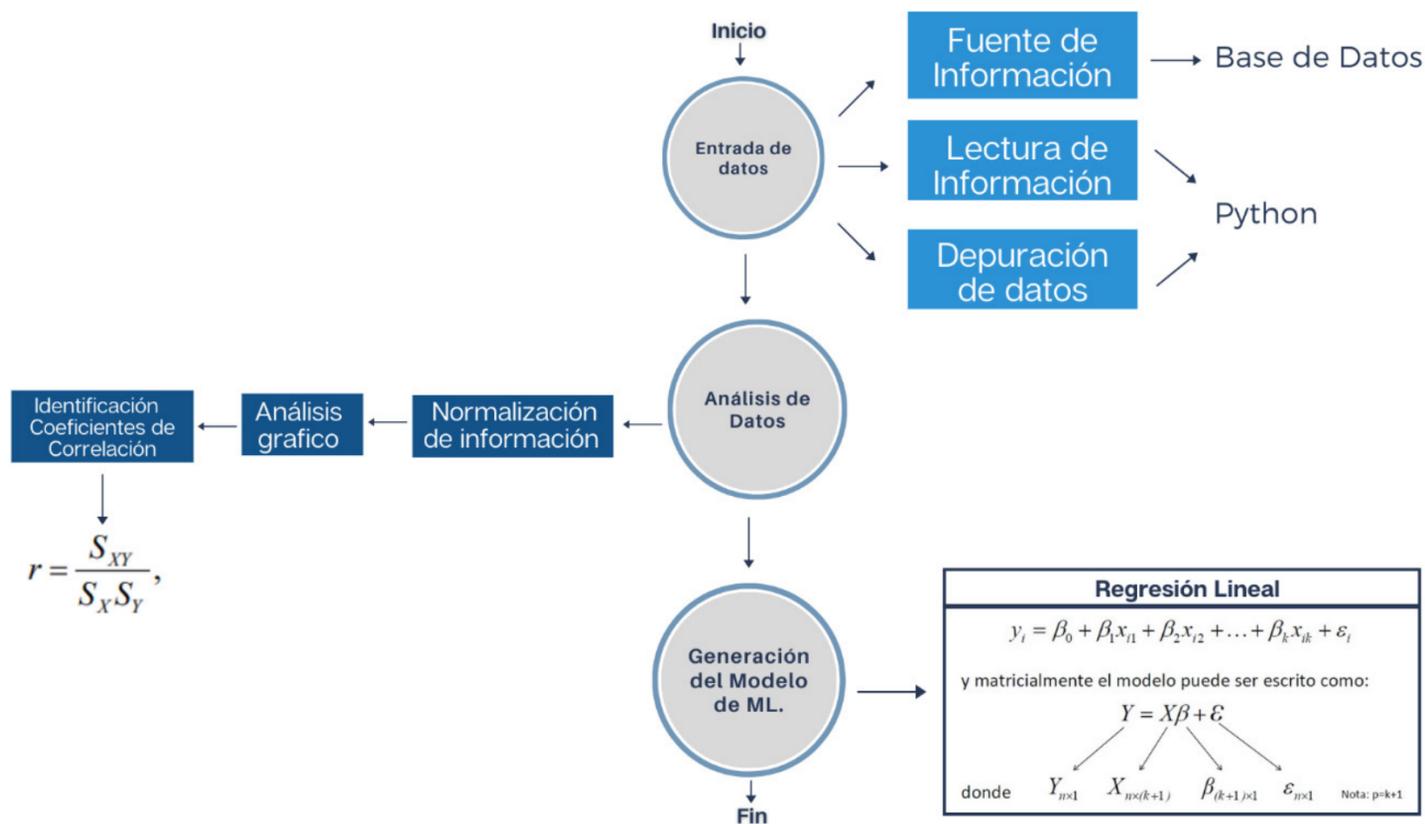


Fig. 2. Flujo de Información y generación del modelo predictor de ML.
Fuente: Elaboración propia.

IV. RESULTADOS

A. Análisis de correlación

La regresión lineal ajusta un modelo lineal entre una variable dependiente que es la variable la cual se desea predecir, y una o más variables independientes que se usarán como datos de entrada del modelo. En nuestro caso la variable independiente corresponde a la suma de los aminoácidos ramificados, y las variables independientes son los aminoácidos no ramificados.

La correlación es una medida que nos da un indicio sobre qué tan relacionadas se encuentran la variable dependiente y las variables independientes, la correlación toma un valor entre 1 y -1, donde los valores más cercanos a 1 indican una relación directa entre los datos en cuestión, mientras que un valor cercano a -1 indica una relación inversa, los valores cercanos a 0 indican que no existe una relación significativa. Con el valor de correlación se puede estimar a priori la importancia de cada variable para el modelo que se va a entrenar. En este trabajo luego de calcular la correlación para todas las variables se descartaron aquellas que no superan un umbral de 0.83 con respecto a la variable dependiente. Descartando variables de entrada es posible reducir los cálculos del modelo y realizar un entrenamiento más eficiente.

El entrenamiento del modelo lineal se realiza a partir de (1):

$$Y = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_n x_n \quad (1)$$

Donde Y es la variable dependiente, x son las variables independientes y b los coeficientes relativos a cada variable independiente.

El entrenamiento usa un algoritmo de optimización el cual busca los valores de los coeficientes que más reducen el error cuadrático medio, y luego de entrenarlo basta con reemplazar en (1) las variables independientes con los valores de entrada, las cuales para nuestro caso son las cantidades de aminoácidos no ramificados, y como resultado se obtendrá la predicción de la cantidad de aminoácidos ramificados [28].

Los coeficientes de correlación mayores a 0.8 generados a partir del grafico de calor usado para identificar correlaciones se presentan en la [Tabla 1](#). Las variables nutricionales que presentan los datos suministrados con mejor correlación fueron; la proteína y los aminoácidos (alanina, arginina, ácido aspártico, cisteína, ácido glutámico, glicina, histidina, lisina, metionina, fenilalanina, prolina, serina, treonina, triptófano, tirosina), los cuales presentan correlaciones superiores a 0.83; la alta correlación permite obtener un modelo predictivo con mayor precisión. Se descartan las variables nutricionales inferiores a 0.8 (carbohidratos, fibra, azúcar, niacina, riboflavina magnesio, fosforo potasio, hierro, zinc) debido a que no tienen una alta correlación para ser utilizados en el modelo de regresión lineal propuesto.

TABLA 1.
VALORES DE CORRELACIÓN DE NUTRIENTES VS BCAA DE LA BASE SUPERIORES A 0.8.

Variable	Correlación	Variable	Correlación
Proteína	0.856270	Lisina	0.974459
Alanina	0.963784	Metionina	0.973257
Arginina	0.941280	Fenilalanina	0.986270
Acido Aspártico	0.943397	Prolina	0.863694
Cisteína	0.913595	Serina	0.974902
Ácido Glutámico	0.957467	Treonina	0.991160
Glicina	0.836438	Triptófano	0.942845
Histidina	0.973936	Tirosina	0.989

Fuente: Elaboración propia.

El resultado de los coeficientes del modelo de ML y el término independiente se encuentran descritos en la [Tabla 2](#). Los coeficientes de cada variable estiman el cambio en la respuesta media por cada aumento de una unidad respecto a cada variable, cuando todos los demás predictores se mantienen constantes y el termino independiente ajusta el modelo, permitiendo relacionar los datos con la predicción.

TABLA 2.
COEFICIENTES DE REGRESIÓN LINEAL, VARIABLES Y TERMINO INDEPENDIENTE.

Variable	Coefficiente de Regresión lineal	Variable	Coefficiente de Regresión lineal
Proteína	-1.86615475e-04	Lisina	1.96327175e-01
Alanina	1.01296070e-01	Metionina	6.17459073e-01
Arginina	-1.30296547e-01	Fenilalanina	1.64684780e+00
Ácido Aspártico	2.46256128e-02	Prolina	1.67611679e-01
Cisteína	4.17694223e-01	Serina	-1.90907849e-02
Ácido Glutámico	-4.40667018e-02	Treonina	1.11330108e+00
Glicina	-9.65945865e-02	Triptofano	-2.58886832e-01
Histidina	3.77019744e-01	Tirosina	5.62038991e-01
Término Independiente: 0.00636913458725985			

Fuente: Elaboración propia.

B. Validación del Modelo

El resultado determinado por el modelo de ML del error cuadrático medio es de 0.02, indicando que el modelo de predicción tiene un valor estadístico altamente significativo.

El análisis posterior determinó la diferencia de aminoácidos ramificados en diversos alimentos y la proteína aislada de Quinoa. La [Fig. 3](#) establece las cantidades de aminoácidos ramificados predichos por el modelo de regresión lineal en g/100 g; los resultados muestran que el mayor contenido aminoacídico se encuentra en la proteína aislada de quinoa (15.9 g/100 g) y el menor en el brócoli (0.37 g/100 g). También

se puede observar la diferencia nutricional de la Quinoa según su presentación, siendo aquella que se extrae de la proteína pura, la que tiene mayor valor de BCAA (15.9 g/100 g), en segundo lugar la quinoa cruda (2.1 g/100 g) y la de menor valor, la quinoa cocida (0.66 g/100 g).

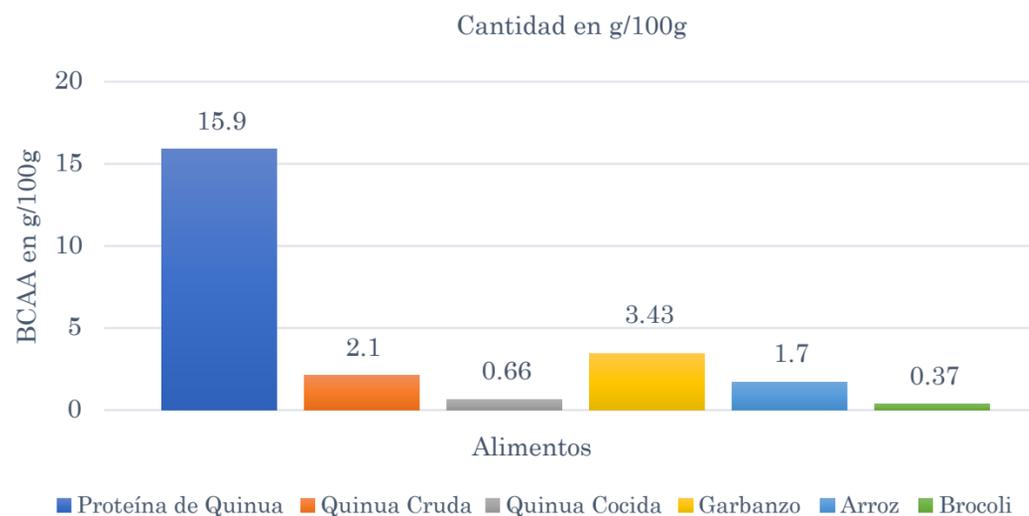


Fig. 3. BCAA presentes en proteína de quinoa, quinoa cruda, quinoa cocida, garbanzo, arroz y brócoli.
Fuente: Elaboración propia.

V. DISCUSIÓN

Con base en la fuente de datos de la USDA se encuentra una alta correlación entre los aminoácidos esenciales ramificados y los aminos esenciales no ramificados, determinando una alta probabilidad de encontrar concentraciones de macronutrientes en diversos alimentos [18], [25], generando una oportunidad en el desarrollo de nuevos productos nutraceuticos, potenciando la industria farmacéutica, alimenticia y agrícola por la posibilidad de identificar diversos componentes presentes en alimentos, que pueden ser usados como principios activos [20], [22], como es el caso de la proteína de Quinoa que tiene una alta concentración de leucina, isoleucina y valina, aminoácidos importantes en el funcionamiento del sistema musculo esquelético [2], [5].

Incluir datos precisos soportados técnicamente, es muy importante para respaldar los resultados del modelo predictor. En este artículo se utiliza la base de datos del USDA porque es una base frecuentemente actualizada y de acceso abierto, por lo tanto, uno de los conjuntos de datos más utilizados en todo el mundo [16]. Estos resultados concuerdan con datos encontrados previamente en la literatura que corroboran la presencia y cantidad de aminoácidos de cadena ramificada (leucina, isoleucina y valina) en la proteína aislada de Quinoa [17].

Nuevos retos alimenticios llevan a que el ser humano consuma alimentos de fuentes vegetales ya que la salud se ha visto desde un concepto más holístico y complejo [29]. Por lo tanto, la presencia de mayor cantidad de BCAA en la proteína aislada de quinoa y no en las semillas de quinoa cruda o cocida, manifiestan la importancia de realizar procesos de extracción óptimos en la Quinoa, y consumir proteína de Quinoa en su forma aislada garantiza una dieta con mayor concentración de aminoácidos ramificados favoreciendo a la salud del sistema musculo esquelético [5], [7].

VI. CONCLUSIONES

Este artículo implementa regresión lineal y analiza los nutrientes presentes en la Quinoa con modelos de ML; uno de los principales retos es comparar la gran cantidad de datos suministrados con una alta precisión, la cual se obtuvo con resultados muy favorables; 15.9 g/100g de BCAA; estos resultados contribuirían a un posible efecto positivo en la salud muscular. Además, el gran aporte de aminoácidos esenciales en su composición nutricional, podría ser una alternativa proteica al cambio dietario del siglo XXI.

Trabajos relacionados [18], [20], [22], incorporan métodos de ML en investigación nutricional, como una herramienta tecnológica, para facilitar la integración de muchas variables de información, esto podría favorecer tanto a ingenieros de alimentos como agricultores, proporcionando análisis predictivos que identifiquen la calidad nutricional de alimentos promoviendo la industria.

Se puede observar que herramientas tecnológicas como ML dan apertura a la creación de plataformas tecnológicas como aplicaciones móviles las cuales permitan disponer información nutricional, promoviendo el desarrollo de fórmulas nutraceuticas con optimas concentraciones, además estas aplicaciones podrían identificar la presencia de macro y micronutrientes, proporcionando parámetros de calidad alimenticia apoyando la toma de decisiones en la ingesta de alimentos, promoviendo dietas integrales que mejoren la salud de las personas.

AGRADECIMIENTOS

Agradecemos a Natural Company, en particular al equipo de investigación y desarrollo por los recursos facilitados durante el desarrollo del presente trabajo.

REFERENCIAS

- [1] A. Bojanic, *Organización de las Naciones Unidas para la Alimentación y la Agricultura, La Quinoa: Cultivo milenario para contribuir a la seguridad alimentaria mundial*, STGO, CL: FAO/RLC, 2011. Disponible en <https://www.fao.org/family-farming/detail/es/c/326239/>
- [2] E. Soto, W. Mercado, R. Estrada, R. Repo, F. Díaz, F. y G. Díaz, *El mercado y la producción de quinua en el Perú*. LI, PE: IICA/INIA, 2015. Disponible en <http://repositorio.inia.gob.pe/handle/20.500.12955/1257>
- [3] M. Oliva, A. Duque y L. García-Alzate, “Caracterización fisicoquímica del cereal y almidón de Quinoa *Chenopodium quinoa*”, *Rev Ion*, vol. 31, no. 1, pp. 25–29, Sep. 2018. <http://dx.doi.org/10.18273/revion.v31n1-2018004>
- [4] D. Ferrier, *Bioquímica*, 6 ed. PHL, USA: Lippincott Williams & Wilkin, 2014.
- [5] R. Rajendram, V. Preedy & V. Patel, *Branched Chain Amino Acids in Clinical Nutrition*, vol. 1. NY, USA: Humana Press, 2015. <https://doi.org/10.1007/978-1-4939-1923-9>
- [6] D. Rojas, F. Figueras y S. Durán, “Ventajas y desventajas nutricionales de ser vegano o vegetariano”, *Rev chil nutr*, vol. 44, no. 3, pp. 218–225, 2017. Disponible en <https://search.bvsalud.org/gim/resource/es/biblio-899824>
- [7] J. Physiol & K. Atherton, “Muscle protein synthesis in response to nutrition and exercise”, *J Physiol*, vol. 590, no. 5, pp. 1049–1057, Jan. 2012. <https://doi.org/10.1113/jphysiol.2011.225003>
- [8] K. Smith, N. Reynolds, S. Downie, A. Patel & M. J. Rennie, “Effects of flooding amino acids on incorporation of labeled amino acids into human muscle protein,” *Am J Physiol*, vol. 275, no. 1, pp. E73–E78, Jul. 1998. <https://doi.org/10.1152/ajpendo.1998.275.1.E73>
- [9] M. S. Brook, P. Scaife, J. J. Bass, J. Cegielski, S. Watanabe, D. J. Wilkinson, K. Smith, B. E. Phillips & P. J. Atherton, “A collagen hydrolysate/milk protein-blend stimulates muscle anabolism equivalently to an isoenergetic milk protein-blend containing a greater quantity of essential amino acids in older men,” *Clin Nutr*, vol. 40, no. 6, pp. 4456–4464, Jun. 2021. <https://doi.org/10.1016/j.clnu.2021.01.002>
- [10] H. Hiraiwa, T. Okumura, T. Kondo, T. Kato, S. Kazama, T. Ishihara, E. Iwata, M. Shimojo, S. Kondo, S. Aoki, Y. Kanzaki, D. Tanimura, H. Sano, Y. Awaji, S. Yamada & T. Murohara, “Usefulness of the plasma branched-chain amino acid/aromatic amino acid ratio for predicting future cardiac events in patients with heart failure,” *J Cardiol*, vol. 75, no. 6, pp. 689–696, Jan. 2020. <https://doi.org/10.1016/j.jjcc.2019.12.016>
- [11] C. Gutiérrez, M. Lares, J. Sandoval y M. Hernández, “Aminoácidos de Cadena Ramificada”, *Rev Dig Postgrado*, vol. 9, no. 2, pp. 1–6, 2018. <https://doi.org/10.37910/RDP.2020.9.2.e224>
- [12] G. R. Morantes-Quintana, G. R. Rincón-Polo & N. A. Pérez-Santodomingo, “Multiple linear regression model to estimate PM 1 concentration”, *Rev Int Contam Ambie*, vol. 35, no. 1, pp. 179–194, 2019. <https://doi.org/10.20937/rica.2019.35.01.13>
- [13] IBM Corporación, *Regresión lineal*, 2021, abril 13. [Online]. Disponible en <https://www.ibm.com/co-es/analytics/learn/linear-regression>
- [14] R. Montero, “Modelos de regresión lineal múltiple”, *Documentos de Trabajo en Economía Aplicada*, 2016. Recuperado de https://www.ugr.es/~montero/matematicas/regresion_lineal.pdf
- [15] M. Gupta, “ML | Linear Regression”, *Geeks for Geeks*, 18 May. 2022. [Online]. Disponible en <https://www.geeksforgeeks.org/ml-linear-regression>
- [16] *Download FoodData Central Data*, USDA, abril, 2022. [Online]. Disponible en <https://fdc.nal.usda.gov/>
- [17] P. Vargas, R. Arteaga y L. Cruz, “Análisis bibliográfico sobre el potencial nutricional de la quinua (*chenopodium quinoa*) como alimento funcional”, *Rev CAZ*, vol. 46, no. 4, pp. 89–100. Dic. 2019. Disponible en http://centroazucar.uclv.edu.cu/index.php/centro_azucar/article/view/19
- [18] D. Kirk, C. Catal & B. Tekinerdogan, “Precision nutrition: A systematic literature review,” *Comp Biol Med*, vol. 133, pp. 104365–104365, Jun. 2021. <https://doi.org/10.1016/j.compbiomed.2021.104365>
- [19] G. K. Golla, J. A. Carlson, J. Huan, J. Kerr, T. Mitchell & K. Borner, “Developing Novel Machine Learning Algorithms to Improve Sedentary Assessment for Youth Health Enhancement”, presented at *2016 ICHI*, IEEE, CH, IL, USA, 04-07 Oct. 2016, pp. 375–379. <https://doi.org/10.1109/ICHI.2016.67>
- [20] J. Meerasri & R. Sothornvit, “Artificial neural networks (ANNs) and multiple linear regression (MLR) for prediction of moisture content for coated pineapple cubes,” *Case Stud Therm Eng*, vol. 33, pp. 101942–101942, May. 2022. <https://doi.org/10.1016/j.csite.2022.101942>
- [21] G. Çakmak & C. Yıldız, “The prediction of seedy grape drying rate using a neural network method”, *Comput Electron Agric*, vol. 75, no. 1, pp. 132–138, Jan. 2011. <https://doi.org/10.1016/j.compag.2010.10.008>
- [22] V. Siy, V. Antonio, C. Siguin, N. Gordoncillo, J. Sescon, C. Go & E. Miro, “Predicting undernutrition among elementary schoolchildren in the Philippines using machine learning algorithms,” *Nutrition*, vol. 96, pp. 111571–111571, Apr. 2022. <https://doi.org/10.1016/j.nut.2021.111571>
- [23] A. Antónov, “Valores nutricionales de alimentos y productos comunes”, *Kaggle*, 2019, julio 18. Disponible en <https://www.kaggle.com/datasets/trolukovich/nutritional-values-for-common-foods-and-products>

- [24] A. Beklemysheva, “Why Use Python for AI and Machine Learning?”, *Steel kiwi*, 2022, enero. Disponible: <https://steelkiwi.com/blog/python-for-ai-and-machine-learning/>
- [25] M. Silva, M. Ribeiro, O. Viegas, Z. Martins, M. Faria, S. Casal, E. Pinto, A. Almeida, O. Pinho & I. Ferreira, “Exploring two food composition databases to estimate nutritional components of whole meals,” *J Food Compos Anal*, vol. 102, no. 5, pp. 104070–104070, Sep. 2021. <https://doi.org/10.1016/j.jfca.2021.104070>
- [26] V. Naoum, D. Kyriopoulos, A. Charonis, K. Athanasakis & J. Kyriopoulos, “The Pareto Principle (“80–20 Rule”) In Healthcare Services In Greece,” *Value Health*, vol. 19, no. 7, pp. 18–18, 2016. <https://doi.org/10.1016/j.jval.2016.09.1563>
- [27] P. Rattan, D. Penrice & D. Simonetto, “Artificial Intelligence and Machine Learning: What You Always Wanted to Know but Were Afraid to Ask,” *Gastro Hep Advances*, vol. 1, no. 1, pp 70–78, Nov. 2022. <https://doi.org/10.1016/j.gastha.2021.11.001>
- [28] A. Banerji, “Gradient Descent in Linear Regression,” *Analytics Vidhya*, 2021, abril 19. Available: <https://www.analyticsvidhya.com/blog/2021/04/gradient-descent-in-linear-regression/>
- [29] B. Gutiérrez & Gómez, “Scoping Review of Complexity Science in Dentistry,” *Dent Hypotheses*, vol. 12, no. 3, pp. 109–117, Nov. 2021. Available: <http://www.dentalhypotheses.com/text.asp?2021/12/3/109/329758>

Nicolas S. Caicedo Cerón recibió su título de Regente de Farmacia de la Universidad de Antioquia (Colombia). También obtuvo el título de Administrador Comercial y de Mercadeo de la Institución Universitaria ESUMER (Colombia). En el año 2021 empieza estudios de Doctorado de Ciencias Aplicadas de la Universidad Santiago de Cali (Colombia). Sus áreas de interés se enmarcan en el desarrollo de productos nutraceúticos.

Juan J. Caicedo Cerón recibió su título de Ingeniero de Sistemas e Informática de la Universidad Nacional (Colombia). Actualmente cuenta con 10 años de experiencia en el área de desarrollo de software.

José F. Oñate Garzón recibió su título de Biólogo de la Universidad del Cauca (Colombia). También obtuvo el título en Magister en Mutagénesis y Doctorado en biología-Bioquímica y biofísica de la Universidad de Antioquia (Colombia).