

Detección de desnutrición crónica en niños menores a cinco años aplicando técnicas multivariantes.

Detection of Chronic Malnutrition in Children Under Five Using Multivariate Techniques

<http://doi.org/10.53358/ideas.v7i1.1072>

Dennys R. Daquilema, Erick P. Herrera

Centro de Posgrado, Universidad Politécnica Estatal del Carchi, Posgrados,
 Av. Universitaria y Antisana, Tulcán, Ecuador

dennys.daquilema@upec.edu.ec, erick.herrera@upec.edu.ec

Fecha de envío, marzo 7/2024 - Fecha de aceptación, abril 11/2024 - Fecha de publicación, enero 30/2025

Resumen: El presente artículo muestra el diseño de cinco clasificadores para la detección de desnutrición crónica en niños menores a cinco años, para lo cual, se procesó una base de datos con 5 variables cuantitativas y 26 cualitativas que fueron sometidas a una disminución de dimensión mediante un análisis de correspondencia múltiple, posteriormente, mediante el uso de redes neuronales se generó el primer clasificador que obtuvo un accuracy de 1 y un error cuadrático medio (mse) de 0.0050. El criterio que determinó el número óptimo de neuronas en la capa oculta diferencia a los tres clasificadores restantes basados en inteligencia artificial, mismos que partieron de un análisis de componentes principales que estableció una arquitectura con una única capa oculta que alberga el 96.3% de la varianza acumulada. La primera propuesta de red neuronal poco profunda utilizó un método de iteración que determinó 51 neuronas en la capa oculta con un accuracy de 0.846 y un mse de 0.145, el segundo clasificador posee 42 neuronas que corresponden a 2/3 del número de neuronas de la capa anterior, mostró 0.769 de accuracy y 0.217 de mse, por último, el método de Xu and Chen estableció que la capa oculta cuente con tres neuronas, generando un accuracy de 1 y un mse de 0.0075. Además, se implementó un modelo basado en regresión logística para comparar la efectividad de las propuestas respecto a esta técnica de la estadística tradicional. El estudio finalizó con un protocolo de pruebas que permitió determinar el mejor clasificador propuesto para esta tarea e identificar las variables más influyentes, se obtuvo que el primer clasificador sin capas ocultas fue el que presentó el mejor rendimiento, demostrando mayor exactitud y menor cantidad de errores frente a los otros clasificadores, con el peso y la talla como las variables más relevantes.

Palabras Clave: Desnutrición Crónica, Redes Neuronales Artificiales, Clasificador, Técnicas Multivariantes, Regresión Logística

Abstrac: This article presents the design of five classifiers for the detection of chronic malnutrition in children under the age of five. To achieve this, a dataset with 5 quantitative variables and 26 qualitative variables was processed, with dimensionality reduction performed through a multiple correspondence analysis. Subsequently, a neural network-based classifier was created, achieving an accuracy of 1 and a mean squared error (MSE) of 0.0050.

The criterion used to determine the optimal number of neurons in the hidden layer differentiated the three remaining artificial intelligence-based classifiers. These classifiers started with a principal component analysis that defined an architecture with a single hidden layer, capturing 96.3% of the accumulated variance. The first shallow neural network proposal used an iterative method that determined 51 neurons in the hidden layer, resulting in an accuracy of 0.846 and an MSE of 0.145. The second classifier, with 42 neurons corresponding to two-thirds of the previous layer's neurons, showed an accuracy of 0.769 and an MSE of 0.217. Finally, the Xu and Chen method established a hidden layer with three neurons, achieving an accuracy of 1 and an MSE of 0.0075.

Additionally, a logistic regression-based model was implemented to compare the effectiveness of these proposals against this traditional statistical technique. The study concluded with a testing protocol that determined the best classifier for this task and identified the most influential variables. The results revealed that the first classifier without hidden layers showed the best performance, demonstrating greater accuracy and fewer errors compared to the other classifiers, with weight and height being the most significant variables.

Keywords: Chronic Malnutrition, Artificial Neural Networks, Classifier, Multivariate Techniques, Logistic Regression

Autor de correspondencia:

Dennys R. Daquilema, dennys.daquilema@upec.edu.ec

Introducción

La desnutrición infantil crónica se ha convertido en uno de los mayores problemas que aqueja a los diferentes gobiernos, ya que generan gastos de forma directa e indirecta. Niños con desnutrición presentan enfermedades que le cuestan altas inversiones a los gobiernos de turno que buscan eliminar o mitigar dichas enfermedades mediante sus centros de salud pública, además, niños con malos hábitos alimenticios no se desarrollan correctamente ni de forma física ni intelectual, con lo cual, disminuyen la capacidad laboral y desarrollo de un país.

De acuerdo al informe sobre desnutrición presentado por la Organización Mundial de la Salud, "se calcula que 149.2 millones de niños padecen de retraso en el crecimiento y 340 millones de menores a 5 años sufren carencias de vitaminas y de otros nutrientes esenciales" [1]. La desnutrición crónica en niños tiene una alta tasa de incidencia en países en vías de desarrollo, así como varios sectores del área rural, es así que existe una mayor probabilidad de mortalidad en un niño con este tipo de enfermedad que uno cuyo estado nutricional es normal, ya que sus defensas debilitadas no pueden hacer frente a las enfermedades y, cuando la desnutrición se complica, el riesgo de muerte es muy alto.

La Encuesta Nacional de Salud y Nutrición (ENSANUT) resalta la existencia de un 23% de menores de cinco años con desnutrición, porcentaje que no mostró diferencia estadística significativa con la realizada en el 2014 [2]. Finalmente, de acuerdo a un estudio realizado por la Universidad Técnica del Norte (UTN), la provincia de Esmeraldas posee un 20% de menores de cinco años con desnutrición, cifra que, a pesar de ser alta, ubica a la provincia por debajo de los valores medios de todo el país [3].

Son incontables los efectos nocivos de la malnutrición infantil, que no solo afectan al infante sino también a los gobiernos zonales que, según la UNICEF los gastos asociados a la malnutrición son del 4,3% del producto interno bruto ecuatoriano [4], además, en el estudio del impacto social y económico de la malnutrición realizado por la CEPAL, revelan cifras alarmantes generadas por la desnutrición infantil, es así que, los costos en salud en nuestro país es de 44 millones de dólares para eliminar o mitigar enfermedades asociadas a una mala alimentación, en educación aumentó el gasto de 4,2 a 27,2 millones producto del bajo rendimiento académico, y en productividad se pierde un aproximado a mil millones de dólares por la disminución de la esperanza de vida [5].

Basado en lo expuesto, la necesidad de detectar la desnutrición crónica a edades tempranas es prioritaria en todo gobierno de turno, con lo actual, se refuerza el objetivo del presente trabajo que compara técnicas multivariantes para la detección de desnutrición crónica, mismas que fueron aplicadas a una base de datos que recoge la información de niños menores a cinco años que presentaron desnutrición crónica en el centro de salud Nuevo San Rafael de la provincia de Esmeraldas desde mayo del 2021 a abril del 2022.

Trabajos Relacionados

El estudio [6] desarrolló un modelo validado con Machine Learning (ML) para predecir el estado de desnutrición de niños con cardiopatías congénitas. La base de datos estudiada consta de 536 niños con cardiopatía congénita 1 año después de su cirugía, desde diciembre de 2017 hasta mayo de 2020. La base de datos se dividió en un grupo de entrenamiento y otro de prueba, se emplearon cinco técnicas de ML, (Regresión logística (LR), support vector machine (SVM), adaptive boosting (ADA), perceptron multicapa (MLP)

y extreme gradient boosting (XGBoost)) y el resultado fue comparado con el área bajo curvas características y SHARP (Shapley Additive exPlanations). Los parámetros fueron ajustados usando validación cruzada 5 veces junto con el método de búsqueda de mallas (grid-search method). La salida es la desnutrición, pero también catalogaron bajo peso, retraso de crecimiento y emaciación. Para disminuir un posible sobre ajuste se seleccionó las características estadísticamente significativas y luego se sintetizaron con una eliminación recursiva (RFE) basado en random forest. El Bootstrap method fue aplicado con 1000 repeticiones para determinar el intervalo de confianza, precisión, y sensibilidad. Para la interpretación se emplea SHARP. Las medias, desviaciones estándar, rangos intercuantiles, se utilizaron para analizar y expresar las variables continuas, las cuales se sometieron al test t o al test de Mann – Whitney (test U). Las categóricas se expresan como porcentajes y se comparan mediante prueba chi-cuadrado, para variables con proporciones faltantes al 20% imputaron las variables categóricas con la moda, y las variables continuas con imputación múltiple. El modelo ML (XGBoost) fue el que obtuvo el mejor rendimiento con un porcentaje alto de (0.87 a 0.91) de AUC, además, concluyeron que un nivel educativo bajo en los padres, y un ingreso económico bajo son factores asociados a los casos que presentaron bajo peso. Uno de los aportes más importantes de esta investigación fue la generación de predictores para determinar el momento más apropiado para realizar una intervención nutricional. La no diferenciación racial y la aplicación únicamente a niños sometidos a una cirugía constituyen como las principales recomendaciones de mejora para este trabajo.

Con el estudio [7] se identificaron los factores de riesgo potenciales en mujeres de Bangladesh mediante un modelo basado en Machine Learning (ML), se empleó una regresión logística multinomial para extraer los posibles factores de desnutrición. Además, se emplearon cinco métodos de inteligencia artificial (Naive Bayes (NB), support vector machine (SVM), árboles de decisión (DT), redes neuronales (ANN), random forest (RF)), cuyos desempeños fueron evaluados por el método denominado accuracy and under the curve (AUC) con tres conjuntos de protocolos de validación cruzada (K2,K5,K10). Para la generación de la base de datos se tomó un total de 15464 muestras de la encuesta demográfica de Bangladesh del 2014 levantada con muestreo por conglomerados estratificados, el estudio fue realizado con un 95% de nivel de confianza, y el 99% de las entrevistadas eran mujeres casadas entre edades de 15 a 49 años. La edad, región, índice de riqueza, nivel de educación, amamanta o no, trabajando o no, educación del esposo, fuente de agua que consume, tipo de aceite para cocinar fueron identificados como factores significativos para determinar la desnutrición, también establecieron factores para determinar obesidad y sobrepeso. El clasificador basado en RF tiene la más alta precisión (81,4% y 82.4%) debido a que posee una gran cantidad de variables categóricas y de tendencia no lineal, y el método de menor rendimiento es el de NB. Un dato de mejor es ampliar el estudio a mujeres de una zona geográfica lejana a la indicada.

En [8] se tuvo como principal objetivo identificar si un niño se encuentra en estado de desnutrición con retraso en el crecimiento, emaciación, bajo peso, o si tiene edema nutricional, los datos para el estudio provinieron de la encuesta SMART de nutrición de Afganistán de 2015 a 2018, que cuenta con 2131 registros que incluyen fecha de nacimiento, fecha de visita, edad, sexo, altura, peso, provincia, circunferencia del brazo medio superior (MUAC), y características del edema nutricional. Las técnicas empleadas fueron Random Forest, Part rule induction, Naïve Bayes, y sus resultados fueron comparados con métodos de regresión logística. Se utilizaron 5 pasos, comprensión del dominio del problema (leer, contar con expertos para dominar el problema), análisis de datos (identificar cada característica de la base de datos), preparación de datos (se establecieron puntajes Z para peso/edad (WAZ),

altura/edad (HAZ), peso/altura (WHZ), se tomaron estas variables basadas en el estándar de la OMS $z_score=(X-M)/\sigma$ para la talla que sigue una distribución normal, mientras que el peso para la edad, y el peso para talla y la longitud se utilizó $Z_Score=((X-M)^L-1)/LS$, minería de datos (Random Forest (RF), part rule induction, Naive Bayes (NB), y Regresión Logística (LR). "WEKA machine learning tool". Además, se dividió la base de datos en dos grupos, 80% entrenamiento, 20% para las pruebas. "Correlation Attribute Eval" se usa para evaluar la importancia de los atributos, y "Synthetic Minority Oversampling Technique" (SMOTE) para equilibrar los datos. La principal conclusión obtenida fue que RF y PART fueron los métodos más precisos con un 99.99%, NB del 95,82%. Dentro de los datos negativos está que los datos atípicos fueron eliminados, al igual que los datos que mostraban obesidad, y los faltantes reemplazados por la media y que no toma en cuenta datos socioeconómicos.

El estudio [9] se centró en determinar la existencia de desnutrición, para lo cual se consideraron como variables, la edad del niño y la madre, sexo de los niños, tamaño de la familia, ingreso mensual, educación y ocupación de la madre y estado civil, altura, peso, orden de nacimiento, y enfermedad infantil, lactancia materna, número de hijos. La base de datos se generó a partir de una encuesta para las variables cualitativas, y capacitaron por dos días al personal que levantó el instrumento, y se encuestaron a padres cuyos hijos estaban en la edad de 6 a 59 meses de edad de la ciudad de Hawassa, en total se tomaron 811 en un plazo de agosto a septiembre de 2012. El método empleado fue el análisis fue una regresión logística bi variada con "Crude Odds Ratio al 95% CI" y para identificar la relación entre variables, un estudio transversal con un $\alpha=5\%$. Establecieron predictores para los tres tipos de desnutrición (crónica, bajo peso y emaciación), pero lastimosamente la publicación no muestra el rendimiento de los predictores, sino que simplemente se centran en comparar lo obtenido con índices en otros países. Además, compara los resultados de cada variable y cuantifican las relaciones y efectos entre ellas.

El diseñar y desarrollar un sistema experto que sea capaz de identificar la desnutrición severa en niños de 0 a 5 años fue el objetivo principal de [10], que empleó un sistema experto basado en el método de encadenamiento directo con base la antropometría que es el estudio de la geometría y la naturaleza de la masa. Para el desarrollo del sistema, experto definieron dos reglas, una premisa y una conclusión. El grupo de estudio fueron niños de una edad menor a 5 años de Indonesia, y como principal resultado fue la interfaz del sistema experto que permite ingresar variables y concluir si hay la posibilidad de que exista desnutrición. Además, emplearon PHP y MySQL y dejaron de lado el análisis estadístico.

En [11] se empleó el método Naive Bayes, cuya principal fortaleza radica en ofrecer excelentes resultados de predicción de parámetros con bases de datos reducidas (pocos datos). Este algoritmo trabaja con la premisa de que la presencia o ausencia de una variable (característica) no está ligada a la presencia o ausencia de otra. La base de estudio contiene datos de niños de entre 1 a 5 años de edad, y para nuevos análisis permite ingresar los datos vía web, lastimosamente no indica el rendimiento y porcentajes de acierto en los entrenamientos.

La investigación [12] empleó una encuesta descriptiva donde se determinó la correlación entre variable mediante lenguaje de programación Python y con algoritmos híbridos AdaBoost y árboles de decisión. El principal objetivo fue desarrollar un sistema de planificación de alimentos recomendados para el control de la desnutrición en niños de 6 a 12 años utilizando algoritmos híbridos de aprendizaje automático. Emplearon una

encuesta descriptiva donde se determinó la correlación entre variables con utilización de lenguaje Python con algoritmos híbridos AdaBoost y árboles de decisión. El conjunto de datos analizados incluye 1001 puntos recopilados de los centros de salud de la ciudad de Mianeh ubicada en el este de Azerbaiyán en Irán, desde el sistema web integrado de Apple. Mediante árboles de decisión llegaron a la conclusión que el número de características en los alimentos ingeridos, es el dato más importante, empleando una precisión del 98,5%.

Adicionalmente, se puede citar la investigación de [13], basada en una encuesta del año 2014 realizada en Bangladesh, donde analizaron a 7079 niños. Los resultados obtenidos de esta investigación reflejan un 35.4% en el retraso del crecimiento, 15,4 % emaciación y el 32.8% bajo peso. Mediante la regresión logística determinaron dentro de los retrasos de crecimiento que cinco son los factores que generan mayor riesgo. Los estudios se han centrado en la población de Bangladesh dado que más de mitad de la población padece de desnutrición, y la mayor diferencia con el estudio expuesto en [7] que realizó un análisis en la misma localidad, radica en el ajuste de los hiperparámetros en los algoritmos de inteligencia artificial mejorando los resultados de precisión de los predictores. Técnicas de Support Vector Machine (SVM), Random Forest (RF) fueron comparados mediante el criterio de área bajo la curva (AUC) que les permitieron validar una mejor precisión con valores de un p-valor menor al 0.001, además, analizaron la relación entre variables explicativas y la desnutrición con el empleo de una prueba chi-cuadrado. Un aporte principal es que lograron determinar las regiones con mayor cantidad de desnutrición (Sylhet) y también la de menor porcentaje (Khulna). Los resultados podrían mejorar con un análisis de datos agrupados y con un análisis posterior de afinamiento con la técnica de Bonferroni.

La investigación [14] contó como principal objetivo detectar la desnutrición en niños menores de 5 años con un grupo de imágenes de 500 niños mediante el uso de redes neuronales convolucionales (CNN). El conjunto de datos al ser analizados por AlexNet con los valores de niños con desnutrición y niños sanos. La arquitectura denominada AlexNet se caracteriza por conseguir una alta precisión y fue empleada en la base de datos de ImageNet, la cual sirve para entrenar la CNN para diferenciar y clasificar un total de 1000 categorías de objetos. Para el estudio se utilizó una población de 500 imágenes de niños menores de cinco años, del cual, del 100% de los datos, el 90% fueron utilizados para el aprendizaje del sistema y el 10% para la realización de la prueba. El CNN mostró un 96% de precisión en la determinación de resultados, la misma que determinaba si los niños estaban sanos, con desnutrición o riesgo de mal nutrición. La principal salida de resultados del sistema es un denominado "dashboard" que representa mediante gráfico de barras y diagramas de pastel los parámetros de desnutrición. El número de predicciones correctas del sistema fue empleado como medida de rendimiento, y se empleó en imágenes de 227x227x3 de 250 niños con desnutrición y 250 niños sanos. Para mejorar el resultado se desearía emplear el método para detectar el tipo de desnutrición.

Materiales y métodos

El estudio que se detalla en el presente documento analiza una base de datos constituida a partir de 122 niños que asistieron al centro de salud Tipo C, en San Rafael - Esmeraldas (CSNSR) desde mayo del 2021 hasta abril del 2022. Dicha base de datos es fruto de [15] donde se tabuló los datos provenientes de un cuestionario con preguntas cerradas de opción múltiple, misma que fue aplicada a los familiares y/o cuidadores de niños que mostraron bajo peso y estatura, además incorpora, registros históricos del personal del centro de salud. El instrumento fue validado por un Alfa de Cronbach igual a 0,918 que es uno de los resultados del anterior trabajo de titulación.

Para el desarrollo y determinación de resultados se emplean métodos que combinan el estudio de variables tanto cuantitativas como cualitativas. La metodología cualitativa es necesaria por el uso de técnicas multivariantes en valores antropométricos, mientras que la metodología cuantitativa para determinar, interpretar y predecir la desnutrición en niños; variables utilizadas por médicos pediatras y autoridades para desarrollar planes de acción en contra de la desnutrición.

El desarrollo del mismo, parte del uso de datos obtenidos en una encuesta aplicada a personas adultas con hijos, los mismos que visitaron el centro de salud por problemas de desnutrición crónica, la información obtenida prevé ser de vital importancia para su estudio, ya que conjuga el estado de salud actual del infante, datos antropométricos, historial clínico, así como también la información socioeconómica e historial clínico de su familia. Como punto de partida se utilizaron los datos antropométricos y somatométricos (Peso, talla, hemoglobina, edad, sexo), que combinados con los datos socio-económicos del responsable del infante (estado civil, dependencia laboral, grado académico, ingreso económico, charlas recibidas, parentesco, número de hijos, consumo de vitaminas del bebe, visitas domiciliarias de personal médico, control médico al infante, número de comidas del infante), se busca determinar las relaciones entre variables y establecer un clasificador que permita predecir la presencia de desnutrición crónica.

Tratamiento de Datos

Para el estudio es necesario la aplicación de un protocolo de análisis de datos, y la primera fase consiste en eliminar la posible presencia de datos atípicos. Las distancias de Mahalanobis es el método que permite detectar atípicos basado en la correlación que existe entre variables, es por tal razón que, a pesar de ser una técnica desarrollada en los años 30, sigue siendo una de las metodologías más utilizadas [16].

$$d_m(\vec{x}_1, \vec{x}_2) = \sqrt{(\vec{x}_1 - \vec{x}_2)^T \Sigma^{-1} (\vec{x}_1 - \vec{x}_2)} \quad (1)$$

La ecuación 1 citada en [17] y [18] muestra la relación del método con la geometría tradicional que incorpora la matriz de covarianza y permite hallar valores denominados distancias, mismas que al ser comparadas, proporcionan una técnica para saber que tan lejos o cerca se encuentra una observación con respecto a un grupo de datos. Semipositividad, simetría, y desigualdad triangular son las propiedades que cumplen las distancias de Mahalanobis, además, en [17] se resalta aplicaciones innovadoras del método en diferentes campos como el monitoreo de la fatiga, y el procesamiento de datos.

Análisis de componentes principales (ACP)

Es una técnica multivariante que se aplica en variables cuantitativas, considerada también como una herramienta de aprendizaje no supervisado, que se dedica al análisis de variables aleatorias inter correlacionadas, y que busca hallar relaciones y subconjuntos entre dichas variables mediante la representación de información en nuevas variables ortogonales, denominadas componentes principales. Según Hervé y Lynne, la extracción de información relevante, la reducción de dimensionalidad de variables, y el análisis de la estructura de las observaciones se constituyen en los objetivos de este método [19].

El ACP no se basa en el supuesto de normalidad multivariante de los datos, y descrito de una forma matemática, no es más que una rotación de los ejes de coordenadas de las variables originales a un nuevo sistema ortogonal cuyos ejes coincide con la dirección de máxima varianza. El método parte de la construcción de una matriz X de dimensión $I \times J$ donde I son las observaciones y J son las variables [19]. Es común que una base de datos cuente con variables medidas en diferentes unidades, razón por la cual, se hace necesario dividir cada observación para la norma de la matriz en un proceso denominado normalización. Mediante la expresión $|X - \lambda I| = 0$ es posible hallar el polinomio característico de la matriz, cuyas raíces son los denominados valores propios (eigenvalores), y que a partir de los mismos se determinan los vectores característicos (eigenvectores).

$$X = U\Delta V^T \quad (2)$$

La ecuación 2, representa la matriz X como un producto de matrices donde U es la matriz ortonormal formada por los vectores propios de XX^T (right singular vectors), Δ es una matriz diagonal cuyas entradas son los valores propios de la matriz X ordenados en forma descendente, y la matriz ortonormal V es la formada por los vectores propios de $X^T X$ (left singular vectors).

$$I = \sum_i^J x_{i,j}^2 \quad (3)$$

$$d_{i,g}^2 = \sum_j^J (x_{i,j} - g_j)^2 \quad (4)$$

Las ecuaciones 3 y 4 obtenidas de [20] muestran la denominada inercia total de las columnas de la base de datos, y la distancia euclídea de las observaciones al centro de gravedad de cada columna (centroide o baricentro). El centro de gravedad no es nada más que el vector de medias de cada columna y que en el instante en que dicho vector sea igual a cero, las ecuaciones 3 y 4 resultan ser idénticas. Las componentes principales son obtenidas a partir de la descomposición en valores singulares (SVD) de la matriz X , dicha técnica no es más que una variación del método de descomposición de valores y vectores propios que le permite ser aplicado tanto a matrices cuadradas como rectangulares.

El ACP a más de ser una roto traslación alrededor del origen, también se lo interpreta como una proyección ortogonal hacia nuevos ejes, \bar{u}_α razón por la cual, cada componente principal ψ se obtiene de una combinación lineal de las variables originales X .

$$\psi_\alpha = u_\alpha X \quad (5)$$

La ecuación 5 muestra a las n componentes principales producto de las n proyecciones sobre el eje canónico u_α , con lo cual, con $\alpha=1$ se genera la primera componente principal de la matriz X que se representa en la ecuación 6.

$$\psi_1 = u_{11}X_1 + u_{21}X_2 + u_{31}X_3 + \dots + u_{1j}X_j \quad (6)$$

Según [20] a los valores del vector u_α se los denominan loadings y se los puede interpretar como la razón entre el peso y la importancia que tiene cada variable en cada componente, además, es importante señalar que cada componente principal establece un modelo simplificado del modelo original, es aquí donde surge la necesidad de incluir el concepto de proporción de varianza explicada acumulada, que permite establecer un criterio de decisión para el número óptimo de componentes principales que minimice la pérdida de información y reduzca el modelo de estudio.

$$\sum_{j=1}^J Var(X_j) = \sum_{j=1}^J \frac{1}{n} \sum_{i=1}^n x_{i,j}^2 \quad (7)$$

$$\frac{1}{n} \sum_{i=1}^n \psi_{i,m}^2 = \frac{1}{n} \sum_{i=1}^n \left(\sum_{j=1}^J u_{jm} x_{ij} \right)^2 \quad (8)$$

$$\frac{\sum_{i=1}^n \left(\sum_{j=1}^J u_{jm} x_{ij} \right)^2}{\sum_{j=1}^J \sum_{i=1}^n x_{i,j}^2} \quad (9)$$

La ecuación 7 representa la varianza total de base de datos normalizados, mientras que 8 indica la varianza explicada por la m-ésima componente principal, y la ecuación 9 la proporción de varianza explicada [21].

Análisis de correspondencias múltiples (ACM)

Es uno de los métodos estadísticos factoriales que se aplica en variables categóricas, que básicamente representa a los datos de análisis en puntos sobre un plano coordinado. Se lo considera también como una adaptación del ACP para datos categóricos, dado que, ambas técnicas convierten a la base de datos en una matriz normalizada para posteriormente descomponerla con base en sus valores y vectores propios en caso de ser cuadrada, o utilizando el SVD para matrices cuyo número de filas son distinto al número de columnas [22].

El método se basa en la construcción de tablas de Burt y de contingencia (cruce de variables consigo misma y de los pares considerados relevantes) [23], para lo cual, a la base de datos se la convierte en una matriz X de dimensión $I \times J$, misma que al ser normalizada $x_{ij}/\|X\|$ toma el nombre de matriz de correspondencia P , posteriormente, se determina las frecuencias marginales relativas (masas) $p_i = x_i/\|X\|$ y $p_j = x_j/\|X\|$ denominadas r_i y c_j respectivamente. Basado en el supuesto de independencia de los residuos, tenemos:

$$s_{ij} = \frac{x_{ij} - r_i c_j}{\sqrt{r_i c_j}} \quad (10)$$

cuya forma matricial es:

$$S = D_r^{-1/2} (X - r c^T) D_c^{-1/2} \quad (11)$$

Las ecuaciones 10 y 11 muestran la matriz de residuos estandarizados, donde, X es la matriz original, r y c son vectores, fila y columnas de las masas, y D_r con D_c matrices diagonales cuyas entradas son los elementos de estas masas respectivamente. La matriz de inercia total (ecuación 12) se obtiene al elevar al cuadrado las ecuaciones anteriores, y que conjuntamente con el método de SVD ($X=U\Sigma V^T$) obtenemos la conexión entre el método SVD y la descomposición de valores propios dados por las ecuaciones 13 y 14 donde las matrices U , V , y Δ son las mismas matrices descritas en el método ACP.

$$\sum_i \sum_j s_{ij}^2 = tr(SS^T) \quad (12)$$

$$S^T S = V\Sigma^2 V^T = V\Delta V^T \quad (13)$$

$$SS^T = U\Sigma^2 U^T = U\Delta U^T \quad (14)$$

Con lo descrito ya es posible construir los mapas de análisis de correspondencia que es la principal salida de esta técnica, es así que, en la ecuación 15 se muestra las coordenadas principales de las filas, en 16 las coordenadas estándar de las filas, y en la 17 y 18 de las columnas respectivamente [22].

$$F = D_r^{-1/2} U \Delta \quad (15)$$

$$A = D_r^{-1/2} U \quad (16)$$

$$G = D_c^{-1/2} V \Delta \quad (17)$$

$$B = D_c^{-1/2} V \quad (18)$$

Regresión logística múltiple

Técnica multivariante clásica que modela el logaritmo natural de una razón de probabilidades, y es ideal para describir el comportamiento de una variable categórica con una o varias variables de tipo cuantitativas o continuas. Para describir el método, basta con fijarse en una regresión lineal con un solo predictor de tipo continuo, y una respuesta con una variable categórica dicotómica (dos niveles), en la cual, la curva de regresión estaría dividida en tres secciones, la mitad con una tendencia lineal, y los extremos asemejándose a curva intentando unirse a cada una de las rectas paralelas generada por la variable categórica [24]. Para solucionar este problema de tendencia, la regresión logística incorpora una función cuya gráfica se asemeja a una S denominada sigmoide (función logística), cuya forma matemática es la mostrada en la ecuación 19.

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (19)$$

En términos sencillos, la regresión logística calcula el logaritmo natural de la probabilidad de que la variable categórica Y adquiera un valor k a partir de que la variable independiente (predictor) X tenga un valor específico x , $p(Y=k|X=x)$ [25], matemáticamente lo expuesto se lo expresa en la ecuación 20.

$$\text{logit } Y = \ln \text{ odds} = \ln \frac{p(Y = k|X = x)}{1 - p(Y = k|X = x)} = \beta_0 + \beta_1 X \quad (20)$$

El modelo logístico simple indicado en la ecuación 20, nace a partir de sustituir los términos de una regresión lineal simple en la función sigmoide, para posteriormente despejar $\beta_0 + \beta_1 X$ mediante el empleo de logaritmos naturales.

$$p(Y = k|X = x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X)}} = \frac{e^{(\beta_0 + \beta_1 X)}}{1 + e^{(\beta_0 + \beta_1 X)}} \quad (21)$$

La razón entre la probabilidad de que un evento sea verdadero y la probabilidad de que sea falso toma el nombre de odds, dichos odds están presentes en la función de regresión logística y generan valores entre 0 y 1, y dado que, estos rangos de probabilidades son muy pequeños se lo amplía a un intervalo de $[-\infty, \infty]$ mediante el uso del logaritmo natural, añadiendo finalmente a la expresión logística el término $\ln \text{ odds}$. Según [25] una regresión logística debe cumplir con los supuestos de independencia de las variables, linealidad entre la variable respuesta y los predictores, y recomienda ser aplicada en análisis que poseen de 50 a 100 observaciones, finalmente, estableciendo un valor umbral, es posible determinar si una nueva observación pertenece a una u otra zona de probabilidades, con lo cual la regresión logística se convierte en un clasificador.

Redes neuronales

Hoy en día Artificial neural network (ANN) o red neuronal artificial es una de las técnicas de inteligencia artificial más utilizada en el diseño de clasificadores, esta técnica cuyo modelo matemático se asemeja una red neuronal biológica, está compuesta por capas o nodos interconectados, las mismas que están estructuradas por unidades que procesan información de la misma forma que las neuronas del cerebro.

En la configuración estructural básica de una red neuronal, los datos a analizar se cargan en forma de un vector multidimensional en la capa de entrada, dicha capa es la encargada de distribuir a las capas ocultas que toman su nombre dado que, normalmente no se visualiza los resultados de sus procesos, finalmente, la arquitectura cuenta con una capa de salida. Cada una de las neuronas que forman la red neuronal cumplen la función de un clasificador binario, y reciben la información proveniente de la capa anterior, misma que se encuentra multiplicada por un peso, y afectada por una función de activación que elimina la tendencia lineal propia de cada neurona. Además, cada neurona cuenta con una entrada adicional llamado sesgo, que es un término independiente que permite ser variada y por consiguiente realizar ajustes a la red neuronal.

Las cuatro funciones de activación más conocidas se muestran en el anexo A, de las cuales la función paso (escalonada) ha perdido posicionamiento, debido a que, cuenta únicamente con segmentos rectos que impiden el descenso de gradiente, dado que sus derivadas parciales son cero. La función sigmoide oscila en el intervalo de 0 a 1, mientras que la tangente hiperbólica al tener un rango de -1 a 1 tiene de ventaja generar una función de salida más o menos centrada en el origen al inicio del entrenamiento con lo cual facilita la convergencia del método, finalmente, la función ReLu presenta problemas en x igual a cero dado que en dicho punto la función no es suave por lo que no es diferenciable, pero evitando ese valor al no tener un valor máximo al final facilita el proceso del descenso de gradiente [26].

Dado que constantemente, las redes neuronales amplían su dominio como herramientas de clasificación y predicción en las diferentes ciencias, las investigaciones que buscan mejorar sus desempeños también aumentan, es así que, para el caso de clasificadores basados en redes neuronales se ha desarrollado una función de activación denominada softmax, que básicamente incorpora unos puntajes por clase y que cuya expresión descrita en la ecuación 22 es muy parecida a la expresión de predicción de una regresión lineal.

$$s_k(x) = x^T \theta^k \quad (22)$$

Una vez definidos los puntajes de softmax, la función estima la probabilidad \hat{p}_k de que una observación pertenezca a una u otra clase mediante la siguiente ecuación:

$$\hat{p}_k = \sigma(s(k))_k = \frac{\exp(s_k(x))}{\sum_{j=1}^K \exp(s_j(x))} \quad (23)$$

Llegado a este punto, es necesario hablar de la característica que ha posicionado a las redes neuronales como uno de los mejores métodos para el diseño de clasificadores, y es su capacidad de autoaprendizaje, dicha característica se basa en el algoritmo de backpropagation que le permite evaluar los errores y ajustar los pesos asociados a cada variable para conseguir el modelo más ajustado a la respuesta esperada [27]. Para expresar matemáticamente el algoritmo de retropropagación, se parte del modelo matemático de una red neuronal definido por las ecuaciones 24, 25 y 26.

$$C_i^L = (a_i^L - y_i^L)^2 \quad (24)$$

$$z_j^L = \sum_i (a_i^{L-1} w_{ij}^L + b_j^L) \quad (25)$$

$$a_i^L = \sigma(z_i^L) \quad (26)$$

En las ecuaciones anteriores, los subíndices i , denotan una neurona específica en una determinada capa L , mientras que a describe el comportamiento de dicha neurona producto de aplicar una función de activación σ a la información recibida de una capa anterior. La sumatoria de los efectos de las capas anteriores $a_i^{(L-1)}$ multiplicada por los pesos (w) respectivos y sumado el sesgo (b) toma la representación de z , además, dicho valor dependiendo de la función de activación varía típicamente entre 0 a 1 o entre -1 y 1. Finalmente, y_i^L es el valor esperado y C_i^L es la función de coste, que se constituye en el error de cada neurona de una capa donde está actuando el algoritmo de backpropagation [27].

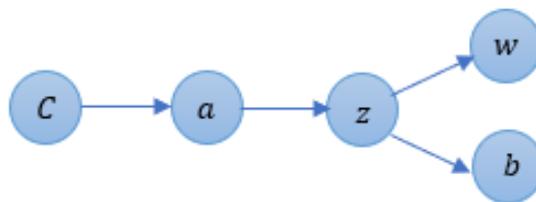


Fig. 1. Relación de variables que sirve como base para el planteamiento de la regla de la cadena.

La figura 1 muestra la manera en la que se encuentran relacionadas las variables unas con otras, y al ser una composición de funciones se debe recurrir a la regla de la cadena para determinar sus derivadas parciales. Es así que, si se desea obtener la razón de cambio de C con respecto a z, el primer paso consiste en definir la ruta (camino) que vaya desde C hasta z, en este ejemplo se parte de C, se pasa por a y finalmente se llega a z, además, recordar que para ir de una variable a otra se debe obtener su derivada parcial, es decir, para ir de C a "a" se debe calcular la derivada parcial de C con respecto a "a", mientras que para ir de "a" a z se determina la derivada de "a" con respecto a z, para finalmente multiplicar estas derivadas parciales y obtener la razón de cambio deseada [27]. Basado en lo expuesto, la razón de cambio de C con respecto a w queda definido como:

$$\frac{\partial C}{\partial w^L} = \frac{\partial C}{\partial a^L} * \frac{\partial a^L}{\partial z^L} * \frac{\partial z^L}{\partial w^L} \quad (27)$$

Además, las razones de cambio del costo con respecto al sesgo, de la sumatoria de los efectos con respecto a los pesos y al sesgo, quedan expresadas como:

$$\frac{\partial C}{\partial b^L} = \frac{\partial C}{\partial a^L} * \frac{\partial a^L}{\partial z^L} * \frac{\partial z^L}{\partial b^L} \quad \frac{\partial C}{\partial b^L} = \frac{\partial C}{\partial a^L} * \frac{\partial a^L}{\partial z^L} * \frac{\partial z^L}{\partial b^L} \quad (28)$$

$$\frac{\partial z^L}{\partial w^L} = a_i^{L-1} \quad (29)$$

$$\frac{\partial z^L}{\partial h} = 1 \quad I = \sum_j x_{i,j}^2 \quad I = \sum_j x_{i,j}^2 \quad (30)$$

Lo que se busca reducir es el error final, pero es obvio que dicho error se genera por la acumulación de errores presente en cada neurona, el mismo que es denotado por δ^L y matemáticamente está constituido por la derivada parcial del costo con respecto a la sumatoria de los efectos [28].

$$\delta^L = \frac{\partial C}{\partial a^L} * \frac{\partial a^L}{\partial z^L} \quad (31)$$

Reemplazando la ecuación 29, 30 y 31 en 27 y 28 tenemos:

$$\frac{\partial C}{\partial w^L} = \delta^L \quad (32)$$

$$\frac{\partial C}{\partial h^L} = \delta^L * a_i^{L-1} \quad (33)$$

Una vez cuantificados los errores de la capa L mediante las ecuaciones 32 y 33, es necesario extender el concepto a todas las capas anteriores que forman la arquitectura de la red neuronal, para esto, se define la matriz de parámetros W^L que determina como cambia la sumatoria de los efectos cuando se varía la función de activación de una neurona en una capa previa [27].

$$\delta^{L-1} = \delta^L W^L \frac{\partial z^L}{\partial a^{L-1}} \quad (34)$$

Generalizando las expresiones de las razones de cambio de la función de costo con respecto a los pesos y al sesgo y sustituyendo la ecuación 34, se obtiene el vector gradiente, mismo que forma parte de un ciclo repetitivo donde se lo resta del error, hasta reducir la diferencia entre la respuesta obtenida y los valores esperados de la base de datos de entrenamiento.

$$\nabla f^L = \left\langle \frac{\partial C}{\partial b^L}, \frac{\partial C}{\partial w_i^L} \right\rangle = \langle \delta^{L-1}, \delta^{L-1} a^{L-2} \rangle \quad (35)$$

Finalmente, el éxito del clasificador radica en los criterios que se empleen para el diseño de la arquitectura de la red neuronal, empezando por determinar la función de activación y el número de etapas de entrenamiento, seguido del número de capas ocultas y del número de neuronas en cada una (deep learning). Es común pensar que, mientras mayor sea el número de etapas de entrenamiento mejor será el resultado de la clasificación, esto no necesariamente es verdad, dado que, a mayor epochs mayor costo computacional aumentando tiempos de entrenamiento y en ocasiones generando sobre ajustes al modelo (overfitting), lastimosamente para este criterio no existe una regla definitiva, es así que, en la mayoría de investigaciones se determina este número con base en recomendaciones de investigaciones con objetivos similares, ensayo y error, o simplemente pericia de un experto. Un criterio que ha mostrado buenos resultados establece que, el número de épocas de entrenamiento debe ser igual al triple del número de variables de la base de datos de estudio, criterio que será empleado para la actual investigación.

Al contrario de lo anterior, el número de capas ocultas será determinado mediante la técnica de análisis de componentes principales, en el cual se buscará que el número de capas contenga un mínimo del 70% de la varianza explicada acumulada, además, el número de neuronas en cada capa será determinado mediante tres criterios. El primero consta en evaluar los errores de un proceso de iteración (ciclo for) que reemplaza en cada capa oculta un número de neuronas definido en un intervalo $[0.5n; 2n]$ donde n es el número de neuronas de la capa anterior; este proceso se repite en todas las capas ocultas; veamos un ejemplo sencillo que permite explicar este criterio, supongamos que disponemos de una base de datos de 6 variables, por lo que, el intervalo del ciclo for será, $[3, 12]$ se inicia el proceso de iteración que arroja una tabla que tabula los errores generados cuando la capa oculta tiene 3, 4, 5 y así sucesivamente hasta probar con 12 neuronas, con base en la tabla de error se define el número de neuronas de la primera capa oculta, es aquí que, se fija el número de neuronas en la primera capa, se cambia el contador a la segunda capa y se repite el proceso de iteración, este proceso se repite en todas las capas de la red neuronal. El segundo criterio, es una recomendación muy sencilla que indica que, cada capa oculta debe contener los dos tercios de la cantidad de neuronas de la capa anterior, y el tercer método, es el método de Xu and Chen, descrito en [29].

$$N_h = \frac{1}{2} \frac{N_t}{N_i \log N_t}, \frac{N_t}{N_i} > 30 \quad \text{o} \quad N_h = \frac{N_t}{N_i}, \frac{N_t}{N_i} \leq 30 \quad (36)$$

La ecuación 36 determina el número de neuronas en una capa oculta que depende del valor de la razón entre el número de etapas de entrenamiento N_t y el número de entradas (dimensión del vector de entrada) N_i .

Definida la red neuronal, para determinar su desempeño es primordial recurrir a métricas como el accuracy que es una medida de la cercanía de un valor experimental frente al valor real propuesto en una matriz, de modo que ayuda a determinar una estimación de la exactitud de las predicciones realizadas en el conjunto de datos con base en la diferencia entre la cantidad teórica y la cantidad definida analíticamente para obtener información sobre posibles errores en la ejecución [30], mientras el (MSE) es un indicador estadístico que muestra el error cuadrado promedio de las predicciones realizadas por la diferencia entre un valor predicho y un valor real con lo que proporciona la función de costo para la regresión lineal desarrollada [31].

$$\text{MSE} = \frac{1}{n} \cdot \sum_{i=1}^n (x_i - \hat{x}_i)^2 \quad (37)$$

La ecuación 37 establece el cálculo del error cuadrático medio donde n es el número de observaciones, x_i es el valor teórico y \hat{x}_i es el valor predicho por el modelo, según lo descrito en [31].

$$\text{accuracy} = \frac{\text{número de predicciones verdaderas}}{\text{número total de predicciones}} = \frac{V_p + V_n}{V_p + V_n + F_p + F_n} \quad (38)$$

En la ecuación 38 se muestra el cálculo de accuracy donde V_p son los valores positivos, V_n los valores negativos, F_p son falsos positivos y F_n son falsos negativos, estos son obtenidos relacionando los valores teóricos y predichos en una matriz, de tal forma que, se determinan los posibles errores que se pueden mostrar [32].

Un factor importante a considerar para el desempeño de la red neuronal es el overfitting, este se define como un comportamiento de aprendizaje automático que se presenta cuando se ha entrenado al modelo con demasiados datos de un solo tipo, por lo que proporciona predicciones precisas únicamente para los datos de entrenamiento, pero no para datos nuevos arrojando resultados equivocados [33]. Para evitar el overfitting se debe diversificar el conjunto de datos de entrenamiento y poner a prueba constantemente distintos datos de entrada.

Resultados

La base de datos en la que se sustentó la actual investigación estuvo constituida por 31 variables, de las cuales 5 son cuantitativas y 26 cualitativas, que recolecta datos antropométricos y datos históricos de un total de 122 individuos. Además, producto de un trabajo de titulación previo, se comprobó la validez del instrumento mediante un Alfa de Cronbach igual a 0,918. La presencia o no de desnutrición crónica fue determinada por médicos expertos y se encuentra representada en la columna desnutrición de la base de datos, dicha columna es el referente para comprobar las predicciones del actual estudio. Los análisis estadísticos clásicos carecen de robustez frente a datos con grandes componentes categóricos, es así que, la principal propuesta de clasificación del presente trabajo se basó en técnicas de deep learning y técnicas multivariantes, además, dichos resultados fueron contrastados con los obtenidos en una regresión logística que se constituye en una de las técnicas de mayor adaptación a elementos no lineales de la estadística convencional.

El lenguaje de programación estadístico R, con su vinculación a Python mediante la paquetería denominada Anaconda, son los programas que permitieron estudiar la base de datos, y que gracias a las extensiones TensorFlow y Keras facilitaron el desarrollo de procesos de deep learning. La primera etapa del estudio se constituyó en la coerción y corrección de variables, seguido de un análisis de datos atípicos y faltantes, que gracias a la carencia de los mismos se soslayó el proceso de imputación.

El anexo B permite visualizar las diferentes categorías de las variables cualitativas, y los valores máximos y mínimos de las variables cuantitativas de las tres primeras variables, además, mediante el comando missing se pudo evidenciar la carencia de datos faltantes [34]. Comprobado el correcto estado de la base de datos, se crea una nueva data excluyendo las cinco variables cuantitativas, para someterla a un análisis de correspondencia múltiple.

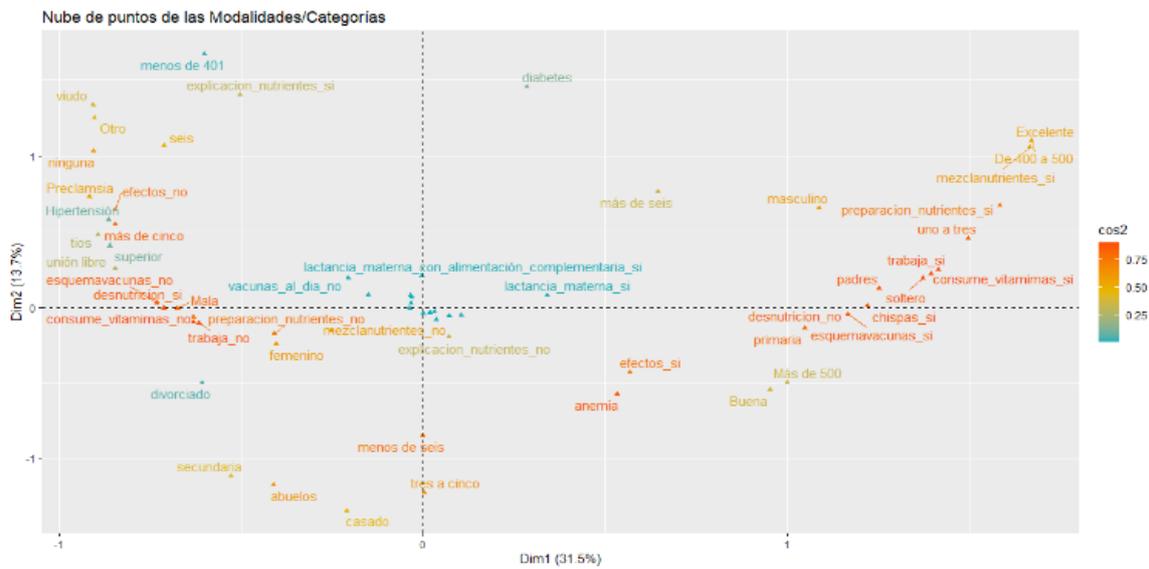


Fig. 3. Análisis de correspondencia múltiple incluyendo el valor de coseno cuadrado

El coseno cuadrado cos^2 mostrado en la figura 3 denota que tan confiable es la representación de las variables en un plano 2D, es así que, la variable explicación nutrientes al poseer un valor muy alejado a 1 imposibilita tomar decisiones sobre ella, por tal razón, las únicas variables que se eliminan del estudio serán las variables dicotómicas efectos, y mezcla nutrientes.

Realizada la disminución de dimensiones, a las variables se las convierte en dummy, con lo cual, la base de datos definitiva queda constituida por 64 variables, separadas en 63 que servirán como regresores (features), y una variable denominada (desnutrición) como única variable de respuesta (label). El primer estudio al cual se sometió la base de datos, es la red neuronal de una capa, seguido de una red neuronal con capas ocultas (Deep learning), dichos estudios inician con la etapa de normalización que permitió la unificación de escalas, evitando que los procesos de aprendizaje eliminen o reduzcan significancia de una o varias variables basadas en sus discrepancias métricas frente a variables que por su misma naturaleza manejen valores mucho mayores. Finalizada la etapa de normalización mediante la librería Keras con su comando `layer_normalization`, se dividió a la base de datos en dos partes, la primera correspondiente al 90% para el entrenamiento de la red neuronal y el 10% restante para la comprobación y validación de la arquitectura neuronal [36].

Clasificador 1 – Modelo sin capas ocultas: Con la base separada y normalizada, se elaboró el primer esquema del clasificador basado en una red neuronal simple, para lo cual, se entrenó a la red neuronal artificial con un total de 189 etapas de aprendizaje, en las cuales se empleó una función de activación softmax acompañada con una optimización Adam con un radio de aprendizaje de 0.1. El anexo D muestra los valores obtenidos a lo largo de las etapas de entrenamiento del loss, accuracy, y mse que evalúan el desempeño del clasificador al comparar sus resultados con el 10% de las observaciones que no fueron empleadas en el aprendizaje, finalmente, la tabla 1 representa las métricas finales con un loss de 0.023, mse de 0.005 y un accuracy de 1, que confirman un muy buen rendimiento de este clasificador.

Tabla 1. Valores de loss, accuracy, y mse para cuantificar el rendimiento del clasificador sin capas ocultas.

| loss | accuracy | mse |
|-----------|----------|-----------|
| 0.0234065 | 1 | 0.0050165 |

Clasificador 2 – Modelo por iteración: El segundo clasificador es la red neuronal con capas ocultas que se constituyen en una técnica de deep learning, y que para determinar la arquitectura de la red neuronal multicapa, se inició con la determinación del número de capas ocultas mediante un ACP (análisis de componentes principales) que posibilita determinar el número mínimo de capas que contengan mínimo el 75% de la varianza acumulada. En el anexo E, se establece que la arquitectura constará de una única capa oculta que contendrá el 96.3% de la varianza acumulada.

La siguiente etapa es establecer el número de neuronas en la capa oculta, para eso se utilizó tres técnicas, la primera consiste en una rutina for que reemplaza en la capa oculta todos los números de neuronas comprendidas en el intervalo de 31 (mitad del número de entradas de la capa anterior) a 126 que es el doble del número neuronas de la capa anterior, en cada iteración se guardan los valores de loss, accuracy y mse para posteriormente compararlos y concluir el número idóneo de neuronas.

Basado en la tabla del anexo F y el anexo G, la capa oculta queda definida con 51 neuronas en la capa oculta (anexo H (a)), dado que, con valores de loss de 0.06525, accuracy de 1 y mse de 0.01398 son las combinaciones con las mejores métricas en el rendimiento del clasificador. Además, el anexo G muestra el desempeño de la red neuronal con una capa oculta, en la cual se visualiza unos valores finales de 0.47 para loss, accuracy de 0.84 y un mse de 0.145, que a pesar de ser bueno, no muestran mejorías con respecto al modelo sin capas ocultas.

Clasificador 3 – Modelo con 2/3 de la capa anterior: El segundo criterio para determinar el número de neuronas en la capa oculta es la que menciona que la capa debe tener los 2/3 del número de variables de la capa anterior, para este caso la capa anterior es la de entrada con 63 variables, por lo que la capa oculta quedaría definida con 42 neuronas. El anexo H (b) resume el desempeño de la red neuronal con la regla de los 2/3 para el número de neuronas de la capa oculta, en la cual, los valores finales arrojan buenos resultados de clasificación con un valor de loss de 0.98, accuracy de 0.76 y un mse de 0.14 con drop out.

Clasificador 4 – Modelo con Xu and Chen: Finalmente, el criterio de Xu and Chen [29] que emplea la razón entre número de etapas de entrenamiento y número de variables de entrada, que para el presente trabajo es 183/53 y al ser menor a 30, se concluye que la capa oculta debe contar con 3 neuronas (anexo H (c)).

La última arquitectura de la red neuronal es la que mejor desempeño mostró de las redes neuronales con capas ocultas, con valores de 0.034 de loss, 1 de accuracy y mse de 0.007, además, todas las configuraciones emplearon la función drop out que evita que todas las neuronas se enciendan al mismo tiempo. El anexo I muestra la arquitectura del clasificador con una capa de entrada, una capa oculta con tres neuronas y una capa de salida.

Para complementar el estudio, se contrasta los resultados de los modelos con inteligencia artificial, con una regresión logística que representa a la estadística tradicional, es así que, se emplea el comando glm en la base de datos con las variables originales antes de convertirlas en dummy, y cuya variable de salida es de tipo categórica dicotómica (presenta o no desnutrición crónica).

El summary de la regresión logística, representada en el anexo J, muestra un valor de 5.62×10^{-10} con 59 grados de libertad de desviación residual, y un criterio de Akaike de 76, valores que permitieron seleccionar este modelo basado en una regresión logística con una distribución binomial. Definido el modelo, se verifica sus supuestos, el criterio de linealidad es evaluado mediante la librería car y su comando boxTidwell cuya hipótesis nula establece el cumplimiento de linealidad.

Tabla 2. Parámetros del test de Box - Tidwell para el análisis del supuesto de linealidad

| | MLE of lambda | Score Statistic (z) | Pr(> z) |
|-------------------|---------------|---------------------|----------|
| train\$edad_meses | -2.92003 | -0.3433 | 0.7314 |
| train\$peso | 6.95545 | -0.2800 | 0.7795 |
| train\$talla | -0.42608 | 0.2901 | 0.7718 |
| train\$edad | 6.21758 | -0.2402 | 0.8102 |

iterations = 12

Basado en la tabla 2 donde todos los estadísticos de prueba de las variables numéricas son mayores a la significancia y, además, la no existencia de patrones en los diagramas de dispersión del anexo K, bastan para aceptar la existencia de linealidad de las variables numéricas.

La independencia de los residuos es el segundo criterio que será evaluado, aunque se conoce que los datos fueron tomados en un lapso de tiempo considerable (un año) se pondrá a prueba este supuesto con la gráfica de los residuos como variable dependiente de los índices de las observaciones (anexo L). Al existir una relación lineal entre regresores se genera inestabilidad en los coeficientes de la regresión debido al aumento de varianza, por lo que el último supuesto a verificar es el de la multicolinealidad, para lo cual, se emplea las funciones corrplot y cor para visualizar la correlación existente entre las variables numéricas (Anexo M).

Por otro lado, el anexo N muestra la correlación entre el peso, edad, y talla, siendo el valor de 0.96 entre edad y peso la correlación más alta, seguida de un valor de 0.87 tanto entre la edad y talla, y la talla peso. Confirmada la existencia de correlación se hace necesario determinar el grado de relación entre las variables, para lo cual se empleará los factores infladores de varianza (VIF) mostrados en el anexo Ñ.

A diferencia de los análisis anteriores, los factores infladores de varianza analizan todas las variables tanto numéricas como categóricas, la tabla 9 muestra los resultados del comando vif de la librería car y presenta tres columnas de resultados, de las cuales concluiremos sobre la columna GVIF^{1/(2Df)} que determina los VIF con base en los grados de libertad, se considera que valores mayores a 5 generan valores de multicolinealidad, por lo que, el modelo presenta multicolinealidad en la variable comidas al día, y aunque los valores de edad en meses, talla y peso no sobrepasan el valor referencial, si se encuentran muy cerca del mismo. Cumplido el objetivo del diseño de los clasificadores, se reforzará la validación

mediante una matriz de confusión y curva roc, para esto, se selecciona al clasificador sin capas ocultas el cual fue que mostró mejor desempeño, y mediante los comandos confusionMatrix y roc de la librería pROC se obtiene del anexo O (32TP y 72TN) y anexo P.

La tabla 3 y anexo P confirman el buen desempeño del clasificador basado en redes neuronales sin capas ocultas, con 2 valores de falsos negativos (FN) y 3 de falsos positivos (FP), mientras que la curva ROC representa una tendencia favorable muy cercana a la ideal.

Con el clasificador definido, se realiza pruebas para reconocer el grado de importancia de cada variable en el clasificador, para lo cual se suprimió variable por variable y mediante los valores de rendimiento del clasificador se establece el grado de importancia de cada una. Mediante un ciclo for se automatizó el proceso de eliminación y se pudo evidenciar que prácticamente no había cambios significativos, excepto al eliminar la variable peso y talla, siendo esta última la que más variación generó. El anexo Q (a) muestra que, a diferencia del clasificador con todas las variables, ahora presenta 5 FP y 7 FN y un menor rendimiento con valores de loss de 0.793, accuracy de 0.846 y mse de 0.152, confirmando que la eliminación de la variable talla afecta en la clasificación. El segundo rendimiento se evaluó al eliminar las variables de peso y talla, resultados que se muestran en el anexo Q (b).

Aunque en la matriz de confusión varía muy pocos los errores, las métricas de loss, accuracy, y mse muestran un menor rendimiento del clasificador con dichas eliminaciones.

Tabla 3. Tabla resumen del rendimiento de cada modelo

| Modelo | > test_results | | |
|---|----------------|-----------|-----------|
| | loss | accuracy | mse |
| Sin capas ocultas | 0.0234065 | 1 | 0.0050165 |
| Capa oculta por iteracion | 0.4720982 | 0.8461539 | 0.1452236 |
| Capa oculta con 2/3 de la capa anterior | 0.9802729 | 0.7692308 | 0.2174405 |
| Capa oculta con Xu and Chen | 0.034536101 | 1.0000000 | 0.007563 |

Discusión

La fase de resultados permitió evaluar dos importantes parámetros, el primero se obtuvo al comparar el rendimiento de la regresión logística como técnica de la estadística tradicional con un clasificador basado en redes neuronales, para lo cual, se analizó 31 variables (5 cualitativas y 26 cuantitativas) que fueron reducidas a 29 gracias a un análisis de correspondencia múltiple que permitió la eliminación de los variables efectos y mezcla nutrientes, posteriormente las variables restantes fueron convertidas a variables dummy que finalmente constituyeron en un total de 62, que sirvieron para diseñar cuatro clasificadores, de los cuales, el clasificador con perceptrón simple (sin capas ocultas) fue el que mejor métricas mostró con valores de 0.02, 1 y 0.005 de loss accuracy y mse respectivamente, seguido del modelo con una capa oculta diseñado bajo el criterio de Xu and Chen [29] que estableció 3 neuronas en la capa oculta y métricas de 0.034, 1 y 0.0075, el tercer clasificador obtuvo una configuración con 51 neuronas en la capa oculta, número que fue obtenido mediante un proceso de iteración y generó valores de 0.065, 1 y 0.0139, el último clasificador

basado en redes neuronales constó con 42 neuronas siguiendo el criterio de que el número de neuronas debe ser los 2/3 del número de neuronas de la capa anterior y mostró el peor desempeño con 0.98, 0.76 y 0.14, cabe indicar que el número de capas ocultas en los tres últimos fue obtenido mediante un análisis de componentes principales que estableció que existía una varianza acumulada igual al 96.3% hasta la primera capa oculta, además, todas fueron sometidas a 189 etapas de entrenamiento. Con respecto a la regresión logística, mostró un correcto desempeño con uno 5.62×10^{-10} de desviación residual y un criterio de Akaike de 76, pero que, al momento de comprobar sus supuestos, incumple con el criterio de multicolinealidad, por lo que, disminuye su confiabilidad confirmando la superioridad de las técnicas de deep learning debido a su mejor adaptabilidad a variables no lineales, y a su disponibilidad de ser aplicarla a bases de datos sin importar su tamaño. El segundo parámetro se obtuvo al tener el clasificador definido, en el cual mediante una eliminación sucesiva de variables se pudo evidenciar que las variables: talla y peso, influyen en mayor proporción en las métricas del clasificador.

El actual estudio generó un clasificador de desnutrición crónica adaptado a todo tipo de niño menores de cinco años a diferencia del desarrollado en [6], que únicamente evalúa desnutrición en niños sometidos a cirugías por cardiopatía congénita, además, un hecho importante es que el presente estudio se encuentra adaptado a las características de Ecuador que lastimosamente, es uno de los países con mayor problemas de desnutrición de Latinoamérica, razón por la cual, el presente trabajo se posiciona por encima de los avances realizados en [7], [8] que adaptan sus condiciones a países de Bangladesh y Afganistán que distan de gran manera con las características de nuestro país. Lo desarrollado en [9] es muy parecido a lo logrado en el actual clasificador, pero la principal diferencia es que trabajan con una menor cantidad de variables y además logran cuantificar de mejor manera las relaciones y efectos de cada variable con la presencia de desnutrición. La investigación [14] que también empleada redes neuronales, pero de tipo convolucionales se encuentra aplica a la clasificación con base en imágenes de los niños, es decir hacen un estudio con base en visión artificial y que a diferencia del actual trabajo no analizan parámetros socioeconómicos ni historial médico, sino que simplemente evalúan los rasgos físicos de niños sanos y con desnutrición. Finalmente, el rendimiento del actual clasificador se encuentra muy cerca de lo alcanzado en [12] y supera a lo investigado en [10] que simplemente lograron establecer una interfaz sin ningún sustento o apoyo estadístico. Aunque se realizó un análisis de rendimiento del clasificador durante la eliminación de cada variable, sería interesante plantear un trabajo futuro en el cual se realice eliminaciones de grupos de variables para establecer el grupo de variables que más influyen en la clasificación de individuos con o sin desnutrición, además se podría experimentar para probar el comportamiento del clasificador para determinar todo tipo de desnutriciones y no simplemente la crónica. La variable que se esperaba que sea más determinante fue la lactancia materna, pero al momento de ser retirada no generó mayor cambio en las métricas del clasificador, sería conveniente ampliar el estudio en esa variable e incluso abordar la composición de la misma, ya que, ha mostrado ser un componente muy importante para eliminar y/o reducción la desnutrición.

Un limitante del actual trabajo fue haber partido de una base de datos de tamaño reducida y aislada a una única ubicación geográfica como lo es Esmeraldas, el siguiente paso, consistiría en ampliar la base de datos a más ciudades y regiones para evidenciar el comportamiento del clasificador, y de ser necesario realizar las correcciones que potencien su trabajo. El presente trabajo de investigación podría convertirse en una importante herramienta para la toma de decisiones de futuros planes gubernamentales que han dedicado recursos para la lucha contra la desnutrición infantil.

Conclusiones

El presente estudio propone un clasificador basado en redes neuronales para establecer un nuevo método de detección de desnutrición crónica a niños menores de cinco años. Fue desarrollado a partir de una base de datos levantada en el centro de salud Tipo C, en San Rafael - Esmeraldas (CSNSR) desde mayo del 2021 hasta abril del 2022, a 122 infantes que presentaron desnutrición. La fase de desarrollo inició con una depuración de la base de datos, seguida de un análisis de correspondencia múltiple que buscó reducir la cantidad de variables. Los clasificadores considerados fueron 4, de los cuales únicamente el primero se basa en una técnica de la estadística clásica como es la regresión logística, los otros tres usan métodos multivariantes, más específicamente redes neuronales. Para el diseño de los clasificadores con capas ocultas se empleó un ACP (análisis de componente principales) que evidenció la necesidad máxima de implementar una capa oculta que conserva el 96.3% de varianza acumulada. La arquitectura fue completada mediante la determinación del número de neuronas en la capa oculta, donde se emplearon tres criterios diferentes. Al comparar los clasificadores con capas ocultas y de perceptrón simple, el de perceptrón simple mostró mejor desempeño, confirmando la presencia de overfitting que, en lugar de generar mejoras, deterioran los resultados de clasificación, confirmando que no siempre una arquitectura más compleja es sinónimo de mayor rendimiento. El clasificador sin capas ocultas fue el que presentó mejor rendimiento con un loss de 0.023, accuracy de 1, y un mse de 0.05, además su matriz de confusión mostró únicamente cinco errores de predicción. Finalmente, con la eliminación consecutiva de variables, se concluyó que las más relevantes son las variables talla y peso.

En conclusión, la actual investigación ha aportado con un clasificador basado en técnicas multivariantes, mostrando una importante aplicación de estas técnicas actuales que complementan a los métodos estadísticos conocidos para resolver problemáticas de todo tipo.

Anexos

Todos los anexos se encuentran debidamente almacenado en Zenodo en la siguiente dirección <https://zenodo.org/records/10957048>

Referencias

- [1]. UNICEF, "Informe sobre nutrición infantil 2021 | Resumen ¿UNA ALIMENTACIÓN PARA," 2022. Accessed: Jul. 29, 2022. [Online]. Available: <https://www.unicef.org/media/107236/file/%20Fed%20to%20Fail%20-%20BRIEF-SPANISH-Final.pdf>
- [2]. INEC, "Encuesta Nacional de Salud y Nutrición ENSANUT," 2018. Accessed: Jul. 29, 2022. [Online]. Available: https://www.ecuadorencifras.gob.ec/documentos/web-inec/Estadisticas_Sociales/ENSANUT/ENSANUT_2018/Principales%20resultados%20ENSANUT_2018.pdf
- [3]. J. Rivera, U. Andina, S. Bolívar, and S. Olarte, "LA MALNUTRICIÓN INFANTIL EN ESMERALDAS: ¿UN ÉXITO RELATIVO?," 2022. Accessed: Jul. 29, 2022. [Online]. Available: <http://revistasoj.s.utn.edu.ec/index.php/lauinvestiga/article/view/457/350>
- [4]. UNICEF Ecuador, "Desnutrición en Ecuador." Accessed: Jul. 29, 2022. [Online]. Available: <https://www.unicef.org/ecuador/desnutrici%C3%B3n-cr%C3%B3nica-infantil>
- [5]. A. Fernández, R. Martínez, I. Carrasco, and A. Palma, "Impacto social y económico de la malnutrición," 2017. Accessed: Jul. 29, 2022. [Online]. Available: https://www.cepal.org/sites/default/files/publication/files/41247/LCTS2017-32_es.pdf
- [6]. H. Shi et al., "Explainable machine learning model for predicting the occurrence of postoperative malnutrition in children with congenital heart disease," *Clinical Nutrition*, vol. 41, no. 1, pp. 202–210, Jan. 2022, doi: 10.1016/j.clnu.2021.11.006.
- [7]. M. M. Islam et al., "Application of machine learning based algorithm for prediction of malnutrition among women in Bangladesh," *International Journal of Cognitive Computing in Engineering*, vol. 3, pp. 46–57, Jun. 2022, doi: 10.1016/j.ijcce.2022.02.002.
- [8]. P. M. P. K. J. H. C. Ziaullah Momand, *Data Mining Prediction of Malnutrition in Afghan Children*. IEEE, 2020. Accessed: Mar. 18, 2023. [Online]. Available: <https://ieeexplore.ieee.org/document/9059388>
- [9]. H. Darsene, A. Geleto, A. Gebeyehu, and S. Meseret, "Magnitude and predictors of undernutrition among children aged six to fifty nine months in Ethiopia: A cross sectional study," *Archives of Public Health*, vol. 75, no. 1, Jul. 2017, doi: 10.1186/s13690-017-0198-4.
- [10]. D. Kartika, R. Liana Gema, and M. Pratiwi, "Expert Systems for Identifying Children's Severe Malnutrition," 2016.
- [11]. R. N. Apriyono, A. Triayudi, and I. D. Sholihati, "Web-Based Expert System Detects Malnutrition in Toddlers with the Naïve Bayes Method," 2020. [Online]. Available: <https://iocscience.org/ejournal/index.php/mantik>
- [12]. S. Najafloo and M. Rabiei, "Recommended System for Controlling Malnutrition in Iranian Children 6 to 12 Years Old using Machine Learning Algorithms," 2021.

- [13]. S. M. J. Rahman et al., "Investigate the risk factors of stunting, wasting, and underweight among under-five Bangladeshi children and its prediction based on machine learning approach," *PLoS One*, vol. 16, no. 6 June 2021, Jun. 2021, doi: 10.1371/journal.pone.0253172.
- [14]. A. R. Lakshminarayanan, B. Pavani, V. Rajeswari, S. Parthasarathy, A. A. Azeez Khan, and K. Javubar Sathick, "Malnutrition Detection using Convolutional Neural Network," in *Proceedings of 2021 IEEE 7th International Conference on Bio Signals, Images and Instrumentation, ICBSII 2021*, Institute of Electrical and Electronics Engineers Inc., Mar. 2021. doi: 10.1109/ICBSII51839.2021.9445188.
- [15]. Meneses Pozo Daniela and Narváez Delgado Angie, "Estrategias de salud e incidencia en desnutrición crónica infantil en menores de cinco años que acuden del Centro de Salud Nuevo San Rafael- Esmeraldas," UNIVERSIDAD POLITÉCNICA ESTATAL DEL CARCHI. FACULTAD DE INDUSTRIAS AGROPECUARIAS Y CIENCIAS AMBIENTALES , 2022.
- [16]. L. Sandoval-Pillajo, M. PUSDÁ, F. Garrido, and E. P. Herrera-Granda, "Embedded system for mobility of blind people," *Revista Ibérica de Sistemas e Tecnologías de Informação*, vol. 2019, no. E19, pp. 328–340, 2019, [Online]. Available: <https://www.proquest.com/openview/841aa93ba3c3df454b7fc833a09a1823/1?pq-origsite=gscholar&cbl=1006393>
- [17]. M. María, T. E. Portillo, J. A. Salas, and P. Mendoza, "MAHALANOBIS Y LAS APLICACIONES DE SU DISTANCIA ESTADÍSTICA."
- [18]. E. P. Herrera-Granda, J. C. Torres-Cantero, A. Rosales, and D. H. Peluffo-Ordóñez, "A Comparison of Monocular Visual SLAM and Visual Odometry Methods Applied to 3D Reconstruction," *Applied Sciences*, vol. 13, no. 15, 2023, doi: 10.3390/app13158837.
- [19]. H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 4. pp. 433–459, Jul. 2010. doi: 10.1002/wics.101.
- [20]. V. M. Gonzalez Rojas, G. Conde Arango, and A. F. Ochoa Muñoz, "Análisis de Componentes Principales en presencia de datos faltantes: el principio de datos disponibles," *Scientia et Technica*, vol. 26, no. 2, pp. 219–228, Jun. 2021, doi: 10.22517/23447214.20591.
- [21]. J. AmatRodrigo, "Análisis de Componentes Principales (Principal Component Analysis, PCA) y t-SNE," 2017. [Online]. Available: <https://github.com/JoaquinAmatRodrigo/Estadistica-con-R>
- [22]. J. B. Greenacre Michael, "Multiple Correspondence Analysis and Related Methods Statistics in the Social and Behavioral Sciences Series," Londres, 2006.
- [23]. J. Enrique and P. Olivares, "Modelo de análisis de correspondencias múltiples," 1996.
- [24]. C. Y. J. Peng, K. L. Lee, and G. M. Ingersoll, "An introduction to logistic regression analysis and reporting," *Journal of Educational Research*, vol. 96, no. 1, pp. 3–14, 2002, doi: 10.1080/00220670209598786.

- [25]. J. Amat Rodrigo, "Regresión logística simple y múltiple," 2016.
- [26]. A. Géron, "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow SECOND EDITION Concepts, Tools, and Techniques to Build Intelligent Systems," 2019. [Online]. Available: <http://oreilly.com>
- [27]. F. Bueno, "Redes Neuronales: Entrenamiento y Comportamiento," 2019.
- [28]. A. E. Jácome Ortega, J. A. Caraguay Procel, E. P. Herrera-Granda, and I. D. Herrera Granda, "Confirmatory Factorial Analysis Applied on Teacher Evaluation Processes in Higher Education Institutions of Ecuador," in *Advances in Intelligent Systems and Computing*, Springer, 2020, pp. 157–170. doi: 10.1007/978-3-030-37221-7_14.
- [29]. K. Yotov, E. Hadzhikolev, and S. Hadzhikoleva, "Determining the Number of Neurons in Artificial Neural Networks for Approximation, Trained with Algorithms Using the Jacobi Matrix," *TEM Journal*, vol. 9, no. 4, pp. 1320–1329, Nov. 2020, doi: 10.18421/TEM94-02.
- [30]. J. M. Betz, P. N. Brown, and M. C. Roman, "Accuracy, precision, and reliability of chemical measurements in natural products research," *Fitoterapia*, vol. 82, no. 1. pp. 44–52, Jan. 2011. doi: 10.1016/j.fitote.2010.09.011.
- [31]. Machaca Apaza Lianne, "Estimación de la evapotranspiración de referencia," 2016. [Online]. Available: <http://repositorio.unap.edu.pe/handle/20.500.14082/4612>
- [32]. A. Zapeta Hernández, G. A. Galindo Rosales, H. J. Juan Santiago, and M. Martínez Lee, "Métricas de rendimiento para evaluar el aprendizaje automático en la clasificación de imágenes petroleras utilizando redes neuronales convolucionales," *Ciencia Latina Revista Científica Multidisciplinar*, vol. 6, no. 5, pp. 4624–4637, Nov. 2022, doi: 10.37811/cl_rcm.v6i5.3420.
- [33]. Poloni Pablo, "Nuevo Algoritmo de clasificación supervisado sin parámetros, no afectado por el desbalanceo y overfitting," 2014. [Online]. Available: https://bibliotecadigital.exactas.uba.ar/download/tesis/tesis_n5839_Poloni.pdf
- [34]. E. P. Herrera-Granda, I. D. Herrera-Granda, L. L. Lorente-Leyva, P. D. Granda-Gudiño, and J. A. Caraguay-Procel, "Implementation of an Artificial Vision System and Monitoring of Human Objectives using a Quadricopter," *Revista Ibérica de Sistemas e Tecnologías de Informação*, vol. 2019, no. E19, pp. 198–211, 2019, [Online]. Available: <https://www.proquest.com/openview/ca355d43bc1ac0d236ca8da769393332/1?pq-origsite=gscholar&cbl=1006393>
- [35]. I. D. Herrera-Granda, W. G. Imbaquingo-Usiña, L. L. Lorente-Leyva, E. P. Herrera-Granda, D. H. Peluffo-Ordóñez, and D. G. Rossit, "Optimization of the Network of Urban Solid Waste Containers: A Case Study BT - Technology Trends," M. Botto-Tobar, G. Pizarro, M. Zúñiga-Prieto, M. D'Armas, and M. Zúñiga Sánchez, Eds., Cham: Springer International Publishing, 2019, pp. 578–589.

- [36]. I. D. Herrera-Granda et al., "Artificial Neural Networks for Bottled Water Demand Forecasting: A Small Business Case Study," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 11507 LNCS, Springer Verlag, 2019, pp. 362–373. doi: 10.1007/978-3-030-20518-8_31.