

Diagnóstico del dengue utilizando redes neuronales artificiales

Buenaventura Ríos Ríos

bventura63@hotmail.com

<https://orcid.org/0000-0001-9777-1757>

Tarapoto – Perú

RESUMEN

El dengue es una enfermedad viral producida por el mosquito *Aedes Aegypti*, de gran impacto en las últimas décadas en América Latina y el Perú, particularmente en la Amazonía que representa el 57% del territorio y el 12% de la población nacional. El objetivo del presente estudio es diagnosticar el dengue utilizando como estrategia inteligente a las redes neuronales artificiales (RNA) *backpropagation*, a partir de los síntomas de la enfermedad, teniendo como datos de entrada 430 patrones con 24 síntomas cada uno, proporcionados por la Dirección Regional de Salud (DIRESA) de la Región San Martín en el Perú. Se trata de una investigación aplicada, desarrollada por objetivos que, a través de iteraciones recurrentes se va determinando la arquitectura de red neuronal más adecuada para el diagnóstico del dengue sin alarma, con alarma y grave. La arquitectura de Red Neuronal Artificial - Perceptrón Multicapa (RNA - MLP) *cascafeedforward* de dos capas ocultas: 24 - 12 - 4 - 1, es la que mejores resultados ofrece con un error cuadrático medio equivalente a 0.0457, factor de regresión $R=0.90715$ y efectividad de 95.35%.

Palabras clave: dengue; red neuronal artificial; retropropagación.

Correspondencia: bventura63@hotmail.com

Artículo recibido 15 de julio del 2022. Aceptado para publicación: 20 de agosto del 2022.

Conflictos de Interés: Ninguna que declarar

Todo el contenido de **Ciencia Latina Revista Científica Multidisciplinar**, publicados en este sitio están disponibles bajo

Licencia [Creative Commons](https://creativecommons.org/licenses/by/4.0/) 

Como citar: Ríos Ríos, B. (2022). Diagnóstico del dengue utilizando redes neuronales artificiales. *Ciencia Latina Revista Científica Multidisciplinar*, 6(4), 5636- 5651. https://doi.org/10.37811/cl_rcm.v6i4.3037

Diagnosing dengue using artificial neural networks

ABSTRACT

Dengue is a viral disease caused by the *Aedes Aegypti* mosquito, which has had a great impact recent decades in Latin America and Peru, particularly in the Amazon region, and represents 57% of the territory and 12% of the national population. The objective of this study is to diagnose dengue using artificial neural networks (ANN) backpropagation as an intelligent strategy, based on the symptoms of the disease, having 430 patterns with 24 symptoms as input data, provided by the Regional Directorate of Health (DIRESA) of the San Martin Region in Peru. This is applied research, developed by objectives that, through recurring iterations, determines the most appropriate neural network architecture for the diagnosis of dengue without alarm, with alarm and severe. The Artificial Neural Network architecture - Multilayer Perceptron (ANN - MLP) cascade feed forward with two hidden layers: 24 - 12 - 4 - 1, is the one that offers the best results with a mean square error equivalent to 0.0457, regression factor $R=0.90715$ and effectiveness of 95.35%.

Keywords: dengue; artificial neural network; backpropagation.

INTRODUCCION

En la región San Martín, enclavada en la Amazonía del Perú, el dengue es una enfermedad viral recurrente, que se presenta todos los años, particularmente en épocas de lluvias (febrero, marzo y abril) que favorece al crecimiento poblacional del mosquito *Aedes Aegypti*, vector para la transmisión de la enfermedad, con consecuencias sociales graves, algunas veces mortales. De allí que, el diagnóstico del dengue, revierte particular importancia para el tratamiento de la enfermedad en forma oportuna.

Según datos de la Organización Panamericana de la Salud (OPS), el año 2015 se presentaron en la América Latina 2 439 527 casos de dengue, 12 495 con dengue hemorrágico (grave) y 1 365 fallecidos (OPS, 2016). En referencia al Perú, según datos del Ministerio de Salud (MINSA, 2020), el año 2017 se presentó un valor pico con 68 290 casos de dengue y 89 defunciones, asociados al fenómeno del niño; según datos recientes de la misma entidad en el año 2018 se presentaron 4 698, el 2019 ascendió a 15 290, el 2020 siguió incrementándose a 47 933, y, el 2021 a 48 029, en ese periodo también se registraron 187 fallecidos (MINSA, 2022).

En lo que respecta a la región San Martín, ubicada en el noreste del Perú, entre los años 2013 al 2021 se registraron 16 678 casos, observando que se ha producido un incremento significativo el año 2020 con 4 955 casos y el 2021 con 5 859 infectados, coincidiendo en el último caso con la impronta de la pandemia del COVID 19, ver Tabla 1.

Tabla 1 Registro de dengue entre los años 2013 al 2021.

Casos registrados de dengue por año en la región San Martín									
Año	2013	2014	2015	2016	2017	2018	2019	2020	2021
N° Casos	1 208	1 574	220	335	460	98	1 969	4 955	5 859

Fuente: MINSA.

El problema se complica si se toma en cuenta que, el *Aedes Aegypti*, también es el vector de otro tipo de enfermedades como la Chicungunya, Zika y Fiebre Amarilla de la Selva, que presentan similares síntomas, no siendo fácil distinguir las unas de las otras, por lo que se requiere contar con laboratorios, no siempre disponibles en las zonas rurales del Perú. Es preciso señalar que la región San Martín cuenta 10 provincias, 78 distritos y aproximadamente 3 000 centros poblados, en la que prestan sus servicios más de 400 centros de salud.

En las últimas décadas, el tema del diagnóstico del dengue ha sido estudiado utilizando estrategias de inteligencia artificial, como las redes neuronales artificiales (RNA), lógica difusa y algoritmos genéticos, sin embargo, mayormente han sido abordados considerando variables ambientales como la temperatura, clima, precipitación pluvial, así como, información genética, imágenes satelitales, entre otros, pero no han sido abordados directamente tomando en cuenta los síntomas de la enfermedad.

Para el presente estudio, se optó por el uso de las RNA como estrategia inteligente para diagnosticar el dengue a partir de los síntomas de la enfermedad, aspecto que no ha sido suficientemente estudiado. Como se sabe, las RNA son sistemas de procesamiento de información cuya estructura y funcionamiento emulan a las redes neuronales biológicas contenidas en el cerebro humano, los mismos que han demostrado tener enormes potencialidades para extraer información a partir de datos complejos e imprecisos, capacidad de aprendizaje y de procesar información en paralelo. De acuerdo con Montaña (2002), las RNA consisten en:

Un gran número de elementos simples de procesamiento llamados nodos o neuronas que están organizados en capas. Cada neurona está conectada con otras neuronas mediante enlaces de comunicación, cada uno de los cuales tiene asociado un peso. Los pesos representan la información que será usada por la red neuronal para resolver un problema determinado. (p.120).

Para el cálculo de los pesos de cada enlace o conexión, las RNA requieren de entrenamiento, el mismo que se realiza a través de un algoritmo de entrenamiento, como el algoritmo *backpropagation*, el más conocido y versátil, que se basa en el cálculo del gradiente negativo o gradiente descendente del error cuadrático medio en la salida de la red neuronal.

Por el entrenamiento, que es un proceso de aprendizaje, el algoritmo modifica el valor de los pesos de las conexiones cada vez que se entrena la red, por eso la clave está en entrenar la red neuronal cuantas veces sea necesaria, hasta encontrar el menor error de entrenamiento y sobre todo de validación, de donde se obtienen los pesos definitivos. Al respecto, Caicedo y López (2017) refieren que “el aprendizaje para la red neuronal artificial, es la habilidad para aprender del entorno y mejorar su desempeño, es un proceso interactivo que permite ajustar los pesos sinápticos” (p.28).

Por lo expuesto, el presente estudio se trata de una *investigación aplicada*, que en términos de Padrón (2006) están orientados a “resolver problemas de la vida cotidiana o a controlar situaciones prácticas”, en ese campo se inscriben la innovación técnica, artesanal o industrial, así como, el uso de teorías científicas previamente validadas para la solución de problemas prácticos. Desde esa perspectiva este estudio constituye una investigación por objetivos, que busca dar respuesta a un problema práctico haciendo uso de una herramienta informática validada como las redes neuronales artificiales.

En suma, el objetivo principal de la investigación ha sido implementar un prototipo de diagnóstico del dengue utilizando redes neuronales artificiales - multilayer propagation (RNA-MLP), es decir, multicapa, en cuya arquitectura la información fluye de la entrada a la salida pasando por capas intermedias (*feedforward*), utilizando como algoritmo de entrenamiento la opción de retropropagación o propagación hacia atrás (*backpropagation*), en la que el error se propaga de la capa de salida hacia las capas ocultas, haciendo posible el cálculo del menor error global y, por ende, la red neuronal de mejor performance. Una opción del MATLAB es la red tipo *cascafeedforward*, que es una variante del *feedforward*, que consiste en ingresar todos los datos de entrada a cada una de las capas ocultas incluida la capa de salida, así como, de cada capa anterior a una capa posterior; esta opción también ha sido probada.

MARCO TEORICO

Las redes neuronales artificiales forman parte de los sistemas inteligentes que, en términos de Paletta (2007), son todos aquellos programas de computación que muestran algún tipo de inteligencia, cualidad a la que se asocia el conocimiento y la capacidad de decidir, en ese sentido alude que todo sistema inteligente debe ser capaz de aprender, ser autónomo, razonar y desarrollar su autoconciencia, en suma, debe ser capaz de percibir, razonar y actuar.

El mismo Paletta (2010) define a la inteligencia artificial como “la ciencia de emular procesos propios de la inteligencia humana tales como la percepción, el razonamiento y el aprendizaje, haciendo uso de medios no naturales” (p.15). Desde esa perspectiva, las RNA, así como, la lógica difusa y los algoritmos genéticos, forman parte de lo que se conoce como inteligencia artificial.

Hilera y Martínez (1995) definen a las RNA como sistemas conexionistas de procesamiento de información cuya estructura y funcionamiento están inspirados en las redes neuronales biológicas, como el que se observa en la Figura 1.

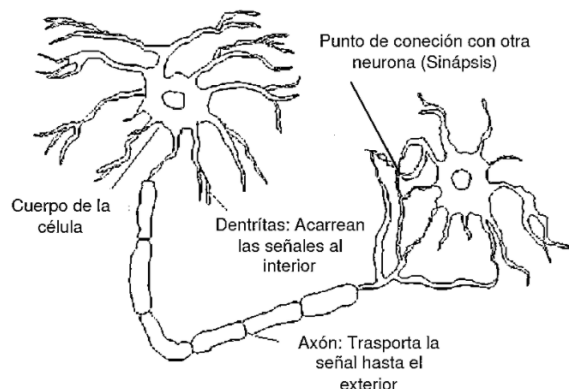


Figura 1. Esquema de una neurona biológica. Copyright © 2013: Rojas, J.

En las neuronas biológicas, las dendritas reciben las señales de entrada excitatorias o inhibitorias de las neuronas vecinas, el cuerpo celular las combina, integra y emite señales de salida que son transportadas por el axón, a través de procesos llamados sinapsis, distribuyendo la información a otras neuronas.

A partir de lo anterior, podemos definir a las RNA como:

Un gran número de elementos simples de procesamiento llamados nodos o neuronas que están organizados en capas. Cada neurona está conectada con otras neuronas mediante enlaces de comunicación, cada uno de los cuales tiene asociado un peso. Los pesos representan la información que será usada por la red neuronal para resolver un problema determinado. (Montaño, 2002, p.120).

Esos elementos simples de procesamiento de información a los que se refiere Montaño, constituyen las *neuronas artificiales* o nodos, a partir del cual se construyen las RNA, cuyo esquema básico se muestra en la Figura 2. En ella, los nodos representan elementos computacionales simples que emulan la respuesta de una neurona a un determinado estímulo, cuando la suma de señales de entrada es suficientemente alta, superando un determinado umbral, la neurona manda una señal a las neuronas con las que mantiene contacto. El proceso desarrollado por el nodo requiere de una función de activación como la sigmoidea, tangente hiperbólica o lineal (purelin).

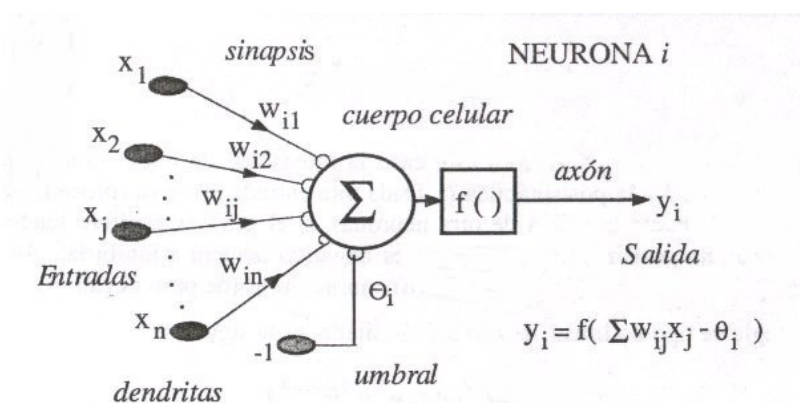


Figura 2. Esquema de una neurona artificial. Copyright © 2005: Palacios, F.

La RNA más simple es el perceptrón, presentado en 1958 por el psicólogo Frank Rosenblat, conocido también como red monocapa, porque solo tiene una capa de procesamiento de información. De acuerdo con Caicedo y López (2017) el perceptrón “posee conectividad total, ya que la neurona de la capa de salida esta conectada a todas las unidades o neuronas de entrada... la función de activación que utiliza la neurona de salida es el tipo escalón binario o bipolar”, (p.38).

Sin embargo, el perceptrón tiene una serie de limitaciones, particularmente su incapacidad para separar regiones que no son linealmente separables, no lineales. Las redes neuronales multicapa MLP (Multilayer Perceptron), permiten superar esas limitaciones gracias al desarrollo del algoritmo *Backpropagation* (propagación hacia atrás), desarrollado por el profesor Paul Werbos en 1974, para el entrenamiento de la red multicapa, haciendo uso de la Regla Delta o *error cuadrático medio* (mse), que hace posible calcular el error global de toda la RNA.

Las redes feedforward constituyen un tipo especial de RNA, caracterizadas por la forma en que fluye la información, de la entrada hacia la salida pasando secuencialmente por las capas intermedias. Normalmente, las redes feedforward trabajan con funciones de activación o transferencia sigmoidea tangencial (tansig) en las capas ocultas y con funciones purelin (lineal) en la capa de salida; la función de entrenamiento por autotomiasia es el Levenberg Marquardt (trainlm). Existen diferentes versiones de las redes feedforward, como las redes de adaptación fitnet, de reconocimiento de patrones patternnet y la red en cascada hacia adelante cascadefeedforward.

El entrenamiento es el aprendizaje de la RNA que, según la Real Academia Española, es “la adquisición por la práctica de una conducta duradera” (2020). Cuando el entrenamiento es supervisado, el proceso de aprendizaje se realiza bajo el control de un

supervisor o maestro externo, quien monitorea el comportamiento del error cuadrático medio resultado de la comparación de la salida objetivo o deseada (target) respecto a la salida real del sistema (output), cuando se le ingresa un conjunto de entradas. En estos procesos los datos de entrada y la salida deseada son conocidos, el problema consiste en encontrar a través de iteraciones sucesivas los pesos de los nodos con el menor error global posible.

El algoritmo de entrenamiento Backpropagation, que se basa en el calculo del gradiente negativo o gradiente descendente del error cuadrático medio en la salida de la RNA, modifica el valor de los pesos de las conexiones entre las neuronas de entrada, ocultas y de salida, cada vez que entrena la red se modifican los pesos, la clave esta en entrenar la red neuronal cuantas veces sea necesaria, hasta encontrar el menor error de entrenamiento (e_p) y sobre todo de validación, de donde se obtienen los pesos definitivos.

$$e_p = \frac{1}{2} \sum_{j=1}^M (d_{pj} - y_{pj})^2$$

Donde:

d_{pj} : datos objetivo (target).

y_{pj} : datos de salida.

Las RNA, en general, han demostrado ser una valiosa estrategia inteligente para la predicción del dengue, a partir de factores ambientales como la temperatura media, humedad relativa media y precipitación total, tal es el caso del estudio realizado por Aburas, Cetiner y Sari (2010) en donde concluyen que “se logró una alta correlación de 0.96 como se esperaba cuando se usa el mismo conjunto de datos (330 patrones) tanto para el entrenamiento como para las pruebas. Se decidió una capa oculta para ser óptima”, (p.4259).

En otro estudio, Ibrahim, Faisal y Taib (2010) utilizaron medidas de fisiología, manifestaciones clínicas y mediciones de BIA (Bioelectrical Impedance Analysis), para construir un modelo de pronóstico del dengue con una RNA MLP con algoritmo Levenberg Marquardt, obtienen una precisión de 74.2 % y 0.02 de error cuadrático

medio, para una arquitectura de red 35 – 26 -1. De los 35 datos de entrada 18 correspondieron a manifestaciones clínicas.

Los mismos autores Ibrahim, Faisal, Taib y Salim (2010), utilizaron una RNA con arquitectura 3 – 6 – 1, para predecir el dengue a partir de investigaciones en la sangre, tales como, el recuento de plaquetas, hematocritos y el nivel de aspartato aminotransferasa, con una precisión de predicción global de 96.27 %.

Un estudio interesante fue el realizado por Gambhira, Malika y Kumarb (2017), para el diagnóstico temprano de la enfermedad del dengue haciendo un movimiento preventivo antes de que se convierta en una enfermedad infecciosa; para ello utilizaron un modelo de diagnóstico al que denominaron PSO (Particle Swarm Optimization), ver Figura 3.

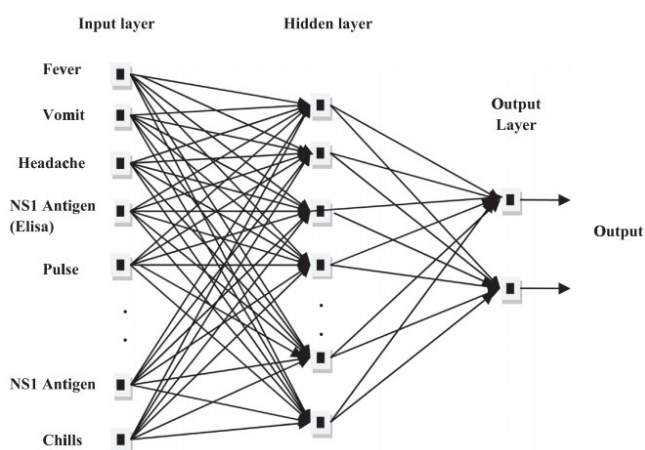


Figura 3. RNA de tres capas y atributos de entrada del dengue. Fuente: Ghambira, 2017.

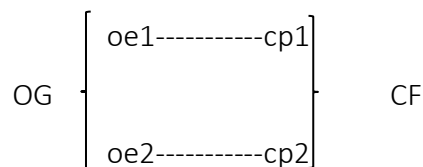
Se utilizaron datos clínicos como fiebre, vómito, dolor abdominal, escalofríos, dolor del cuerpo, cefalea y debilidad, también datos externos edad y sexo, además de inmunoglobulinas IgM, IgG, NS1 y Elisa. El sistema PSO-ANN demostró tener una alta precisión equivalente a 82.3 %.

METODOLOGIA

El presente estudio corresponde a una investigación aplicada ex post facto, descriptiva y transversal, como ya se dijo, orientada a resolver problemas de la vida cotidiana o a controlar situaciones prácticas, cuya estructura implica partir de una situación problemática (el diagnóstico del dengue a partir de los síntomas de la enfermedad), seleccionar una teoría de carácter reconocida y suficientemente aceptada como las redes neuronales artificiales, de la que se derivan una serie de acciones y de previsiones en el

marco de un prototipo y, finalmente, el ensayo y puesta a prueba del prototipo determinando la probabilidad del modelo aplicativo de resolver la problemática planteada, Padrón (2006).

El prototipo de la investigación es la que se muestra a continuación:



En la que el objetivo general (OG) es el diseño e implementación de un prototipo inteligente para el diagnóstico del dengue a partir de los síntomas de la enfermedad; el objetivo específico uno (oe1) es el análisis, tratamiento y adecuación de los datos de entrada; el objetivo específico dos (oe2) el desarrollo del prototipo; para arribar a conclusiones parciales (cp1 y cp2) y la conclusión final (CF).

Los datos fueron obtenidos de la Dirección Regional de Salud de San Martín (DIRESA) y comprende 807 casos, que corresponden a los años 2017 al 2019, de los que se obtuvo una muestra modificada de 215 patrones aplicando la fórmula de Wittkoski (2010) para una población finita:

$$n = \frac{N \cdot Z^2 \cdot p \cdot q}{e^2 \cdot (N - 1) + Z^2 \cdot p \cdot q}$$

Donde: n = Tamaño de la muestra; N = Población o universo; Z = Nivel de confianza (95%) = 1.96; p = Probabilidad a favor (50%) = 0.5; q = Probabilidad en contra = (1-p) = 0.5; e = error muestral (5%) = 0.05.

Los datos obtenidos, en formato Excel, corresponden a diversas enfermedades como el dengue, la chicungunya, zika y fiebre amarilla de la selva, presentando los diferentes síntomas que corresponde a cada enfermedad. La muestra ha sido seleccionada por patrones de 24 síntomas cada uno, en forma aleatoria, distinguiendo aquellas que corresponden al dengue de las otras enfermedades que no son dengue.

Para el entrenamiento de la RNA-MLP se siguió el siguiente procedimiento de ajuste: Primero, inicializar la red y entrenar de nuevo, en cada reinicio se asignan aleatoriamente nuevos pesos y vías a la conexiones neurales; segundo, aumentar el número de neuronas en la capa oculta, considerando el menor número de neuronas posible para no saturar la red; tercero, cambiar el algoritmo de entrenamiento de *trainlm* a *trainbr* (bayesiano) y, cuarto, usar datos de entrenamiento adicionales en caso no se de la convergencia.

El procesamiento de la información, sea para entrenamiento, validación y test de la RNA-MLP, se ha utilizado el MATLAB versión 2017b y sus diferentes librerías.

RESULTADOS Y DISCUSIÓN

Se realizó el análisis, tratamiento y adecuación de los datos proporcionados por la Dirección Regional de Salud de San Martín, distinguiendo los síntomas del dengue sin alarma (artralgias, artralgias de manos, artralgias de pie, mialgias, cefalea, dolor ocular, dolor lumbar, exantema, conjuntivitis no purulenta y nauseas), dengue con alarma (dolor abdominal intenso, dolor torácico, derrame seroso, vómito persistente e ictericia), dengue grave (pulso débil, extremidades frías, diferencial de presión arterial, sangrado grave, apertura ocular, respuesta motora y respuesta verbal), y, de manera transversal la diarrea. Esta diferenciación fue meramente didáctica, toda vez que el sistema de red neural utilizado no distingue a las unas de las otras, a todas las combinaciones posibles las reconoce por igual como verdaderos positivos según corresponda.

Los 807 casos registrados por la DIRESA y entregados en formato Excel, fueron adaptados y adecuados en forma matricial (24 x 215), también en formato Excel, de 24 filas (síntomas) por 215 columnas (patrones); como se muestra en la Tabla 2 para los primeros 36 patrones. Cada patrón es el resultado de una combinación particular de síntomas. La última fila corresponde al Target, es decir, a la salida objetivo, que muestra la presencia (1) o ausencia (0) del dengue.

Tabla 2

Modelo de matriz de 24x215, con los primeros 36 patrones de entrada.

CODIGO MINSA	6	12	25	39	44	51	59	67	71	78	83	86	90	93	111	122	130	136	138	144	149	150	165	182	190	193	200	206	214	220	228	240					
SINTOMAS/PATRONES	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21	P22	P23	P24	P25	P26	P27	P28	P29	P30	P31	P32	P33	P34	P35	P36	
1 Fiebre	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		
2 Artralgias	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		
3 Artralgias_ma	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
4 Artralgias_pi	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
5 Mialgias	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
6 Cefalea	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
7 Dolor_ocular	0	1	0	1	1	1	1	0	1	1	0	0	1	0	0	1	0	0	1	1	0	1	0	0	1	0	0	0	1	0	0	0	1	0	0	1	1
8 Dolor_lumbar	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
9 Rash_exantema	0	1	0	0	1	1	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	1	0	0	1	0	0	0	1	1	0	0	1	1	0	0
10 Conjuntivitis	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	
11 Nauseas_vomitos	0	0	1	1	0	1	0	1	0	0	0	0	0	0	0	0	1	0	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
12 Dolor_abdominal	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
13 Dolor_toracico	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
14 Derrame_seroso	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
15 Vómito_persistente	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
16 Ictericia	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
17 Pulso_débil	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
18 Extremidad	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19 Diferpresion	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20 Sangrado_grave	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21 Apertura_ocular	0	1	0	1	1	0	1	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
22 Respuesta_motora	0	1	0	1	1	1	0	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
23 Respuesta_verbal	0	1	0	1	1	1	0	1	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
24 Diarrea	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TARGET	1	1	1	0	1	1	0	1	0	1	1	1	0	0	1	1	1	0	1	0	1	0	0	0	0	0	1	1	1	1	0	1	1	0	1	1	

Esta adecuación fue necesaria, toda vez que la herramienta de MATLAB aplicada a redes neuronales artificiales, requiere que los datos de entrada y salida se organicen de esa manera para los procesos de entrenamiento, validación y test.

Luego se desarrolló el prototipo de la red neural a utilizar. Inicialmente, se optó por una arquitectura de red neuronal artificial multicapas – multilayer propagation (RNA-MLP) de 24 neuronas en la capa de entrada, 10 neuronas en la capa oculta y una neurona en la capa de salida para simular la presencia o ausencia del dengue. Por defecto, el comando `feedforwardnet` del MATLAB trabaja con funciones *sigmoidea tangente hiperbólica* (`tansig`) en las capas ocultas y la *función de transferencia lineal* (`purelin`) en la capa de salida, sin embargo, se obtienen mejores resultados cuando se aplica en la capa de salida funciones `tansig`, favoreciendo a la convergencia de la salida.

Para el entrenamiento de la RNA se trabajó con el algoritmo de entrenamiento *trainlm* (Levenberg – Marquardt), que actualiza los valores de peso y sesgo de cada conexión neuronal de acuerdo con los parámetros de Levenberg – Marquardt, para cuyo efecto se utilizó la secuencia de comandos que se muestra a continuación.

```
Input = xlsread('inputDengue.xlsx','C6:PP29');
Target = xlsread('inputDengue.xlsx','C31:PP31')
Red = forwardnet([10]);
red.trainFcn = 'trainlm';
[red tr] = train(red,input,target);
Output = sim(red,input);
```

Cuando se realizaron las pruebas con 215 patrones de entrada se observó que la red no tenía convergencia, por lo que se duplico el número de patrones de entrada a 430, con los que si se logró importantes niveles de convergencia. Se hicieron muchas iteraciones durante el entrenamiento, variando el número de neuronas de la capa oculta, partiendo de 10 neuronas (por convención), observando que los mejores niveles de convergencia se presentaron cuando se trabaja con 12 neuronas.

Con el propósito de mejorar la performance de la RNA-MLP backpropagation y alcanzar la convergencia, se decidió incrementar el número de capas ocultas a dos, observando que la arquitectura de red **24 – 12 – 4 – 1** (24 neuronas en la capa de entrada, 12 neuronas en la primera capa oculta, 4 neuronas en la segunda capa oculta y una neurona en la capa de salida), presenta una mejor convergencia con un *error cuadrático medio* (`mse`) equivalente a 0.0601 y un *factor de regresión* $R=0.92$.

Finalmente, se decidió hacer pruebas con la red tipo `cascafeedforward`, con la misma arquitectura de red **24 – 12 – 4 -1**, utilizando `tansig` como función de transferencia de las

dos capas ocultas y de salida, y, la función *trainlm* como algoritmo de aprendizaje. El resultado fue una mejora en el error de validación equivalente a 0.0457 y un factor R que se mantuvo por encima de 0.9. La convergencia gráfica se observa en la Figura 4 a continuación.

Figura 4. Convergencia de datos target (azul) y datos output (rojo).

El MATLAB, cuyo acrónimo proviene del concepto Laboratory Mathematical, nos permite evaluar el comportamiento de la red cascadefeedforward en forma gráfica. Así, en este formato, se puede observar el comportamiento de la curva de validación (verde), ver Figura 5. Se muestra que la mejor performance se obtiene con un valor mínimo equivalente a 0.045682, al cumplirse once épocas o iteraciones. Como se observará la curva de validación y de prueba (rojo) son muy similares, no existiendo separación en aumento, lo que demuestra que no existe sobreajuste.

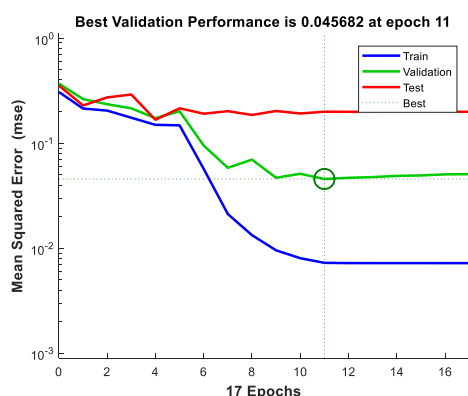


Figura 5. Mejor curva de validación para un error de 0.045682.

Ahora veamos el estado de entrenamiento (training state), el mismo que se puede observar en la Figura 6. Se puede notar que el gradiente, luego de alcanzar un valor máximo en la época 4, desciende sostenidamente hasta la época 17 donde alcanza un valor mínimo de 9.0163×10^{-5} . Lo mismo sucede con el factor de corrección μ que presenta una tendencia descendente.

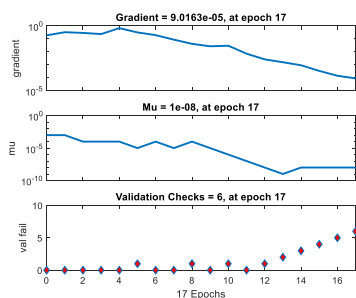


Figura 6. Estado del entrenamiento de la red cascadefeedforward.

En cuanto al Histograma de Error, se puede observar que los datos de entrenamiento, validación y prueba se concentran alrededor del error, lo que significa que no hay mayor dispersión y que existe una gran convergencia entre los datos objetivo (target) y los datos de salida (output), ver Figura 7.

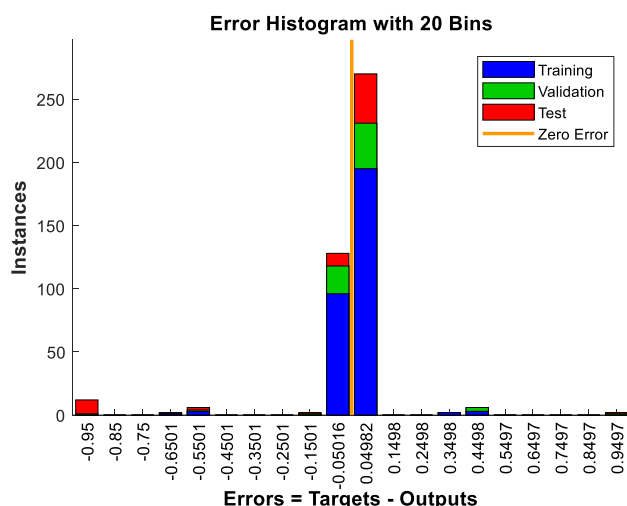


Figura 7. Histograma de error, muestra la convergencia de la red.

A continuación, se muestra la Gráfica de Regresión (ver Figura 8), que pone de manifiesto la correlación existente entre la salida objetivo (target) y la salida output obtenida luego del entrenamiento de la red cascadefeedforward.

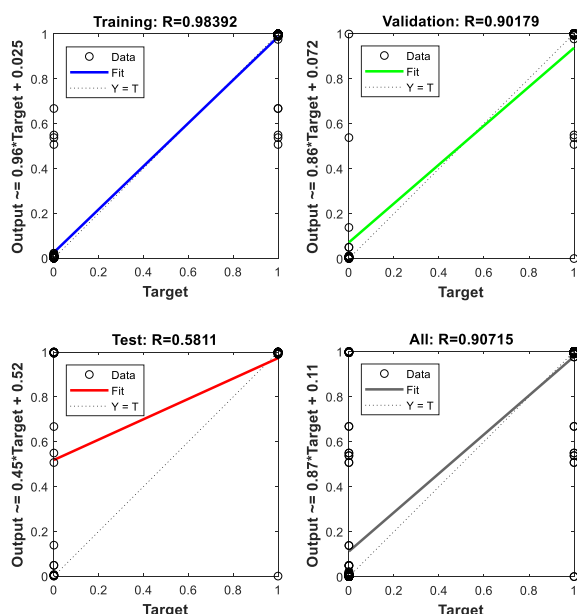


Figura 8. Mejor factor de correlación de la red cascadefeedforward, $R=0.90715$.

Se observa que la correlación en el bloque de entrenamiento y de validación son buenos, por encima de 0.9, sin embargo, el de mayor relevancia es sin duda el factor de

correlación “All”, que expresa a todas las correlaciones juntas, con un valor $R=0.90715$, que significa una importante correlación de 90.715 %, que demuestra la versatilidad de la RNA-MLP cascadefeedforward considerada en el estudio.

CONCLUSIONES

La red neuronal artificial – Multilayer Propagation (RNA-MLP), tipo cascadefeedforward, con arquitectura de red 24 – 12 – 4 -1 (24 neuronas en la capa de entrada, 12 neuronas en la primera capa oculta, 4 neuronas en la segunda capa oculta y una neurona en la capa de salida), con función de activación *tansig* en las capas ocultas y de salida, algoritmo de entrenamiento *trainlm* (Levenberg Marquardt), presenta un alto valor de factor de correlación $R=90.175$, lo que demuestra su validez para el diagnóstico del dengue a partir de los síntomas de la enfermedad en la región San Martín, con una efectividad equivalente a 90.175 %.

REFERENCIAS

- Aburas, H., Cetiner, B. y Sari, M. (2010). *Dengue confirmed-cases prediction: A neural network model*. Elsevier Ltd.
- Caicedo, E. López, J. (2017). *Una Aproximación Práctica a las Redes Neuronales Artificiales*. Cali: Universidad del Valle. Programa Editorial.
- Gambhira, S., Malika, S., y Kumar, Y. (2017). *PSO-ANN based diagnostic model for the early detection of dengue disease*. New Horizons in Translational Medicine. DOI: 10.1016/j.nhtm.2017.10.001.
- Hilera, J. Martínez, V. (1995). *Redes Neuronales Artificiales: Fundamentos, modelos y aplicaciones*. Madrid: RA-MA Editorial.
- Ibrahim, F., Nasir, M., Abu, W., Chong, Ch. y Sulaiman, S. (2005). *A novel dengue fever (DF) and dengue hemorrhagic fever (DHF) analysis using artificial neural network (ANN)*. Elsevier, Malaysia.
- Ibrahim, F., Faisal, T., Salim, M., y Taib, M. (2010). *Non-invasive diagnosis of risk in dengue patients using bioelectrical impedance analysis and artificial neural network*. Medical and Biological Engineering. DOI: 10.1007/s11517-010-0669-z.
- MINSA, (2020). *Casos de Dengue por Departamentos -2020*. Centro Nacional de Epidemiología, Prevención y Control de Enfermedades.
<https://www.dge.gob.pe/portal/docs/vigilancia/sala/2020/SE01/dengue.pdf>.

- MINSA, (2022). *Número de casos de dengue, 2018 – 2022**. Centro Nacional de Epidemiología, Prevención y Control de Enfermedades. StuDocu: <https://www.studocu.com/pe/document/universidad-senor-de-sipan/salud-publica/norma-tecnica-actualizada-del-minsa-sobre-dengue-en-peru/22783529>.
- Montaño, J. (2002). *Redes neuronales artificiales aplicadas al análisis de datos*. Palma de Mallorca, Universitat de les Illes Balears.
- OPS. (2016). *Number of Reported Cases of Dengue and Severe Dengue (SD) in the Americas, by Country*. Organización Panamericana de la Salud: Epidemiological Week / EW 41.
- Padrón, J. (2006). *Bases del concepto de Investigación Aplicada*. Caracas. Recuperado de: <http://padron.entretemas.com.ve/InvAplicada/index.htm>.
- Paletta, M. (2007). *¿Sistemas Inteligentes, realidad o fantasía?* Venezuela, Universidad Católica Andrés Bello: Revista GUAYANA. Recuperado de: http://books.google.com.pe/books?id=BCJm9jigHr4C&pg=PA77&lpg=PA77&dq=sistema+inteligente+mauricio+paletta&source=bl&ots=3UJ5FX-ZNJ&sig=Ju3dnlfXEBCZfH7P8nKdbBIJ-s&hl=es&ei=O-xFTsX7Jajm0QHwrLHxBw&sa=X&oi=book_result&ct=result&resnum=1&ved=0CB8Q6AEwAA#v=onepage&q&f=false.
- Paletta, M. (2010). *Inteligencia Artificial Básica*. Puerto Ordaz: Fondo Editorial UNEG.
- Wittkoski D. (2010). *Manual para la Formulación y Evaluación de Proyectos de Inversión*. Cooperación Técnica del BID. Lima.