

MODELO DE REDES NEURONALES OPTIMIZADAS CON ALGORITMOS GENÉTICOS: UNA APLICACIÓN PARA PROYECCIONES SOBRE PACIENTES CON CÁNCER HOSPITALIZADOS EN LA C.C.S.S.

ANDREY ALONSO MORA BERMÚDEZ
Maestría en Administración de Negocios
Universidad Estatal a Distancia, Costa Rica
mora.andrey@gmail.com

JUAN ANTONIO SANCHO CHACÓN
Maestría en Administración de Negocios
Universidad Estatal a Distancia, Costa Rica
jasancho2@gmail.com

RESUMEN

En el campo de la salud, es fundamental disponer de proyecciones efectivas de la demanda de los servicios, con el fin de realizar una adecuada planificación, garantizando la disposición oportuna de dichos servicios y una utilización eficiente de los recursos. Este artículo pretende ampliar el conocimiento teórico en relación con la utilización de técnicas de inteligencia artificial, específicamente las redes neuronales y los algoritmos genéticos, en el desarrollo modelos de proyección y optimización, con el fin de evaluar su aplicación práctica en la estimación del número de pacientes con cáncer atendidos en hospitales de la Caja Costarricense de Seguro Social (CCSS) así como en la estimación de la estancia hospitalaria esperada para este tipo de pacientes.

PALABRAS CLAVES: REDES NEURONALES, ALGORITMOS GENÉTICOS, CÁNCER.

ABSTRACT

In the health field, it is important to have effective demand estimates in order to better plan health services, to ensure the punctual provision of these services and to use resources efficiently. This paper increase the theoretical knowledge related to the use of artificial intelligence techniques, specifically neural net-

works and genetic algorithms, for the development of forecasting and optimization models, and it evaluates their practical application for estimating the number of cancer patients assisted by the Caja Costarricense de Seguro Social and for estimating the expected hospital stay time for this kind of patients.

KEYWORDS: NEURAL NETWORKS, GENETIC ALGORITHMS, CANCER.

INTRODUCCIÓN

Las proyecciones anticipadas del valor de una o más variables, constituyen una herramienta fundamental que disponen las organizaciones para realizar una adecuada planificación estratégica y tomar mejores decisiones.

Con el paso del tiempo se han ido desarrollando diversas técnicas de pronósticos, unas cualitativas y otras cuantitativas, que toman en consideración distintas variables relacionadas con el comportamiento de los datos. Asimismo, con el avance tecnológico, se ha diseñado aplicaciones que permiten realizar pronósticos con el fin de facilitarle esta tarea a las organizaciones. Sin embargo, aun-

que muchas técnicas utilizan estos programas computacionales, el grado de efectividad de los mismos no es constante, ya que los supuestos bajo los cuales trabajan no siempre se adecuan a los datos de entrada.

Los esfuerzos por mejorar la calidad de los pronósticos han dado paso a nuevas técnicas. Valenzuela (2008) se refiere a algunas de estas técnicas como sistemas basados en la inteligencia, los cuales se fundamentan en la forma en que funciona el cerebro humano o en la lógica de la evolución. Entre estas herramientas se incluyen los algoritmos genéticos (AG) y las redes neuronales artificiales (RNA).

De acuerdo con Toro, Mejía y Salazar (2004) las redes neuronales artificiales son modelos matemáticos que tratan de hacer una pequeña emulación del funcionamiento del cerebro humano. Son metodologías novedosas que permiten hacer pronósticos donde hay cierto comportamiento no lineal.

Las redes neuronales "aprenden" o modelan la dinámica del fenómeno en estudio y son capaces de estar constantemente cambiando para adaptarse a las nuevas condiciones (Nojek, Britos, Rossi y García, 2003).

Uno de los elementos clave para el funcionamiento de las redes neuronales es el proceso de aprendizaje, se han realizado algunos estudios que indican que el aprendizaje de las redes neuronales puede optimizarse utilizando algoritmos genéticos, obteniendo de esta forma resultados más robustos y cercanos a la realidad.

Para Herrera, Lozano y Verdegay (1994) "los Algoritmos Genéticos son procedimientos adaptativos para la búsqueda de soluciones en espacios complejos inspirados en la evolución biológica, con patrones de operaciones basados en el principio darwiniano de reproducción y supervivencia de los individuos que mejor se adaptan al entorno en el que viven". Esta dinámica es uti-

lizada por estos algoritmos para ir creando soluciones que vayan optimizando los valores de ciertas variables, con los que se puedan obtener mejores resultados para el problema que se está analizando.

En los últimos años, con el desarrollo más avanzado de las técnicas de los AG y las RNA, a nivel mundial se han realizado algunas investigaciones que relacionan estos dos temas, utilizándolos de manera conjunta para buscar obtener mejores resultados.

Orozco y Medina (2004) utilizaron la asociación de los AG y las RNA para crear un modelo de evolución del conocimiento en individuos en un sistema autorregulado, el cual permitía entender y manipular la evolución de las ideas y su transformación en conocimiento dentro de la mente humana. Así mismo se pretendía investigar cómo hacer para que un colectivo de individuos pudiera transmitir sus conocimientos a otros; de tal manera que lo que se buscaba era cómo hacer que los organismos pudieran comenzar su vida con conocimientos preestablecidos que ya han sufrido un proceso de selección por sus antecesores y que por lo tanto fueran mejorados con el pasar del tiempo.

Otros investigadores han utilizado estas técnicas para crear robots que se adaptan al ambiente cambiante gracias a algoritmos evolutivos y controladores neuronales, tal es el caso de Nolfi (citado por Orozco y Medina [2004]), quien junto con su equipo desarrollaron Khepera, un robot con estas características, para el cual desarrolló un sistema de auto aprendizaje con redes neuronales artificiales, las cuales seleccionaban sus unidades de entrada mediante algoritmos genéticos.

Por otra parte, existen casos en nuestro país donde se han desarrollado aplicaciones que combinan los pronósticos con los algoritmos genéticos, tal es el caso de Calvo (2004) y de Pacheco, Sancho y Zoch (2006), quienes utilizaron los al-

goritmos genéticos para generar pronósticos de variables económicas y ventas, respectivamente.

De forma separada estas técnicas han sido utilizadas en diferentes aplicaciones incluyendo el pronóstico de diversas variables que afectan el accionar de las organizaciones, sin embargo no se encontró evidencia de investigaciones donde utilicen redes neuronales optimizadas con algoritmos genéticos para realizar proyecciones.

La finalidad de este trabajo es ampliar el conocimiento teórico en relación con la utilización de las redes neuronales y los algoritmos genéticos en el desarrollo de modelos de proyección y optimización. Evaluando su aplicación práctica y sus resultados con respecto a otras metodologías de pronóstico que se han desarrollado y son utilizadas a nivel mundial.

Para ello se han desarrollado dos redes neuronales, las cuales fueron optimizadas mediante algoritmos genéticos, una utilizada para proyectar el número de casos nuevos de pacientes con cáncer que se estarían atendiendo en los hospitales de la CCSS, y la otra para determinar la estancia hospitalaria de este tipo de pacientes.

METODOLOGÍA

De acuerdo con Grajales (1996) la investigación se puede clasificar desde tres perspectivas diferentes: según la fuente de datos, según la finalidad y según la profundidad del estudio. Esta investigación es de tipo documental, aplicada, y explicativa.

Es documental porque para iniciarla fue necesario hacer una revisión detallada de libros, tesis y artículos con el fin de establecer los elementos teóricos y metodológicos relacionados con redes neuronales y algoritmos genéticos. Es aplicada, ya que después de la investigación teórica se desarrollaron dos aplicaciones prácticas para proyecciones sobre pacientes con cáncer hospi-

talizados en la CCSS. Finalmente, es explicativa, dado que el modelo busca explicar cómo las variables se relacionan para generar una salida determinada, es decir, tiene un componente de inferencia estadística.

Por otra parte, la investigación es predominantemente cuantitativa, ya que el objetivo fue el desarrollo de un modelo de pronósticos mediante redes neuronales, lo cual implica la relación de diferentes variables (categóricas y cuantitativas) para obtener un resultado numérico de la proyección; además se optimizaron las redes con algoritmos genéticos y se compararon con otros modelos tradicionales analizando los errores de pronósticos.

Los sujetos de esta investigación fueron los pacientes con cáncer internados en los diferentes hospitales de la Caja Costarricense de Seguro Social, considerando pacientes de todas las edades, de ambos sexos, cuyo diagnóstico principal para la hospitalización fuera cualquiera de los posibles tumores malignos que existen, de acuerdo con la Clasificación Internacional de Enfermedades, en su décima edición (CIE 10)¹.

La fuente de información principal correspondió a la base de datos de egresos hospitalarios de la Caja Costarricense de Seguro Social, la cual estaba constituida por 98 variables y con información sobre 119 132 egresos de pacientes² con cáncer, para el período comprendido entre los años 1997 y 2010.

Adicionalmente, para el estudio de las variables macroeconómicas y sociales que afectan el fenómeno en estudio se consultaron fuentes como el Ministerio de Planificación (MIDEPLAN), el Instituto Nacional de Estadística y Censos

1. OMS/OPS (2008). Clasificación Internacional de Enfermedades. Ginebra.
2. Si una misma persona ingresó y egresó del hospital en más de una ocasión durante el año, se contabiliza según el número de veces que egresó del centro de salud.

(INEC) y el Programa de Naciones Unidas para el Desarrollo (PNUD). Los datos fueron obtenidos por medio de Internet, en las páginas Web de estas instituciones.

Para la construcción del modelo de redes neuronales de casos nuevos se utilizaron todos los datos mensuales disponibles sobre la cantidad de pacientes nuevos con cáncer hospitalizados la CCSS en el período 1997-2010³, así como las variables macroeconómicas indicadas para este mismo período de estudio. En total se consideraron 168 datos.

En el caso del modelo de redes neuronales para estimación de la estancia hospitalaria, se trabajó con los datos disponibles sobre los pacientes con cáncer egresados únicamente en el año 2010, ya que a diferencia de la primera red, la variable tiempo no tiene incidencia en el fenómeno de estudio.

DESARROLLO

Tal como se mencionó anteriormente, para alcanzar el objetivo de la investigación se desarrollaron dos redes neuronales, una para la proyección

de casos nuevos de pacientes con cáncer hospitalizados por mes (Red de casos nuevos) y otra para determinar la estancia hospitalaria de este tipo de pacientes (Red de estancia).

Para desarrollar estas redes neuronales se siguieron una serie de pasos: análisis del fenómeno en estudio, definición de las variables de entrada, estructuración de la red, aprendizaje de la red y optimización con los algoritmos genéticos.

Análisis del fenómeno de estudio

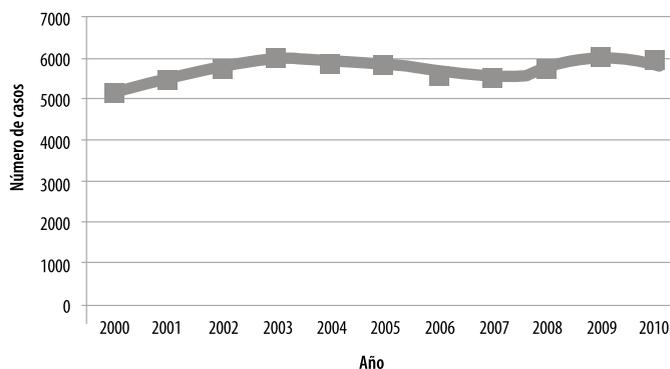
El primer paso fue analizar el comportamiento de estos dos fenómenos de estudio, así como las variables relacionadas con estos.

En la Figura 1 se observa cómo el número de casos nuevos de pacientes con cáncer hospitalizados ha variado entre los 5000 y 6000 anuales, con los años 2003 y 2009 como los de mayor número.

En cuanto a la estancia hospitalaria, en la Tabla 1 se puede observar cómo en su mayoría la estancia fue menor a 5 días y los plazos más largos correspondieron a pocos casos.

FIGURA 1

EVOLUCIÓN DEL NÚMERO DE CASOS NUEVOS DE PACIENTES CON CÁNCER HOSPITALIZADOS EN LA CCSS ENTRE LOS AÑOS 2000 Y 2010



FUENTE: Base de datos de egresos hospitalarios, Área de Estadísticas en Salud, CCSS (2010).

3. Para la red de casos nuevos se contabilizó cada paciente solo la primera vez que fue hospitalizado.

TABLA 1
DISTRIBUCIÓN DEL NÚMERO DE CASOS DE PACIENTES CON CÁNCER HOSPITALIZADOS EN LA CCSS, SEGÚN ESTANCIA TOTAL EN EL PERÍODO 2000-2010

Estancia total	2000		2005		2010	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje	Frecuencia	Porcentaje
Menos de 5 días	1356	28	2181	37	2183	37
De 5 a 10 días	1227	25	1396	24	1309	22
De 10 a 15 días	836	17	837	14	828	14
De 15 a 20 días	523	11	466	8	568	10
De 20 a 25 días	317	7	322	6	364	6
De 25 a 30 días	269	6	185	3	242	4
De 30 a 45 días	213	4	293	5	323	5
De 45 a 60 días	77	1,6	99	1,7	88	1,5
De 60 a 90 días	41	0,8	55	0,9	36	0,6
Más de 90 días	8	0,2	12	0,2	5	0,1
Total	5133	100	5846	100	5946	100

Fuente: Base de datos de egresos hospitalarios, Área de Estadísticas en Salud, CCSS (2010).

Selección de variables de entrada

Una vez que se realizó el análisis anterior, se procedió a identificar una serie de variables relacionadas con este fenómeno de estudio y que podrían servir para proyectarlo. A estas variables se les aplicó un análisis de correlaciones, con el fin de determinar si existe dependencia entre las mismas, de modo que se justifique su incorporación en el modelo de redes neuronales.

Para realizar el análisis se utilizó el estadístico Chi cuadrado⁴ para las relaciones entre variables

$$4. \chi^2 = \sum_i \frac{(\text{observada}_i - \text{teórica}_i)^2}{\text{teórica}_i}$$

categorías, los estadísticos F y D de Cohen⁵ para las relaciones entre una variable categórica y una variable cuantitativa, y el índice de Pearson⁶ para relaciones entre variables cuantitativas. Para calcular los diferentes índices se utilizó el programa estadístico SPSS versión 15.

En la Red de casos nuevos, las variables de entrada seleccionadas están compuestas por variables

$$5. d = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{(\sigma_1^2 + \sigma_2^2)/2}}$$

donde x_1 and x_2 son las medias del grupo 1 y grupo 2, y σ_1^2 y σ_2^2 son las varianzas respectivas

$$6. \rho_{x,y} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{E[(X - \mu_x)(Y - \mu_y)]}{\sigma_x \sigma_y}$$

Dónde: $\rho_{x,y}$ es el coeficiente de correlación de Pearson, σ_{xy} es la covarianza de (X,Y); σ_x es la desviación típica de la variable X, σ_y es la desviación típica de la variable Y

macroeconómicas y sociales como: tasa de mortalidad, índice de desarrollo humano, inversión social en salud, cobertura régimen de enfermedad y maternidad, tasa de desempleo y nivel de pobreza. Asimismo, se incluyeron algunas variables propias del fenómeno que mejoraran los resultados, las cuales fueron: mes, promedio de casos nuevos de pacientes con cáncer hospitalizados en los tres últimos años, promedio de los dos últimos años, promedio del último año, promedio de los últimos seis meses y promedio de los últimos tres meses.

Es importante mencionar que de las 12 variables seleccionadas, 3 no tenían una correlación significativa con la variable de salida (número de casos nuevos), sin embargo debido a que una de las características de las Redes Neuronales es que consideran comportamientos no lineales, se determinó incluirlas por criterio de experto. Estas variables son: tasa de mortalidad, inversión social en salud y nivel de pobreza.

En el caso de la Red de estancia, se seleccionaron las siguientes ocho variables de entrada: edad del paciente, sexo, residencia, centro de salud, área de ingreso, servicio de ingreso, ingresos previos y diagnóstico principal, todas relacionadas con el paciente como tal.

Estructuración de las redes

Solano (1997), Hilera y Martínez (2000) y Gallardo (2009) coinciden en los elementos en que se basa la arquitectura de una RNA. Se le llama arquitectura de una red neuronal a la estructura de interconexión del conjunto de nodos que componen la red, y que se organizan en capas.

Se conoce como capa o nivel a un conjunto de neuronas cuyas entradas provienen de la misma fuente (que puede ser otra capa de neuronas) y cuyas salidas se dirigen al mismo destino (que puede ser otra capa de neuronas).

Para el caso de los modelos más simples de RNA se puede encontrar una capa de entrada, y una capa de salida. Conforme fue avanzando el estudio de las RNA se vio la necesidad de incorporar capas intermedias, también llamadas ocultas, las cuales procesan la información dentro del sistema, pero no son visibles desde fuera del mismo.

El número de nodos en la capa de entrada depende de las variables de entradas consideradas para la representación del fenómeno de estudio. El número de nodos en la capa intermedia varía estableciéndose en muchos casos mediante prueba y error para elegir el número más conveniente. El número de nodos en la capa de salida depende del número de categorías que tenga la misma.

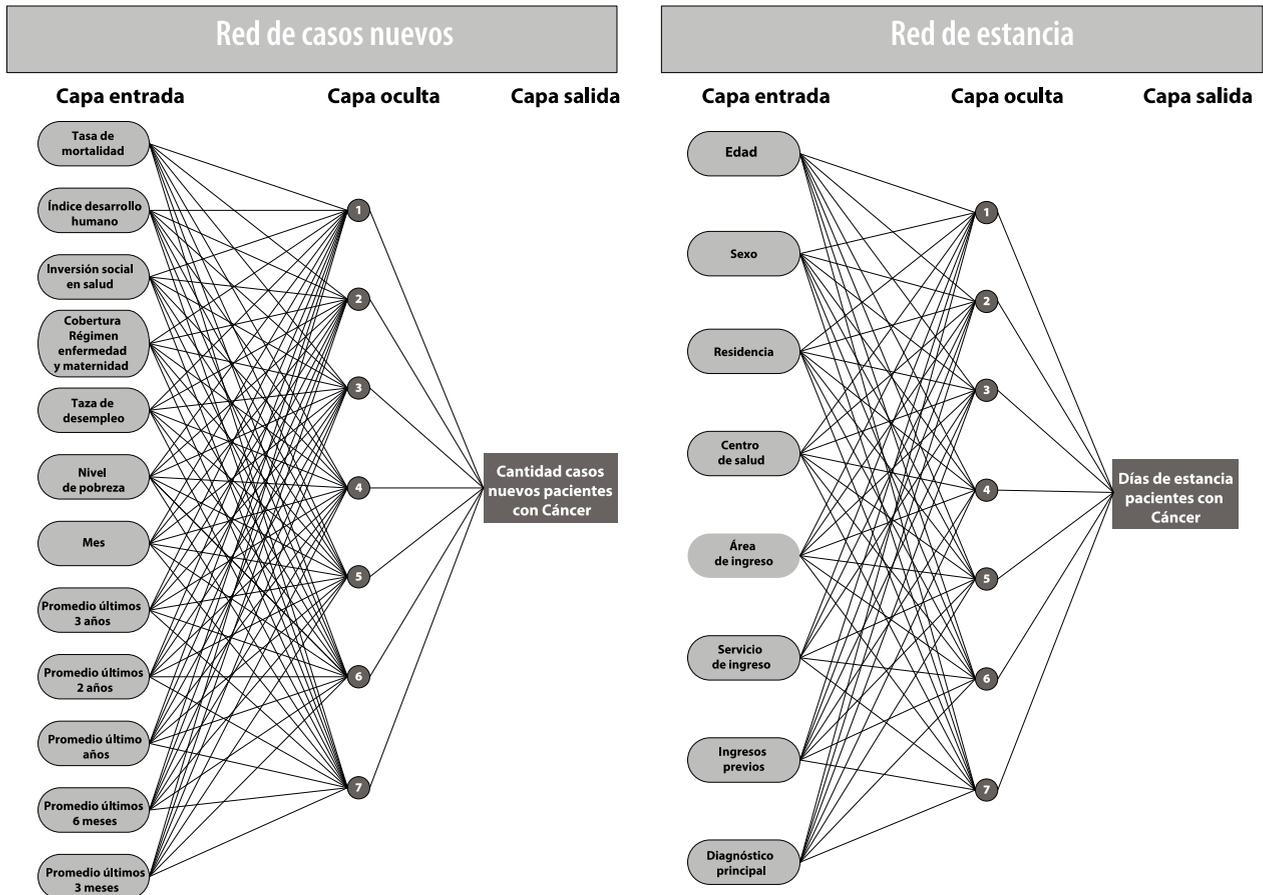
Para determinar la arquitectura de las redes neuronales que son objeto de esta investigación, se utilizaron los programas estadísticos SPSS 15 y Statistica 19; ambos tienen un módulo sobre redes neuronales, en el cual se pueden incluir diferentes escenarios sobre el comportamiento de las variables de entrada y de salida, para identificar el tipo de red que mejor se ajusta a los datos así como la cantidad de capas y neuronas ocultas recomendadas, seleccionando la combinación que genera el menor error de la red.

Con base en lo anterior, se determinó que la Red de casos nuevos está compuesta por una capa de entrada con 12 variables o neuronas, una capa oculta con 7 neuronas y una capa de salida con una neurona. Esto corresponde a una estructuración tipificada como 12:7:1 (ver Figura 2).

Por su parte la estructuración de la Red de estancia recomendada es de tipo 8:7:1 (ver Figura 2).

En la figura 2 se puede observar la estructuración de ambas redes, donde están representadas las variables de entrada en círculos, las cuales se interrelacionan con las variables ocultas (en negro) y luego como estas definen la variable de salida en cuadrado.

FIGURA 2
ESTRUCTURACIÓN DE LAS REDES NEURONALES: CASOS NUEVOS Y ESTANCIA



Fuente: Elaboración propia.

Funcionamiento de la red

Cada neurona de la capa de entrada está caracterizada en cualquier instante por un valor numérico denominado valor o estado de activación $a_i(t)$. Este corresponde al valor de cada variable de entrada que alimenta la red. Ver ejemplo para la Red casos nuevos en la Tabla 2.

A estos valores de entrada se les aplica una función de salida que transforma el estado actual de activación en una señal de salida. Dicha señal

es enviada a través de los canales de comunicación unidireccionales a otras unidades de la red (neuronas ocultas). En estos canales la señal se modifica de acuerdo con la sinapsis (el peso w_{ih}) asociada a cada uno de ellos.

En el caso de las dos redes objeto de esta investigación se definió que la función de activación sería la función identidad $F(y) = y$, por lo que el parámetro que se le pasa a la neurona oculta será directamente el valor de la variable de entrada ($a_i(t) = y_i$) multiplicada por un peso w .

Tabla 2
EJEMPLO DE NEURONAS DE ENTRADA EN LA RED DE CASOS NUEVOS

Neurona de entrada	Valor o estado de activación $a_i(t)$
Mes	5 (mayo)
Promedio 3 años	527 casos nuevos
Promedio 2 años	527 casos nuevos
Promedio 1 año	527 casos nuevos
Promedio 6 meses	527 casos nuevos
Promedio 3 meses	509 casos nuevos
Tasa general de mortalidad	4
Índice de Desarrollo Humano	0,85
Inversión Social en Salud	4,8%
Cobertura Régimen de Enfermedad y Maternidad	86,2%
Tasa de desempleo	4,2%
Nivel de pobreza	20%

Fuente: Elaboración propia.

Cada neurona de la capa de entrada tiene un peso w asociado a cada una de las neuronas ocultas (w_{ih})⁷, y cada neurona oculta tiene un peso asociado a la neurona de salida (w_{ho}); estos pesos son clave ya que, son los que permiten que la red aprenda del fenómeno de estudio.

En el caso de la Red de casos nuevos, se tienen 12 neuronas de entrada y 7 neuronas ocultas, por lo que en total se requieren 91 pesos w ($12 \cdot 7 + 7$). Por su parte en la Red de estancia se tienen un total de 63 pesos w ($8 \cdot 12 + 7$). Estos pesos se inicializan con un valor aleatorio, y conforme la red evoluciona (aprende) adquieren valores positivos, negativos o cero. Ver ejemplo para la Red de casos nuevos en la Tabla 3.

7. El subíndice i se refiere a la capa de entrada (input en inglés), h se refiere a la capa oculta (hidden en inglés) y el subíndice o se refiere a la capa de salida (output en inglés).

Si w_{ih} es positivo, indica que la interacción entre las neuronas i y h es excitadora; es decir, siempre que la neurona i esté activada, la neurona h recibirá una señal de i que tenderá a activarla. Si w_{ih} es negativo, la sinapsis será inhibitoria. En este caso, si i está activada, enviará una señal a h que tenderá a desactivar a ésta. Finalmente, si $w_{ih} = 0$, se supone que no hay conexión entre ambas. Lo mismo sucede con los pesos w_{ho} .

Los valores de las neuronas de entrada se dividieron entre un factor común (entre 1000), ya que, las pruebas iniciales realizadas en los modelos demostraron que al hacer esto se obtenían pesos relativos con valores manejables y comprensibles, lo cual permitía determinar diferencias entre las diversas estructuraciones de la red, pero sin alterar la relación de importancia que tuvieran estas variables entre sí en el mismo modelo.

TABLA 3
EJEMPLO DE PESOS Y ENTRADAS NETAS EN LA RED DE CASOS NUEVOS

Neuronas de entrada	Valor Y_i	Pesos						
		w_{ih1}	w_{ih2}	w_{ih3}	w_{ih4}	w_{ih5}	w_{ih6}	w_{ih7}
Mes	0,005	-10,05	-4,93	40,67	1,55	58,28	24,44	-25,82
Promedio 3 años	0,527	3,15	-28,27	3,04	-21,53	29,73	-20,24	59,07
Promedio 2 año	0,527	25,28	20,01	15,13	-37,54	17,31	36,27	-3,85
Promedio 1 año	0,527	-12,83	-2,62	29,95	9,90	-40,34	-13,43	-6,37
Promedio 6 meses	0,527	-1,58	12,68	-7,23	21,84	-56,84	-17,96	16,05
Promedio 3 meses	0,509	11,16	4,58	25,71	-121,30	-8,40	0,31	42,13
Tasa general de mortalidad	0,004	148,85	19,50	5,11	20,91	-11,98	7,70	-24,17
Índice de Desarrollo Humano	0,00085	-5,70	-42,19	31,84	-26,72	97,09	37,37	1,94
Inversión Social en Salud	0,0048	1,64	-9,29	-31,65	-4,64	-8,76	0,10	10,79
Cobertura Régimen de Enfermedad y Maternidad	0,0862	-57,27	48,04	4,95	8,52	19,79	-23,87	7,26
Tasa de desempleo	0,0042	-13,11	-21,05	-6,47	-5,51	-23,60	46,61	-10,42
Nivel de pobreza	0,02	78,65	-61,20	239,78	1,40	15,03	26,31	18,13
	Valor (Net_h)	10,20	6,08	39,92	-75,36	-28,51	-9,09	56,42
	w_{ho}	59,70	13,67	69,70	24,96	12,58	117,29	4,53
	Neuronas ocultas	O_{h1}	O_{h2}	O_{h3}	O_{h4}	O_{h5}	O_{h6}	O_{h7}

Fuente: Elaboración propia.

Cada neurona en la capa oculta recibe señales de cada una de las neuronas de entrada. Las señales que llegan a cada neurona oculta se combinan entre ellas, generando así la entrada total, Net_h . ($Net_h = \sum y_i w_{ih}$). Ver ejemplo para la Red de casos nuevos en la Tabla 4.

Al valor Net_h de cada neurona oculta se le aplica nuevamente una función de activación para enviar una señal a la neurona de salida. En este caso, para ambas redes, se utilizó la función identidad $F(y) = y$, por lo que Net_h es igual a y_h

Multiplicando los valores de salida de las neuronas ocultas por los pesos relativos w_{ho} se generó la entrada total para la neurona de salida Net_o ($Net_o = \sum_h y_h w_{ho}$).

A esta entrada total se le aplicó una función de salida $F(Net_o)$ para generar el valor de salida final de la red (ya sea el número de casos nuevos de pacientes con cáncer hospitalizados en el mes, para la Red de casos nuevos; o los días de hospitalización del paciente con cáncer para la Red de estancia). Nuevamente en ambas redes se utilizó la función identidad $F(y) = y$. Ver ejemplo para la Red de casos nuevos en la Tabla 5.

TABLA 4
EJEMPLO DE CÁLCULO DE LA ENTRADA NETA PARA LA NEURONA OCULTA OH1 EN LA RED DE CASOS NUEVOS

Valor de entrada (x_i)	w_{ih1}	$y_i * w_{ih1}$
0,005	-10,051	-0,050
0,527	3,149	1,659
0,527	25,275	13,320
0,527	-12,830	-6,762
0,527	-1,579	-0,832
0,509	11,161	5,681
0,004	148,853	0,595
0,00085	-5,702	-0,005
0,0048	1,643	0,008
0,0862	-57,267	-4,936
0,0042	-13,106	-0,055
0,02	78,650	1,573
$Net_{h1} (\sum_i y_i w_{ih})$.		10,197

Fuente: Elaboración propia.

TABLA 5
EJEMPLO DE CÁLCULO DE UNA SALIDA FINAL EN LA RED DE CASOS NUEVOS

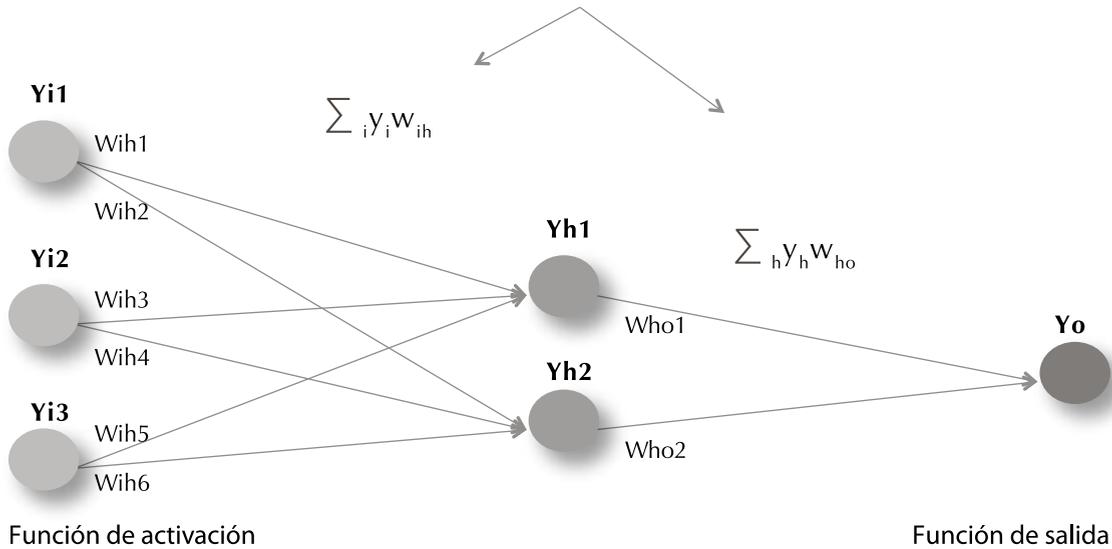
$Net_h (y_h)$	w_{ho}	$y_h * w_{ho}$
10,197	59,702	608,8
6,078	13,675	83,1
39,923	69,700	2782,6
-75,359	24,956	-1880,7
-28,508	12,582	-358,7
-9,090	117,294	-1066,2
56,417	4,528	255,5
Valor de la neurona salida (Net_o)		424,404

Fuente: Elaboración propia.

El proceso de funcionamiento de la red neuronal explicado anteriormente se esquematiza en la Figura 3.

Con base en esta estructuración y el funcionamiento descrito, la red neuronal genera la primera salida, en este caso, el valor pronosticado. A partir de esta salida la red inicia un proceso de aprendizaje, para lo cual se utiliza una regla determinada. El proceso consiste en comparar el valor de salida (pronosticado) con el valor real y ajustar los pesos de las conexiones entre neuronas; esto implica que la red se debe entrenar.

FIGURA 3
FUNCIONAMIENTO DE LA RED NEURONAL



Fuente: Elaboración propia.

Entrenamiento de la red

De acuerdo con Sotolongo y Guzmán (2001), el proceso de aprendizaje consiste en hallar los pesos que codifican los conocimientos (disminuyen el error). Una regla de aprendizaje hace variar el valor de los pesos de una red hasta que estos adoptan un valor constante ($dw/dt=0$), cuando esto ocurre se dice que la red ya ha aprendido.

Para el caso de las redes neuronales de casos nuevos y de estancia, se utilizó el modelo de aprendizaje conocido como "propagación hacia atrás"⁸.

En el modelo de propagación hacia atrás, todas las neuronas interfieren en una disminución del error por medio de la modificación de sus pesos; de ahí que el trabajo es dividido en muchas pequeñas tareas que cada neurona realiza.

Para aplicar este algoritmo primero se compara el resultado obtenido en las neuronas de salida con la salida real de la red y se calcula un valor del error para cada neurona oculta y de salida.

Para las neuronas de la capa de salida, el valor del error (delta) es:

$$\delta_o = (d_o - y_o) f'_o(Net_o)$$

donde d_o es el valor real y y_o es el valor de salida de la red ($y_o = Net_o$).

Continuando con el ejemplo de la tabla 6, asumiendo que la salida real era de 515 casos nuevos y la calculada por la red fue de 424 casos nuevos, el error de la red sería de 91 casos (515-424).

Tanto para la Red de estancia como para la Red de casos nuevos, la derivada de la función de salida $f'_o(Net_o)$ es igual a 1, ya que, f es la función identidad y su derivada es 1.

8. En inglés se conoce como *back propagation*.

Para las neuronas ocultas, el error (delta) está dado por:

$$\delta_h = f'_h(\text{Net}_h) \sum \delta_o w_{ho}$$

Donde se observa que el error en las capas ocultas depende de todos los términos de error de la capa de salida. De aquí surge el término propagación hacia atrás. Ver ejemplo para la Red de casos nuevos en la Tabla 6.

TABLA 6
EJEMPLO DE CÁLCULO DEL ERROR PARA LAS NEURONAS OCULTAS RED DE CASOS NUEVOS

Neurona oculta	δ_o	W_{ho}	$\delta_h (\delta_o * W_{ho})$
O _{h1}	90,596	59,702	5408,710
O _{h2}	90,596	13,675	1238,850
O _{h3}	90,596	69,700	6314,481
O _{h4}	90,596	24,956	2260,909
O _{h5}	90,596	12,582	1139,886
O _{h6}	90,596	117,294	10626,291
O _{h7}	90,596	4,528	410,232

Fuente: Elaboración propia.

Nuevamente, en el caso de las redes de casos nuevos y de estancia, $f'_h(\text{Net}_h)$ es igual a 1, ya que se utilizó la función identidad.

Posteriormente, estos errores se transmiten hacia atrás por medio de un ajuste en los pesos de la red, partiendo de la capa de salida, hacia todas las neuronas de la capa intermedia que contribuyan directamente a la salida, recibiendo el porcentaje de error aproximado a la participación de la neurona intermedia en la salida original.

Los pesos se ajustan de la forma siguiente:

Para los pesos de las neuronas de la capa oculta:

$$w_{ho}(t + 1) = w_{ho}(t) + \Delta w_{ho} t + 1$$

$$\Delta w_{ho}(t + 1) = \alpha \delta_o y_h$$

donde δ_o es el error de salida, y_h es el valor de salida de las neuronas (que es igual a Net_h) y α es una constante que corresponde a la tasa de aprendizaje, es decir a la velocidad con que se espera que la red neuronal aprenda.

En el caso de las redes neuronales de estancia y de casos nuevos, al inicio se utilizó un valor aleatorio para la constante α entre 0 y 1. Sin embargo, se determinó que se generan mejores resultados iniciando con un aleatorio de orden de 10^{-5} .

Continuando con el ejemplo, utilizando un α de $2,81 \times 10^{-5}$ y considerando que δ_o es igual 90, el valor del diferencial $\Delta w_{ho}(t + 1)$ para las neuronas ocultas se presenta en la Tabla 7.

TABLA 7
CÁLCULO DEL ERROR PARA LAS NEURONAS OCULTAS RED DE CASOS NUEVOS

fa	y_h ó Net_h	δ_o	Δw_{ho}
2,81E-05	10,197	90,596	0,0259
2,81E-05	6,078	90,596	0,0155
2,81E-05	39,923	90,596	0,1015
2,81E-05	-75,359	90,596	-0,1916
2,81E-05	-28,508	90,596	-0,0725
2,81E-05	-9,090	90,596	-0,0231
2,81E-05	56,417	90,596	0,1434

Fuente: Elaboración propia.

El proceso se repite, capa por capa, hasta que todas las neuronas de la red hayan recibido un error que describa su aportación relativa al error total.

Para los pesos de las neuronas de la capa de entrada, el error (delta) estaría dado por:

$$W_{ih}(t + 1) = W_{ih}(t) + \Delta W_{ih}(t + 1);$$

$$\Delta W_{ih}(t + 1) = \alpha \delta_{hi} x_i$$

En las tablas 8 y 9 se presentan los cálculos obtenidos para el ejemplo de la Red de casos nuevos.

TABLA 8
EJEMPLO DE CÁLCULO DEL ERROR PARA LAS NEURONAS DE ENTRADA EN LA RED DE CASOS NUEVOS

Neuronas entrada	Valor x_i	Diferencial de los pesos de las variables de entrada Δw_{ih}						
		Δw_{ih1}	Δw_{ih2}	Δw_{ih3}	Δw_{ih4}	Δw_{ih5}	Δw_{ih6}	Δw_{ih7}
Mes	0,005	0,000759	0,000174	0,000886	0,000317	0,000160	0,001491	0,000058
Prom 3 años	0,527	0,079985	0,018320	0,093379	0,033435	0,016857	0,157143	0,006067
Prom 2 años	0,527	0,079985	0,018320	0,093379	0,033435	0,016857	0,157143	0,006067
Prom 1 año	0,527	0,079985	0,018320	0,093379	0,033435	0,016857	0,157143	0,006067
Prom 6 meses	0,527	0,079985	0,018320	0,093379	0,033435	0,016857	0,157143	0,006067
Prom 3 meses	0,509	0,077253	0,017695	0,090190	0,032293	0,016281	0,151776	0,005859
TM	0,004	0,000607	0,000139	0,000709	0,000254	0,000128	0,001193	0,000046
IDH	0,00085	0,000129	0,000030	0,000151	0,000054	0,000027	0,000253	0,000010
ISS	0,0048	0,000729	0,000167	0,000851	0,000305	0,000154	0,001431	0,000055
CREM	0,0862	0,013083	0,002997	0,015274	0,005469	0,002757	0,025703	0,000992
TD	0,0042	0,000637	0,000146	0,000744	0,000266	0,000134	0,001252	0,000048
Np	0,02	0,003035	0,000695	0,003544	0,001269	0,000640	0,005964	0,000230
fa		0,0000281	0,0000281	0,0000281	0,0000281	0,0000281	0,0000281	0,0000281
δ_{ho}		5408,71	1238,85	6314,48	2260,91	1139,89	10 626,29	410,23
Neuronas ocultas		O_{h1}	O_{h2}	O_{h3}	O_{h4}	O_{h5}	O_{h6}	O_{h7}

Fuente: Elaboración propia.

TABLA 9
**EJEMPLO DE CÁLCULO DEL Δw_{ih} PARA
 LA NEURONA DE ENTRADA 2 (PROM 3AÑOS)
 EN LA RED DE CASOS NUEVOS**

	fa	δ_{ho}	x_i	Δw_{ih}
O_{h1}	0,0000281	5408,71	0,527	0,07998
O_{h2}	0,0000281	1238,85	0,527	0,01832
O_{h3}	0,0000281	6314,48	0,527	0,09338
O_{h4}	0,0000281	2260,91	0,527	0,03343
O_{h5}	0,0000281	1139,89	0,527	0,01686
O_{h6}	0,0000281	10 626,29	0,527	0,15714
O_{h7}	0,0000281	410,23	0,527	0,00607

Fuente: Elaboración propia.

Finalmente, para determinar los pesos a utilizar en la siguiente corrida de la red, se le suma el diferencial de cada neurona de entrada y cada neurona oculta al peso anterior:

$$W(t + 1) = W(t) + \Delta W$$

El proceso se repite hasta que el término de error resulta aceptablemente pequeño para cada uno de los patrones aprendidos.

$$E_p = \frac{1}{2} \sum_{k=1}^M \delta_{pk}^2$$

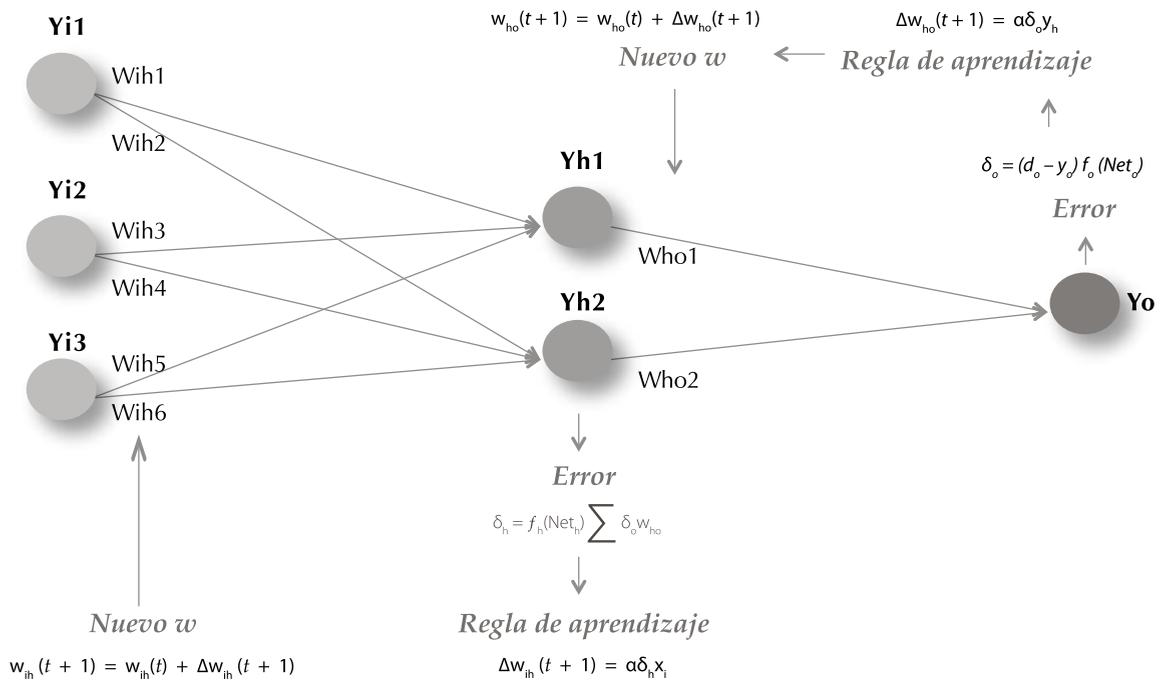
El proceso de aprendizaje descrito se realizó con los datos de 168 meses disponibles para la Red de casos nuevos de pacientes con cáncer.

En el caso de la Red de estancia, aunque se disponía información sobre 71 631 casos de los últimos 14 años, se trabajó únicamente con los datos de pacientes egresados en el 2010 ya que, a diferencia de la otra red, la variable tiempo no tiene incidencia en el fenómeno de estudio. En este caso podrían afectar otras variables relacionadas con proceso, gestión interna o las características propias del paciente, por lo que al usar datos del 2010 se busca que el modelo represente la realidad más reciente de los hospitales y de la salud de la población en general.

De los casos del 2010 que se disponía, se seleccionaron los últimos 1000 casos, ya que, el programa informático utilizado para la creación de los modelos de la red (Microsoft Excel) tiene una capacidad limitada para manejar grandes volúmenes de datos relacionados con múltiples fórmulas matemáticas y macros, por lo que si se utilizaban más casos el programa no funcionaba adecuadamente.

En la figura 4 se esquematiza el entrenamiento de una red neuronal.

FIGURA 4
ENTRENAMIENTO DE LA RED NEURONAL



Fuente: Elaboración propia.

Optimización con algoritmos genéticos

Una vez construidos los modelos de redes neuronales, el siguiente paso fue la optimización mediante algoritmos genéticos.

El uso de la técnica de algoritmos genéticos está más generalizado en problemas de optimización, ya sea para maximizar beneficios o minimizar errores. Su aplicación implica la definición de conjuntos de solución del problema y de una función de evaluación que permita discriminar entre las diferentes soluciones planteadas.

Esta técnica de simple aplicación pero con resultados robustos, probados en diferentes situaciones, se utiliza para optimizar los resultados de las dos redes neuronales, objeto de esta investigación. Lo anterior se lleva a cabo básicamente para buscar un conjunto de soluciones iniciales que permitan a la red generar resultados satis-

factorios en un tiempo más corto, o sea, acelerando la adaptación de la red a las condiciones de las diferentes variables que influyen en el resultado esperado.

Para la aplicación de esta técnica se tienen que definir una serie de elementos, los cuales conforman el algoritmo, estos son definidos por Moujahid, Inza y Larrañaga (s.f.) y comprenden la población inicial, la función de evaluación, los procedimientos de selección de individuos, las condiciones de los cruces y la mutación, que van a permitir al algoritmo moverse a través del espacio de solución, donde se pueda encontrar la solución óptima.

El algoritmo definido tanto para la Red de casos nuevos como para la de estancia hospitalaria, fue un algoritmo de minimización del error de pronóstico o estimación.

Es importante destacar que para la elaboración del algoritmo genético las soluciones son una combinación de los valores de los diferentes pesos, que relacionan las variables de entrada con las neuronas ocultas (w_{ih}) y a su vez estas con la capa de salida (w_{ho}), e incluyendo la tasa de aprendizaje (α), ya que en las pruebas realizadas con las redes neuronales se determinó que esta tasa de aprendizaje influía de manera importante en los resultados.

Algo necesario de resaltar es que la optimización se hizo con base en toda la red. Es decir, incluyendo tanto la etapa de construcción como la etapa de entrenamiento, para que no se perdiera la consistencia de la red de aprender de todos los casos históricos y con ello generara una mejor adaptación.

Como primer paso para la optimización se define el tamaño de la población. Para ambas redes el tamaño se definió en ocho individuos, de manera que se analizara un importante número de soluciones, pero, sin hacer demasiado lento y complicado el proceso, ya que entre más individuos se evalúen más se va a tardar en tiempo de procesamiento.

Posteriormente se establece la población inicial, que es el conjunto de soluciones sobre el cual se inicia el algoritmo genético. En el caso de la Red de casos nuevos, la población inicial se definió con números aleatorios con valores entre -50 y 100 para los pesos que relacionen las diferentes neuronas, y un aleatorio entre 0 y 1, multiplicado por 10^{-4} para la tasa de aprendizaje.

En el caso de la Red de estancia, la población inicial se definió con valores entre -2 y 2 para los pesos de las relaciones entre las neuronas y un aleatorio entre 0 y 1 multiplicado por 10^{-3} para la tasa de aprendizaje.

Estos valores fueron obtenidos analizando los resultados del aprendizaje que se había realizado para ambas redes, tratando de utilizar rangos de posibles cantidades que incluyeran los valores observados en las variables posterior al entrenamiento efectuado a la red.

Para ilustrar el proceso de optimización se desarrolla un ejemplo para la Red de casos nuevos, pero con sólo una parte de los pesos que se deben ir modificando. La población inicial se presenta en la tabla 10.

Tabla 10
EJEMPLO DE DEFINICIÓN DE LA POBLACIÓN INICIAL

Cromosoma	FA	W1	W2	W3	W4	W5
1	1,7329E-05	-6,55979	-19,594	-9,7086	-46,472	9,8176
2	2,1713E-05	14,9468	-3,37171	-46,188	3,451	13,756
3	3,6472E-05	-29,2447	40,4086	-47,805	46,352	-34,428
4	2,3098E-05	-20,9298	-19,0834	49,8861	-15,016	-45,583
5	4,0554E-05	7,07829	35,8608	-9,2617	-33,74	-19,701
6	1,4175E-05	-37,1374	27,3395	-35,429	10,617	-6,686
7	5,6424E-06	-15,5153	29,6089	-30,053	-41,071	39,307
8	8,3497E-05	46,811	29,0125	31,2843	-21,942	33,752

Fuente: Elaboración propia.

El siguiente paso es realizar la evaluación de los individuos, para lo cual se utiliza la función objetivo o función de evaluación. Para ambos modelos es la función que define el error de pronóstico conocido como Error Porcentual Medio Absoluto o MAPE (por sus siglas en inglés).

$$MAPE = \frac{\sum |PE|}{n}$$

Para aplicar esta función, se utilizó el error que se presenta en los datos correspondientes al entre-

namiento de cada red, o sea el último 20% en la Red de casos nuevos y 15% en el caso de la Red de estancia.

El valor se obtuvo definiendo cada una de las soluciones como el conjunto inicial de pesos de la red neuronal, con lo cual se generaba el error del entrenamiento, que es el valor que se deseaba minimizar. Ver la evaluación de la población inicial en la tabla 11.

TABLA 11
EJEMPLO DE EVALUACIÓN DE LA POBLACIÓN INICIAL

Cromosoma	Error	FA	W1	W2	W3	W4	W5
1	28,94%	1,7329E-05	-6,55979	-19,594	-9,7086	-46,472	9,8176
2	9,63%	2,1713E-05	14,9468	-3,37171	-46,188	3,451	13,756
3	11,48%	3,6472E-05	-29,2447	40,4086	-47,805	46,352	-34,428
4	18,53%	2,3098E-05	-20,9298	-19,0834	49,8861	-15,016	-45,583
5	17,39%	4,0554E-05	7,07829	35,8608	-9,2617	-33,74	-19,701
6	13,67%	1,4175E-05	-37,1374	27,3395	-35,429	10,617	-6,686
7	110,85%	5,6424E-06	-15,5153	29,6089	-30,053	-41,071	39,307
8	22,13%	8,3497E-05	46,811	29,0125	31,2843	-21,942	33,752

Fuente: Elaboración propia.

Seguidamente se realiza la selección de los individuos. De acuerdo con la evaluación de los mismos, y dada la naturaleza de la función de evaluación, lo que se busca es que los individuos minimicen el error de pronóstico, por lo tanto se seleccionan aquellos que generan los errores menores, de manera que estos continúen en el proceso. Para efectos de ambas redes se definió la selección del 50% de los individuos de la población para continuar en el proceso y para que funcionaran como padres al momento de rea-

lizar los cruces. Esto permitió que el algoritmo genético en esta etapa renovara la mitad de los individuos de la población en estudio, creando un balance entre lo que se tiene y lo nuevo por explorar.

Para el ejemplo que se está desarrollando, se necesitaría ordenar la población de acuerdo con el error y seleccionar los individuos con mejores resultados, tal como se muestra en la tabla 12.

Tabla 12
EJEMPLO DE SELECCIÓN DE LOS INDIVIDUOS

Cromosoma	Error	FA	W1	W2	W3	W4	W5
2	9,63%	2,1713E-05	14,9468	-3,37171	-46,188	3,451	13,756
3	11,48%	3,6472E-05	-29,2447	40,4086	-47,805	46,352	-34,428
6	13,67%	1,4175E-05	-37,1374	27,3395	-35,429	10,617	-6,686
5	17,39%	4,0554E-05	7,07829	35,8608	-9,2617	-33,74	-19,701
4	18,53%	2,3098E-05	-20,9298	-19,0834	49,8861	-15,016	-45,583
8	22,13%	8,3497E-05	46,811	29,0125	31,2843	-21,942	33,752
1	28,94%	1,7329E-05	-6,55979	-19,594	-9,7086	-46,472	9,8176
7	110,85%	5,6424E-06	-15,5153	29,6089	-30,053	-41,071	39,307

Fuente: Elaboración propia.

Posteriormente, se realiza la combinación de los individuos con mejores resultados (padres) que busca que se generen nuevos individuos (hijos) con mejores resultados que aquellos que les dieron origen.

Para ambas redes se realizaron dos cruces, donde se generaron cuatro nuevos individuos, de manera que se volviera a completar el tamaño de población a ocho.

En cuanto a la forma del cruce, se especificó una función de combinación definida como una proporción entre los valores de cada uno de los padres, para cada uno de los elementos que conforman el individuo, la proporción estuvo definida por la fórmula:

$$H_1 = P + \alpha(P - M)$$

$$H_2 = P - \alpha(P - M)$$

Donde:

H_1 = Valor del elemento del primer hijo.

H_2 = Valor del elemento del segundo hijo.

P = Valor del elemento del padre.

M = Valor del elemento de la madre.

α = Número aleatorio entre 0 y 1.

Para efectos de la escogencia de individuos que funcionaran como padre o madre se definió que de los cuatro mejores resultados, que se seleccionaban para continuar en el proceso, el mejor de ellos iba a funcionar como padre en ambos cruces; mientras que en el primer cruce el segundo mejor resultado funcionaba de madre y para el segundo cruce, la madre era el tercer mejor resultado. Esta combinación permitía darle mayor énfasis al mejor resultado encontrado, buscando nuevos espacios de solución en la zona cercana a este, que trataran de converger más rápidamente a resultados mejores.

De estos dos cruces surgen cuatro nuevos individuos, los cuales pasan a formar parte de la población de estudio y se evalúan en la red neuro-

nal, para determinar el error que puedan proporcionar. Además, estos individuos son ordenados según el error generado, con tal de diferenciar los mejores individuos que se van obteniendo.

Para el ejemplo, los datos de los nuevos individuos, debidamente evaluados, se muestran en la tabla 13.

TABLA 13
EJEMPLO DE EVALUACIÓN DE LOS NUEVOS INDIVIDUOS

Cromosoma	Error	FA	W1	W2	W3	W4	W5
2	9,63%	2,1713E-05	14,9468	-3,37171	-46,188	3,451	13,756
3	11,48%	3,6472E-05	-29,2447	40,4086	-47,805	46,352	-34,428
6	13,67%	1,4175E-05	-37,1374	27,3395	-35,429	10,617	-6,686
5	17,39%	4,0554E-05	7,07829	35,8608	-9,2617	-33,74	-19,701
9	16,50%	1,2793E-05	41,6544	-29,8308	-45,211	-22,477	42,877
10	13,88%	3,0633E-05	-11,7608	23,0874	-47,165	29,379	-15,364
11	12,41%	2,3541E-05	27,5759	-10,8184	-48,797	1,7134	18,713
12	8,60%	1,9885E-05	2,31769	4,07499	-43,579	5,1886	8,7996

Fuente: Elaboración propia.

Luego se realiza la mutación, o sea la aplicación de pequeños cambios aleatorios a las soluciones, de manera que se puedan explorar otras zonas del conjunto solución. Para efectuar el proceso de mutación en ambas redes se utilizó una tasa del 33%, la cual responde a la necesidad de equilibrar entre la ampliación de las posibilidades de encontrar mejores soluciones en otras áreas del conjunto solución, con el control de la eficiencia del proceso, de manera que este no se haga muy lento y complicado.

Para efectos de definir los individuos que se iban a mutar, se respetó los dos mejores individuos

que se tenían en la población de estudio, los cuales no se modificaron, y se aplicó la mutación a los demás individuos.

Para definir los individuos y los pesos exactos que se iban a estar mutando de cada conjunto solución, se aplicó un proceso de escogencia aleatoria; adicionalmente el valor a mutar se sustituyó con un valor aleatorio tal como los definidos en la población inicial. Ver ejemplo de mutación en los elementos marcados de la tabla 14.

TABLA 14
EJEMPLO DE MUTACIÓN DE LOS INDIVIDUOS

Cromosoma	FA	W1	W2	W3	W4	W5
12	1,9885E-05	2,31769	4,07499	-43,579	5,1886	8,7996
2	2,1713E-05	14,9468	-3,37171	-46,188	3,451	13,756
3	3,6472E-05	-29,2447	40,4086	-47,805	46,352	-34,428
11	2,1488E-05	27,5759	-10,8184	-48,797	1,7134	18,713
6	1,4175E-05	-42,5268	-6,44925	-35,429	10,617	31,335
10	3,0633E-05	-11,7608	23,0874	-47,165	39,7	49,102
9	1,2793E-05	41,6544	-49,3113	-0,4376	-22,477	42,877
5	4,0554E-05	7,07829	35,8608	-9,2617	-12,281	-19,701

Fuente: Elaboración propia.

Una vez que se ha completado la mutación se vuelve a evaluar los seis individuos que fueron mutados en la red neuronal correspondiente, para determinar si alguno de ellos genera por-

centajes de error menores a los dos mejores resultados que se tenía y se mantuvieron sin cambio.

La evaluación de la nueva generación se observa en tabla 15.

TABLA 15
EJEMPLO DE EVALUACIÓN DE LA NUEVA GENERACIÓN

Cromosoma	Error	FA	W1	W2	W3	W4	W5
12	8,60%	1,9885E-05	2,31769	4,07499	-43,579	5,1886	8,7996
2	9,63%	2,1713E-05	14,9468	-3,37171	-46,188	3,451	13,756
6	11,13%	1,4175E-05	-42,5268	-6,44925	-35,429	10,617	31,335
5	16,23%	4,0554E-05	7,07829	35,8608	-9,2617	-12,281	-19,701
9	22,01%	1,2793E-05	41,6544	-49,3113	-0,4376	-22,477	42,877
10	22,43%	3,0633E-05	-11,7608	23,0874	-47,165	39,7	49,102
3	24,95%	3,6472E-05	-29,2447	40,4086	-47,805	46,352	-34,428
11	28,74%	2,1488E-05	27,5759	-10,8184	-48,797	1,7134	18,713

Fuente: Elaboración propia.

La secuencia de tareas descrita anteriormente de selección, cruce, evaluación, mutación y re-evaluación, constituye una iteración del algoritmo, esta secuencia se debe seguir ejecutando de forma continua, de manera que cada vez se vayan obteniendo mejores resultados, hasta un punto en donde ya no se disminuya el error o se esté satisfecho con el resultado obtenido.

En el caso de las dos redes neuronales objeto de esta investigación, se aplicaron bloques de 45 iteraciones para evaluar hasta donde se optimizaba los resultados, siempre que estos procesos no implicaran un consumo excesivo de tiempo y esfuerzo que no estuvieran presentando una mejora significativa en los resultados.

RESULTADOS Y DISCUSIÓN

Como parte del desarrollo de la investigación, se diseñó una herramienta automatizada para efectuar las proyecciones tanto para la Red de casos nuevos, como para la de estancia.

Esta herramienta, que fue desarrollada en Microsoft Excel 2007, incluye los procesos necesarios en cada una de las tres fases del modelo de la red neuronal, tanto para el aprendizaje, con base en los datos históricos; como la optimización, a través del algoritmo genético; y la operación, donde se generan los pronósticos, basándose en los datos de las variables de entrada.

Entrenamiento inicial

La Red de casos nuevos, se entrenó con un total de 120 casos (71% de los disponibles). Para analizar la efectividad del entrenamiento se utilizaron 36 datos (21%) y el restante 8% se utilizó para comparar las proyecciones. El resultado del porcentaje de error medio absoluto (MAPE, por sus siglas en inglés) durante el entrena-

miento fue de 11,76% y después del entrenamiento el error fue de 7,53%

En el caso de la Red de estancia, para el entrenamiento se utilizaron los últimos 1000 casos disponibles. De estos, 700 (70%) se utilizaron para el entrenamiento, 150 (15%) para la evaluación y 150 (15%) para comparar las proyecciones. El resultado del error total durante el entrenamiento fue de 84,13% y después del entrenamiento el error fue de 47,37%.

Optimización con algoritmos genéticos.

Después de la optimización, el resultado del error para la Red de casos nuevos pasó de 7,53% a 5,07%. En el caso de la Red de estancia el resultado del error pasó de 47,37% a 40,79%. En ambos casos se logra apreciar la efectividad del proceso de optimización. Lo anterior debido a que se lograron reducciones importantes en los porcentajes de error de ambas redes.

Efectividad del Modelo

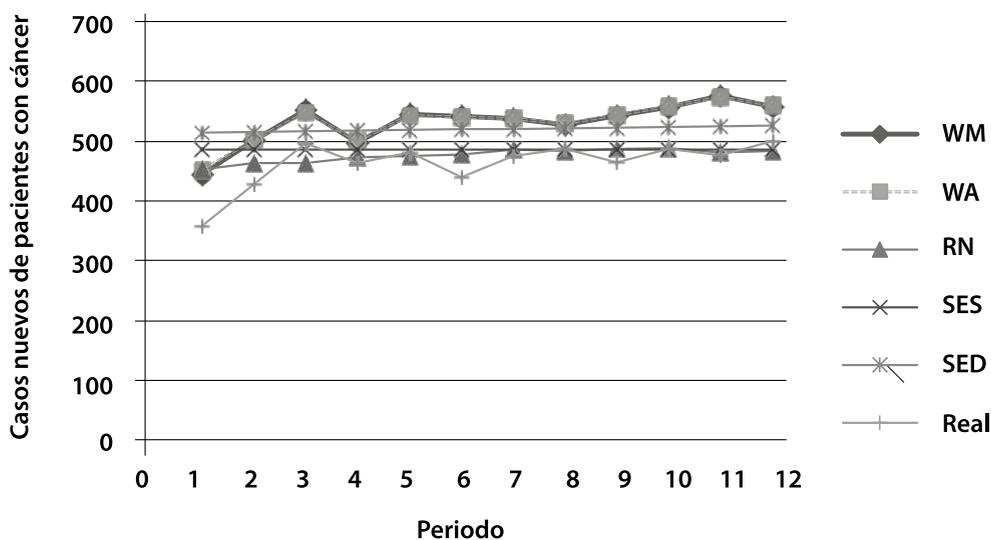
Para analizar la efectividad del modelo de red neuronal de casos nuevos, se generaron proyecciones para los siguientes 12 períodos y se compararon con las proyecciones obtenidas mediante cuatro modelos tradicionales de pronósticos: **z** exponencial simple, suavización exponencial doble, "winters" aditivo y "winters" multiplicativo. Con esto se determinó que el modelo de redes neuronales genera los mejores resultados en términos del error de pronóstico (5,07%). El más cercano de los otros métodos fue el Winters Aditivo con un error de 6,24%. Los resultados se muestran en la tabla 16 y figura 5.

Tabla 16
**COMPARACIÓN DE RESULTADOS DE PRONÓSTICOS GENERADOS POR LOS
 MODELOS TRADICIONALES Y EL MODELO DE REDES NEURONALES**

Período	1	2	3	4	5	6	7	8	9	10	11	12	Errores totales		
													MAPE	MAD	MSE
SES	481	481	481	481	481	481	481	481	481	481	481	481	7,53	32,42	1758
SED	510	511	513	514	515	516	517	518	519	520	521	523	9,57	41,19	2806
WM	440	497	549	493	541	538	535	524	539	554	573	556	6,34	27,93	1253
WA	447	498	545	498	540	537	535	527	541	555	570	557	6,24	27,39	1207
RN	447	458	459	468	471	473	481	479	482	484	477	479	5,07	22,9	818
Real	351	422	492	458	477	434	470	483	460	483	472	496			

Fuente: Elaboración propia.

Figura 5
**COMPARACIÓN DE RESULTADOS DE PRONÓSTICOS GENERADOS POR LOS MODELOS
 TRADICIONALES Y EL MODELO DE REDES NEURONALES**



Fuente: Elaboración propia.

Cabe destacar que aunque el porcentaje de diferencia es bajo, solo 1.1%, esto se puede traducir en unos 60 casos nuevos de pacientes hospitalizados con cáncer al año, lo cual ayudaría a realizar una mejor preparación de los recursos humanos y materiales necesarios para su atención. Adicionalmente, el costo de crear un modelo de esta naturaleza no difiere en forma importante con respecto a uno que utilice un método tradicional de pronóstico.

Horizonte de pronóstico

Finalmente, se evaluó el horizonte de pronóstico de la Red de casos nuevos analizando tres escenarios: un año, dos años, y tres años. Los resultados de la red neuronal son más efectivos que los obtenidos con los modelos tradicionales en los tres años analizados, donde por ejemplo en el primer año el MAPE de la red es de 5,63%, por debajo de 6,9% del SES que es el mejor de los tradicionales, y para el tercer año el error de la red neuronal es de 5,22%, siendo menor que el obtenido por los modelos tradicionales, cuyo mejor resultado fue un 7,2%. Los resultados se presentan en la tabla 17.

CONCLUSIONES

Como conclusión principal de la investigación se determina que la utilización de un modelo de redes neuronales, optimizadas con algoritmos genéticos, para proyectar la cantidad de casos nuevos de pacientes con cáncer hospitalizados en la CCSS, genera mejores resultados en términos del menor error de pronóstico, que los obtenidos utilizando algunas de las técnicas tradicionales de pronósticos.

El modelo desarrollado presenta mejores resultados que los métodos tradicionales considerando un horizonte de pronósticos de uno, dos o tres años. Después de este período la recomendación es hacer una actualización de datos de entrada de la red, así como un nuevo entrenamiento y optimización, ya que, los cambios en variables macroeconómicas y sociales incorporadas en el modelo se manifiestan con un rezago de tres años.

La construcción de un modelo de redes neuronales conlleva una comprensión profunda del fenómeno que se quiere representar, no solo desde el punto de vista matemático, sino desde el punto de vista lógico. Tal como se determinó a

TABLA 17
EVALUACIÓN DEL HORIZONTE DE PRONÓSTICOS DEL MODELO DE REDES NEURONALES EN COMPARACIÓN CON LOS MODELOS TRADICIONALES

Modelo	Un año			Dos años			Tres años		
	MAPE	MAD	MSE	MAPE	MAD	MSE	MAPE	MAD	MSE
SES	6,9%	27,8	1998	7,3%	65,7	3136	8,3%	114,9	6006
SED	13,6%	58,3	4706	5,9%	49,9	2938	13,1%	184,1	14280
WM	15,6%	70,1	5314	5,8%	51,6	2134	7,3%	103,1	5170
WA	15,8%	70,9	5417	5,8%	51,2	2116	7,2%	102,2	5084
RN	5,63%	25,33	941	5,39%	25,1	848	5,22%	21,8	1036

Fuente: Elaboración propia.

través del trabajo realizado, la determinación del tipo de red, la cantidad de capas y del número de neuronas en la capa oculta, son elementos clave para la generación de resultados aceptables; no obstante, aunque estos elementos algunas veces se determinan mediante prueba y error, hoy día se tienen a disposición programas informáticos estadísticos, que contienen módulos para modelar redes neuronales, que pueden utilizarse como guía para establecer los elementos en cuestión.

La utilización de técnicas avanzadas de predicción y optimización, como lo son las redes neuronales y los algoritmos genéticos, permiten analizar otros factores que intervienen en los fenómenos en estudio, establecer relaciones que no parecen muy claras al entendimiento y razonabilidad matemática y lógica; sin embargo, como se pudo comprobar en el proceso de investigación realizado, a pesar de que se pueden crear estos modelos, la tecnología disponible para el manejo de los mismos, es aún limitada, por lo que se debe restringir a modelos simples y con poco manejo de datos; lo cual con el desarrollo de las tecnologías y los programas computacionales, permitirá modificarse y avanzar en su uso y aprovechamiento en todos los sectores organizacionales.

REFERENCIAS

- Calvo, A. (2004). *Un modelo de redes neuronales aplicado al Sistema Bancario Costarricense: período 1998-2002*. Tesis de Maestría. Universidad de Costa Rica, San José, Costa Rica.
- CCSS (2010). *Base de datos de egresos hospitalarios*. Área de Estadísticas en Salud, San José, Costa Rica.
- Gallardo, E. (2009). *Modelo de redes neuronales artificiales aplicado a un problema de clasificación de estudiantes de secundaria según actitud hacia la aculturación*. Tesis de Maestría. Universidad de Costa Rica, San José, Costa Rica.
- Grajales, T. (1996). *Conceptos Básicos para la Investigación Social*. Nuevo León, México: Publicaciones Universidad de Monterrey.
- Herrera, F. Lozano, M. y Verdegay, J. (1994). *Algoritmos Genéticos: Fundamentos, Extensiones y Aplicaciones*. ETS de Ingeniería Informática, Universidad de Granada, España. Consultado en la dirección ftp://decsai.ugr.es/pub/arai/tech_rep/lozano/esp/arbor.ps.Z
- Hilera, J. y Martínez, V. (2000). *Redes neuronales artificiales: fundamentos, modelos y aplicaciones*. México: Alfaomega.
- Moujahid, A., Inza, I. y Larrañaga, P. (s.f.). *Algoritmos Genéticos*. Departamento de Ciencias de la Computación e Inteligencia Artificial. Universidad del País Vasco. Consultado desde la dirección electrónica: <http://www.sc.ehu.es/ccwbayes/docencia/mmcc/docs/t2geneticos.pdf>
- Nojek, S., Britos, P., Rossi, B. y García Martínez, R. (2003). Pronóstico de Ventas: Comparación de Predicción basada en Redes Neuronales versus Método Estadístico. *Reportes Técnicos en Ingeniería del Software*. Vol 5 (1): pp. 1-12. Instituto Tecnológico de Buenos Aires-Argentina.
- Orozco, C. y Medina, R. (2004). *Modelo de la evolución del conocimiento en individuos en un sistema autorregulado, basado en redes neuronales artificiales y algoritmos genéticos*. Tesis para optar por el título de Ingeniero en Sistemas. Universidad EAFI. Colombia. Disponible en la dirección <http://bdigital.eafit.edu.co/bdigital/PROYECTO/P006.32CDO747/capitulo1.pdf>
- Pacheco, J. Sancho, J y Zoch, A. (2006). *Rediseño de los procesos de administración de inventarios y almacenaje en la empresa Café El Rey*. Tesis para optar por el grado de Licenciatura en Ingeniería Industrial. Universidad de Costa Rica, San José, Costa Rica.
- Solano, Y. (1997). *Improvement of convergence speed and generalization capability in artificial neural networks by normalizing the Back-propagation algorithm*. Chiba, Japan: s.n.e.
- Sotolongo, G., Guzmán (2001). MV. Aplicaciones de las redes neuronales. El caso de la bibliometría. *Ciencias de la Información*. Vol 32(1): pp. 27-34. Consultado desde la dirección electrónica: <http://www.idict.cu/cinfo99/v32n1a2001/bibliometria.htm>
- Toro, O. Mejía, D. Salazar, H. (2004). Pronóstico de ventas usando redes neuronales. *Scientia Et Technica*. Vol X (26): pp. 25-30. Universidad Tecnológica de Pereira-Colombia.
- Valenzuela, M (2008). *Logística Inteligente*. Centro de Computación Inteligente y Robótica Tecnológico de Monterrey, Campus Monterrey. Recuperado de http://strauss.ing.puc.cl/laccir/Portals/0/research_teams/itesm/optimization_and_logistics.pdf

Recibido: 23 de setiembre de 2013

Aceptado: 30 de octubre de 2013