

**TAMIZACIÓN DE PACIENTES CON SOSPECHA DE CÁNCER DE MAMA
MEDIANTE REDES NEURONALES ARTIFICIALES UTILIZANDO VARIABLES
CLÍNICAS**

ELIANA XIMENA GONZÁLEZ MORALES

**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FÍSICO-MECÁNICAS
ESCUELA DE INGENIERÍAS ELÉCTRICA, ELECTRÓNICA Y
TELECOMUNICACIONES
BUCARAMANGA
2013**

**TAMIZACIÓN DE PACIENTES CON SOSPECHA DE CÁNCER DE MAMA
MEDIANTE REDES NEURONALES ARTIFICIALES UTILIZANDO VARIABLES
CLÍNICAS**

ELIANA XIMENA GONZÁLEZ MORALES

**Trabajo de grado para optar al título de
Ingeniera Electrónica**

Director

PhD. DANIEL ALFONSO SIERRA BUENO

Codirectora

MD. OLGA MERCEDES ÁLVAREZ OJEDA

**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FÍSICO-MECÁNICAS
ESCUELA DE INGENIERÍAS ELÉCTRICA, ELECTRÓNICA Y
TELECOMUNICACIONES
BUCARAMANGA**

2013

A Dios, mi guía e inspiración para la realización de este proyecto.

A mis padres Elizabeth y Carlos S., por su amor, dedicación y apoyo incondicional.

A mi hermana Carolina y mi cuñado Elkin, por brindarme sus consejos y motivación.

A mis familiares y amigos que siempre me apoyaron.

AGRADECIMIENTOS

Agradecimientos especiales al Doctor Daniel Alfonso Sierra Bueno, director de este proyecto, y a la Doctora Olga Mercedes Álvarez Ojeda, codirectora, por sus excelentes aportes y colaboración que permitieron el desarrollo de este trabajo

Agradecimiento a la Dra. Martha Lucía Silva Mora, por su asesoría y aportes en el área de epidemiología.

Agradecimientos al doctor Óscar Gualdrón González, por su colaboración en la parte inicial de este proyecto.

Agradecimientos a la Universidad Industrial de Santander y a la Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones, por mi formación como ingeniera.

TABLA DE CONTENIDO

INTRODUCCIÓN	14
1. MARCO CONCEPTUAL	15
1.1. Aplicación de la inteligencia artificial en el diagnóstico médico	15
1.2. Cáncer de mama y factores de riesgo	15
1.3. Redes Neuronales Artificiales	17
1.4. Métodos Bioestadísticos	21
1.5. Medidas de validez de las pruebas diagnósticas	22
1.6. Validación cruzada de k iteraciones	23
2. METODOLOGÍA	23
2.1. Adquisición de datos	23
2.2. Variables de entrada y salida	24
2.3. Preselección de covariables	24
2.4. Selección de las covariables	24
2.5. Construcción de la RNA	24
3. RESULTADOS	25
3.1. Determinación de entradas y salida de la RNA	25
3.2. Implementación de las redes	26
3.3. Validación cruzada de k iteraciones	27
4. OBSERVACIONES Y CONCLUSIONES	27
REFERENCIAS	28
ANEXOS	30

LISTA DE FIGURAS

Figura 1. Vista lateral del seno.	16
Figura 2. Modelo de una neurona.	17
Figura 3. Ejemplos de arquitecturas neuronales.	18
Figura 4. Arquitectura Perceptrón Multicapa.	19
Figura 5. Fases del algoritmo de aprendizaje BP.	19
Figura 6. Ejemplo del comportamiento del error durante el entrenamiento	20
Figura 7. Arquitectura red RBF.	20
Figura 8. Ejemplo de curva ROC	23
Figura 9. Metodología	24

LISTA DE TABLAS

Tabla 1 Factores de riesgo para el cáncer de mama	16
Tabla 2. Relación prueba diagnóstica / enfermedad	22
Tabla 3. Resultado de la Regresión Logística Binaria	26
Tabla 4 Resumen de resultados de la simulación de las redes BP	26
Tabla 5 Resumen de resultados de la simulación de RBFN	27
Tabla 6 Resultado de la Validación cruzada de 5 iteraciones	27

LISTA DE ANEXOS

Anexo A. FORMATOS DE RECOLECCIÓN DE LA INFORMACIÓN	30
Anexo B. RESULTADO DE LA SIMULACIÓN REDES BACKPROPAGATION	39
Anexo C. RESULTADO DE LA SIMULACIÓN REDES RBFN	43

GLOSARIO

Área bajo la curva ROC (AUC): estima la capacidad de una prueba diagnóstica de discriminar entre enfermos y no enfermos.

Especificidad: capacidad de detectar correctamente los individuos que no presentan enfermedad.

Factor de riesgo: es cualquier característica o exposición de un individuo que aumenta la probabilidad de que presente una enfermedad.

Falsos negativos (FN): individuos que se encuentran enfermos y la prueba los clasifica como no enfermos.

Falsos Positivos (FP): individuos no enfermos y la prueba los clasifica como enfermos.

Generalización: capacidad de encontrar respuestas razonables a patrones que no fueron utilizados durante el proceso de entrenamiento de una red neuronal artificial.

Incidencia: es el número de veces que ocurre un evento en una población dada.

Redes Neuronales Artificiales (RNA): son modelos cuyo objetivo es emular las habilidades del cerebro al ejecutar tareas específicas a partir de un conjunto de patrones de aprendizaje.

Sensibilidad: capacidad de clasificar correctamente a los individuos enfermos.

Tamización: prueba diagnóstica aplicada con el fin de identificar una población, aparentemente sana, en mayor riesgo de tener una determinada enfermedad, que hasta ese momento no se les ha diagnosticado.

Verdaderos Negativos (VN): individuos no enfermos y la prueba los clasifica como correctamente.

Verdaderos Positivos (VP): individuos que se encuentran enfermos y la prueba los clasifica correctamente.

RESUMEN

TÍTULO: TAMIZACIÓN DE PACIENTES CON SOSPECHA DE CÁNCER DE MAMA MEDIANTE REDES NEURONALES ARTIFICIALES UTILIZANDO VARIABLES CLÍNICAS*

AUTORA: Eliana Ximena González Morales**

PALABRAS CLAVES: Cáncer de mama, Factores de Riesgo, Red neuronal artificial (RNA), *Backpropagation*, RBFN, Validación cruzada

CONTENIDO:

El cáncer de mama es uno de los tumores más frecuentes en el mundo y ha presentado un incremento en las tasas de incidencia y mortalidad. Estudios realizados indican que estas cifras se podrían reducir mediante programas de tamización para detección temprana. Estos resultados han promovido la búsqueda de técnicas alternativas y complementarias como sistemas de Inteligencia Artificial, los cuales han demostrado mejorar la precisión de la detección y diagnóstico del cáncer. En este trabajo se utilizaron las Redes Neuronales Artificiales (RNA) en la tamización de 120 pacientes con sospecha de cáncer de mama utilizando variables clínicas. Se realizó un análisis de regresión logística binaria (LR) para determinar los factores de riesgo asociados al cáncer de mama. Se crearon dos modelos de patrones de entrada a las redes, el primero conformado por los resultados obtenidos en la LR y el segundo fue la combinación del modelo 1 y otros factores de riesgo potencial. Se utilizaron las redes *Backpropagation* (BP) y Redes de Función de Base Radial (RBFN) y se obtuvieron diversos tipos de configuraciones variando el algoritmo de aprendizaje, la función de transferencia y número de neuronas ocultas. El funcionamiento de las redes se evaluó usando la sensibilidad, especificidad y área bajo la curva ROC (AUC). Se seleccionaron las configuraciones que presentaron en la simulación valor de AUC igual o superior a 0.8. Posteriormente se les practicó la validación cruzada de 5 iteraciones y se encontró que la red que mostró mejor desempeño e independencia ante los datos fue la RNA-9 con una sensibilidad del 71.78%, especificidad del 78.29% y AUC de 0.7503. Es una red tipo BP con algoritmo de aprendizaje *Levenberg-Marquardt* y función de transferencia sigmoidea tangente hiperbólica en la capa oculta y de salida.

* Proyecto de grado

** Facultad de Ingenierías Físico-Mecánicas. Escuela de Ingenierías Eléctrica, Electrónica y Telecomunicaciones.
Director: PhD. Daniel Alfonso Sierra Bueno. Codirectora: Md. Olga Mercedes Álvarez Ojeda

ABSTRACT

TITLE: SELECTIONS OF PATIENS WITH SOSPECTED BREAST CANCER USING ARTIFICIAL NEURAL NETWORKS AND CLINICAL VARIABLES*

AUTHOR: Eliana Ximena González Morales**

KEYWORDS: Breast cancer, Risk Factors, Artificial Neural Network (ANN), Backpropagation, RBFN, cross-validation

CONTENT:

Breast cancer is one of the most common tumors in the world and has shown an increase in incidence and mortality rates. Studies indicate that these figures could be reduced by screening programs for early detection. These results have prompted the search for alternative and complementary techniques such as artificial intelligence systems, which have been shown to improve the accuracy of detection and diagnosis of cancer. In this work using Artificial Neural Networks (ANN) in the screening of 120 patients with suspected breast cancer using clinical variables. Was performed a binary logistic regression analysis (LR) to determine the risk factors associated with breast cancer. Two models were created input patterns to the network, the first consisting of the results of the LR and the second was the combination of model 1 and other potential risk factors. Backpropagation (BP) and Networks Radial Basis Function (RBFN) are used, and various types of configurations obtained by varying the learning algorithm, the transfer function and the number of hidden neurons. The network performance was evaluated using sensitivity, specificity and receiver operating characteristics area (AUC). The models were selected in the simulation had AUC value equal to or greater than 0.8. Subsequently we performed cross-validation 5 iterations and found that the network showed better performance and independence before the data was the RNA-9 with a sensitivity of 71.78%, specificity 78.29% and AUC of 0.7503. It is a type BP network with learning algorithm Levenberg-Marquardt and sigmoid transfer function hyperbolic tangent in the hidden and output layer.

* Minor Degree Project

** Physics Mechanical Engineering Faculty. Electric. Electronic and Telecommunications School. Director: PhD. Daniel Alfonso Sierra Bueno, Co-Director: Md. Olga Mercedes Álvarez Ojeda

Tamización de Pacientes con Sospecha de Cáncer de Mama Mediante Redes Neuronales Artificiales Utilizando Variables Clínicas

González Morales, Eliana Ximena

Resumen— El cáncer de mama es uno de los tumores más frecuentes en el mundo y ha presentado un incremento en las tasas de incidencia y mortalidad. Estudios realizados indican que estas cifras se podrían reducir mediante programas de tamización para detección temprana. Estos resultados han promovido la búsqueda de técnicas alternativas y complementarias como sistemas de Inteligencia Artificial, los cuales han demostrado mejorar la precisión de la detección y diagnóstico del cáncer. En este trabajo se utilizaron las Redes Neuronales Artificiales (RNA) en la tamización de 120 pacientes con sospecha de cáncer de mama utilizando variables clínicas. Se realizó un análisis de regresión logística binaria (LR) para determinar los factores de riesgo asociados al cáncer de mama. Se crearon dos modelos de patrones de entrada a las redes, el primero conformado por los resultados obtenidos en la LR y el segundo adicionando otros factores de riesgo potenciales. Se utilizaron las redes *Backpropagation* (BP) y Redes de Función de Base Radial (RBFN) y se obtuvieron diversos tipos de configuraciones variando el algoritmo de aprendizaje, la función de transferencia y número de neuronas ocultas. El funcionamiento de las redes se evaluó usando la sensibilidad, especificidad y área bajo la curva ROC¹ (AUC). Se seleccionaron las configuraciones que presentaron en la simulación valor de AUC igual o superior a 0.8. Posteriormente se les practicó la validación cruzada de 5 iteraciones y se encontró que la red que mostró mejor desempeño e independencia ante los datos fue la RNA-9 con una sensibilidad del 71.78%, especificidad del 78.29% y AUC de 0.7503. Es una red tipo BP con algoritmo de aprendizaje *Levenberg-Marquardt* y función de transferencia sigmoidea tangente hiperbólica en la capa oculta y de salida.

Palabras clave— Cáncer de mama, Biodatos, Factores de Riesgo, Redes Neuronales Artificiales (RNA), Curva ROC

INTRODUCCIÓN

El cáncer de mama es el tumor maligno más frecuente en las mujeres con más de 1 millón de nuevos casos anuales en el mundo. En Colombia, las tasas de incidencia y mortalidad se han incrementado, con un diagnóstico cercano a 7000 casos nuevos y más de 2000 muertes por año [1].

Se ha demostrado que la detección temprana a través de los programas de tamización ha reducido la mortalidad por este tipo de cáncer [2] [3] [4] [5]. Sin embargo, en países en vías de desarrollo, se observa una tendencia contraria debido a la dificultad en el acceso a los servicios de salud de la población y falta de implementación de esta clase de programas [6] [7].

Técnicas de Inteligencia Artificial como las Redes Neuronales Artificiales (RNA) se han venido utilizando de manera exitosa en aplicaciones médicas, y sirven de apoyo a los profesionales del área de la salud, en el diagnóstico efectivo de diferentes tipos de enfermedades [8] [9]. Debido a que este tipo de técnicas son aplicaciones tipo software, que manejan información como los datos de la historia clínica, no son invasivas, no causan dolor, ni exponen a radiación al paciente, además de ser una aplicación de bajo costo.

El presente trabajo de investigación propone la selección y validación de un modelo basado en RNA utilizando variables clínicas, que permita la tamización de pacientes sospechosos de cáncer de mama y sirva de ayuda para el diagnóstico objetivo de los médicos.

Se utilizaron las redes neuronales *Backpropagation* (BPN) y Redes de Función de Base Radial (RBFN) con diferentes tipos de combinaciones de algoritmo de aprendizaje, función de transferencia y número de neuronas ocultas. Posteriormente con el fin de evaluar la capacidad discriminatoria de cada una de las redes se realizó una validación mediante el cálculo del área bajo la curva ROC (AUC). Para finalizar se presentan los resultados y las conclusiones de este trabajo.

¹ Receiver Operating Characteristic

1. MARCO CONCEPTUAL

1.1. Aplicación de la inteligencia artificial en el diagnóstico médico

El incremento de los conocimientos y el desarrollo de nuevas técnicas diagnósticas ha hecho cada vez más compleja la toma de decisiones por parte de los médicos tratantes, quienes han continuado realizando el análisis de la información de forma manual y de manera subjetiva. Por esto el uso de sistemas informáticos se ha convertido en una de las áreas de gran interés en la medicina.

La informática médica basada en Inteligencia Artificial (IA) tuvo sus inicios a finales de los años 60 [10]. Los primeros sistemas médicos fueron desarrollados por universidades de Estados Unidos, los cuales mostraban la posibilidad de realizar diagnósticos de diferentes enfermedades: En *Rutgers University* se diseñó CASNET² para el diagnóstico y tratamiento de enfermedades. A mediados de 1970, en Standford se desarrolló MYCIN, para diagnosticar infecciones microbianas y dar una recomendación de tratamiento con medicamentos. El *Massachusetts Institute of Technology* diseñó en 1970, PIP³, un simulador para clasificación de la historia médica de un paciente dentro de un grupo de enfermedades renales conocidas. En 1980 se desarrolla en Pittsburg, INTERNIST, para el diagnóstico de medicina interna a través de la investigación de métodos heurísticos, utilizando el diagnóstico diferencial en la toma de decisiones clínicas [8].

Posteriormente en los años 90 aparecieron nuevos modelos, entre ellos las RNA, las cuales han tenido gran éxito en el desarrollo de ayudas médicas. Se destacan aplicaciones sobre tratamientos de la arteria coronaria, infartos miocárdicos, diagnóstico de cáncer, neumonía y desórdenes cerebrales, entre otros [10].

Para el diagnóstico de cáncer de mama se han estudiado e implementado diferentes tipos de aplicaciones que han servido de complemento a métodos tradicionales como la mamografía y la termografía, los cuales han mostrado resultados interesantes.

En 1992, Wu et al. [11] aplicaron RNA a los resultados de mamografías digitales para detectar microcalcificaciones

agrupadas y obtuvieron una reducción del 50% en falsos positivos manteniendo el 95% de verdaderos positivos.

Lo J. Y. et al. [12] evaluaron la contribución de variables de la historia médica a la predicción de cáncer de mama con RNA y resultados de mamografías, encontrando que para una especificidad dada de 98%, adicionar la edad a la RNA entrenada con los hallazgos mamográficos, aumentaba la sensibilidad de 39% a 42%. Para esta misma especificidad la impresión del radiólogo había obtenido una sensibilidad sólo del 12%

Ng. Et al. [13] evaluaron una técnica que integra termografía, RNA y métodos bio-estadísticos, la cual obtuvo una tasa de precisión de 80.95%, con 100% de sensibilidad y 70.6% de especificidad en la identificación de cáncer de mama. Lo resultados son exitosos al compararlos con el examen clínico que obtuvo una tasa de precisión entre el 60% y el 70%.

En el análisis de termografía de mama con RNA, Koay J. et al. [14], demuestran que una RNA *backpropagation* fue capaz de generar una salida bastante precisa, al ser entrenada con parámetros estadísticos del termograma.

1.2. Cáncer de mama y factores de riesgo

El cáncer de mama es una enfermedad causada por el crecimiento descontrolado de células anormales debido a alteraciones en el ADN. La mayoría de las veces puede formar un tumor o irse extendiendo a diferentes partes del cuerpo.

Se origina en el tejido del seno (Figura 1) principalmente en los ácinos de los lobulillos mamarios y menos frecuente en tejido adiposo, conectivo y vasos sanguíneos y linfáticos [15].

Diferentes estudios realizados [5] [6] [15] han encontrado que existen factores de riesgo que aumentan la probabilidad de desarrollar cáncer de seno (Tabla 1), entre los cuales se encuentran

- *Género*: Es 100 veces más frecuente en mujeres que en hombres [15].
- *Edad*: El riesgo de desarrollar cáncer de mama aumenta con la edad. Tienen mayor riesgo las mujeres con edad alrededor de los 50 años [5] [15].

² Causal Associational Networks

³ Present Illmes Program

GONZÁLEZ: TAMIZACIÓN DE PACIENTES CON SOSPECHA DE CÁNCER DE MAMA MEDIANTE REDES NEURONALES ARTIFICIALES UTILIZANDO VARIABLES CLÍNICAS

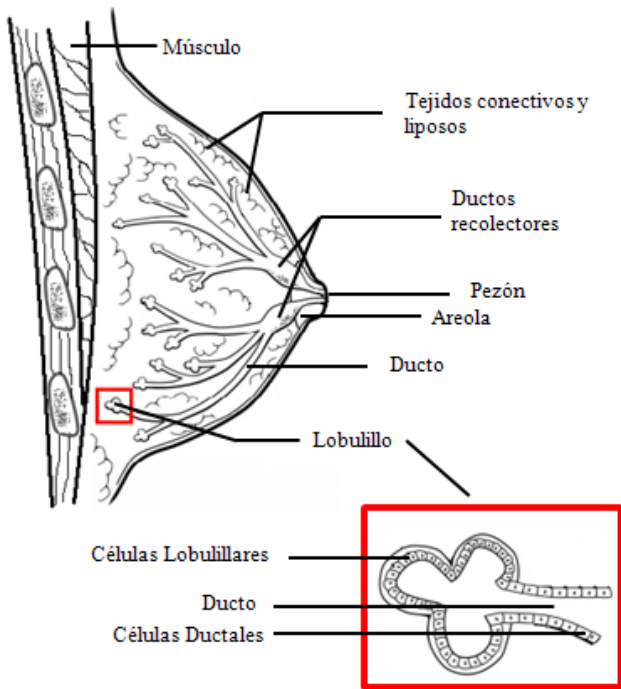


Figura 1. Vista lateral del seno.
Adaptado de [15]

Tabla 1 Factores de riesgo para el cáncer de mama

Factor	Riesgo Relativo	Grupo de riesgo
Género	>100	Mujeres
Edad	>10	Mayor o igual a 50 años
Genético	5-10	Alguno de los padres con mutación de los genes BRCA1 ^a y BRCA2 ^b
Historia familiar de cáncer de seno	>2	Cáncer de seno en familiares de primer grado de consanguinidad.
Antecedentes personales de cáncer de seno	3-4	
Previa enfermedad benigna de seno	4-5	Hiperplasia atípica
Menarquia	3	Antes de los 12 años
Menopausia	2	Después de los 55 años
Edad del primer embarazo	3	Embarazo a término después de los 30 años
Número de hijos	-	Por cada parto se disminuye el riesgo en 7%.
Anticonceptivos orales	1.24	Uso actual
Terapia de reemplazo hormonal	1.35	Uso mayor a 10 años
Lactancia	-	Es un factor protector. Disminuye el riesgo en 3.4% por cada 12 meses de lactancia.
Sobrepeso u obesidad	2	Índice de Masa Corporal > 35

^a BRCA1: Breast Cancer Type 1

^b BRCA2: Breast Cancer Type 2

- **Factores de riesgo genéticos:** Del 5 al 10 por ciento de las personas con cáncer de seno presentan cambios genéticos hereditarios [6] [15].
- **Historia familiar:** Tener familiares de primer grado de consanguinidad (madre, hermanas o hijas) con cáncer de seno, aumenta el riesgo de padecerlo [5] [6] [15] [16].
- **Antecedentes personales de cáncer de seno** [5] [6] [15]
- **Antecedentes personales de problemas benignos de seno:** algunos de estos problemas pueden convertirse en cáncer de seno, aumentando el riesgo de desarrollar esta enfermedad [5] [6] [15].
- **Edad de la menarquia y la menopausia:** Tienen mayor riesgo de tener cáncer de seno las mujeres que empezaron a menstruar antes de los 12 años y que tuvieron la menopausia después de los 55 años. Según la *American Cancer Society* en [15] esto se puede deber a que la mujer tendría mayor tiempo de exposición hormonal [5] [6] [15].
- **Numero de hijos:** A mayor número de hijos disminuye el riesgo de tener cáncer de seno [6] [16].
- **Edad del primer embarazo:** Las mujeres que no han tenido hijos o tuvieron el primero después de los 30 años tienen mayor probabilidad de desarrollar cáncer de seno [5] [15].
- **Anticonceptivos orales:** algunos estudios indican que las mujeres que toman anticonceptivos orales tienen un riesgo levemente mayor de desarrollar cáncer de seno a aquellas que no los han usado, pero que este desaparece al dejar de usarlas [5] [15] [16].
- **Terapia de reemplazo hormonal:** el uso de este tipo de tratamientos para disminuir los efectos producidos por la menopausia logra incrementar el riesgo de padecer cáncer de seno [5] [6] [15].

GONZÁLEZ: TAMIZACIÓN DE PACIENTES CON SOSPECHA DE CÁNCER DE MAMA MEDIANTE REDES NEURONALES ARTIFICIALES UTILIZANDO VARIABLES CLÍNICAS

- *Lactancia*: Un periodo de lactancia comprendido entre 1.5 y 2 años puede disminuir el riesgo del cáncer de seno [6] [15].
- *Sobrepeso u obesidad*: Incrementa el riesgo de tener cáncer de seno especialmente si el aumento de peso ocurre después de la menopausia [5] [6] [15] [16].

1.3. Redes Neuronales Artificiales

Las Redes Neuronales Artificiales (RNA) son modelos cuyo objetivo es emular las habilidades del cerebro al ejecutar tareas específicas a partir de un conjunto de patrones de aprendizaje.

Entre las propiedades de las redes neuronales biológicas deseadas en las RNA se destacan: la generalización, habilidad de adaptación, tolerancia a fallos y procesamiento paralelo, entre otras.

Para las RNA o para los sistemas de inteligencia artificial, la generalización es la capacidad de encontrar respuestas razonables a patrones que no fueron utilizados durante el proceso de aprendizaje [17]. Si el sistema resuelve una nueva situación se dice que la red presenta una buena generalización.

Las RNA están compuestas por unidades simples de procesamiento o neuronas que se encuentran interconectadas entre sí para la transmisión de la información.

Las neuronas dentro de la red se encuentran organizadas en grupos o capas, las cuales junto con señales de entrada y la salida forman una arquitectura de red.

La conexión entre dos neuronas se denomina *sinapsis*. La intensidad de una sinapsis representa el conocimiento, por lo tanto es variable y se asocia a un peso sináptico.

El funcionamiento del cerebro se puede modelar a través de la creación de nuevas conexiones o ruptura de estas, eliminación de neuronas o la modificación de las intensidades sinápticas. Este último método es el más utilizado en las RNA y es conocido como *plasticidad* [17].

En las RNA el conocimiento se adquiere a través de un algoritmo de aprendizaje el cual lo logra generalmente a través de la modificación de los pesos sinápticos.

Modelo general de una neurona artificial

En la Figura 2 se presenta un modelo de neurona estándar la cual está conformada por [17] [18] :

a. *Entradas*: x_i

b. *Conexiones o Sinapsis*: asociadas a un peso sináptico, el cual indica la intensidad de interacción entre la neurona presináptica i y la neurona postsináptica j . Dependiendo del signo, el peso sináptico es excitatorio cuando es positivo, inhibitorio cuando es negativo, o tomar valor 0 cuando no existe conexión.

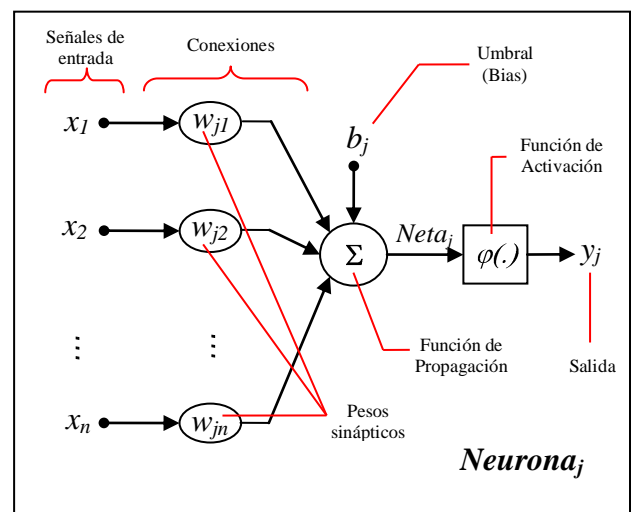


Figura 2. Modelo de una neurona. Adaptado de [17]

c. *Función o regla de propagación*: Calcula el valor del potencial postsináptico de la neurona conocido como $Neta_j$. Por lo general se utiliza como función de propagación, la suma ponderada de las señales de entrada por sus respectivos pesos sinápticos, incluyendo su umbral o *Bias*, denotado por b_i , si existiera. Se puede describir como:

$$Neta_j = \sum_i^n w_{ji} x_i + b_i \quad (1)$$

En donde x_i son las señales de entrada provenientes de otras neuronas, w_{ji} son los pesos sinápticos de la neurona j , b_i es el umbral y n el número de señales de entrada.

d. *Función de Activación*: Representa el estado actual de la neurona j .

$$y_j = \varphi(Neta_j) \quad (2)$$

GONZÁLEZ: TAMIZACIÓN DE PACIENTES CON SOSPECHA DE CÁNCER DE MAMA MEDIANTE REDES NEURONALES ARTIFICIALES UTILIZANDO VARIABLES CLÍNICAS

Las funciones de activación comúnmente utilizadas en aplicaciones prácticas son: función Identidad, Escalón, Lineal a tramos, Sigmoidea, Gaussiana, entre otras.

e. *Función de salida*: Ocasionalmente se tiene en cuenta cuando ésta es diferente a la función identidad, la cual es la utilizada con mayor frecuencia.

Arquitecturas de las redes neuronales

La arquitectura es la disposición de las conexiones entre neuronas en una red. La agrupación de neuronas del mismo tipo se denomina capa y una o más capas en conjunto con las entradas y la salida forma un sistema de red neuronal artificial [18].

Se distinguen dos tipos de capas: capas ocultas o intermedias y capa de salida. Algunos autores denominan también como “capa” al grupo de patrones de entrada a la red, pero al nombrar la estructura de red no se toma en cuenta como tal debido a que en esta parte de la arquitectura no se realiza ningún tipo de cómputo [17]. Por ejemplo, una red de tres capas estaría conformada por dos capas ocultas y una capa de salida. En este trabajo se hará referencia a la capa de entrada como nodos de entrada o simplemente entradas para evitar este tipo de confusiones.

Las redes neuronales se pueden clasificar en [17]:

- *Redes monocapa*: formadas por los nodos de entrada y una capa de neuronas de salida. Su flujo de señales es unidireccional (*feedforward*) (Figura 3a).
- *Redes Unidireccionales Multicapa*: formadas por los nodos de entrada, una o más capas de neuronas ocultas o intermedias y la capa de neuronas de salida (Figura 3b).
- *Redes recurrentes o feedback*: son redes multicapa que tienen por lo menos una conexión de realimentación. El flujo de señal se puede dar en cualquier sentido, incluido el de entrada-salida. También puede existir realimentación de la misma neurona. En la Figura 3c se puede observar un ejemplo de red recurrente de una sola capa.

Tipos de aprendizaje

El proceso de aprendizaje se puede dar por la creación o variación de los pesos sinápticos, modificación de la topología al eliminar neuronas o cambiar el número de conexiones.

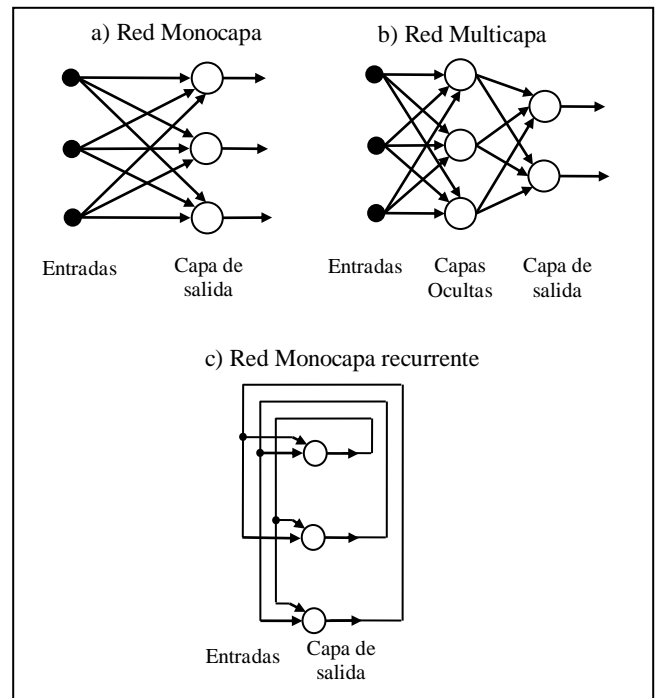


Figura 3. Ejemplos de arquitecturas neuronales.
Adaptado de [17]

Entre los algoritmos de aprendizaje más conocidos se encuentran [18]:

- a. *Aprendizaje supervisado*: consiste en ingresar un grupo de patrones a la red junto a la salida deseada. La red ajusta los pesos de las neuronas, corrigiendo el error, de forma que al presentarle posteriormente esos patrones los relacione con la salida memorizada. Se encuentran en este grupo el Perceptrón, Adalina, *Backpropagation*, entre otros.
- b. *Aprendizaje no supervisado o autoorganizado*: se presentan los patrones a la red y ésta los agrupa de acuerdo a rasgos similares. Por ejemplo los Mapas de Kohonen y *Neocognitrón*.
- c. *Aprendizaje híbrido*: Combina el aprendizaje supervisado y el autoorganizado, los cuales se presentan en diferentes capas de neuronas. Entre ellos se destaca *Radial Basis Function* (RBF).
- d. *Aprendizaje reforzado*: Se encuentra entre el aprendizaje supervisado y el autoorganizado. Determina un error que indica el rendimiento global de la red pero no se le proporciona la salida deseada. Ejemplo de este tipo de aprendizaje es el Premio-castigo asociativo.

GONZÁLEZ: TAMIZACIÓN DE PACIENTES CON SOSPECHA DE CÁNCER DE MAMA MEDIANTE REDES NEURONALES ARTIFICIALES UTILIZANDO VARIABLES CLÍNICAS

Redes propuestas para el sistema de clasificación

Red Neuronal Backpropagation

Se conoce como red neuronal *Backpropagation* (BPN) al conjunto formado por la arquitectura Perceptrón Multicapa (MLP), y el algoritmo de aprendizaje *Backpropagation* (BP) [18].

Es una red multicapa unidireccional en la cual las neuronas de una capa proporcionan la entrada a las neuronas de la siguiente. En una misma capa las neuronas no se interconectan entre si. La Figura 4 muestra un ejemplo de un MPL de una capa oculta y una capa de salida:

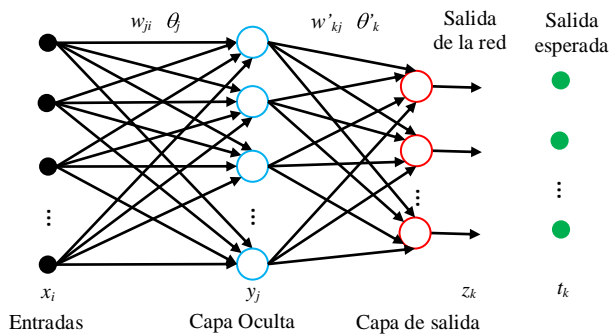


Figura 4. Arquitectura Perceptrón Multicapa. Adaptado de [18]

Al ingresar a una capa, cada entrada es ponderada dentro de la neurona por un peso ajustable. Posteriormente todas las entradas son sumadas y transformadas por una función de activación para obtener la salida.

Una de las arquitecturas del MPL más utilizadas se puede expresar de la siguiente manera [18]

$$z_k = g\left(\sum_j w'_{kj} y_j - \theta'_k\right) = g\left(\sum_j w'_{kj} f\left(\sum_i w_{ji} x_i - \theta_j\right) - \theta'_k\right) \quad (3)$$

En donde x_i son las entradas de la red, y_j las salidas de la capa oculta, z_k las salidas de la capa final y t_k serán las salidas objetivo o esperadas; Los pesos de la capa oculta son w_{ji} y sus umbrales θ_j , y los pesos de la capa de salida son w'_{kj} y sus umbrales θ'_k . La función de activación de la capa oculta f de tipo sigmoidea puede ser

$$f(x) = \frac{1}{1 + e^{-x}} \quad f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (4)$$

La función de activación de la capa de salida g puede ser sigmoidea pero por lo general se considera la función identidad.

El algoritmo de aprendizaje BP es un tipo de aprendizaje supervisado, que utiliza una función de error o coste para medir el rendimiento actual de la red. Un método de optimización buscará llevar este error al mínimo a través de la actualización iterativa de los pesos sinápticos.

Generalmente se utiliza como función de coste el error cuadrático medio (MSE)

$$E(w_{ji}, \theta_j, w'_{kj}, \theta'_k) = \frac{1}{2} \sum_{\mu} \sum_k [t_k^{\mu} - g(w'_{kj} y_j^{\mu} - \theta'_k)]^2 \quad (5)$$

Donde t_k^{μ} son los resultados esperados para la neurona.

Como método de optimización la técnica de gradiente descendente define los pesos de la capa de salida w'_{kj} como en los de la capa oculta w_{ji}

$$\delta w'_{kj} = -\varepsilon \frac{\delta E}{\delta w'_{kj}} \quad \delta w_{ji} = -\varepsilon \frac{\delta E}{\delta w_{ji}} \quad (6)$$

El algoritmo se basa en un ciclo de propagación-adaptación de dos fases [19] (Figura 5):

La primera fase es el aprendizaje “hacia adelante”, en donde la señal de entrada se propaga a través de la red, capa por capa, hasta producir una señal de salida. Esta salida se compara con la respuesta deseada y se calcula la señal de error para cada una de las neuronas de la capa de salida.

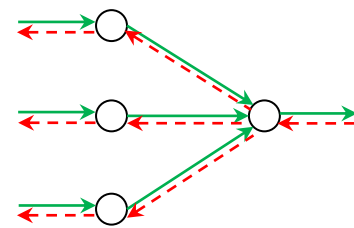


Figura 5. Fases del algoritmo de aprendizaje BP. Adaptado de [17]

GONZÁLEZ: TAMIZACIÓN DE PACIENTES CON SOSPECHA DE CÁNCER DE MAMA MEDIANTE REDES NEURONALES ARTIFICIALES UTILIZANDO VARIABLES CLÍNICAS

La segunda fase o aprendizaje “hacia atrás”, inicia en la capa de salida hasta la primera capa oculta, propagando la señal de error a través de la red en sentido contrario al flujo de señal. En este proceso se calcula el aporte de error de cada una de las neuronas de las capas ocultas y con esto se realiza el ajuste de los pesos sinápticos.

El aprendizaje BP presenta dos formas básicas de funcionamiento [17]:

- Modo *batch* o aprendizaje por lotes: en este modo los pesos se actualizan solamente después de haberse presentado a la red todos los patrones de entrada, lo cual constituye una época.
- Modo *on-line* o aprendizaje secuencial: en este modo los pesos se van actualizando a medida que se presentan cada uno de los patrones a la red.

Uno de los inconvenientes que puede tener la red al momento del entrenamiento es el denominado sobreentrenamiento o sobreajuste (*overfitting*) [18] [20]. Este se presenta cuando al tratar de alcanzar un error de aprendizaje muy pequeño, el algoritmo ajusta tanto los parámetros que lleva a la red a memorizarse los patrones de entrada en vez de realizar un verdadero entrenamiento para generalización. Como respuesta tendrá un error de generalización grande al ingresarle un grupo de datos nuevo.

Lo ideal es entrenar la red hasta que el error de generalización sea el mínimo. Esto se puede lograr utilizando el método denominado validación cruzada (*cross validation*) [18].

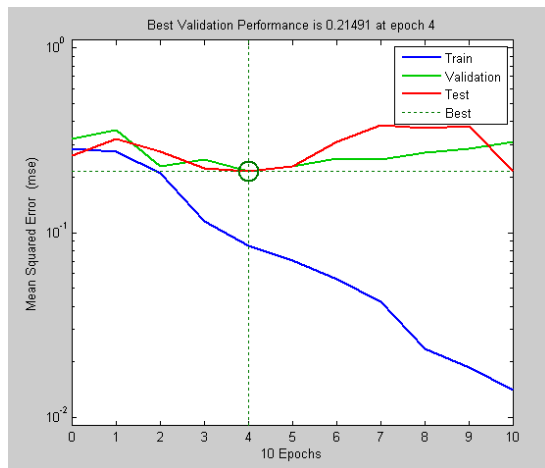


Figura 6. Ejemplo del comportamiento del error durante el entrenamiento

La validación cruzada consiste en dividir el grupo de patrones de entrada en tres subgrupos: entrenamiento, validación y prueba, con el fin de ir entrenando y validando la red hasta encontrar ese punto óptimo donde el error de generalización sea el mínimo.

Sin embargo también se puede presentar que el entrenamiento se detenga cuando el error se encuentre en un mínimo local y no en uno global. Esto se podría controlar utilizando la técnica de parada temprana (*early stopping*) [18]. Esta técnica realiza el entrenamiento hasta alcanzar un error de aprendizaje dado, mientras va guardando distintas configuraciones de la red para luego buscar entre estas la que obtuvo el menor error de generalización (Figura 6).

Radial Basis Function (RBF)

Se conoce como Red de Función de Base Radial o *Radial Basis Function Network* (RBFN) a la red que utiliza como función de activación en la capa oculta una función de base radial $\phi(r)$ [18] [19]. Este tipo de funciones se denominan así por su simetría radial las cuales dan una salida igual a los patrones que presenten la misma distancia a un centroide.

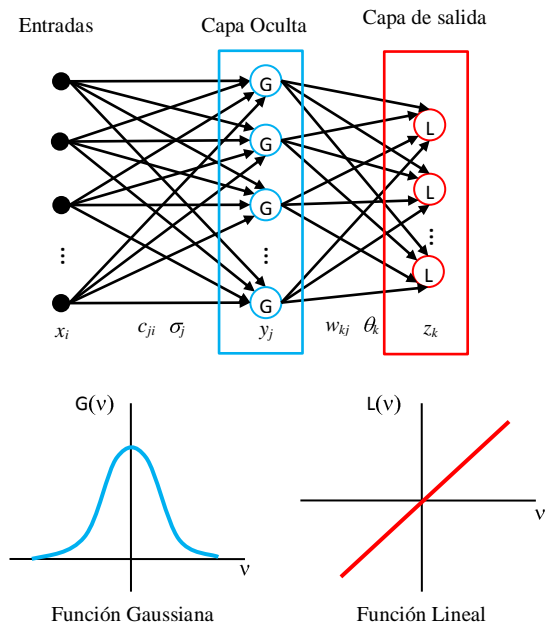


Figura 7. Arquitectura red RBF. Adaptado de [18]

GONZÁLEZ: TAMIZACIÓN DE PACIENTES CON SOSPECHA DE CÁNCER DE MAMA MEDIANTE REDES NEURONALES ARTIFICIALES UTILIZANDO VARIABLES CLÍNICAS

Es una red unidireccional que utiliza aprendizaje supervisado y no supervisado. Su arquitectura está formada por las entradas, una capa oculta y una capa de salida (Figura 7).

En la capa de salida la función de activación es la suma ponderada y se aplica por lo general una función de tipo sigmoideo

El proceso en la capa oculta se basa en el cálculo del centroide, dado por la distancia entre las entradas y el vector de pesos asociados a cada neurona.

Cada una de las neuronas de la capa oculta almacena un centroide c_{ji} , y calcula la distancia euclídea r_j que separa las entradas x_i de su centroide [18]

$$r_j^2 = \|\bar{x} - \bar{c}_j\|^2 = \sum_i (x_i - c_{ji})^2 \quad (7)$$

Luego se calcula la salida de la capa oculta y_j a través de la función de base radial. Entre las más utilizadas se encuentra la función Gaussiana la cual está dada por [21]

$$\phi(r) = e^{-r^2/2\sigma^2} \quad (8)$$

En donde σ es un parámetro de normalización que mide el ancho de la función gaussiana o *spread* el cual corresponde al radio de influencia de la neurona en el espacio de entradas. El valor de σ es directamente proporcional al tamaño de la región afectada por la neurona.

Entonces la salida de la capa oculta y_j quedaría [18]

$$y_j = e^{\frac{-r_j^2}{2\sigma_j^2}} = e^{\frac{-\sum_i (x_i - c_{ji})^2}{2\sigma_j^2}} \quad (9)$$

Una neurona se activa si las entradas se aproximan a su centroide, si es muy diferente, significa que no reconoce el patrón de entradas y la salida tiende a cero.

La respuesta de la capa de salida z_k sería entonces [18]

$$z_k = \sum_j w_{kj} y_j + \theta_k = \sum_j w_{kj} \phi(r_j) + \theta_k \quad (10)$$

En esta red se utiliza comúnmente el aprendizaje por etapas.

En la primera etapa se introducen todos los patrones de entrada para entrenar los grupos neuronales de las capas ocultas. Estos actualizan su centroide a través de iteraciones sucesivas hasta la convergencia.

El valor de los centroides c_{ji} se calcula por medio de un algoritmo de agrupamiento no supervisado por ejemplo el algoritmo *k-means* [18]

$$c_{ji} = \frac{1}{N_j} \sum_{\bar{x} \in \text{neurona } j} \bar{x} \quad (11)$$

En donde N_j representa el número de entradas a la neurona j y x son las entradas a esa neurona.

Los parámetros de normalización o escala se calculan a través de criterios heurísticos, como procedimiento que calcula el radio de influencia de las entradas de cada neurona con respecto a las demás [18]

$$\sigma_j^2 = \frac{1}{N} \sum_{l=1}^N \|\bar{c}_l - \bar{c}_j\|^2 = \frac{1}{N} \sum_{l=1}^N \sum_k (c_{lk} - c_{jk})^2 \quad (12)$$

En donde N es el número de neuronas más cercanas a la neurona j .

Finalmente, en la segunda fase se ajustan los pesos de las conexiones de la capa de salida [18]

$$z_k = \sum_j w_{kj} \phi(r_j) + \theta_k \quad (13)$$

la cual resulta en

$$w_{kj}(t+1) = w_{kj}(t) + \varepsilon(t_k - z_k) \phi(r_j) \quad (14)$$

En donde t_k son las salidas objetivo.

Entre las redes más comunes que utilizan funciones de base radial se encuentran la Red Neuronal Probabilística (PNN) utilizada en aplicaciones de clasificación para datos con variable objetivo categórica, y la General Regression Neural Networks (GRNN) utilizada para regresión y variable objetivo tipo continua.

1.4. Métodos Bioestadísticos

Regresión Logística Binaria

La regresión logística busca determinar la relación entre la variable dependiente o respuesta y las variables independientes (predictoras) o covariables. Su objetivo es encontrar esos factores que tienen un efecto sistemático sobre la probabilidad de pertenecer al grupo que posee el atributo de interés [22], como por ejemplo presentar una enfermedad o no, y así establecer la capacidad de predicción de una prueba.

GONZÁLEZ: TAMIZACIÓN DE PACIENTES CON SOSPECHA DE CÁNCER DE MAMA MEDIANTE REDES NEURONALES ARTIFICIALES UTILIZANDO VARIABLES CLÍNICAS

La regresión logística binaria es indicada cuando la variable dependiente o diagnóstico de la enfermedad es dicotómica (enfermo o no) y las observaciones son variables aleatorias independientes que siguen una distribución binomial.

El modelo de regresión logística binaria mide la influencia que tienen las covariables sobre la respuesta en términos de *Odds Ratio* u Oportunidad Relativa.

La Oportunidad (*Odds*) es la probabilidad de que se presente un evento entre la probabilidad de que no ocurra [22]:

$$Odds = \frac{p_A}{(1 - p_A)} \quad (15)$$

De manera general, la *Odds Ratio* (OR) es la división entre la oportunidad de que ocurra un evento en el grupo expuesto y la oportunidad de que ocurra en el grupo de los no expuestos [22]:

$$Odds Ratio = \frac{p_A}{(1 - p_A)} \bigg/ \frac{p_B}{(1 - p_B)} \quad (16)$$

La OR puede tomar valores entre cero e infinito. La interpretación se puede dar de la siguiente manera:

- Para $OR = 1$ no existe factor de riesgo puesto que la oportunidad es igual tanto para los expuestos como para los no expuestos.
- Para $OR > 1$ indica un aumento en la probabilidad de ocurrencia de un evento por lo tanto la existencia de un factor de riesgo.
- Para $OR < 1$ indica una disminución en la probabilidad de ocurrencia del evento, es decir, que hay un factor protector que reduce la presencia de este.

El modelo de regresión logística está dado por [23]:

$$\log \left(\frac{p(y = 1|x)}{1 - p(y = 1|x)} \right) = b_0 + \sum_i^n b_i x_i \quad (17)$$

En donde,

- p es la probabilidad de que la variable independiente y tome el valor de 1 (presente el atributo de interés)
- x_i las covariables o factores de riesgo
- n es el número de covariables
- b_0 es el término independiente
- b_i los coeficientes de las covariables

Variable confusora

Es la variable que afecta la variable independiente al estar relacionada con alguna independiente. Este tipo de variables deben formar parte del modelo así no presenten mucha significancia estadística [22].

1.5. Medidas de validez de las pruebas diagnósticas

Una prueba diagnóstica puede dar un resultado positivo o negativo así los individuos puedan estar o no enfermos [24]. Para evaluar su capacidad de clasificación se tienen que comparar los resultados obtenidos en la prueba contra el patrón de oro definido para la condición analizada. Esto nos proporciona las siguientes cuatro posibilidades:

- Verdaderos Positivos (VP): individuos que se encuentran enfermos y la prueba los clasifica correctamente (prueba positiva).
- Falsos negativos (FN): individuos que se encuentran enfermos y la prueba los clasifica como no enfermos.
- Verdaderos Negativos (VN): individuos no enfermos y la prueba los clasifica correctamente (prueba negativa).
- Falsos Positivos (FP): individuos no enfermos y la prueba los clasifica como enfermos.

Estas posibilidades se pueden presentar en una tabla de 2x2 para facilitar su visualización (Tabla 2).

Los índices para medir la validez de una prueba son la sensibilidad y la especificidad [24].

Tabla 2. Relación prueba diagnóstica / enfermedad

	Enfermos	No enfermos	
Prueba +	VP	FP	VP+FP
Prueba -	FN	VN	FN+VN
	VP+FN	VN+FP	

La sensibilidad es la capacidad de clasificar correctamente a los individuos enfermos. También se le conoce como Tasa de Verdaderos Positivos (TVP) y está dada por

$$Sensibilidad = \frac{VP}{VP + FN} \quad (18)$$

GONZÁLEZ: TAMIZACIÓN DE PACIENTES CON SOSPECHA DE CÁNCER DE MAMA MEDIANTE REDES NEURONALES ARTIFICIALES UTILIZANDO VARIABLES CLÍNICAS

La especificidad es la capacidad de detectar correctamente los individuos que no presentan enfermedad. Es la Tasa de Verdaderos Negativos (TVN) y está dada por

$$\text{Especificidad} = \frac{VN}{VN + FP} \quad (19)$$

La exactitud es el porcentaje de individuos clasificados correctamente, es decir, la suma de los verdaderos positivos y los verdaderos negativos

$$\text{Exactitud} = \frac{VP + VN}{VP + FP + FN + VN} \quad (20)$$

Se desea de una prueba diagnóstica que presenten valores altos tanto de sensibilidad como de especificidad. En los casos en donde su relación es dependiente se debe elegir un punto de corte para diferenciar los individuos enfermos de los no enfermos.

Una forma para analizar la relación entre la sensibilidad y la especificidad es a partir la curva de Características Operativas para el Receptor (ROC)⁴. En esta gráfica el eje y corresponde a la sensibilidad y el eje x a [1-especificidad] (Figura 8).

Como se puede ver a medida que se mueve el punto de corte se confirma la relación inversa entre la sensibilidad y la especificidad.

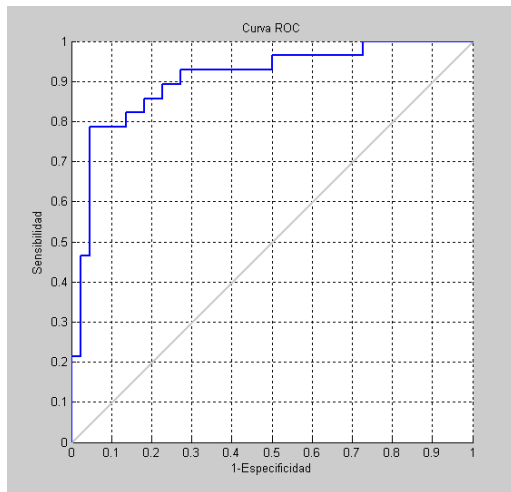


Figura 8. Ejemplo de curva ROC

El Área bajo la curva ROC (AUC) estima la capacidad de una prueba diagnóstica de discriminar entre enfermos y no enfermos. Los valores que toma están comprendidos entre 1, que se interpretaría como una discriminación perfecta, y 0.5 que equivale a no discriminar [25]. Para valores superiores de 0.7 se considera que la prueba tiene un nivel aceptable de discriminación [26].

1.6. Validación cruzada de k iteraciones

La validación cruzada es una técnica utilizada para analizar modelos de inteligencia artificial y minería de datos con el fin de evaluar su capacidad de generalización y estabilidad ante los datos.

En la validación cruzada de k iteraciones, el grupo original de datos se divide en k sub-grupos, de los cuales se escoge uno como el grupo de datos de validación y los $k-1$ restantes constituyen el grupo de entrenamiento. El proceso se repite k veces de tal forma que cada uno de los subgrupos será en cada iteración el grupo de validación [19].

Posteriormente se calcula la media aritmética de los k resultados de las iteraciones para obtener una sola respuesta.

Con este método se garantiza que de forma aleatoria todos los datos se utilicen para entrenamiento y que cada uno de ellos pertenezca sólo una vez al grupo de validación.

2. METODOLOGÍA

En la Figura 9 se muestra el diagrama de la metodología seguida en este proyecto de investigación.

2.1. Adquisición de datos

Los datos para la realización de este trabajo hacen parte del proyecto de COLCIENCIAS *Evaluación de la Termografía Infrarroja en la Detección del Carcinoma de Glándula Mamaria*⁵. La información pertenece a 120 mujeres mayores de edad, quienes por sospecha de cáncer de mama habían sido remitidas a consulta médica especializada en este tipo de enfermedad⁶.

⁴ Receiver Operating Characteristic

⁵ Estudio adelantado por la Universidad Industrial de Santander a través del Grupo de Investigación en Conectividad y Procesado de Señal (CPS) en conjunto con el Grupo de Investigación en Patología Oncológica ONCOPAT.

⁶ Entre septiembre de 2008 y agosto de 2011.

GONZÁLEZ: TAMIZACIÓN DE PACIENTES CON SOSPECHA DE CÁNCER DE MAMA MEDIANTE REDES NEURONALES ARTIFICIALES UTILIZANDO VARIABLES CLÍNICAS

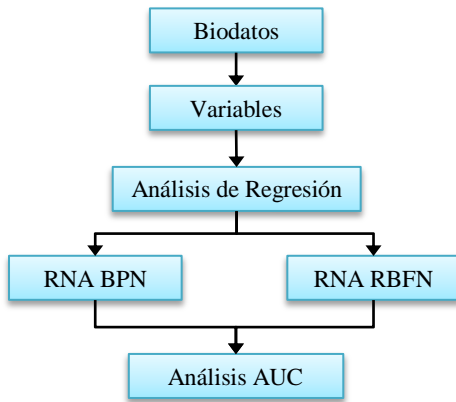


Figura 9. Metodología

La edad del grupo de participantes oscila entre los 18 y 89 años. De acuerdo al resultado de la biopsia en seno, 46 de ellas tienen cáncer de seno mientras que las 74 restantes no tienen.

La información se encontró recopilada en dos encuestas (Anexo A): la primera incluye la información personal, datos sociodemográficos, historia de salud familiar, cobertura y acceso a la atención médica. La segunda contiene los datos del examen clínico y el resultado histopatológico o biopsia de mama. Estas encuestas fueron digitalizadas en Microsoft Excel para facilitar la manipulación y posterior exportación a Matlab®.

2.2. Variables de entrada y salida

Los patrones de entrada a la RNA son las covariables o factores de riesgo asociados al cáncer de mama. La salida deseada es el diagnóstico obtenido en la biopsia de seno cuyo resultado es tener o no cáncer.

2.3. Preselección de covariables

Dado que las encuestas hacen parte de un estudio de investigación más amplio se obtuvo información adicional a la requerida en este trabajo. Por tal razón fue necesario realizar un análisis de estos datos y sólo tener en cuenta los factores que, de acuerdo a estudios realizados y el criterio médico, puedan aumentar el riesgo de incidencia de cáncer de mama.

A continuación se relacionan los factores de riesgo contemplados para determinar las entradas al sistema:

- Datos sociodemográficos
 - Edad de la paciente
 - Área de residencia
 - Raza

- Historia familiar de cáncer de mama
 - De primer grado: Mamá, hermanas, hijas
 - De segundo grado: Tías y primas
- Antecedentes Hormonales
 - Edad de la menarquia
 - Número de embarazos
 - Edad del primer embarazo
 - Lactancia
 - Edad de la menopausia
 - Terapia de reemplazo hormonal
- Valoración clínica
 - Dolor en los senos
 - Asimetría en los senos
 - Presencia de masas en los senos (percepción del médico)
 - Presencia de masas en los senos (percepción de la paciente)
 - Telorrea
 - Anormalidad en la piel de los senos
- Medidas antropométrica
 - Talla
 - Peso
 - Índice de Masa Corporal (IMC)

2.4. Selección de las covariables

Es importante identificar las covariables fuertemente asociadas al cáncer de mama con el fin de aumentar la velocidad de entrenamiento de la red y su capacidad de discriminación. Para esto se utilizó el modelo de regresión logística binaria debido a la naturaleza dicotómica de la respuesta.

Se definió como atributo de interés la presencia de cáncer de mama de acuerdo al resultado de la biopsia. El análisis se realizó con el *software* estadístico Stata⁷ versión 8.2.

2.5. Construcción de la RNA

Con el fin de elegir la red que presente mejor desempeño, se crearon diferentes tipos de configuraciones basadas en la combinación de especificaciones como algoritmo de aprendizaje, función de transferencia y número de neuronas ocultas.

⁷ Stata © Copyright 1996–2013 StataCorp LP.

GONZÁLEZ: TAMIZACIÓN DE PACIENTES CON SOSPECHA DE CÁNCER DE MAMA MEDIANTE REDES NEURONALES ARTIFICIALES UTILIZANDO VARIABLES CLÍNICAS

La construcción de las redes neuronales artificiales se realizó en el software Matlab®⁸. Se utilizaron las redes neuronales *backpropagation* y RBFN.

Red Neuronal Backpropagation

Los parámetros utilizados para la construcción de la red fueron los siguientes:

- *Estructura de la red*: Se creó una red de dos capas: una capa oculta y una capa de salida. En la capa oculta se probó un rango de 5 a 20 neuronas. La capa de salida estaba conformada por una neurona.
- *Función de Activación*: En la capa oculta se utilizó la función sigmoide tangente hiperbólica. En la capa de salida se probaron dos funciones: la lineal y la sigmoide tangente hiperbólica.
- *Algoritmo de aprendizaje*: se utilizaron para el entrenamiento los algoritmos *Levenberg-Marquardt* y *Resilient backpropagation*.
- *Función de error o coste*: se utilizó el error cuadrático medio (MSE)
- *Entradas a la red*: Cada grupo de patrones de entrada fue separado aleatoriamente en tres subgrupos:
 - Grupo de entrenamiento: 60%
 - Grupo de validación: 20%
 - Grupo de prueba: 20%

El grupo de entrenamiento se utilizó para determinar los pesos de la RNA. El grupo de validación se creó para implementar la técnica de parada temprana y evitar el sobreajuste. Con el grupo de prueba se probó el rendimiento final de la red.

Red Neuronal RBFN

Se utilizó la Red Neuronal Probabilística, *Probabilistic Neural Network* (PNN), la cual utiliza funciones de base radial y es indicada para problemas de clasificación. Se utilizó el *spread* como el parámetro que determinó las diferentes configuraciones de red.

Los parámetros utilizados para la construcción de la red fueron los siguientes:

- *Estructura de la red*: Es una red de dos capas: una capa oculta y una capa de salida. El algoritmo determina el número de neuronas ocultas de acuerdo al que presente mejor comportamiento para cada valor de *spread*. La capa de salida tiene una neurona.
- *Spread*: Se probó un rango entre 0.1 y 3 con incrementos de 0.1.
- *Función de transferencia*: En la capa oculta se utilizó la función de base radial *radbas* y la capa de salida la función *compet*.
- *Función de error o coste*: se utilizó el error cuadrático medio (MSE)
- *Entradas a la red*: Se utilizaron los mismos grupos de entrenamiento obtenidos para las redes BP con el fin de comparar el rendimiento en las diferentes redes.

Validación de los algoritmos

Para analizar la capacidad de clasificación, se calculó la sensibilidad, especificidad y AUC de cada una de las configuraciones a partir del resultado de simulación y la respuesta deseada.

Debido a que los pesos iniciales son aleatorios y por tanto cada vez que se entrena una red da un resultado diferente, fue necesario realizar 5 entrenamientos para cada una de las configuraciones propuestas.

Se eligieron las configuraciones cuyo valor de AUC fue igual o superior a 0.8, y se les realizó la validación cruzada de 5 iteraciones para analizar su estabilidad ante los datos.

3. RESULTADOS

3.1. Determinación de entradas y salida de la RNA

Se realizó una prueba de independencia basada en el contraste *Chi-cuadrado* y se determinó que no existe colinealidad entre las covariables que se utilizaron.

El análisis de regresión logística binaria presentó el modelo con la mejor significancia estadística el cual está compuesto por los factores de riesgo que se muestran en la Tabla 3.

⁸ Matlab® Version 7.6.0.324 (R2008a). The MathWorks, Inc.

GONZÁLEZ: TAMIZACIÓN DE PACIENTES CON SOSPECHA DE CÁNCER DE MAMA MEDIANTE REDES NEURONALES ARTIFICIALES UTILIZANDO VARIABLES CLÍNICAS

Tabla 3. Resultado de la Regresión Logística Binaria

Covariable	Odds Ratio	P>z
Menopausia	-0,61264	0,381
Edad	0,13544	0,000
Número de embarazos	-0,29550	0,034
Índice de Masa Corporal - IMC	0,06633	0,210
Primer embarazo antes de los 20 años	0,73523	0,235
Terapia de reemplazo hormonal	0,53024	0,546
Masa mayor a 2cm	1,49461	0,010
_cons	-8,79611	0,000

El logaritmo de la razón de tener cáncer de mama es -8,76: -0,61 si es menopáusica, +0,13 por cada año de la pacientes, -0,29 por cada embarazo, +0,6 por cada punto del índice de masa corporal, +0,73 si el primer embarazo fue antes de los 20 años, +0,53 si ha recibido terapia hormonal y +1,49 si la masa tiene más de 2 cm.

Aunque de manera individual algunas de las covariables no son significativas al nivel de significación habitual ($p < 0.05$) no se excluyeron debido a que se identificaron como variables confusoras.

Con los resultados de la selección, se crearon dos modelos para el entrenamiento de la red. Estos son:

- Modelo 1: conformado por las variables obtenidas en la regresión logística. Estas son: menopausia, edad, número de embarazos, IMC, primer embarazo antes de los 20 años, terapia de reemplazo hormonal y masa mayor a 2 cm. Con este modelo la estructura de la

RNA es 7-N-1, en donde N es el número de neuronas ocultas.

- Modelo 2: conformado por el grupo 1 y las variables que según la literatura médica son factores de riesgo de cáncer de seno pero que no tuvieron significancia estadística según el análisis realizado. Estas son: menarquia, lactancia y antecedentes de primer grado de cáncer de mama y de órganos del sistema reproductor femenino. La estructura de la RNA con este modelo es 10-N-1.

La salida objetivo también conocida como *target* es el resultado de la biopsia, el cual es positiva (1) para cáncer de seno o negativa (0) si no presenta cáncer.

3.2. Implementación de las redes

Para cada uno de los dos modelos de entrenamiento se obtuvieron 60 configuraciones de la red BP y 30 de la red RBFN para un total de 180 configuraciones. A cada una se le realizó 5 entrenamientos de los cuales se tuvo en cuenta el que presentó mayor valor de área ROC en la simulación (Anexo B).

Se seleccionaron las configuraciones que tuvieron el valor de AUC igual o superior a 0.8 (Tabla 4). Cumplieron con este criterio tres configuraciones tipo BP para el grupo 1 de entrenamiento (RNA-1 a RNA-3) y seis configuraciones tipo de BP para el grupo 2 (RNA-4 a RNA-9).

Tabla 4 Resumen de resultados de la simulación de las redes BP

Estructura de red	Tipo RNA	Función de transferencia		Algoritmo de aprendizaje	Neuronas ocultas	Acierto (%)	Sensibilidad (%)	Especificidad (%)	AUC	
		capa oculta	capa salida							
Configuraciones con mejor comportamiento (AUC \geq 0.8):										
RNA-1	7-11-1	BP	tansig	purelin	trainlm	11	86,96	77,78	92,86	0,8556
RNA-2	7-12-1	BP	tansig	purelin	trainrp	12	83,33	66,67	93,33	0,8000
RNA-3	7-20-1	BP	tansig	tansig	trainlm	20	83,33	66,67	93,33	0,8000
RNA-4	10-5-1	BP	tansig	purelin	trainlm	5	87,50	66,67	100,00	0,8333
RNA-5	10-6-1	BP	tansig	purelin	trainlm	6	83,33	77,78	86,67	0,8222
RNA-6	10-9-1	BP	tansig	purelin	trainrp	9	83,33	66,67	93,33	0,8000
RNA-7	10-8-1	BP	tansig	tansig	trainlm	8	87,50	66,67	100,00	0,8333
RNA-8	10-14-1	BP	tansig	tansig	trainlm	14	83,33	66,67	93,33	0,8000
RNA-9	10-17-1	BP	tansig	tansig	trainlm	17	87,50	77,78	93,33	0,8630
Configuraciones con peor comportamiento:										
RNA-10	7-8-1	BP	tansig	purelin	trainrp	8	70,83	33,33	93,33	0,6333
RNA-11	10-2-1	BP	tansig	tansig	trainlm	19	70,83	33,33	93,33	0,6333

Abreviaturas: BP: *Backpropagation*, RNA: Red Neuronal Artificial, tansig: Función de transferencia sigmoidea tangente hiperbólica, purelin: Función de transferencia Lineal, trainlm: Algoritmo de aprendizaje *Levenberg-Marquardt*, trainrp: Algoritmo de entrenamiento *Resilient Backpropagation*, AUC: Área bajo la curva ROC

GONZÁLEZ: TAMIZACIÓN DE PACIENTES CON SOSPECHA DE CÁNCER DE MAMA MEDIANTE REDES NEURONALES ARTIFICIALES UTILIZANDO VARIABLES CLÍNICAS

Las redes RBFN fueron entrenadas y se simularon inicialmente con los datos de entrenamiento para probar que clasificaran correctamente. Se presentó una correcta clasificación para un *spread* inferior a 1,5 para el grupo de entradas 1, y para el grupo de entradas 2 las configuraciones con un *spread* inferior a 2,1.

Para las redes BP, el mayor valor de AUC obtenido fue 0.8630. Se obtuvo una sensibilidad alrededor del 70% y una especificidad mayor al 90% la cual alcanzó el 100% en algunas configuraciones.

La configuración que presentó mejor desempeño fue la RNA-9 con un valor de AUC de 0.8630, 77.78% de sensibilidad y 93.33% de especificidad. Las que presentaron peor comportamiento fueron las redes RNA-10 y RNA-11 con un AUC de 0.633.

Las redes tipo RBFN presentaron valores de AUC inferiores a 0.8 (Anexo C). Como se observó en los resultados que era posible que al ampliar algunos rangos del *spread* se obtuvieran mejores resultados, se entrenaron redes con esos valores. Aunque se encontraron configuraciones con mejor comportamiento (AUC de 0.7111), ninguna cumplió con el criterio de selección (Tabla 5). Las redes que presentaron el peor comportamiento obtuvieron valores de AUC alrededor de 0.5000.

Tabla 5 Resumen de resultados de la simulación de RBFN

Modelo entrada	Spread	Acierto (%)	Sensibilidad (%)	Especificidad (%)	AUC
Configuraciones con mejor comportamiento:					
1	0,0810	66,67	88,89	53,33	0,7111
2	0,1040	66,67	88,89	53,33	0,7111
Configuraciones con peor comportamiento:					
1	0,0300	37,50	100,00	0,00	0,5000
2	0,2000	54,17	22,22	73,33	0,4778

Abreviaturas: AUC: Área bajo la curva ROC

3.3. Validación cruzada de k iteraciones

Finalmente para observar la estabilidad frente a los datos, se realizó a cada una de las configuraciones seleccionadas una validación cruzada de 5 iteraciones, la cual dio como resultado la media aritmética de la sensibilidad, especificidad y AUC (Tabla 6).

Se obtuvo una sensibilidad entre el 60% y el 70%, especificidad mayor al 75% y AUC alrededor del 0.7.

Aunque algunas configuraciones obtuvieron mayores valores de especificidad, la RNA-9 fue la que presentó el mejor comportamiento con la mayor sensibilidad (71.78%), una especificidad del 78.29% y AUC de 0.7503.

Tabla 6 Resultado de la Validación cruzada de 5 iteraciones

	Acierto ^a (%)	Sensibilidad ^a (%)	Especificidad ^a (%)	AUC ^a
RNA-1	72,50	58,00	80,76	0,6938
RNA-2	73,33	47,56	89,24	0,6840
RNA-3	72,50	51,33	84,86	0,6810
RNA-4	74,17	58,67	83,62	0,7114
RNA-5	71,56	63,28	76,95	0,7040
RNA-6	71,67	58,67	79,71	0,6919
RNA-7	74,17	58,44	83,71	0,7108
RNA-8	75,00	63,11	82,57	0,7284
RNA-9	75,83	71,78	78,29	0,7503

^a Media Aritmética

Abreviaturas: AUC: Área bajo la curva ROC

En relación a los patrones de entrada, las configuraciones que mostraron mayor independencia de los datos fueron las pertenecientes al modelo 2. Respecto a las especificaciones de red, obtuvieron mejores resultados las configuraciones que utilizaron función de transferencia sigmoidea tangente hiperbólica en las dos capas y algoritmo de aprendizaje *Levenberg-Marquardt*.

4. OBSERVACIONES Y CONCLUSIONES

Este trabajo de investigación evalúa una posible herramienta para el diagnóstico clínico de cáncer de mama basada en redes neuronales artificiales en conjunto con variables clínicas. La metodología de este trabajo puede ser aplicada al estudio de otros tipos de cáncer.

El análisis de regresión logística binaria identificó siete factores de riesgo asociados al cáncer de mama: ser menopáusica, edad de la paciente, número de embarazos, índice de masa corporal, primer embarazo antes de los 20 años, haber recibido terapia de remplazo hormonal y masa en seno con tamaño mayor a 2 cm.

Los resultados obtenidos en la validación cruzada mostraron que los modelos basados en redes neuronales artificiales se pueden considerar como pruebas diagnósticas con un nivel aceptable de discriminación.

GONZÁLEZ: TAMIZACIÓN DE PACIENTES CON SOSPECHA DE CÁNCER DE MAMA MEDIANTE REDES NEURONALES ARTIFICIALES UTILIZANDO VARIABLES CLÍNICAS

La configuración que muestra el mejor desempeño con el análisis de validación cruzada fue la RNA-9 tipo BP, algoritmo de aprendizaje *Levenberg-Marquardt* y función de transferencia sigmoidea tangente hiperbólica en las dos capas.

En general, las redes que presentaron los mejores resultados fueron tipo BP con función de transferencia sigmoidea tangente hiperbólica en las dos capas y algoritmo de aprendizaje *Levenberg-Marquardt*. Esto concuerda con diferentes estudios en donde se ha encontrado que este tipo de funciones son más eficientes para la arquitectura perceptrón multicapa.

Aunque parecía que las variables recomendadas en la literatura médica no presentaban una fuerte correlación como lo indicaban los resultados de la regresión logística binaria, mostraron buen desempeño en la validación cruzada. El modelo 2 es el más adecuado como entradas a la RNA.

Se mostró la implementación de las redes neuronales como alternativa para la tamización de cáncer de mama con un desempeño bastante aceptable, superando al diagnóstico realizado en el examen médico el cual presenta baja sensibilidad (48.3 - 59.8%) y depende del examinador [27].

En relación a otros estudios que implementaron redes BP con características similares a las desarrolladas en este trabajo [28] [29], se observó que las configuraciones presentaron valores semejantes de sensibilidad (50-60%) y especificidad (80%).

Las redes BP desarrolladas basadas en variables clínicas tienen alta especificidad pero baja sensibilidad.

Para este trabajo ninguna de las configuraciones de red tipo RBFN obtuvo valores de AUC iguales o superiores a 0,8.

La mayor limitante que se presentó en el desarrollo de este trabajo fue en la recopilación de la información debido a la dificultad de encontrar personas que quisieran participar en el estudio y tuvieran resultado de la biopsia en seno.

Las conclusiones de este proyecto de investigación estuvieron basadas en los datos de mujeres con sospecha de cáncer de mama quienes ya habían sido remitidas para biopsia. Estos casos fueron seleccionados debido a que proporcionaban el estado definitivo de la enfermedad. Los resultados de este trabajo no deben ser aplicados a otro tipo de población ya que se pueden obtener resultados diferentes.

Para trabajos futuros se recomienda ampliar el grupo de muestra a la población en general. La implementación de otros modelos de redes neuronales artificiales, aumentando el número de capas ocultas y variación de parámetros permitirá comparar la eficiencia de este tipo de aplicaciones.

REFERENCIAS

- [1] J. Ferlay, H. Shin, F. Bray, D. Forman, C. Mathers y D. Parkin, «GLOBOCAN 2008 v2.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 10. Lyon, France: International Agency for Research on Cancer», 2010. [En línea]. Available: <http://globocan.iarc.fr>. [Último acceso: 2013 04 15].
- [2] P. Boyle y B. Levin, «World Cancer Report 2008», IARC Press, Lyon, 2008.
- [3] R. K. F. Hogben, «Screening for breast cancer in England: a review», *Obstetrics and Gynecology*, vol. 20, pp. 545-549, 2008.
- [4] R. W. Blamey, A. R. M. Wilson y J. Patnick, «ABC of breast diseases Screening for breast cancer», *BMJ*, vol. 321, n° 7262, p. 689–693, 2000.
- [5] K. McPherson, C. M. Steel y J. M. Dixon, «ABC of breast diseases. Breast cancer-epidemiology, risk factors and genetics», *BMJ*, vol. 321, p. 624–628, 2000 Disponible en: <http://www.bmj.com/cgi/content/full/321/7261/624?view=long&pmid=10977847>.
- [6] Instituto Nacional de Cancerología, «Recomendaciones para la tamización y la detección temprana del cáncer de mama en Colombia», INC, Bogotá, 2006.
- [7] S. Díaz, M. Piñeros y O. Sánchez, «Detección temprana del cáncer de mama: aspectos críticos para un programa de tamizaje organizado en Colombia», *Rev Colom Cancerol*, vol. 9, n° 3, pp. 93-105, 2005.
- [8] R. Trias Capella, «Inteligencia artificial en medicina: estado actual y perspectivas», *Medicina Clínica*, vol. 100, n° (Supl. 1), pp. 45-46, 1993.
- [9] S. C. Fok, E. Y. K. Ng y G. L. Thimm, «Developing Case-Based Reasoning for Discovery of Breast Cancer», *Journal of Mechanics in Medicine and Biology*, vol. 3, n° 3 y 4, pp. 231-245, 2003.
- [10] E. Bernal, «Inteligencia Artificial Aplicada al Diagnóstico Médico. Estado del arte Tesis de Maestría en Ingeniería, Ingeniería de Sistemas y Computación», Universidad Nacional de Colombia, 2006.
- [11] Y. Wu, K. Doi, M. L. Giger y R. M. Nishikawa, «Computerized detection of clustered microcalcifications in digital mammograms: applications of artificial neural networks», *Medical physics*, vol. 19, n° 3, pp. 555-60, 1992.
- [12] J. Y. Lo, J. A. Baker, P. J. Kornguth y C. E. Floyd, «Effect of patient history data on the prediction of breast cancer from mammographic findings with Artificial Neural Networks», *Acad Radiol*, vol. 6, pp. 10-15, 1999.

GONZÁLEZ: TAMIZACIÓN DE PACIENTES CON SOSPECHA DE CÁNCER DE MAMA MEDIANTE REDES NEURONALES ARTIFICIALES UTILIZANDO VARIABLES CLÍNICAS

- [13] E. Ng y E. Kee, «Advanced integrated technique in breast cancer thermography», *Journal of Medical Engineering & Technology*, vol. 32, n° 2, pp. 103-114, 2008 Disponible en: <http://informahealthcare.com/doi/abs/10.1080/03091900600562040>.
- [14] J. Koay, C. Herry y M. Frize, «Analysis of breast thermography with an artificial neural network», *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, vol. 2, pp. 1159-62, 2004.
- [15] American Cancer Society, «Breast Cancer», Última revisión médica: 24/09/2012. Última revisión: 24/09/2012. [En línea]. Available: <http://www.cancer.org/acs/groups/cid/documents/webcontent/003090-pdf.pdf>. [Último acceso: Septiembre 2012].
- [16] B. Norsa'adah, B. Rusli y A. Imran, «Risk factors of breast cancer in women in Kelantan, Malaysia», *Singapore Medical Journal*, vol. 46, n° 12, pp. 698-705, 2005. Disponible en: <http://www.sma.org.sg/smj/4612/4612a3.pdf>.
- [17] S. Haykin, *Neural Networks: A comprehensive foundation*, New Jersey: Prentice-Hall, Inc., 1999.
- [18] B. Martín del Brío y A. Sanz Molina, *Redes neuronales y sistemas borrosos*, Mexico: Alfaomega Grupo Editor, S.A. de C.V., 2007.
- [19] R. Flórez y J. M. Fernández, *Las redes neuronales artificiales*, Oleiros (La Coruña): Editorial Netbiblo, 2008.
- [20] H. Demuth, M. Beale y M. Hagan, *Neural Network Toolbox 6. User's Guide*, The MathWorks, Inc., 2009.
- [21] S. Chen, C. Cowan y P. M. Grant, «Orthogonal least squares learning algorithm for radial basis», *IEEE Transactions on Neural Networks*, vol. 2, n° 2, pp. 302-309, 1991.
- [22] F. J. Barón López y F. Téllez Montiel, *Apuntes de Bioestadística: Tercer ciclo en ciencias de la salud y medicina*, Málaga: Uma - Universidad de Málaga, 2004.
- [23] F. J. Caballero Granado, «Modelos de regresión logística incondicional», ©Sociedad Andaluza de Enfermedades Infecciosas (SAEI), 2010. [En línea]. Available: <http://saei.org/hemero/epidemiol/nota4.html>. [Último acceso: 2013].
- [24] B. Bermejo Fraile, *Epidemiología clínica aplicada a la toma de decisiones en medicina*, 2da ed., Navarra: Gobierno de Navarra, Departamento de Salud, 2006, p. 228.
- [25] A. K. Akobeng, «Understanding diagnostic tests 3: receiver operating characteristic curves», *Acta Paediatrica*, vol. 96, pp. 644-647, 2007.
- [26] Biostatistics and Biomathematics Research Group, «Biostatistics and Biomathematics Research Group - Institut de Recerca Biomèdica de Lleida», 25 09 2006. [En línea]. Available: <http://web.udl.es/Biomath/Bioestadistica/Dossiers/Temas%20especiales/Intervalos%20Referencia/Diagnostico%20y%20ROC.pdf>. [Último acceso: 2013 04 26].
- [27] T. Tan, C. Quek, G. Ng y E. Ng, «A novel cognitive interpretation of breast cancer thermography with complementary learning fuzzy neural memory structure», *J. Expert Systems with Applications*, vol. 33, pp. 652-666, 2007.
- [28] S. Fok, E. Ng y K. Tai, «Early detection and visualization of breast tumor with thermogram and neural network», *Journal of Mechanics in Medicine and Biology*, vol. 2, n° 2, pp. 185-195, 2002.
- [29] E. Ng, S. Fok, Y. Peh, F. Ng y L. Sim, «Computerized detection of breast cancer with artificial intelligence and thermograms», *Journal of Medical Engineering & Technology*, vol. 26, n° 4, pp. 152-157, 2002.
- [30] G. Van Belle, F. Lloyd D., P. J. Heagerty y T. Lumley, *Biostatistics: a methodology for the health sciences*, New Jersey: John Wiley & Sons, Inc., 2004.
- [31] D. Armesto, «Pruebas diagnósticas: Curvas ROC», *Rev Electron Biomed*, vol. 1, pp. 77-82, 2011.

ANEXOS

Anexo A. FORMATOS DE RECOLECCIÓN DE LA INFORMACIÓN

REGISTRO No. 4 CONSENTIMIENTO INFORMADO



Estimada Paciente:

Las enfermedades de los senos son cada vez más frecuentes en las mujeres en todo el mundo. Por esto, los investigadores del proyecto de investigación **“EVALUACIÓN DE LA TERMOGRAFÍA INFRARROJA EN LA DETECCIÓN DEL CARCINOMA DE GLÁNDULA MAMARIA”** realizado por miembros del Departamento de Patología de la Universidad Industrial de Santander (UIS) y de la Unidad de Oncología del Hospital Universitario (HUS) de Santander, en unión con el Grupo de Investigación de Conectividad y Procesado de Señal (CPS) y el grupo de investigación en Innovación y Desarrollo Tecnológico de Unisangil (IDENTUS) preocupados por la ésta situación estamos realizando este estudio que tiene como objetivo conocer la temperatura de las distintas partes del seno y sobre todo la temperatura específica del trastorno que pudiera tener el seno enfermo. Esto permite saber si éste método es útil para detectar el cáncer de seno en mujeres de cualquier edad.

La técnica de termografía infrarroja, es un examen que se puede realizar en cualquier parte del cuerpo, pero en éste estudio lo vamos a realizar a los senos únicamente. Se hace mediante el uso de una cámara diseñada especialmente para esto. La cámara se utiliza como una cámara fotográfica, de manera que registra la temperatura superficial de los senos. Lo que se observa después de la toma es una imagen de los senos con distintos colores en donde las zonas más calientes se ven de un color diferente al de las zonas más frías. Eso hace que si en el seno hay alguna masa, ésta pudiera resultar más fría o más caliente que el resto del seno y haría pensar que existe enfermedad en ese sitio.

La toma de la imagen es únicamente de los senos y no de su rostro ni de otro sitio del cuerpo. La cámara no estará en contacto físico con usted, no genera rayos x y es un procedimiento rápido e indoloro.

Este estudio se realizará en 200 pacientes, que es el número de personas necesarias para poder hacer un análisis valedero de los resultados. Su participación es completamente voluntaria y la realización de la técnica no implica ningún costo para usted como tampoco habrá compensación económica alguna. Los miembros del grupo investigador estarán en disposición de brindarle ahora y en el futuro cualquier información o pregunta que le surja acerca de los resultados o del procedimiento; para eso se suministrará en el momento de su valoración los datos de los responsables de aclarar dudas al respecto.

La información generada por este estudio es estrictamente confidencial y se mantendrá su privacidad. La información del estudio no será utilizada para generar beneficios económicos. Usted es libre de rehusar a participar en este estudio en cualquier momento sin que esto conlleve a cambios en su futuro cuidado.

Usted tiene derecho a conocer los resultados de los estudios realizados cuando lo desee, una vez se haya realizado el análisis de las imágenes así como de solicitar que no sean incluidos en las conclusiones del trabajo. De igual forma el grupo investigador podrá tomar la decisión de retirarla del estudio si lo considera conveniente.

PARTICIPANTE

Yo _____ Firma _____

Cédula No. _____ de _____

He leído y recibido copia del presente consentimiento informado. Habiendo comprendido el significado de la investigación declaro estar debidamente informada y consiento en participar en este el estudio.

Ciudad: _____ Fecha: _____ Hora: _____

TESTIGOS

Nombre _____ Firma _____

Nombre: _____ Firma _____

Ciudad: _____ Fecha: _____ Hora: _____

INVESTIGADOR QUE BRINDA EL CONSENTIMIENTO

Nombre: _____ Firma _____

Teléfono: _____

Ciudad: _____ Fecha: _____ Hora: _____



Universidad Industrial de Santander

FORMATO DE RECOLECCIÓN DE LA INFORMACIÓN

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE SALUD
ESCUELA DE MEDICINA

REGISTRO No. 5

ENCUESTA SOBRE LOS FACTORES ASOCIADOS SOCIODEMOGRÁFICOS, HEREDITARIOS Y HORMONALES ENDÓGENOS EN LAS PACIENTES CON CARCINOMA INFILTRANTE DE LA GLÁNDULA MAMARIA EN EL DEPARTAMENTO DE SANTANDER

Se guardará la confidencialidad de los datos y en ningún momento se revelará la identificación de los pacientes.

NOTA: Es importante que al registrar los datos tenga en cuenta las unidades de medida, marque con una X la respuesta en la casilla que corresponda y evite dejar espacios en blanco.

Nombre (s) y Apellidos:	
Dirección:	Barrio:
Tél.	Fecha hoy: dd/mm/aa (/ /)
Código de las Imágenes: Frontal: _____ Lateral Der _____ Lateral Izq _____ Oblicua	
Der. _____ Oblicua Izq _____	
No. de Historia Clínica:	Cédula:
Talla del Brasier: <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
DATOS SOCIODEMOGRAFICOS	
1. ¿Cuántos años cumplidos tiene?: <input type="checkbox"/> <input type="checkbox"/> años	2. ¿Cuál es su fecha de nacimiento? (/ /) dd/mm/aa
3. En el último año, ¿dónde ha residido? Nombre del lugar (municipio y/o vereda) _____ Área urbana <input type="checkbox"/> Área rural <input type="checkbox"/>	
4. De acuerdo a su recibo de luz, ¿A cual estrato corresponde su vivienda? 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/>	
5. En su opinión, ¿a cuál de las siguientes razas pertenece usted? 1. Blanca <input type="checkbox"/> 2. Mestiza <input type="checkbox"/> 3. Negra <input type="checkbox"/> 4. No sabe <input type="checkbox"/> 5. No responde <input type="checkbox"/>	
6. ¿Cuál es su Estado Civil?:	
Soltero	<input type="checkbox"/>
Casado	<input type="checkbox"/>
Divorciado	<input type="checkbox"/>
Unión Libre	<input type="checkbox"/>
Otro	<input type="checkbox"/>



Universidad
Industrial de
Santander

7. ¿Convive con pareja estable y permanente? Si <input type="checkbox"/> No <input type="checkbox"/>	
8. ¿Cuál fue el último grado de estudios que usted aprobó?	
Tipo de enseñanza	No. de años
Ninguna	0 <input type="checkbox"/>
Primaria	1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/>
Secundaria	6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 <input type="checkbox"/> 10 <input type="checkbox"/> 11 <input type="checkbox"/>
Técnica o Universitaria, incluyendo Postgrados	1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 <input type="checkbox"/> 10 <input type="checkbox"/>
9. ¿A qué se dedicó la mayor parte del tiempo en el último año?	
Trabajó	<input type="checkbox"/>
Trabajó y estudió	<input type="checkbox"/>
Estudió (a)	<input type="checkbox"/>
Actividades del hogar	<input type="checkbox"/>
Buscó trabajo	<input type="checkbox"/>
Pensionado (a)	<input type="checkbox"/>
Retirado sin pensión	<input type="checkbox"/>
Otra	<input type="checkbox"/> ¿Cuál?
10. ¿En cuál de los siguientes rangos está el ingreso mensual de su familia (personas que aportan económicamente para el sostenimiento de su hogar)? 1. \$0-\$496.800 <input type="checkbox"/> 2. \$496.800-\$993.600 <input type="checkbox"/> 3. \$993.600-\$1987.200 <input type="checkbox"/> 4. \$1987.200-\$3.974.400 <input type="checkbox"/> 5. \$3.974.400 o más <input type="checkbox"/> 6. No sabe <input type="checkbox"/> 7. Rehúsa contestar <input type="checkbox"/>	
COBERTURA Y ACCESO A LA ATENCIÓN MÉDICA	
11. ¿A su familia alguna vez le aplicaron la encuesta del SISBEN? Si <input type="checkbox"/> No <input type="checkbox"/> → Si "No" pase a la pregunta 14 No sabe/No recuerda <input type="checkbox"/> → Si "No" pase a la pregunta 14	
12. Después del año 2001, ¿Le han aplicado la encuesta del SISBEN a su familia? Si <input type="checkbox"/> No <input type="checkbox"/> No sabe/No recuerda <input type="checkbox"/>	
13. ¿En que nivel del SISBEN está clasificado? <input type="checkbox"/>	
14. ¿En el último año ha estado o estuvo asegurado o afiliado a un plan de salud como cotizante? Si <input type="checkbox"/> No <input type="checkbox"/> Si, pero no siempre <input type="checkbox"/> No sabe/No recuerda <input type="checkbox"/>	
15. ¿Actualmente esta asegurado o afiliado a un plan de salud como cotizante? Si <input type="checkbox"/> No <input type="checkbox"/> No sabe/No recuerda <input type="checkbox"/>	
16. ¿En el último año ha estado o estuvo asegurado o afiliado a un plan de salud como beneficiario? Si <input type="checkbox"/> No <input type="checkbox"/> Si, pero no siempre <input type="checkbox"/> No sabe/No recuerda <input type="checkbox"/>	
17. ¿Actualmente esta asegurado o afiliado a un plan de salud como beneficiario? Si <input type="checkbox"/> No <input type="checkbox"/> No sabe/No recuerda <input type="checkbox"/>	
18. ¿Actualmente a que entidad de salud está afiliado o es beneficiario?	
1. Nueva EPS (ISS) <input type="checkbox"/>	6. Fuerzas Militares, Policía Nacional <input type="checkbox"/>
2. Administradora de régimen Subsidiado (ARS) <input type="checkbox"/>	7. ECOPETROL <input type="checkbox"/>

3. Empresa promotora de Salud (EPS)	<input type="checkbox"/>	8. Magisterio	<input type="checkbox"/>
4. Empresa de Medicina prepagada	<input type="checkbox"/>	9. Ninguna	<input type="checkbox"/>
5. Empresa Solidaria	<input type="checkbox"/>	10. Otra, ¿Cuál?	
19. Nombres de la entidad de salud a la cual está afiliado			
HISTORIA DE SALUD FAMILIAR			
20. ¿Su mamá biológica tiene o tuvo cáncer? Si <input type="checkbox"/> No <input type="checkbox"/> → Si "No" pase a la pregunta 23 No sabe <input type="checkbox"/> → Si "No sabe" Pase a la pregunta 23			
21. ¿Qué edad tenía su mamá cuando le diagnosticaron cáncer? <input type="checkbox"/> <input type="checkbox"/> años			
22. ¿En qué sitio/órgano su mamá tiene o tuvo cáncer? _____			
23. ¿Su mamá biológica está viva? Si <input type="checkbox"/> No <input type="checkbox"/> No sabe <input type="checkbox"/>			
24. ¿Su papá biológico tiene o tuvo cáncer? Si <input type="checkbox"/> No <input type="checkbox"/> → Si "No" pase a la pregunta 27 No sabe <input type="checkbox"/> → Si "No" pase a la pregunta 27			
25. ¿Qué edad tenía su papá cuando le diagnosticaron cáncer? <input type="checkbox"/> <input type="checkbox"/> años			
26. ¿En qué sitio/órgano su papá tiene o tuvo cáncer? _____			
27. ¿Cuántas hermanas tiene usted? <input type="checkbox"/> <input type="checkbox"/> Si no tiene hermanas pase a la pregunta 31			
28. ¿Alguna de sus hermanas tiene o tuvo cáncer de mama? Si <input type="checkbox"/> No <input type="checkbox"/> → Si "No" pase a la pregunta 31 No sabe <input type="checkbox"/>			
29. ¿Cuántas hermanas tuvieron cáncer de mama? <input type="checkbox"/> <input type="checkbox"/>			
30. ¿Qué edad tenía(n) su(s) hermana(s) cuando le diagnosticaron cáncer de mama? 1. <input type="checkbox"/> <input type="checkbox"/> años 2. <input type="checkbox"/> <input type="checkbox"/> años 3. <input type="checkbox"/> <input type="checkbox"/> años			
31. ¿Algunas de sus hijas tiene o tuvo cáncer de mamá? Si <input type="checkbox"/> No <input type="checkbox"/> → Si "No" pase a la pregunta 34 No sabe <input type="checkbox"/>			
32. ¿Cuántas hijas tienen o tuvieron cáncer de mama? <input type="checkbox"/> <input type="checkbox"/>			
33. ¿Qué edad tenía(n) su(s) hija(s) cuando le(s) diagnosticaron cáncer de mama? 1. <input type="checkbox"/> <input type="checkbox"/> años 2. <input type="checkbox"/> <input type="checkbox"/> años 3. <input type="checkbox"/> <input type="checkbox"/> años			
34. ¿Alguna de sus tías o primas tiene o tuvo cáncer de mama? Si <input type="checkbox"/> No <input type="checkbox"/> → Si "No" pase a la pregunta 39 No sabe <input type="checkbox"/>			
35. ¿Cuántas tías tienen o tuvieron cáncer de mama? <input type="checkbox"/> <input type="checkbox"/>			
36. ¿Qué edad tenía(n) su(s) tía(s) cuando le(s) diagnosticaron cáncer de mama? 1. <input type="checkbox"/> <input type="checkbox"/> años 2. <input type="checkbox"/> <input type="checkbox"/> años 3. <input type="checkbox"/> <input type="checkbox"/> años			
37. ¿Cuántas primas tienen o tuvieron cáncer de mama? <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Ninguna → Si "Ninguna" pase a la pregunta 39			
38. ¿Qué edad tenía(n) su(s) prima(s) cuando le(s) diagnosticaron cáncer de mama? 1. <input type="checkbox"/> <input type="checkbox"/> años 2. <input type="checkbox"/> <input type="checkbox"/> años 3. <input type="checkbox"/> <input type="checkbox"/> años			



Universidad Industrial de Santander

ANTECEDENTES PERSONALES

39. ¿Qué edad tenía usted cuando tuvo su primera menstruación, regla o periodo? <input type="text"/> <input type="text"/> años
40. ¿Ha tenido embarazos? Si <input type="checkbox"/> No <input type="checkbox"/> → Si "No" pase a la pregunta 45 ¿Cuántos? <input type="text"/> <input type="text"/>
41. ¿Ha tenido recién nacidos muertos? Si <input type="checkbox"/> No <input type="checkbox"/> Si la respuesta es "sí" ¿cuantos? _____
42. ¿Ha tenido abortos? Si <input type="checkbox"/> No <input type="checkbox"/> Si la respuesta es "sí" cuantos abortos? _____
43. ¿A que edad tuvo su primer embarazo? <input type="text"/> <input type="text"/> años
44. ¿Alimentó con leche materna a su(s) hijo(s)? Si <input type="checkbox"/> No <input type="checkbox"/> ¿Cuánto tiempo alimento con leche materna a su(s) hijo(s)? 1. Hijo _____ (meses) 2. Hijo _____ (meses) 3. Hijo _____ (meses) 4. Hijo _____ (meses) 5. Hijo _____ (meses) 6. Mas? Si <input type="checkbox"/> No <input type="checkbox"/> Si la respuesta es "sí", sumé el tiempo (aproximadamente) de los restantes en que los amamanto _____
45. ¿Ha sido diagnosticada alguna vez alguno de estos canceres?: Mama Si <input type="checkbox"/> No <input type="checkbox"/> Ovario Si <input type="checkbox"/> No <input type="checkbox"/> Útero Si <input type="checkbox"/> No <input type="checkbox"/>
46. ¿Ha transcurrido más de 12 meses desde su última menstruación? Si <input type="checkbox"/> No <input type="checkbox"/>
47. ¿Cuál fue la fecha de su última regla (el primer día de sangrado o menstruación)? (/ /) dd/mm/aa
48. ¿Recibió tratamiento hormonal para la menopausia? Si <input type="checkbox"/> No <input type="checkbox"/> Si su respuesta es "Si", ¿Durante cuanto tiempo los uso o los ha venido usando? _____ Si recuerda, qué tipo de medicamento ¿usó? _____
49. ¿Recibe tratamiento hormonal para la menopausia? Si <input type="checkbox"/> No <input type="checkbox"/> Si su respuesta es "Si", ¿Durante cuanto tiempo los ha venido usando? _____ Si recuerda, qué tipo de medicamento ¿usó? _____
50. ¿Le han tomado mamografías en los últimos 3 años? Si <input type="checkbox"/> No <input type="checkbox"/> Si su respuesta es "Si", ¿la última mamografía fue tomada? En el último año <input type="checkbox"/> Entre 1 y 2 años <input type="checkbox"/> Entre 2 y 3 años <input type="checkbox"/>



Universidad
Industrial de
Santander

51. ¿Alguna vez le han practicado cirugía o biopsia en los senos? Si <input type="checkbox"/> No <input type="checkbox"/> → Si "No" pase a la pregunta 54
52. ¿Hace cuanto tiempo le realizaron la última intervención (cirugía o biopsia)? En los últimos 3 meses <input type="checkbox"/> Entre 3 y 12 meses <input type="checkbox"/> Entre 1 y 2 años <input type="checkbox"/> ¿Más años?¿Cuantos? _____
53. ¿En que seno le realizaron el procedimiento (cirugía o biopsia)? Derecho <input type="checkbox"/> Izquierdo <input type="checkbox"/>
54. ¿Ha recibido alguna vez tratamiento de radioterapia? Si <input type="checkbox"/> No <input type="checkbox"/> Si la respuesta es "si", por cuanto tiempo? _____
INFORMACION DEL ESTADO DURANTE LA TOMA DE LA TERMOGRAFIA
55. ¿Ha realizado alguna actividad (caminata, bronceado, etc.) donde se haya expuesto prolongadamente al sol durante los últimos 5 días? Si <input type="checkbox"/> No <input type="checkbox"/> Si la respuesta es "si" ¿Por cuánto tiempo? _____ horas
56. ¿Hoy usó lociones, cremas, polvos o algún tipo de maquillaje en el área? Si <input type="checkbox"/> No <input type="checkbox"/> ¿Cuál? _____
57. ¿Hoy usó desodorante o antitranspirante? Si <input type="checkbox"/> No <input type="checkbox"/>
58. ¿Realizó alguna terapia física en las últimas 24 horas? Si <input type="checkbox"/> No <input type="checkbox"/>
59. ¿Realizó algún ejercicio físico 4 horas antes del examen? Si <input type="checkbox"/> No <input type="checkbox"/>
60. ¿Ha tomado algún medicamento para el dolor o vaso dilatador el día del examen? Si <input type="checkbox"/> No <input type="checkbox"/>



**REGISTRO NO 6
REGISTRO DE LA INFORMACIÓN CLÍNICA**

**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE SALUD
ESCUELA DE MEDICINA**

Se guardará la confidencialidad de los datos y en ningún momento se revelará la identificación de los pacientes.

NOTA: Es importante que al registrar los datos tenga en cuenta las unidades de medida, marque con una X la respuesta en la casilla que corresponda y evite dejar espacios en blanco.

Nombre (s) y Apellidos:																			
Dirección:	Barrio:																		
Tél:	Fecha hoy dd/mm/aa (/ /)																		
Código de las Imágenes: Frontal: _____ Lateral Der. _____ Lateral Izq _____ Oblicua Der. _____ Oblicua Izq _____																			
No. de Historia Clínica:	Cédula:																		
Estatura: _____	Peso: _____																		
IMC: _____																			
DATOS DEL EXÁMEN CLÍNICO																			
1. ¿La paciente presenta dolor en las glándulas mamarias? Si <input type="checkbox"/> No <input type="checkbox"/> Si la respuesta es "sí", ¿En cuál glándula mamaria? y en que región de la glándula mamaria?																			
<table border="0"> <tr> <td align="center" colspan="4">Derecha</td> <td align="center" colspan="4">Izquierda</td> </tr> <tr> <td>CSE <input type="checkbox"/></td> <td>CSI <input type="checkbox"/></td> <td>CIE <input type="checkbox"/></td> <td>CII <input type="checkbox"/></td> <td>Centro <input type="checkbox"/></td> <td>CSE <input type="checkbox"/></td> <td>CSI <input type="checkbox"/></td> <td>CIE <input type="checkbox"/></td> <td>CII <input type="checkbox"/></td> <td>Centro <input type="checkbox"/></td> </tr> </table>		Derecha				Izquierda				CSE <input type="checkbox"/>	CSI <input type="checkbox"/>	CIE <input type="checkbox"/>	CII <input type="checkbox"/>	Centro <input type="checkbox"/>	CSE <input type="checkbox"/>	CSI <input type="checkbox"/>	CIE <input type="checkbox"/>	CII <input type="checkbox"/>	Centro <input type="checkbox"/>
Derecha				Izquierda															
CSE <input type="checkbox"/>	CSI <input type="checkbox"/>	CIE <input type="checkbox"/>	CII <input type="checkbox"/>	Centro <input type="checkbox"/>	CSE <input type="checkbox"/>	CSI <input type="checkbox"/>	CIE <input type="checkbox"/>	CII <input type="checkbox"/>	Centro <input type="checkbox"/>										
2. ¿La paciente se ha palpado masa(s) en la(s) glándula(s) mamaria(s)? Si <input type="checkbox"/> No <input type="checkbox"/> Si la respuesta es "sí", ¿En cuál glándula mamaria? y en que región de la glándula mamaria?																			
<table border="0"> <tr> <td align="center" colspan="4">Derecha</td> <td align="center" colspan="4">Izquierda</td> </tr> <tr> <td>CSE <input type="checkbox"/></td> <td>CSI <input type="checkbox"/></td> <td>CIE <input type="checkbox"/></td> <td>CII <input type="checkbox"/></td> <td>Centro <input type="checkbox"/></td> <td>CSE <input type="checkbox"/></td> <td>CSI <input type="checkbox"/></td> <td>CIE <input type="checkbox"/></td> <td>CII <input type="checkbox"/></td> <td>Centro <input type="checkbox"/></td> </tr> </table>		Derecha				Izquierda				CSE <input type="checkbox"/>	CSI <input type="checkbox"/>	CIE <input type="checkbox"/>	CII <input type="checkbox"/>	Centro <input type="checkbox"/>	CSE <input type="checkbox"/>	CSI <input type="checkbox"/>	CIE <input type="checkbox"/>	CII <input type="checkbox"/>	Centro <input type="checkbox"/>
Derecha				Izquierda															
CSE <input type="checkbox"/>	CSI <input type="checkbox"/>	CIE <input type="checkbox"/>	CII <input type="checkbox"/>	Centro <input type="checkbox"/>	CSE <input type="checkbox"/>	CSI <input type="checkbox"/>	CIE <input type="checkbox"/>	CII <input type="checkbox"/>	Centro <input type="checkbox"/>										
3. La paciente presenta asimetría en la inspección de las glándulas mamarias? Si <input type="checkbox"/> No <input type="checkbox"/>																			
4. ¿La paciente presenta Telorrea? Si <input type="checkbox"/> No <input type="checkbox"/> Si la respuesta es "sí" ¿La Telorrea es espontánea? Si <input type="checkbox"/> No <input type="checkbox"/> Si hay telorrea, tipo de secreción: Sanguinolenta <input type="checkbox"/> Verdosa <input type="checkbox"/> Blanquecina <input type="checkbox"/> Serosa <input type="checkbox"/>																			
5. ¿La paciente presenta masa(s) con dimensiones superiores a 2 cm en la glándula mamaria? Si <input type="checkbox"/> Posiblemente <input type="checkbox"/> No <input type="checkbox"/> Si la respuesta es "sí" o "Posiblemente", ¿En cuál glándula mamaria? y en que región de la glándula mamaria?																			



Universidad Industrial de Santander

	Derecha	Izquierda
	CSE <input type="checkbox"/> CSI <input type="checkbox"/> CIE <input type="checkbox"/> CII <input type="checkbox"/> Centro <input type="checkbox"/>	CSE <input type="checkbox"/> CSI <input type="checkbox"/> CIE <input type="checkbox"/> CII <input type="checkbox"/> Centro <input type="checkbox"/>
	¿Las masa(s) es (son)? Móvil(es) <input type="checkbox"/> Fija(s) <input type="checkbox"/>	
6.	¿La paciente presenta alteraciones en la piel de la glándula mamaria? Si <input type="checkbox"/> No <input type="checkbox"/> Si la respuesta es "sí", ¿Qué tipo de anomalía existe? Eritema <input type="checkbox"/> Edema(piel de naranja) <input type="checkbox"/> Nodulaciones <input type="checkbox"/> Retracción <input type="checkbox"/> Ulceración <input type="checkbox"/> ¿En cuál mama y en qué región presenta la anomalía de la piel? Derecha Izquierda CSE <input type="checkbox"/> CSI <input type="checkbox"/> CIE <input type="checkbox"/> CII <input type="checkbox"/> Centro <input type="checkbox"/> CSE <input type="checkbox"/> CSI <input type="checkbox"/> CIE <input type="checkbox"/> CII <input type="checkbox"/> Centro <input type="checkbox"/>	
7.	¿La paciente presenta anomalías en los nódulos linfoides axilares? Si <input type="checkbox"/> No <input type="checkbox"/> Si la respuesta es "sí", ¿Qué nivel de anomalía se encontró? N1(menores a 2 cm) <input type="checkbox"/> N2 (mayores a 2 cm) <input type="checkbox"/> N3 <input type="checkbox"/>	
8.	¿La paciente presenta nódulo (s) supraclavicular(es)? Si <input type="checkbox"/> No <input type="checkbox"/> Si la respuesta es "sí", ¿Qué nivel de anomalía se encontró? N1(menores a 2 cm) <input type="checkbox"/> N2 (mayores a 2 cm) <input type="checkbox"/> N3 <input type="checkbox"/>	
9.	Tiene resultado de mamografía? Si <input type="checkbox"/> No <input type="checkbox"/> Si la respuesta es "sí", Cuál es el resultado de la mamografía? BIRADS 0 <input type="checkbox"/> BIRADS I <input type="checkbox"/> BIRADS II <input type="checkbox"/> BIRADS III <input type="checkbox"/> BIRADS IV <input type="checkbox"/> BIRADS V <input type="checkbox"/>	

Anexo B. RESULTADO DE LA SIMULACIÓN REDES BACKPROPAGATION

Grupo de entrenamiento	Tipo RNA	Función de transferencia		Algoritmo de aprendizaje	Neuronas ocultas	No. Épocas	Mejor Época	Error Validación	Tiempo Entrenamiento (s)	Acierto (%)	Sensibilidad (%)	Especificidad (%)	AUC
		Capa oculta	Capa de salida										
1	BP	tansig	purelin	trainlm	5	8	2	0,2403	0,578	75,00	33,33	100,00	0,6667
1	BP	tansig	purelin	trainlm	6	15	9	0,2698	0,406	75,00	66,67	80,00	0,7333
1	BP	tansig	purelin	trainlm	7	10	4	0,2707	0,344	79,17	77,78	80,00	0,7889
1	BP	tansig	purelin	trainlm	8	10	4	0,2390	0,329	75,00	55,56	86,67	0,7111
1	BP	tansig	purelin	trainlm	9	9	3	0,2438	0,296	73,91	55,56	85,71	0,7259
1	BP	tansig	purelin	trainlm	10	8	2	0,1855	0,297	79,17	44,44	100,00	0,7222
1	BP	tansig	purelin	trainlm	11	13	7	0,2156	0,391	86,96	77,78	92,86	0,8556
1	BP	tansig	purelin	trainlm	12	10	4	0,2205	0,343	79,17	77,78	80,00	0,7889
1	BP	tansig	purelin	trainlm	13	9	3	0,2329	0,328	79,17	77,78	80,00	0,7889
1	BP	tansig	purelin	trainlm	14	10	4	0,2766	0,344	79,17	44,44	100,00	0,7222
1	BP	tansig	purelin	trainlm	15	8	2	0,2899	0,328	70,83	55,56	80,00	0,6778
1	BP	tansig	purelin	trainlm	16	7	1	0,2929	0,282	75,00	55,56	86,67	0,7111
1	BP	tansig	purelin	trainlm	17	8	2	0,3006	0,313	79,17	66,67	86,67	0,7667
1	BP	tansig	purelin	trainlm	18	9	3	0,2348	0,328	83,33	55,56	100,00	0,7778
1	BP	tansig	purelin	trainlm	19	8	2	0,2566	0,313	75,00	66,67	80,00	0,7333
1	BP	tansig	purelin	trainlm	20	9	3	0,2912	0,343	70,83	66,67	73,33	0,7000
1	BP	tansig	tansig	trainlm	5	9	3	0,2215	0,3130	79,17	44,44	100,00	0,7222
1	BP	tansig	tansig	trainlm	6	12	6	0,2290	0,3600	75,00	66,67	80,00	0,7333
1	BP	tansig	tansig	trainlm	7	14	8	0,2354	0,3910	75,00	66,67	80,00	0,7333
1	BP	tansig	tansig	trainlm	8	8	2	0,2094	0,2970	75,00	44,44	93,33	0,6889
1	BP	tansig	tansig	trainlm	9	9	3	0,2187	0,3130	79,17	66,67	86,67	0,7667
1	BP	tansig	tansig	trainlm	10	9	3	0,2896	0,2970	70,83	55,56	80,00	0,6778
1	BP	tansig	tansig	trainlm	11	7	1	0,2241	0,2660	75,00	66,67	80,00	0,7333
1	BP	tansig	tansig	trainlm	12	15	9	0,3780	0,4060	70,83	66,67	73,33	0,7000
1	BP	tansig	tansig	trainlm	13	2	1	0,4018	0,2660	75,00	77,78	73,33	0,7556
1	BP	tansig	tansig	trainlm	14	7	1	0,2136	0,2970	70,83	55,56	80,00	0,6778
1	BP	tansig	tansig	trainlm	15	11	5	0,2650	0,3600	79,17	55,56	93,33	0,7444
1	BP	tansig	tansig	trainlm	16	11	5	0,2651	0,3440	75,00	33,33	100,00	0,6667
1	BP	tansig	tansig	trainlm	17	9	3	0,2440	0,3440	70,83	44,44	86,67	0,6556
1	BP	tansig	tansig	trainlm	18	10	4	0,2850	0,3590	66,67	55,56	73,33	0,6444
1	BP	tansig	tansig	trainlm	19	10	4	0,2479	0,3440	70,83	66,67	73,33	0,7000
1	BP	tansig	tansig	trainlm	20	11	5	0,2770	0,3910	83,33	66,67	93,33	0,8000
1	BP	tansig	purelin	trainrp	5	12	6	0,2327	0,328	79,17	55,56	93,33	0,7444

Grupo de entrenamiento	Tipo RNA	Función de transferencia		Algoritmo de aprendizaje	Neuronas ocultas	No. Épocas	Mejor Época	Error Validación	Tiempo Entrenamiento (s)	Acierto (%)	Sensibilidad (%)	Especificidad (%)	AUC
		Capa oculta	Capa de salida										
1	BP	tansig	purelin	trainrp	6	10	4	0,2020	0,297	75,00	44,44	93,33	0,6889
1	BP	tansig	purelin	trainrp	7	12	6	0,1776	0,312	79,17	55,56	93,33	0,7444
1	BP	tansig	purelin	trainrp	8	10	4	0,2182	0,281	70,83	33,33	93,33	0,6333
1	BP	tansig	purelin	trainrp	9	12	6	0,2303	0,312	79,17	55,56	93,33	0,7444
1	BP	tansig	purelin	trainrp	10	15	9	0,2175	0,36	66,67	55,56	73,33	0,6444
1	BP	tansig	purelin	trainrp	11	20	14	0,2092	0,406	70,83	55,56	80,00	0,6778
1	BP	tansig	purelin	trainrp	12	18	12	0,2463	0,391	83,33	66,67	93,33	0,8000
1	BP	tansig	purelin	trainrp	13	11	5	0,1833	0,313	79,17	55,56	93,33	0,7444
1	BP	tansig	purelin	trainrp	14	24	18	0,2458	0,438	75,00	66,67	80,00	0,7333
1	BP	tansig	purelin	trainrp	15	10	4	0,2158	0,313	83,33	55,56	100,00	0,7778
1	BP	tansig	purelin	trainrp	16	13	7	0,2176	0,328	70,83	33,33	93,33	0,6333
1	BP	tansig	purelin	trainrp	17	17	11	0,2411	0,391	79,17	55,56	93,33	0,7444
1	BP	tansig	purelin	trainrp	18	19	13	0,1962	0,406	75,00	77,78	73,33	0,7556
1	BP	tansig	purelin	trainrp	19	16	10	0,2163	0,375	66,67	55,56	73,33	0,6444
1	BP	tansig	purelin	trainrp	20	12	6	0,1818	0,313	75,00	66,67	80,00	0,7333
1	BP	tansig	tansig	trainrp	5	13	7	0,1698	0,312	79,17	55,56	93,33	0,7444
1	BP	tansig	tansig	trainrp	6	6	0	0,2255	0,265	70,83	55,56	80,00	0,6778
1	BP	tansig	tansig	trainrp	7	18	12	0,2002	0,391	75,00	66,67	80,00	0,7333
1	BP	tansig	tansig	trainrp	8	16	10	0,1697	0,359	75,00	55,56	86,67	0,7111
1	BP	tansig	tansig	trainrp	9	18	12	0,2016	0,391	70,83	55,56	80,00	0,6778
1	BP	tansig	tansig	trainrp	10	22	16	0,1821	0,406	75,00	66,67	80,00	0,7333
1	BP	tansig	tansig	trainrp	11	22	16	0,2420	0,421	70,83	55,56	80,00	0,6778
1	BP	tansig	tansig	trainrp	12	17	11	0,2108	0,375	75,00	66,67	80,00	0,7333
1	BP	tansig	tansig	trainrp	13	11	5	0,1913	0,297	75,00	66,67	80,00	0,7333
1	BP	tansig	tansig	trainrp	14	16	10	0,2186	0,375	79,17	55,56	93,33	0,7444
1	BP	tansig	tansig	trainrp	15	12	6	0,2008	0,312	70,83	66,67	73,33	0,7000
1	BP	tansig	tansig	trainrp	16	15	9	0,1843	0,359	75,00	55,56	86,67	0,7111
1	BP	tansig	tansig	trainrp	17	9	3	0,2388	0,282	66,67	77,78	60,00	0,6889
1	BP	tansig	tansig	trainrp	18	17	11	0,2400	0,39	70,83	66,67	73,33	0,7000
1	BP	tansig	tansig	trainrp	19	11	5	0,2242	0,312	70,83	55,56	80,00	0,6778
1	BP	tansig	tansig	trainrp	20	16	10	0,2466	0,375	66,67	55,56	73,33	0,6444
2	BP	tansig	purelin	trainlm	5	11	5	0,2385	0,343	87,50	66,67	100,00	0,8333
2	BP	tansig	purelin	trainlm	6	11	5	0,2326	0,343	83,33	77,78	86,67	0,8222
2	BP	tansig	purelin	trainlm	7	10	4	0,2476	0,359	79,17	55,56	93,33	0,7444
2	BP	tansig	purelin	trainlm	8	9	3	0,2242	0,312	75,00	55,56	86,67	0,7111
2	BP	tansig	purelin	trainlm	9	9	3	0,3119	0,313	75,00	66,67	80,00	0,7333

Grupo de entrenamiento	Tipo RNA	Función de transferencia		Algoritmo de aprendizaje	Neuronas ocultas	No. Épocas	Mejor Época	Error Validación	Tiempo Entrenamiento (s)	Acierto (%)	Sensibilidad (%)	Especificidad (%)	AUC
		Capa oculta	Capa de salida										
2	BP	tansig	purelin	trainlm	10	9	3	0,2529	0,344	75,00	33,33	100,00	0,6667
2	BP	tansig	purelin	trainlm	11	10	4	0,3214	0,328	79,17	66,67	86,67	0,7667
2	BP	tansig	purelin	trainlm	12	10	4	0,2247	0,36	83,33	55,56	100,00	0,7778
2	BP	tansig	purelin	trainlm	13	14	8	0,2607	0,422	68,18	55,56	76,92	0,7074
2	BP	tansig	purelin	trainlm	14	8	2	0,3467	0,312	66,67	88,89	53,33	0,7111
2	BP	tansig	purelin	trainlm	15	8	2	0,2774	0,313	83,33	55,56	100,00	0,7778
2	BP	tansig	purelin	trainlm	16	8	2	0,3279	0,313	79,17	77,78	80,00	0,7889
2	BP	tansig	purelin	trainlm	17	11	5	0,3740	0,375	69,57	77,78	64,29	0,7296
2	BP	tansig	purelin	trainlm	18	8	2	0,2391	0,313	79,17	55,56	93,33	0,7444
2	BP	tansig	purelin	trainlm	19	9	3	0,2658	0,406	75,00	44,44	93,33	0,6889
2	BP	tansig	purelin	trainlm	20	9	3	0,3570	0,375	75,00	44,44	93,33	0,6889
2	BP	tansig	tansig	trainlm	5	12	6	0,1906	0,359	75,00	55,56	86,67	0,7111
2	BP	tansig	tansig	trainlm	6	9	3	0,2398	0,313	79,17	55,56	93,33	0,7444
2	BP	tansig	tansig	trainlm	7	22	16	0,3409	0,531	75,00	44,44	93,33	0,6889
2	BP	tansig	tansig	trainlm	8	11	5	0,2223	0,359	87,50	66,67	100,00	0,8333
2	BP	tansig	tansig	trainlm	9	25	19	0,1922	0,594	79,17	66,67	86,67	0,7667
2	BP	tansig	tansig	trainlm	10	9	3	0,2482	0,312	62,50	77,78	53,33	0,6556
2	BP	tansig	tansig	trainlm	11	10	4	0,2373	0,328	83,33	55,56	100,00	0,7778
2	BP	tansig	tansig	trainlm	12	9	3	0,2198	0,328	75,00	55,56	86,67	0,7111
2	BP	tansig	tansig	trainlm	13	10	4	0,2386	0,329	79,17	44,44	100,00	0,7222
2	BP	tansig	tansig	trainlm	14	10	4	0,2385	0,344	83,33	66,67	93,33	0,8000
2	BP	tansig	tansig	trainlm	15	9	3	0,2419	0,343	70,83	66,67	73,33	0,7000
2	BP	tansig	tansig	trainlm	16	11	5	0,2028	0,375	75,00	55,56	86,67	0,7111
2	BP	tansig	tansig	trainlm	17	11	5	0,2349	0,407	87,50	77,78	93,33	0,8630
2	BP	tansig	tansig	trainlm	18	11	5	0,2138	0,406	83,33	55,56	100,00	0,7778
2	BP	tansig	tansig	trainlm	19	10	4	0,2158	0,406	70,83	33,33	93,33	0,6333
2	BP	tansig	tansig	trainlm	20	14	8	0,3424	0,485	79,17	55,56	93,33	0,7444
2	BP	tansig	purelin	trainrp	5	25	19	0,2068	0,468	75,00	55,56	86,67	0,7111
2	BP	tansig	purelin	trainrp	6	13	7	0,2486	0,328	83,33	55,56	100,00	0,7778
2	BP	tansig	purelin	trainrp	7	17	11	0,2324	0,375	75,00	44,44	93,33	0,6889
2	BP	tansig	purelin	trainrp	8	15	9	0,2058	0,36	79,17	55,56	93,33	0,7444
2	BP	tansig	purelin	trainrp	9	17	11	0,2329	0,375	83,33	66,67	93,33	0,8000
2	BP	tansig	purelin	trainrp	10	17	11	0,2019	0,375	79,17	66,67	86,67	0,7667
2	BP	tansig	purelin	trainrp	11	36	30	0,2178	0,578	79,17	55,56	93,33	0,7444

Grupo de entrenamiento	Tipo RNA	Función de transferencia		Algoritmo de aprendizaje	Neuronas ocultas	No. Épocas	Mejor Época	Error Validación	Tiempo Entrenamiento (s)	Acierto (%)	Sensibilidad (%)	Especificidad (%)	AUC
		Capa oculta	Capa de salida										
2	BP	tansig	purelin	trainrp	12	9	3	0,2412	0,281	75,00	55,56	86,67	0,7111
2	BP	tansig	purelin	trainrp	13	10	4	0,2308	0,297	79,17	66,67	86,67	0,7667
2	BP	tansig	purelin	trainrp	14	28	22	0,2063	0,516	73,91	50,00	86,67	0,7185
2	BP	tansig	purelin	trainrp	15	15	9	0,2406	0,375	79,17	44,44	100,00	0,7222
2	BP	tansig	purelin	trainrp	16	25	19	0,2890	0,453	75,00	66,67	80,00	0,7333
2	BP	tansig	purelin	trainrp	17	12	6	0,2668	0,312	65,22	50,00	73,33	0,6593
2	BP	tansig	purelin	trainrp	18	17	11	0,2198	0,375	79,17	55,56	93,33	0,7444
2	BP	tansig	purelin	trainrp	19	12	6	0,2281	0,328	75,00	44,44	93,33	0,6889
2	BP	tansig	purelin	trainrp	20	13	7	0,2530	0,359	75,00	66,67	80,00	0,7333
2	BP	tansig	tansig	trainrp	5	37	31	0,1909	0,578	66,67	66,67	66,67	0,6667
2	BP	tansig	tansig	trainrp	6	16	10	0,1950	0,375	75,00	55,56	86,67	0,7111
2	BP	tansig	tansig	trainrp	7	22	16	0,2263	0,437	79,17	66,67	86,67	0,7667
2	BP	tansig	tansig	trainrp	8	21	15	0,2196	0,422	79,17	55,56	93,33	0,7444
2	BP	tansig	tansig	trainrp	9	13	7	0,2187	0,343	75,00	44,44	93,33	0,6889
2	BP	tansig	tansig	trainrp	10	34	28	0,1922	0,547	75,00	44,44	93,33	0,6889
2	BP	tansig	tansig	trainrp	11	14	8	0,2259	0,328	75,00	66,67	80,00	0,7333
2	BP	tansig	tansig	trainrp	12	14	8	0,1963	0,344	79,17	44,44	100,00	0,7222
2	BP	tansig	tansig	trainrp	13	33	27	0,2015	0,547	79,17	66,67	86,67	0,7667
2	BP	tansig	tansig	trainrp	14	21	15	0,2008	0,421	75,00	44,44	93,33	0,6889
2	BP	tansig	tansig	trainrp	15	35	29	0,1927	0,563	75,00	55,56	86,67	0,7111
2	BP	tansig	tansig	trainrp	16	18	12	0,2089	0,39	75,00	55,56	86,67	0,7111
2	BP	tansig	tansig	trainrp	17	16	10	0,2189	0,375	79,17	44,44	100,00	0,7222
2	BP	tansig	tansig	trainrp	18	30	24	0,2259	0,531	75,00	55,56	86,67	0,7111
2	BP	tansig	tansig	trainrp	19	13	7	0,2118	0,344	79,17	44,44	100,00	0,7222
2	BP	tansig	tansig	trainrp	20	14	8	0,1975	0,343	83,33	55,56	100,00	0,7778

Abreviaturas: BP: *Backpropagation*, RNA: Red Neuronal Artificial, tansig: Función de transferencia sigmoidea tangente hiperbólica, purelin: Función de transferencia Lineal, trainlm: Algoritmo de aprendizaje *Levenberg-Marquardt*, trainrp: Algoritmo de entrenamiento *Resilient Backpropagation*, AUC: Área bajo la curva ROC

Anexo C. RESULTADO DE LA SIMULACIÓN REDES RBFN

Grupo de entrenamiento	Tipo de RNA	Spread	Verificación de entrenamiento								Simulación grupo de prueba							
			VP	FP	VN	FN	Acierto (%)	Sensibilidad (%)	Especificidad (%)	AUC	VP	FP	VN	FN	Acierto (%)	Sensibilidad (%)	Especificidad (%)	AUC
1	PNN	0,030	28	0	44	0	100,00	100,00	100,00	1	9	15	0	0	37,50	100,00	0,00	0,5000
1	PNN	0,033	28	0	44	0	100,00	100,00	100,00	1	9	14	1	0	41,67	100,00	6,67	0,5333
1	PNN	0,045	28	0	44	0	100,00	100,00	100,00	1	9	13	2	0	45,83	100,00	13,33	0,5667
1	PNN	0,054	28	0	44	0	100,00	100,00	100,00	1	9	12	3	0	50,00	100,00	20,00	0,6000
1	PNN	0,069	28	0	44	0	100,00	100,00	100,00	1	8	11	4	1	50,00	88,89	26,67	0,5778
1	PNN	0,075	28	0	44	0	100,00	100,00	100,00	1	8	10	5	1	54,17	88,89	33,33	0,6111
1	PNN	0,081	28	0	44	0	100,00	100,00	100,00	1	8	7	8	1	66,67	88,89	53,33	0,7111
1	PNN	0,084	28	0	44	0	100,00	100,00	100,00	1	7	7	8	2	62,50	77,78	53,33	0,6556
1	PNN	0,087	28	0	44	0	100,00	100,00	100,00	1	5	6	9	4	58,33	55,56	60,00	0,5778
1	PNN	0,090	28	0	44	0	100,00	100,00	100,00	1	5	5	10	4	62,50	55,56	66,67	0,6111
1	PNN	0,100	28	0	44	0	100,00	100,00	100,00	1	5	5	10	4	62,50	55,56	66,67	0,6111
1	PNN	0,200	28	0	44	0	100,00	100,00	100,00	1	4	3	12	5	66,67	44,44	80,00	0,6222
1	PNN	0,300	28	0	44	0	100,00	100,00	100,00	1	3	3	12	6	62,50	33,33	80,00	0,5667
1	PNN	0,400	28	0	44	0	100,00	100,00	100,00	1	3	3	12	6	62,50	33,33	80,00	0,5667
1	PNN	0,500	28	0	44	0	100,00	100,00	100,00	1	3	3	12	6	62,50	33,33	80,00	0,5667
1	PNN	0,600	28	0	44	0	100,00	100,00	100,00	1	3	3	12	6	62,50	33,33	80,00	0,5667
1	PNN	0,700	28	0	44	0	100,00	100,00	100,00	1	3	3	12	6	62,50	33,33	80,00	0,5667
1	PNN	0,800	28	0	44	0	100,00	100,00	100,00	1	3	2	13	6	66,67	33,33	86,67	0,6000
1	PNN	0,900	28	0	44	0	100,00	100,00	100,00	1	3	2	13	6	66,67	33,33	86,67	0,6000
1	PNN	1,000	28	0	44	0	100,00	100,00	100,00	1	3	2	13	6	66,67	33,33	86,67	0,6000
1	PNN	1,100	28	0	44	0	100,00	100,00	100,00	1	2	2	13	7	62,50	22,22	86,67	0,5444
1	PNN	1,200	28	0	44	0	100,00	100,00	100,00	1	2	2	13	7	62,50	22,22	86,67	0,5444
1	PNN	1,300	28	0	44	0	100,00	100,00	100,00	1	2	2	13	7	62,50	22,22	86,67	0,5444
1	PNN	1,400	28	0	44	0	100,00	100,00	100,00	1	2	2	13	7	62,50	22,22	86,67	0,5444
1	PNN	1,500	27	0	44	1	98,61	96,43	100,00	0,9821	2	2	13	7	62,50	22,22	86,67	0,5444
1	PNN	1,600	27	0	44	1	98,61	96,43	100,00	0,9821	2	2	13	7	62,50	22,22	86,67	0,5444
1	PNN	1,700	27	0	44	1	98,61	96,43	100,00	0,9821	2	2	13	7	62,50	22,22	86,67	0,5444
1	PNN	1,800	27	0	44	1	98,61	96,43	100,00	0,9821	2	2	13	7	62,50	22,22	86,67	0,5444
1	PNN	1,900	25	0	44	3	95,83	89,29	100,00	0,9464	2	2	13	7	62,50	22,22	86,67	0,5444
1	PNN	2,000	24	0	44	4	94,44	85,71	100,00	0,9286	2	2	13	7	62,50	22,22	86,67	0,5444
1	PNN	2,100	24	0	44	4	94,44	85,71	100,00	0,9286	2	2	13	7	62,50	22,22	86,67	0,5444
1	PNN	2,200	24	0	44	4	94,44	85,71	100,00	0,9286	2	2	13	7	62,50	22,22	86,67	0,5444

Grupo de entrenamiento	Tipo de RNA	Spread	Verificación de entrenamiento								Simulación grupo de prueba							
			VP	FP	VN	FN	Acierto (%)	Sensibilidad (%)	Especificidad (%)	AUC	VP	FP	VN	FN	Acierto (%)	Sensibilidad (%)	Especificidad (%)	AUC
1	PNN	2,300	23	0	44	5	93,06	82,14	100,00	0,9107	2	2	13	7	62,50	22,22	86,67	0,5444
1	PNN	2,400	23	0	44	5	93,06	82,14	100,00	0,9107	2	2	13	7	62,50	22,22	86,67	0,5444
1	PNN	2,500	21	1	43	7	88,89	75,00	97,73	0,8636	2	2	13	7	62,50	22,22	86,67	0,5444
1	PNN	2,600	21	1	43	7	88,89	75,00	97,73	0,8636	2	2	13	7	62,50	22,22	86,67	0,5444
1	PNN	2,700	21	1	43	7	88,89	75,00	97,73	0,8636	2	2	13	7	62,50	22,22	86,67	0,5444
1	PNN	2,800	19	1	43	9	86,11	67,86	97,73	0,8279	2	2	13	7	62,50	22,22	86,67	0,5444
1	PNN	2,900	17	1	43	11	83,33	60,71	97,73	0,7922	2	2	13	7	62,50	22,22	86,67	0,5444
1	PNN	3,000	16	1	43	12	81,94	57,14	97,73	0,7744	2	2	13	7	62,50	22,22	86,67	0,5444
2	PNN	0,0030	28	0	44	0	100,00	100,00	100,00	1	9	15	0	0	37,5	100	0	0,5000
2	PNN	0,0540	28	0	44	0	100,00	100,00	100,00	1	9	14	1	0	41,67	100,00	6,67	0,5333
2	PNN	0,0570	28	0	44	0	100,00	100,00	100,00	1	9	13	2	0	45,83	100,00	13,33	0,5667
2	PNN	0,0600	28	0	44	0	100,00	100,00	100,00	1	9	12	3	0	50,00	100,00	20,00	0,6000
2	PNN	0,0690	28	0	44	0	100,00	100,00	100,00	1	8	12	3	1	45,83	88,89	20,00	0,5444
2	PNN	0,0720	28	0	44	0	100,00	100,00	100,00	1	8	11	4	1	50,00	88,89	26,67	0,5778
2	PNN	0,0810	28	0	44	0	100,00	100,00	100,00	1	8	10	5	1	54,17	88,89	33,33	0,6111
2	PNN	0,0840	28	0	44	0	100,00	100,00	100,00	1	8	10	5	1	54,17	88,89	33,33	0,6111
2	PNN	0,0900	28	0	44	0	100,00	100,00	100,00	1	8	9	6	1	58,33	88,89	40,00	0,6444
2	PNN	0,1000	28	0	44	0	100,00	100,00	100,00	1	8	8	7	1	62,50	88,89	46,67	0,6778
2	PNN	0,1020	28	0	44	0	100,00	100,00	100,00	1	8	8	7	1	62,50	88,89	46,67	0,6778
2	PNN	0,1040	28	0	44	0	100,00	100,00	100,00	1	8	7	8	1	66,67	88,89	53,33	0,7111
2	PNN	0,1060	28	0	44	0	100,00	100,00	100,00	1	7	7	8	2	62,50	77,78	53,33	0,6556
2	PNN	0,1120	28	0	44	0	100,00	100,00	100,00	1	6	6	9	3	62,50	66,67	60,00	0,6333
2	PNN	0,1200	28	0	44	0	100,00	100,00	100,00	1	5	6	9	4	58,33	55,56	60,00	0,5778
2	PNN	0,1400	28	0	44	0	100,00	100,00	100,00	1	5	5	10	4	62,50	55,56	66,67	0,6111
2	PNN	0,2000	28	0	44	0	100,00	100,00	100,00	1	2	4	11	7	54,17	22,22	73,33	0,4778
2	PNN	0,3000	28	0	44	0	100,00	100,00	100,00	1	2	3	12	7	58,33	22,22	80,00	0,5111
2	PNN	0,4000	28	0	44	0	100,00	100,00	100,00	1	2	3	12	7	58,33	22,22	80,00	0,5111
2	PNN	0,5000	28	0	44	0	100,00	100,00	100,00	1	2	3	12	7	58,33	22,22	80,00	0,5111
2	PNN	0,6000	28	0	44	0	100,00	100,00	100,00	1	2	3	12	7	58,33	22,22	80,00	0,5111

Grupo de entrenamiento	Tipo de RNA	Spread	Verificación de entrenamiento								Simulación grupo de prueba							
			VP	FP	VN	FN	Acierto (%)	Sensibilidad (%)	Especificidad (%)	AUC	VP	FP	VN	FN	Acierto (%)	Sensibilidad (%)	Especificidad (%)	AUC
2	PNN	0,7000	28	0	44	0	100,00	100,00	100,00	1	2	3	12	7	58,33	22,22	80,00	0,5111
2	PNN	0,8000	28	0	44	0	100,00	100,00	100,00	1	2	3	12	7	58,33	22,22	80,00	0,5111
2	PNN	0,9000	28	0	44	0	100,00	100,00	100,00	1	2	3	12	7	58,33	22,22	80,00	0,5111
2	PNN	1,0000	28	0	44	0	100,00	100,00	100,00	1	2	3	12	7	58,33	22,22	80,00	0,5111
2	PNN	1,1000	28	0	44	0	100,00	100,00	100,00	1	2	3	12	7	58,33	22,22	80,00	0,5111
2	PNN	1,2000	28	0	44	0	100,00	100,00	100,00	1	2	3	12	7	58,33	22,22	80,00	0,5111
2	PNN	1,3000	28	0	44	0	100,00	100,00	100,00	1	2	3	12	7	58,33	22,22	80,00	0,5111
2	PNN	1,4000	28	0	44	0	100,00	100,00	100,00	1	2	3	12	7	58,33	22,22	80,00	0,5111
2	PNN	1,5000	28	0	44	0	100,00	100,00	100,00	1	2	3	12	7	58,33	22,22	80,00	0,5111
2	PNN	1,6000	28	0	44	0	100,00	100,00	100,00	1	2	3	12	7	58,33	22,22	80,00	0,5111
2	PNN	1,7000	28	0	44	0	100,00	100,00	100,00	1	2	3	12	7	58,33	22,22	80,00	0,5111
2	PNN	1,8000	28	0	44	0	100,00	100,00	100,00	1	2	3	12	7	58,33	22,22	80,00	0,5111
2	PNN	1,9000	28	0	44	0	100,00	100,00	100,00	1	2	3	12	7	58,33	22,22	80,00	0,5111
2	PNN	2,0000	28	0	44	0	100,00	100,00	100,00	1	2	3	12	7	58,33	22,22	80,00	0,5111
2	PNN	2,1000	28	1	43	0	98,61	100,00	97,73	0,9886	2	3	12	7	58,33	22,22	80,00	0,5111
2	PNN	2,2000	28	1	43	0	98,61	100,00	97,73	0,9886	2	3	12	7	58,33	22,22	80,00	0,5111
2	PNN	2,3000	28	1	43	0	98,61	100,00	97,73	0,9886	3	3	12	6	62,50	33,33	80,00	0,5667
2	PNN	2,4000	27	1	43	1	97,22	96,43	97,73	0,9708	3	3	12	6	62,50	33,33	80,00	0,5667
2	PNN	2,5000	27	1	43	1	97,22	96,43	97,73	0,9708	4	3	12	5	66,67	44,44	80,00	0,6222
2	PNN	2,6000	26	1	43	2	95,83	92,86	97,73	0,9529	4	3	12	5	66,67	44,44	80,00	0,6222
2	PNN	2,7000	26	2	42	2	94,44	92,86	95,45	0,9416	4	3	12	5	66,67	44,44	80,00	0,6222
2	PNN	2,8000	26	2	42	2	94,44	92,86	95,45	0,9416	3	3	12	6	62,50	33,33	80,00	0,5667
2	PNN	2,9000	25	2	42	3	93,06	89,29	95,45	0,9237	3	3	12	6	62,50	33,33	80,00	0,5667
2	PNN	3,0000	24	2	42	4	91,67	85,71	95,45	0,9058	3	3	12	6	62,50	33,33	80,00	0,5667

Abreviaturas: PNN: Red Neuronal Probabilística, RNA: Red Neuronal Artificial, AUC: Área bajo la curva ROC